



STRUCTURAL AND DYNAMIC ASPECTS OF PROTEIN FUNCTION AND ALLOSTERY

EDITED BY: George Lisi, Ivan Rivalta and Vincenzo Venditti
PUBLISHED IN: Frontiers in Molecular Biosciences



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88974-821-1

DOI 10.3389/978-2-88974-821-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

STRUCTURAL AND DYNAMIC ASPECTS OF PROTEIN FUNCTION AND ALLOSTERY

Topic Editors:

George Lisi, Brown University, United States

Ivan Rivalta, University of Bologna, Italy

Vincenzo Venditti, Iowa State University, United States

Citation: Lisi, G., Rivalta, I., Venditti, V., eds. (2022). Structural and Dynamic Aspects of Protein Function and Allostery. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-821-1

Table of Contents

- 05 Editorial: Structural and Dynamic Aspects of Protein Function and Allostery**
George P. Lisi, Ivan Rivalta and Vincenzo Venditti
- 07 The *Vibrio cholerae* SpeG Spermidine/Spermine N-Acetyltransferase Allosteric Loop and $\beta 6$ - $\beta 7$ Structural Elements are Critical for Kinetic Activity**
Van Thi Bich Le, Sofiya Tsimbalyuk, Ee Qi Lim, Allan Solis, Darwin Gawat, Paloma Boeck, Ee Qing Lim, Rosselini Renolo, Jade K. Forwood and Misty L. Kuhn
- 23 Modulating Enzyme Function via Dynamic Allostery Within Biliverdin Reductase B**
Jasmina S. Redzic, Michael R. Duff, Ashley Blue, Todd M. Pitts, Pratul Agarwal and Elan Zohar Eisenmesser
- 37 Substitution of a Surface-Exposed Residue Involved in an Allosteric Network Enhances Tryptophan Synthase Function in Cells**
Rebecca N. D'Amico, Yuliana K. Bosken, Kathleen F. O'Rourke, Alec M. Murray, Woudasie Admasu, Chia-en A. Chang and David D. Boehr
- 51 A Single Point Mutation Controls the Rate of Interconversion Between the g^+ and g^- Rotamers of the Histidine 189 χ^2 Angle That Activates Bacterial Enzyme I for Catalysis**
Jeffrey A. Purslow, Jolene N. Thimmesch, Valeria Sivo, Trang T. Nguyen, Balabhadra Khatiwada, Rochelle R. Dotas and Vincenzo Venditti
- 63 Understanding the Binding Transition State After the Conformational Selection Step: The Second Half of the Molecular Recognition Process Between NS1 of the 1918 Influenza Virus and Host p85 β**
Alyssa Dubrow, Iktae Kim, Elias Topo and Jae-Hyun Cho
- 71 Reactivity of Thiol-Rich Zn Sites in Diacylglycerol-Sensing PKC C1 Domain Probed by NMR Spectroscopy**
Taylor R. Cole and Tatyana I. Igumenova
- 82 Cholesterol Is a Dose-Dependent Positive Allosteric Modulator of CCR3 Ligand Affinity and G Protein Coupling**
Evan van Aalst and Benjamin J. Wylie
- 96 Ligand Binding Introduces Significant Allosteric Shifts in the Locations of Protein Fluctuations**
Ambuj Kumar and Robert L. Jernigan
- 104 Dissecting Monomer-Dimer Equilibrium of an RNase P Protein Provides Insight Into the Synergistic Flexibility of 5' Leader Pre-tRNA Recognition**
Danyun Zeng, Ainur Abzhanova, Benjamin P. Brown and Nicholas J. Reiter
- 116 The Role of Substrate Mediated Allostery in the Catalytic Competency of the Bacterial Oligosaccharyltransferase PglB**
Brittany R. Morgan and Francesca Massi

- 134** *Structural Characterization of the RNA-Binding Protein SERBP1 Reveals Intrinsic Disorder and Atypical RNA Binding Modes*
Antoine Baudin, Alma K. Moreno-Romero, Xiaoping Xu, Emily E. Selig, Luiz O. F. Penalva and David S. Libich
- 149** *A Conserved Structural Role for the Walker-A Lysine in P-Loop Containing Kinases*
Fatlum Hajredini and Ranajeet Ghose
- 158** *Mapping Function From Dynamics: Future Challenges for Network-Based Models of Protein Structures*
Lorenza Pacini, Rodrigo Dorantes-Gilardi, Laurent Vuillon and Claire Lesieur
- 164** *Molecular Level Insights Into the Structural and Dynamic Factors Driving Cytokine Function*
Jennifer Y. Cui and George P. Lisi
- 173** *Elucidating the Activation Mechanism of AMPK by Direct Pan-Activator PF-739*
Elnaz Aledavood, Aria Gheeraert, Alessia Forte, Laurent Vuillon, Ivan Rivalta, F. Javier Luque and Carolina Estarellas
- 189** *Improving Small-Molecule Force Field Parameters in Ligand Binding Studies*
Stefano Raniolo and Vittorio Limongelli
- 199** *A Cysteine Variant at an Allosteric Site Alters MIF Dynamics and Biological Function in Homo- and Heterotrimeric Assemblies*
Erin Skeens, Georgios Pantouris, Dilip Shah, Ramu Manjula, Michael J. Ombrello, N. Karl Maluf, Vineet Bhandari, George P. Lisi and Elias J. Lolis



Editorial: Structural and Dynamic Aspects of Protein Function and Allostery

George P. Lisi^{1*}, Ivan Rivalta^{2,3*} and Vincenzo Venditti^{4,5*}

¹Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, United States, ²Dipartimento di Chimica Industriale "Toso Montanari," Alma Mater Studiorum, Università di Bologna, Bologna, Italy, ³Université de Lyon, École Normale Supérieure de Lyon, Université Lyon 1, CNRS UMR 5182, Laboratoire de Chimie, Lyon, France, ⁴Department of Chemistry, Iowa State University, Ames, IA, United States, ⁵Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA, United States

Keywords: protein dynamics, spectroscopy, simulations, allostery, enzymes

Editorial on the Research Topic

Structural and Dynamic Aspects of Protein Function and Allostery

Chemical signals are routinely propagated through biomacromolecules to modulate the structure of active sites and protein-protein interfaces. However, our understanding of the fundamental mechanisms coupling disparate regions of proteins is limited by the lack of accurate, atomic resolution information on their intrinsic dynamics. Indeed, the pathways most critical to chemical information flow are often intimately linked with the conformational ensembles populated by biomolecules, and spatially distant from traditional catalytic or ligand binding sites. A "holy grail" of biophysical chemistry has been to understand, at the molecular level, how ligand binding information is transmitted through a protein matrix to induce a functional response. Though a majority of information about dynamically-driven biological function comes from lower molecular weight proteins, an exploding number of studies have taken advantage of advances in spectroscopy, electron microscopy, and molecular simulations to synergistically map dynamic pathways that underlie long-range communication in multidomain systems.

The advantage of visualizing the solution ensembles of large biomolecules lies in the potential to leverage flexible hotspots for drug discovery or *de novo* spatial and temporal regulation of protein function. The articles in this Research Topic tackle this knowledge gap by reporting on the structural and dynamic components that govern intra- and inter-domain crosstalk.

Redzic et al. demonstrated an intricate link between micro-millisecond protein dynamics and allostery in biliverdin reductase β (BLVRB). Strikingly, evolutionary differences in amino acid sequence distal to the catalytic site induce a substantial variability in molecular motions that work in concert to regulate BLVRB function. A thermodynamic and kinetic investigation by Dubrow et al. offers insight into the role of protein motions in organizing biomolecular interfaces. Through site-directed mutagenesis, this work captures the per-residue impact on conformational selection during the binding transition state of influenza A nonstructural protein 1 and human p85 β . The allosteric interactions of cholesterol with the chemokine receptor CCR3, critical to immune cell trafficking, is characterized by van Aalst and Wylie with circular dichroism and fluorescence polarization. The authors identified cholesterol as a critical mediator of substrate binding and receptor activation that drives signal transduction in GTPase assays. Cole and Igumenova used NMR spectroscopy to explore the interplay between Cd²⁺, Zn²⁺ and the conserved homology 1 (C1) zinc finger domain of protein kinase C, revealing the atomistic structural and thermodynamic properties of the C1 coordination sphere that facilitate competition between divalent metals.

OPEN ACCESS

Edited and reviewed by:

Ralf Metzler,
University of Potsdam, Germany

*Correspondence:

George P. Lisi
george_lisi@brown.edu
Ivan Rivalta
i.rivalta@unibo.it
Vincenzo Venditti
venditti@iastate.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 15 February 2022

Accepted: 18 February 2022

Published: 10 March 2022

Citation:

Lisi GP, Rivalta I and Venditti V (2022)
Editorial: Structural and Dynamic
Aspects of Protein Function
and Allostery.
Front. Mol. Biosci. 9:876499.
doi: 10.3389/fmolb.2022.876499

In another study, Purslow et al. highlighted the influence of protein dynamics on the activity of bacterial phosphotransferase Enzyme I by dissecting the contributions of active site flexibility to enzymatic turnover, most notably through the rotameric equilibrium of the catalytic His. Another elegant NMR relaxation study, performed by Zeng et al., revealed an RNA-driven disorder-to-order transition in the N-terminus of a bacterial RNase P, which serves as a dynamic checkpoint to ensure substrate alignment and enzyme activation. Baudin et al. reported an NMR structural study of the SERPINE1 mRNA binding protein (SERBP1), revealing SERBP1 to be intrinsically disordered but capable of sampling several compact conformations. The authors further define its RNA binding preferences and propensity for liquid-liquid phase separation, providing seminal molecular details of its mechanism.

Le et al. carried out a novel structural study of the SpeG N-acetyltransferase, defining the structural basis for an allosteric mechanism that is unique within this enzyme family. This work implicates a dynamic loop, along with several β -strands, as mediators of enzyme activity and lays the foundation for expanded structure-function studies of SpeG. Long-range allostery in α -tryptophan synthase (α TS) is shown by D'Amico et al. to involve networks of flexible residues that propagate ~ 25 Å chemical signals. Here, a novel role for surface-exposed residues in modulating dynamic crosstalk in α TS is revealed by NMR and molecular simulations. Skeens et al. and Cui and Lisi explored the intrinsic dynamics of cytokines as a driver for promiscuous and non-overlapping functions. Site-directed mutagenesis, novel structural engineering, and receptor binding demonstrated an intimate link between multi-timescale conformational dynamics and several biological activities.

The computational works contributing to this Research Topic covered various flavors and challenges of modern atomistic simulations, demonstrating the benefits of obtaining information with atomistic resolution that can be directly compared with experimental evidence and highlighting the potential of *in silico* predictions. For instance, Raniolo and Limongelli combined quantum-mechanics and free-energy calculations to improve standard ligand parametrization, allowing enhanced sampling (Funnel-Metadynamics) simulations of the paradigmatic benzamidine/trypsin molecular binding system that elegantly reproduced the high-resolution crystallographic ligand binding mode, providing a very accurate description of the binding mechanism. Hajredini and Ghose, instead, showed how enhanced sampling MD simulations could provide insight into the structural mediating role of a conserved “catalytic” residue that inactivates two distantly related kinase families, i.e., bacterial tyrosine and shikimate kinases.

Estarellas et al. computationally assessed changes in the structural and dynamical properties of distinct isoforms of the adenosine monophosphate-activated protein kinase complexes. The comprehensive analysis of molecular dynamics simulations,

also involving network theory tools, enabled characterization of key molecular factors that mediated activation of pan-activator PF-739, identifying distinctive features that correlate with the affinities of different isoforms. Massi and Morgan also combined molecular dynamics simulations with network analysis to show that substrate specificity in the enzymatic activity of the oligosaccharyltransferase of *Campylobacter lari* is regulated by modulation of dynamic allosteric pathways. Finally, Pacini et al. provided a perspective on the future challenges of network theory calculations aimed at elucidating the link between the information encoded in protein primary sequences, their dynamics and functions. Jernigan and Kumar proposed, instead, the use of elastic network models to infer the dynamics of a variety of proteins (with known bound and unbound structures) by observing the transfer of fluctuations among distant regions upon binding of an allosteric ligand.

Overall, the collection of manuscripts selected for this special issue of *Frontiers in Molecular Biosciences* highlights once more the power of combining computational and experimental approaches to characterize protein allostery with atomic resolution. We expect that further strengthening and exploiting such synergy will be instrumental toward a complete dissection of the structure/dynamics/function relationship in proteins and the development of predictive tools to identify allosteric networks and hot spots for drug design.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

GPL acknowledges funds from National Institutes of Health Grants P20 GM109035 and R01 GM136815 and National Science Foundation Grant MCB 2143760. VV acknowledges funds from National Institutes of Health Grant R35 GM133488.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Lisi, Rivalta and Venditti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The *Vibrio cholerae* SpeG Spermidine/Spermine N-Acetyltransferase Allosteric Loop and $\beta 6$ - $\beta 7$ Structural Elements Are Critical for Kinetic Activity

Van Thi Bich Le¹, Sofiya Tsimbalyuk², Ee Qi Lim¹, Allan Solis¹, Darwin Gawat¹, Paloma Boeck¹, Ee Qing Lim¹, Rosselini Renolo¹, Jade K. Forwood² and Misty L. Kuhn^{1*}

¹Department of Chemistry & Biochemistry, San Francisco State University, San Francisco, CA, United States, ²School of Biomedical Science, Charles Sturt University, Wagga Wagga, NSW, Australia

OPEN ACCESS

Edited by:

George Lisi,
Brown University, United States

Reviewed by:

Andrey Cherstvy,
University of Potsdam, Germany
Qin Xu,
Shanghai Jiao Tong University, China

*Correspondence:

Misty L. Kuhn
mkuhn@sfsu.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 23 December 2020

Accepted: 16 February 2021

Published: 13 April 2021

Citation:

Le VTB, Tsimbalyuk S, Lim EQ, Solis A, Gawat D, Boeck P, Lim EQ, Renolo R, Forwood JK and Kuhn ML (2021) The *Vibrio cholerae* SpeG Spermidine/Spermine N-Acetyltransferase Allosteric Loop and $\beta 6$ - $\beta 7$ Structural Elements Are Critical for Kinetic Activity. *Front. Mol. Biosci.* 8:645768. doi: 10.3389/fmolb.2021.645768

Polyamines regulate many important biological processes including gene expression, intracellular signaling, and biofilm formation. Their intracellular concentrations are tightly regulated by polyamine transport systems and biosynthetic and catabolic pathways. Spermidine/spermine N-acetyltransferases (SSATs) are catabolic enzymes that acetylate polyamines and are critical for maintaining intracellular polyamine homeostasis. These enzymes belong to the Gcn5-related N-acetyltransferase (GNAT) superfamily and adopt a highly conserved fold found across all kingdoms of life. SpeG is an SSAT protein found in a variety of bacteria, including the human pathogen *Vibrio cholerae*. This protein adopts a dodecameric structure and contains an allosteric site, making it unique compared to other SSATs. Currently, we have a limited understanding of the critical structural components of this protein that are required for its allosteric behavior. Therefore, we explored the importance of two key regions of the SpeG protein on its kinetic activity. To achieve this, we created various constructs of the *V. cholerae* SpeG protein, including point mutations, a deletion, and chimeras with residues from the structurally distinct and non-allosteric human SSAT protein. We measured enzyme kinetic activity toward spermine for ten constructs and crystallized six of them. Ultimately, we identified specific portions of the allosteric loop and the $\beta 6$ - $\beta 7$ structural elements that were critical for enzyme kinetic activity. These results provide a framework for further study of the structure/function relationship of SpeG enzymes from other organisms and clues toward the structural evolution of members of the GNAT family across domains of life.

Keywords: Gcn5-related N-acetyltransferase, spermidine/spermine N-acetyltransferase, polyamine, allosteric regulation, acetylation, spermidine, spermine, chimeric

INTRODUCTION

Polyamines are organic molecules that are fully protonated at physiological pH and bind to negatively charged sites of DNA, RNA, proteins, and phospholipid membranes (Tabor and Tabor, 1984; Finger et al., 2014). Examples of polyamines include putrescine (put), cadaverine (cad), spermidine (spd), spermine (spm), and norspermidine (nspd), but a variety of other

polyamines exist (Hamana and Matsuzaki, 1992; Michael, 2016). They play many important roles in cellular processes, including the regulation of gene expression (Childs et al., 2003), intracellular signaling (Karatan et al., 2005; Sobe et al., 2017), virulence (Jelsbak et al., 2012), and biofilm formation (Karatan et al., 2005; Nesse et al., 2015; Sobe et al., 2017). Furthermore, polyamines are present in all domains of life and some viruses (Michael, 2016; Firpo and Mounce, 2020); however, some organisms within these groups do not contain any polyamines (Hamana and Matsuzaki, 1992; Pegg and Casero, 2011). The most common polyamines in bacteria are spd and put, but others such as nspd and spm can also be present (Hamana and Matsuzaki, 1992; Michael, 2016). In general, spd is found in high concentrations in bacteria, but some bacteria, including *Escherichia coli*, have a higher prevalence of put than spd. While most bacteria contain both of these polyamines, some bacteria either do not have spd or put, or they have different polyamines. For example, Vibrionaceae family members, including *Vibrio*, *Photobacterium*, and *Listonella* species, have nspd as the predominant polyamine (Hamana and Matsuzaki, 1992).

The concentrations of polyamines in bacteria are typically in the millimolar range (from 0.1 to 30 mM) (Shah and Swiatlo, 2008) and are tightly regulated by a variety of means, including polyamine import/export systems, and intracellular anabolic and catabolic processes (Casero and Pegg, 1993; Wallace et al., 2003; Pegg and Casero, 2011; Schneider and Wendisch, 2011). Catabolic enzymes that are critical for regulating polyamine concentrations include the spermidine/spermine N-acetyltransferases (SSATs). These enzymes acetylate polyamines using acetyl coenzyme A (AcCoA) as an acetyl donor (Zhu et al., 2006; Hegde et al., 2007; Filippova et al., 2015a; Tsimbalyuk et al., 2020). Once acetylated, the overall charge of the polyamine is reduced and the molecules are exported (Kramer et al., 2008) or recycled (Floris and Finazzi Agrò, 2013). Several different SSATs have been investigated, including hSSAT1 (*Homo sapiens*) (Zhu et al., 2006), PaiA (*Thermoplasma acidophilum*, *Bacillus subtilis*) (Forouhar et al., 2005; Filippova et al., 2011), BldD (*Bacillus subtilis*) (Woolridge et al., 1999), and SpeG (*Vibrio cholerae*, *E. coli*, *Coxiella burnetii*, *Staphylococcus aureus*, and *Bacillus thuringiensis*) (Fukuchi et al., 1994; Filippova et al., 2015a; Franklin et al., 2015; Li et al., 2019; Tsimbalyuk et al., 2020). All SSATs that have been kinetically characterized have been shown to acetylate spm more readily or at least as well as other polyamines. For quite some time it was thought that spm was not produced by bacteria and was instead imported into bacterial cells because they lack a spermine synthase (Ikeguchi et al., 2006). However, it has been shown that spm can indeed be produced in some bacteria, including some species of Actinobacteria, Clostridiales, and Bacilliales (Pegg and Michael, 2010; Michael, 2016), using a different mechanism that involves a carboxyspermidine dehydrogenase and carboxyspermidine decarboxylase (Kim et al., 2016).

While SSAT enzymes catalyze the same reaction and exhibit relatively similar substrate specificity, their 3D structures can differ quite significantly, but all retain a core Gcn5-related

N-acetyltransferase (GNAT) structural fold. SpeG from *Vibrio cholerae* (VcSpeG) was the first GNAT enzyme shown to exhibit a dodecameric structure, which is unusual for the GNAT family of enzymes since they are predominantly monomers or lower ordered oligomers. VcSpeG is also the first SSAT enzyme to exhibit an allosteric site where polyamines can bind (Filippova et al., 2015a). The allosteric site is located between monomers that form the hexamers of the dodecamer (Filippova et al., 2015a), and some evidence has been provided that shows when polyamine binds to these sites, the dodecamer tightens (Filippova et al., 2015b). When apo- and polyamine-bound structures of VcSpeG are compared, there is a large conformational change of the allosteric loop that occurs (Filippova et al., 2015a).

To learn more about the importance of the allosteric loop and other regions of VcSpeG on its kinetic activity, we designed point mutants and a series of chimeric constructs between the structurally divergent hSSAT1 and VcSpeG proteins. These specific enzymes were selected because they were the most extreme cases of structural diversity of SSATs that have been characterized (Figure 1). Moreover, the importance of the allosteric loop of SpeG proteins has not been kinetically elucidated. Here, we identified specific residues of the allosteric loop and the $\beta 6$ - $\beta 7$ region of the VcSpeG protein that are required for its enzymatic activity, which provides new insight into the structure/function relationship of SSAT proteins.

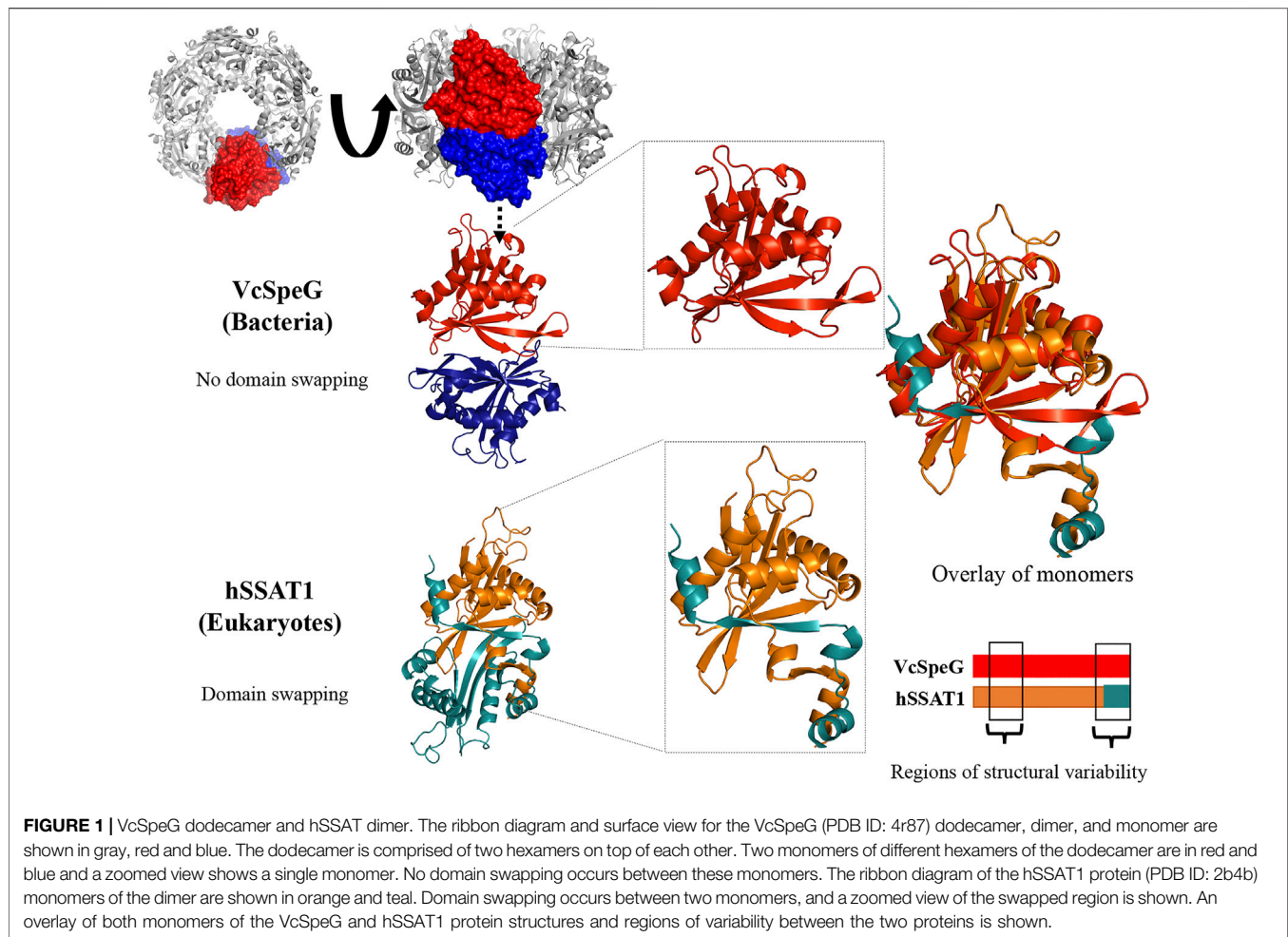
MATERIALS AND METHODS

Materials

All chemicals were purchased at the highest quality from either Millipore Sigma or Thermo Fisher Scientific. Spermine dihydrate and acetyl coenzyme A trilithium salts were purchased from Millipore Sigma.

Cloning

Different constructs (outlined in Figure 2, Supplementary Figure S1 and Table S1) were made using the Q5® site-directed mutagenesis kit (New England Biolabs, Inc.) and the VcSpeG plasmid template (*V. cholerae* O1 biovar El Tor str. N16961 speG (VC_A0947) gene, Uniprot ID: Q9KL03) described in (Majorek et al., 2014; Filippova et al., 2015a). Forward and reverse primers (Supplementary Table S1) were designed using the NEBasechanger tool (<https://nebasechanger.neb.com/>) and purchased from Integrated DNA Technologies, Inc. A 50 μ L Polymerase Chain Reaction (PCR) (1 ng VcSpeG template DNA; 0.25 μ M forward and reverse primers; and 1X master mix) was performed using an Axygen® MaxyGene II Thermal Cycler. Denaturation, annealing and extension times and temperatures were selected based on the manufacturer's recommendations and the NEBasechanger tool output for each set of primers for individual constructs. The size of PCR products was confirmed on a 1% agarose gel and then samples were subjected to the kinase, ligase, and DpnI reaction (1 μ L PCR product, 1X KLD buffer, and 1X KLD enzyme in a final volume of 10 μ L) for 5 min at RT. Then 0.5 μ L of this product was used to transform 100 μ L of DH5 α competent cells that were previously



prepared using the Zymo Research Mix & Go transformation kit. We spread 20 μ L of the transformation mixture and 20 μ L of SOC media onto an LB-agar plate containing 100 μ g/ml ampicillin. The plate was incubated at 37°C overnight. To confirm correct DNA sequences, single colonies were sent for direct colony sequencing (Genewiz). Correct clones were then purified and transformed into *E. coli* BL21 (DE3) competent cells using a similar transformation procedure as described above. Glycerol stocks were made and stored at –80°C until ready to use.

Heterologous Protein Expression

The BL21 (DE3) glycerol stock of each construct was used to inoculate a 5 ml culture of 1X LB media with 100 μ g/ml ampicillin in a 15 ml sterile test tube. The culture was grown at 37°C overnight with shaking in a GeneMate Mini shaker at 200 rpm. The next day, 2 ml of the overnight culture was added into a baffled 1 L flask containing 200 ml of 1X Terrific Broth with 100 μ g/ml ampicillin. Cells were grown at 37°C with shaking in a Thermo Scientific MaxQ 4450 shaker at 200 rpm until the OD_{630nm} reached 0.6–0.8. The flask was then placed on ice until cool and 0.5 mM isopropyl β -D-thiogalactopyranoside (IPTG) was added to induce protein expression. The culture was then shaken on a Thermo Scientific MaxA 2000 shaker at

150 rpm at RT for 16 h. Cells were harvested by centrifugation at 4500 rpm in a Thermo Scientific ST 8R centrifuge at 4°C for 30 min. The supernatant was discarded and the pellet was resuspended in 30 ml of lysis buffer (0.01 M Tris HCl pH 8.3, 0.5 M NaCl, 5 mM imidazole, 5% glycerol, 5 mM beta-mercaptoethanol (BME)). This mixture was kept on ice and sonicated using a Branson 450 Digital Sonifier. The program was set for a total of 5 min of sonication with 10 s on and 10 s off at 10% amplitude. Afterwards, the crude extracts were stored at –80°C.

Protein Purification

Proteins were purified using nickel affinity chromatography via the following procedure. Crude extracts were thawed and centrifuged at 15,000 rpm in a Sorvall SS-34 rotor at 4°C for 50 min. The ÄKTA™ Start FPLC was equilibrated with 10 CVs of buffer A (10 mM Tris HCl pH 8.3, 500 mM NaCl, 5 mM BME) and the supernatant was loaded onto a 1 ml HisTrap FF column using the sample pump. The column was washed with 5 CVs of buffer A with 25 mM imidazole to remove non-specific binding proteins. The tagged protein was eluted with 5 CVs of buffer B (10 mM Tris HCl pH 8.3, 500 mM NaCl, 5 mM BME, 500 mM imidazole) and then desalted using a PD-10 column (GE

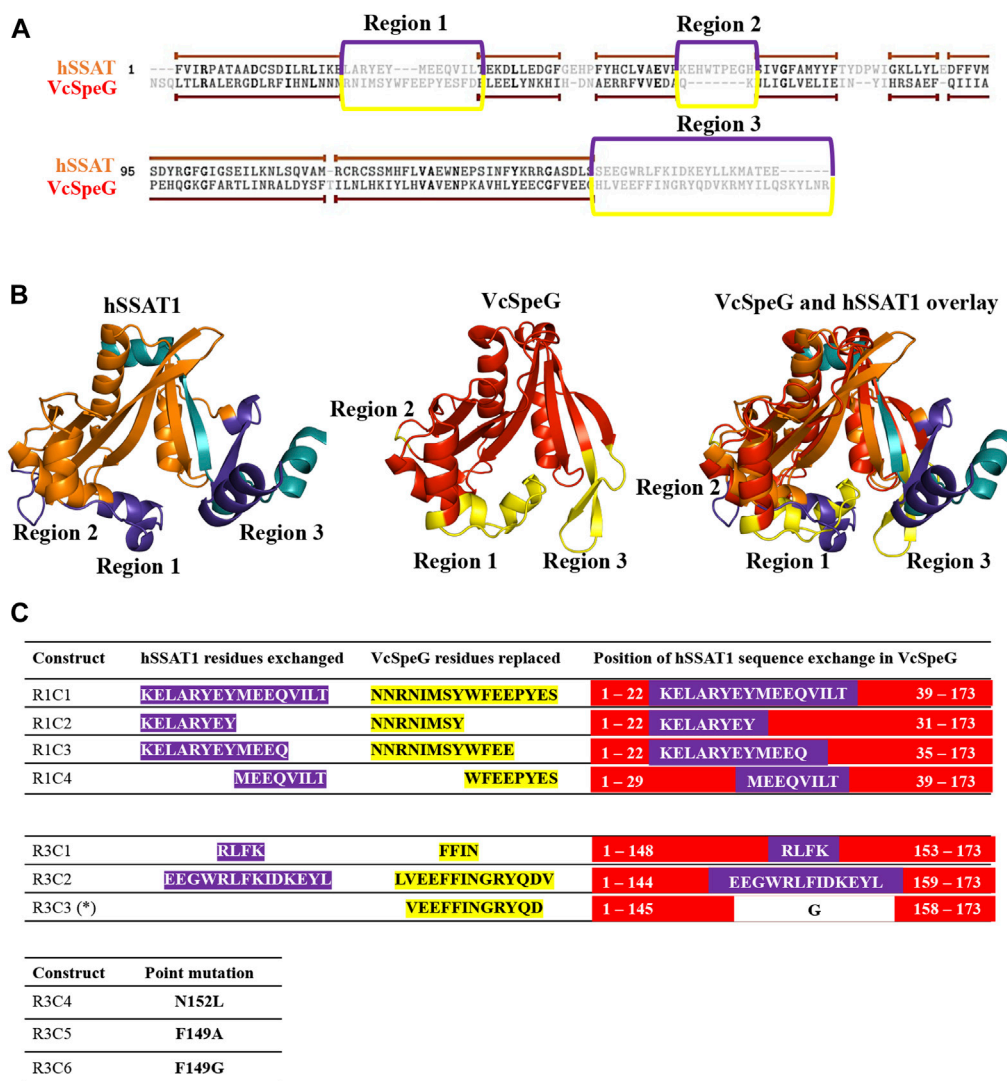


FIGURE 2 | Comparison of sequences and structures of VcSpeG and hSSAT1 and defining regions and constructs for this study. **A.** Dissimilar regions between VcSpeG (PDB ID: 4r87) and hSSAT1 (PDB ID: 2b4b) structures shown on a primary sequence alignment generated using TopMatch. Gray residues framed in purple (hSSAT1) and yellow (VcSpeG) show unaligned regions of the structures during structural alignment. We named these regions of structural variability in sequential order. **B.** Three different regions highlighted on the monomer VcSpeG and hSSAT1 structures. All three regions are shown in purple on the hSSAT1 structure (orange); the teal secondary structure shown is from the adjacent monomer of the dimer. Only the complete monomer is shown for clarity. Similarly, all three regions of the VcSpeG protein monomer (red) are highlighted in yellow. **C.** Construct names and residues substituted or altered in the ten VcSpeG constructs for this study. Residues highlighted in yellow are removed from VcSpeG and replaced by hSSAT1 residues in purple. The position of hSSAT1 exchange on each VcSpeG cartoon linear protein sequence is also shown. An asterisk indicates the deletion construct.

Healthcare) and buffer A. Next, Tobacco Etch Virus (TEV) protease (purified as described previously (Czub et al., 2018)) was used to cleave the polyhistidine tag from the protein construct. A total of 1 ml of sample containing buffer A, the tagged protein and TEV (mass ratio was 20:1 for tagged protein: TEV) was added into the top cup of a Slide-a-Lyzer mini dialysis tube (Thermo Fisher Scientific) and 14 ml of cleavage buffer (50 mM Tris HCl pH 8.3, 300 mM NaCl, 0.5 mM EDTA, 5% glycerol, 2.5 mM DTT) was added to the bottom of the dialysis tube. This tube was placed on an oscillator at 37°C for 2 h and the buffer in the bottom of the tube was exchanged and placed at 4°C overnight. The protein was centrifuged at 4500 rpm at 4°C for

10 min to remove precipitate. The supernatant was combined with buffer A to a total volume of 3 ml to dilute the DTT and then loaded onto a 1 ml HisTrap FF column using a 3 ml injection loop. The cleaved protein was collected as two different peaks: one was the eluate after 5 CVs of buffer A was flowed over the column, and the second was during a 0–30% gradient of buffer B over 10 CVs. Proteins were desalted into buffer A without BME and concentrated using Amicon 10kD MWCO centrifugation tubes (Millipore). Protein from the second peak was used for all kinetic assays. Absorbances of proteins at $A_{280\text{nm}}$ were measured using a Thermo Scientific Nanodrop Spectrophotometer and protein concentrations were calculated using the monomer

molecular weight of 20,675 Da, and an extinction coefficient of $20,400 \text{ M}^{-1}\text{cm}^{-1}$. The purity of the proteins and cleavage of the tag were confirmed by SDS-PAGE and proteins were stored at -80°C . All constructs were soluble and were purified to near homogeneity (**Supplementary Figure S2**).

Steady-State Enzyme Kinetics

All constructs were screened for enzymatic activity toward spermine using a previously described colorimetric assay (Kuhn et al., 2013) with the following modifications. Each 50 μL reaction contained 70 mM bicine pH 8.0, 20 mM NaCl, 0.5 mM AcCoA, 3 mM spm, and 10 μL of enzyme (0.1 $\mu\text{g}/\mu\text{L}$) was used to initiate the reaction in a 96 well clear polystyrene microplate. The reaction proceeded at 37°C for 5 min and 50 μL of solution (6 M guanidine HCl and 0.1 M Tris HCl pH 8.0) was added to terminate the reaction. Next, 200 μL of Ellman's reagent (0.2 mM 5,5-dithio-bis-(2-nitrobenzoic acid) (DNTB), 0.1 M Tris HCl pH 8.0, 1 mM EDTA) was added and incubated for 10 min at RT. The amount of CoA produced from each reaction was determined spectrophotometrically by measuring the absorbance of the samples at 415 nm and standard solutions of cysteine. Substrate saturation curves were performed on all constructs that exhibited activity during the screening assay, by varying spm concentrations (0–3 mM) and holding AcCoA concentration constant (0.5 mM). Protein concentrations used for substrate saturation curves were: WT (0.07 μM), R1C1 (1.2 μM), R1C2 (0.14 μM), R1C3 (3.6 μM), R3C1 (2.5 μM), R3C2 (17 μM), R3C4 (0.1 μM), R3C5 (0.1 μM), and R3C6 (0.2 μM). During screening, R3C3 and R1C4 had very low to no activity above baseline, so we did not perform a substrate saturation curve for these constructs. Data were collected for at least two biological replicates for each construct, except R1C3 and were fitted to the Hill equation using Prism 9.0.0. The $S_{0.5}$ is the concentration of substrate at half the maximal velocity.

Crystallization

Proteins were expressed and purified using procedures previously described in (Tsimbalyuk et al., 2020). All crystallization trials were performed using the hanging drop vapor diffusion method in 48-well plates (Hampton Research). Each coverslip contained 1.5 μL of protein and 1.5 μL of reservoir solution suspended over 300 μL of reservoir solution. A detailed list of final crystallization conditions used are outlined in **Supplementary Table S2**. Crystals were cryoprotected in reservoir solution with 20% glycerol before being flash-cooled at 100 K in liquid nitrogen.

Data Collection, Data Processing and Structure Determination

X-ray diffraction data were collected using the MX1 crystallography beamline (Eiger 2 9M detector) at the Australian Synchrotron (McPhillips et al., 2002; Cowieson et al., 2015). The data were indexed and integrated in iMosflm (Battye et al., 2011), merged and scaled using AIMLESS (Evans and Murshudov, 2013). Phaser (McCoy, 2007) was used for phase determination and molecular replacement, and the *V. cholerae*

SpeG protein structure (PDB ID: 4mi4) was used as a search model. The model was rebuilt in Coot (Emsley et al., 2010), and refined using Refmac5 (Murshudov et al., 2011; Kovalevskiy et al., 2018) and Phenix (Adams et al., 2010; Afonine et al., 2012). Structures were validated using Phenix. R1C2, R1C3, R3C1, R3C4, R3C5 and R3C6 structures were deposited into the Protein Data Bank with accession codes 7kwh, 7kwj, 7kwq, 7kwx, 7kx2, and 7kx3, respectively.

Homology Modeling

Two homology models for each construct were built using the Swiss-Model server (<https://swissmodel.expasy.org/interactive#structure>) (Waterhouse et al., 2018). One model was based on the apo- WT VcSpeG structure PDB ID: 4jjx and the other model was based on the spm-bound VcSpeG structure PDB ID: 4mi4. The quality of each model was evaluated using QMEAN (Qualitative Model Energy Analysis) and GMQE (Global Model Quality Estimation) scores (**Supplementary Table S3**).

RESULTS

Defining Key Regions of VcSpeG for Investigation

The VcSpeG and hSSAT1 proteins both acetylate spm and spd, but they have different quaternary structures, i.e. VcSpeG is a dodecamer (Filippova et al., 2015a), whereas hSSAT1 is a dimer (Zhu et al., 2006). The hSSAT1 protein is a domain-swapped dimer and the active site contains residues from both protomers (Zhu et al., 2006), but the dimer within the VcSpeG dodecamer does not exhibit swapping and the active site is contained within each monomer (Filippova et al., 2015a) (**Figure 1**). Unlike VcSpeG, the hSSAT1 protein does not exhibit an allosteric polyamine binding site.

This allosteric site in VcSpeG is located between adjacent monomers of the hexamer and interacts with spm using residues N22, E33, E34, Y36, E37, E41 of one monomer and H49, I50, D52, and E55 of an adjacent monomer; there is a loop in the allosteric site (residues 23–38) that transitions to a helix upon binding spm (Filippova et al., 2015a). While the hSSAT1 and VcSpeG enzymes have low sequence identity (21%), they exhibit significant structural similarity, with the exception of three main regions: Region 1 is the VcSpeG allosteric loop ($\alpha 1$ – $\alpha 2$). Region two is an insertion between $\beta 2$ and $\beta 3$ in the hSSAT1 protein that is not present in the VcSpeG enzyme, and Region 3 is the C-terminal portion of the protein ($\beta 6$ – $\beta 7$ in VcSpeG and $\beta 7$ – $\alpha 6$ in hSSAT1) that interacts with the allosteric loop of VcSpeG (**Figures 2, 3**). Therefore, we investigated the contribution of the allosteric loop (Region 1) and portions of the $\beta 6$ – $\beta 7$ residues (Region 3) of VcSpeG on its kinetic activity. This was accomplished by substituting these portions of the VcSpeG protein with comparable residues of the hSSAT1 protein (i.e. constructing chimeric proteins); we also created selected point mutations of the VcSpeG enzyme (**Figure 2**). Our primary focus was on Regions 1 and 3 of the VcSpeG protein (**Figure 2**) since we hypothesized the combination

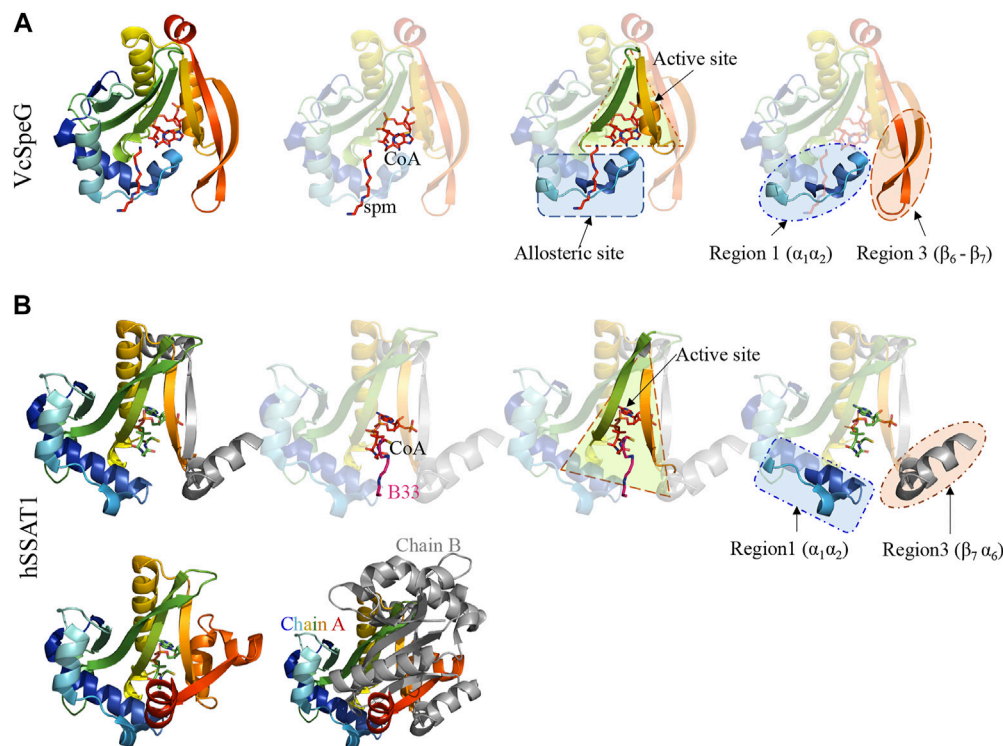


FIGURE 3 | Comparison of VcSpeG and hSSAT1 structures. Each structure is colored in rainbow from N-terminus to C-terminus from blue to red. **A.** A monomer of the VcSpeG PDB ID: 4r87 structure is shown as a ribbon diagram and ligands CoA and spm are shown as sticks. The allosteric site, active site, and Regions 1 and 3 of the protein structure are shown. Spm is bound to the allosteric site, whereas CoA is bound to the active site. **B.** A single monomer (chain A) of the hSSAT1 protein structure PDB ID: 2b4b is shown in rainbow colors and the second monomer (chain B) of the dimer is in gray. The second monomer (chain B) except the secondary structure that is swapped into the first monomer has been removed for clarity to show the location of the active site. The active site of this protein has CoA and the inhibitor $N^1 N^{11}$ -bis-(ethyl)-norspermine shown as sticks. The regions that correspond to Regions 1 and 3 of the VcSpeG protein structure are shown.

of these two regions would be important for forming the active conformation of VcSpeG.

Strategy for Generating VcSpeG Constructs

To investigate these previously unstudied regions of the VcSpeG protein, we used a structure-guided approach. Our strategy was to first select and test a subset of key point mutants and chimeric constructs. We created six chimeras (R1C1-R1C4, R3C1-R3C2), one construct with a deletion of a specific region of the protein (R3C3), and three single point mutants (R3C4-R3C6) (Figure 2). These constructs were named based on the region where the residue(s) are located in the VcSpeG protein and were numbered sequentially, e.g. R3C2 corresponds to a change in Region 3 and is the second construct made to query the importance of this region of the VcSpeG protein. A cartoon version of the protein sequences is shown in Figure 2 and full protein sequences of these ten constructs can be found in Supplementary Figure S1 and Table S1.

Chimeric Constructs in Region 1 of VcSpeG: The Importance of Specific Regions of the Allosteric Loop (R1C1-R1C4)

The allosteric loop of VcSpeG is sixteen residues in length. To determine which residues of the hSSAT1 protein should be

substituted for VcSpeG residues of this region (Region 1), we overlaid the spm-bound VcSpeG structure (PDB ID: 4r87) and the hSSAT1 structure (PDB ID: 2b4b) and performed a sequence alignment between these two proteins (Figure 2). Residues of the $\alpha 1$ - $\alpha 2$ helices of the hSSAT1 protein overlaid with the allosteric loop (Region 1) of the VcSpeG protein (Figure 3). Since this loop was relatively long, we created a series of four constructs to investigate the importance of this region on the activity of VcSpeG. R1C1 replaces all residues of this region with the hSSAT1 protein sequence, R1C2 replaces the first half of the residues, R1C3 replaces the first three quarters of the residues, and R1C4 replaces the latter half of the residues of this region (Figure 2).

Chimeric Constructs and Deletion in Region 3 of VcSpeG: The Importance of the Beta-Turn Between $\beta 6$ - $\beta 7$ (R3C1-R3C3)

The C-terminal portions of the protein sequences of VcSpeG and hSSAT1 adopt quite different conformations. This portion of the hSSAT1 protein is intercalated in the opposite monomer of the dimer (domain-swapped), whereas the entire sequence of the single polypeptide chain of VcSpeG is contained within a single monomer (no domain-swapping) (Figure 1). In the linear protein sequence, Region 3 of these proteins corresponds to $\beta 6$ - $\beta 7$ of

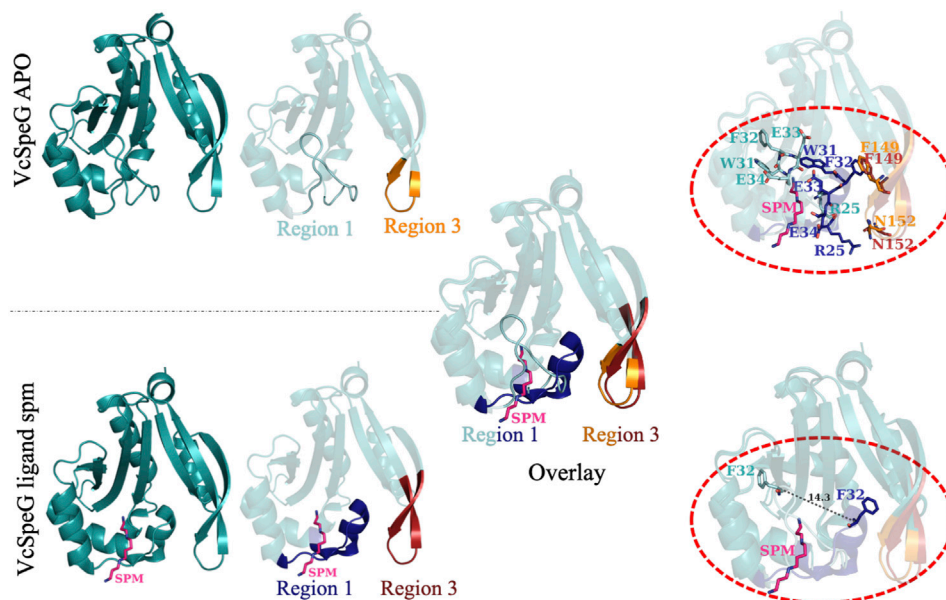


FIGURE 4 | Comparison of structural changes of the allosteric loop of the VcSpeG protein in absence and presence of spermine. The VcSpeG monomer in presence (PDB ID: 4mi4) and absence of spm (PDB ID: 4jjx) is shown in ribbon diagrams. The spm ligand is shown as pink sticks and nitrogens are in dark blue. Region 1 in the apo form is shown in cyan, but the same region in the spm-bound structure is shown in dark blue. Region 3 from the apo structure is in orange and is in red for the spm-bound structure. An overlay of these two structures is shown to indicate movement of Region 1 when spm is present. Residues that alter their conformation significantly upon this transition (R25, W31, F32, E33, E34, F149, and N152) are shown in the top dotted oval in sticks and are colored according to structures and regions from which they are found. The bottom dotted oval shows the distance between the C α of F32 in the apo- and spm-bound structures is 14.3 Å.

VcSpeG and $\beta 7$ - $\alpha 6$ of hSSAT1 (**Figure 3**). We created three constructs to investigate the importance of the $\beta 6$ - $\beta 7$ region of VcSpeG on its activity. The first construct (R3C1) removes the residues that form a portion of $\beta 7$ and a portion of the beta-turn of the $\beta 6$ - $\beta 7$ region of the protein and replaces them with the corresponding residues from hSSAT1 in its linear sequence (**Figure 2**). Since the N152 residue of VcSpeG is replaced with a positively charged lysine from hSSAT1 and it normally forms an H-bond with R25 of Region 1, it is possible this substitution would repel the interaction between these two regions. The second construct (R3C2) substitutes a larger portion of the $\beta 6$ - $\beta 7$ region of VcSpeG with the hSSAT1 sequence (**Figure 2**). This construct would enable us to determine if the protein can withstand large alterations in this region and still be active, and/or whether these types of alterations are sufficient to cause the VcSpeG protein to adopt domain-swapping like the hSSAT1 protein. The final construct (R3C3) is a deletion of the entire portion of the VcSpeG protein sequence that was substituted in the R3C2 construct and has glycine residue insertion (**Figure 2**).

Point Mutations in Region 3 of VcSpeG: The Importance of Residues That Potentially Stabilize Interactions Between Regions 1 and 3 (R3C4-R3C6)

When polyamine binds to the allosteric site of VcSpeG, a significant conformational change of the allosteric loop (Region 1; $\alpha 1$ - $\alpha 2$) occurs compared to the apo-structure, and Region 1 moves closer to Region 3 ($\beta 6$ - $\beta 7$) (**Figure 4**). Indeed, the distance between the alpha carbon of F32 on this allosteric loop in

the apo- (PDB ID: 4jjx) and the alpha carbon of F32 on the spm-bound (PDB ID: 4mi4) structures is 14.3 Å (**Figure 4**). Our analysis of the spm-bound structure of VcSpeG showed these two regions form potentially significant contacts: π - π interactions between F32 and F149 and an H-bond between R25 and N152. To elucidate whether these residues could be important for stabilizing the interaction between these two regions of VcSpeG, we created three point mutants: N152L, F149A, and F149G (R3C4-R3C6; **Figure 2**). We substituted leucine for asparagine at position 152 to retain the overall structural similarity of the residue but remove the ability to H-bond, and we substituted alanine and glycine for phenylalanine at position 149 to potentially decrease the ability of this residue to form hydrophobic interactions.

Crystallization Trials and Structure Determination

In order to confirm the structural integrity and assess the conformation of the chimeric and mutant proteins, we attempted to determine the structure of all ten constructs using X-ray crystallography. All ten constructs were solubly expressed in *E. coli* and purified using a two step purification method. Purified proteins were screened for conditions that induce crystallization, and a total of six crystal structures were determined: R1C2 (PDB ID: 7kwh), R1C3 (PDB ID: 7kwj), R3C1 (PDB ID: 7kwq), R3C4 (PDB ID: 7kwx), R3C5 (PDB ID: 7kx2), and R3C6 (PDB ID: 7kx3) at the resolution 2.9Å, 2.58Å, 2.3Å, 2.42Å, 2.6Å, and 2.67Å, respectively. Detailed data

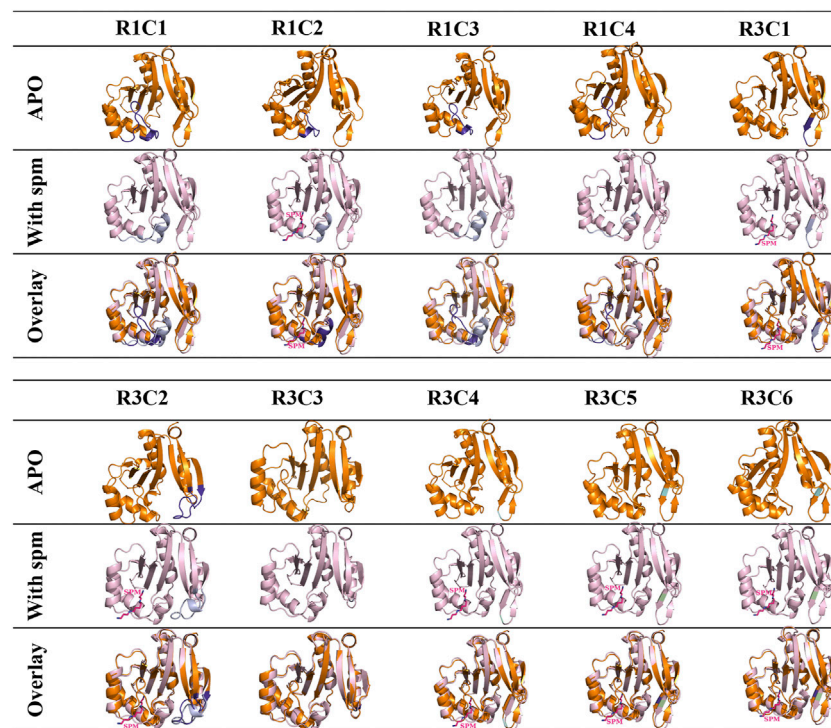


FIGURE 5 | Analysis of apo and spm-bound homology models of 10 VcSpeG constructs. Homology models of all constructs were generated using Swiss model to obtain apo- and spm-bound models since some proteins did not crystallize. Apo-homology models are in orange ribbon diagrams and the WT-apo structure (PDB ID: 4jix) was used as the template. The spm-bound homology models are in light pink and were based on the WT with spm-bound structure (PDB ID: 4mi4) as the template. Substitutions are shown in purple for the apo models and violet for the spm-bound models. Light green indicates the location of a point mutation and spm is shown as pink sticks. Spm was present in the R1C2, R3C1, R3C2, R3C4, R3C5, and R3C6 models.

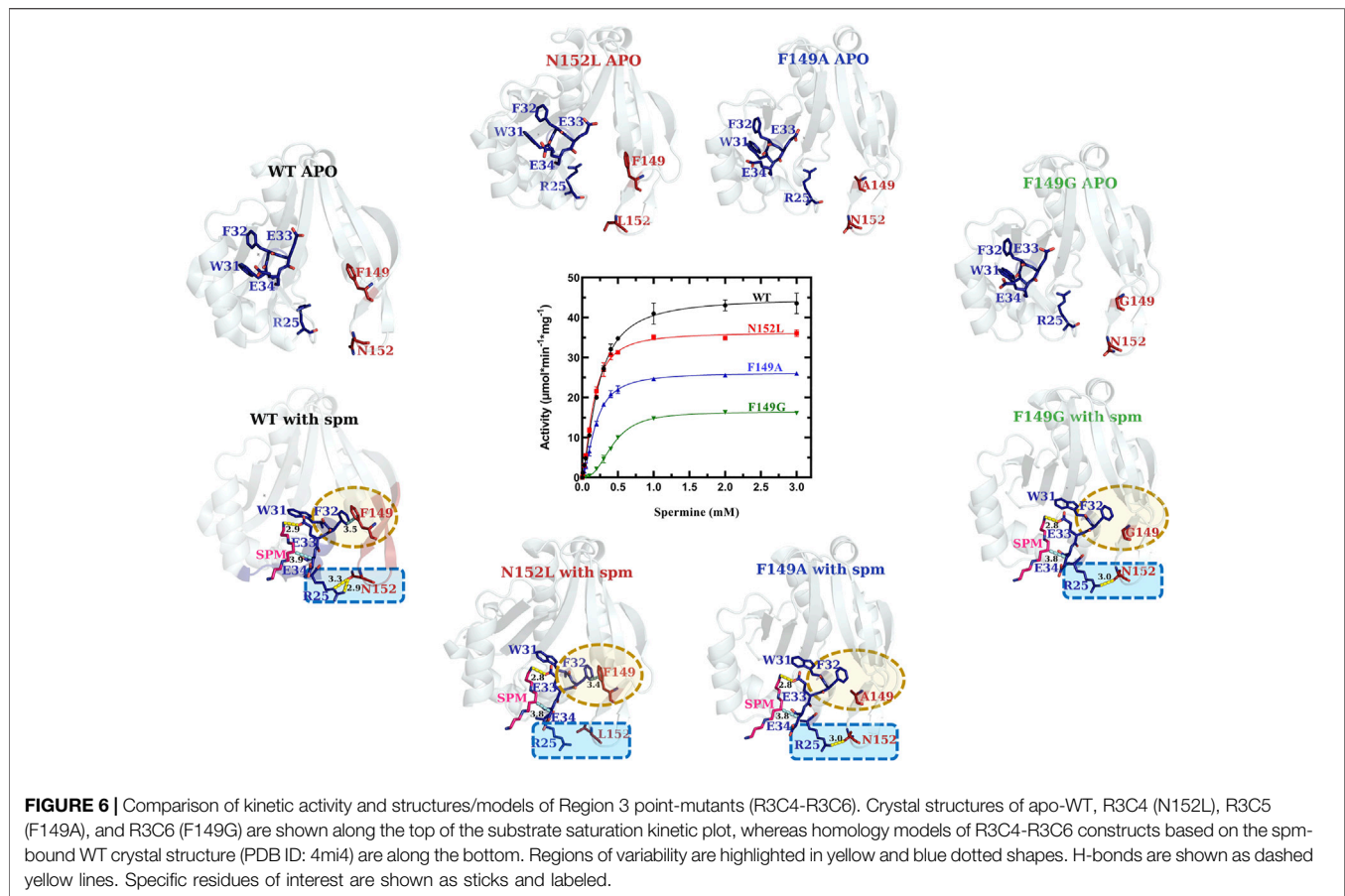
collection and refinement statistics of the solved structures can be found in the Supplemental Information **Supplementary Tables S4, S5**. All substituted residues of the VcSpeG protein sequences of these constructs were ordered in the structures, with the exception of R1C3, which had ARYEYME residues of the substitution that were disordered (**Figure 2** and **Supplementary Figure S3**). Despite crystallizing in a range of crystallography space groups (**Supplementary Table S4, S5**), six structures adopted the apo-conformation of VcSpeG (**Figure 4** and **Supplementary Figure S3**). Since all constructs were soluble and the majority were able to be crystallized, it appears the alterations to the VcSpeG sequence did not cause the protein to improperly fold.

Generation of Homology Models of Constructs in Presence and Absence of Ligand

Since we did not obtain crystals for all constructs and none had spm bound, we built homology models of each construct in the presence and absence of spm (**Figure 5**). The homology models of the apo-constructs were generated using the PDB ID: 4jix WT VcSpeG structure as the template, and the homology models of the spm-bound constructs used the spm-bound PDB ID: 4mi4 WT VcSpeG structure as the template. All models were of good

quality as defined by their QMEAN and GMQE scores (**Supplementary Table S3**). Our rationale for generating homology models of the apo constructs was three-fold: 1) some residues in the crystal structures were disordered and the placement of certain residues could not be determined, 2) the structures of some constructs could not be determined or they did not crystallize, and 3) we wanted to determine if the homology models generated were similar to the crystal structures of constructs as a way to validate the homology model predictions.

Our analysis of the apo-crystal structures and corresponding homology models of the monomer showed good root mean square deviations (rmsd) between the two (R1C2 rmsd 1.45 Å, R3C1 rmsd 1.17 Å, R3C4 rmsd 1.01 Å, R3C5 rmsd 1.25 Å, and R3C7 rmsd 1.20 Å), with the exception of R1C3 which had an rmsd of 4.33 Å. Minor differences in residue positions were observed in the overlaid R1C2 (K23-E29) and R3C1 (R149, L150, F151, K152) crystal structures and models. However, the R1C3 crystal structure had a disordered region (K23-Q34) which contributed to the higher rmsd value for the overlaid apo-structure and homology model (**Supplementary Figure S3**). We also generated homology models for the remaining constructs that had not been crystallized in the apo-form and found their rmsd values ranged from 0.36–1.56 Å (**Supplementary Table S6**).



Next, we generated homology models of the constructs based on the spm-bound VcSpeG crystal structure. The R1C2, R3C1, R3C2, R3C4, R3C5, R3C6 homology models retained spm bound to the allosteric site in the models, whereas R1C1, R1C3, R1C4, R3C3 did not. Similar to the apo crystal structures and homology models, the rmsd values for the homology models based on the spm-bound WT crystal structure were quite good. They ranged from 0.86–1.06 Å for all constructs, with the exception of R3C2, which had an rmsd of 1.84 Å (**Supplementary Table S6**). This discrepancy was based on an alteration in secondary structure predicted in the homology model compared to the WT crystal structure (described below). Based on these results, and the fact that the WT VcSpeG structure in the presence of spm has been determined previously, we felt confident the homology models could be used as a reasonably accurate tool for rationalizing our kinetic results and predicting potential structural effects of altering these regions of the VcSpeG protein in absence of crystal structures. Of course, crystal structures, binding studies, and/or molecular dynamics simulations of these constructs in the presence of spm are needed to confirm or disprove our predictions based on the homology models.

Kinetic Characterization of all VcSpeG Constructs

All proteins, including the WT enzyme, were screened for their enzymatic activity toward spm. With the exception of R3C3, all constructs exhibited measurable activity above baseline under saturating conditions. Therefore, substrate saturation curves were constructed and fits were performed to obtain estimates of the kinetic parameters of the constructs.

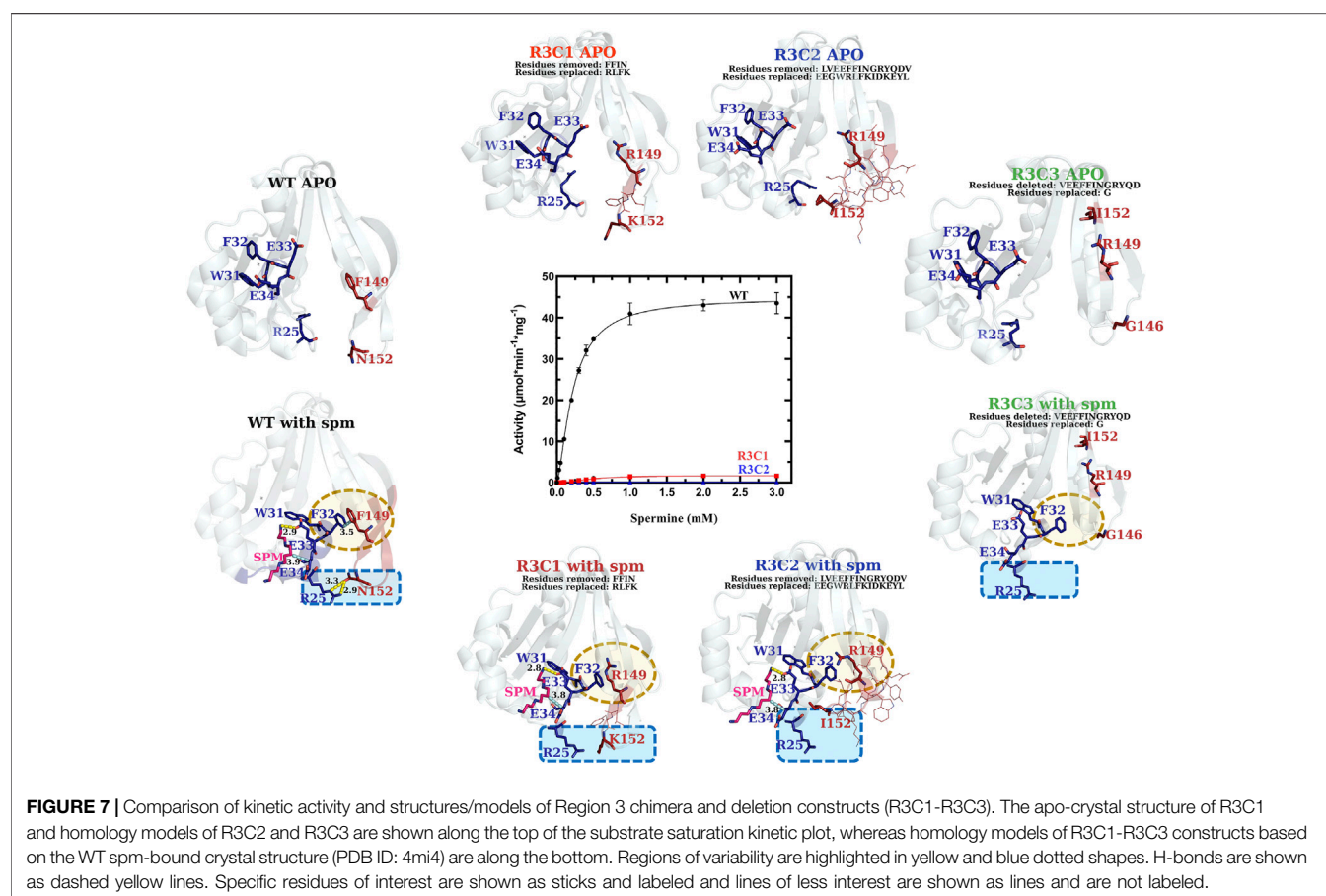
VcSpeG F149A, F149G and N152L Point Mutations Exhibit Relatively Minor Effects on Enzyme Kinetic Activity

In the WT VcSpeG structure, two main interactions between Regions 1 and 3 occur: an H-bond between R25 and N152, and π - π interactions between F32 and F149. During our kinetic analyses we found the VcSpeG N152L (R3C4), F149A (R3C5), and F149G (R3C6) mutant enzymes still retained reasonably high activity toward spm (**Figure 6** and **Table 1**). The catalytic efficiency of the N152L mutant increased slightly (1.2-fold), and the F149A and F149G mutants decreased 1.4- and 5.3-fold, respectively compared to WT. The relatively minor changes in catalytic efficiency were driven predominately by alterations to k_{cat} .

TABLE 1 | Kinetic parameters of WT VcSpeG and ten constructs.

		$S_{0.5}$ (mM)	k_{cat} (s^{-1})	$k_{cat}/S_{0.5}$ ($M^{-1} s^{-1}$)	n
Allosteric loop chimera	WT	0.222 ± 0.005	15.5	6.97×10^4	1.47 ± 0.05
	R1C1	2.53 ± 0.71	1.2	4.58×10^2	0.91 ± 0.06
	R1C2	0.364 ± 0.008	9.3	2.54×10^4	2.17 ± 0.11
	R1C3	0.244 ± 0.020	0.3	1.33×10^3	1.01 ± 0.06
	R1C4	ND*			
$\beta 6$ - $\beta 7$ region chimera	R3C1	0.459 ± 0.016	0.6	1.29×10^3	1.97 ± 0.12
	R3C2	0.098 ± 0.008	0.1	5.24×10^2	1.23 ± 0.10
	R3C3	ND*			
	R3C4	0.153 ± 0.004	12.5	8.17×10^4	1.59 ± 0.06
$\beta 6$ - $\beta 7$ region point mutants	R3C5	0.187 ± 0.003	9	4.83×10^4	1.68 ± 0.04
	R3C6	0.430 ± 0.006	5.7	1.32×10^4	2.60 ± 0.09

*None-detected.



Our analysis of the spm-bound homology models for the R3C4-R3C6 constructs showed all three retained spm in the allosteric site (Figure 6). Based on these models, a substitution of leucine for asparagine at position 152 (R3C4) caused the H-bond between side chains of R25 and N152 to be disrupted, while the π - π interactions between F32 and F149 side chains were retained. On the other hand, when we replaced phenylalanine with alanine (R3C5) and with glycine (R3C6) at position 149, the π - π interactions between F32 and F149 were removed and

the H-bond between R25 and N152 side chains remained. While this H-bond is still possible in the R3C6 homology model, it may ultimately be disrupted due to the greater flexibility of a glycine substitution at position 149. In contrast, an alanine substitution at position 149 still has hydrophobic character and may be sufficient for the H-bond between R25 and N152 to be retained as shown in the homology model. Overall, these results provide a basis for understanding why the mutations at these positions did not completely disrupt the enzyme's ability to catalyze the reaction.

The VcSpeG $\beta 6$ - $\beta 7$ Region Is Critical for Enzyme Kinetic Activity

In contrast to the single point mutations we tested, exchanging larger portions of Region 3 (R3C1-R3C3 constructs) caused dramatic effects on enzyme activity (**Figure 7** and **Table 1**). The replacement of the beta-turn (four amino acids; FFIN (F149-N152)) of Region 3 of the VcSpeG protein (R3C1) caused the activity to decrease 26-fold compared to WT. Additionally, replacement of a more significant portion (13 amino acids) of the $\beta 6$ - $\beta 7$ region of the protein (R3C2), caused the activity to decrease 155-fold compared to WT. Comparatively, the catalytic efficiency of these two constructs (R3C1 and R3C2) decreased 54-fold and 133-fold, respectively compared to WT, indicating k_{cat} exhibited a larger influence on catalytic efficiency than $S_{0.5}$. The R3C3 construct was inactive.

We analyzed the homology models that were generated based on the spm-bound VcSpeG structure for these constructs and saw both R3C1 and R3C2 had spm bound in the allosteric site, but R3C3 did not (**Figure 7**). As expected, the homology model of the R3C3 construct showed the core of the protein structure remained unchanged, but 14 amino acids of Region 3 of the protein were not present. The R3C1 model showed the secondary structures of this region generally adopted the same conformation as WT. When we examined the residues of this construct compared to the WT protein more closely, the homology models showed substitutions of F149 to arginine and F150 to leucine affect their interactions with Region 1. For example, the models showed R149 formed an H-bond with the adjacent E148 residue, which disrupted the π - π interactions originally held by F32 and F149. Additionally, when F150 was replaced by leucine in the model, the T-shaped π - π interaction originally formed between F150 and Y155 in the WT structure was disrupted. Substitution of N152 to lysine also removed the H-bond typically seen between R25 and N152 in the WT structure (**Supplementary Figure S4**).

In the R3C2 homology model, the substituted region of this construct adopts a more helical character for residues EGWRL (146–150), which is in contrast to the longer beta-strands in this region of the WT protein (**Supplementary Figure S5**). This indicates larger substitutions within this region of the protein could potentially cause changes to secondary structure. The conformation of the secondary structure that contains this substitution appears to be stabilized by H-bonds between the side chain of E155 and backbone amines of E126, N127, and D153, and an H-bond between the backbone oxygen of Y156 and backbone amine of V125. Note, E155 and Y156 correspond to tyrosine and glutamine, respectively in the WT structure and are located on opposite sides of the beta-strand. They do not form these H-bonds seen in the R3C2 construct homology model (**Supplementary Figure S5**).

Specific Portions of the VcSpeG Allosteric Loop Are Critical for Enzyme Kinetic Activity

Our examination of systematic allosteric loop substitutions (Region 1) on VcSpeG kinetic activity showed: 1) substituting large regions of this loop caused dramatic decreases in activity and 2) using this approach enabled us to narrow down the region

of the loop that had the most significant effect on activity (**Figure 8** and **Table 1**). Substituting the entire 16 residue allosteric loop with residues from the hSSAT1 protein (R1C1) decreased activity 13-fold compared to WT. On the other hand, only substituting the first eight residues of this loop (R1C2) caused a minimal (1.7-fold) decrease in activity compared to WT. Alternatively, when the first three quarters of the loop or latter half of the loop residues were substituted (R1C3 and R1C4, respectively), the enzyme activity decreased by 52-fold for R1C3 and no activity was detected for the R1C4 construct. A more drastic decrease in catalytic efficiency was observed for the R1C1 construct (two orders of magnitude compared to WT) compared to the R1C3 construct (one order of magnitude compared to WT). This difference is due to the fact that the R1C1 enzyme did not reach complete saturation under the described reaction conditions and exhibited a one order of magnitude lower apparent affinity for spm compared to the R1C3 and WT enzymes.

Based on the kinetic results, it became clear that the region of the allosteric loop that was critical for activity was between residues 31–34 (WFEE). Therefore, we examined the homology models for the Region 1 constructs to determine whether we could glean information from the structures that might explain why this region was important (**Figure 8**). In the WT structure, this portion of the allosteric loop forms several important interactions, including an NH-arene interaction between Q86 and W31 (**Supplementary Figure S6**), π - π interactions between F32 of Region 1 and F149 of Region 3, an H-bond between the terminal amine of spm and the sidechain of E33 in the allosteric site, and a hydrophobic interaction between E34 and the methylenes of spm in the allosteric site. The R1C1, R1C3, and R1C4 constructs all have MEEQ in place of WFEE, which would disrupt these interactions and may explain the reduced activity of these constructs. On the other hand, the R1C2 construct retained these residues and may explain why this construct retained activity.

DISCUSSION

Chimeric Constructs as a Tool to Study SSAT Enzymes

Our strategy for using chimeric constructs to probe the importance of different regions of a protein of interest is not new. In fact, chimeric proteins are used in all sorts of applications, including bioengineering (Adlakha et al., 2011), vaccine development (Nuccitelli et al., 2011), cancer treatments (Qiu et al., 2008), and as drugs for type 2 diabetes (Findeisen et al., 2019). Chimeric genes arise naturally during evolution and enable formation of different protein functions and the diversification of protein structures (Rogers and Hartl, 2012). Here, we have used chimera to provide a broad overview of potential regions of the VcSpeG protein that are important for its activity and structure. We specifically investigated the contribution of the VcSpeG allosteric loop and adjacent $\beta 6$ - $\beta 7$ region on its activity. Our study was primarily centered on enzyme kinetics and homology models of the constructs based on the spm-bound VcSpeG structure. It is possible that some constructs are unable to bind spm, and/or they do not adopt the conformation of the

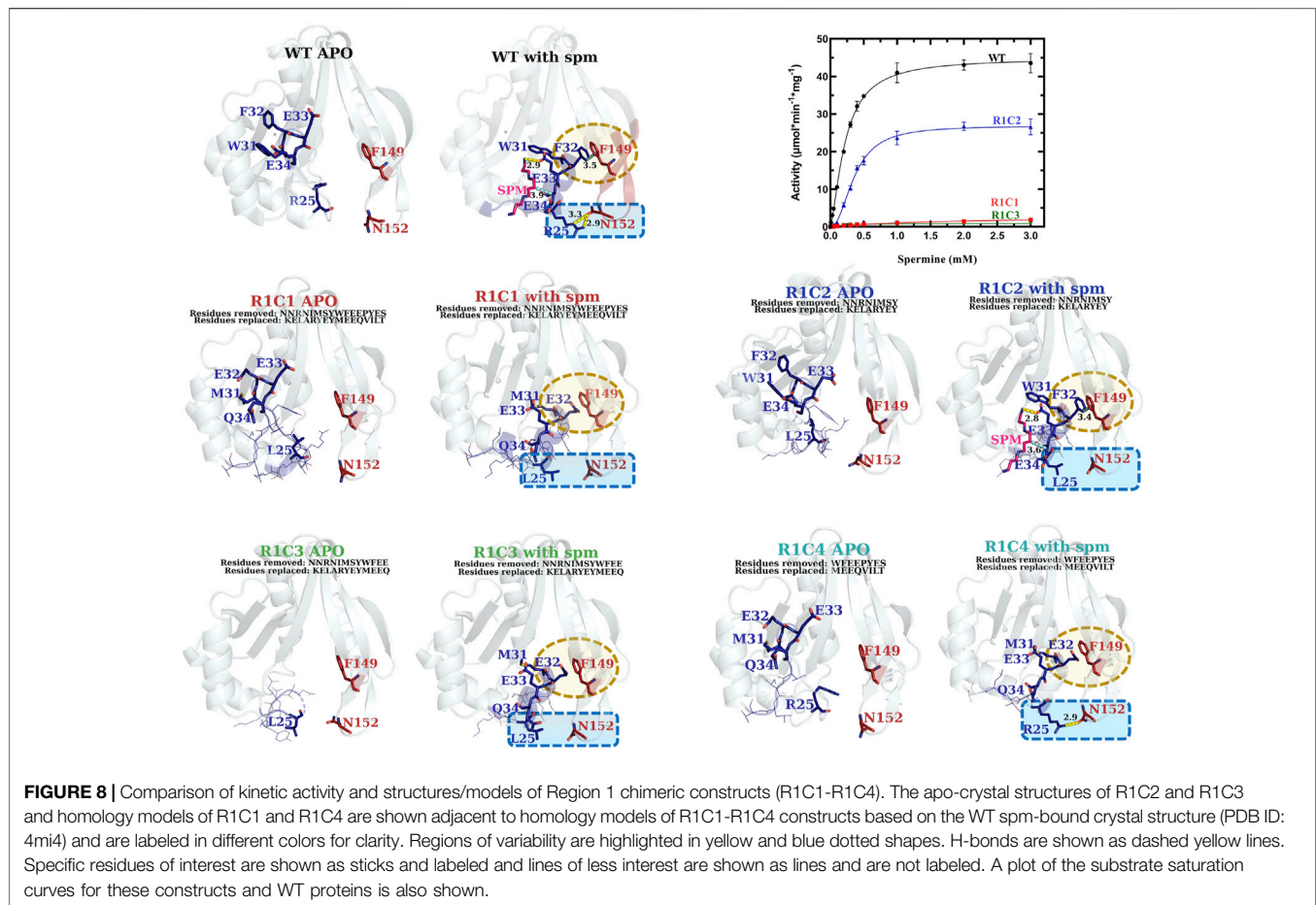


FIGURE 8 | Comparison of kinetic activity and structures/models of Region 1 chimeric constructs (R1C1-R1C4). The apo-crystal structures of R1C2 and R1C3 and homology models of R1C1 and R1C4 are shown adjacent to homology models of R1C1-R1C4 constructs based on the WT spm-bound crystal structure (PDB ID: 4mi4) and are labeled in different colors for clarity. Regions of variability are highlighted in yellow and blue dotted shapes. H-bonds are shown as dashed yellow lines. Specific residues of interest are shown as sticks and labeled and lines of less interest are shown as lines and are not labeled. A plot of the substrate saturation curves for these constructs and WT proteins is also shown.

allosteric loop observed in the models. Regardless, these models show the types of interactions that may be important for an active state of VcSpeG and form the basis for future queries across different polyamine acetyltransferases and the broader GNAT family.

Two Previously Unexplored Regions of VcSpeG Are Critical for Activity

Our results show both the allosteric loop (Region 1) and the $\beta 6$ - $\beta 7$ residues (Region 3) of the VcSpeG protein are important for VcSpeG activity. Furthermore, substituting these regions with hSSAT1 residues, with the exception of R1C2, caused major decreases in enzyme activity. The investigation of alterations to Region 1 of the VcSpeG protein (R1C1-R1C4) showed we could replace residues in the first half of the allosteric loop (R1C2) and retain moderate activity, but exchanging residues in either the second half of the loop (R1C4), the first three quarters of the loop (R1C3) or the entire loop (R1C1) caused the enzyme to be inactive. Therefore, residues within the WFEET portion of the allosteric loop are critical for activity, but we do not know whether all residues or only a subset of these four residues are required for activity toward spm for the VcSpeG enzyme. These four residues are conserved across all structurally characterized SpeG enzymes, and Sugiyama *et al.* previously showed point mutants of the *E. coli* SpeG enzyme in this comparable region

were critical for activity toward spd (Sugiyama *et al.*, 2016). However, we currently do not know whether the VcSpeG allosteric site must bind polyamine for turnover to occur and which residues in the allosteric site might transmit this signal when spm or spd is bound. Regardless, it is clear from our homology models that the conformation of the allosteric loop is important for forming adequate interactions with Region 3. Furthermore, a tightening between these two regions of the VcSpeG protein occurs in the WT structure that appears to be disrupted by substitutions to either Region 1 or Region 3 in our chimeric constructs (Figure 9).

The single point mutants we tested in Region 3 (F149, N152; R3C4-R3C6) were insufficient to totally inhibit enzyme activity. However, when we examined the chimera with a four-residue substitution (R3C1), including both of the F149 and N152 residues exchanged with positively charged residues, we saw a much more substantial decrease in activity. While it is possible a single amino acid substitution to this region may be enough to alter kinetic activity, it appears the contribution of multiple residues, largely driven by F149, is responsible for decreased activity. In the homology models, we saw these substitutions altered H-bonds and π - π interactions with residues in Region 1, which may be important for forming a more closed conformation of these two regions of the protein (Figure 9). H-bonds are important for protein structure and function, including protein

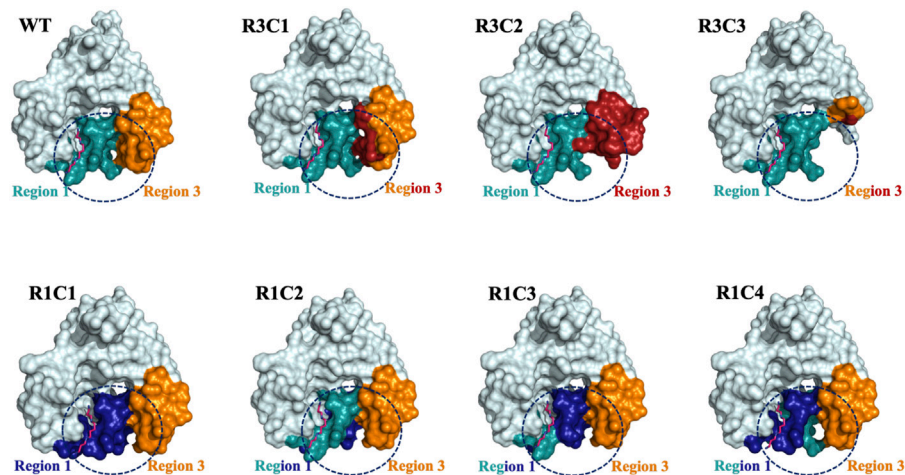


FIGURE 9 | Surface representations of selected Region 1 and 3 constructs. The surface effects of residue substitutions for R1C1-R1C4 and R3C1-R3C3 constructs compared to the WT protein are shown. These diagrams are based on the homology models of the spm-bound WT VcSpeG structure (PDB ID: 4mi4). Region 1 (allosteric loop) is shown in teal and Region 3 is in orange in all models and the WT structure. Residues exchanged from hSSAT1 in Region 1 are shown in dark blue and residues exchanged in Region 3 are in red. The dashed circle emphasizes the surface alterations to these regions of all constructs compared to the same area in the WT VcSpeG protein.

folding, protein-ligand recognition, and enzymatic activity (Jeffrey and Saenger, 1991; Myers and Pace, 1996; Nick Pace et al., 2014). Additionally, when residues are buried in non-polar environments, π - π , cation- π , and amino- π interactions have been recognized for their vital roles in protein structure and stability (Burley and Petsko, 1986; Gallivan and Dougherty, 1999; Dougherty, 2007; Dalkas et al., 2014; Vernon et al., 2018; Thakuria et al., 2019). Therefore, the combination of π - π and H-bond interactions between Regions 1 and 3 appear to be quite important for VcSpeG enzyme activity and are driven by the proper conformation of the allosteric loop.

The replacement or deletion of the entire Region 3 of the VcSpeG protein completely inhibits the enzyme's ability to turnover. Our homology models showed large substitutions in Region 3 caused changes to the secondary structure and surfaces of this region (Figure 9). Thus, the reduction in activity for these constructs is most likely due to removing key interactions with Region 1. Moreover, the R3C2 construct model shows exchanging these hSSAT1 residues opens up the space between Regions 1 and 3, and creates new interactions in Region 3 not observed in the VcSpeG WT structure or other construct models (Figure 9). In the greater context of GNAT evolution, these results indicate that relatively small changes to this region of GNAT proteins could alter the structure. However, it is possible that substitutions to Regions 1 and/or 3 may be necessary to create specific types of interfaces within or between monomers of the protein.

Further qualitative examination of the electrostatic surfaces of Regions 1 and 3 in the VcSpeG WT and construct models may provide an explanation as to why specific alterations to these two regions of the protein significantly affect enzyme activity (Figure 10). In constructs where the allosteric loop (Region 1) was altered, we observed two key changes compared to WT: 1) the acidity of the allosteric site where polyamine binds was

significantly decreased for R1C1, R1C3 and R1C4 constructs and 2) the interface between Regions 1 and 3 became more neutral except for the R1C4 construct. Since the R1C2 construct still maintained significant activity, it appears retention of acidity in the allosteric site may enable polyamine binding and subsequent catalysis, especially when significant van der Waals and complementary electrostatic interactions between Regions 1 and 3 were maintained. This was also observed in constructs where the β 6- β 7 residues (Region 3) were altered. If the basicity of Region 3 increased (R3C1, R3C2) or the region was completely removed (R3C3), the enzyme activity was obliterated or drastically decreased. In cases where only a single residue in this region was altered (R3C4-R3C6), the disruption to the activity was far less. Thus, it is likely that both the acidity of the allosteric site and complementary interfacial residues between Regions 1 and 3 must be maintained for productive catalysis.

Rationale for Effects of Allosteric Loop on VcSpeG Activity

When polyamine is bound to the VcSpeG allosteric site, the allosteric loop adopts a different conformation and forms an alpha helix. This conformational change opens up the acceptor site within the active site, creating a pocket for the polyamine acceptor substrate to bind. Therefore, substitutions to residues of the allosteric loop may affect VcSpeG activity in a variety of ways. One possibility is the allosteric loop loses or lessens its ability to bind polyamine at the allosteric site since some residues on this loop protrude into this site. These substitutions could: 1) potentially change the overall surface charge of the predominantly negatively charged binding site, or 2) create steric effects that could prevent or alter polyamine binding, affecting the allosteric signal. A second possibility is the allosteric loop may still adopt the

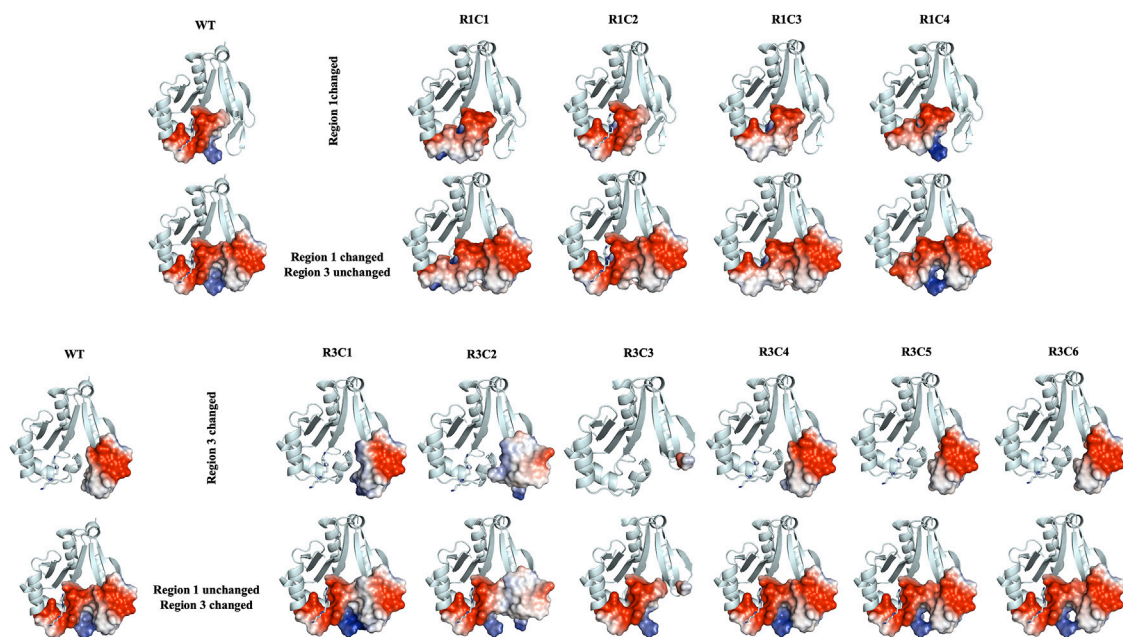


FIGURE 10 | Electrostatic surface representations for Regions 1 and 3 constructs compared to WT VcSpeG. Ribbon diagrams of homology models for each construct are shown and are based on the VcSpeG WT PDB ID: 4mi4 crystal structure with spm bound in the allosteric site. Electrostatic surfaces were calculated in Pymol for Regions 1 and 3 of each construct, where blue represents positive charge, red is negative charge and white is neutral. The upper panel of the figure shows constructs that were modified in Region 1, with Region 3 remaining unchanged, and the bottom panel shows constructs modified in Region 3, with Region 1 remaining unchanged.

helical conformation without requiring polyamine binding, but lessen turnover due to absence of an allosteric signal. Alternatively, these changes could cause the allosteric loop to block the active site, either through inability to adopt the proper loop conformation or inability to bind polyamine, thus inhibiting enzyme activity. Ultimately, a specific conformation of the allosteric loop appears to be required for interaction with Region 3 of the VcSpeG protein for catalysis to proceed efficiently. Molecular dynamics simulations may help to clarify these effects and should be explored in the future. Clearly, much remains to be learned about this enzyme system.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <http://www.wwpdb.org/>, 7kwh <http://www.wwpdb.org/>, 7kwj <http://www.wwpdb.org/>, 7kwq <http://www.wwpdb.org/>, 7kwx <http://www.wwpdb.org/>, 7kx2 <http://www.wwpdb.org/>, 7kx3.

AUTHOR CONTRIBUTIONS

VL and MK wrote the first draft of the manuscript. Experiments were performed by VL, ST, EQiL, AS, DG, PB, EQingL, and RR. VL, ST, JF and MK performed data analysis. The study was

conceptualized by MK. VL, ST, JF, and MK edited the final draft of the manuscript.

FUNDING

Research reported in this work was supported in part by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM133506 and a CSUPERB New Investigator Grant (to MLK). Additional student support for Van Le was provided by the following agencies: NSF LSAMP: HRD-1302873 and NIH MBRS-RISE: R25-GM059298. This research was undertaken in part using the MX1 and MX2 beamlines at the Australian Synchrotron, part of ANSTO, and using the Australian Cancer Research Foundation (ACRF) detector.

ACKNOWLEDGMENTS

We express our sincere appreciation to Dr. George Gassner and Dr. Anton Guliaev at San Francisco State University for their critical reading of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.645768/full#supplementary-material>.

REFERENCES

- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Cryst.* 66, 213–221. doi:10.1107/S0907444909052925
- Adlakha, N., Rajagopal, R., Kumar, S., Reddy, V. S., and Yazdani, S. S. (2011). Synthesis and characterization of chimeric proteins based on cellulase and xylanase from an insect gut bacterium. *Appl. Environ. Microbiol.* 77, 4859–4866. doi:10.1128/AEM.02808-10
- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., et al. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Cryst.* 68, 352–367. doi:10.1107/S0907444912001308
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Cryst.* 67, 271–281. doi:10.1107/S0907444910048675
- Burley, S. K., and Petsko, G. A. (1986). Dimerization energetics of benzene and aromatic amino acid side chains. *J. Am. Chem. Soc.* 108, 7995–8001. doi:10.1021/ja00285a019
- Casero, R. A., and Pegg, A. E. (1993). Spermidine/spermine N1-acetyltransferase--the turning point in polyamine metabolism. *FASEB J.* 7, 653–661. doi:10.1096/fasebj.7.8.8500690
- Childs, A. C., Mehta, D. J., and Gerner, E. W. (2003). Polyamine-dependent gene expression. *Cell Mol Life Sci* 60, 1394–1406. doi:10.1007/s00018-003-2332-4
- Cowieson, N. P., Aragao, D., Clift, M., Ericsson, D. J., Gee, C., Harrop, S. J., et al. (2015). MX1: a bending-magnet crystallography beamline serving both chemical and macromolecular crystallography communities at the Australian Synchrotron. *J. Synchrotron Radiat.* 22, 187–190. doi:10.1107/S1600577514021717
- Czub, M. P., Zhang, B., Chiarelli, M. P., Majorek, K. A., Joe, L., Porebski, P. J., et al. (2018). A Gcn5-related N-acetyltransferase (GNAT) capable of acetylating polymyxin B and colistin antibiotics *in vitro*. *Biochemistry* 57, 7011–7020. doi:10.1021/acs.biochem.8b00946
- Dalkas, G. A., Teheux, F., Kwasigroch, J. M., and Rooman, M. (2014). Cation- π , amino- π , π - π , and H-bond interactions stabilize antigen-antibody interfaces. *Proteins* 82, 1734–1746. doi:10.1002/prot.24527
- Dougherty, D. A. (2007). Cation- π interactions involving aromatic amino acids. *J. Nutr.* 137, 1504S–1508S. doi:10.1093/jn/137.6.1504S
- Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Cryst.* 66, 486–501. doi:10.1107/S0907444910007493
- Evans, P. R., and Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Cryst.* 69, 1204–1214. doi:10.1107/S0907444913000061
- Filippova, E. V., Kuhn, M. L., Osipiuk, J., Kiryukhina, O., Joachimiak, A., Ballicora, M. A., et al. (2015a). A novel polyamine allosteric site of SpeG from *Vibrio cholerae* is revealed by its dodecameric structure. *J. Mol. Biol.* 427, 1316–1334. doi:10.1016/j.jmb.2015.01.009
- Filippova, E. V., Weigand, S., Osipiuk, J., Kiryukhina, O., Joachimiak, A., and Anderson, W. F. (2015b). Substrate-induced allosteric change in the quaternary structure of the spermidine N-acetyltransferase SpeG. *J. Mol. Biol.* 427, 3538–3553. doi:10.1016/j.jmb.2015.09.013
- Filippova, W. V., Shuvalova, L., Minasov, G., Kiryukhina, O., Zhang, Y., and Clancy, S., et al. (2011). Crystal structure of the novel PaiA N-acetyltransferase from *Thermoplasma acidophilum* involved in the negative control of sporulation and degradative enzyme production. *Proteins* 79, 2566–2577.
- Findeisen, M., Allen, T. L., Henstridge, D. C., Kammoun, H., Brandon, A. E., Baggio, L. L., et al. (2019). Treatment of type 2 diabetes with the designer cytokine IC7Fc. *Nature* 574, 63–68. doi:10.1038/s41586-019-1601-9
- Finger, S., Schwieger, C., Arouri, A., Kerth, A., and Blume, A. (2014). Interaction of linear polyamines with negatively charged phospholipids: the effect of polyamine charge distance. *Biol. Chem.* 395, 769–778. doi:10.1515/hsz-2014-0126
- Firpo, M. R., and Mounce, B. C. (2020). Diverse functions of polyamines in Virus infection. *Biomolecules* 10, 628. doi:10.3390/biom10040628
- Floris, G., and Finazzi Agrò, A. (2013). “Amine oxidases,” in *In Encyclopedia of biological chemistry*. Editors W. J. Lennarz and M. D. Lane. 2nd Edn (Waltham: Academic Press), 87–90. doi:10.1016/B978-0-12-378630-2.00182-1
- Forouhar, F., Lee, I.-S., Vujcic, J., Vujcic, S., Shen, J., Vorobiev, S. M., et al. (2005). Structural and functional evidence for *Bacillus subtilis* PaiA as a novel N1-spermidine/spermine acetyltransferase. *J. Biol. Chem.* 280, 40328–40336. doi:10.1074/jbc.M505332200
- Franklin, M. C., Cheung, J., Rudolph, M. J., Burshteyn, F., Cassidy, M., Gary, E., et al. (2015). Structural genomics for drug design against the pathogen *Coxiella burnetii*. *Proteins* 83, 2124–2136. doi:10.1002/prot.24841
- Fukuchi, J., Kashiwagi, K., Takio, K., and Igarashi, K. (1994). Properties and structure of spermidine acetyltransferase in *Escherichia coli*. *J. Biol. Chem.* 269, 22581–22585. doi:10.1016/s0021-9258(17)31685-x
- Gallivan, J. P., and Dougherty, D. A. (1999). Cation- π interactions in structural biology. *Proc. Natl. Acad. Sci. U S A* 96, 9459–9464. doi:10.1073/pnas.96.17.9459
- Hamana, K., and Matsuzaki, S. (1992). Polyamines as a chemotaxonomic marker in bacterial systematics. *Crit. Rev. Microbiol.* 18, 261–283. doi:10.3109/10408419209113518
- Hegde, S. S., Chandler, J., Vetting, M. W., Yu, M., and Blanchard, J. S. (2007). Mechanistic and structural analysis of human spermidine/spermine N1-acetyltransferase [hSSAT]. *Biochemistry* 46, 7187–7195. doi:10.1021/bi700256z
- Ikeguchi, Y., Bewley, M. C., and Pegg, A. E. (2006). Aminopropyltransferases: function, structure and genetics. *J. Biochem.* 139, 1–9. doi:10.1093/jb/mvj019
- Jeffrey, G. A., and Saenger, W. (1991). *Hydrogen bonding in biological structures*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-85135-3
- Jelsbak, L., Thomsen, L. E., Wallrodt, I., Jensen, P. R., and Olsen, J. E. (2012). Polyamines are required for virulence in *Salmonella enterica* serovar typhimurium. *PLoS One* 7, e36149. doi:10.1371/journal.pone.0036149
- Karatan, E., Duncan, T. R., and Watnick, P. I. (2005). NspS, a predicted polyamine sensor, mediates activation of *Vibrio cholerae* biofilm formation by norspermidine. *J. Bacteriol.* 187, 7434–7443. doi:10.1128/JB.187.21.7434-7443.2005
- Kim, S. H., Wang, Y., Khomutov, M., Khomutov, A., Fuqua, C., and Michael, A. J. (2016). The essential role of spermidine in growth of *Agrobacterium tumefaciens* is determined by the 1,3-diaminopropane moiety. *ACS Chem. Biol.* 11, 491–499. doi:10.1021/acschembio.5b00893
- Kovalevskiy, O., Nicholls, R. A., Long, F., Carlon, A., and Murshudov, G. N. (2018). Overview of refinement procedures within REFMAC5: utilizing data from different sources. *Acta Cryst. Sect D Struct. Biol.* 74, 215–227. doi:10.1107/S2059798318000979
- Kramer, D. L., Diegelman, P., Jell, J., Vujcic, S., Merali, S., and Porter, C. W. (2008). Polyamine acetylation modulates polyamine metabolic flux, a prelude to broader metabolic consequences. *J. Biol. Chem.* 283, 4241–4251.
- Kuhn, M. L., Majorek, K. A., Minor, W., and Anderson, W. F. (2013). Broad-substrate screen as a tool to identify substrates for bacterial Gcn5-related N-acetyltransferases with unknown substrate specificity. *Protein Sci.* 22, 222–230. doi:10.1002/pro.2199
- Li, B., Maezato, Y., Kim, S. H., Kurihara, S., Liang, J., and Michael, A. J. (2019). Polyamine-independent growth and biofilm formation, and functional spermidine/spermine N-acetyltransferases in *Staphylococcus aureus* and *Enterococcus faecalis*. *Mol. Microbiol.* 111, 159–175. doi:10.1111/mmi.14145
- Majorek, K. A., Kuhn, M. L., Chruszcz, M., Anderson, W. F., and Minor, W. (2014). Double trouble-Buffer selection and His-tag presence may be responsible for nonreproducibility of biomedical experiments. *Protein Sci.* 23, 1359–1368. doi:10.1002/pro.2520
- McCoy, A. J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* 63, 32–41. doi:10.1107/S0907444906045975
- McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., et al. (2002). Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* 9, 401–406. doi:10.1107/s0909049502015170
- Michael, A. J. (2016). Polyamines in eukaryotes, bacteria, and archaea. *J. Biol. Chem.* 291, 14896–14903. doi:10.1074/jbc.R116.734780

- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., et al. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Cryst.* 67, 355–367. doi:10.1107/S0907444911001314
- Myers, J. K., and Pace, C. N. (1996). Hydrogen bonding stabilizes globular proteins. *Biophys. J.* 71, 2033–2039. doi:10.1016/S0006-3495(96)79401-8
- Nesse, L. L., Berg, K., and Vestby, L. K. (2015). Effects of norspermidine and spermidine on biofilm formation by potentially pathogenic *Escherichia coli* and *Salmonella enterica* wild-type strains. *Appl. Environ. Microbiol.* 81, 2226–2232. doi:10.1128/AEM.03518-14
- Nick Pace, C., Scholtz, J. M., and Grimsley, G. R. (2014). Forces stabilizing proteins. *FEBS Lett.* 588, 2177–2184. doi:10.1016/j.febslet.2014.05.006
- Nuccitelli, A., Cozzi, R., Gourlay, L. J., Donnarumma, D., Necchi, F., Norais, N., et al. (2011). Structure-based approach to rationally design a chimeric protein for an effective vaccine against Group B Streptococcus infections. *Proc. Natl. Acad. Sci. U S A.* 108, 10278–10283. doi:10.1073/pnas.1106590108
- Pegg, A. E., and Casero, R. A. (2011). Current status of the polyamine research field. *Methods Mol. Biol.* 720, 3–35. doi:10.1007/978-1-61779-034-8_1
- Pegg, A. E., and Michael, A. J. (2010). Spermine synthase. *Cell. Mol. Life Sci.* 67, 113–121. doi:10.1007/s00018-009-0165-5
- Qiu, X.-C., Xu, Y.-M., Wang, F., Fan, Q.-Y., Wang, L.-F., Ma, B.-A., et al. (2008). Single-chain antibody/activated BID chimeric protein effectively suppresses HER2-positive tumor growth. *Mol. Cancer Ther.* 7, 1890–1899. doi:10.1158/1535-7163.MCT-07-2235
- Rogers, R. L., and Hartl, D. L. (2012). Chimeric genes as a source of rapid evolution in *Drosophila melanogaster*. *Mol. Biol. Evol.* 29, 517–529. doi:10.1093/molbev/msr184
- Schneider, J., and Wendisch, V. F. (2011). Biotechnological production of polyamines by Bacteria: recent achievements and future perspectives. *Appl. Microbiol. Biotechnol.* 91, 17–30. doi:10.1007/s00253-011-3252-0
- Shah, P., and Swiatlo, E. (2008). A multifaceted role for polyamines in bacterial pathogens. *Mol. Microbiol.* 68, 4–16. doi:10.1111/j.1365-2958.2008.06126.x
- Sobe, R. C., Bond, W. G., Wotanis, C. K., Zayner, J. P., Burriss, M. A., Fernandez, N., et al. (2017). Spermine inhibits *Vibrio cholerae* biofilm formation through the NspS-MbaA polyamine signaling system. *J. Biol. Chem.* 292, 17025–17036. doi:10.1074/jbc.M117.801068
- Sugiyama, S., Ishikawa, S., Tomitori, H., Niiyama, M., Hirose, M., Miyazaki, Y., et al. (2016). Molecular mechanism underlying promiscuous polyamine recognition by spermidine acetyltransferase. *Int. J. Biochem. Cell Biol.* 76, 87–97. doi:10.1016/j.biocel.2016.05.003
- Tabor, C. W., and Tabor, H. (1984). Polyamines. *Annu. Rev. Biochem.* 53, 749–790. doi:10.1146/annurev.bi.53.070184.003533
- Thakuria, R., Nath, N. K., and Saha, B. K. (2019). The nature and applications of π - π interactions: a perspective. *Cryst. Growth Des.* 19, 523–528. doi:10.1021/acs.cgd.8b01630
- Tsimbalyuk, S., Shornikov, A., Thi Bich Le, V., Kuhn, M. L., and Forwood, J. K. (2020). SpeG polyamine acetyltransferase enzyme from *Bacillus thuringiensis* forms a dodecameric structure and exhibits high catalytic efficiency. *J. Struct. Biol.* 210, 107506. doi:10.1016/j.jsb.2020.107506
- Vernon, R. M., Chong, P. A., Tsang, B., Kim, T. H., Bah, A., Farber, P., et al. (2018). Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* 7. doi:10.7554/eLife.31486
- Wallace, H. M., Fraser, A. V., and Hughes, A. (2003). A perspective of polyamine metabolism. *Biochem. J.* 376, 1–14. doi:10.1042/BJ20031327
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi:10.1093/nar/gky427
- Woolridge, D. P., Martinez, J. D., Stringer, D. E., and Gerner, E. W. (1999). Characterization of a novel spermidine/spermine acetyltransferase, BltD, from *Bacillus subtilis*. *Biochem. J.* 340, 753–758. doi:10.1042/bj3400753
- Zhu, Y.-Q., Zhu, D.-Y., Yin, L., Zhang, Y., Vonnrhein, C., and Wang, D.-C. (2006). Crystal structure of human spermidine/spermine N1-acetyltransferase (hSSAT): the first structure of a new sequence family of transferase homologous superfamily. *Proteins: Struct. Funct. Bioinformatics* 63, 1127–1131. doi:10.1002/prot.20965

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Le, Tsimbalyuk, Lim, Solis, Gawat, Boeck, Lim, Renolo, Forwood and Kuhn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Modulating Enzyme Function *via* Dynamic Allostery within Biliverdin Reductase B

Jasmina S. Redzic¹, Michael R. Duff², Ashley Blue³, Todd M. Pitts⁴, Pratul Agarwal⁵ and Elan Zohar Eisenmesser^{1*}

¹Department of Biochemistry and Molecular Genetics, School of Medicine, University of Colorado Denver, Denver, CO, United States, ²Biochemistry and Cellular and Molecular Biology Department, University of Tennessee, Knoxville, TN, United States, ³National High Magnetic Field Laboratory, Tallahassee, FL, United States, ⁴Division of Medical Oncology, School of Medicine, University of Colorado, Aurora, CO, United States, ⁵Department of Physiological Sciences and High Performance Computing Center, Oklahoma State University, Stillwater, OK, United States

OPEN ACCESS

Edited by:

Vincenzo Venditti,
Iowa State University, United States

Reviewed by:

Magnus Wolf-Watz,
Umeå University, Sweden
Giuseppe Melacini,
McMaster University, Canada

*Correspondence:

Elan Zohar Eisenmesser
Elan.Eisenmesser@ucdenver.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 05 April 2021

Accepted: 30 April 2021

Published: 20 May 2021

Citation:

Redzic JS, Duff MR, Blue A, Pitts TM,
Agarwal P and Eisenmesser EZ (2021)
Modulating Enzyme Function via
Dynamic Allostery within Biliverdin
Reductase B.
Front. Mol. Biosci. 8:691208.
doi: 10.3389/fmolb.2021.691208

The biliverdin reductase B (BLVRB) class of enzymes catalyze the NADPH-dependent reduction of multiple flavin substrates and are emerging as critical players in cellular redox regulation. However, the role of dynamics and allostery have not been addressed, prompting studies here that have revealed a position 15 Å away from the active site within human BLVRB (T164) that is inherently dynamic and can be mutated to control global micro-millisecond motions and function. By comparing the inherent dynamics through nuclear magnetic resonance (NMR) relaxation approaches of evolutionarily distinct BLVRB homologues and by applying our previously developed Relaxation And Single Site Multiple Mutations (RASSMM) approach that monitors both the functional and dynamic effects of multiple mutations to the single T164 site, we have discovered that the most dramatic mutagenic effects coincide with evolutionary changes and these modulate coenzyme binding. Thus, evolutionarily changing sites distal to the active site serve as dynamic “dials” to globally modulate motions and function. Despite the distal dynamic and functional coupling modulated by this site, micro-millisecond motions span an order of magnitude in their apparent kinetic rates of motions. Thus, global dynamics within BLVRB are a collection of partially coupled motions tied to catalytic function.

Keywords: enzyme, dynamics, allostery, coupling, reductase, NMR

INTRODUCTION

The allosteric regulation of enzyme function is now recognized to occur through multiple mechanisms, which can be mechanical where physical interactions induce long-range conformational changes or dynamic where motions influence networks of partially coupled movements (Motlagh et al., 2014). Methods that are reliant on NMR chemical shifts or evolutionary substitutions have been powerful tools to identify allosteric networks (Selvaratnam et al., 2012a; Selvaratnam et al., 2012b; Gagne et al., 2015; Salinas and Ranganathan, 2018). In contrast, identifying dynamically coupled networks that underly allosteric communication has been

Abbreviations: BLVRA, biliverdin reductase A; BLVRB, biliverdin reductase B; CPMG, Car-Purcell-Meiboom-Gill; NMR, nuclear magnetic resonance.

more challenging. This is in part due to the fact that dynamics may reflect changes in sampled conformations that are small and are therefore reliant on sensitive experimental methods to directly monitor such changes to what has been called “invisible states” (Kleckner and Foster, 2011). Enzymes are often particularly reliant on dynamics to perform their catalytic functions and dynamically coupled networks that underlie allosteric communication has been identified within several enzymes to date (Doucet, 2011; Duff et al., 2018). Thus, controlling such dynamic allostery may be a powerful way to engineer or fine-tune catalytic function (Doucet, 2011). However, it has also become increasingly recognized that protein dynamics are segmental in nature that lead to partial couplings and often span large distances within proteins (Schlegel et al., 2009; Torbeev et al., 2011; McDonald et al., 2012). Such complicated networks of partially coupled dynamics make it difficult to understand the role of allostery in enzyme function. To address such partially coupled networks, we have previously developed a simple but straightforward strategy that utilizes several criteria for the selection of distally coupled dynamic residues that are then mutated to help understand their roles in allostery and function (Holliday et al., 2017). The first criterion is that these distal residues are inherently dynamic and the second criterion is that they are coupled to active site perturbations, such as exhibiting chemical shift perturbations (CSPs) or dynamic changes upon substrate binding or mutagenesis to the active site. In our initial application of this approach to the isomerase cyclophilin-A, mutations to an inherently dynamic residue 20 Å away from the active site, termed a dynamic “hot spot”, could be mutated in order to control substrate binding in an isomer-specific manner (Holliday et al., 2017). Such “hot spot” mutations induced global dynamic effects that controlled catalytic efficiency and could also be used to map specific networks of dynamically coupled interactions (Holliday et al., 2017). We therefore referred to this approach as Relaxation And Single Site Multiple Mutations (RASSMM) in order to emphasize the fact that we could monitor distal relaxation effects and changes to function by specifically engineering a panel of mutations to one single site. R2-Car-Purcell-Meiboom-Gill (CPMG) dispersion experiments have proven to be a particularly powerful relaxation experiment in monitoring such global changes to dynamics. R2-CPMG dispersion monitors R2 relaxation as a function of an imparted CPMG refocusing field, which provides information regarding rates of motions, populations sampled, and even structural information (Kleckner and Foster, 2011; Alderson and Kay, 2020). Here, we have applied this RASSMM approach to biliverdin reductase B (BLVRB) in order to determine whether global motions are allosterically coupled to function within this enzyme and whether enzyme motions may be modulated from a distance to control function in a residue-specific manner.

The biliverdin reductase B (BLVRB) family of enzymes are emerging as critical flavin reductases in multiple organisms that range from pathogenic bacteria to humans, as their flavin substrates act as redox sensors and coenzymes for many other enzymes (Vervoort and Rietjens, 1996; McDonagh, 2001; Becker et al., 2011; Huijbers et al., 2014; Adak and Begley, 2017). The BLVRB enzyme family catalyze the NADPH-reduction of bilirubin but also multiple flavins (**Figure 1A**), which includes flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN). In

humans, redox regulation via BLVRB (NCBI Reference Sequence: NP_000704.1) is so important that this enzyme alone can dictate hematopoietic cell fate (Wu et al., 2016). This critical role is also consistent with our discovery of high levels of BLVRB expression in red blood cells (Paukovich et al., 2018). Mechanistically, we have previously discovered that coenzyme binding is orders of magnitude tighter than substrate binding (Paukovich et al., 2018), which is in contrast to the well-known dihydrofolate reductase (DHFR) where both coenzyme and substrate bind relatively tightly (Fierke et al., 1987). However, both BLVRB and DHFR share several common catalytic features. For example, both enzymes share similar mechanisms where bulk water is used for an initial protonation step of the substrate followed by hydride transfer from the coenzyme to the substrate (Smith et al., 2008; Liu et al., 2014). Another similarity is that despite their completely different tertiary structures, both DHFR and BLVRB rely on loop closure for function. Specifically, the “M20 loop” of DHFR dynamically modulates its activity (Boehr et al., 2006; Bhabha et al., 2013; Singh et al., 2015; Hughes et al., 2017) and we have recently shown that the “R78 loop” of BLVRB dynamically modulates its activity (Paukovich et al., 2018; **Figure 1B**). Thus, loop closure within these reductases provides a unique environment for their catalytic transfer of a hydride from one molecule (the coenzyme) to another (the substrate) and their activities are not necessarily reliant on a single residue for catalysis. This is likely why attempts to identify a “catalytic residue” within BLVRB have been unsuccessful, which include BLVRB active site mutants of S111 and H153 that result in diminished function but not functional knockouts (Smith et al., 2008; Chu et al., 2017). DHFR is similar in that catalysis is a collective of multiple residues that contribute to the active site environment, as elucidated by elegant studies that have shown residues such as DHFR D27 and Y100 exhibit synergistic effects (Liu et al., 2014). However, while distal dynamic residues within DHFR modulate active site motions and function (Watt et al., 2007; Mauldin et al., 2012; Duff et al., 2018), such global coupling remains unknown for BLVRB.

To determine whether distally coupled dynamic residues modulate active site motions and function in BLVRBs, here we used the RASSMM approach on human BLVRB along with comparative relaxation studies of lemur BLVRB (NCBI Reference Sequence: XP_020138941.1). Our previous studies had suggested that a distal site, T164, is both inherently dynamic and coupled to the active site, which would fulfill the requirements for selecting a residue to then mutate in accord with the RASSMM approach. For example, coenzyme binding induces chemical shift perturbations (CSPs) to active site residues such as S111 and H153, but also CSPs to distal sites that included T164 within the C-terminal lobe (**Figure 1C**). Here, both ensemble calculations and active sites mutants corroborated the inherent dynamics and active site coupling of this T164 site, which are the requirements for the selection of residues to select for the RASSMM approach. Considering that the second step in the RASSMM approach requires mutating this identified site to a series of residues, we used evolution as a guide to help select site-specific mutations. Specifically, we discovered that position 164 within mammalian BLVRBs toggles between a threonine and serine and that dynamics are completely quenched within the lemur BLVRB homologue that comprises only one single substitution within this region, which is

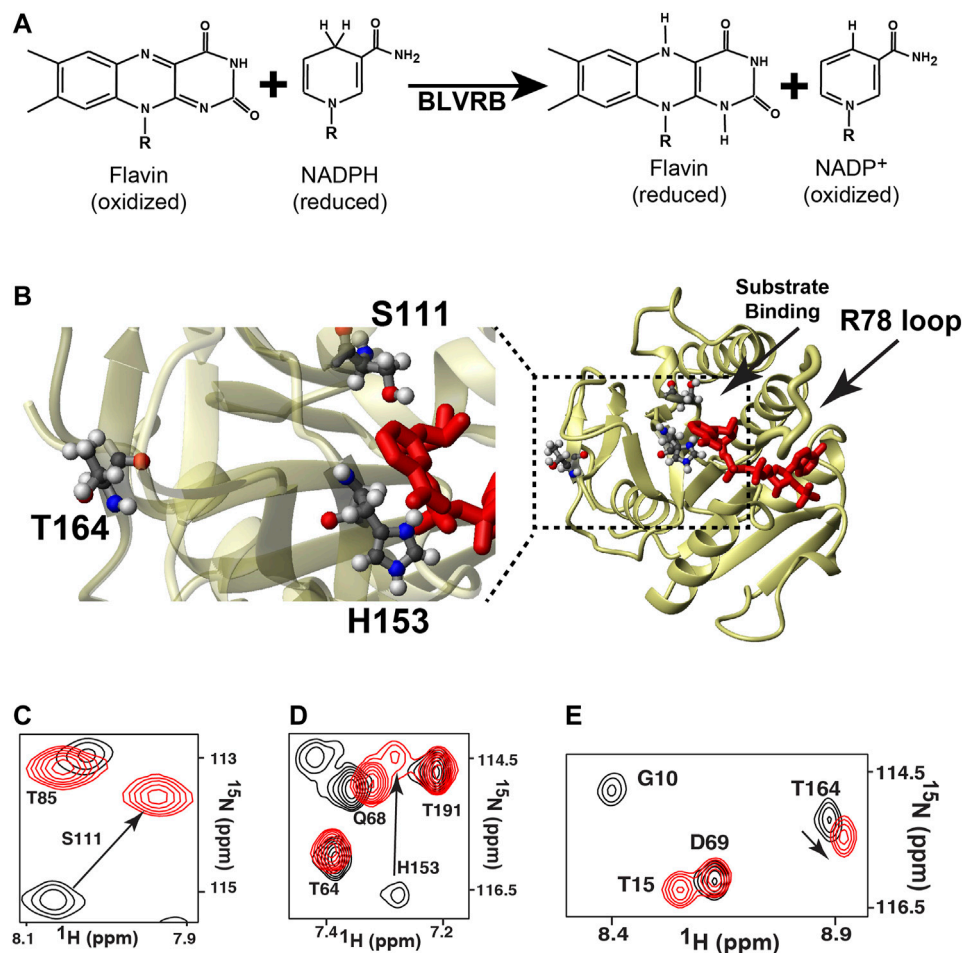


FIGURE 1 | The general reaction mediated by BLVRB and the global impact mediated by coenzyme binding. **(A)** The catalytic reaction of BLVRB comprises the NADPH-dependent reduction of flavins. For flavins, “R” corresponds to different moieties that define the flavin such as FAD or FMN. For NADPH/NADP⁺, only the nicotinamide moiety is shown, and “R” corresponds to the remaining molecule. **(B)** X-ray crystal structure of WT BLVRB, accession number 1HDO (Pereira et al., 2001), along with a blow-up of the active site with residues S111 and H153 and residue T164 that is 15 Å away from the coenzyme. ¹⁵N-HSQC of apo BLVRB (black) and holo BLVRB (red) of the amides for **(C)** S111, **(D)** H153, and **(E)** T164.

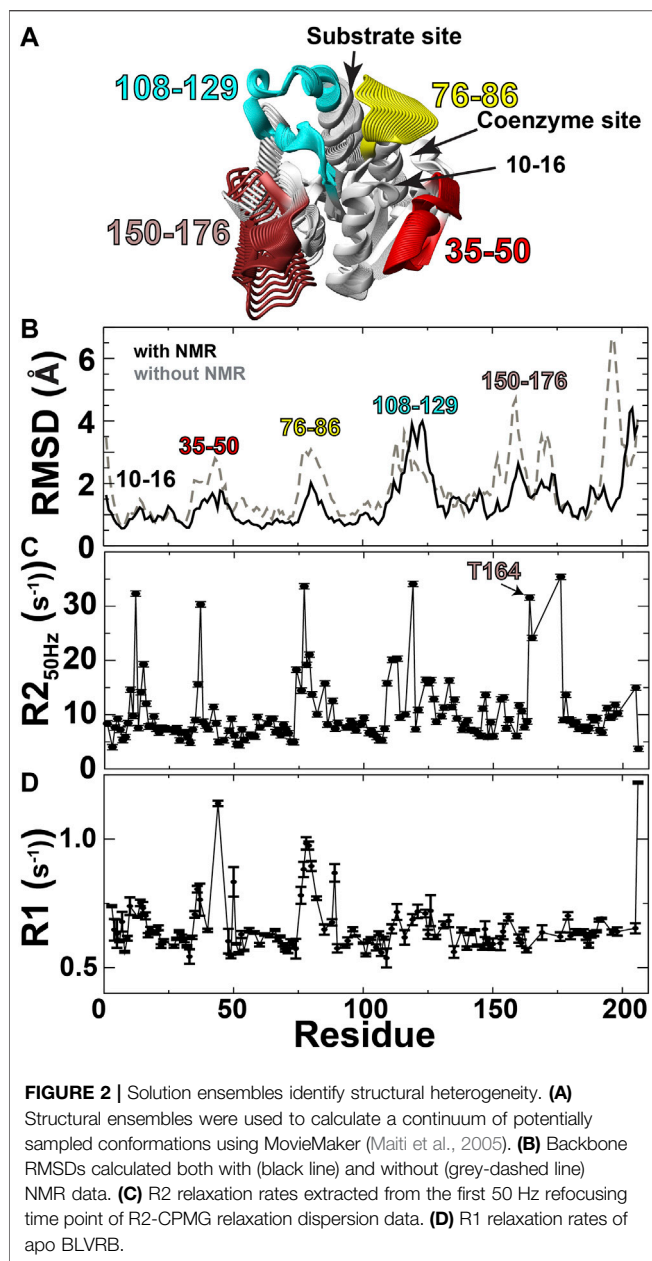
S164. Thus, in accord with the second step of the RASSMM approach, multiple mutations to human BLVRB T164 that included a serine were engineered and further corroborated coupling to the active site through both CSPs and R2-CPMG dispersions. Such allosteric changes that could be altered by mutations to T164 were found to increase coenzyme affinity. We have therefore discovered that this distal site may be modulated by either evolution or mutagenesis to allosterically control function.

RESULTS

Ensemble Calculations Identify the Conformational Plasticity Measured via NMR Relaxation

We first sought an atomic-resolution description of BLVRB dynamics through structural studies in order to identify distal

residues to the active site that are inherently dynamic, which is the first criterion of the RASSMM approach. While previous relaxation studies indicate that several regions within BLVRB exhibit chemical exchange (Paukovich et al., 2019), the underlying physical exchange process that induces chemical exchange may be due to dynamics of these regions themselves or neighboring regions. Structural studies are particularly challenging with inherently dynamic proteins such as BLVRB, as high-resolution structural data are difficult to obtain. For example, the difficulty in crystallizing apo BLVRB has been ascribed to its dynamic nature (Chu et al., 2017), which is consistent with our previous studies that have identified millisecond-microsecond (μ s-ms) timescale exchange monitored through R2-CPMG relaxation dispersion largely quenched within the holo enzyme (Paukovich et al., 2018). In fact, the resonances from several regions such as residues 167–175 and 199–204 are simply not observed, potentially owing to exchange on an intermediate timescale (Paukovich et al.,



2018). Thus, as an alternative structural approach, chemical shift-based methods in conjunction with sparse NOEs were used here to guide solution ensemble calculations, which have been shown to produce structures with remarkable accuracy and precision as we and others have shown (Rosato et al., 2012; Kendrick et al., 2014; Holliday et al., 2017). Specifically, we used CS-Rosetta that utilizes fragment libraries to build structural ensembles and has been successfully used for proteins almost twice as large as BLVRB (Lange et al., 2012). While such calculations may therefore be biased toward known structures, they represent plausible conformations consistent with structural data (chemical shifts, NOEs), which can be directly compared to relaxation data that monitors motions on multiple timescales. To obtain these low-resolution structural ensembles (Figure 2;

Table 1), we employed several experimental restraints described in Materials and Methods, which included sparse NOEs (Supplementary Table S1).

Ensemble calculations of apo BLVRB identify potential dynamic regions that predominantly include the active site along with residues 156–176 within the C-terminal lobe that comprises residues 150–206 (Figure 2). Specifically, CS-Rosetta was used to calculate solution ensembles of apo BLVRB using NMR chemical shifts (BMRB accession 27,462) and supplemented with sparse NOEs (Supplementary Table S1). These ensemble calculations are shown here as a continuum (Figure 2A), which pictorially highlights the dynamic regions that include the active site and residues 156–176 within the C-terminal lobe. Ensemble calculations using experimental restraints result in better precision (i.e., smaller RMSDs) than those calculated in the absence of any experimental NMR data (Figure 2B, black line vs. grey-dashed line, respectively), indicating that the experimental data helps define the solution ensembles. The RMSD of these experimentally driven structural ensembles was then compared to NMR relaxation data. These comparisons include both μ s-ms motions monitored via R2 relaxation (Figure 2C) and ps-ns motions monitored via R1 relaxation, as previously measured (Paukovich et al., 2018; Figure 2D). Conformationally heterogeneous regions within the calculated ensembles were largely corroborated by these NMR relaxation measures, as they were confirmed to be mobile in either the faster ps-ns timescale (residues 35–50 and 76–86) and/or slower μ s-ms timescale (residues 76–86, 108–129, and 156–176). Residues 10–16 likely exhibit chemical exchange due to the R78 loop that samples both open/closed conformations within the solution ensembles. Finally, Thus, in accord with the RASSM approach that first seeks to identify distally dynamic regions to the active site, both our previous relaxation experiments and these data-driven ensembles indicate that residues within this region of 156–176 satisfy this first criterion.

TABLE 1 | Structural statistics for the apo BLVRB CS-rosetta structures. NOE restraints were exported from CCPNmr as exact distances used for calculations. The conserved SDR domain comprises residues 1–150.

No. of residues	206
Chemical Shift restraints	1,338
N/HN	83/182
CA/HA	193/175
CB/HB	174/177
NOE distance restraints	143
Avg violation per structure	23 ± 3
Avg violation distance (Å)	0.27 ± 0.01
Medium-range ($ i-j < 4$)	73
Long-range ($ i-j > 5$)	70
Average rmsd	
Backbone, 1–206 (Å)	2.58 ± 0.68
Backbone, 1–150 (Å)	2.00 ± 0.68
Ramachandran plot summary (%)	
Most favored regions	83.4
Allowed regions	16.6

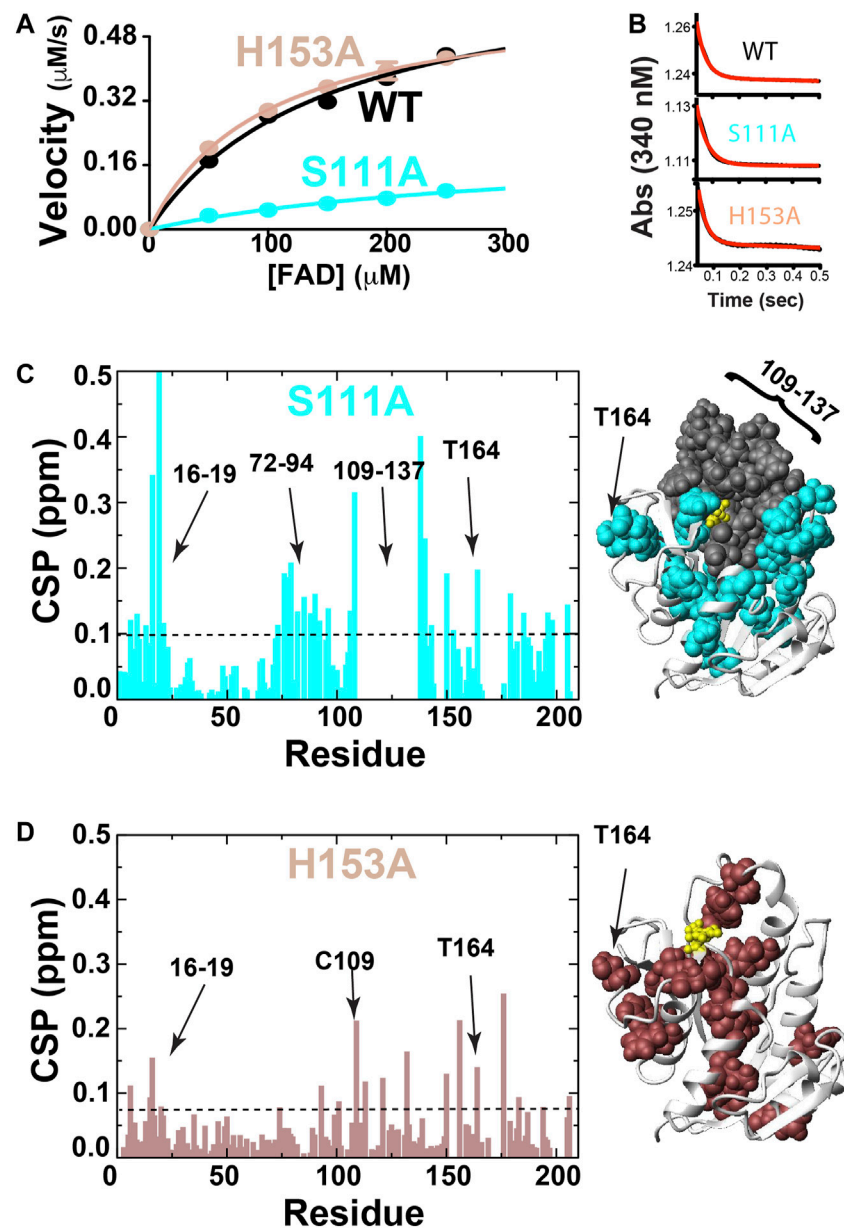


FIGURE 3 | BLVRB active site mutations S111A and H153A. **(A)** Michaelis-Menten kinetics monitored for BLVRB WT, S111A, and H153A. **(B)** Pre-steady-state kinetics for BLVRB WT, S111A, and H153A. **(C)** CSPs between apo BLVRB WT and apo BLVRB S111A with the line indicating 0.1 ppm that is the average plus 1/2 of the standard deviation. Residues for which their corresponding resonances were not found are also shown (gray). **(D)** CSPs between apo BLVRB WT and apo BLVRB H153A with the line indicating 0.075 ppm that is the average plus 1/2 of the standard deviation.

Active Site Mutants are Globally Coupled

In order to address the second criterion of the RASSMM approach that seeks to identify inherently dynamic regions coupled to the active site, we sought to mutate specific residues to monitor their distal effects. Already, we knew that coenzyme binding induces changes to distal sites, yet the conformational rearrangements are quite extensive that provided the impetus here to make site-specific changes. We selected BLVRB S111A and H153A, which have previously been shown to alter BLVRB function through kinetics studies that

indicate the underlying reasons are due to diminished substrate affinity for S111A and diminished coenzyme affinity for H153A (Chu et al., 2017; Smith et al., 2008). We confirmed that both mutants are active in reducing FAD, which resulted in the expected reduction to substrate affinity solely for the S111A mutation (Figure 3A; Table 2). Interestingly, neither mutation significantly alters the chemical step of hydride transfer (k_{hyd} , Figure 3B and Table 2), which was assessed through pre-steady-state kinetics as originally described by Farnum et al. (1991) and more recently for our studies with BLVRB (Duff et al., 2020).

TABLE 2 | Individual catalytic parameters for BLVRB mutations.

Human BLVRB	k_{hyd}^a (s^{-1})	K_M^b for FAD (μM)	k_{cat}^c (s^{-1})
WT ^c	28 ± 5	151 ± 18	0.068 ± 0.004
S111A	33 ± 8	325 ± 96	0.021 ± 0.004
H153A	30 ± 7	98 ± 11	0.059 ± 0.003
T164A	36 ± 7	153 ± 18	0.063 ± 0.002
T164I	33 ± 12	156 ± 18	0.049 ± 0.003
T164S	34 ± 5	127 ± 15	0.039 ± 0.003

^aHydride transfer rates were determined from pre-steady-state UV-kinetics.

^bMichaelis-Menten constants and turnover-rates were determined from steady-state UV-kinetics.

^cInitially reported in Duff et al. (2020).

Thus, while H153A and S111A mutations have been shown to modulate function through variations in coenzyme or substrate binding affinities, respectively (Smith et al., 2008; Chu et al., 2017), we show here that they do not affect the chemical step. The reason that we emphasize this is that our previous application of the RASSMM approach identified a distally dynamic site within the proline isomerase cyclophilin-A that also altered modulated substrate affinity but did not appear to modulate the chemical step (Holliday et al., 2017).

Both BLVRB mutations S111A and H153A impart large-scale changes, as monitored through CSPs relative to the WT BLVRB (Figures 3C,D). Interestingly, much of the active site that we have previously shown undergoes conformational exchange in human wild type (WT) BLVRB, exhibits a complete loss in signal within the S111A mutation that includes residues 109–137 (Figure 3C). Thus, while we have previously shown that much of the active site of WT BLVRB exhibits an inherent chemical exchange on the fast timescale through R2-CPMG relaxation dispersion (Paukovich et al., 2018), S111A shifts the timescale of this regime from a fast to an intermediate exchange. This means that S111A is globally coupled through dynamics in addition to structural perturbations monitored through CSPs. The H153A mutation also imparts distal changes (Figure 3D). Although the CSPs induced by H153A are smaller than those induced by S111A, many of these distally coupled residues are similar. Specifically, the resonance of T164 is perturbed in the context of both mutations (Figures 3C,D), indicating that these active site residues are both coupled to T164. Thus, in addition to its inherent dynamics, T164 further fulfills the criteria of the RASSMM approach in that it is also coupled to the active site.

Dynamic Differences between Human and Lemur BLVRBs Include Position 164

While identifying human BLVRB T164 as an ideal site to mutate for the RASSMM approach, we also noticed that this position is evolutionarily dynamic and realized that evolution may help serve as a guide for the specific selection of residues to mutate. Specifically, a sampling of mammalian sequences at sequential branch points according to Hallstrom and Janke (2010) reveals that position 164 toggles between a threonine and serine (Figure 4A). Thus, prior to specifically mutating human T164 in order to address its global changes through the RASSMM approach, we first sought

evolutionary insight by comparing the dynamics between homologues that differ in this position. We used lemur BLVRB that comprises a serine at position 164 (S164) instead of a threonine as its human counterpart (T164). Our previous studies revealed that the lemur BLVRB exhibits similar fast timescale dynamics to human BLVRB identified via R1 relaxation and a nearly identical holo enzyme structure (Figure 4B; Duff et al., 2020). However, we have previously discovered that a single mutation to cyclophilin-A can induce global dynamic changes to slower μs -ms timescale motions that modulate function (Doshi et al., 2016; Holliday et al., 2017). As lemur BLVRB comprises 16 substitutions within its 206 residues relative to human BLVRB that includes position 164, which is inherently dynamic and coupled to the active site within human BLVRB (Figure 4C), we sought here to compare the μs -ms dynamics between the two homologues.

Despite similar fast timescale motions and X-ray crystal structures, R2 relaxation measured here reveals that human and lemur BLVRBs differ in their inherent μs -ms timescale motions. Specifically, R2-CPMG dispersions were collected for lemur BLVRB and compared to human BLVRB (Supplementary Figure S1). Overall differences are illustrated by comparing R2 relaxation rates from the lowest R2-CPMG refocusing field of 50 Hz that comprises the highest contribution from exchange (Figure 4D). One of the most dramatic differences to the inherent motions include a near complete quenching of R2s within lemur BLVRB S164 relative to human BLVRB T164 that is discussed further below. The localized nature of motions can also be observed within the R2-CPMG dispersion profiles themselves that are a mixture between fast movements that give rise to linear dispersions and slower motions that give rise to sigmoidal profiles (Supplementary Figure S1). Even neighboring residues often exhibit a complicated mixture of differential dynamics that suggests caution in accurately extracting the explicit exchange rates when such nuclei are likely sensing multiple chemical exchange phenomena. However, single nuclei were fit in order to provide an estimate of their exchange rates that are mapped onto their respective the X-ray crystal structures of their holo forms (Figures 4E,F). For example, several exchange phenomena of neighboring residues within the active site exhibit similar motions in both BLVRB homologues, such as C109/S111, I133/M135, and residues 125–127 that result in similar kinetic rates for both homologues of $\sim 1500 \text{ s}^{-1}$ (Supplementary Figure S1). Despite the similar exchange for residues 125–127 that comprise the substrate binding site, residues 124–126 appear to have two exchange phenomena in lemur BLVRB. Specifically, residues 124–126 have an R2-CPMG field dependence that indicate μs -ms timescale motions but retain a high R2 at the highest imparted CPMG field that indicates an additional faster exchange process within the μs timescale (Supplementary Figure S1). These findings alone are meaningful, as they reveal a shift to an additional exchange phenomenon within lemur BLVRB relative to that of human BLVRB. The selection of residues chosen to globally fit are described in later sections upon applying the RASSMM approach to site 164 that identifies potentially coupled motions. However, it is important to point out here that similar fast timescale motions with similar structures across homologues but with very different μs -ms timescale motions have been reported for the cyclophilin-A family

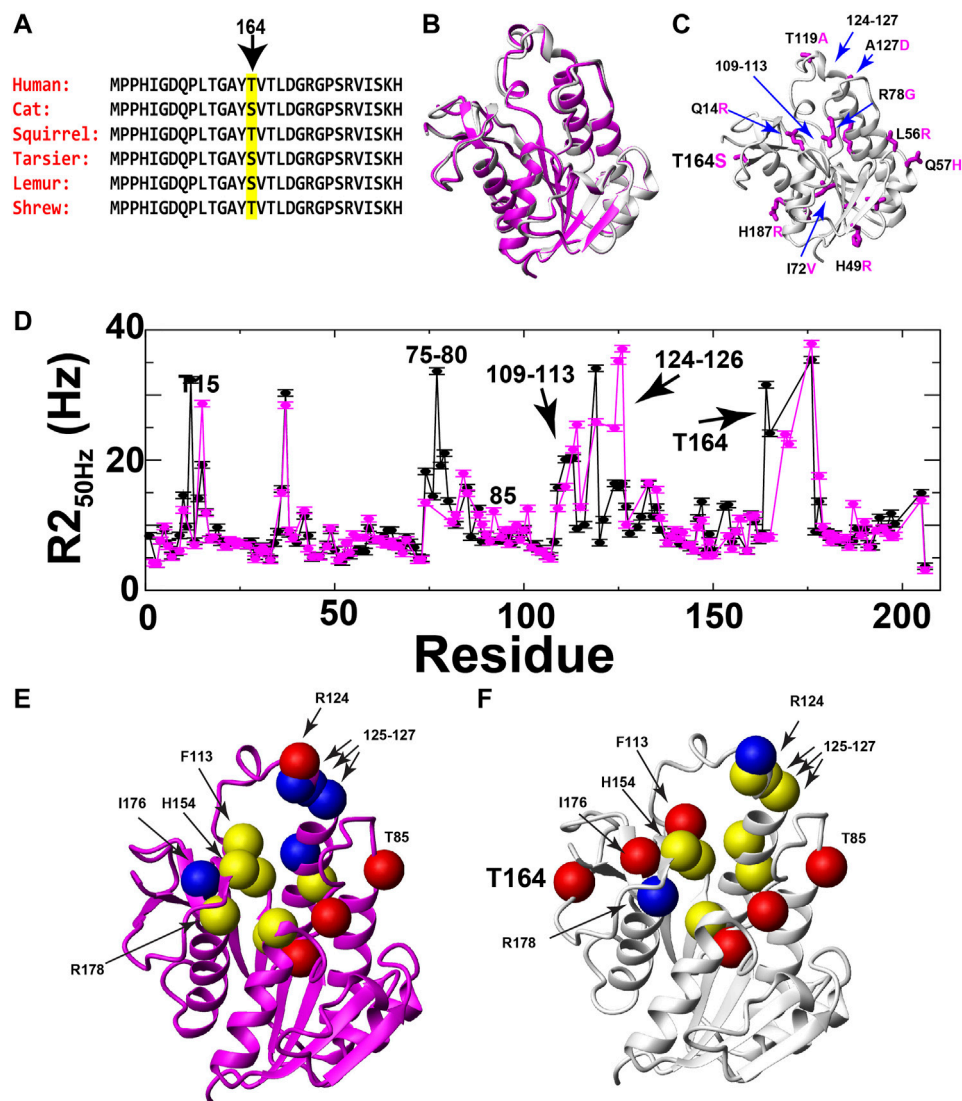


FIGURE 4 | Evolutionary changes of BLVRB position 164 and comparative dynamics of human and lemur BLVRB family members. **(A)** Sequence comparison of six mammalian BLVRBs for residues 151–169 shown from top (most recent) to bottom as a function of their most recent evolutionary branch point as described by Hallstrom and Janke (2010). Position 164 is highlighted (yellow). **(B)** Superposition of human and lemur BLVRBs (PDB accession numbers 1HDO and 6OQG, respectively). **(C)** Lemur BLVRB comprises 16 amino acid changes mapped onto human BLVRB (magenta) with several of these amino acids highlighted that include position 164 (human in black, lemur in magenta). **(D)** R2 relaxation rate at 50 Hz CPMG refocusing for human (black) and lemur (magenta) as a function of residue. **(E)** Individually extracted rates of exchange (k_{ex}) for lemur BLVRB amides are shown as balls for rates less than 1000 s^{-1} (blue), from 1000 to 3000 s^{-1} (yellow), and higher than 3000 s^{-1} (red). **(F)** Individually extracted rates of exchange (k_{ex}) for the same residues for human BLVRB amides are shown as balls for rates less than 1000 s^{-1} (blue), from 1000 to 3000 s^{-1} (yellow), and higher than 3000 s^{-1} (red). Only residue human T15 is not shown as the exchange contribution was too low to fit (Supplementary Figure S1).

just as we report here for BLVRB family members (Holliday et al., 2015a).

Potentially the most striking change in R2 relaxation identified here is that measured at position 164 where exchange is observed within human BLVRB T164 yet is completely quenched in the context of lemur BLVRB S164 (Figures 4E,F; Supplementary Figure S1). There are no evolutionary substitutions within close proximity to position 164 between these two homologues (Figure 4C), suggesting that this is a true change to the inherent dynamics of this region caused by its substitution.

Evolutionarily Guided Mutations to Human BLVRB T164 Incur the Largest Functional Impact and Suggests That Coenzyme Binding is Coupled to the Rate-Limiting Step

As the first step of the RASSMM approach identified human BLVRB T164 as both inherently dynamic and coupled to the active site, we next sought to address the second step of the RASSMM approach that imparts multiple mutations to this single site. While we engineered a

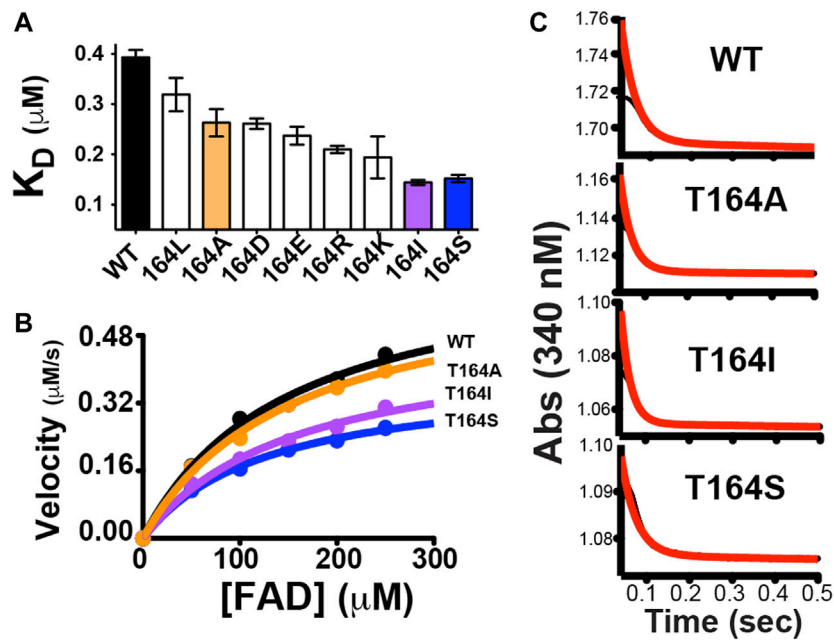


FIGURE 5 | Functional changes induced by point mutations to human BLVRB T164. **(A)** A panel of mutations was screened for their dissociation constants (K_D) to NADP⁺ using ITC, as previously described. Mutations that were selected for further analysis are colored for WT BLVRB (black), T164A (orange), T164I (purple), and T164S (blue). **(B)** Steady-state kinetics for BLVRB WT (black), T164A (orange), T164I (purple), and T164S (blue). **(C)** Pre-steady-state kinetics BLVRB WT, T164A, T164I, and T164S at the lowest FAD concentration used.

panel of mutations to this site that include hydrophobic residues and hydrophilic residues, evolutionarily guided mutations were also considered based on our findings above with lemur BLVRB. Specifically, the conservative mutation of T164→S was also engineered based on the evolutionary comparisons within mammalian BLVRBs. We first probed the panel of BLVRB single-site mutants for their role in coenzyme binding using isothermal titration calorimetry (ITC). Surprisingly, one of the largest imparted changes to coenzyme affinity that incurred an approximately 2-fold increase was for that of the conservative mutation of T164S (Figure 5A). The other mutation with a similar functional impact on coenzyme binding is T164I. Interestingly, an isoleucine is found within multiple insect species that include *Aedes aegypti* BLVRB (NCBI Reference Sequence: XP_001649677.1), although insect BLVRBs are dramatically different from their mammalian counterparts in primary sequence that make it more difficult to pinpoint the role of specific evolutionary changes (<50% sequence identity). Nonetheless, these binding studies immediately identify an allosteric role for position 164 within human BLVRB with the most dramatic changes to coenzyme affinity also being identical to evolutionary changes.

In order to further examine the functional role of position 164 and simultaneously use these allosteric changes to potentially help address the rate-limiting step of BLVRB catalysis, we selected a subset of mutations that represent a minimum change in coenzyme affinity (T164A) and a larger change in coenzyme affinity (T164I, T164S). Because the coenzyme binds to BLVRB several orders of magnitude tighter than the flavin substrates, the rate-limiting step has been presumed to be coenzyme binding (Paukovich et al., 2018).

This is supported here by mutations at the T164 site using steady-state kinetics of substrate turnover (Figure 5B). Specifically, substrate turnover was similar between WT BLVRB and the T164A mutant, yet turnover was significantly diminished for BLVRB, T164I, and T164S. Considering that the substrate K_M values were similar for all T164 mutations (Table 2), this suggests that the tighter binding of the coenzyme (and likely a slower off-rate) is rate-limiting for substrate turnover. Mutations to T164 do not measurably alter the chemical step of hydride transfer, as monitored through pre-steady-state kinetics (Figure 5C), thereby further narrowing the role of allosteric coupling through T164 to the changes monitored in its coenzyme interactions. Thus, analogous to our first application of RASSMM where distal mutations were coupled to substrate binding (Holliday et al., 2017), a similar phenomenon is identified here where distal mutations are coupled to coenzyme binding in BLVRB.

Mutations to Human BLVRB T164 Modulate Global Dynamics

At the heart of the RASSMM approach is the potential for identifying similar relaxation effects to multiple residues imparted by variations at the single mutation site (i.e., networks), such as those imparted by mutations to human BLVRB T164. Thus, we once again chose to further probe human BLVRB T164A (minimum functional change) and both BLVRB, T164I, and T164S (maximum functional changes). CSPs can also provide a means to identify coupled networks, yet many of the perturbed residues are similar for all three mutations with variations primarily to their magnitudes (Figures 6A–C).

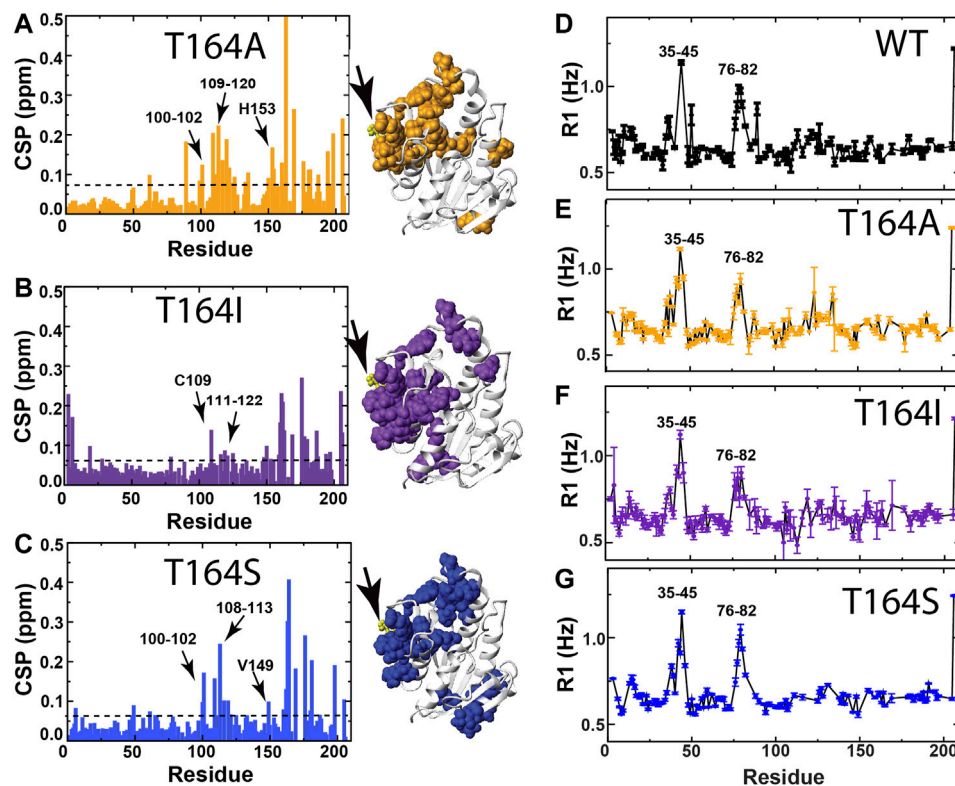


FIGURE 6 | CSPs and R1 relaxation rates of BLVRB T164 mutants. **(A)** CSPs between apo BLVRB WT and apo BLVRB T164A with the line indicating 0.08 ppm that is the average plus $\frac{1}{2}$ of the standard deviation. **(B)** CSPs between apo BLVRB WT and apo BLVRB T164I with the line indicating 0.07 ppm that is the average plus $\frac{1}{2}$ of the standard deviation. **(C)** CSPs between apo BLVRB WT and apo BLVRB T164S with the line indicating 0.07 ppm that is the average plus $\frac{1}{2}$ of the standard deviation. **(D)** R1 relaxation rates for BLVRB WT previously published (Paukovich et al., 2018). **(E)** R1 relaxation rates for BLVRB T164A. **(F)** R1 relaxation rates for BLVRB T164I. **(G)** R1 relaxation rates for BLVRB T164S. All CSPs above the line are mapped onto the X-ray crystal structure (PDB accession 1HDO) and all data were collected at 900 MHz at 20°C.

Furthermore, CSPs induced by T164 mutations do not appear to induce similar changes to that of coenzyme binding. For example, we employed an NMR chemical shift projection analysis (CHESPA) that has been successfully used to monitor global changes that reflect conformational shifts to sampling of active states, which includes changes induced by mutations (Selvaratnam et al., 2012a; Selvaratnam et al., 2012b; Gagne et al., 2015). Specifically, by comparing CSPs induced upon human WT BLVRB binding to the coenzyme to CSPs induced by the T164 mutations (Supplementary Table S2), only a subset of residues largely localized near position 164 within the C-terminal lobe exhibited any covariance at all (Supplementary Table S3). Interestingly, there is a negative correlation for several residues within the T164A active site (Supplementary Table S3, residues 119–120 and H153). This could theoretically suggest that dynamic changes imparted by T164 mutations for T164S and T164I that increase coenzyme affinity may be negated by structural changes within T164A. However, overall there is no clear direct relationship between function induced by T164 mutations and CSPs. Additionally, R1 relaxation rates are largely similar for these three mutations that indicate similar ps–ns motions as WT BLVRB (Figures 6D–G). For example, locally elevated R1 relaxation rates for residues 35–45 and 76–82 are consistent to that previously reported for human and lemur BLVRBs (Paukovich et al.,

2018; Duff et al., 2020 site). Thus, we next focused on μ s–ms timescale motions using R2-CPMG dispersion that has previously proven successful in identifying coupled networks within cyclophilin-A (Holliday et al., 2017).

All three single site mutations induce changes to R2-CPMG dispersion profiles for residues that primarily surround the active site (Figure 7 and Supplementary Figure S2 for full R2 CPMG dispersions). Most interestingly, the T164S mutation quenches exchange at position 164 itself (Figure 7, left), which is exactly what is observed in lemur BLVRB that comprises an endogenous S164 (Figure 4D; Supplementary Figure S1). Overall, induced changes can broadly be separated into two groups. The first group includes largely localized changes with no clear patterns induced by each mutation (Figure 7, colored green and Supplementary Figure S2A that also illustrates several sites that do not change). For example, while active site residues such as T110, S111, F113, and H153 are adjacent to each other, each T164 mutation induces different changes to all of their R2-CPMG dispersion profiles. The second group includes several residues on both sides of the active site that exhibit similar patterns for each mutation (Figure 7, colored red and Supplementary Figure S2B). These include residues 125–127, H132, K178, and Y205 that exhibit similar profiles for both T164A and WT BLVRB but have higher

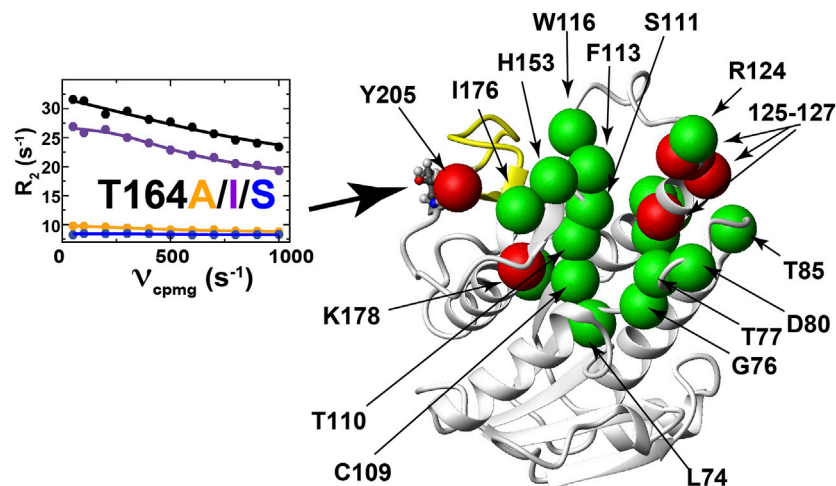


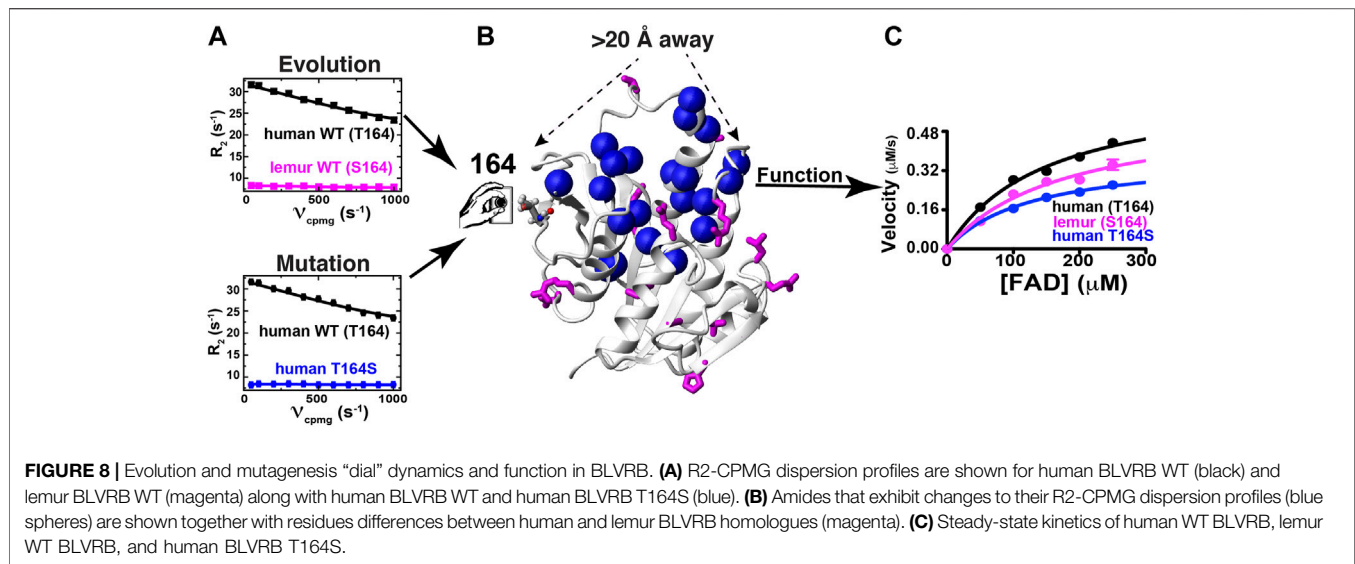
FIGURE 7 | Mutations to BLVRB T164 induce long-range dynamic changes. R₂-CPMG dispersion profiles for BLVRB WT (black), T164A (orange), T164I (purple), and T164S (blue) at position 164 are shown with the amide of distal changes mapped onto the X-ray crystal structure shown as either green spheres (R₂-CPMG profiles that are differentially altered upon mutation to T164 along with S111 that is not differentially perturbed) or red spheres (R₂-CPMG profiles that exhibit similar responses upon mutation to T164).

amplitudes for T164I and lower amplitudes for T164S. While these amides of this second group are located on the periphery of the active site (**Figure 7**, red), it is important to note that their conformational dynamics could be coupled through side chain dynamics that include R124, F113, and H153. Thus, the RASSM approach has identified imparted changes *via* position 164 that includes similar changes to this second group of residues.

Considering that both T164I and T164S mutants impact function with opposing effects to R₂-CPMG dispersions, such findings reveal the localized nature (or segmental nature) of dynamics within this μs-ms timescale that has been shown to dominate dynamics in many systems that are still partially coupled (McDonald et al., 2012; Holliday et al., 2017). The similar perturbation patterns induced for the second group of R₂-CPMG dispersions could theoretically be used as a rationale for globally fitting their R₂-CPMG dispersions. Global fits of these residues or a subset of these residues of 125–127 indicate that this group moves slightly faster within both T164I and T164S relative to the WT BLVRB, T164A, and lemur BLVRB (**Supplementary Table S4**). It is plausible that faster motions may underlie the higher affinity of T164I and T164S for the coenzyme by driving sampling of binding competent conformations. However, further ensemble-based approaches to the panel of mutations produced here will likely be necessary to address the explicit mechanism by which these alterations to conformational sampling leads to higher coenzyme affinity. Finally, it is also important to note that a challenge in identifying networks in human BLVRB through changes to R₂-CPMG dispersions is that several key residues surrounding the T164 site are simply not observed, which includes residues 167–175 and 199–204 (**Figure 7**, yellow ribbon). However, based on the imparted changes that can be measured, it is clear that while dynamics are coupled over large distances, they are also largely localized for many residues.

DISCUSSION

The dynamic basis of enzyme catalysis is still an emerging field that has benefited from the advent of NMR relaxation studies such as those used here to begin identifying allosterically coupled networks within BLVRB. Allosteric regulation of enzyme function is increasingly recognized to occur through multiple mechanisms, which include mechanical coupling such as the classic case of hemoglobin but also dynamic coupling where motions influence networks of partially coupled movements. Unfortunately, the basis of long-range allosteric regulation is complicated owing to the nature of motions that are often segmental. For example, even after decades of studies on DHFR, studies are still revealing how global motions modulate this enzyme's function (Boehr et al., 2006; Mauldin et al., 2012; Singh et al., 2015). Here, we have discovered that the evolutionarily changing position of 164 within BLVRBs also modulates function by a comparative analysis that includes multiple BLVRB homologues (i.e., human and lemur BLVRBs) and the application of the RASSM approach that has identified both functional and dynamic changes. Strikingly, whether it is evolution or mutagenesis that positions a serine at position 164, the specific changes to dynamics at this site are identically quenched (**Figure 8A**). However, the segmental nature of motions identified within other enzymes is also evident in BLVRB (McDonald et al., 2012; Holliday et al., 2017), as monitored through the differential effects to R₂-CPMG dispersions for different mutations to human BLVRB T164. None-the-less, networks of coupled motions such as those for human BLVRB T164S can be measured all the way to the active site (**Figure 8B**) and impart a 2-fold change to turnover (**Figure 8C**). Considering that mutations to active site residues such as BLVRB S111 impart approximately a 3-fold change to substrate turnover that dictates hematopoietic cell fate (Wu et al.,



2016; Chu et al., 2017), the fact that catalysis can be modulated distally by a 2-fold change is significant. Our studies here therefore reveal that the dynamics of BLVRB can fine-tune function *via* dynamic allostery.

While the use here of multiple BLVRB homologues and multiple human BLVRB mutants illustrate how enzyme function may be “dialed” through allostery (Figure 8), there remain multiple questions that must be further addressed to understand the underlying mechanisms. For example, the dynamic effects imparted by human BLVRB T164 mutants are similar for residues on opposite sides of the active site and thus, side chain dynamics may be important to determine whether these serve to bridge dynamics. Most of these active site residues are aromatics, so the use of ^{13}C -aromatic R2-CPMG experiments may be important to identify their differences (Weininger et al., 2012; Kasinath et al., 2013; Raum et al., 2018). Future studies that probe potential shifts to faster μs timescales using R1rho-type experiments may also reveal how these mutations allosterically modulate dynamics in the holo forms as well and how these may be related to faster timescale dynamics within the apo forms. Swapping evolutionarily changing sites identified *via* RASSM such as those at position 164 within other homologues and monitoring their effects on dynamics and catalytic function will also be important to understand how such networks have evolved beyond human BLVRB. Finally, ensemble-based structural studies of these homologous and mutants may be especially informative to address the physical basis of allosteric regulation. While solution-based approaches would require an enormous amount of data, such as resonance assignments and NOEs for each mutant, newly developed X-ray crystallography approaches that directly visualize these ensembles may offer an alternative means (Keedy et al., 2018; Yabukarski et al., 2020).

In regard to BLVRB function, the human T164 mutations engineered here provides evidence that coenzyme-release is coupled to the rate-limiting step. Specifically, recent attempts to identify the rate-limiting step for BLVRB have been hampered by the complexity of ternary complex formation and a uniquely evolving

active site. For example, coenzyme-release can be controlled *via* evolutionarily guided substitutions to the BLVRB active site, yet the FAD substrate appears to slow coenzyme-release in a manner that is dependent on these very same active site substitutions (Duff et al., 2020). In contrast, mutations to T164 here have allowed us to allosterically modulate function without directly perturbing the active site through. Specifically, as the coenzyme affinity becomes tighter for human BLVRB T164 mutations (Figure 5A), FAD substrate turnover is slowed (Figure 5B), which suggests that coenzyme binding is either the rate-limiting step or coupled to the rate-limiting step. Thus, we show here that allosterically modulating enzyme function may provide a general means for probing catalytic mechanism.

MATERIALS AND METHODS

Protein Expression and Purification

All BLVRB proteins were purified as previously described (Paukovich et al., 2018; Duff et al., 2020). Briefly, all sequences encoding human BLVRB (NCBI Reference Sequence: NP_000704.1) and lemur BLVRB (NCBI Reference Sequence: XP_020138941.1) were codon optimized and synthesized as a Gibson block (Integrated DNA Technologies, Inc.,). For insertion into pET21b. All human BLVRB mutants were also constructed the same way. As human BLVRB T164S did not express in pET21b, it was recloned into pJ401k for expression. All constructs encoded an N-terminal 6xHis tag and thrombin cleavage site. Unlabeled proteins were grown in luria broth (LB) while labeled proteins were grown in M9 minimal media (6 g/L Na_2HPO_4 , 3 g/L KH_2PO_4 , 0.5 g/L NaCl, 1 g/L NH_4Cl , 2 g/L glucose, 2 ml of 1 M MgSO_4 , 100 ml of 1 M NaCl CaCl_2 , 10 mg/L thiamine). ^{15}N -labeled proteins were grown in M9 media with ^{15}N ammonium chloride and ^2H , ^{15}N -labeled proteins grown in 100% D_2O with ^{15}N ammonium chloride. Cell growths were induced with 0.1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG). All proteins were purified in denaturing buffer (5 M guanidine, 50 mM Na_2HPO_4 , pH 7.5, 500 mM NaCl, 10 mM imidazole) *via* Ni affinity (Sigma) and

followed by application to a 3 ml Resource 15RPC (GE Healthcare Life Sciences) to remove any residual coenzyme prior to refolding through dialysis into 1 M arginine buffer (100 mM Tris, pH 7.5, 100 mM NaCl, 1 mM DTT). Finally, proteins were dialyzed to NMR buffer (50 mM bis-Tris, pH 6.5, 50 mM NaCl) and concentrated proteins were cleaved with thrombin (Sigma) to remove the 6x-His tag and applied to a Superdex-75 (GE). Proteins were concentrated and frozen until further use.

Thermodynamic and Kinetic Analysis

Isothermal titration calorimetry (ITC), steady-state, and pre-steady state experiments for BLVRB were performed identically to previous studies (Paukovich et al., 2018; Duff et al., 2020). Briefly, a MicroCal VP-ITC was used with samples containing 100 mM enzyme and NADP⁺ as the titrant at 1 mM with all buffers identical to that used for NMR (50 mM bis-Tris, pH 6.5, 50 mM NaCl). ITC experiments were performed in duplicate at 20°C and processed using Origin software provided with the MicroCal VP-ITC. All reported values are the average and standard deviations of these duplicates. For UV-steady state kinetics that monitored initial velocities, a Biotek Synergy 2 multi-mode detection plate reader was used with 200 µl total volumes (pathlength of 0.625 cm), which monitored NADPH conversion to NADP⁺ at 340 nm, as previously described (Cunningham et al., 2000). Initial velocities were fit to the Michaelis-Menten equation using GraphPad Prism version 4.0 (GraphPad Software Inc, San Diego, CA). Velocities corrections to mM*s⁻¹ were calculated using the extinction coefficient of NADPH of 6,222 M⁻¹*cm⁻¹. Pre-steady-state kinetics experiments were performed using an SX20 model Applied PhotoPhysics stopped-flow with 1 µs dead time. The pre-steady-state burst was fit to an exponential decay (the pre-steady-state rate of hydride transfer, k_{hyd}) and a linear equation (the steady-state rate) using 3 FAD concentrations of 0.5, 1.0, and 1.5 mM using GraphPad Prism. As hydride transfer rates are concentration-independent (Fierke et al., 1987; Maglia et al., 2003), reported hydride transfer rates were averaged between these FAD concentrations with uncertainties calculated as their standard deviations.

NMR Spectroscopy and Data Analysis

All BLVRB samples were prepared in 50 mM bis-Tris, pH 6.5, 50 mM NaCl, 1 mM DTT at 500 mM enzyme with 5% D₂O and data were collected on a Varian 900 spectrometer equipped with a cryo-probe at 20°C. Resonance assignments for the WT BLVRB previously published were used (Paukovich et al., 2018) and amide assignments for T164 mutants were confirmed via 3D¹⁵N-NOESY-HSQC. Spectra were processed using NMRPipe (Delaglio et al., 1995) and all data were analyzed using CCPNmr software (Vranken et al., 2005). R2-CPMG dispersions were collected on ²H,¹⁵N-labeled proteins with TROSY selection as previously described (Schlegel et al., 2009; Holliday et al., 2017) and dispersions were fit to the full Carver-Richards equation using CPMG_fit as previously described (Schlegel et al., 2009). R1 relaxation experiments were collected on ¹⁵N-labeled proteins using the standard BioPack sequence with 0.01, 0.1, 0.3, 0.5, 0.7, 0.9, and 1.1 s delays with R1 relaxation rates calculated in CCPNmr.

NMR Chemical shift projection analysis (CHESPA) was used to probe for potential covariance between human BLVRB T164

mutations and WT BLVRB coenzyme binding, as previously described (Selvaratnam et al., 2012a; Selvaratnam et al., 2012b; Gagne et al., 2015). Briefly, amide chemical shifts were exported from CCPNmr (**Supplementary Table S2**) and the cosine angles were calculated between CSPs induced for human WT BLVRB apo→holo previously published (BMRB accession 27,462 and 27,463) and each apo→T164 mutation described in this study (**Supplementary Table S3**).

NMR Ensemble Calculations

Apo BLVRB ensembles were calculated with chemical shifts previously obtained (Paukovich et al., 2018) and supplemented with NOEs for structure determination using Resolution Adapted Structural REcombination (RASREC) in CS-Rosetta (Lange and Baker, 2012), as we have previously described (Holliday et al., 2015b). CCPNmr was used for spectral analysis with chemical shifts and NOEs exported for CS-Rosetta calculations. NOEs were derived from ¹⁵N-NOESY and ¹³C-NOESY spectra (**Supplementary Table S1**). Co-evolution restraints were also used as calculated from the GREMLIN server using the human BLVRB sequence (Kamisetty et al., 2013). Fragment libraries for CS-Rosetta were calculated without any homologues in order to avoid further biasing with final statistics of the 10 lowest energy structures shown in **Table 1**.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JR and EE cloned, expressed, and purified all proteins and processed all NMR data. MD and PA collected all kinetics data for hydride transfer experiments. TP helped collect kinetics data for substrate turnover. AB helped collect NMR data.

FUNDING

PA was supported by application number GM105978. EE was supported by NSF application number 1807326, NIH application number R56 CA230069, and NIH application number, R21AI146295. The National High Magnetic Field Laboratory is supported by National Science Foundation through NSF/DMR-1644779 and the State of Florida.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.691208/full#supplementary-material>

REFERENCES

- Adak, S., and Begley, T. P. (2017). Flavin-N5-oxide: A New, Catalytic Motif in Flavoenzymology. *Arch. Biochem. Biophys.* 632, 4–10. doi:10.1016/j.abb.2017.08.001
- Alderson, T. R., and Kay, L. E. (2020). Unveiling Invisible Protein States with NMR Spectroscopy. *Curr. Opin. Struct. Biol.* 60, 39–49. doi:10.1016/j.sbi.2019.10.008
- Becker, D. F., Zhu, W., and Moxley, M. A. (2011). Flavin Redox Switching of Protein Functions. *Antioxid. Redox Signaling* 14, 1079–1091. doi:10.1089/ars.2010.3417
- Bhabha, G., Ekiert, D. C., Jennewein, M., Zmasek, C. M., Tuttle, L. M., Kroon, G., et al. (2013). Divergent Evolution of Protein Conformational Dynamics in Dihydrofolate Reductase. *Nat. Struct. Mol. Biol.* 20, 1243–1249. doi:10.1038/nsmb.2676
- Boehr, D. D., McElheny, D., Dyson, H. J., and Wright, P. E. (2006). The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis. *Science* 313, 1638–1642. doi:10.1126/science.1130258
- Chu, W.-T., Nesbitt, N. M., Gnatenko, D. V., Li, Z., Zhang, B., Seeliger, M. A., et al. (2017). Enzymatic Activity and Thermodynamic Stability of Biliverdin IX β Reductase Are Maintained by an Active Site Serine. *Chem. Eur. J.* 23, 1891–1900. doi:10.1002/chem.201604517
- Cunningham, O., Gore, M. G., and Mantle, T. J. (2000). Initial-rate Kinetics of the Flavin Reductase Reaction Catalysed by Human Biliverdin-IX β Reductase (BVR-B). *Biochem. J.* 345, 393–399. doi:10.1042/0264-6021:3450393
- Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* 6, 277–293. doi:10.1007/bf00197809
- Doshi, U., Holliday, M. J., Eisenmesser, E. Z., and Hamelberg, D. (2016). Dynamical Network of Residue-Residue Contacts Reveals Coupled Allosteric Effects in Recognition, Catalysis, and Mutation. *Proc. Natl. Acad. Sci. U.S.A.* 113, 4735–4740. doi:10.1073/pnas.1523573113
- Doucet, N. (2011). Can Enzyme Engineering Benefit from the Modulation of Protein Motions? Lessons Learned from NMR Relaxation Dispersion Experiments. *Ppl* 18, 336–343. doi:10.2174/092986611794653950
- Duff, M. R., Borreguero, J. M., Cuneo, M. J., Ramanathan, A., He, J., Kamath, G., et al. (2018). Modulating Enzyme Activity by Altering Protein Dynamics with Solvent. *Biochemistry* 57, 4263–4275. doi:10.1021/acs.biochem.8b00424
- Duff, M., Redzic, J., Ryan, L., Paukovich, N., Zhao, R., Nix, J., et al. (2020). Structure, Dynamics, and Function of the Evolutionarily Changing Biliverdin. *J. Biochem.* 168, 191–202. doi:10.1093/jb/mvaa039
- Farnum, M. F., Magde, D., Howell, E. E., Hirai, J. T., Warren, M. S., Grimsley, J. K., et al. (1991). Analysis of Hydride Transfer and Cofactor Fluorescence Decay in Mutants of Dihydrofolate Reductase: Possible Evidence for Participation of Enzyme Molecular Motions in Catalysis. *Biochemistry* 30, 11567–11579. doi:10.1021/bi00113a012
- Fierke, C. A., Johnson, K. A., and Benkovic, S. J. (1987). Construction and Evaluation of the Kinetic Scheme Associated with Dihydrofolate Reductase from *Escherichia coli*. *Biochemistry* 26, 4085–4092. doi:10.1021/bi00387a052
- Gagné, D., Narayanan, C., and Doucet, N. (2015). Network of Long-Range Concerted Chemical Shift Displacements upon Ligand Binding to Human Angiogenin. *Protein Sci.* 24, 525–533. doi:10.1002/pro.2613
- Hallstrom, B. M., and Janke, A. (2010). Mammalian Evolution May Not Be Strictly Bifurcating. *Mol. Biol. Evol.* 27, 2804–2816. doi:10.1093/molbev/msq166
- Holliday, M. J., Armstrong, G. S., and Eisenmesser, E. Z. (2015a). Determination of the Full Catalytic Cycle Among Multiple Cyclophilin Family Members and Limitations on the Application of CPMG-RD in Reversible Catalytic Systems. *Biochemistry* 54, 5815–5827. doi:10.1021/acs.biochem.5b00746
- Holliday, M. J., Camilloni, C., Armstrong, G. S., Isern, N. G., Zhang, F., Vendruscolo, M., et al. (2015b). Structure and Dynamics of GeoCyp: A Thermophilic Cyclophilin with a Novel Substrate Binding Mechanism that Functions Efficiently at Low Temperatures. *Biochemistry* 54, 3207–3217. doi:10.1021/acs.biochem.5b00263
- Holliday, M. J., Camilloni, C., Armstrong, G. S., Vendruscolo, M., and Eisenmesser, E. Z. (2017). Networks of Dynamic Allostery Regulate Enzyme Function. *Structure* 25, 276–286. doi:10.1016/j.str.2016.12.003
- Hughes, R. L., Johnson, L. A., Behiry, E. M., Loveridge, E. J., and Allemann, R. K. (2017). A Rapid Analysis of Variations in Conformational Behavior during Dihydrofolate Reductase Catalysis. *Biochemistry* 56, 2126–2133. doi:10.1021/acs.biochem.7b00045
- Huijbers, M. M. E., Montersino, S., Westphal, A. H., Tischler, D., and van Berkel, W. J. H. (2014). Flavin Dependent Monooxygenases. *Arch. Biochem. Biophys.* 544, 2–17. doi:10.1016/j.abb.2013.12.005
- Kamisetty, H., Ovchinnikov, S., and Baker, D. (2013). Assessing the Utility of Coevolution-Based Residue-Residue Contact Predictions in a Sequence- and Structure-Rich Era (Vol 110, Pg 15674, 2013). *Proc. Natl. Acad. Sci. United States America* 110, 18734. doi:10.1073/pnas.1319550110
- Kasinath, V., Valentine, K. G., and Wand, A. J. (2013). A ¹³C Labeling Strategy Reveals a Range of Aromatic Side Chain Motion in Calmodulin. *J. Am. Chem. Soc.* 135, 9560–9563. doi:10.1021/ja4001129
- Keedy, D. A., Hill, Z. B., Biel, J. T., Kang, E., Rettenmaier, T. J., Brandão-Neto, J., et al. (2018). An Expanded Allosteric Network in PTP1B by Multitemperature Crystallography, Fragment Screening, and Covalent Tethering. *Elife* 7, e36307. doi:10.7554/eLife.36307
- Kendrick, A. A., Holliday, M. J., Isern, N. G., Zhang, F., Camilloni, C., Huynh, C., et al. (2014). The Dynamics of Interleukin-8 and its Interaction with Human CXCR1 Receptor I Peptide. *Protein Sci.* 23, 464–480. doi:10.1002/pro.2430
- Kleckner, I. R., and Foster, M. P. (2011). An Introduction to NMR-Based Approaches for Measuring Protein Dynamics. *Biochim. Biophys. Acta (Bba) - Proteins Proteomics* 1814, 942–968. doi:10.1016/j.bbapap.2010.10.012
- Lange, O. F., and Baker, D. (2012). Resolution-adapted Recombination of Structural Features Significantly Improves Sampling in Restraint-guided Structure Calculation. *Proteins* 80, 884–895. doi:10.1002/prot.23245
- Lange, O. F., Rossi, P., Sgourakis, N. G., Song, Y., Lee, H.-W., Aramini, J. M., et al. (2012). Determination of Solution Structures of Proteins up to 40 kDa Using CS-Rosetta with Sparse NMR Data from Deuterated Samples. *Proc. Natl. Acad. Sci.* 109, 10873–10878. doi:10.1073/pnas.1203013109
- Liu, C. T., Francis, K., Layfield, J. P., Huang, X., Hammes-Schiffer, S., Kohen, A., et al. (2014). *Escherichia Coli* Dihydrofolate Reductase Catalyzed Proton and Hydride Transfers: Temporal Order and the Roles of Asp27 and Tyr100. *Proc. Natl. Acad. Sci. USA* 111, 18231–18236. doi:10.1073/pnas.1415940111
- Maglia, G., Javed, M. H., and Allemann, R. K. (2003). Hydride Transfer during Catalysis by Dihydrofolate Reductase from *Thermotoga Maritima*. *Biochem. J.* 374, 529–535. doi:10.1042/bj20030412
- Maiti, R., Van Domselaar, G. H., and Wishart, D. S. (2005). MovieMaker: a Web Server for Rapid Rendering of Protein Motions and Interactions. *Nucleic Acids Res.* 33, W358–W362. doi:10.1093/nar/gki485
- Mauldin, R. V., Sapienza, P. J., Petit, C. M., and Lee, A. L. (2012). Structure and Dynamics of the G121V Dihydrofolate Reductase Mutant: Lessons from a Transition-State Inhibitor Complex. *Plos One* 7, e33252. doi:10.1371/journal.pone.0033252
- McDonagh, A. F. (2001). Turning Green to Gold. *Nat. Struct. Biol.* 8, 198–200. doi:10.1038/84915
- McDonald, L. R., Boyer, J. A., and Lee, A. L. (2012). Segmental Motions, Not a Two-State Concerted Switch, Underlie Allostery in CheY. *Structure* 20, 1363–1373. doi:10.1016/j.str.2012.05.008
- Motlagh, H. N., Wrabl, J. O., Li, J., and Hilser, V. J. (2014). The Ensemble Nature of Allostery. *Nature* 508, 331–339. doi:10.1038/nature13001
- Paukovich, N., Redzic, J. S., Chi, Y. C., Rahkola, J. T., Issaian, A., Blue, A., et al. (2019). *Streptococcus Pneumoniae* G5 Domains Bind Different Ligands. *Protein Sci.* 28, 1797–1805. doi:10.1002/pro.3693
- Paukovich, N., Xue, M., Elder, J. R., Redzic, J. S., Blue, A., Pike, H., et al. (2018). Biliverdin Reductase B Dynamics Are Coupled to Coenzyme Binding. *J. Mol. Biol.* 430, 3234–3250. doi:10.1016/j.jmb.2018.06.015
- Pereira, P. J. B., Macedo-Ribeiro, S., Párraga, A., Pérez-Luque, R., Cunningham, O., Darcy, K., et al. (2001). Structure of Human Biliverdin IX β Reductase, an Early Fetal Bilirubin IX β Producing Enzyme. *Nat. Struct. Biol.* 8, 215–220. doi:10.1038/84948
- Raum, H. N., Dreydoppel, M., and Weininger, U. (2018). Conformational Exchange of Aromatic Side Chains by ¹H CPMG Relaxation Dispersion. *J. Biomol. NMR* 72, 105–114. doi:10.1007/s10858-018-0210-5
- Rosato, A., Aramini, J. M., Arrowsmith, C., Bagaria, A., Baker, D., Cavalli, A., et al. (2012). Blind Testing of Routine, Fully Automated Determination of Protein Structures from NMR Data. *Structure* 20, 227–236. doi:10.1016/j.str.2012.01.002

- Salinas, V. H., and Ranganathan, R. (2018). Coevolution-based Inference of Amino Acid Interactions Underlying Protein Function. *Elife* 7, e34300. doi:10.7554/eLife.34300
- Schlegel, J., Armstrong, G., Redzic, J. S., Zhang, F., and Eisenmesser, E. Z. (2009). Characterizing and Controlling the Inherent Dynamics of Cyclophilin-A. *Protein Sci.* 18, NA. doi:10.1002/pro.89
- Selvaratnam, R., Mazhab-Jafari, M. T., Das, R., and Melacini, G. (2012a). The Auto-Inhibitory Role of the EPAC Hinge Helix as Mapped by NMR. *Plos One* 7, e48707. doi:10.1371/journal.pone.0048707
- Selvaratnam, R., VanSchouwen, B., Fogolari, F., Mazhab-Jafari, M. T., Das, R., and Melacini, G. (2012b). The Projection Analysis of NMR Chemical Shifts Reveals Extended EPAC Autoinhibition Determinants. *Biophysical J.* 102, 630–639. doi:10.1016/j.bpj.2011.12.030
- Singh, P., Francis, K., and Kohen, A. (2015). Network of Remote and Local Protein Dynamics in Dihydrofolate Reductase Catalysis. *ACS Catal.* 5, 3067–3073. doi:10.1021/acscatal.5b00331
- Smith, L. J., Browne, S., Mulholland, A. J., and Mantle, T. J. (2008). Computational and Experimental Studies on the Catalytic Mechanism of Biliverdin-Ix β Reductase. *Biochem. J.* 411, 475–484. doi:10.1042/bj20071495
- Torbeev, V. Y., Raghuraman, H., Hamelberg, D., Tonelli, M., Westler, W. M., Perozo, E., et al. (2011). Protein Conformational Dynamics in the Mechanism of HIV-1 Protease Catalysis. *Proc. Natl. Acad. Sci.* 108, 20982–20987. doi:10.1073/pnas.1111202108
- Vervoort, J., and Rietjens, I. M. C. M. (1996). Unifying Concepts in Flavin-dependent Catalysis. *Biochem. Soc. Trans.* 24, 127–130. doi:10.1042/bst0240127
- Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., et al. (2005). The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins* 59, 687–696. doi:10.1002/prot.20449
- Watt, E. D., Shimada, H., Kovrigin, E. L., and Loria, J. P. (2007). The Mechanism of Rate-Limiting Motions in Enzyme Function. *Proc. Natl. Acad. Sci.* 104, 11981–11986. doi:10.1073/pnas.0702551104
- Weininger, U., Diehl, C., and Akke, M. (2012). ¹³C Relaxation Experiments for Aromatic Side Chains Employing Longitudinal- and Transverse-Relaxation Optimized NMR Spectroscopy. *J. Biomol. NMR* 53, 181–190. doi:10.1007/s10858-012-9650-5
- Wu, S., Li, Z., Gnatenko, D. V., Zhang, B., Zhao, L., Malone, L. E., et al. (2016). BLVRB Redox Mutation Defines Heme Degradation in a Metabolic Pathway of Enhanced Thrombopoiesis in Humans. *Blood* 128, 699–709. doi:10.1182/blood-2016-02-696997
- Yabukarski, F., Biel, J. T., Pinney, M. M., Doukov, T., Powers, A. S., Fraser, J. S., et al. (2020). Assessment of Enzyme Active Site Positioning and Tests of Catalytic Mechanisms through X-Ray-Derived Conformational Ensembles. *Proc. Natl. Acad. Sci. USA* 117, 33204–33215. doi:10.1073/pnas.2011350117

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Redzic, Duff, Blue, Pitts, Agarwal and Eisenmesser. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Substitution of a Surface-Exposed Residue Involved in an Allosteric Network Enhances Tryptophan Synthase Function in Cells

Rebecca N. D'Amico^{1†}, Yuliana K. Bosken^{2†}, Kathleen F. O'Rourke¹, Alec M. Murray¹, Woudasie Admasu¹, Chia-en A. Chang² and David D. Boehr^{1*}

¹Department of Chemistry, The Pennsylvania State University, University Park, PA, United States, ²Department of Chemistry, The University of California Riverside, Riverside, CA, United States

OPEN ACCESS

Edited by:

George Lisi,
Brown University, United States

Reviewed by:

Ivaylo Ivanov,
Georgia State University,
United States
Patrick Loria,
Yale University, United States

*Correspondence:

David D. Boehr
ddb12@psu.edu

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 12 March 2021

Accepted: 11 May 2021

Published: 26 May 2021

Citation:

D'Amico RN, Bosken YK,
O'Rourke KF, Murray AM, Admasu W,
Chang CA and Boehr DD (2021)
Substitution of a Surface-Exposed
Residue Involved in an Allosteric
Network Enhances Tryptophan
Synthase Function in Cells.
Front. Mol. Biosci. 8:679915.
doi: 10.3389/fmolb.2021.679915

Networks of noncovalent amino acid interactions propagate allosteric signals throughout proteins. Tryptophan synthase (TS) is an allosterically controlled bienzyme in which the indole product of the alpha subunit (α TS) is transferred through a 25 Å hydrophobic tunnel to the active site of the beta subunit (β TS). Previous nuclear magnetic resonance and molecular dynamics simulations identified allosteric networks in α TS important for its function. We show here that substitution of a distant, surface-exposed network residue in α TS enhances tryptophan production, not by activating α TS function, but through dynamically controlling the opening of the indole channel and stimulating β TS activity. While stimulation is modest, the substitution also enhances cell growth in a tryptophan-auxotrophic strain of *Escherichia coli* compared to complementation with wild-type α TS, emphasizing the biological importance of the network. Surface-exposed networks provide new opportunities in allosteric drug design and protein engineering, and hint at potential information conduits through which the functions of a metabolon or even larger proteome might be coordinated and regulated.

Keywords: allostery, enzyme regulation, nuclear magnetic resonance, molecular dynamics, TIM barrel, substrate channeling, conformational dynamics

INTRODUCTION

Allosteric regulation of protein function is critical for several biological processes, including metabolism, oxygen transport and signal transduction (Wodak et al., 2019). Many enzymes are allosterically regulated, where a binding event distal to the active site impacts catalytic function (Monod et al., 1965; Koshland et al., 1966). Amino acid interaction networks have been proposed to connect different functional sites on a protein to propagate these allosteric signals (Dokholyan, 2016). Such networks might also bridge across proteins in multi-enzyme complexes, allowing efficient coordination of various enzyme functions (Cong et al., 2019; East et al., 2020). A better understanding of these protein-spanning networks would provide insight into how protein function is allosterically regulated in multi-enzyme complexes, and may offer novel avenues for protein engineering (Lee et al., 2008; Reynolds, et al., 2011; Gorman et al., 2019) and drug design (Nussinov and Chung, 2015; Greener and Sternberg, 2018).

TS has emerged as a model system for understanding allosteric regulation and functional coordination in multi-enzyme complexes (Dunn, 2012). TS contains both α TS and β TS subunits

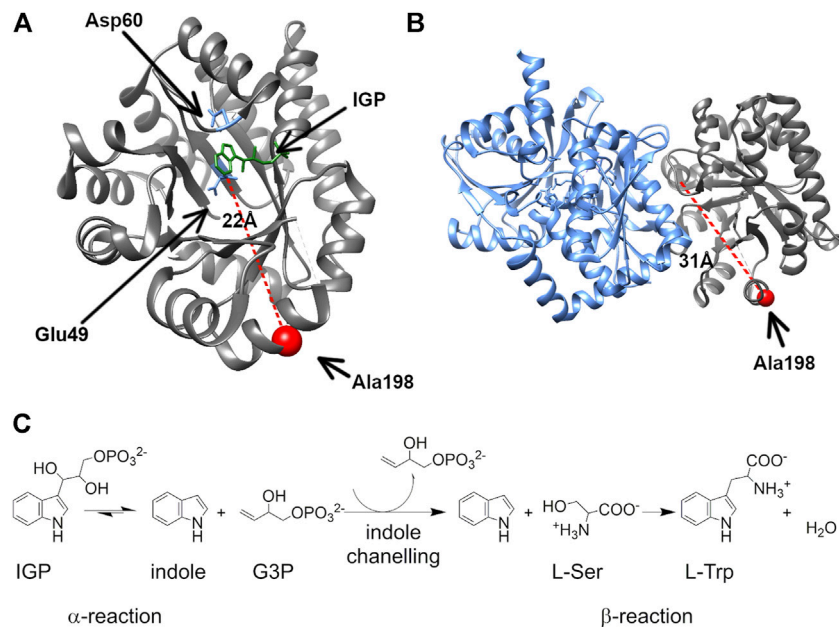


FIGURE 1 | The surface-exposed, network residue Ala198 is remote from both the α TS active site and the α TS- β TS interface. Structures of **(A)** the α TS subunit (PDB ID Code: 2RH9), highlighting catalytic residues Glu49 and Asp60, the substrate IGP, and Ala198, and **(B)** the full tryptophan synthase complex with the β TS subunit pictured in blue **(C)** Reactions catalyzed by alpha and beta subunits of TS.

in a linear $\alpha\beta\alpha$ arrangement (Figure 1). The TS enzyme is an anti-microbial drug target in *Mycobacterium tuberculosis* (Abrahams et al., 2017; Wellington et al., 2017) and has been implicated in antibiotic resistant strains of *Chlamydia trachomatis* (Somboonna et al., 2019). There has also been interest in engineering β TS for the production of novel tryptophan analogs (Buller et al., 2015; Murciano-Calles et al., 2016; Buller et al., 2018). TS has been of particular interest due to the channeling of the intermediate, indole, between the two subunits through a 25 Å hydrophobic tunnel (Dunn et al., 1990). Structural changes in TS appear to be highly coordinated as structural changes in one subunit affects the structure and catalytic activity of the other subunit (Dunn, 2012). Notably, both subunits experience a marked decrease in catalytic activity in the absence of the other subunit (Niks et al., 2013). Understanding how structural changes and function are coordinated between the TS subunits may provide insight into similar interactions in other multi-subunit enzyme complexes.

We previously delineated amino acid interaction networks for α TS in the absence of β TS (Axe et al., 2014; O'Rourke et al., 2019) using the solution-state nuclear magnetic resonance (NMR) method known as CHESCA (chemical shift covariance analysis) (Selvaratnam et al., 2011). Interestingly, these networks included residues at the α TS/ β TS binding interface, which suggested that they might be involved in coordinating function between α TS and β TS. A particularly interesting residue was Ala198, a surface-exposed network residue that is conformationally dynamic throughout the catalytic cycle of α TS (O'Rourke et al., 2018), and is distant from both the α TS active site and the α TS/ β TS binding interface. Previous studies

indicated that the A198G substitution induces a modest decrease in α TS catalytic activity, and changed the structure/dynamics of other network residues, according to NMR studies (Axe et al., 2014). Here, we have developed a cell-based assay to screen the phenotypic consequences of network substitutions of α TS, in particular, evaluating network substitutions at position 198. Surprisingly, we show that the A198W substitution induces more rapid bacterial growth in a tryptophan auxotrophic strain of *Escherichia coli*, likely by increasing catalytic efficiency of TS. Several computational approaches have been developed in an effort to map allosteric networks which commonly rely on tracing significant changes in residues conformation and interaction (Feher et al., 2014; Johnson et al., 2018; Botello-Smith and Luo, 2019; Wang et al., 2020). Similarly, in this study we evaluated pairwise forces to distinguish specific interaction changes resulting from mutation. The A198W substitution induces structural and dynamic changes to the allosteric network that connects to the α TS/ β TS interface resulting in new connections, opening of the indole channel and more efficient tryptophan biosynthesis. These studies underscore the importance of network residues not only to enzyme function but to coordination of functions between enzymes.

MATERIALS AND METHODS

Escherichia coli Growth Curves

Growth curves were determined using *E. coli* K-12 BW25113 cells with the *trpA* gene knocked out. Cells were provided by the Keio

Collection at Yale University (Baba et al., 2006). The kanamycin cassette was then eliminated using the lambda red recombinase method (Datsenko and Wanner, 2000). Cells were grown for 16 h in Vogel-Bonner minimal media supplemented with 1X MEM vitamin mix, and 7.5 μ M tryptophan. Cell lines containing a trpA plasmid were also grown in the presence of 50 μ g/ml kanamycin. After 16 h, the resulting cultures were diluted 1:10 in identical medium containing 7.5 μ M tryptophan and were grown for 8 h. Cell cultures were then diluted 1:25 in identical medium lacking tryptophan. Growth over a period of 50–60 h was measured using OD₆₀₀ readings. Resulting curves were plotted and fit using the following equation:

$$N_t = \frac{K}{1 + \left(\frac{K-N_0}{N_0}\right)e^{-rt}}$$

where the N_0 represents population at time zero and N_t represents the population size at time t . K represents the carrying capacity—or the maximum population size that could be supported with no limitations. This equation was adapted from the Growthcurver package (Sprouffske and Wagner, 2016) for use in MatLab (MathWorks, Natick, MA, United States). Error for each fit was calculated using residual bootstrapping with 1,000 repetitions.

Generation of Position 198 Variants

The A198C, A198F, A198G, A198N, A198Q, and A198W variants were obtained using QuikChange Lightning Site Directed Mutagenesis Kit (Agilent Technologies). Primers were designed by using Agilent's primer design tool. The remaining variants (A198D, A198V, A198R, A198L, A198K) were obtained using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs, NEB) using a site-saturation method. These primers were designed using NEB's supplied primer design tool.

Expression and Purification of α TS and β TS

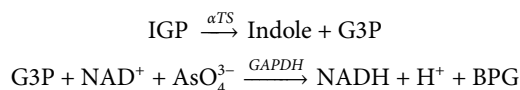
All proteins were expressed using *E. coli* BL21 (DE3⁺) cells. Expression of unlabeled proteins was carried out using Luria-Bertrani (EMD Millipore, Billerica, MA, United States) medium, and expression of isotopically labeled proteins was carried out using ²H₂O-based M9 medium with ¹⁵NH₄Cl by previously described methods (Axe and Boehr, 2013; Axe et al., 2014). The expressed β TS subunit contained a N-terminal hexahistidine tag.

The α TS subunit and variants were purified as previously (Axe and Boehr, 2013; Axe et al., 2014; O'Rourke et al., 2019). The β TS subunit was purified using a Ni-NTA resin using column buffer (25 mM HEPES, pH 7.5, 1 mM Na₂EDTA, 200 mM NaCl) and a gradient of imidazole from 50–500 mM. The β TS subunit eluted at 250 mM imidazole. Samples were then concentrated using a Corning Spin-X UF centrifugal concentrator and applied to a size exclusion S200 column (GE Healthcare) and eluted with column buffer.

Enzyme Kinetic Assays

The chemoenzymatic production of indole-3-glycerol phosphate (IGP) substrate was carried out as previously described (O'Rourke

et al., 2019). For enzyme assays, the IGP concentration was determined using $\epsilon = 5.54 \text{ mM}^{-1} \text{ cm}^{-1}$. Assays of the α TS subunit alone were performed using a previously described assay (Creighton, 1970; Lane and Kirschner, 1991), which utilizes a coupled reaction scheme as follows:



where G3P is glyceraldehyde-3-phosphate, NAD⁺ and NADH are the oxidized and reduced forms of nicotinamide adenine dinucleotide, and BPG is 1,3 bisphosphoglycerate. Production of NADH was measured at 340 nm using $\epsilon = 6.22 \text{ mM}^{-1} \text{ cm}^{-1}$. The assay was performed in assay buffer (100 mM HEPES pH 7.5, 100 mM KCl, and 40 μ M pyridoxal 5' phosphate) plus 20 mM sodium arsenate. The concentration of the α TS subunit was 60 μ M and the IGP concentration was 0.075–2.25 mM.

The catalytic activity of the β TS subunit (Faeder and Hammes, 1970) was assayed in the presence of wild-type α TS or variants by measuring the production of tryptophan at 290 nm using $\epsilon = 1.89 \text{ mM}^{-1} \text{ cm}^{-1}$. The concentration of the $\alpha_2\beta_2$ TS complex was 0.5 μ M. The reaction was initiated using 1–50 μ M indole (Thermo Fisher). Assays were performed in the same assay buffer as indicated above, plus 60 mM serine.

The full TS complex reaction (Brzovic et al., 1992) was measured by tracking production of tryptophan from IGP at 290 nm using $\epsilon = 1.89 \text{ mM}^{-1} \text{ cm}^{-1}$. The reaction was initiated with 0.05–3 mM of IGP. The concentration of the $\alpha_2\beta_2$ TS complex used was 50 μ M and assays were performed in assay buffer, plus 60 mM serine.

Given that some of the kinetic data was collected under conditions in which the total substrate concentration $[(S_T)]$ approached the total enzyme concentration $[(E_T)]$, the kinetic data was fit to the Morrison quadratic equation (Morrison, 1969):

$$v_0 = V_{\max} \frac{([E_T] + [S_T] + K_M) - \sqrt{([E_T] + [S_T] + K_M)^2 - 4[E_T][S_T]}}{2[E_T]}$$

where v_0 is the initial reaction velocity, V_{\max} is the maximum reaction velocity and K_M is equivalent to the Michaelis–Menten constant. The maximum catalytic rate constant (k_{cat}) is determined by dividing V_{\max} by $[E_T]$ as normal.

For each initial velocity vs. $[S]$ curve, kinetic data was collected for eight different $[S]$ data points, in which these data points represent three technical replicates (i.e., using the same batch of purified enzyme), and each curve was generated in totality at least three times with different biological replicates (i.e., using different batches of purified enzyme).

NMR Experiments

The NMR experiments were conducted and analyzed as previously described (Axe and Boehr, 2013; O'Rourke et al., 2018). Protein samples were exchanged into NMR buffer (50 mM potassium phosphate, pH 7.8, 2 mM DTT, 0.2 mM EDTA, and 10% ²H₂O) and contained 1.0 mM ²H, ¹⁵N-labeled α TS, 10 mM indole (Thermo Fisher) and/or 20 mM G3P (Sigma

Aldrich) where appropriate. ^{15}N R_2 relaxation dispersion experiments were collected and analyzed according to previously established procedures (Loria et al., 2008; O'Rourke et al., 2018). Briefly, data was collected at 283 K on 600 and 850 MHz Bruker Avance III spectrometers using previously described pulse sequences (Loria et al., 1999) and data analyzed using the computer program GLOVE (Sugase et al., 2013).

Molecular Dynamics Simulations and Analysis

Due to the lack of X-ray crystal structure for the *E. coli* $\alpha_2\beta_2$ TS complex, wild type and A198W mutant molecular systems were based on the X-ray crystal structure of *Salmonella typhimurium* TS PDB ID 2CLK (Ngo, Harris et al., 2007) and corresponding residues were substituted to match the *E. coli* sequence. The coordinates for the active conformation of Glu49 and the substrates were taken from PDB entry 1QOQ (Weyand and Schlichting, 2000).

Four independent simulations were performed for each system (different random seed, starting from the same equilibrated system)—wild type (WT) and variant (A198W). MD simulations were performed using standard Amber package with GPU acceleration (Salomon-Ferrer et al., 2013; Case et al., 2018). The protein was parameterized using Amber Force Field FF14SB (Maier et al., 2015). General Amber force field (GAFF) was applied to ligands and charges were assigned using AM1-bcc model (Jakalian et al., 2002). All systems were prepared by a three-step minimization process (hydrogens, sidechains and all atoms), solvated with TIP3P water model with counter ions in a rectangular box with edges at minimum 12 Å from any atom (Jorgensen et al., 1983). The solvated systems were minimized and equilibrated from 0 to 298 K at 25 K intervals. MD trajectories were collected over 200 ns at 1 ps interval with 2 fs timestep under constant pressure and temperature. Particle mesh Ewald was used for long range electrostatics and SHAKE algorithm for fixed heavy atom-hydrogen bond lengths (Ryckaert et al., 1977; Sagui et al., 2004).

The systems were visualized and analyzed using Visual Molecular Dynamics (Humphrey et al., 1996) and Molecular Operating Environment. The trajectory output files were processed with PTRAJ software to contain 2,000 frames, each representing 0.1 ns timestep. Dihedral data was collected with T-Analyst software (Ai et al., 2010; Roe and Cheatham, 2013). Dihedral entropy was also calculated using T-Analyst and residues with difference in entropy between WT and variant higher than 0.2 kcal/mol were selected for further analysis (sidechain and backbone dihedral angles were considered). Four MD runs were analyzed for the WT and A198W variant systems. The force distribution analysis tool was used to identify significant contacts and persistent interactions throughout the 200 ns trajectories. The pair wise atom forces are represented as a scalar value with negative values indicating attraction and positive values showing repulsion (Stacklies et al., 2011).

RESULTS

The A198W Network Substitution Engenders a Fast Growth Phenotype to *E. coli* Cells

We had previously identified network substitutions that are modestly detrimental to the kinetic parameters of αTS (Axe et al., 2014). We sought to develop a cell growth-based assay that would allow us to analyze αTS substitutions in a much higher throughput manner, and especially evaluate any phenotypic responses to network substitutions. Towards this goal, we obtained *E. coli* K12 cells with the *trpA* gene knockout from the Keio collection at Yale University (Baba et al., 2006), and analyzed growth in the presence and absence of tryptophan, with and without complementing plasmid expressing *E. coli* αTS (Supplementary Figure S1). As expected, cells that were not supplemented with tryptophan and did not carry the αTS -expressing plasmid failed to grow. Interestingly, the cells supplemented with plasmid but not tryptophan experienced a delay in the start of the growth phase. This phenomenon has been previously observed in *E. coli* K12 strains, including isoleucine and serine auxotrophs (Fisher et al., 2011). Fisher and coworkers were able to rescue auxotroph viability with *de novo* proteins for several amino acid auxotrophs. The authors attributed the slower cell growth to the *de novo* proteins being less efficient than the endogenous protein. However, we observed this lag even when the cells were supplemented with the wild-type (WT) version of the knocked-out enzyme. We attributed this lag to the additional metabolic stress induced by the required plasmid. This lag was observed consistently in cells that required the plasmid to produce tryptophan for cell growth.

We were especially interested in evaluating substitutions at position 198 of αTS . Ala198 is an allosteric network residue (Axe et al., 2014; O'Rourke et al., 2019) that is distant from both the αTS active site (>20 Å) and the $\alpha\text{TS}/\beta\text{TS}$ binding interface (>30 Å) (Figure 1). Nonetheless, a substitution at position 198 has a modest effect on the αTS kinetic parameters (Axe et al., 2014). Moreover, Ala198 has been shown by NMR ^{15}N R_2 relaxation dispersion experiments to be conformationally dynamic on the millisecond timescale in the presence and absence of substrate/products (O'Rourke et al., 2018). In order to investigate the structural and/or functional relevance (if any) of Ala198, especially within a biologically relevant context, we generated several point mutations at this position and tested each of these point variants using the *E. coli* K12 knockout cells. We found that several amino acid substitutions caused a statistically significant decrease in cellular growth rate (i.e., A198C and A198Q) while others were not able to rescue growth (i.e., A198K and A198V) (Figure 2A). Most surprisingly, production of the A198W variant resulted in an increase in cellular growth rate (Figure 2B). These results indicated that position 198 was indeed structurally/functionally important for TS. We note that expression levels of WT and A198W αTS were very similar (Supplementary Figure S2), suggesting that cell growth differences were likely related to functional differences induced by the A198 substitutions. As it is atypical for amino acid

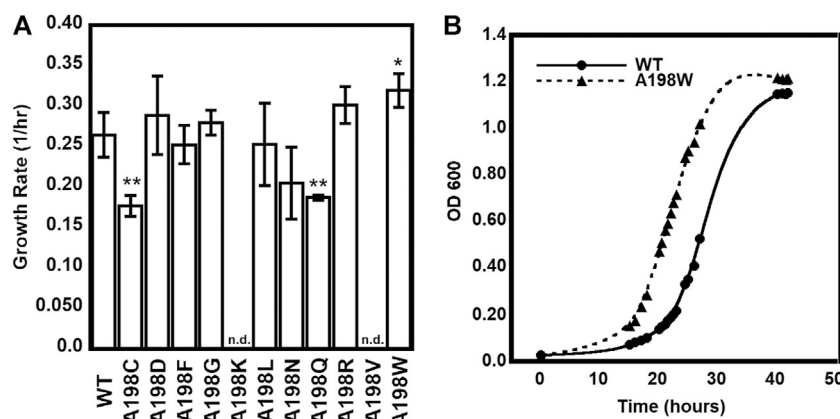


FIGURE 2 | Growth complementation assays of *trpA*-deleted *Escherichia coli* K12 cells. **(A)** Tryptophan-auxotrophic *E. coli* cells were complemented with plasmid expressing α TS WT or A198 variants, and growth rates determined. Asterisks indicate that the growth rate for *E. coli* expressing the A198 variant was statistically different than the growth rate for *E. coli* expressing WT α TS, according to *p* value ($0.01 < * < 0.05$; $0.001 < ** < 0.01$; $n = 3$). n. d. not determined because *E. coli* cells expressing the A198K and A198V variants did not grow in the absence of tryptophan **(B)** Representative growth curves demonstrating the A198W variant's increased growth rate.

TABLE 1 | Kinetic parameters for WT and A198W TS.

Assay	Protein	k_{cat} (s^{-1})	$K_{M,IGP}$ (mM)	$K_{M,indole}$ (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)	$(k_{cat}/K_M)^{A198W}/(k_{cat}/K_M)^{WT}$
α TS alone	WT	$(2.88 \pm 0.54) \times 10^{-4}$	0.94 ± 0.28		0.31 ± 0.11	
	A198W	$(3.81 \pm 10.16) \times 10^{-4}$	1.31 ± 0.09		0.29 ± 0.02	0.94 ± 0.34
TS complex	WT	2.83 ± 0.24	0.49 ± 0.10		$(5.8 \pm 1.2) \times 10^3$	
	A198W	4.14 ± 0.25	0.42 ± 0.06		$(9.9 \pm 1.4) \times 10^3$	1.7 ± 0.4
β TS assay	WT	3.81 ± 0.38		25 ± 5	$(1.5 \pm 0.3) \times 10^5$	-
	A198W	3.08 ± 0.26		16 ± 4	$(1.9 \pm 0.5) \times 10^5$	1.3 ± 0.4

substitutions to enhance enzyme function, most of the remaining studies focused on understanding the functional, structural and dynamic consequences of the A198W substitution. We note first that we repeated the growth curve experiments for WT ($n = 13$) and A198W ($n = 9$) to determine the extent of this increased rate. The average growth rate for cells expressing the A198W variant (0.3375 ± 0.0400) was consistently higher than that for cells expressing WT α TS (0.2664 ± 0.0549 ; $p = 0.0035$).

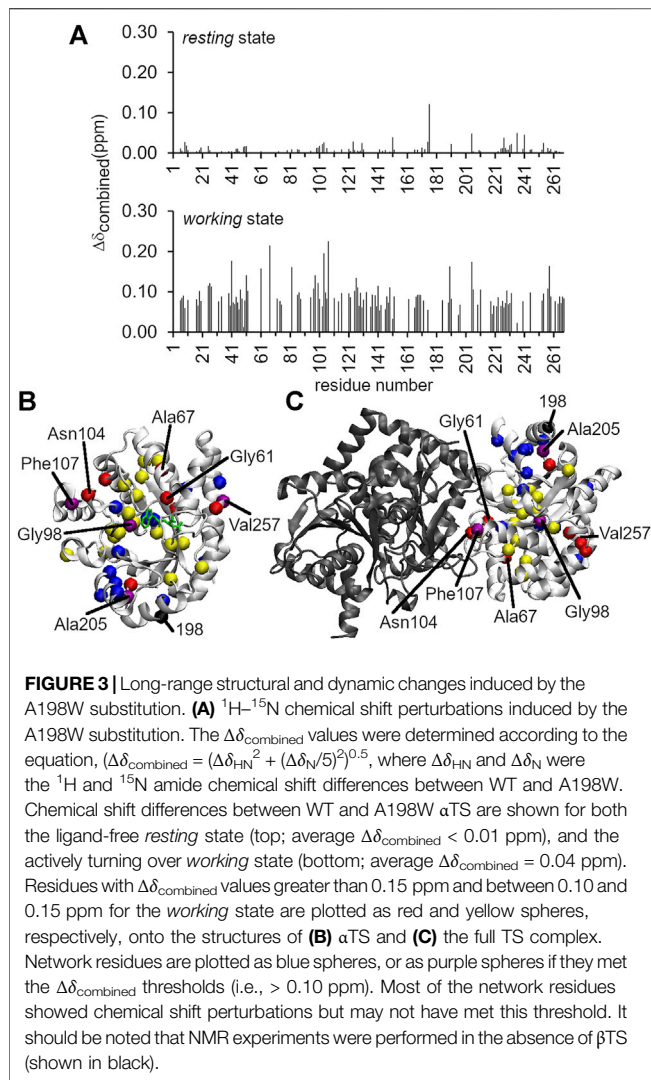
The A198W Substitution Increases Catalytic Efficiency of The Full TS Complex

To better understand the effects of the A198 substitutions on the structure and function of the TS enzyme, we expressed and purified the A198K, A198V, and A198W variants. The A198K and A198V NMR spectra were very different from that of WT α TS (Supplementary Figure 3), including many chemical shift and peak intensity changes. Unfortunately, these changes precluded ready NMR assignments for these spectra. We note in the next section that the A198W variant also led to changes in the NMR spectrum, but these changes were not nearly as severe as those induced by the A198K and A198V variants (Supplementary Figure 3). Given these results, the remaining studies were performed only with the A198W variant.

We conducted a number of kinetic experiments to better elucidate the effects of the A198W substitution on the function of the TS enzyme (Table 1, Supplementary Figure 4). We note that some of the substrate concentrations approached the total enzyme concentration. In these cases, it is better to model the steady-state kinetics using the Morrison quadratic equation (Morrison, 1969), which also provides the Michaelis-Menten constant (K_M) and the catalytic turnover rate constant (k_{cat}). However, the kinetic parameters determined by the Morrison and Michaelis-Menten equations were nearly identical and not statistically different. We report here only the kinetic parameters determined through the Morrison equation for simplification.

We first analyzed the effects of the A198W substitution on the catalytic activity of α TS in the absence of β TS. The activity of α TS alone was measured via a coupled reaction (Lane and Kirschner, 1991), in which production of product G3P from the α TS forward reaction was coupled to reduction of NAD^+ , generating a measurable absorbance change at 340 nm. Given the results of the cell-based assay, it was surprising that there did not appear to be any significant difference for the catalytic efficiencies (k_{cat}/K_M) between WT ($0.31 M^{-1} s^{-1}$) and A198W ($0.29 M^{-1} s^{-1}$) α TS.

We then compared the kinetic parameters of the entire WT and A198W TS complexes by measuring the production of



tryptophan from IGP at 290 nm. Tryptophan production requires reactions to occur in both subunits, with the product of the αTS reaction (indole) being channeled to the βTS subunit. In contrast to the αTS assays, there was a modest increase in the catalytic efficiency of the A198W TS variant ($9.9 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$) compared to WT TS ($5.8 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$). We also tested the ability of the A198W αTS subunit to stimulate βTS activity by measuring the production of tryptophan starting from indole at 290 nm. Here, there was also a slight increase in the catalytic efficiency in the presence of A198W αTS (A198W, $1.9 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$; WT, $1.5 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$), although this appeared to be driven by a small decrease in the apparent K_M , whereas the change in the catalytic efficiency for the full TS complex assay was driven primarily by an increase in k_{cat} in the presence of A198W αTS (see **Table 1**).

While the kinetic effects of the A198W substitution were admittedly modest (1.4–1.7 fold), these effects were on par with the faster growth rate (~ 1.3 fold) in the *E. coli* cells expressing A198W αTS . For comparisons, we reference some studies by the Arnold lab in trying to engineer stand-alone βTS enzymes for chemoenzymatic processes. As indicated, αTS and

βTS have decreased enzyme activity in the absence of the other subunit (Niks et al., 2013). However, the Arnold lab was able to use directed evolution on the *Pyrococcus furiosus* βTS to identify βTS enzymes with higher catalytic activities, comparable to those found with the full TS complex (Buller et al., 2015). A recombination library of these original mutations enabled the discovery of mutations that also enhanced the catalytic activity of the isolated *Thermotoga maritima* βTS , such that the engineered βTS had catalytic activities ~ 2.5 – 4.5 greater than that of the full TS complex (Murciano-Calles et al., 2016). However, it is notable that similar strategies failed to identify *E. coli* βTS variants with enhanced activity (Murciano-Calles et al., 2016).

The A198W Substitution Does not Affect Binding Affinity Between the αTS and βTS Subunits

One potential explanation for the increased activity of the TS complex was that the A198W substitution could be impacting the binding affinity between αTS and βTS . To investigate this, isothermal titration calorimetry (ITC) studies were performed on the WT and the A198W variant αTS subunit binding to the βTS subunit (**Supplementary Figure 5**). WT and A198W αTS bound βTS with nearly identical nanomolar dissociation constants ($\sim 25 \text{ nM}$). These experiments showed that there was no significant thermodynamic difference between the two binding events, implying that functional changes due to the A198W substitution were not simply due to differences in subunit association thermodynamics.

The A198W Substitution Attenuates Conformational Exchange Events Including at the $\alpha\text{TS}/\beta\text{TS}$ Interface

Considering that we had previously identified Ala198 as a potentially important residue by NMR (Axe et al., 2014; O'Rourke et al., 2018; O'Rourke et al., 2019), we also evaluated structure/dynamic changes induced by A198W through NMR methods. We note that these NMR experiments were conducted in the absence of βTS . We collected ^1H - ^{15}N HSQC spectra (**Figure 3**) and ^{15}N R_2 relaxation dispersion (**Figure 4**) data for the ligand-free *resting* state and the *working* state of αTS . The working state represents active turnover conditions to reach dynamic conformational equilibrium between the E:IGP and E:indole:G3P states in a 4:1 ratio (Axe and Boehr, 2013).

Intriguingly, the chemical shift perturbations induced by the A198W substitution were more substantial in the *working* state than in the ligand-free *resting* state. The average chemical shift perturbations for the *resting* and *working* state conditions were < 0.01 and ~ 0.04 ppm, respectively [as measured by $\Delta\delta_{\text{total}} = [\Delta\delta_{\text{HN}}^2 + (\Delta\delta_{\text{N}}/5)^2]^{0.5}$, where $\Delta\delta_{\text{HN}}$ and $\Delta\delta_{\text{N}}$ were the ^1H and ^{15}N amide chemical shift differences between WT and A198W αTS]. In the *resting* state, Leu176 experienced the largest chemical shift perturbation with $\Delta\delta_{\text{total}} = 0.12$ ppm (**Figure 3**). In the *working* state, the A198W substitution induced chemical shift perturbations throughout αTS (**Figure 3**), with the largest

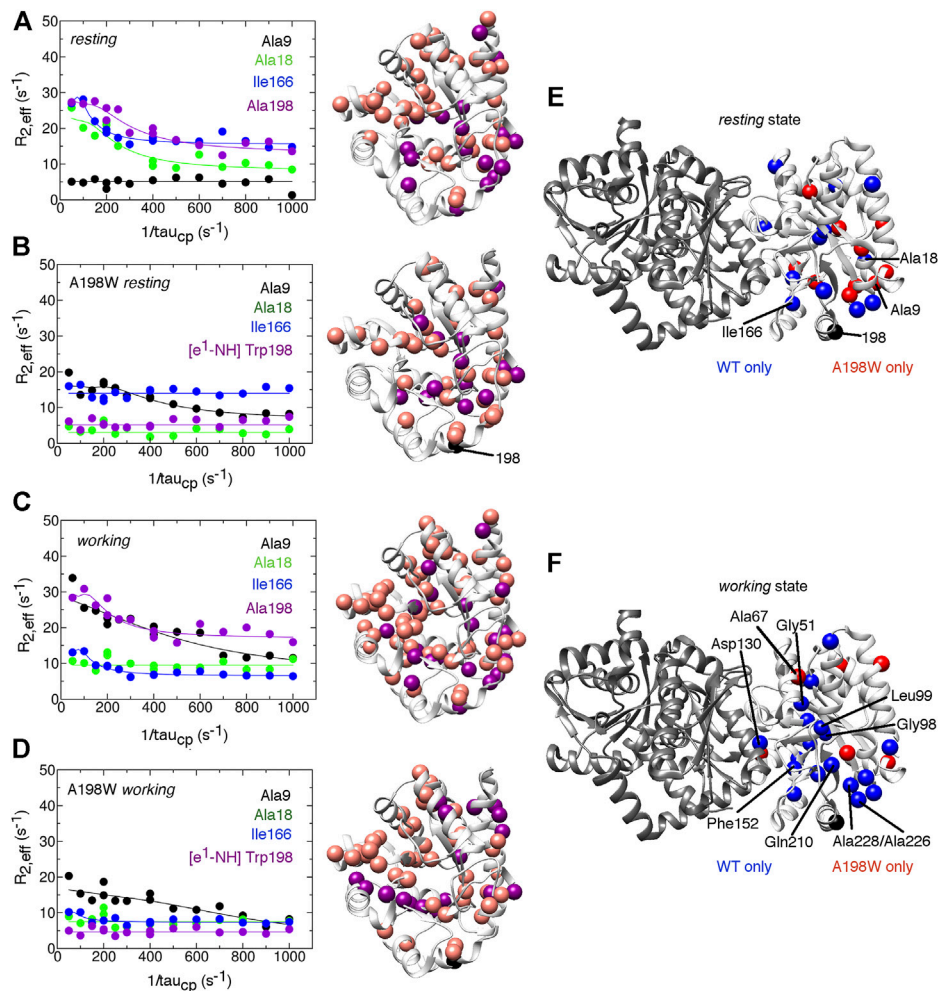


FIGURE 4 | The A198W surface-exposed, network substitution suppresses millisecond conformational motions in α TS. Conformational exchange events in *E. coli* α TS enzyme for (A) WT ligand-free resting, (B) A198W ligand-free resting, (C) WT working and (D) A198W working states. (Left) Example ^{15}N R_2 relaxation dispersion curves collected at a ^1H Larmor frequency of 850 MHz for resonances belonging to Ala9 (black), Ala18 (green), Ile166 (blue), and the sidechain ϵ^1 -NH group of Trp198 (purple). (Right) Locations of conformational exchange according to the R_2 relaxation dispersion curves plotted as spheres onto the α TS structure. Purple spheres indicate that associated R_2 relaxation dispersion curves can be fit to two-site exchange, while pink spheres indicate exchange broadening, but the R_2 relaxation dispersion curves cannot be fit reliably to two-site exchange. Comparison of the conformational exchange events between WT and A198W α TS for (E) resting and (F) working states is also shown. Blue (red) spheres indicate conformational exchange events present in the WT (A198W) enzyme but not in the A198W (WT) enzyme. It should be noted that the NMR data was collected in the absence of β TS. The ^{15}N R_2 relaxation dispersion experiments were conducted at 283 K using a buffer consisting of 50 mM potassium phosphate, pH 7.8, 2 mM DTT, 0.2 mM Na_2EDTA , and 10% $^2\text{H}_2\text{O}$, and 0.5–1.0 mM protein with 10 mM indole, and/or 20 mM G3P, where appropriate.

chemical shift perturbations associated with the $\beta_2\alpha_2$ active site loop (i.e., Ala67) and residues at the α TS/ β TS interface (i.e., Asn104, Phe107). Our previous NMR studies (Axe et al., 2014; O'Rourke et al., 2019) identified two network clusters, and most residues experiencing substantial chemical shift perturbations (i.e., $\Delta\delta_{\text{total}} > 0.10$ ppm) belonged to one of these clusters, or were next to a network residue.

The A198W substitution also led to changes in millisecond conformational exchange events in both the resting and working states of α TS according to the ^{15}N R_2 relaxation dispersion experiments (Figure 4). In the working state, the A198W substitution mostly suppressed protein motions that were present in the WT enzyme. These suppressed motions

included those near the substitution site (e.g., Ala226, Ala228), near the active site (e.g., Gly51, Gly98, Leu99, Phe152, Gln210) and near the α TS/ β TS interface (e.g., Asp130). The rate of conformational exchange also slowed for those residues still undergoing exchange in A198W α TS (Supplementary Table 1). Together the chemical shift perturbations and ^{15}N R_2 relaxation dispersion data suggest that there are dynamic connections between position 198, the active site and the α TS/ β TS interface, thus providing potential means through which substitutions at position 198 can influence TS function. Conversely, it is remarkable that despite all of these dynamic changes, there is little effect on the kinetic parameters of α TS (Table 1). The function of α TS in isolation may be robust to these

dynamic changes, although it is noted that α TS acts as a very poor catalyst in the absence of β TS.

Bioinformatics Analyses of the NMR-Derived Allosteric Networks

The A198W substitution appeared to influence TS function by modulating interactions between α TS and β TS, especially changing interactions directly involving or nearby previously identified allosteric network residues. Other biophysical and computational methods exist that can also help to delineate allosteric networks (O'Rourke et al., 2016). For example, the concept of “frustration” can provide insight into how energy is distributed in protein structures and how mutations or conformational changes shift these energy distributions (Ferreiro et al., 2007). In the algorithm developed by Ferreiro et al. (2007), local frustration is determined by mutating single residues or pairs of residues *in silico* and computing energy changes. If the native residue or pair is destabilizing compared to alternatives, this residue or pair interaction is “frustrated”. Sites with high frustration are often associated with binding or allosteric sites, and may be important for guiding functionally-relevant dynamics (Ferreiro et al., 2014; Ferreiro et al., 2018; Freiburger et al., 2019). We used the AWSEM-MD Frustratometer (Parra et al., 2016) to identify highly frustrated residues in the related *Salmonella typhimurium* α TS, which has 85% sequence identity to *E. coli* α TS. We evaluated frustration for a number of complexes, including TS bound with the α TS substrate mimic N-[1H-indol-3-yl-acetyl] aspartic acid (PDB 1K3U), TS bound with the α TS transition state analog 4-(2-hydroxy-4-fluorophenylthio)-butylphosphonic acid (PDB 1C9D), TS bound with α TS inhibitor F9 and L-tryptophan in the beta site (PDB 5CGQ) and TS bound with the β TS quinoid intermediate (PDB 3CEP). The most highly frustrated residues were similar for the different TS structures. Specifically, residues Glu2, Pro28, Gln32, Asp46, Asp56, Asn66, Glu83, Asp130, Glu135, and Lys263 (all conserved between *E. coli* and *S. typhimurium* α TS) were identified as highly frustrated in all evaluated structures (Supplementary Figure 6). It is noteworthy that Gln32, Asp46, Asp130, and Lys263 were all identified as network residues according to the previous NMR studies (O'Rourke et al., 2019); the backbone resonances for Glu2 and Asn66 were unassigned, the ^1H - ^{15}N based NMR experiments could not provide information about Pro28 although Asp27 was a network residue, and Ser136 (i.e., next to Glu135) was also identified as a network residue. The A198W substitution induced ^1H - ^{15}N backbone amide chemical shift changes (e.g., Asp27, Ala47) and/or motional changes (e.g., Asp130, Val131) to some of these or nearby residues.

Another highly used method to identify allosteric networks in proteins involves identifying covarying or coevolving residue pairs. In this method, statistical analyses are performed on large multiple sequence alignments to identify if two residues covary/coevolve across the alignment (dos Santos et al., 2019; Morcos and Onuchic, 2019). Residues that covary likely interact, and may be involved in the same allosteric network (Rivoire et al., 2016). For our purposes, we used the RaptorX-Complex Contact webserver (Zeng et al., 2018) to identify covarying residues within α TS and between α TS

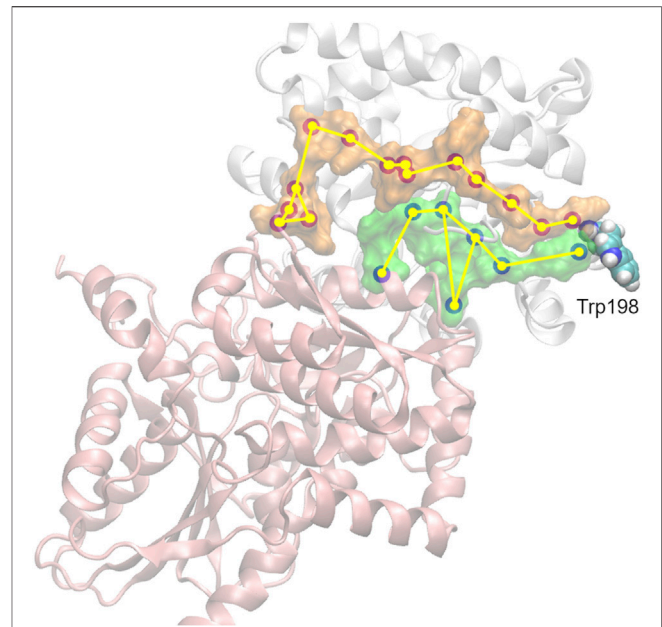
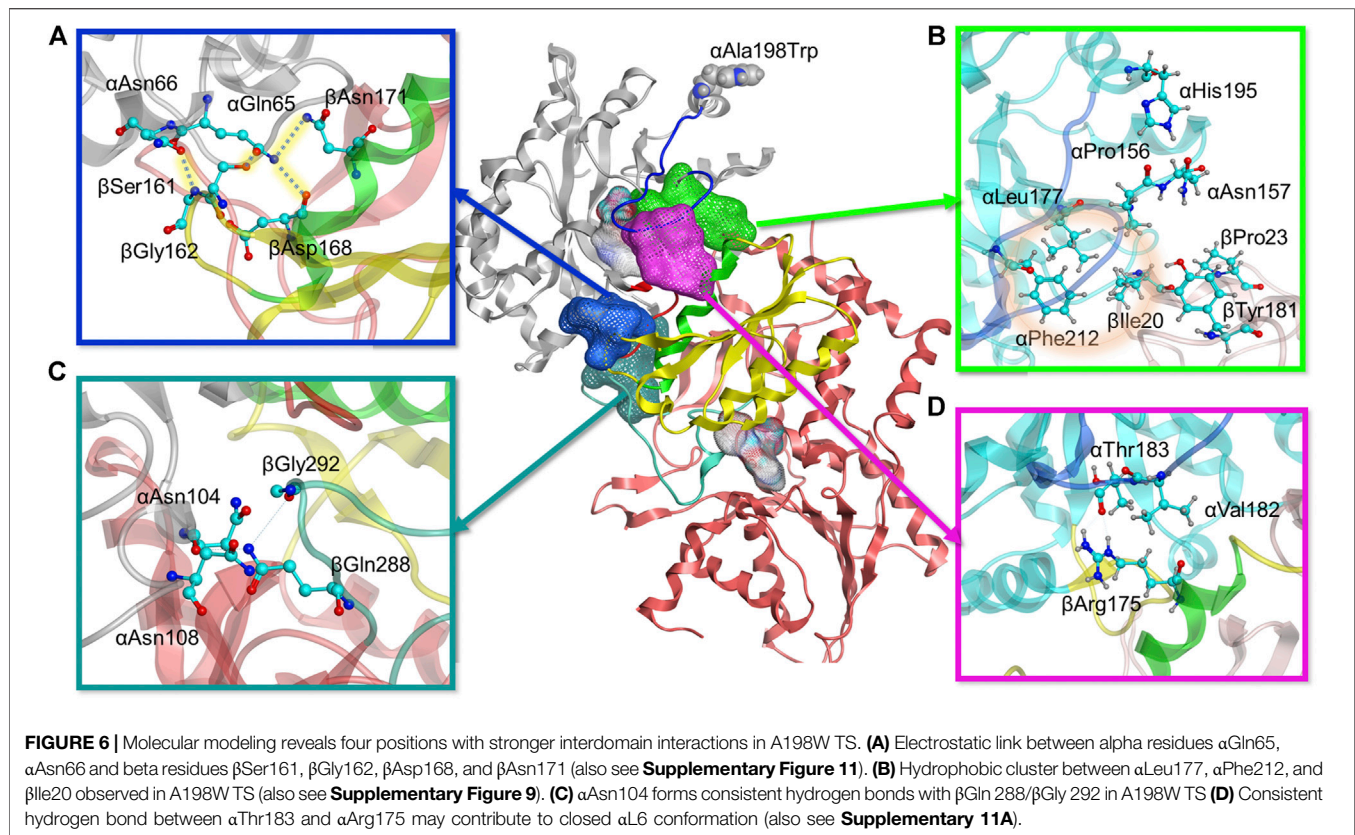


FIGURE 5 | Propagation of interaction changes induced by the A198W substitution were examined in two directions leading to the α TS (β TS (pink) interface). α C of residues with affected conformation are traced, based on pair-wise force distribution analysis, from the substitution site as follows: in orange— α Asn194, α Leu193, α Leu191, α Ala189, α Arg188, α Glu186, α Ala185, α Ser235, α Lys239, α Glu242, α Asn66, α Gln65, β Ser161, β Gly162; in green— α His195, α Pro156, β Ile20, α Leu177, α Phe212, α Thr183, β Arg175.

and β TS using the *E. coli* sequences as starting points. We identified the top ten covarying pairs of residues within α TS (see Supplementary Figure 6). For seven of these pairs, one or both residues are NMR-derived network residues (O'Rourke et al., 2019). The other three pairs contain a residue that is next to a network residue (e.g., covarying residue Ser125 in two of the pairs is next to Asp124; covarying residue Ala265 is between the network residues Ala264 and Thr266). We also identified the top covarying pairs of residues between α TS and β TS (see Supplementary Figure 6). Perhaps not surprisingly, and likely owing to the power of the algorithm, all of these pairs are at the α TS- β TS interface. These residues include those previously identified as network residues (i.e., Ala103, Asn104 and Phe107) (O'Rourke et al., 2019). The A198W substitution induces ^1H - ^{15}N backbone amide chemical shift changes (e.g., Ala47, Gly51, Gly98, Leu100, Asn104, Phe107) and/or motional changes (e.g., Gly51, Gly98, Asp124, Val148, Gln210, Ala231, Ala265) to many of these covarying residues or nearby residues in α TS.

Molecular Dynamics Simulations Indicate that the A198W Substitution Induces a More Open Indole Channel

While the NMR data provided some insight into communication between position 198 and the α TS/ β TS interface, the NMR data was limited in providing full context for the $\alpha\beta$ TS complex. As such, we analyzed molecular dynamics (MD) trajectories for α - β heterodimers, including for WT, A198K, A198V, and A198W



α TS. The MD simulations for the A198K and A198V variants indicated a dramatic movement of active site loop 6 (also known as the β 6 α 6 loop or as α L6) into a position that completely opens the active site and which is likely not conducive for catalysis (**Supplementary Figure 7**); this α L6 conformation was not observed for either WT or A198W α TS. The large conformational change in α L6 may help to explain the large chemical shift and peak intensity changes in the A198K and A198V NMR spectra.

The MD simulations for the A198W variant were perhaps more insightful, and so, we focused more attention on analyzing these changes compared to what was observed for WT α TS. Briefly, the MD simulations indicated that local changes at the substitution site propagated structural dynamic changes throughout α TS (**Figure 5**), including decreasing the overall flexibility of α L6. It is important to note that large scale conformational changes captured in our NMR experiments on millisecond timescale are induced by side chain motions and backbone fluctuations which could be detected within the nanoseconds timeframe. Detailed atomistic analysis of our trajectories explains how changes in forces and interactions lead to different dynamics and conformational behavior in the WT and A198W system. Our data shows that such changes led to the establishment of contacts at the α TS/ β TS interface at four key positions in the A198W variant within the 200 ns simulation time, while such links were not formed or maintained in the WT enzyme on the same timescale (**Figure 6**). Furthermore, interactions observed for key residues within the indole

channel (β Tyr279, β Phe280) in the mutant system suggest that substrate channeling may be more efficient in the A198W variant. We detail some of the conformational dynamics and contact changes induced by the A198W substitution below.

As the simulations included both α TS and β TS subunits, we differentiate α TS and β TS residues and secondary structures by including the α and β designation in front of the secondary structure or residue. For example, the A198W substitution led to a displacement of α -helix 6 (α H6) towards α L6 (**Supplementary Figure 8**). This displacement allows α His195 to form a dynamic but consistent hydrogen bond with α Pro156 and a recurrent contact with α Asn157. A group of hydrophobic residues, α Leu177, α Phe212, and β Ile20 were also in consistent close contact. While it is not clear if this set of interactions induces the helix shift or vice versa, lack of it, as observed in the WT system, resulted in a formation of a small cluster between α Asn157, β Ile20, β Pro23 and β Tyr181. while the interactions between β Ile20 and α Leu177 and α Phe212 were interrupted (**Figure 6B**, **Supplementary Figure 9**). This collection of interactions may be important as the absence of the hydrophobic contact between β Ile20 and α Leu177 and α Phe212 may contribute to disordering of α L6, allowing it to move towards a more open conformation and providing a possible means of ligand escape from α TS (**Supplementary Figure 9**).

There were also additional changes to interactions involving α L6. For example, α Asn194 δ N formed a more stable hydrogen bond with α Ala226 O of helix 7 (α H7) in the A198W variant

compared to WT TS. A backbone hydrogen bond between α Leu191 O and α Gln210 N appeared to also be very strong in the A198W variant, while it was rarely observed in WT TS (**Supplementary Figure 10**). These findings were interesting considering that α Gln210 and α Ala226 appear to have suppressed conformational exchange on the millisecond timescale, at least according to the NMR studies on α TS alone (**Figure 4**). Consistent contact was also observed between α Ala185, α Ser215, and α Ser235 which may favor the closed conformation of α L6 in the A198W variant, considering that these interactions were interrupted in the WT enzyme.

A very critical aspect of the dynamics of α L6 in the A198W system was its consistent interactions with helix 6 of β TS (β H6), an important part of the COMM domain, well known to play a key role in the interdomain communication and activation of α TS (Dunn, 2012). Pairwise force data and hydrogen bond analysis showed a consistent hydrogen bond between α Thr183 O and β Arg175 guanidino group in over 70% of the frames analyzed. This hydrogen bond was not maintained in any of the trajectories of the simulated WT system (**Figure 6D**, **Supplementary Figure 11**). In addition to this hydrogen bond, a hydrophobic interaction was established between α Val182 and the hydrocarbon portion of the β Arg175 sidechain. Loss of these interactions as observed in the WT system and in one of the trajectories for the A198W system resulted in displacement and disordering of α L6. In the A198W trajectory with the absent hydrogen bond between α Thr183 and β Arg175, a hydrophobic interaction between α Val182 and β Ser178 and β Gly179 was maintained but it was not sufficient to prevent the loop displacement.

Another series of interactions correlated with the events occurring at the α L6 and β H6 interface - interactions of α Gln65 and α Asn66 with β Ser161 and β Gly162 (**Figure 6A**). Dihedral analysis showed one distinct conformation for the C – α C – β C – γ C angle of α Gln65 and α Asn66 in the A198W variant (**Supplementary Figure 11**). This conformation was associated with a hydrogen bond between α Gln65 and β Ser161 (68% in A198W and less than 3% in WT). In addition, electrostatic (salt-bridge like) interaction between α Gln65 and beta residues β Asp168 and β Asn171 was observed in the A198W variant. β Asp168 and β Asn171 were also part of β H6. A hydrogen bond between δ O of α Asn66 and the backbone of β Gly162 was observed in over 50% of all frames for the A198W system, and less than 20% in the WT enzyme. Another hydrogen bond between the backbone nitrogen of α Gln65 and α Gly61 was also more stable in the A198W variant. These results are interesting considering that the A198W substitution induced substantial chemical shift changes in both α Gly61 and α Ala67 (**Figure 3**). Unfortunately, the backbone resonances of α Gln65 and α Asn66 are unassigned. Nonetheless, the NMR results indicate that there are substantial structural dynamic changes in this region even in the absence of β TS. The A198W substitution also induced millisecond conformational exchange in α Ala67 (**Figure 4**), suggesting that this region may be seeking alternative binding interactions.

The interactions of α Gln65 and α Asn66 with the small loop between beta sheet 5 (β S5) and beta helix 6 (β H6) and α Thr183-

β Arg175 hydrogen bond appeared to also affect the position of β H6 relative to α TS. In the A198W variant the displacement occurred in the direction of α TS whereas in WT enzyme, lack of consistent interactions led to displacement in the opposite direction away from the α TS (**Supplementary Figure 12**). The difference in the β H6 position also affected important residues lining the indole channel, namely β Tyr279 and β Phe280. Force analysis showed that both residues in the WT enzyme have stronger interactions with residues from β H6. A more consistent link between β Tyr279 and β Lys167 was observed in WT TS. A hydrogen bond between β Tyr279 and β Asn171 was also observed in one of the WT simulations. Such interactions position the phenol ring of β Tyr279 within the tunnel possibly interfering with the indole transfer (Weyand and Schlichting, 2000). Similarly, β Phe280 in WT enzyme showed stronger interactions with β Cys170 of β H6 which brought the phenyl sidechain within the indole channel. In the A198W variant, these residues favored interactions with beta residues 306–308 on the channel “wall.” A hydrogen bond between β Tyr279 and α Asp56 was observed in 40% of MD frames in the A198W variant compared to 10% in WT TS. Such interaction may stabilize the phenol ring of β Tyr279 in an “open channel” conformation (**Figure 7**).

Other interactions induced by the A198W substitution may also be important for α TS- β TS communication, although the mechanism is unclear. For example, α Asn104 favors interactions with β TS in the A198W variant (**Figure 6C**). In WT TS, α Asn104 formed a very consistent hydrogen bond with α Asp130 (> 60% frames analyzed) and in much smaller extent with the backbone of β Ile278. In the A198W variant, a dynamic hydrogen bond network formed between α Asn104 and β Gln 288/ β Gly 292; these beta residues form part of the allosteric “metal binding loop” (Weyand and Schlichting, 1999). NMR studies indicated that the A198W substitution induced chemical shift perturbations for α Asn104 and α Asp130.

DISCUSSION

Amino acid interactions networks have been proposed to be the means through which allosteric signals are propagated across a protein's structure (Dokholyan, 2016). Understanding such networks has practical importance in protein engineering (Lee et al., 2008; Reynolds, McLaughlin and Ranganathan, 2011; Gorman et al., 2019) and allosteric drug design (Nussinov and Chung, 2015; Greener and Sternberg, 2018). It should be noted that Hilser and colleagues have proposed allosteric models that do not require “networks” to transmit information across a protein framework. Instead, these models focus on changes within the conformational ensemble using a free energy landscape framework (Motlagh et al., 2014). However, we believe these views of allostery can be compatible (D'Amico et al., 2020). In fact, the CHESCA approach identifies network residues as those being involved in the same conformational change (Selvaratnam et al., 2011), but does not necessitate that there are direct interactions between such residues.

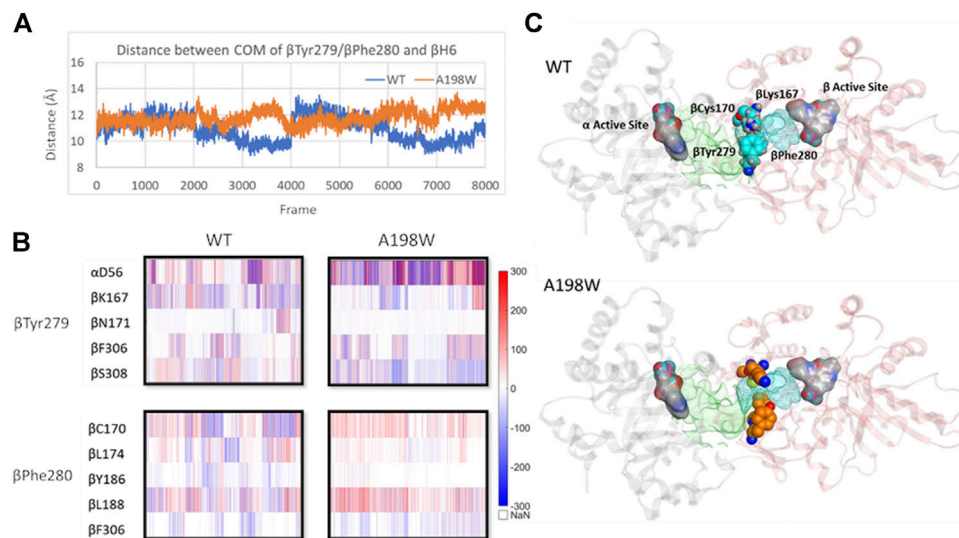


FIGURE 7 | Displacement of β H6 towards α TS leads to widening of indole channel which is further supported by weaker interactions between the channel lining residues β Tyr279 and β Phe280 with residues of β H6 in A198W. **(A)** Distance between center of mass of residues β Tyr279/ β Phe280 and β H6 lining the affected tunnel portion **(B)** Pairwise force distribution analysis indicates that interactions for residues β Tyr279 and β Phe280 have different trends in WT and A198W. **(C)** Interactions between β Tyr279 and β Lys167 position the aromatic ring within the indole channel. Active sites in both subunits (grey- α subunit, pink- β subunit) are represented as atom colored surface. Indole path is shown in green surface and the affected portion of the channel is mapped in cyan. See also **Supplementary Figure 12**.

We have previously identified networks in α TS using a CHESCA-type approach (Axe and Boehr, 2013; Axe et al., 2014; O'Rourke et al., 2019), and have found that network substitutions can have detrimental effects on α TS catalytic activity (Axe et al., 2014). We developed a cell-based assay to screen additional α TS variants by connecting function to cellular growth rates. We found that the surface-exposed, residue position 198, distant from both the α TS active site and the α TS/ β TS interface (**Figure 1**), was sensitive to amino acid substitutions. Specifically, we discovered α TS variants (A198K, A198V) that failed to support growth of tryptophan auxotrophic *E. coli* cells and discovered a variant (A198W) that surprisingly enhanced the growth rate of these cells (**Figure 2**). While the enhanced function was modest, it was similar to the enhancement of β TS activity by previous protein engineering efforts by the Arnold lab (Buller et al., 2015; Murciano-Calles et al., 2016). The A198W substitution likely enhances TS activity by modulating conformational dynamics involving other network residues (**Figures 3, 4**) to communicate with β TS (**Figures 5, 6**) and help open the indole channel (**Figure 7**). It is remarkable that the effect of the A198W substitution is only realized in the full TS complex, suggesting that the intrinsic networks we previously identified in α TS alone have functional roles in the full TS enzyme. It will be likewise informative to evaluate networks in β TS using the same methods we have developed for α TS, as well as further test the α TS network using the methods developed here.

We note that the rate limiting step for the α TS reaction seems to be a conformational change that occurs after IGP binds, and this step is sped up significantly in the presence of

serine-bound β TS (Anderson et al., 1991). For the β TS reaction, the rate limiting step is the proton abstraction from the external aldimine, which is also highly influenced by the presence of ligand-bound α TS (Ngo, et al., 2007). For the complete TS reaction, it is postulated that the release of tryptophan is at least partially rate limiting. In this context, and given the kinetic, NMR and MD simulation results, it would seem that the A198W substitution enhances TS function by modulating dynamic interactions at the α TS- β TS interface, which promotes β TS activity, possibly enhancing release of tryptophan product. It is also enticing to suggest that the substitution boosts indole channeling, due to the higher enhancements of the full TS reaction compared to the β TS reaction alone, and the indication in the MD simulations that the indole channel is more open, more often in the A198W variant. In a similar manner, it has been suggested that the enhancing mutations in *P. furiosus* β TS allosterically modulate conformational dynamics related to the conformational cycle necessary for the complex β TS reaction (Maria-Solano et al., 2019). We also note that most of the amino acids involved in the α TS- β TS interaction and/or form the indole channel are conserved (see **Supplementary Figures 13 and 14**) or have co-evolved to maintain interactions (see **Supplementary Figure 6**). While Ala198 itself is not conserved, other residues in the NMR-derived network are conserved or co-evolve (see **Supplementary Figure 6**), and so similar networks might exist in these other TS enzymes. Such conformational dynamics and allosteric networks may be important for the evolution of the communication between the α TS and β TS subunits

(Schupfner et al., 2020). As such, other network substitutions in *E. coli* α TS and other α TS enzymes might likewise activate TS function through a similar mechanism.

More generally, these results indicate that amino acid interaction networks may not only be important for function within an enzyme subunit, but they may be used to bridge communication between subunits in a multienzyme complex, and only reveal such enhancements in the full complex. This finding reveals new possibilities for controlling function in such complexes; both individual enzyme function and communication between subunits might be modulated through engineering network residues. Similar studies have been performed on imidazole glycerol phosphate synthase (Rivalta et al., 2012; Lisi et al., 2016; Lisi et al., 2017), which is a histidine biosynthetic, bifunctional enzyme also demonstrating substrate channeling. Our results reported here are unique in that the network substitution enhanced enzyme function, atypical for any mutational study, and the substitution was surface-exposed. Modifying surface-exposed network residues are unlikely to change the protein fold or substantially affect protein stability, and thus, offer a novel and potentially useful strategy in enzyme design for overall rate enhancement.

Networks in α TS not only connect the α TS active site to the α TS/ β TS interface, but also connect to other protein surfaces (Axe et al., 2014; O'Rourke et al., 2019). While these network surfaces might be targeted by novel allosteric modulators (Nussinov and Chung, 2015; Greener and Sternberg, 2018) or be used to graft novel regulatory units (Lee et al., 2008; Reynolds et al., 2011; Dokholyan, 2016), it is also intriguing to ask why these network surfaces even exist. Are allosteric networks simply intrinsic to all proteins (Gunasekaran et al., 2004)? Or has evolution shaped these allosteric surfaces for other purposes? For example, metabolites or other enzymes involved in tryptophan metabolism might interact, albeit weakly, with these network surfaces to modulate TS function. The fast cell growth phenotype engendered by the A198W substitution might partially or fully arise because it modifies unidentified protein-protein interactions, instead of owing solely to changes to enzyme activity. Other "quinary" interactions (Cohen and Pielak, 2017) may likewise exert their effects through network surfaces on other proteins. NMR, MD, high throughput mutational screening (Wrenbeck et al., 2017) and other network methods (O'Rourke et al., 2016) can be used to evaluate the physical basis and molecular evolution of networks in other enzyme complexes, potentially revealing how allosteric networks connect the larger proteome (Cong et al., 2019).

REFERENCES

- Abrahams, K. A., Fütterer, K., Rullas, J., Ortega-MuroLoman, F. N. J., Moynihan, P. J., et al. (2017). Inhibiting Mycobacterial Tryptophan Synthase by Targeting the Inter-subunit Interface. *Sci. Rep.* 7 (1), 9430. doi:10.1038/s41598-017-09642-y
- Ai, R., Qaiser Fatmi, M., and Chang, C. E. A. (2010). T-analyst: A Program for Efficient Analysis of Protein Conformational Changes by Torsion Angles.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

RD, YB, KO, CC, and DB contributed to conception and design of the study. RD and AM performed and developed the cell-based assays. RD analyzed the cell-based assays and performed statistical analysis. RD and KO generated the protein variants, expressed and purified proteins. RD performed the kinetic assays. KO collected the NMR data. KO and DB analyzed the NMR data. WA performed the amino acid covariation and frustratometer studies. WA and DB analyzed the amino acid covariation and frustratometer results. YB collected the MD data. YB and CC analyzed the MD data. RD, YB, CC, and DB wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

We acknowledge the support from the NSF National Supercomputer Centers (TG-CHE130009). This work was supported by NSF Grants MCB 1932984 (CAC) and MCB 1615032 (DDB).

ACKNOWLEDGMENTS

We would like to thank Emery Usher, Dr. Scott Gorman and Dr. Scott Showalter for assistance with the ITC experiments, Dennis Winston for assistance with MatLab, and the former and current staff at the Penn State NMR Facility (Drs. Emmanuel Hatzakis, Tapas Mal, Debashish Sahu) for assistance with the NMR experiments. We would also like to thank Alyson Boehr and other members of the Boehr and Chang labs for helpful comments on this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.679915/full#supplementary-material>

J. Computer-Aided Mol. Des. 24 (10), 819–827. doi:10.1007/s10822-010-9376-y

Anderson, K. S., Miles, E. W., and Johnson, K. A. (1991). Serine Modulates Substrate Channeling in Tryptophan Synthase: A Novel Intersubunit Triggering Mechanism. *J. Biol. Chem.* 266 (13), 8020–8033. doi:10.1016/s0021-9258(18)92934-0

Axe, J. M., and Boehr, D. D. (2013). Long-Range Interactions in the Alpha Subunit of Tryptophan Synthase Help to Coordinate Ligand Binding, Catalysis, and

- Substrate Channeling. *J. Mol. Biol.* 425 (9), 1527–1545. doi:10.1016/j.jmb.2013.01.030
- Axe, J. M., Yezdimer, E. M., F. Orourke, K., Kerstetter, N. E., You, W., Chang, C. E. A., et al. (2014). Amino Acid Networks in a (β/α)8 Barrel Enzyme Change during Catalytic Turnover. *J. Am. Chem. Soc.* 136 (19), 6818–6821. doi:10.1021/ja501602tKerstetter
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of Escherichia Coli K-12 In-Frame, Single-Gene Knockout Mutants: The Keio Collection. *Mol. Syst. Biol.* 2, 218. doi:10.1038/msb4100050
- Botello-Smith, W. M., and Luo, Y. (2019). Robust Determination of Protein Allosteric Signaling Pathways. *J. Chem. Theor. Comput.* 15 (4), 2116–2126. doi:10.1021/acs.jctc.8b01197
- Brzovic, P. S., Ngo, K., and Dunn, M. F. (1992). Allosteric Interactions Coordinate Catalytic Activity between Successive Metabolic Enzymes in the Tryptophan Synthase Bienenzyme Complex. *Biochemistry* 31 (15), 3831–3839. doi:10.1021/bi00130a014
- Buller, A. R., Brinkmann-ChenRomney, S. D. K., Herger, M., Murciano-Calles, J., Frances, H., and Arnold, J. (2015). Directed Evolution of the Tryptophan Synthase β -Subunit for Stand-Alone Function Recapitulates Allosteric Activation. *Proc. Natl. Acad. Sci. U. S. A.* 112 (47), 14599–14604. doi:10.1073/pnas.1516401112
- Buller, A. R., Paul Van Roye, R., Jackson, K. B., Remkes, A., Herger, M., and Arnold, F. H. (2018). Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble. *J. Am. Chem. Soc.* 140 (23), 7256–7266. doi:10.1021/jacs.8b03490
- Case, D. A., Walker, R. C., Cheatham, T. E., Simmerling, C., Roitberg, A., and Merz, K. M. (2018). *Amber 2018*. San Francisco: University of California. Available at: <http://ambermd.org/doc12/Amber18.pdf>
- Cohen, R. D., and Pielak, G. J. (2017). A Cell Is More Than the Sum of its (Dilute) Parts: A Brief History of Quinary Structure. *Protein Sci.* 26 (3), 403–413. doi:10.1002/pro.3092
- Cong, Q., Anisichenko, I., Ovchinnikov, S., and Baker, D. (2019). Protein Interaction Networks Revealed by Proteome Coevolution. *Science* 365 (6449), 185–189. doi:10.1126/science.aaw6718
- Creighton, T. E. (1970). A Steady-State Kinetic Investigation of the Reaction Mechanism of the Tryptophan Synthetase of *Escherichia Coli*. *Eur. J. Biochem.* 13 (1), 1–10. doi:10.1111/j.1432-1033.1970.tb00892.x
- D'Amico, R. N., Murray, A. M., and Boehr, D. D. (2020). Driving Protein Conformational Cycles in Physiology and Disease: 'Frustrated' Amino Acid Interaction Networks Define Dynamic Energy Landscapes: Amino Acid Interaction Networks Change Progressively along Alpha Tryptophan Synthase's Catalytic Cycle. *BioEssays* 42 (9), 2000092. doi:10.1002/bies.202000092
- Datsenko, K. A., and Wanner, B. L. (2000). One-Step Inactivation of Chromosomal Genes in *Escherichia Coli* K-12 Using PCR Products. *Proc. Natl. Acad. Sci. United States America* 97 (12), 6640–6645. doi:10.1073/pnas.120163297
- Dokholyan, N. V. (2016). Controlling Allosteric Networks in Proteins. *Chem. Rev.* 116 (11), 6463–6487. doi:10.1021/acs.chemrev.5b00544
- Dos Santos, R. N., Jiang, X., Martínez, L., and Morcos, F. (2019). Coevolutionary Signals and Structure-Based Models for the Prediction of Protein Native Conformations. *Methods Mol. Biol.* 1851, 83–103. doi:10.1007/978-1-4939-8736-8_5
- Dunn, M. F., Aguilar, V., Peter, B., Catherine, A., William, F., Houben, K. F., et al. (1990). The Tryptophan Synthase Bienenzyme Complex Transfers Indole between the α - and (β -Sites via a 25–30 Å Long Tunnel. *Biochemistry* 29 (37), 8598–8607. doi:10.1021/bi00489a015
- Dunn, M. F. (2012). Allosteric Regulation of Substrate Channeling and Catalysis in the Tryptophan Synthase Bienenzyme Complex. *Arch. Biochem. Biophys.* 519 (2), 154–166. doi:10.1016/j.abb.2012.01.016
- East, K. W., Skeens, E., Jennifer, Y., Brandon, M., Hsu, R., Batista, V. S., et al. (2020). NMR and Computational Methods for Molecular Resolution of Allosteric Pathways in Enzyme Complexes. *Biophysical Rev.* 12 (1), 155–174. doi:10.1007/s12551-019-00609-zCui
- Faeder, E. J., and Hammes, G. G. (1970). Kinetic Studies of Tryptophan Synthetase. Interaction of Substrates with the B Subunit. *Biochemistry* 9 (21), 4043–4049. doi:10.1021/bi00823a003
- Feher, V. A., Durrant, J. D., Van Wart, A. T., and Amaro, R. E. (2014). Computational Approaches to Mapping Allosteric Pathways. *Curr. Opin. Struct. Biol.* 25, 98–103. doi:10.1016/j.sbi.2014.02.004
- Ferreiro, D. U., Hegler, J. A., Elizabeth, A., and Wolynes, P. G. (2007). Localizing Frustration in Native Proteins and Protein Assemblies. *Proc. Natl. Acad. Sci. United States America* 104 (50), 19819–19824. doi:10.1073/pnas.0709915104
- Ferreiro, D. U., Komives, E. A., and Wolynes, P. G. (2014). Frustration in Biomolecules. *Q. Rev. Biophys.* 47 (4), 285–363. doi:10.1017/S0033583514000092
- Ferreiro, D. U., Komives, E. A., and Wolynes, P. G. (2018). Frustration, Function and Folding. *Curr. Opin. Struct. Biol.* 48, 68–73. doi:10.1016/j.sbi.2017.09.006
- Fisher, M. A., McKinley, K. L., Bradley, L. H., Viola, S. R., and Hecht, M. H. (2011). De Novo Designed Proteins from a Library of Artificial Sequences Function in *Escherichia Coli* and Enable Cell Growth. *PLoS ONE* 6 (1), e15364. doi:10.1371/journal.pone.0015364
- Freiberger, M. I., Brenda Guzovsky, A., Wolynes, P. G., Gonzalo Parra, R., and Ferreiro, D. U. (2019). Local Frustration Around Enzyme Active Sites. *Proc. Natl. Acad. Sci. U. S. A.* 116 (10), 4037–4043. doi:10.1073/pnas.1819859116
- Gorman, S. D., D'Amico, R. N., Winston, D. S., and Boehr, D. D. (2019). Engineering Allostery into Proteins. *Adv. Exp. Med. Biol.* 1163, 359–384. doi:10.1007/978-981-13-8719-7_15
- Greener, J. G., and Sternberg, M. J. E. (2018). Structure-Based Prediction of Protein Allostery. *Curr. Opin. Struct. Biol.* 50, 1–8. doi:10.1016/j.sbi.2017.10.002
- Gunasekaran, K., Ma, B., and Nussinov, R. (2004). Is Allostery an Intrinsic Property of All Dynamic Proteins?. *Proteins: Struct. Funct. Genet.* 57 (3), 433–443. doi:10.1002/prot.20232
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual Molecular Dynamics. *J. Mol. Graphics* 14 (1), 33–38. doi:10.1016/0263-7855(96)00018-5
- Jakalian, A., Jack, D. B., and Bayly, C. I. (2002). Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* 23 (16), 1623–1641. doi:10.1002/jcc.10128
- Johnson, Q. R., Lindsay, R. J., and Shen, T. (2018). CAMERRA: An Analysis Tool for the Computation of Conformational Dynamics by Evaluating Residue-Residue Associations. *J. Comput. Chem.* 39 (20), 1568–1578. doi:10.1002/jcc.25192
- Jorgensen, W. L., Chandrasekhar, J., Jeffrey, D., Roger, W. I., and Klein, M. L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79 (2), 926–935. doi:10.1063/1.445869
- Koshland, D. E., Nemethy, J. G., and Filmer, D. (1966). Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* 5 (1), 365–385. doi:10.1021/bi00865a047
- Lane, A. N., and Kirschner, K. (1991). Mechanism of the Physiological Reaction Catalyzed by Tryptophan Synthase from *Escherichia Coli*. *Biochemistry* 30 (2), 479–484. doi:10.1021/bi00216a025
- Lee, J., Natarajan, M., Nashine, V. C., Michael, S., Vo, T., Russ, W. P., et al. (2008). Surface Sites for Engineering Allosteric Control in Proteins. *Science* 322 (5900), 438–442. doi:10.1126/science.1159052
- Lisi, G. P. P., Manley, G. A. A., Hendrickson, H., Rivalta, I., Batista, V. S. S., and Patrick Loria, J. (2016). Dissecting Dynamic Allosteric Pathways Using Chemically Related Small-Molecule Activators. *Structure* 24 (7), 1155–1166. doi:10.1016/j.str.2016.04.010
- Lisi, G. P., East, K. W., Batista, V. S., and Patrick Loria, J. (2017). Altering the Allosteric Pathway in IGPS Suppresses Millisecond Motions and Catalytic Activity. *Proc. Natl. Acad. Sci. U. S. A.* 114 (17), E3414–E3423. doi:10.1073/pnas.1700448114
- Loria, J., Patrick, Berlow, Rebecca, B., and Watt, Eric, D. (2008). Characterization of Enzyme Motions by Solution NMR Relaxation Dispersion. *Acc. Chem. Res.* 41 (2), 214–221. doi:10.1021/ar700132n
- Loria, J., Rance, M., and Palmer, A. G. (1999). A TROSY CPMG Sequence for Characterizing Chemical Exchange in Large Proteins. *J. Biomol. NMR* 15 (2), 151–155. doi:10.1023/A:1008355631073
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theor. Comput.* 11 (8), 3696–3713. doi:10.1021/acs.jctc.5b00255
- Maria-Solano, A., Miguel, A., Iglesias-Fernández, J., and Osuna, S. (2019). Deciphering the Allosterically Driven Conformational Ensemble in Tryptophan Synthase Evolution. *J. Am. Chem. Soc.* 141 (33), 13049–13056. doi:10.1021/jacs.9b03646

- Monod, J., Wyman, J., and Changeux, J. P. (1965). On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* 12 (1), 88–118. doi:10.1016/S0022-2836(65)80285-6
- Morcos, F., and Onuchic, J. N. (2019). The Role of Coevolutionary Signatures in Protein Interaction Dynamics, Complex Inference, Molecular Recognition, and Mutational Landscapes. *Curr. Opin. Struct. Biol.* 56, 179–186. doi:10.1016/j.sbi.2019.03.024
- Morrison, J. F. (1969). Kinetics of the Reversible Inhibition of Enzyme-Catalysed Reactions by Tight-Binding Inhibitors. *BBA - Enzymol.* 185 (2), 269–286. doi:10.1016/0005-2744(69)90420-3
- Motlagh, H. N., Wrabl, J. O., Jing, L., and Hilser, V. J. (2014). The Ensemble Nature of Allostery. *Nature* 508 (7496), 331–339. doi:10.1038/nature13001
- Murciano-Calles, J., Romney, D. K., Brinkmann-Chen, S., Buller, A. R., and Arnold, F. H. (2016). A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations that Mimic Allosteric Activation. *Angew. Chem. - Int. Edition* 55 (38), 11577–11581. doi:10.1002/anie.201606242
- Ngo, H., Harris, R., Kimmich, N., Casino, P., Niks, D., Blumenstein, L., et al. (2007a). Synthesis and Characterization of Allosteric Probes of Substrate Channeling in the Tryptophan Synthase Bienenzyme Complex. *Biochemistry* 46 (26), 7713–7727. doi:10.1021/bi700385f
- Ngo, H., Kimmich, N., Harris, R., Niks, D., Blumenstein, L., Kulik, V., et al. (2007b). Allosteric Regulation of Substrate Channeling in Tryptophan Synthase: Modulation of the L-Serine Reaction in Stage I of the β -Reaction by α -Site Ligands. *Biochemistry* 46 (26), 7740–7753. doi:10.1021/bi7003872
- Niks, D., Hilario, E., Adam, D., Ngo, H., Borchardt, D., Neubauer, T. J., et al. (2013). Allostery and Substrate Channeling in the Tryptophan Synthase Bienenzyme Complex: Evidence for Two Subunit Conformations and Four Quaternary States. *Biochemistry* 52 (37), 6396–6411. doi:10.1021/bi400795e
- Nussinov, R., and Chung, J. (2015). The Design of Covalent Allosteric Drugs. *Annu. Rev. Pharmacol. Toxicol.* 55 (1), 249–267. doi:10.1146/annurev-pharmtox-010814-124401
- O'Rourke, K. F., Gorman, S. D., and Boehr, D. D. (2016). Biophysical and Computational Methods to Analyze Amino Acid Interaction Networks in Proteins. *Comput. Struct. Biotechnol. J.* 14, 245–251. doi:10.1016/j.csbj.2016.06.002
- O'Rourke, K. F., Jennifer, M., D'Amico, R. N., Sahu, D., and Boehr, D. D. (2018). Millisecond Timescale Motions Connect Amino Acid Interaction Networks in Alpha Tryptophan Synthase. *Front. Mol. Biosciences* 5, 92. doi:10.3389/fmolb.2018.00092
- O'Rourke, K. F., Sahu, D., Yuliana, K., D'Amico, R. N., Chia, en, A. C., and Boehr, D. D. (2019). Coordinated Network Changes across the Catalytic Cycle of Alpha Tryptophan Synthase. *Structure* 27 (9), 1405–1415. doi:10.1016/j.str.2019.05.013
- Parra, R. G., Schafer, N. P., Radusky, L. G., Tsai, M.-Y., Guzovsky, A. B., Wolynes, P. G., et al. (2016). Protein Frustratometer 2: A Tool to Localize Energetic Frustration in Protein Molecules, Now with Electrostatics. *Nucleic Acids Res.* 44 (W1), W356–W360. doi:10.1093/nar/gkw304
- Reynolds, K. A., McLaughlin, R. N., and Rama, R. (2011). Hot Spots for Allosteric Regulation on Protein Surfaces. *Cell* 147 (7), 1564–1575. doi:10.1016/j.cell.2011.10.049
- Rivalta, I., Sultan, M. M., Lee, N. S., Manley, G. A., Patrick Loria, J., and Batista, V. S. (2012). Allosteric Pathways in Imidazole Glycerol Phosphate Synthase. *Proc. Natl. Acad. Sci. U. S. A.* 109 (22), E1428–E1436. doi:10.1073/pnas.1120536109
- Rivoire, O., Reynolds, K. A., and Rama, R. (2016). Evolution-Based Functional Decomposition of Proteins. *PLoS Comput. Biol.* 12 (6), e1004817. doi:10.1371/journal.pcbi.1004817
- Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theor. Comput.* 9 (7), 3084–3095. doi:10.1021/ct400341p
- Ryckaert, J. P., Ciccotti, G., Herman, J., and Berendsen, C. (1977). Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* 23 (3), 327–341. doi:10.1016/0021-9991(77)90098-5
- Sagui, C., Pedersen, L. G., and Darden, T. A. (2004). Towards an Accurate Representation of Electrostatics in Classical Force Fields: Efficient Implementation of Multipolar Interactions in Biomolecular Simulations. *J. Chem. Phys.* 120 (1), 73–87. doi:10.1063/1.1630791
- Salomon-Ferrer, R., Götz, A. W., Duncan, P., Scott, L., Ross, C., and Walker, A. (2013). Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theor. Comput.* 9 (9), 3878–3888. doi:10.1021/ct400314y
- Schupfner, M., Straub, K., Busch, F., Merkl, R., and Reinhard Sterner, D. (2020). Analysis of Allosteric Communication in a Multienzyme Complex by Ancestral Sequence Reconstruction. *Proc. Natl. Acad. Sci. U. S. A.* 117 (1), 346–354. doi:10.1073/pnas.1912132117
- Selvaratnam, R., Chowdhury, S., Bryan, V. S., and Melacini, G. (2011). Mapping Allostery through the Covariance Analysis of NMR Chemical Shifts. *Proc. Natl. Acad. Sci. U. S. A.* 108 (15), 6133–6138. doi:10.1073/pnas.1017311108
- Somboonna, N., Ziklo, N., Ferrin, T. E., Jung, H., and Dean, D. (2019). Clinical Persistence of Chlamydia Trachomatis Sexually Transmitted Strains Involves Novel Mutations in the Functional $\alpha\beta\beta\alpha$ Tetramer of the Tryptophan Synthase Operon. *MBio* 10 (4), e01464–19. doi:10.1128/mBio.01464-19Suh
- Sprouffske, K., and Wagner, A. (2016). Growthcurver: An R Package for Obtaining InterpretMetrics from Microbial Growth Curves. *BMC Bioinformatics* 17 (1), 172. doi:10.1186/s12859-016-1016-7
- Stacklies, W., Seifert, C., and Graeter, F. (2011). Implementation of Force Distribution Analysis for Molecular Dynamics Simulations. *BMC Bioinformatics* 12, 101. doi:10.1186/1471-2105-12-101
- Sugase, K., Konuma, T., Lansing, J. C., and Wright, P. E. (2013). Fast and Accurate Fitting of Relaxation Dispersion Data Using the Flexible Software Package GLOVE. *J. Biomol. NMR* 56 (3), 275–283. doi:10.1007/s10858-013-9747-5
- Wang, J., Jain, A., McDonald, L. R., Craig, G., Lee, A. L., and Dokholyan, N. V. (2020). Mapping Allosteric Communications within Individual Proteins. *Nat. Commun.* 11 (1), 13. doi:10.1038/s41467-020-17618-2
- Wellington, S., Nag, P. P., Michalska, K., Johnston, S. E., Robert, P. J., Kaushik, V. K., Clatworthy, A. E., et al. (2017). Small-Molecule Allosteric Inhibitor of Mycobacterium Tuberculosis Tryptophan Synthase. *Nat. Chem. Biol.* 13 (9), 943–950. doi:10.1038/nchembio.2420
- Weyand, M., and Schlichting, I. (1999). Crystal Structure of Wild-type Tryptophan Synthase Complexed with the Natural Substrate Indole-3-Glycerol Phosphate. *Biochemistry* 38 (50), 16469–16480. doi:10.1021/bi9920533
- Weyand, M., and Schlichting, I. (2000). Structural Basis for the Impaired Channeling and Allosteric Inter-subunit Communication in the BA169L/BC170W Mutant of Tryptophan Synthase. *J. Biol. Chem.* 275 (52), 41058–41063. doi:10.1074/jbc.C000479200
- Wodak, S. J., Paci, E., Nikolay, V., Igor, N. H., Horowitz, A., Jing, L., et al. (2019). Allostery in its Many Disguises: From Theory to Applications. *Structure* 27 (4), 566–578. doi:10.1016/j.str.2019.01.003
- Wrenbeck, E. E., Faber, M. S., and Whitehead, T. A. (2017). Deep Sequencing Methods for Protein Engineering and Design. *Curr. Opin. Struct. Biol.* 45, 36–44. doi:10.1016/j.sbi.2016.11.001
- Zeng, H., Wang, S., Zhou, T., Zhao, F., Li, X., Wu, Q., et al. (2018). ComplexContact: A Web Server for Inter-protein Contact Prediction Using Deep Learning. *Nucleic Acids Res.* 46 (W1), W432–W437. doi:10.1093/nar/gky420

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 D'Amico, Bosken, O'Rourke, Murray, Admasu, Chang and Boehr. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Single Point Mutation Controls the Rate of Interconversion Between the g^+ and g^- Rotamers of the Histidine 189 χ_2 Angle That Activates Bacterial Enzyme I for Catalysis

Jeffrey A. Purslow¹, Jolene N. Thimmesch¹, Valeria Sivo², Trang T. Nguyen¹, Balabhadra Khatiwada¹, Rochelle R. Dotas¹ and Vincenzo Venditti^{1,3*}

¹Department of Chemistry, Iowa State University, Ames, IA, United States, ²Department of Environmental, Biological and Pharmaceutical Sciences and Technologies, Università Degli Studi Della Campania, Caserta, Italy, ³Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA, United States

OPEN ACCESS

Edited by:

Silvina Ponce Dawson,
University of Buenos Aires, Argentina

Reviewed by:

Felipe Trajtenberg,
Institut Pasteur de Montevideo,
Uruguay
Srabani Taraphder,
Indian Institute of Technology
Kharagpur, India

*Correspondence:

Vincenzo Venditti
venditti@iastate.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 23 April 2021

Accepted: 29 June 2021

Published: 08 July 2021

Citation:

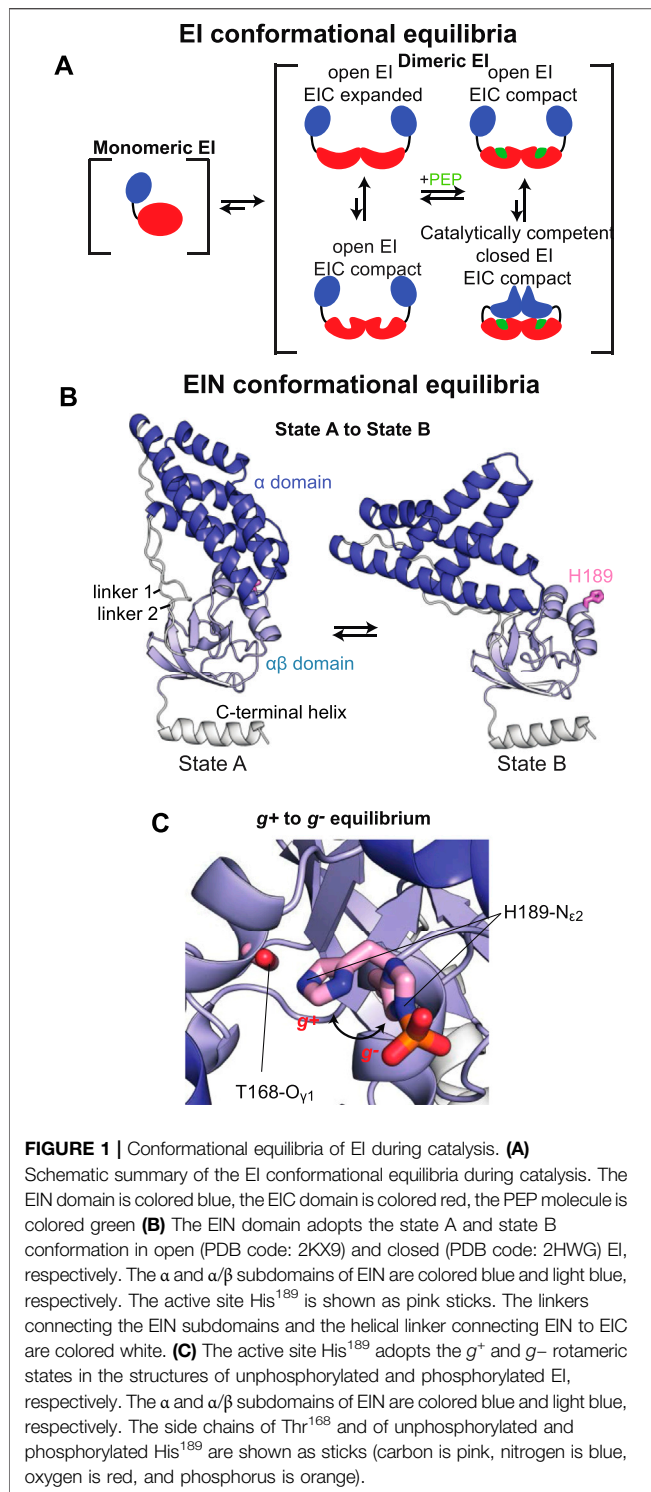
Purslow JA, Thimmesch JN, Sivo V, Nguyen TT, Khatiwada B, Dotas RR and Venditti V (2021) A Single Point Mutation Controls the Rate of Interconversion Between the g^+ and g^- Rotamers of the Histidine 189 χ_2 Angle That Activates Bacterial Enzyme I for Catalysis. *Front. Mol. Biosci.* 8:699203. doi: 10.3389/fmolb.2021.699203

Enzyme I (EI) of the bacterial phosphotransferase system (PTS) is a master regulator of bacterial metabolism and a promising target for development of a new class of broad-spectrum antibiotics. The catalytic activity of EI is mediated by several intradomain, interdomain, and intersubunit conformational equilibria. Therefore, in addition to its relevance as a drug target, EI is also a good model for investigating the dynamics/function relationship in multidomain, oligomeric proteins. Here, we use solution NMR and protein design to investigate how the conformational dynamics occurring within the N-terminal domain (EIN) affect the activity of EI. We show that the rotameric g^+ -to- g^- transition of the active site residue His¹⁸⁹ χ_2 angle is decoupled from the state A-to-state B transition that describes a $\sim 90^\circ$ rigid-body rearrangement of the EIN subdomains upon transition of the full-length enzyme to its catalytically competent closed form. In addition, we engineered EIN constructs with modulated conformational dynamics by hybridizing EIN from mesophilic and thermophilic species, and used these chimeras to assess the effect of increased or decreased active site flexibility on the enzymatic activity of EI. Our results indicate that the rate of the autophosphorylation reaction catalyzed by EI is independent from the kinetics of the g^+ -to- g^- rotameric transition that exposes the phosphorylation site on EIN to the incoming phosphoryl group. In addition, our work provides an example of how engineering of hybrid mesophilic/thermophilic chimeras can assist investigations of the dynamics/function relationship in proteins, therefore opening new possibilities in biophysics.

Keywords: NMR, conformational dynamics, thermophile, protein design, enzyme regulation

INTRODUCTION

Enzyme I (EI) is the first protein in the bacterial phosphotransferase system (PTS), a signal transduction pathway that controls multiple cellular functions including sugar uptake, catabolic gene expression, interactions between carbon and nitrogen metabolisms, chemotaxis, biofilm formation, and virulence, *via* phosphorylation-dependent protein-protein interactions (Postma



et al., 1993; Clore and Venditti, 2013). The phosphorylation state of EI dictates the phosphorylation state of all other downstream components of the PTS (Deutscher et al., 2014) and malfunction of EI has been linked to reduced growth-rate and attenuated virulence in several bacterial species (Edelstein et al., 1999; Jones et al., 2000; Lau et al., 2001; Kok et al., 2003). Given its central role

in controlling bacterial metabolism, EI has been proposed as a target for antimicrobial design (Kok et al., 2003; Huang et al., 2013; Nguyen and Venditti, 2020) or for metabolic engineering efforts aimed at developing more efficient systems for microbial production of chemicals from biomass feedstocks (Doucette et al., 2011; Venditti et al., 2013).

In addition to its relevance for pharmaceutical and biotech applications, EI is an ideal model system for investigating the interplay between ligand binding, post-translational modifications, and conformational dynamics that determine the activity of complex multidomain proteins. Indeed, EI is a 128 kDa dimeric enzyme (Chauvin and Brand, 1996) whose activity depends on the synergistic action of at least four conformational equilibria that results in a series of large intradomain, interdomain, and intersubunit structural rearrangements modulated by substrate binding and two subsequent protein phosphorylation steps (from the substrate to EI and from EI to HPr, the second protein of the PTS) (Figure 1A). The N-terminal phosphoryl-transfer domain (EIN, residues 1–249) contains the phosphorylation site (His¹⁸⁹) and the binding site for the phosphocarrier protein, HPr. The C-terminal domain (EIC, residues 261–575) is responsible for dimerization and contains the binding site for the substrate phosphoenolpyruvate (PEP) and the small molecule regulator α -ketoglutarate (α KG) (Chauvin and Brand, 1996; Venditti et al., 2013). A long helical linker connects the EIN and EIC domains. In the absence of substrate, EI adopts an open conformation in which the EIN domains of the two monomeric subunits are more than 60 Å apart (Schwieters et al., 2010). Binding of PEP induces a transition to the catalytically competent closed form of EI (Venditti et al., 2015a; Venditti et al., 2015b). In the closed structure, the EIN domains of the two monomeric subunits are in direct contact and the active site residue, His¹⁸⁹, is inserted in the catalytic pocket on EIC (Figure 1A) (Teplyakov et al., 2006).

In recent years, we have published several studies revealing that progressive quenching of the intradomain EIC dynamics is an important source of functional regulation that can be exploited to design allosteric inhibitors of EI. In particular, by using high-pressure NMR we have shown that dimerization of EI promotes substrate binding by providing structural stabilization to the EIC catalytic pocket (Nguyen et al., 2021). Coupling NMR relaxation experiments with Small Angle X-ray Scattering, we showed that binding of PEP results in further quenching of μ s-ms dynamics at the EIC catalytic loops that triggers the open-to-close interdomain rearrangement and activates EI for catalysis (Venditti et al., 2015b). Finally, by combining NMR with Molecular Dynamics (MD) simulations, we noticed that residual conformational heterogeneity at the EIC active site in the activated enzyme-substrate complex determines the enzymatic turnover (Dotas et al., 2020) and that perturbing conformational dynamics at the active site loops is an effective strategy to inhibit the phosphoryl-transfer reaction (Nguyen and Venditti 2020).

Despite the wealth of knowledge we possess about the coupling between EIC conformational flexibility and enzymatic activity, very little is known about if and how EIN conformational

dynamics impact the function of the enzyme. Indeed, while a comparison of the experimental atomic-resolution structures of EI indicates that the open-to-close conformational change is coupled to a rigid body reorientation of the α subdomain relative to the α/β subdomain of EIN (commonly referred to as state A-to-state B equilibrium, **Figure 1B**) and that protein phosphorylation shifts the χ^2 angle of His¹⁸⁹ from the g^+ to g^- rotameric state (**Figure 1C**), it is not clear if these conformational equilibria are active in isolated EIN and if their external perturbation can impact turnover. Addressing these questions will advance our understanding of how synergistic couplings among intradomain, interdomain, and intersubunit conformational equilibria affect the function of a multidomain oligomeric protein such as EI, and will provide new perspectives toward the development of EI inhibitors that act on the EIN domain.

Here we investigate the structure and dynamics of isolated EIN in its native and phosphorylated forms by solution NMR spectroscopy. While we do not detect evidence of an active state A/state B equilibrium, relaxation dispersion experiments indicate that the conformational exchange between the g^+ and g^- rotameric states of His¹⁸⁹ is active in the isolated EIN and modulated by protein phosphorylation. Furthermore, we engineered EIN constructs with modulated thermostability and conformational flexibility by hybridizing EIN from mesophilic and thermophilic organisms. Biophysical characterization of the wild-type and hybrid EIN constructs indicates that the rotameric equilibrium is slower in the thermophilic enzyme than in the mesophilic protein and that a single serine to alanine mutation is responsible for the increased activation energy in the thermophilic species. Finally, we performed functional characterization of the wild-type mesophilic and thermophilic EI, as well as of EI constructs that incorporate the hybridizing mutations. Our data show that, although the His¹⁸⁹ rotameric equilibrium is required for the correct functioning of EI, the rate of the phosphoryl transfer reaction is independent on the kinetics of the conformational change, indicating that the g^+ -to- g^- transition is not the rate limiting step for catalysis.

MATERIALS AND METHODS

Protein Expression and Purification

E. coli and *T. tengcongensis* EI, EIN, and HPr were expressed and purified as previously reported (Suh et al., 2008; Venditti and Clore, 2012; Dotas and Venditti, 2019). Single point mutations were introduced using the QuikChange site-directed mutagenesis.

Thermal Stability and Circular Dichroism

Thermal-induced unfolding of EIN was investigated in a 1 mm, 400 μ l, quartz cuvette sample cell using a Jasco J-710 spectropolarimeter. Samples were prepared in H₂O at a protein concentration of \sim 0.5 mg/ml. Ellipticity ($\theta_{222\text{nm}}$) at the 222 nm wavelength was monitored over a 1°C/min temperature gradient ranging from 35 to 75°C and 65–95°C for mesophilic and

thermophilic EIN, respectively. The melting temperature (T_m) was calculated as the maximum value of the derivative of $\theta_{222\text{nm}}$ with respect to temperature.

Nuclear Magnetic Resonance Spectroscopy

NMR samples were prepared in 20 mM Tris-HCl buffer (pH 7.4), 100 mM NaCl, 4 mM MgCl₂, 1 mM ethylenediaminetetraacetic acid (EDTA), 2 mM dithiothreitol (DTT), and 90% H₂O/10% D₂O (v/v). For protein phosphorylation, samples were incubated for 1 h at 37°C with <1 μ M of mesophilic EI (eEI) or thermophilic EI (tEI), <1 μ M of mesophilic HPr (eHPr) or thermophilic HPr (tHPr), and \sim 30 mM of PEP prior to acquisition. Completion of phosphorylation reactions were confirmed by the disappearance of the NMR peaks of the unphosphorylated species from the ¹H-¹⁵N TROSY spectrum of the proteins. The protein concentration, in subunits, was \sim 1 mM for all NMR experiments, unless stated otherwise.

NMR spectra were acquired on Bruker 800, 700, and 600 MHz spectrometers equipped with z-shielded gradient triple resonance cryoprobes. Spectra were processed using NMRPipe (Delaglio et al., 1995) and analyzed using NMRAM-SPARKY (Lee et al., 2015). ¹H-¹⁵N TROSY (transverse relaxation-optimized spectroscopy) (Pervushin et al., 1998) and methyl-TROSY (Tugarinov et al., 2003) experiments were acquired using previously described pulse schemes. Resonance assignments of the ¹H-¹⁵N TROSY spectra for eEIN and tEIN were transferred from previous reports (BMRB accession codes 4,106 and 27,833, respectively) (Garrett et al., 1999; Dotas and Venditti, 2019). Resonance assignments of the ¹H-¹⁵N TROSY spectra for phosphorylated eEIN (eEIN-P) was kindly provided by Drs. Clore and Suh (Suh et al., 2008). Sequential ¹H/¹⁵N/¹³C backbone assignments of phosphorylated tEIN (tEIN-P) were achieved using TROSY versions of conventional 3D triple resonance correlation experiments [HNCO, HNCA, HNCACB, HN(CO)CA, and HN(CO)CACB] (Clore and Gronenborn, 1998). Assignment of the ¹H-¹³C_{methyl} correlations of tEIN-P was performed using out and back experiments (Tugarinov et al., 2014). NMR resonance assignments for tEIN-P were deposited on the BioMagResBank (Ulrich et al., 2008) (accession code 50386).

The weighted combined ¹H/¹⁵N chemical shift perturbations ($\Delta_{H/N}$) resulting on the ¹H-¹⁵N TROSY spectra of EIN from phosphorylation of His¹⁸⁹ were calculated using the following equation (Mulder et al., 1999):

$$\Delta_{H/N} = \sqrt{(\Delta\delta_H W_H)^2 + (\Delta\delta_N W_N)^2} \quad (1)$$

where W_H (=1) and W_N (=0.154) are the weighting factors for the ¹H and ¹⁵N amide chemical shifts and $\Delta\delta_H$ and $\Delta\delta_N$ symbolize the ¹H and ¹⁵N chemical shift differences in ppm between the unphosphorylated and phosphorylated states.

¹⁵N- R_1 and $R_{1\rho}$ experiments were recorded at 40°C and a ¹H frequency of 800 MHz, utilizing heat-compensating pulse schemes with a TROSY readout (Lakomek et al., 2012). A recycle delay of 1.5 s was used for both R_1 and $R_{1\rho}$

experiments, with the spin-lock field (ω_1) for the $R_{1\rho}$ experiments set to 1 kHz. Relaxation delay durations were 0, 120, 280, 440, 640, 800, 1,040, and 1,200 ms for R_1 and 0.2, 4.2, 7.2, 15, 23.4, 32.4, 42, 52.2, and 60 ms for $R_{1\rho}$, respectively. R_1 and $R_{1\rho}$ values were determined by fitting time-dependent exponential restoration of peak intensities at increasing relaxation delays. R_2 values were extracted from the measured R_1 and $R_{1\rho}$ values. Global rotational correlation times (τ_c) were estimated from the mean R_1 and R_2 values, excluding residues displaying enhanced local dynamics on the ps-ns timescale, using the following equation (Kay et al., 1989):

$$\tau_c \approx \frac{1}{4\pi\nu_N} \sqrt{6 \frac{R_2}{R_1} - 7} \quad (2)$$

where ν_N is the ^{15}N resonance frequency in Hz, and R_1 and R_2 are the average determined values of the ^{15}N relaxation rates.

^{15}N and $^{13}\text{C}_{\text{methyl}}$ relaxation dispersion (RD) experiments were conducted at 5, 10, 15, and 20°C using a well-established protocol (Singh et al., 2021). In brief, a pulse sequence that measures the exchange contribution for the TROSY component of the ^{15}N magnetization (Loria et al., 1999) and a pulse scheme for ^{13}C single-quantum CPMG (Carr-Purcell-Meinboom-Gill) RD described by Kay and co-workers (Lundström et al., 2007) were employed. Off-resonance effects and pulse imperfections were minimized using a four-pulse phase scheme (Yip and Zuiderweg, 2004). CPMG RD experiments were performed at a ^1H frequency of 800 and 600 MHz with fixed relaxation delays (T_{relax}) of 60 and 30 ms, for ^{15}N and $^{13}\text{C}_{\text{methyl}}$ experiments, respectively. Different numbers of refocusing pulses were implemented to produce effective CPMG fields (ν_{CPMG}) varying from 50 to 1,000 Hz (Mulder et al., 2001). Experimental errors on the transverse relaxation rates (R_2) were estimated from the noise level estimated with the NMRFAM-SPARKY software. The resulting RD curves acquired at multiple temperatures and magnetic fields were globally fit to a two-site exchange model using the Carver-Richard equation (Carver and Richards, 1972), as described by Dots et al. (2020).

Backbone amide $^1\text{D}_{\text{NH}}$ residual dipolar couplings (RDCs) were measured at 40°C by taking the difference in $^1\text{J}_{\text{NH}}$ scalar couplings in isotropic and alignment media. Phage pf1 (16 mg/ml; ASLA Biotech) was the employed alignment media and the $^1\text{J}_{\text{NH}}$ couplings were measured using the RDCs by TROSY pulse scheme (Fitzkee and Bax, 2010). Xplor-NIH (Schwieters et al., 2003) was used to compute singular value decomposition (SVD) analysis of the RDC values.

Enzyme Kinetic Assays

eEI, eEI_{S191A}, tEI, and tEI_{A191S} were investigated for their ability to catalyze the transfer of the phosphoryl group from PEP to HPr. Assays were performed on a Bruker 700 MHz spectrometer at 25°C using ^1H - ^{15}N Selective Optimized Flip Angle Short Transient (SOFAST) NMR experiments (Schanda et al., 2005), using a protocol previously described (Nguyen et al., 2018). The reaction buffer was 20 mM Tris-HCl buffer (pH 7.4), 100 mM NaCl, 4 mM MgCl_2 , 1 mM ethylenediaminetetraacetic acid (EDTA), 2 mM dithiothreitol (DTT), and 90% H_2O /10% D_2O

(v/v). The reaction volume was 500 μl . All enzymatic assays were run at fixed concentrations of enzyme ($\sim 0.005 \mu\text{M}$), HPr (600 μM), and PEP (1 mM). The initial velocities (v_0) for the phosphoryl transfer reactions were determined by plotting the concentration of unphosphorylated HPr determined by the ^1H - ^{15}N cross-peak intensities, as described by Nguyen et al. (2018) vs. time (Figure 5D). All assays were performed in triplicates to estimate the experimental error.

Mass Spectrometry

A binary ACQUITY UPLC H-Class system coupled with Synapt G2-Si HDMS system (Waters, Milford, MA) and electrospray ionization (ESI) source was used to determine the intact masses of phosphorylated and non-phosphorylated tEIN. Starting samples were prepared by diluting 10 μl of the NMR samples of phosphorylated and non-phosphorylated tEIN with HPLC grade H_2O to final concentration of 2 μM . 1 μl of each sample was injected in the mass spectrometer.

UPLC separations were performed using a Restek Ultra C4 column (5 μm 50 mm \times 1 mm) with a flow rate of 0.4 ml/min. Solvents used were 0.1% formic acid in HPLC grade H_2O (solvent A) and 0.1% formic acid in acetonitrile (solvent B, mobile phase). The gradient used started with an initial condition of 5% B for 1 min, followed by a 7 min gradient of 5–100% solvent B. This was held for 4 min before dropping back to the initial 5% buffer B in 1 min and held for the remainder of the run (total 20 min).

The eluant from the UPLC was introduced to the Waters Synapt G2-Si HDMS with TOF mass analyzer using a Waters Lockspray Source (300–5,000 Da mass range). Finally, the intact mass was determined by deconvolution of mass spectra using the MassLynx 4.2 software.

RESULTS

In this contribution, we investigate the structure and conformational dynamics of native and phosphorylated EIN from two organisms: a mesophilic bacterium (*Escherichia coli*) and a thermophilic organism (*Thermoanaerobacter tengcongensis*). The two proteins are referred to as eEIN and tEIN in the unphosphorylated state, and eEIN-P and tEIN-P in the phosphorylated state throughout the manuscript, respectively. Further, we examined the effect of two single-point mutations, Ser¹⁹¹→Ala¹⁹¹ within eEIN and Ala¹⁹¹→Ser¹⁹¹ within tEIN, on protein structure and dynamics. These mutants are denoted eEIN_{S191A} and tEIN_{A191S}, respectively. Similar notations are used for HPr and the full-length EI (eHPr, tHPr, eEI, tEI, eEI_{S191A}, and tEI_{A191S}). The full-length eEI and tEI share similar sequence (overall identity = 54% and active site identity = 100%) (Supplementary Figure 1) and 3D structure (Oberholzer et al., 2005; Teplyakov et al., 2006; Navdaeva et al., 2011; Evangelidis et al., 2018), but have been shown to be optimally active at 37 and 65°C, respectively, (Navdaeva et al., 2011). The sequence identity of isolated eEIN and tEIN is 48% (Supplementary Figure 1).

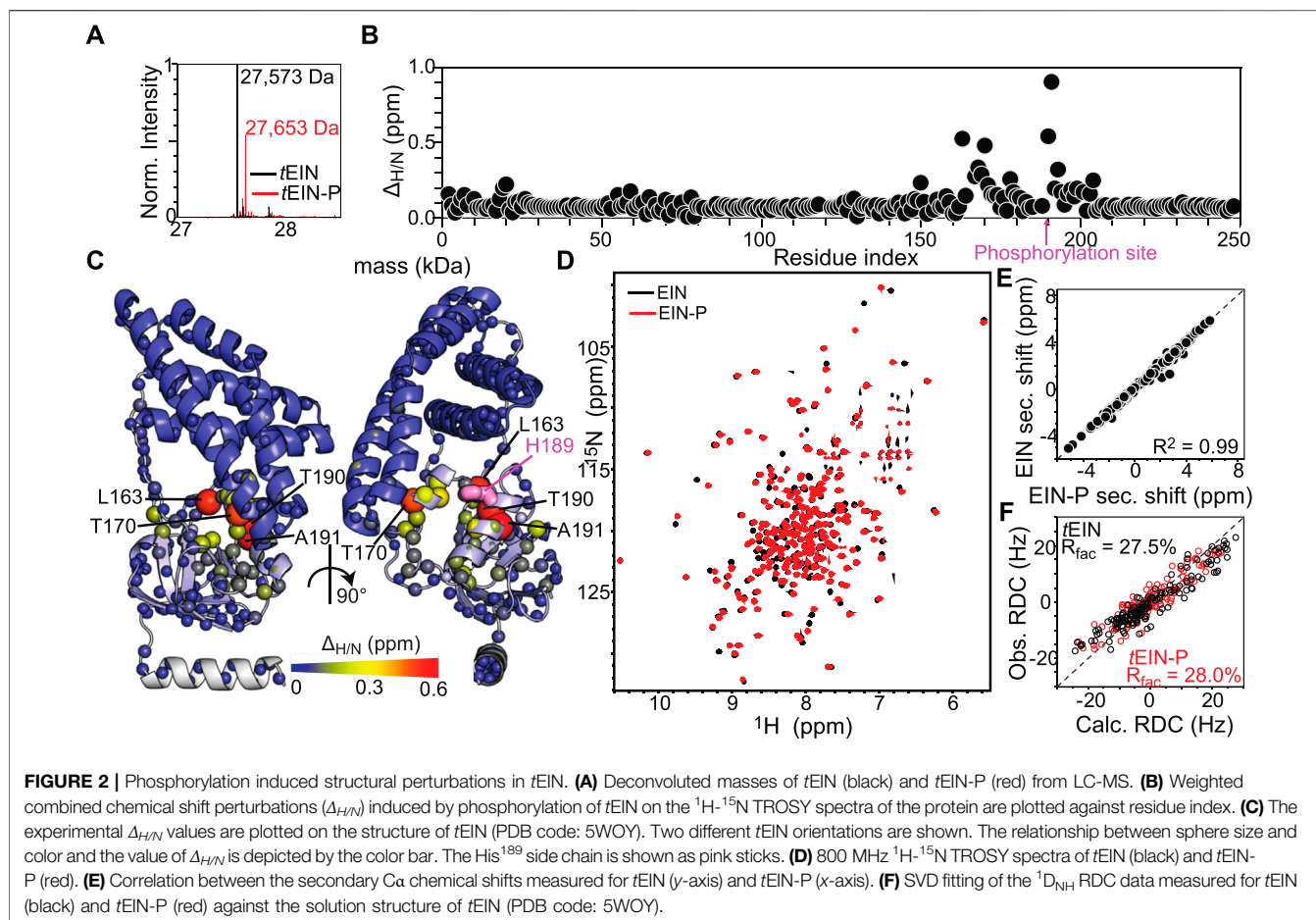


TABLE 1 | Kinetic and thermodynamic parameters for the μs -ms dynamics detected by CPMG.

	k_{ab}/k_{ba} (s^{-1}) ^a				$\Delta^*H_{ab}/\Delta^*H^{b,a}$ (kJ mol^{-1}) ^b	$\Delta^*S_{ab}/\Delta^*S^{b,a}$ ($\text{J K}^{-1} \text{mol}^{-1}$) ^b	$p^b(\%)$ ^c				ΔH (kJ mol^{-1}) ^d	ΔS ($\text{J K}^{-1} \text{mol}^{-1}$) ^d
	5°C	10°C	15°C	20°C			5°C	10°C	15°C	20°C		
tEIN	171	265	406	613	55 ± 1	-2 ± 1	8.1	8.4	8.7	9.0	6 ± 3	0 ± 39
eEIN _{S191A}	361	554	837	1,249	54 ± 1	-2 ± 1	8.9	9.3	9.6	10.0	5 ± 3	-2 ± 22
	3,686	5,425	7,879	11,297	48 ± 1	-3 ± 1						

^aThe major and minor states of the equilibrium are referred to as a and b, respectively. k_{ab} and k_{ba} are the rate constants for the transition from a to b and from b to a, respectively, and are calculated from the values of the optimized parameters k_{ex} ($= k_{ab} + k_{ba}$) and p_b . The upper and lower numbers refer to k_{ab} and k_{ba} , respectively. Errors on rate constants are < 15% of the reported values.

^bActivation enthalpies and entropies for the a to b and b to a transitions were calculated by fitting the temperature dependence of k_{ab} and k_{ba} to the Eyring equation, respectively. The upper and lower numbers refer to k_{ab} and k_{ba} , respectively.

^cErrors on populations are <5% of the reported values.

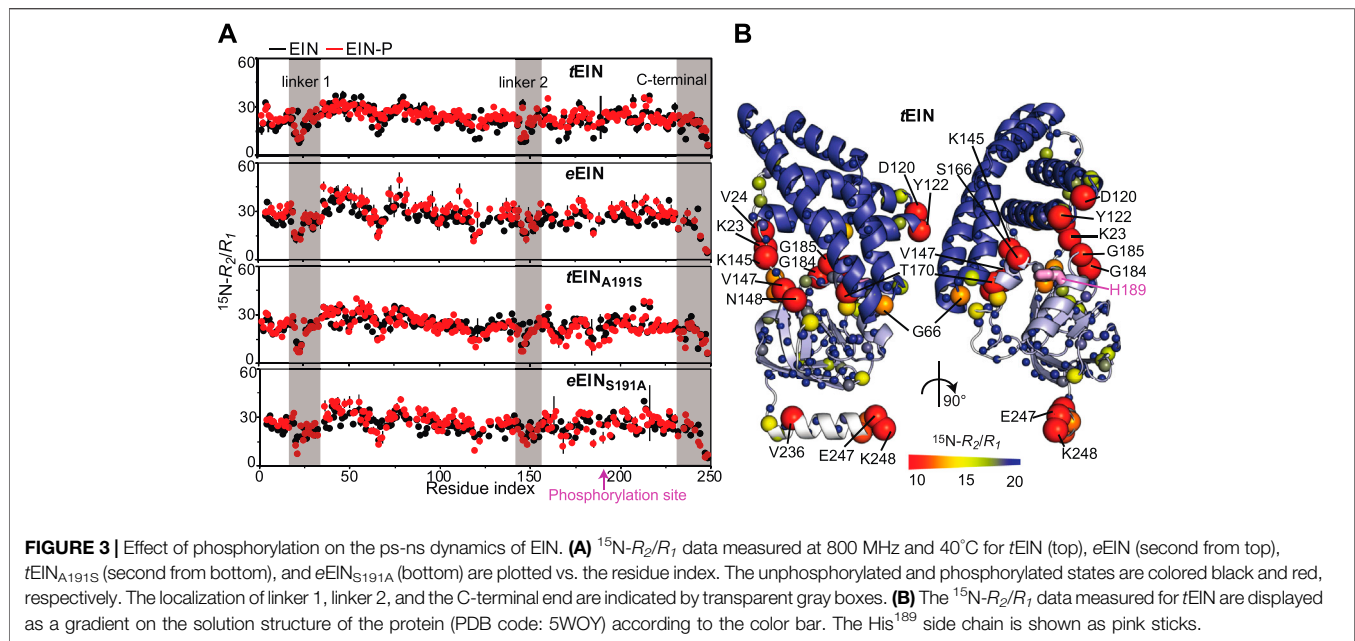
^dEnthalpy and entropy changes associated with the conformational equilibrium were calculated by using the van't Hoff equation. The equilibrium constant (K_{eq}) at each temperature was calculated using the formula $K_{\text{eq}} = p_b/(1-p_b)$.

Effect of Phosphorylation on the Structure of EIN

Previous structural investigations on eEIN have shown that phosphorylation does not affect the conformation of the eEIN backbone, but results in a transition of the χ_2 angle of His¹⁸⁹ from the g^+ to the g^- rotameric state. Such conformational change breaks the hydrogen bond between the Ne2 atom of His¹⁸⁹ and

the hydroxyl group of Thr¹⁶⁸ and grants the inbound phosphoryl group access to the Ne2 atom (Figure 1C) (Garrett et al., 1999; Suh et al., 2008).

Here, the effect of phosphorylation at the His¹⁸⁹ position on the structure of tEIN has been evaluated by NMR chemical shift perturbations and backbone amide residual dipolar coupling ($^1\text{D}_{\text{NH}}$ RDC) data. tEIN-P samples were prepared by adding



catalytic amounts ($<1\ \mu\text{M}$) of *tEI* and *tHPr*, and a large excess ($\sim 30\ \text{mM}$) of PEP directly to an NMR tube containing $\sim 1\ \text{mM}$ of ^{15}N -labeled *tEIN*. The phosphoryl transfer reaction is slow on the chemical shift time scale, and distinct NMR peaks are observed for *tEIN* and *tEIN*-P (**Figure 2D**). Therefore, completion of the phosphorylation reaction was determined by disappearance of the NMR peaks of the unphosphorylated species from the ^1H - ^{15}N TROSY spectrum of the protein. In addition, small aliquots ($\sim 10\ \mu\text{l}$) of the NMR sample were taken before and after addition of EI, and analyzed by Liquid Chromatography with tandem mass spectrometry (LC-MS-MS). The data indicate that the *tEIN* mass increases by 80 Da upon incubation with *tEI*, *tHPr*, and PEP (**Figure 2A**), which is consistent with the addition of a single phosphoryl group.

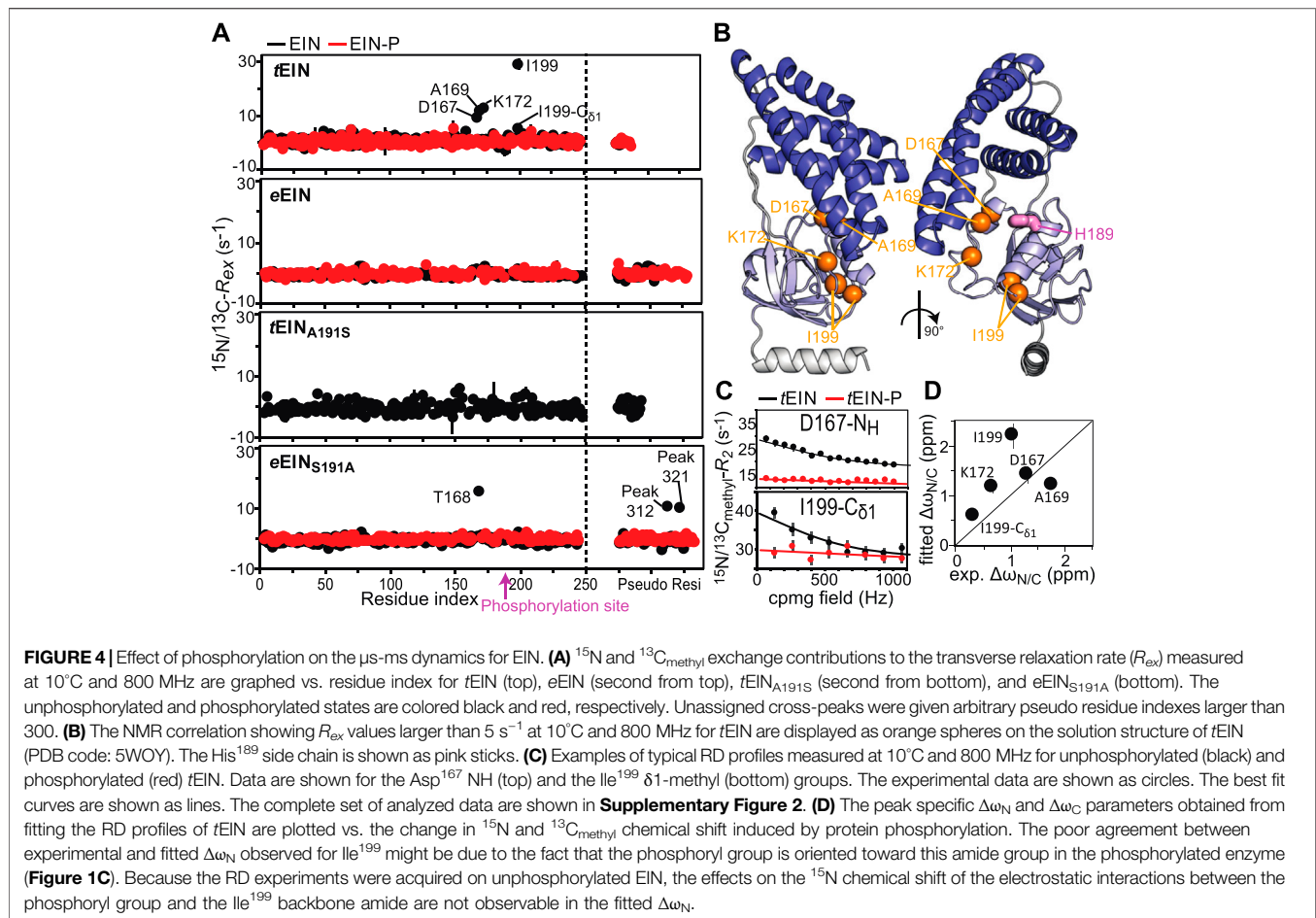
NMR resonance assignments for *tEIN* have been previously reported (BMRB code: 27833) (Dotas and Venditti, 2019). Assignments for *tEIN*-P were performed using conventional triple resonance correlation experiments (see “Materials and Methods”) and deposited in the BioMagResBank (BMRB code: 50386). $^1\text{H}/^{15}\text{N}$ chemical shift perturbations ($\Delta_{H/N}$) induced by phosphorylation in the ^1H - ^{15}N TROSY spectrum of *tEIN* are plotted against residue index in **Figure 2B** and displayed as a gradient on the protein structure in **Figure 2C**. Significant ($>0.3\ \text{ppm}$) $\Delta_{H/N}$ values are observed exclusively in the vicinity of the phosphorylation site on the α/β subdomain and are presumably a result of electronic effects that arise from the presence of the phosphoryl group as well as ring current effects from the change in the χ_2 angle of His¹⁸⁹ to accommodate the phosphoryl group at the Ne2 position. Such hypothesis is supported by the excellent agreement between the secondary Ca chemical shifts measured for phosphorylated and unphosphorylated *tEIN* (**Figure 2E**), which demonstrates that no transition in backbone conformation occurs upon phosphorylation. Further insight into the effect of

phosphorylation on the structure of EIN was obtained by the analysis of $^1\text{D}_{\text{NH}}$ RDC data measured for *tEIN* and *tEIN*-P aligned in a dilute liquid crystalline medium of phage pf1. RDCs of fixed bond vectors, such as the backbone N-H bond vector, are dependent on the orientation of the bond vectors relative to the alignment tensor and thus provide a very sensitive indicator of changes in relative domain orientations (Tjandra and Bax, 1997; Venditti et al., 2016). Singular value decomposition (SVD) fitting of the experimental RDCs measured for *tEIN* and *tEIN*-P to the coordinates of the solution structure of unphosphorylated *tEIN* (PDB code: 5WOY) (Evangelidis et al., 2018) yields R-factors (Clare and Garrett, 1999) of 27.5 and 28.0%, respectively, indicating good agreement between experimental and back-calculated data (**Figure 2F**). Consequently, one can conclude that the relative orientation of the α and α/β subdomains remains unperturbed by phosphorylation.

In summary, consistent with structural investigations performed previously on *eEIN* (Suh et al., 2008), the solution NMR data presented in this section indicate that phosphorylation of *tEIN* results in a localized transition of the rotameric state of the His¹⁸⁹ side chain and does not propagate into larger conformational rearrangements that involve the backbone of the protein.

Effect of Phosphorylation on the ps-ns Dynamics

ps-ns timescale dynamics were investigated for *eEIN*, *eEIN*-P, *tEIN*, and *tEIN*-P using NMR relaxation experiments. NMR samples of *eEIN*-P were produced as described previously (Suh et al., 2008). Residue-specific ^{15}N - R_1 and ^{15}N - R_2 values were obtained at 800 MHz and 40°C by acquisition of TROSY-detected R_1 and $R_{1\rho}$ experiments (Lakomek et al., 2012) on



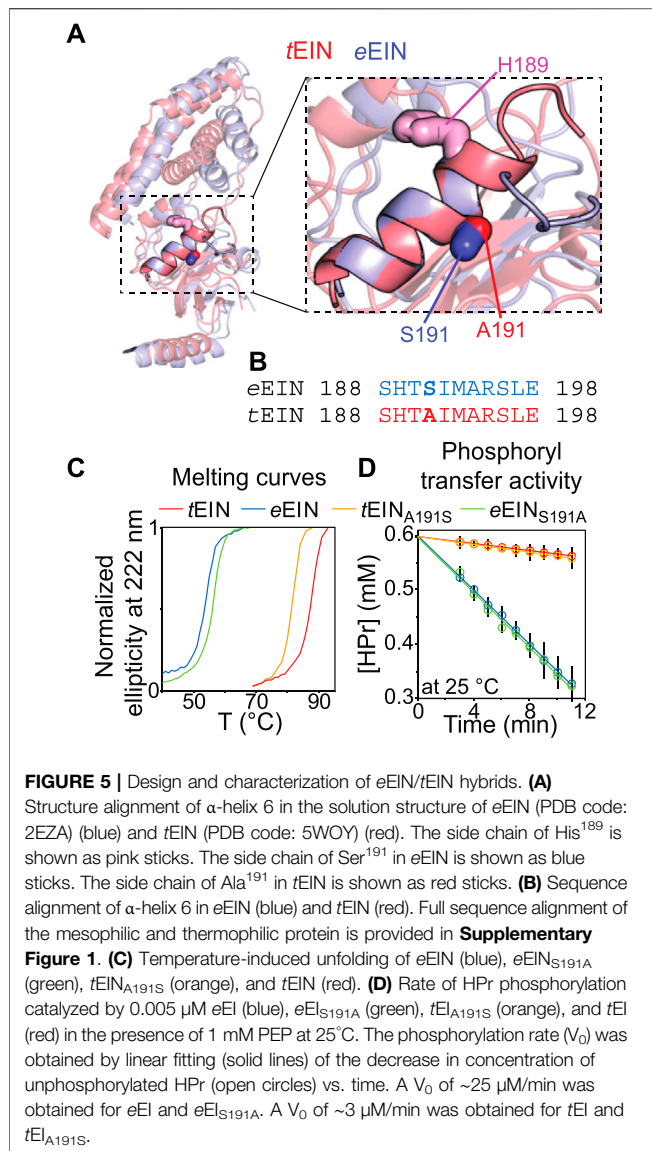
uniformly ^{15}N -labeled protein. ^{15}N - R_2/R_1 ratios are graphed as a function of residue index in **Figure 3A** and depicted as a gradient on the solution structure of *t*EIN in **Figure 3B**. For globular diamagnetic proteins, global tumbling is the only significant contribution to ^{15}N relaxation and the R_2/R_1 values are expected to be constant throughout the amino acid sequence and proportional to the rotational correlation time (τ_c) (Kay et al., 1989). Therefore, residues that produce lower than average R_2/R_1 values likely undergo additional local motion on the ps-ns timescale that decrease the effective correlation time experienced by the N-H bond. Analysis of the NMR relaxation data in **Figure 3A** indicates that *e*EIN, *e*EIN-P, *t*EIN, and *t*EIN-P tumble with a $\tau_c \sim 11\text{ ns}$, which is consistent with the theoretical τ_c calculated for a globular protein of the EIN size ($\sim 11\text{ ns}$). Notably, several regions of the protein exhibit lower than average R_2/R_1 values, suggesting the presence of local backbone dynamics on the ps-ns timescale. These regions cluster into the unstructured linkers connecting the α and α/β subdomains (linker 1: residues 19–32; linker 2: residues 143–156) and at the C-terminal end of EIN (**Figure 3B**). However, phosphorylation of His¹⁸⁹ elicits no observable change in the distribution of the R_2/R_1 values, indicating that the phosphoryl

transfer reaction does not affect the ps-ns timescale dynamics of EIN (**Figure 3A**).

Effect of Phosphorylation on the μ s-ms Dynamics

μ s-ms timescale dynamics in *e*EIN, *e*EIN-P, *t*EIN, and *t*EIN-P were investigated by ^{15}N and $^{13}\text{C}_{\text{methyl}}$ Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion (RD) spectroscopy (Mittermaier and Kay 2006). Experiments were acquired on U- $(^2\text{H}, ^{15}\text{N})$ /Ile (d_1)- $^{13}\text{CH}_3$ /Val, Leu- $(^{13}\text{CH}_3/^{12}\text{C}_2\text{H}_5)$ -labeled samples at two different static fields (600 and 800 MHz) and four different temperatures (5, 10, 15, and 20°C). Simultaneous investigation of RD profiles measured at multiple temperatures returns a comprehensive characterization of the kinetics and thermodynamics of conformational exchange processes between species with distinct chemical shifts occurring on a timescale ranging from ~ 0.1 to $\sim 10\text{ ms}$, by providing enthalpy (ΔH), entropy (ΔS), activation enthalpy ($\Delta^\ddagger H$), and activation entropy ($\Delta^\ddagger S$) for the conformational equilibrium (Purslow et al., 2018).

Exchange contributions toward the transverse relaxation rates (R_{ex}) are plotted against residue index in **Figure 4A**. Large



(>5 s⁻¹) R_{ex} values were detected for amino acids that cluster in the vicinity of the active site His¹⁸⁹ on the α/β subdomain of tEIN (**Figures 4A,B**). In particular, Asp¹⁶⁷, Ala¹⁶⁹, Lys¹⁷² localize on the partially structured helix that is in direct contact with His¹⁸⁹, while Ile¹⁹⁹ is located at the C-terminal end of the short helix that comprises the phosphorylation site. Contrarily, eEIN revealed no significant R_{ex} values (**Figure 4A**), suggesting the observed dynamics in tEIN may be too fast to be detected by CPMG within its mesophilic analogue. Interestingly, all the ¹⁵N and ¹³C_{methyl} RD curves measured at multiple temperatures and static magnetic fields for Asp¹⁶⁷, Ala¹⁶⁹, Lys¹⁷², and Ile¹⁹⁹ within tEIN could be fit simultaneously to a model describing the interconversion between two conformational states (**Figure 4C**; **Supplementary Figure 2**). In this global fitting procedure, the activation ($\Delta^\ddagger G$) and standard (ΔG) free energy of the conformational equilibrium were optimized as global parameters, whereas the ¹⁵N and ¹³C chemical shift differences

between the two conformational states ($\Delta\omega_N$ and $\Delta\omega_C$, respectively) were treated as peak-specific and temperature independent parameters. The exchange rate (k_{ex}) and the fractional population of the minor conformational state (p_b) were calculated at each temperature from the fitted values of $\Delta^\ddagger G$ and ΔG using the general form of the Eyring and reaction isotherm equations, respectively. This fitting procedure reduces the number of optimized parameters from 28 (k_{ex} , p_b , and $\Delta\omega$ for five NMR peaks at four experimental temperatures) to 7 ($\Delta^\ddagger G$, ΔG , and $\Delta\omega$ for five NMR peaks), and it is justified if the heat capacity of activation remains constant over the experimental temperature range (5–20°C) (Nguyen et al., 2017). For completeness, it should be noted that the intrinsic ¹⁵N and ¹³C_{methyl} transverse relaxation rates were also optimized as peak-specific parameters, therefore increasing the overall number of fitted parameters. Also, in order to improve convergence of the fitting algorithm, the $\Delta\omega_N$ parameters were restrained to be larger than 1 ppm. Recently we have used a similar fitting protocol to model the temperature dependence of the μ s-ms dynamics in the EIC domain of the enzyme (Dotas et al., 2020).

A summary of the optimized parameters is reported in **Table 1**. Examples of the global fit are provided in **Figure 4C**. Curves for all the analyzed RD profiles are provided in **Supplementary Figure 2**. The optimized $\Delta^\ddagger G$ is $50,228 \pm 227$ mol⁻¹, which translates into exchange rate constants (sum of forward and backward rate constants, k_{ab} and k_{ba} , respectively) of $2,117 \pm 236$, $3,162 \pm 347$, $4,660 \pm 503$, and $6,780 \pm 719$ s⁻¹ at 5, 10, 15, and 20°C, respectively. The optimized ΔG is $5,601 \pm 330$ J mol⁻¹, resulting in p_b values of 8.0 ± 0.4 , 8.4 ± 0.4 , 8.7 ± 0.4 , and $9.0 \pm 0.4\%$ at 5, 10, 15, and 20°C, respectively (**Table 1**). k_{ab} and k_{ba} were calculated from the values of k_{ex} and p_b , and their temperature dependence was modeled using the van't Hoff and Eyring equations to obtain ΔH , ΔS , $\Delta^\ddagger H$, and $\Delta^\ddagger S$ of the tEIN conformational equilibrium (**Table 1**).

From the analysis of the kinetic, thermodynamic, and NMR parameters obtained by the RD study at multiple temperatures it is apparent that 1) tEIN is in equilibrium between two conformational states, 2) the relative thermodynamic stability of the two states is dictated by enthalpic contributions to the free energy (**Table 1**), and 3) the ¹⁵N and ¹³C chemical shift differences between the two conformational states correlates with the change in ¹⁵N and ¹³C chemical shift (Δ_N and Δ_C , respectively) induced by phosphorylation of tEIN (**Figure 4D**). These findings suggest that the μ s-ms dynamics detected in tEIN by RD experiments report on the equilibrium between the g^+ and g^- rotameric states of His¹⁸⁹ that breaks the hydrogen bond between the Thr¹⁶⁸ and His¹⁸⁹ side-chains and makes the Ne2 atom accessible to the incoming phosphoryl group. Consistent with this hypothesis, the fitted value for ΔH (6 kJ mol⁻¹) is comparable with the reported energies for the weak intramolecular hydrogen bonds involving the hydroxyl group on Ser or Thr side-chains (~7 kJ mol⁻¹) (Pace et al., 2014a; Pace et al., 2014b).

To further test this model, the effect of phosphorylation on the μ s-ms dynamics of tEIN was investigated at 10°C by acquisition of RD experiments on tEIN-P. tEIN-P was

prepared enzymatically as described above. However, in this case, *tEI*, *tHPr*, and excess PEP were purified out of the final NMR sample by anion exchange chromatography. This additional purification step is required to remove any possible complex between *tEI*N and other molecules in the sample that, even at very dilute concentrations (<1% of the total *tEI*N concentration), might cause artifacts in the observed RD profiles (Singh et al., 2021). Since phosphorylated histidine is a labile post-translational modification that decays over the time, the phosphorylation state of the purified *tEI*N-P sample was checked by ^1H - ^{15}N TROSY before and after acquisition of the RD experiments at 10°C. Both NMR spectra show no sign of unphosphorylated *tEI*N cross-peaks, confirming that *tEI*N remained fully phosphorylated during NMR data acquisition.

As expected, the RD data measured for *tEI*N-P at 800 MHz shows that phosphorylation completely suppresses μs -ms dynamics within *tEI*N (Figures 4A,C). Indeed, addition of the bulky phosphoryl group at the Ne2 position hampers formation of the Thr¹⁶⁸-His¹⁸⁹ hydrogen bond and locks the side-chain of His¹⁸⁹ in the g^- rotameric state (Figure 1C). It is worth reinstating that no R_{ex} value larger than 5 s^{-1} was measured for *eEI*N and, as expected, phosphorylation of *eEI*N resulted in no observable change in the R_{ex} distribution (Figure 4A). Being that the experimental structures of phosphorylated and unphosphorylated *eEI*N indicate that the protein must undergo the His¹⁸⁹ rotameric transition to accommodate the phosphoryl group (Figure 1C), our data suggests the g^+/g^- exchange in *eEI*N is faster than in *tEI*N and, therefore, not detected by RD experiments.

Engineering *eEI*N/*tEI*N Hybrids With Modulated Conformational Dynamics

We have recently shown that hybridizing proteins from mesophilic and thermophilic bacteria is an effective strategy to produce active enzymes with modulated conformational dynamics and biological function (Dotas et al., 2020). Indeed, by merging the scaffold of EIC from *T. tengcongensis* with the active site loops of the *E. coli* enzyme we engineered a hybrid EIC variant that displays the thermal stability of the thermophilic protein and the high active site flexibility and low-temperature activity of the mesophilic enzyme. In contrast, implanting the active site loops from *T. tengcongensis* EIC onto the scaffold of *E. coli* EIC resulted in a construct that is more rigid and less active than the mesophilic enzyme (Dotas et al., 2020). Here, *eEI*N/*tEI*N hybrids are engineered to investigate the relationship between the kinetics of the His¹⁸⁹ rotameric equilibrium and turnover number. Comparison of the experimental atomic-resolution structures shows that the N-terminal end of α -helix 6 (which comprises the His¹⁸⁹) in *tEI*N is two residues longer than in *eEI*N (Figure 5A). Alignment of the amino acid sequences reveals a single Ser¹⁹¹Ala mutation within α -helix 6 moving from the mesophilic to the thermophilic construct (Figure 5B; Supplementary Figure 1). As alanine residues are known to promote helix formation in proteins (Panja et al., 2015), we hypothesize that the Ser¹⁹¹Ala mutation provides structural stabilization to α -helix six and is responsible for the slower rotameric equilibrium observed for His¹⁸⁹ in *tEI*N. To test this hypothesis, we investigated the structure, dynamics, and thermal stability of *eEI*N_{S191A} and *tEI*N_{A191S} by solution NMR and circular dichroism (CD).

Although a comparison of the ^1H - ^{15}N TROSY spectra measured for the wild type and mutant proteins shows that mutations at the 191 position provides minimal perturbations to the NMR spectra and, therefore, to the solution fold of EIN (Supplementary Figure 3), the temperature-induced unfolding data acquired by CD reveal that the introduced mutations result in sizable and opposing effects on the thermostability of *eEI*N and *tEI*N (Figure 5C). In particular, melting temperatures (T_m) of 54.0, 56.5, 82.0, and 88.0°C were determined for *eEI*N, *eEI*N_{S191A}, *tEI*N_{A191S}, and *tEI*N from the first derivative of the unfolding sigmoidal curves, respectively. This pattern of T_m values confirms that introducing an Ala residue at position 191 increases the thermal stability of *eEI*N, while the Ala¹⁹¹Ser mutation results in destabilization of *tEI*N.

Analysis of the ^{15}N - R_2/R_1 vs. residue plots reveals a similar pattern for the wild type and mutant proteins, indicating that the introduced mutations do not affect the ps-ns dynamics in native and phosphorylated EIN (Figure 3A). On the other hand, mutations of residue 191 generate observable changes in the μs -ms timescale dynamics of the protein. Indeed, no NMR peak with $R_{\text{ex}} > 5\text{ s}^{-1}$ is detected in the NMR spectra of *tEI*N_{A191S} (Figure 4A), which is consistent with the hypothesis that Ala¹⁹¹Ser mutation in *tEI*N speeds up the rotameric equilibrium of His¹⁸⁹. In contrast, the Ser¹⁹¹Ala mutation in *eEI*N introduces $R_{\text{ex}} > 5\text{ s}^{-1}$ at three ^1H - ^{15}N TROSY correlations. Although we were able to confidently assign only one of these three NMR signals, we ascribe the appearance of exchange induced effects on the spectra of *eEI*N_{S191A} to the rotameric equilibrium of His¹⁸⁹ for the following reasons: 1) The assigned NMR correlation with $R_{\text{ex}} > 5\text{ s}^{-1}$ (Thr¹⁶⁸) localizes in the same region observed to experience exchange contributions to R_2 in *tEI*N (Figure 4A), 2) Global fitting of the RD profiles acquired for the three NMR signals at multiple temperatures and static fields (Supplementary Figure 4) produces kinetic and thermodynamic parameters that are similar to the ones obtained for the His¹⁸⁹ rotameric equilibrium in *tEI*N (Table 1), and 3) Phosphorylation of *eEI*N_{S191A} results in a complete quenching of μs -ms dynamics (Figure 4A).

Overall, the NMR and CD data reported above support the hypothesis that the identity of the residue at position 191 controls helix 6 stability and the dynamics of the g^+/g^- equilibrium of His¹⁸⁹. To test the dependency of the EI biological function on the kinetics of the His¹⁸⁹ rotameric transition, the Ser¹⁹¹Ala and Ala¹⁹¹Ser mutations were incorporated into the full-length *eEI* and *tEI*, respectively. Then, the ability of *eEI*, *eEI*_{S191A}, *tEI*, and *tEI*_{A191S} to transfer the phosphoryl group from PEP to HPr was investigated at 25°C by ^1H - ^{15}N SOFAST-TROSY spectra, as recently described (Nguyen et al., 2018). As expected, the data indicate that at room temperature *eEI* catalyzes the phosphoryl transfer reaction ~10 times faster than *tEI* (Figure 5D). Interestingly, the Ser¹⁹¹Ala and Ala¹⁹¹Ser mutations produce no detectable changes in the activity of *eEI* and *tEI*, respectively (Figure 5D), indicating that the conformational transition from the g^+ to g^- rotameric state of His¹⁸⁹ is not rate limiting for catalysis.

DISCUSSION

Protein conformational transitions are fundamental to signaling, enzyme catalysis, and assembly of cellular

structures. Yet, our understanding of how the interconversion among different folded structures affects function continues to lag. One technical challenge limiting our ability to interrogate the dynamics/function relationship is the lack of universal and straightforward strategies to selectively perturb conformational equilibria in complex biomolecular systems. Here, we have shown that it is possible to perturb protein conformational dynamics without dramatically affecting their thermal stability by hybridizing the amino acid sequence of a mesophilic and a thermophilic analogue. In particular, we have investigated the structure and dynamics of the N-terminal domain of EI from a mesophilic (*e*EIN) and a thermophilic (*t*EIN) bacterium. We found that the two proteins adopt the same fold and undergo a rotameric equilibrium at the His¹⁸⁹ side chain that exposes the phosphorylation site to react with the EI substrate, PEP (**Figure 1C**). Interestingly, CPMG RD experiments revealed that the rotameric transition in *t*EIN occurs on a slower time scale than in *e*EIN (**Figure 4A**). By comparing the primary structures of the mesophilic and thermophilic proteins (**Figures 5A,B**) we identified a single point mutation in *e*EIN (Ser¹⁹¹Ala) and *t*EIN (Ala¹⁹¹Ser) that swaps the observed kinetics for this conformational change, with the *e*EIN_{S191A} mutant exchanging between the rotameric states of His¹⁸⁹ at a rate similar to the one measured for the wild type *t*EIC, and the *t*EIN_{A191S} mutant undergoing the same rotameric equilibrium on a faster time scale, comparable to wild type *e*EIN (**Figure 4, Table 1**). Intriguingly, we have recently used the same mesophilic/thermophilic hybridization strategy described here to engineer constructs of the C-terminal domain of EI (EIC) with modulated active site dynamics (Dotas et al., 2020). In contrast to the EIN case that allows for a single-point hybridizing mutation, design of the EIC hybrids required swapping of the entire active site (composed of three catalytic loops) between mesophilic and thermophilic species. Nonetheless, as for the EIN case presented here, the engineering effort resulted in production of two enzymatically active hybrids with mixed properties: One hybrid displayed the high thermal stability of the thermophilic enzyme and the increased active site flexibility and low-temperature activity of the mesophilic analogue; The second hybrid showed the low thermal stability of the mesophilic enzyme and the rigid active site and low activity at room temperature of the thermophilic protein (Dotas et al., 2020). Therefore, hybridizing homologue proteins from mesophilic and thermophilic bacteria is emerging as a powerful tool in biophysics by providing a straightforward approach to produce functional proteins with modulated internal flexibility.

In addition of serving as a demonstration of the mesophilic/thermophilic hybridization strategy for protein design, the EIN constructs engineered here allowed us to investigate the relationship between enzymatic turnover and the kinetics of the His¹⁸⁹ rotameric equilibrium. Indeed, by introducing the hybridizing mutations at position 191 into the sequence of the full-length enzyme we have demonstrated that increasing the rate of the g^+ -to- g^- transition of the His¹⁸⁹ χ^2 angle does not affect turnover for the phosphoryl transfer reaction catalyzed

by *t*EI (**Figure 4, Figure 5**). Similarly, increasing the activation energy for the His¹⁸⁹ rotameric transition in *e*EI by introducing an Ala residue at position 191 does not affect its enzymatic activity (**Figure 4, Figure 5**). These results indicate that the His¹⁸⁹ conformational change is not rate limiting for catalysis and, therefore, regulation of the EI activity cannot be achieved by slight perturbations of the His¹⁸⁹ conformational dynamics.

Finally, it is important to highlight that the data reported in this manuscript show no evidence for an active state A/state B equilibrium in the isolated EIN. This observation implies that the latter equilibrium is either on a timescale that is not compatible with RD experiments (i.e., outside the μ s-ms regime) or completely inactive in the isolated EIN domain. Considering that in the full-length dimeric EI transition to state B is required to avoid steric overlap between the EIN and EIC domains, and that state B is structurally stabilized by intersubunit EIN-EIN interactions (**Figure 1A**), we deduce that state B is inaccessible by the isolated EIN domain investigated here. In any case, our data indicate that the His¹⁸⁹ g^+ -to- g^- transition that exposes the EI phosphorylation site to PEP (**Figure 1C**) is decoupled from the EIN state A/state B equilibrium and is not triggered by transition of the full-length enzyme to the catalytically active closed conformation (**Figure 1A**).

DATA AVAILABILITY STATEMENT

The names of the repository/repositories and accession number(s) can be found below: BioMagResBank. Accession number: 50386

AUTHOR CONTRIBUTIONS

JP, JT, and VV designed the research; JP, JT, VS, TN, RD, and BK, performed the experiments; JP, JT, VS, BK, and VV analyzed the data; JP and VV wrote the article.

FUNDING

This work was supported by funds from NIGMS R35GM133488 and from the Roy J. Carver Charitable Trust to VV

ACKNOWLEDGMENTS

We thank Marius Clore and Jeong-Yong Suh for providing the resonance assignments of phosphorylated *e*EIN.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.699203/full#supplementary-material>

REFERENCES

- Carver, J. P., and Richards, R. E. (1972). A General Two-Site Solution for the Chemical Exchange Produced Dependence of T₂ upon the Carr-Purcell Pulse Separation. *J. Magn. Reson.* (1969) 6, 89–105. doi:10.1016/0022-2364(72)90090-x
- Chauvin, F., Brand, L., and Roseman, S. (1996). Enzyme I: the First Protein and Potential Regulator of the Bacterial Phosphoenolpyruvate: Glycose Phosphotransferase System. *Res. Microbiol.* 147 (6), 471–479. doi:10.1016/0923-2508(96)84001-0
- Clare, G. M., and Garrett, D. S. (1999). R-factor, FreeR, and Complete Cross-Validation for Dipolar Coupling Refinement of NMR Structures. *J. Am. Chem. Soc.* 121, 9008–9012. doi:10.1021/ja991789k
- Clare, G. M., and Gronenborn, A. M. (1998). Determining the Structures of Large Proteins and Protein Complexes by NMR. *Trends Biotechnol.* 16 (1), 22–34. doi:10.1016/s0167-7799(97)01135-9
- Clare, G. M., and Venditti, V. (2013). Structure, Dynamics and Biophysics of the Cytoplasmic Protein-Protein Complexes of the Bacterial Phosphoenolpyruvate: Sugar Phosphotransferase System. *Trends Biochemical Sciences* 38 (10), 515–530. doi:10.1016/j.tibs.2013.08.003
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* 6 (3), 277–293. doi:10.1007/bf00197809
- Deutscher, J., Aké, F. M. D., Derkaoui, M., Zébré, A. C., Cao, T. N., Bouraoui, H., et al. (2014). The Bacterial Phosphoenolpyruvate:Carbohydrate Phosphotransferase System: Regulation by Protein Phosphorylation and Phosphorylation-dependent Protein-Protein Interactions. *Microbiol. Mol. Biol. Rev.* 78 (2), 231–256. doi:10.1128/mmb.00001-14
- Dotas, R. R., Nguyen, T. T., Stewart, C. E., Ghirlando, R., Potoyan, D. A., and Venditti, V. (2020). Hybrid Thermophilic/Mesophilic Enzymes Reveal a Role for Conformational Disorder in Regulation of Bacterial Enzyme I. *J. Mol. Biol.* 432 (16), 4481–4498. doi:10.1016/j.jmb.2020.05.024
- Dotas, R. R., and Venditti, V. (2019). Resonance Assignment of the 128 kDa Enzyme I Dimer from *Thermoanaerobacter tengcongensis*. *Biomol. NMR Assign* 13 (2), 287–293. doi:10.1007/s12104-019-09893-y
- Doucette, C. D., Schwab, D. J., Wingreen, N. S., and Rabinowitz, J. D. (2011). α -Ketoglutarate Coordinates Carbon and Nitrogen Utilization via Enzyme I Inhibition. *Nat. Chem. Biol.* 7 (12), 894–901. doi:10.1038/nchembio.685
- Edelstein, P. H., Edelstein, M. A. C., Higa, F., and Falkow, S. (1999). Discovery of Virulence Genes of *Legionella pneumophila* by Using Signature Tagged Mutagenesis in a guinea Pig Pneumonia Model. *Proc. Natl. Acad. Sci.* 96 (14), 8190–8195. doi:10.1073/pnas.96.14.8190
- Evangelidis, T., Nerli, S., Nováček, J., Brereton, A. E., Karplus, P. A., Dotas, R. R., et al. (2018). Automated NMR Resonance Assignments and Structure Determination Using a Minimal Set of 4D Spectra. *Nat. Commun.* 9 (1), 384. doi:10.1038/s41467-017-02592-z
- Fitzkee, N. C., and Bax, A. (2010). Facile Measurement of 1H-15N Residual Dipolar Couplings in Large Perdeuterated Proteins. *J. Biomol. NMR* 48 (2), 65–70. doi:10.1007/s10858-010-9441-9
- Garrett, D. S., Seok, Y. J., Peterkofsky, A., Gronenborn, A. M., and Clare, G. M. (1999). Solution Structure of the 40,000 Mr Phosphoryl Transfer Complex between the N-Terminal Domain of Enzyme I and HPr. *Nat. Struct. Biol.* 6 (2), 166–173. doi:10.1038/5854
- Huang, K.-J., Lin, S.-H., Lin, M.-R., Ku, H., Szkaradek, N., Marona, H., et al. (2013). Xanthone Derivatives Could Be Potential Antibiotics: Virtual Screening for the Inhibitors of Enzyme I of Bacterial Phosphoenolpyruvate-dependent Phosphotransferase System. *J. Antibiot.* 66 (8), 453–458. doi:10.1038/ja.2013.30
- Jones, A. L., Knoll, K. M., and Rubens, C. E. (2000). Identification of *Streptococcus agalactiae* Virulence Genes in the Neonatal Rat Sepsis Model Using Signature-Tagged Mutagenesis. *Mol. Microbiol.* 37 (6), 1444–1455. doi:10.1046/j.1365-2958.2000.02099.x
- Kay, L. E., Torchia, D. A., and Bax, A. (1989). Backbone Dynamics of Proteins as Studied by Nitrogen-15 Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* 28 (23), 8972–8979. doi:10.1021/bi00449a003
- Kok, M., Bron, G., Erni, B., and Mukhija, S. (2003). Effect of Enzyme I of the Bacterial Phosphoenolpyruvate : Sugar Phosphotransferase System (PTS) on Virulence in a Murine Model. *Microbiology* 149 (9), 2645–2652. doi:10.1099/mic.0.26406-0
- Lakomek, N.-A., Ying, J., and Bax, A. (2012). Measurement of 15N Relaxation Rates in Perdeuterated Proteins by TROSY-Based Methods. *J. Biomol. NMR* 53 (3), 209–221. doi:10.1007/s10858-012-9626-5
- Lau, G. W., Haataja, S., Lonetto, M., Kensit, S. E., Marra, A., Bryant, A. P., et al. (2001). A Functional Genomic Analysis of Type 3 *Streptococcus pneumoniae* Virulence. *Mol. Microbiol.* 40 (3), 555–571. doi:10.1046/j.1365-2958.2001.02335.x
- Lee, W., Tonelli, M., and Markley, J. L. (2015). NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics* 31 (8), 1325–1327. doi:10.1093/bioinformatics/btu830
- Loria, J. P., Rance, M., and Palmer, III, A. G. (1999). A TROSY CPMG Sequence for Characterizing Chemical Exchange in Large Proteins. *J. Biomol. NMR* 15 (2), 151–155. doi:10.1023/a:1008355631073
- Lundström, P., Vallurupalli, P., Religa, T. L., Dahlquist, F. W., and Kay, L. E. (2007). A Single-Quantum Methyl 13C-Relaxation Dispersion experiment with Improved Sensitivity. *J. Biomol. NMR* 38 (1), 79–88. doi:10.1007/s10858-007-9149-7
- Mittermaier, A., and Kay, L. E. (2006). New Tools Provide New Insights in NMR Studies of Protein Dynamics. *Science* 312 (5771), 224–228. doi:10.1126/science.1124964
- Mulder, F. A. A., Schipper, D., Bott, R., and Boelens, R. (1999). Altered Flexibility in the Substrate-Binding Site of Related Native and Engineered High-Alkaline *Bacillus subtilis* 1.1 Edited by P. E. Wright. *J. Mol. Biol.* 292 (1), 111–123. doi:10.1006/jmbi.1999.3034
- Mulder, F. A. A., Skrynnikov, N. R., Hon, B., Dahlquist, F. W., and Kay, L. E. (2001). Measurement of Slow (μ s–ms) Time Scale Dynamics in Protein Side Chains by 15N Relaxation Dispersion NMR Spectroscopy: Application to Asn and Gln Residues in a Cavity Mutant of T4 Lysozyme. *J. Am. Chem. Soc.* 123 (5), 967–975. doi:10.1021/ja003447g
- Navdaeva, V., Zurbriggen, A., Waltersperger, S., Schneider, P., Oberholzer, A. E., Bähler, P., et al. (2011). Phosphoenolpyruvate: Sugar Phosphotransferase System from the Hyperthermophilic *Thermoanaerobacter tengcongensis*. *Biochemistry* 50 (7), 1184–1193. doi:10.1021/bi101721f
- Nguyen, T. T., Ghirlando, R., Roche, J., and Venditti, V. (2021). Structure Elucidation of the Elusive Enzyme I Monomer Reveals the Molecular Mechanisms Linking Oligomerization and Enzymatic Activity. *Proc. Natl. Acad. Sci.* 118, e2100298118. doi:10.1073/pnas.2100298118
- Nguyen, T. T., Ghirlando, R., and Venditti, V. (2018). The Oligomerization State of Bacterial Enzyme I (EI) Determines EI's Allosteric Stimulation or Competitive Inhibition by α -ketoglutarate. *J. Biol. Chem.* 293 (7), 2631–2639. doi:10.1074/jbc.RA117.001466
- Nguyen, T. T., and Venditti, V. (2020). An Allosteric Pocket for Inhibition of Bacterial Enzyme I Identified by NMR-Based Fragment Screening. *J. Struct. Biol.* X 4, 100034. doi:10.1016/j.jysbx.2020.100034
- Nguyen, V., Wilson, C., Hoemberger, M., Stiller, J. B., Agafonov, R. V., Kutter, S., et al. (2017). Evolutionary Drivers of Thermoadaptation in Enzyme Catalysis. *Science* 355 (6322), 289–294. doi:10.1126/science.aah3717
- Oberholzer, A. E., Bumann, M., Schneider, P., Bächler, C., Siebold, C., Baumann, U., et al. (2005). Crystal Structure of the Phosphoenolpyruvate-Binding Enzyme I-Domain from the *Thermoanaerobacter tengcongensis* PEP: Sugar Phosphotransferase System (PTS). *J. Mol. Biol.* 346 (2), 521–532. doi:10.1016/j.jmb.2004.11.077
- Pace, C. N., Scholtz, J. M., and Grimsley, G. R. (2014b). Forces Stabilizing Proteins. *FEBS Lett.* 588 (14), 2177–2184. doi:10.1016/j.febslet.2014.05.006
- Pace, C. N., Fu, H., Lee Fryar, K., Landua, J., Trevino, S. R., Schell, D., et al. (2014a). Contribution of Hydrogen Bonds to Protein Stability. *Protein Sci.* 23 (5), 652–661. doi:10.1002/pro.2449
- Panja, A. S., Bandopadhyay, B., and Maiti, s. (2015). Protein Thermostability Is Owing to Their Preferences to Non-polar Smaller Volume Amino Acids, Variations in Residual Physico-Chemical Properties and More Salt-Bridges. *PLoS One* 10 (7), e0131495. doi:10.1371/journal.pone.0131495
- Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1998). Transverse Relaxation-Optimized Spectroscopy (TROSY) for NMR Studies of Aromatic Spin Systems

- in13C-Labeled Proteins. *J. Am. Chem. Soc.* 120, 6394–6400. doi:10.1021/ja980742g
- Postma, P. W., Lengeler, J. W., and Jacobson, G. R. (1993). Phosphoenolpyruvate: carbohydrate Phosphotransferase Systems of Bacteria. *Microbiol. Rev.* 57 (3), 543–594. doi:10.1128/mmbr.57.3.543-594.1993
- Purslow, J. A., Nguyen, T. T., Egner, T. K., Dots, R. R., Khatiwada, B., and Venditti, V. (2018). Active Site Breathing of Human Alkbh5 Revealed by Solution NMR and Accelerated Molecular Dynamics. *Biophysical J.* 115 (10), 1895–1905. doi:10.1016/j.bpj.2018.10.004
- Schanda, P., Kupče, Ě., and Brutscher, B. (2005). SOFAST-HMQC Experiments for Recording Two-Dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds. *J. Biomol. NMR* 33 (4), 199–211. doi:10.1007/s10858-005-4425-x
- Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Marius Clore, G. (2003). The Xplor-NIH NMR Molecular Structure Determination Package. *J. Magn. Reson.* 160 (1), 65–73. doi:10.1016/s1090-7807(02)00014-9
- Schwieters, C. D., Suh, J.-Y., Grishaev, A., Ghirlando, R., Takayama, Y., and Clore, G. M. (2010). Solution Structure of the 128 kDa Enzyme I Dimer from *Escherichia Coli* and its 146 kDa Complex with HPr Using Residual Dipolar Couplings and Small- and Wide-Angle X-ray Scattering. *J. Am. Chem. Soc.* 132 (37), 13026–13045. doi:10.1021/ja105485b
- Singh, A., Purslow, J. A., and Venditti, V. (2021). ¹⁵N CPMG Relaxation Dispersion for the Investigation of Protein Conformational Dynamics on the μs-ms Timescale. *J. Vis. Exp.* 170, e62395. doi:10.3791/62395
- Suh, J.-Y., Cai, M., and Clore, G. M. (2008). Impact of Phosphorylation on Structure and Thermodynamics of the Interaction between the N-Terminal Domain of Enzyme I and the Histidine Phosphocarrier Protein of the Bacterial Phosphotransferase System. *J. Biol. Chem.* 283 (27), 18980–18989. doi:10.1074/jbc.m802211200
- Tepljakov, A., Lim, K., Zhu, P.-P., Kapadia, G., Chen, C. C. H., Schwartz, J., et al. (2006). Structure of Phosphorylated Enzyme I, the Phosphoenolpyruvate:sugar Phosphotransferase System Sugar Translocation Signal Protein. *Proc. Natl. Acad. Sci.* 103 (44), 16218–16223. doi:10.1073/pnas.0607587103
- Tjandra, N., and Bax, A. (1997). Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium. *Science* 278 (5340), 1111–1114. doi:10.1126/science.278.5340.1111
- Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E., and Kay, L. E. (2003). Cross-Correlated Relaxation Enhanced 1H–13C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *J. Am. Chem. Soc.* 125 (34), 10420–10428. doi:10.1021/ja030153x
- Tugarinov, V., Venditti, V., and Marius Clore, G. (2014). A NMR experiment for Simultaneous Correlations of Valine and Leucine/isoleucine Methyls with Carbonyl Chemical Shifts in Proteins. *J. Biomol. NMR* 58 (1), 1–8. doi:10.1007/s10858-013-9803-1
- Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., et al. (2008). BioMagResBank. *Nucleic Acids Res.* 36 (Database issue), D402–D408. doi:10.1093/nar/gkm957
- Venditti, V., Tugarinov, V., Schwieters, C. D., Grishaev, A., and Clore, G. M. (2015b). Large Interdomain Rearrangement Triggered by Suppression of Micro- to Millisecond Dynamics in Bacterial Enzyme I. *Nat. Commun.* 6 (1), 5960. doi:10.1038/ncomms6960
- Venditti, V., and Clore, G. M. (2012). Conformational Selection and Substrate Binding Regulate the Monomer/Dimer Equilibrium of the C-Terminal Domain of *Escherichia coli* Enzyme I. *J. Biol. Chem.* 287 (32), 26989–26998. doi:10.1074/jbc.m112.382291
- Venditti, V., Egner, T. K., and Clore, G. M. (2016). Hybrid Approaches to Structural Characterization of Conformational Ensembles of Complex Macromolecular Systems Combining NMR Residual Dipolar Couplings and Solution X-ray Scattering. *Chem. Rev.* 116 (11), 6305–6322. doi:10.1021/acs.chemrev.5b00592
- Venditti, V., Ghirlando, R., and Clore, G. M. (2013). Structural Basis for Enzyme I Inhibition by α-Ketoglutarate. *ACS Chem. Biol.* 8 (6), 1232–1240. doi:10.1021/cb400027q
- Venditti, V., Schwieters, C. D., Grishaev, A., and Clore, G. M. (2015a). Dynamic Equilibrium between Closed and Partially Closed States of the Bacterial Enzyme I Unveiled by Solution NMR and X-ray Scattering. *Proc. Natl. Acad. Sci. USA* 112 (37), 11565–11570. doi:10.1073/pnas.1515366112
- Yip, G. N. B., and Zuiderweg, E. R. P. (2004). A Phase Cycle Scheme that Significantly Suppresses Offset-dependent Artifacts in the R2-CPMG 15N Relaxation experiment. *J. Magn. Reson.* 171 (1), 25–36. doi:10.1016/j.jmr.2004.06.021

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Purslow, Thimmesch, Sivo, Nguyen, Khatiwada, Dots and Venditti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Understanding the Binding Transition State After the Conformational Selection Step: The Second Half of the Molecular Recognition Process Between NS1 of the 1918 Influenza Virus and Host p85 β

Alyssa Dubrow, Iktae Kim, Elias Topo and Jae-Hyun Cho *

Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX, United States

OPEN ACCESS

Edited by:

Vincenzo Venditti,
Iowa State University, United States

Reviewed by:

Jung Ho Lee,
Seoul National University, South Korea
Steven Van Doren,
University of Missouri, United States

*Correspondence:

Jae-Hyun Cho
jaehyuncho@tamu.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 28 May 2021

Accepted: 28 June 2021

Published: 08 July 2021

Citation:

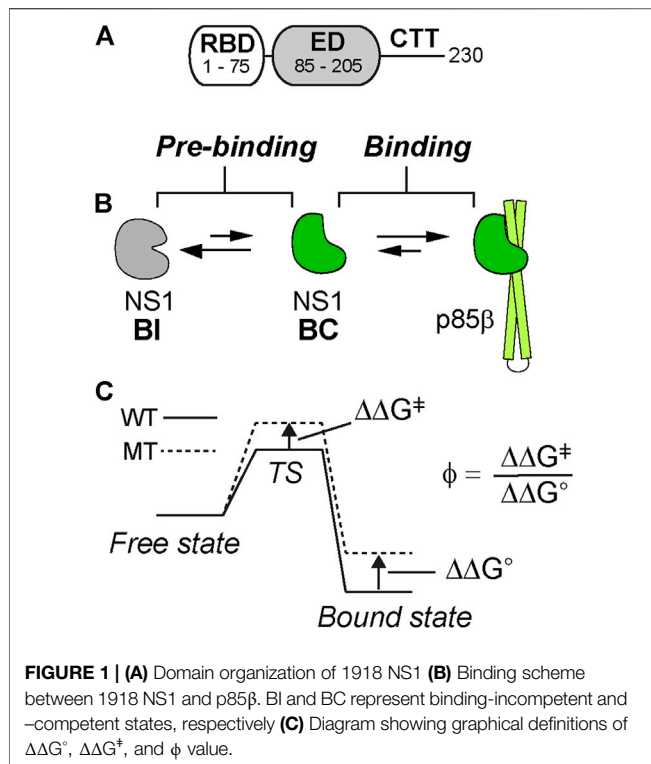
Dubrow A, Kim I, Topo E and Cho J-H
(2021) Understanding the Binding
Transition State After the
Conformational Selection Step: The
Second Half of the Molecular
Recognition Process Between NS1 of
the 1918 Influenza Virus and
Host p85 β .
Front. Mol. Biosci. 8:716477.
doi: 10.3389/fmolb.2021.716477

Biomolecular recognition often involves conformational changes as a prerequisite for binding (i.e., conformational selection) or concurrently with binding (i.e., induced-fit). Recent advances in structural and kinetic approaches have enabled the detailed characterization of protein motions at atomic resolution. However, to fully understand the role of the conformational dynamics in molecular recognition, studies on the binding transition state are needed. Here, we investigate the binding transition state between nonstructural protein 1 (NS1) of the pandemic 1918 influenza A virus and the human p85 β subunit of PI3K. 1918 NS1 binds to p85 β via conformational selection. We present the free-energy mapping of the transition and bound states of the 1918 NS1:p85 β interaction using linear free energy relationship and ϕ -value analyses. We find that the binding transition state of 1918 NS1 and p85 β is structurally similar to the bound state with well-defined binding orientation and hydrophobic interactions. Our finding provides a detailed view of how protein motion contributes to the development of intermolecular interactions along the binding reaction coordinate.

Keywords: nonstructural protein 1, molecular recognition, transition state, protein-protein interaction, conformational selection, influenza virus

INTRODUCTION

Understanding the role of conformational dynamics remains a central topic in the mechanistic study of biomolecular recognition (Boehr et al., 2009; Wand and Sharp, 2018). A particular interest is how the conformational change is coupled to the binding process. For example, a conformational adaptation of proteins may occur via conformational selection or induced-fit (Changeux and Edelstein, 2011; Vogt and Di Cera, 2012). To distinguish between the two limiting models, diverse kinetic approaches have been proposed (Hammes et al., 2009; Vogt and Di Cera, 2012; Gianni et al., 2014; Paul and Weikl, 2016). Moreover, recent advancement of structural biology techniques, especially NMR relaxation dynamics, has enabled detailed characterization of the intrinsic conformational dynamics in the pre-binding step (Sugase et al., 2007; Lange et al., 2008; Cho et al., 2010; Kovermann et al., 2017; Sekhar et al., 2018; Cho et al., 2020). Thus, combining the kinetic and structural dynamics approaches has deepened our understanding of the



functional role of intrinsic conformational dynamics in molecular recognition (Phillips et al., 2013; Chakrabarti et al., 2016; Cho et al., 2020).

Despite much progress, however, there remain outstanding questions as to the role of conformational dynamics in binding processes. For example, many protein-protein interaction (PPI) interfaces consist of heterogeneous regions in which some parts require a major conformational change while other parts only need a minor or no conformational change. Then, which region contributes to the binding transition state more significantly? In addition, what is the role of the “conformationally selected” residues in the binding step? Do they significantly contribute to stabilizing the binding transition state (TS), or is the conformational change needed only to avoid a steric clash during binding? To address these questions, we focus, in the present study, on the molecular recognition of NS1 from the 1918 influenza A virus (IAV).

The 1918 IAV caused the worst flu pandemic (a. k. a. Spanish flu) in recorded human history (Taubenberger, 2006; Taubenberger and Morens, 2006). NS1 of IAV is a major virulence factor and is responsible for suppressing host innate immune responses during the infection cycle (Min and Krug, 2006; Das et al., 2008; Hale et al., 2010a; Krug, 2015). Moreover, it was indicated that 1918 NS1 is an effective interferon antagonist (Basler et al., 2001; Geiss et al., 2002). Structurally, IAV NS1 consists of an RNA-binding domain (RBD), an effector domain (ED), followed by a structurally disordered C-terminal tail (Figure 1A) (Carrillo et al., 2014; Hale, 2014). NS1-ED is a particularly interesting domain because it binds to many host

factors involved in innate immune responses, such as phosphoinositide 3-kinase (PI3K) (Hale et al., 2010b), 30-kDa cleavage and polyadenylation specificity factor 30 (CPSF30) (Das et al., 2008), and protein kinase R (PKR) (Bergmann et al., 2000).

Binding of 1918 NS1 to PI3K results in the inhibition of apoptosis or triggering cation conductance in the infected lung epithelium (Ehrhardt et al., 2007; Gallacher et al., 2009). Interestingly, we have recently found that the free 1918 NS1 has a p85β binding-incompetent (BI) conformation as a major population and undergoes a conformational change to form a transiently populated, binding-competent (BC) form in sub-millisecond timescale (Figure 1B) (Cho et al., 2020). Therefore, the binding mechanism is best characterized as a conformational selection model. Moreover, NS1 exploits most of its surface area in order to bind multiple host proteins, suggesting that the conformational change in the p85β-binding site allosterically affects other bindings (Cho et al., 2020).

Here we characterize the binding TS between 1918 NS1-ED and p85β using linear free energy relationship (LFER) (Leffler, 1953) and ϕ -value analyses (Figure 1C) (Goldenberg et al., 1989; Fersht and Sato, 2004) combined with Ala-scanning. We find that the binding TS is structurally similar to the bound state. We identify that a hydrophobic cluster surrounding a buried hydrogen bond is critical for stabilizing both the transition and bound states. Our result also suggests that the binding process involves characteristics of both conformational selection and induced fit models.

RESULTS

Selection Strategy of Interface Residues for Mutagenesis

The 1918 NS1-ED (hereinafter NS1) residues on the p85β-binding interface are identified by the change in solvent accessible surface area of individual residues upon complexation ($\Delta\text{SASA}_{\text{bind}}$) using the structure of the complex (PDB ID: 6U28) (Figure 2A). The total interface area was 1,658.15 Å²; changes in the polar and apolar surface area were 531.35 Å² and 1,126.81 Å², respectively. Thus, the binding interface is mainly hydrophobic. The interface includes 20 residues on 1918 NS1, among which we selected ten residues for Ala-scanning mutagenesis (Figure 2B). The selected residues involve all hydrophobic residues and polar and charged residues whose $\Delta\text{SASA}_{\text{bind}}$ is larger than 20 Å² between the free and complex structures. The selected residues exhibit varying degrees of conformational changes upon binding to p85β (Figure 2C). The most significant change occurs in the residues of β1 strand. For example, χ_1 angle of Y89 in the β1 strand changes from −54.8° in the free state to 177.2° in the bound state (Figure 2D). It is also important to note that Ala-substitution of the selected residues in the present study was previously applied to NS1 of Puerto Rico 8 (PR8) IAV strain for a cell-based study on its interaction with p85β (Lopes et al., 2017). Thus, the present study on 1918 NS1 enables the comparison with the result of cell-based assays (see Discussion).

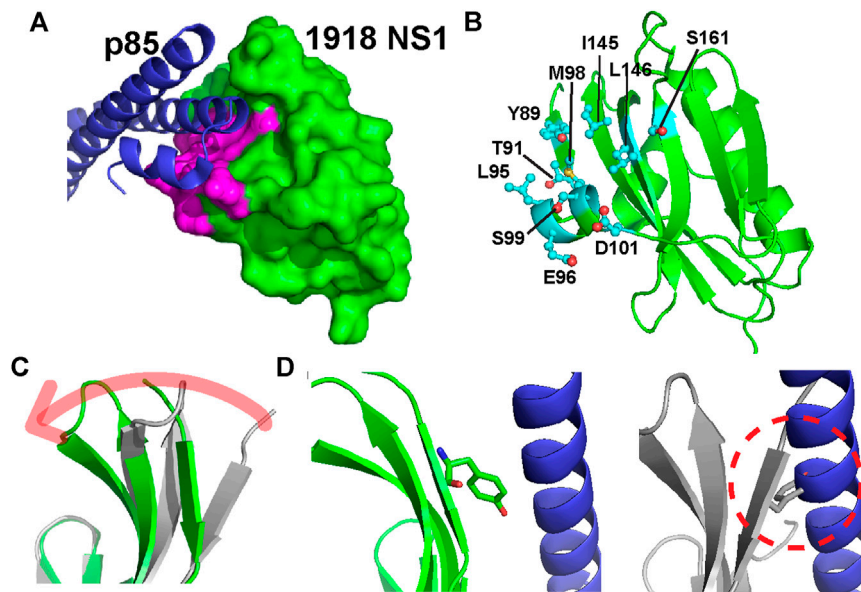


FIGURE 2 | (A) Structure of the 1918 NS1:p85 β complex (PDB ID: 6U28). Binding interface residues on 1918 NS1 included in the present study are shown in magenta **(B)** The Ala-substituted residues on 1918 NS1 are shown in a stick model. The orientation of 1918 NS1 is the same as in panel A **(C)** Overlay of 1918 NS1 structures of the free (gray) and p85 β -bound (green) states. Arrow shows the direction of conformational change upon binding to p85 β **(D)** Conformations of Y89 in the bound (left panel) and free (right panel) states of 1918 NS1. Dashed circle shows the hypothetical steric clash of Y89 with p85 β when free 1918 NS1 is docked to the complex structure.

Energetic Contribution of Interface Residues to the Binding Affinity

In the present study, we employed 1918 NS1 (residues 80–205) in which the C-terminal tail (CTT) is deleted (**Figure 1A**). The crystal structure of the 1918 NS1:p85 β complex showed that the CTT does not interact with p85 β (Cho et al., 2020). Indeed, we found that the K_D value of NS1 without CTT is highly similar to the previously reported value of NS1 with CTT (**Figures 3A,B**) (Cho et al., 2020). In addition, we incorporated W187R substitution to prevent homodimerization of 1918 NS1 (Aramini et al., 2011); thus, wild type (WT) in the present study corresponds to 1918 NS1 W187R. All mutants were prepared on W187R background. We have previously shown that W187 is exposed to solvent, located on the opposite side of the p85 β -binding site, and W187R mutation does not affect the binding to p85 β (Cho et al., 2020).

Table 1 shows the biolayer interferometry (BLI)-measured binding parameters of 1918 NS1 WT and mutants involved in this study. We noticed that some mutants bind to p85 β with a dramatically low affinity, and consequently, K_D values measured by the BLI steady-state method was not reliable. In contrast, mutants without the large change in binding affinity showed a good agreement between kinetics-derived and steady-state-derived K_D values. (**Figure 4A**). For the consistency, we used K_D values measured by the kinetic method (i.e., $K_D = k_{off}/k_{on}$). A linear relationship between k_{obs} and (p85 β) suggests that an apparent two-state binding scheme is applicable to calculate k_{on} and k_{off} values (**Figure 3C**). We also confirmed that the

BLI-derived K_D is comparable with the true equilibrium value obtained by ITC (**Supplementary Figure 1**).

Overall, the Ala-substitution of hydrophobic interface residues significantly affected the binding affinity. However, we found no statistically significant correlation between the change in ΔSASA_{bind} upon Ala-substitution of individual residues ($\Delta\Delta\text{SASA}_{bind}$) and the change in K_D ($\Delta\Delta G^\circ$) (**Figure 4B**). Rather, we found that $\Delta\Delta G^\circ$ correlates with the change in SASA of free 1918 NS1 by Ala-substitution of individual residues ($\Delta\text{SASA}_{X/A}$, where X and A correspond to any residues and Ala, respectively) (**Figure 4C**). These results indicate that hydrophobic interactions and the extent of packing at the interface might play a critical role in the binding affinity of the 1918NS1:p85 β complex.

Characterization of the Binding Transition State

We first assessed the development of global intermolecular interactions at the TS using LFER analysis (**Figure 5A**). The slope of the LFER plot is often called Leffler α -value and is an estimate of the TS position along a reaction coordinate, although α -value can differ from the true TS position in a complicated reaction (Leffler, 1953; Fersht, 2004). All data points in the LFER plot exhibited good linearity, suggesting that the binding reaction occurs through a major TS (Fersht, 2004). In contrast, when binding occurs through highly distinct, multiple TSs, the LFER plot can show a non-linear relationship (Sánchez and Kiefhaber, 2003; Bokhovchuk et al., 2020).

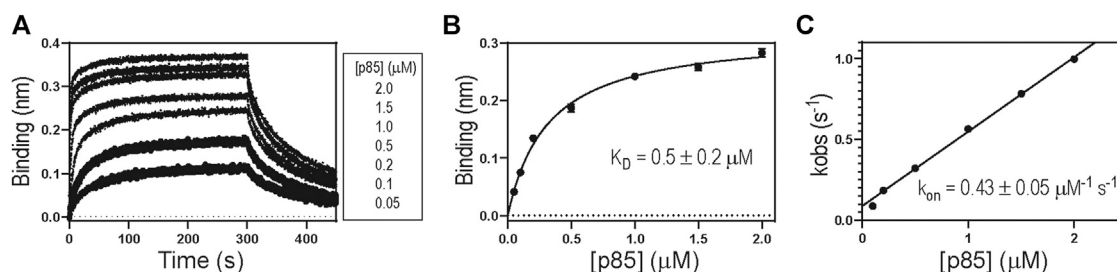


FIGURE 3 | (A) A representative BLI sensorgram showing binding between 1918 NS1 WT and p85β. Representative **(B)** binding isotherm and **(C)** binding kinetics between 1918 NS1 WT and p85β.

TABLE 1 | Binding parameters of 1918 NS1 wild type and mutants.

Proteins	ϕ -values ^a	K_D (μM) ^{a,b}	k_{on} ($\mu\text{M}^{-1} \text{s}^{-1}$) ^b	k_{off} (s^{-1}) ^b
Wild type	–	0.5 ± 0.1	0.43 ± 0.05	0.23 ± 0.02
Y89A	0.7 ± 0.2	46 ± 21	0.02 ± 0.01	1.0 ± 0.3
T91A	N. D ^c	0.7 ± 0.3	0.4 ± 0.2	0.3 ± 0.1
L95A	0.6 ± 0.1	7 ± 2	0.09 ± 0.01	0.7 ± 0.1
E96A	N. D ^c	0.6 ± 0.1	0.24 ± 0.02	0.15 ± 0.03
M98A	0.7 ± 0.1	41 ± 13	0.02 ± 0.01	0.8 ± 0.1
S99A	0.2 ± 0.2	1.3 ± 0.3	0.34 ± 0.04	0.5 ± 0.1
D101A	0.5 ± 0.5	1.7 ± 0.7	0.2 ± 0.1	0.4 ± 0.1
I145A	0.8 ± 0.1	33 ± 11	0.02 ± 0.01	0.6 ± 0.1
L146A	0.6 ± 0.4	14 ± 11	0.05 ± 0.04	0.8 ± 0.2
S161A	N. D ^c	0.4 ± 0.2	0.5 ± 0.1	0.2 ± 0.1

^aUncertainties represent propagated errors.

^bAll values are represented by average \pm standard deviation of 3 repeats. Average values were rounded to have the same number of significant figures with standard deviation.

^cNot Determined due to small $\Delta\Delta G^\circ$.

The Leffler α -value of the 1918 NS1:p85 β interaction was measured to be 0.67 ± 0.06 (Figure 5A), indicating that the binding TS is located at a relatively later stage in the reaction coordinate. This result also suggests that the overall TS structure is similar to that of the bound state. 1918 NS1 forms the BC (i.e., bound-like) conformation that is, a marginally populated, on-pathway intermediate in the binding reaction coordinate (Cho et al., 2020). Therefore, it is reasonable that the binding TS has a similar structure to the bound state.

To characterize the binding TS on a residue level, we conducted a ϕ -value analysis (Matouschek et al., 1989). While LFER reveals the position of the TS along with the reaction coordinate, ϕ -value analysis provides the extent of native (or bound)-like interaction of individual side chains at the TS relative to the native state (Fersht, 2004). Although the ϕ -value analysis was originally developed for studying the protein folding process, it can also characterize the TS of PPIs (Horn et al., 2009; Schreiber et al., 2009).

A ϕ -value close to 0 suggests the mutated residue does not form intermolecular interaction at the TS; i.e., the interaction at the TS is similar to that in the free state. On the contrary, a ϕ -value close to 1 suggests that the mutated residue develops a TS interaction similar to that in the bound state. Intermediate ϕ -values are often interpreted as partial bound-like interaction, although alternative interpretations are possible (Fersht and Sato, 2004; Cho and Raleigh, 2006; Cho et al., 2014). In this analysis, we

only included mutations of which $\Delta\Delta G^\circ$ ($= \Delta G^\circ_{WT} - \Delta G^\circ_{MT}$) is larger than $0.5 \text{ kcal mol}^{-1}$ to avoid artifactual ϕ -values (Fersht and Sato, 2004); seven out of the ten mutants included in this study showed a change in binding free energy of $>0.5 \text{ kcal mol}^{-1}$ (Table 1 and Figure 5B).

The Ala-substitution of hydrophobic residues showed significant effects on both binding affinity and kinetics, indicating their primary role in the binding process. The average ϕ -value of the hydrophobic residues was 0.66 (Table 1 and Figure 5B). Structurally, the hydrophobic residues are clustered on the binding surface of 1918 NS1 (Figure 5C). The cluster of the hydrophobic residues with high-intermediate ϕ -values (0.5–0.7) suggests cooperativity of the residues in the binding TS. Moreover, the hydrophobic clustering seems critical for the functionally critical interaction mediated by Y89 (vide infra).

Y89 is of special interest because the residue is highly conserved in NS1 proteins of almost all human IAVs (Hale et al., 2006; Cho et al., 2020). The importance of this evolutionarily conserved residue was well-demonstrated by functional studies (Hale et al., 2006; Hale et al., 2010b). The side chain of Y89 undergoes a noticeable conformational change upon binding to p85 β (Figure 2C), without which 1918 NS1 would sterically clash with p85 β (Figure 2D). Y89A resulted in large $\Delta\Delta G^\circ$ and $\Delta\Delta G^\ddagger$, yielding a ϕ -value of 0.65; this indicates that Y89 forms a substantial interaction at the binding TS. In the complex, Y89 forms a buried hydrogen bond to D575 of p85 β in the middle of the hydrophobic cluster at the binding interface (Figure 5D). We previously showed that deleting the hydroxyl group by Y89F also impaired the binding significantly (Cho et al., 2020).

The large energetic contribution of the hydrogen bond to the binding energetics might be due to the burial of the bond in the hydrophobic interface. The strength of a hydrogen bond can be significantly higher in the hydrophobic environment than on the solvent-exposed surface (Hendsch and Tidor, 1994; Albeck et al., 2000; Dai et al., 2019). Substantial $\Delta\Delta G^\ddagger$ of Y89A indicates that hydrophobic residues already surround the hydrogen bond at the binding TS similarly to the bound state. Moreover, while hydrophobic interactions are generally non-specific, the buried hydrogen bond might play a role in binding orientation. In other words, the TS structure might not be randomly oriented because of the Y89^{NS1}-D575^{p85} interaction.

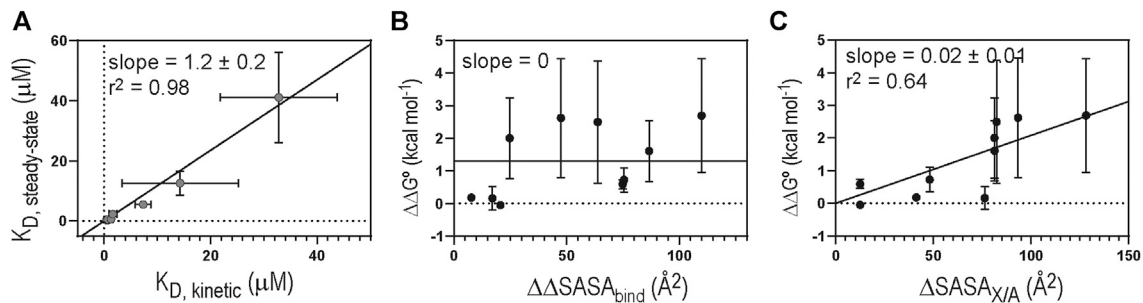


FIGURE 4 | (A) Correlation between K_D values derived by steady-state and kinetic methods. Solid diagonal line represents the best linear regression curve. Plots of **(B)** $\Delta\Delta G^\circ$ vs. $\Delta\Delta\text{SASA}_{\text{bind}}$ and **(C)** $\Delta\Delta G^\circ$ vs. $\Delta\text{SASA}_{\text{X/A}}$. Solid lines represent the linear regression curves selected by hypothesis testing between the null (slope = 0) and alternative (slope \neq 0) hypotheses. Critical α value was 0.05 for the hypothesis test (F-test).

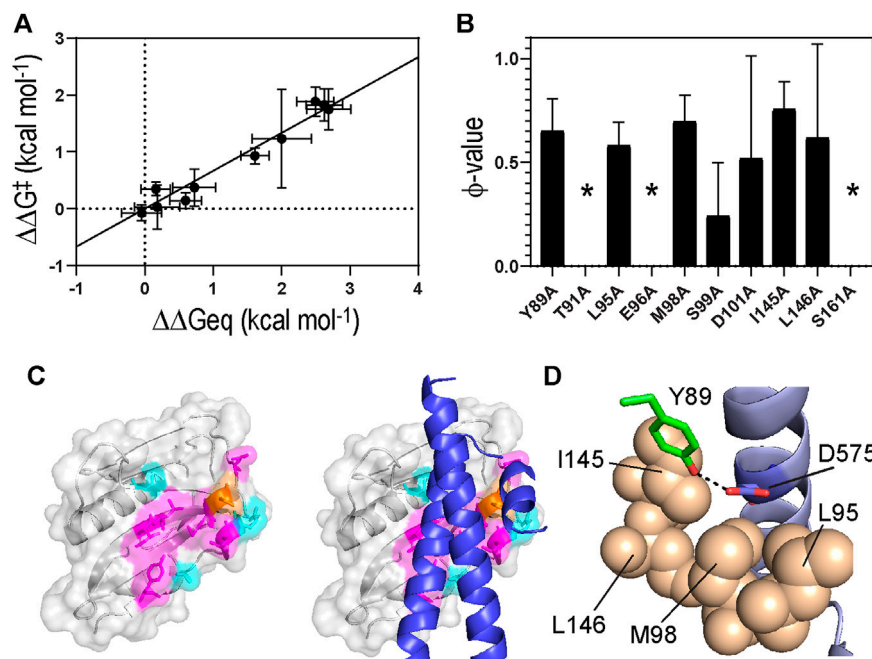


FIGURE 5 | (A) A linear free energy plot showing $\Delta\Delta G^\circ$ against $\Delta\Delta G^\circ$. **(B)** ϕ -values obtained for 1918 NS1 binding to p85 β . Asterisks correspond to the residues whose ϕ -values were not determined due to small $\Delta\Delta G^\circ$. **(C) (Left)** Color-coded ϕ -values on the surface of 1918 NS1. Residues shown in magenta and orange have high-intermediate (0.5–0.8) and low (<0.5) ϕ -values, respectively. Residues showed small $\Delta\Delta G^\circ$ (<0.5 kcal mol $^{-1}$) upon Ala substitution are shown in cyan **(Right)** Structure of p85 β is superimposed on the 1918 NS1 structure. **(D)** The hydrogen bond between Y89 of NS1 and D575 of p85 β . Hydrophobic residues surrounding the hydrogen bond are shown in sphere representation.

In contrast to hydrophobic residues, Ala-mutagenesis of polar and charged residues exhibited a marginal effect on both binding affinity and kinetics. Only two residues, S99A and D101A, changed $\Delta\Delta G^\circ > 0.5 \text{ kcal mol}^{-1}$. Consequently, ϕ -values of other polar and charged residues were not calculated due to small $\Delta\Delta G^\circ$. However, this result indicates that the energetic contribution of polar and charged residues to the binding affinity and kinetics is generally small.

However, D101 was exceptional, whose ϕ -value was measured to be 0.5. Although its uncertainty seems large, additional statistical test supports that the following structural

interpretation of the ϕ -value is plausible (Supplementary Text). The side chain of D101 interacts with the backbone amide of Q588 of p85 β in the bound state (Supplementary Figure 2). Considering the reasonably high ϕ -values of Y89 and D101, the specific binding orientation seems to be well-defined at the TS (Supplementary Figure 2). Interestingly, we previously showed that the 1918NS1:p85 β interaction is not significantly affected by a high salt concentration (Dubrow et al., 2020). Although this seems contrasting with the result of D101A, the difference is reconciled by considering that the ϕ -value is a relative parameter. Namely, the absolute free energy

change upon D101A mutation is significantly less than those by mutations of hydrophobic residues. Moreover, a salt-dependent study is not effective in screening short-range electrostatic interactions (Luisi et al., 2003).

DISCUSSION

Lopes et al. reported the result of Ala-scanning of p85 β -binding interface on PR8 NS1 using cell-based assays (Lopes et al., 2017). Because the p85 β -binding interface is completely conserved between PR8 and 1918 NS1s, it is worth comparing their result with ours. The authors found that Y89A, M98A, and I145A decreased binding affinity to p85 β by 10 to 100-fold compared to that of PR8 NS1-WT. Additionally, L95A, S99A, D101A, and L146A were identified to decrease the affinity by 2 to 3-fold relative to wild type. These results are remarkably consistent with our observation. Therefore, our result is likely to provide mechanistic details of the functional molecular recognition by 1918 NS1.

Our result indicates that hydrophobic interaction is a major driving force for stabilizing both TS and bound state between 1918 NS1 and p85 β . Similar results were reported for other PPIs. For example, Bokhovchuk et al. showed that the binding TS between YAP (Yes-associated protein) and TEAD (TEA/ATTS domain) forms a cluster of hydrophobic residues with high ϕ -values (Bokhovchuk et al., 2020). Horn et al. showed that the binding TS between hGHs (human growth hormones) and their receptor (hGHR) contains a cluster of hydrophobic residues exhibiting high ϕ -values (Horn et al., 2009). The authors also found that the TS structure adopts a native-like orientational specificity. This is analogous to ours; moderately high ϕ -values of Y89^{NS1} and D101^{NS1} indicate that the binding TS has a specific binding orientation similar to that of the bound state.

We also note the cases contrasting to our result. Extensive studies on barnase-barstar and TEM1-BLIP interactions demonstrated that hydrophobic residues have low ϕ -values (<0.1), while charged residues located outside the binding site have high ϕ -values (>0.8). Interestingly, charged residues within the binding site tend to have mixed values (0–0.6) (Schreiber et al., 2009). Remarkably, some of the experimental results were reproduced in recent MD simulations. Pan et al. conducted MD simulations of five PPIs, including barnase-barstar, TEM1-BLIP, and Ras-RAF (Pan et al., 2019). Their simulation result showed that the TSs of the PPIs interfaces are highly hydrated with a low level of native contacts (<20%), implying weak hydrophobic interactions at the TS.

Our previous study revealed a dynamic equilibrium of 1918 NS1 between BI and BC conformers in the pre-binding step (Cho et al., 2020). Thus, 1918 NS1 binds to p85 β most likely via conformational selection. However, it is also reasonable to assume that the BC conformer might need further conformational adjustment along the binding process. Our present result demonstrates that p85 β -binding interface on 1918 NS1 form a significant level of bound-like interactions at the TS. Nevertheless, it is worth noting that most of the polar

and charged residues in the interface only have weak intermolecular interactions at the TS, suggesting the need for further conformational fine-tuning after the TS. In this respect, the 1918 NS1:p85 β interaction might be viewed as a combination of conformational selection and induced fit. On a similar note, Wlodarski, and Zagrovic reported that ubiquitin undergoes conformational selection followed by induced fit in the binding process (Wlodarski and Zagrovic, 2009).

Interestingly, studies on the PPIs mediated by IDPs also demonstrated the coupling of conformational selection and induced fit mechanisms. For example, ϕ -value and LFER analyses revealed that the binding TS of c-Myb:KIX complex has a high content of bound-like structure (Giri et al., 2013). A detailed NMR study further showed that the Nand C-terminal region of c-Myb binds to KIX through conformational selection and induced fit mechanisms (Arai et al., 2015), respectively. Similar coupling of the two binding mechanisms was also observed for the binding of the YAP:TEAD complex (Bokhovchuk et al., 2020).

In conclusion, the present study provides insights into how conformational selection plays the dual role along a binding reaction, avoiding steric clash by populating binding-competent form and stabilizing the binding TS. Testing the generality of our observation in other conformational selection-based PPIs would be highly desirable in the future. In addition, we observed that the binding of 1918 NS1 and p85 β is best characterized by a coupled mechanism of conformational selection and induced fit, which is found in the binding of IDPs. Although the present study does not include a direct measurement of protein motion, the induced fit model implies protein motion or conformational adaptation after the binding transition state. Our previous study showed that 1918 NS1 undergoes conformational dynamics between BI and BC states in the pre-binding step (Cho et al., 2020). Taken together, we infer that protein motions play a key role in both pre-binding and binding steps of 1918 NS1 and p85 β .

MATERIALS AND METHODS

Protein Sample Preparation

1918 NS1 (residues 80–205) WT and mutants involved in this study were expressed in BL21 (DE3) *Escherichia coli* cells with His₆ and SUMO tags and purified by Ni²⁺ NTA column followed by gel-filtration chromatography. The p85 β (residues 435–599) was also expressed in BL21 (DE3) *E. coli* and purified in the same way as 1918 NS1 proteins.

Isothermal Titration Calorimetry

Data was acquired at 25°C using a MicroCal PEAQ-ITC (Malvern Panalytical) instrument. 100 μ M of p85 β was in the syringe and 10 μ M NS1 was in the cell. Both samples were prepared in the same buffer; 20 mM sodium phosphate (pH 7.0), 1 mM TCEP, and 100 mM NaCl. The K_D value was directly obtained from fitting the data to a 1:1 binding model.

Binding Parameters and ϕ -Values

The binding affinity between 1918 NS1 and p85 β was measured at 25°C using an Octet RED biolayer interferometer (Pall ForteBio). His₆-SUMO-tagged NS1 proteins were immobilized on the Ni-NTA biosensor. The buffer contains 20 mM sodium phosphate (pH 7.0), 150 mM NaCl, 1% bovine serum albumin, and 50 mM imidazole. Representative sensorgrams of binding between all 1918 NS1 mutants and p85 β are shown in **Supplementary Figure 3**. All measurements were repeated at least three times. Reported values are the average and standard error of the mean calculated using the repeated measurements.

In BLI binding experiments, the 1918 NS1:p85 β interaction exhibited bi-phasic association and dissociation curves due to non-specific binding. The k_{obs} was measured by fitting a fast-phase of the association curve using a single-exponential function. The k_{on} was calculated using linear regression of p85-dependent association rate [k_{obs} vs. (p85 β)]. The k_{off} was calculated from the direct measurement of the fast-phase in a dissociation curve.

ϕ -values were calculated by dividing $\Delta\Delta G^\ddagger$ by $\Delta\Delta G^\circ$ which were calculated using following equations:

$$\Delta\Delta G^\circ = \Delta G_{Kd}^{WT} - \Delta G_{Kd}^{MT} = -RT \ln(K_D^{WT}/K_D^{MT}) \quad (1)$$

$$\Delta\Delta G^\ddagger = \Delta G_{kon}^{MT} - \Delta G_{kon}^{WT} = -RT \ln(k_{on}^{MT}/k_{on}^{WT}) \quad (2)$$

DATA AVAILABILITY STATEMENT

The datasets analyzed in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found in the **Supplementary Material** and <http://www.wwpdb.org/>, 6U28.

AUTHOR CONTRIBUTIONS

JC conducted data analysis and wrote the manuscript. AD collected all BLI data. IK conducted and analyzed ITC data. AD and ET prepared protein samples.

FUNDING

Research reported in the publication was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health under grant R01GM127723, R01GM127723-02S1, and by the Welch Foundation (A-2028-20200401).

ACKNOWLEDGMENTS

We thank Nowlan Savage for this contribution to the design of 1918 NS1 mutants.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.716477/full#supplementary-material>

REFERENCES

- Albeck, S., Unger, R., and Schreiber, G. (2000). Evaluation of Direct and Cooperative Contributions towards the Strength of Buried Hydrogen Bonds and Salt Bridges. *J. Mol. Biol.* 298, 503–520. doi:10.1006/jmbi.2000.3656
- Arai, M., Sugase, K., Dyson, H. J., and Wright, P. E. (2015). Conformational Propensities of Intrinsically Disordered Proteins Influence the Mechanism of Binding and Folding. *Proc. Natl. Acad. Sci. USA* 112, 9614–9619. doi:10.1073/pnas.1512799112
- Aramini, J. M., Ma, L.-C., Zhou, L., Schauder, C. M., Hamilton, K., Amer, B. R., et al. (2011). Dimer Interface of the Effector Domain of Non-structural Protein 1 from Influenza A Virus. *J. Biol. Chem.* 286, 26050–26060. doi:10.1074/jbc.M111.248765
- Basler, C. F., Reid, A. H., Dybing, J. K., Janczewski, T. A., Fanning, T. G., Zheng, H., et al. (2001). Sequence of the 1918 Pandemic Influenza Virus Nonstructural Gene (NS) Segment and Characterization of Recombinant Viruses Bearing the 1918 NS Genes. *Proc. Natl. Acad. Sci.* 98, 2746–2751. doi:10.1073/pnas.031575198
- Bergmann, M., Garcia-Sastre, A., Carnero, E., Pehamberger, H., Wolff, K., Palese, P., et al. (2000). Influenza Virus NS1 Protein Counteracts PKR-Mediated Inhibition of Replication. *J. Virol.* 74, 6203–6206. doi:10.1128/jvi.74.13.6203-6206.2000
- Boehr, D. D., Nussinov, R., and Wright, P. E. (2009). The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* 5, 789–796. doi:10.1038/nchembio.232
- Bokhovchuk, F., Mesrouze, Y., Meyerhofer, M., Zimmermann, C., Fontana, P., Erdmann, D., et al. (2020). An Early Association between the α -Helix of the TEAD Binding Domain of YAP and TEAD Drives the Formation of the YAP:
- TEAD Complex. *Biochemistry* 59, 1804–1812. doi:10.1021/acs.biochem.0c00217
- Carrillo, B., Choi, J.-M., Bornholdt, Z. A., Sankaran, B., Rice, A. P., and Prasad, B. V. V. (2014). The Influenza A Virus Protein NS1 Displays Structural Polymorphism. *J. Virol.* 88, 4113–4122. doi:10.1128/jvi.03692-13
- Chakrabarti, K. S., Agafonov, R. V., Pontiggia, F., Otten, R., Higgins, M. K., Schertler, G. F. X., et al. (2016). Conformational Selection in a Protein-Protein Interaction Revealed by Dynamic Pathway Analysis. *Cel. Rep.* 14, 32–42. doi:10.1016/j.celrep.2015.12.010
- Changeux, J. P., and Edelstein, S. (2011). Conformational Selection or Induced Fit? 50 Years of Debate Resolved. *F1000 Biol. Rep.* 3, 19. doi:10.3410/b3-19
- Cho, J.-H., Meng, W., Sato, S., Kim, E. Y., Schindelin, H., and Raleigh, D. P. (2014). Energetically Significant Networks of Coupled Interactions within an Unfolded Protein. *Proc. Natl. Acad. Sci.* 111, 12079–12084. doi:10.1073/pnas.1402054111
- Cho, J.-H., O'Connell, N., Raleigh, D. P., and Palmer, A. G. (2010). ϕ -Value Analysis for Ultrafast Folding Proteins by NMR Relaxation Dispersion. *J. Am. Chem. Soc.* 132, 450–451. doi:10.1021/ja909052h
- Cho, J.-H., and Raleigh, D. P. (2006). Denatured State Effects and the Origin of Nonclassical ϕ Values in Protein Folding. *J. Am. Chem. Soc.* 128, 16492–16493. doi:10.1021/ja0669878
- Cho, J.-H., Zhao, B., Shi, J., Savage, N., Shen, Q., Byrnes, J., et al. (2020). Molecular Recognition of a Host Protein by NS1 of Pandemic and Seasonal Influenza A Viruses. *Proc. Natl. Acad. Sci. USA* 117, 6550–6558. doi:10.1073/pnas.1920582117
- Dai, S., Funk, L.-M., von Pappenheim, F. R., Sautner, V., Paulikat, M., Schröder, B., et al. (2019). Low-barrier Hydrogen Bonds in Enzyme Cooperativity. *Nature* 573, 609–613. doi:10.1038/s41586-019-1581-9
- Das, K., Ma, L.-C., Xiao, R., Radvansky, B., Aramini, J., Zhao, L., et al. (2008). Structural Basis for Suppression of a Host Antiviral Response by Influenza A Virus. *Proc. Natl. Acad. Sci.* 105, 13093–13098. doi:10.1073/pnas.0805213105

- Dubrow, A., Lin, S., Savage, N., Shen, Q., and Cho, J.-H. (2020). Molecular Basis of the Ternary Interaction between NS1 of the 1918 Influenza A Virus, PI3K, and CRK. *Viruses* 12, 338. doi:10.3390/v12030338
- Ehrhardt, C., Wolff, T., Pleschka, S., Planz, O., Beermann, W., Bode, J. G., et al. (2007). Influenza A Virus NS1 Protein Activates the PI3K/Akt Pathway to Mediate Antiapoptotic Signaling Responses. *J. Virol.* 81, 3058–3067. doi:10.1128/jvi.02082-06
- Fersht, A. R. (2004). Relationship of Leffler (Bronsted) Values and Protein Folding Values to Position of Transition-State Structures on Reaction Coordinates. *Proc. Natl. Acad. Sci.* 101, 14338–14342. doi:10.1073/pnas.0406091101
- Fersht, A. R., and Sato, S. (2004). -Value Analysis and the Nature of Protein-Folding Transition States. *Proc. Natl. Acad. Sci.* 101, 7976–7981. doi:10.1073/pnas.0402684101
- Gallacher, M., Brown, S. G., Hale, B. G., Fearn, R., Olver, R. E., Randall, R. E., et al. (2009). Cation Currents in Human Airway Epithelial Cells Induced by Infection with Influenza A Virus. *J. Physiol.* 587, 3159–3173. doi:10.1113/jphysiol.2009.171223
- Geiss, G. K., Salvatore, M., Tumpey, T. M., Carter, V. S., Wang, X., Basler, C. F., et al. (2002). Cellular Transcriptional Profiling in Influenza A Virus-Infected Lung Epithelial Cells: the Role of the Nonstructural NS1 Protein in the Evasion of the Host Innate Defense and its Potential Contribution to Pandemic Influenza. *Proc. Natl. Acad. Sci.* 99, 10736–10741. doi:10.1073/pnas.112338099
- Gianni, S., Dogan, J., and Jemth, P. (2014). Distinguishing Induced Fit from Conformational Selection. *Biophysical Chem.* 189, 33–39. doi:10.1016/j.bpc.2014.03.003
- Giri, R., Morrone, A., Toto, A., Brunori, M., and Gianni, S. (2013). Structure of the Transition State for the Binding of C-Myb and KIX Highlights an Unexpected Order for a Disordered System. *Proc. Natl. Acad. Sci.* 110, 14942–14947. doi:10.1073/pnas.1307337110
- Goldenberg, D. P., Frieden, R. W., Haack, J. A., and Morrison, T. B. (1989). Mutational Analysis of a Protein-Folding Pathway. *Nature* 338, 127–132. doi:10.1038/338127a0
- Hale, B. G., Albrecht, R. A., and García-Sastre, A. (2010a). Innate Immune Evasion Strategies of Influenza Viruses. *Future Microbiol.* 5, 23–41. doi:10.2217/fmb.09.108
- Hale, B. G. (2014). Conformational Plasticity of the Influenza A Virus NS1 Protein. *J. Gen. Virol.* 95, 2099–2105. doi:10.1099/vir.0.066282-0
- Hale, B. G., Jackson, D., Chen, Y.-H., Lamb, R. A., and Randall, R. E. (2006). Influenza A Virus NS1 Protein Binds P85beta and Activates Phosphatidylinositol-3-Kinase Signaling. *Proc. Natl. Acad. Sci.* 103, 14194–14199. doi:10.1073/pnas.0606109103
- Hale, B. G., Kerry, P. S., Jackson, D., Precious, B. L., Gray, A., Killip, M. J., et al. (2010b). Structural Insights into Phosphoinositide 3-kinase Activation by the Influenza A Virus NS1 Protein. *Proc. Natl. Acad. Sci. USA* 107, 1954–1959. doi:10.1073/pnas.0910715107
- Hammes, G. G., Chang, Y.-C., and Oas, T. G. (2009). Conformational Selection or Induced Fit: a Flux Description of Reaction Mechanism. *Proc. Natl. Acad. Sci.* 106, 13737–13741. doi:10.1073/pnas.0907195106
- Hendsch, Z. S., and Tidor, B. (1994). Do salt Bridges Stabilize Proteins? A Continuum Electrostatic Analysis. *Protein Sci.* 3, 211–226. doi:10.1002/pro.5560030206
- Horn, J. R., Sosnick, T. R., and Kossiakoff, A. A. (2009). Principal Determinants Leading to Transition State Formation of a Protein-Protein Complex, Orientation Trumps Side-Chain Interactions. *Pnas* 106, 2559–2564. doi:10.1073/pnas.0809800106
- Kovermann, M., Grundström, C., Sauer-Eriksson, A. E., Sauer, U. H., and Wolf-Watz, M. (2017). Structural Basis for Ligand Binding to an Enzyme by a Conformational Selection Pathway. *Proc. Natl. Acad. Sci. USA* 114, 6298–6303. doi:10.1073/pnas.1700919114
- Krug, R. M. (2015). Functions of the Influenza A Virus NS1 Protein in Antiviral Defense. *Curr. Opin. Virol.* 12, 1–6. doi:10.1016/j.coviro.2015.01.007
- Lange, O. F., Lakomek, N.-A., Farès, C., Schröder, G. F., Walter, K. F. A., Becker, S., et al. (2008). Recognition Dynamics up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* 320, 1471–1475. doi:10.1126/science.1157092
- Leffler, J. E. (1953). Parameters for the Description of Transition States. *Science* 117, 340–341. doi:10.1126/science.117.3039.340
- Lopes, A. M., Domingues, P., Zell, R., and Hale, B. G. (2017). Structure-Guided Functional Annotation of the Influenza A Virus NS1 Protein Reveals Dynamic Evolution of the P85β-Binding Site during Circulation in Humans. *J. Virol.* 91, e01081–01017. doi:10.1128/jvi.01081-17
- Luisi, D. L., Snow, C. D., Lin, J.-J., Hendsch, Z. S., Tidor, B., and Raleigh, D. P. (2003). Surface Salt Bridges, Double-Mutant Cycles, and Protein Stability: an Experimental and Computational Analysis of the Interaction of the Asp 23 Side Chain with the N-Terminus of the N-Terminal Domain of the Ribosomal Protein L9†. *Biochemistry* 42, 7050–7060. doi:10.1021/bi027202n
- Matouschek, A., Kellis, J. T., Serrano, L., and Fersht, A. R. (1989). Mapping the Transition State and Pathway of Protein Folding by Protein Engineering. *Nature* 340, 122–126. doi:10.1038/340122a0
- Min, J.-Y., and Krug, R. M. (2006). The Primary Function of RNA Binding by the Influenza A Virus NS1 Protein in Infected Cells: Inhibiting the 2'-5' Oligo (A) synthetase/RNase L Pathway. *Proc. Natl. Acad. Sci.* 103, 7100–7105. doi:10.1073/pnas.0602184103
- Pan, A. C., Jacobson, D., Yatsenko, K., Sritharan, D., Weinreich, T. M., and Shaw, D. E. (2019). Atomic-level Characterization of Protein-Protein Association. *Proc. Natl. Acad. Sci. USA* 116, 4244–4249. doi:10.1073/pnas.1815431116
- Paul, F., and Weikl, T. R. (2016). How to Distinguish Conformational Selection and Induced Fit Based on Chemical Relaxation Rates. *Plos Comput. Biol.* 12 (9), e1005067. doi:10.1371/journal.pcbi.1005067
- Phillips, A. H., Zhang, Y., Cunningham, C. N., Zhou, L., Forrest, W. F., Liu, P. S., et al. (2013). Conformational Dynamics Control Ubiquitin-Deubiquitinase Interactions and Influence *In Vivo* Signaling. *Proc. Natl. Acad. Sci.* 110, 11379–11384. doi:10.1073/pnas.1302407110
- Sánchez, I. E., and Kiefhaber, T. (2003). Non-linear Rate-Equilibrium Free Energy Relationships and Hammond Behavior in Protein Folding. *Biophys. Chem.* 100, 397–407.
- Schreiber, G., Haran, G., and Zhou, H.-X. (2009). Fundamental Aspects of Protein-Protein Association Kinetics. *Chem. Rev.* 109, 839–860. doi:10.1021/cr800373w
- Sekhar, A., Velyvis, A., Zoltsman, G., Rosenzweig, R., Bouvignies, G., and Kay, L. E. (2018). Conserved Conformational Selection Mechanism of Hsp70 Chaperone-Substrate Interactions. *eLife* 7, e32764. doi:10.7554/eLife.32764
- Sugase, K., Dyson, H. J., and Wright, P. E. (2007). Mechanism of Coupled Folding and Binding of an Intrinsically Disordered Protein. *Nature* 447, 1021–1025. doi:10.1038/nature05858
- Taubenberger, J. K. (2006). The Origin and Virulence of the 1918 “Spanish” Influenza Virus. *Proc. Am. Philos. Soc.* 150, 86–112.
- Taubenberger, J. K., and Morens, D. M. (2006). 1918 Influenza: the Mother of All Pandemics. *Emerg. Infect. Dis.* 12, 15–22. doi:10.3201/eid1209.05-0979
- Vogt, A. D., and Di Cera, E. (2012). Conformational Selection or Induced Fit? A Critical Appraisal of the Kinetic Mechanism. *Biochemistry* 51, 5894–5902. doi:10.1021/bi3006913
- Wand, A. J., and Sharp, K. A. (2018). Measuring Entropy in Molecular Recognition by Proteins. *Annu. Rev. Biophys.* 47, 41–61. doi:10.1146/annurev-biophys-060414-034042
- Wlodarski, T., and Zagrovic, B. (2009). Conformational Selection and Induced Fit Mechanism Underlie Specificity in Noncovalent Interactions with Ubiquitin. *Proc. Natl. Acad. Sci.* 106, 19346–19351. doi:10.1073/pnas.0906966106

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Dubrow, Kim, Topo and Cho. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Reactivity of Thiol-Rich Zn Sites in Diacylglycerol-Sensing PKC C1 Domain Probed by NMR Spectroscopy

Taylor R. Cole and Tatyana I. Igumenova*

Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX, United States

OPEN ACCESS

Edited by:

Vincenzo Venditti,
Iowa State University, United States

Reviewed by:

Francesca Massi,
University of Massachusetts Medical
School, United States
Justin Lorieau,
University of Illinois at Chicago,
United States

*Correspondence:

Tatyana I. Igumenova
tigumenova@tamu.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 21 June 2021

Accepted: 27 July 2021

Published: 10 August 2021

Citation:

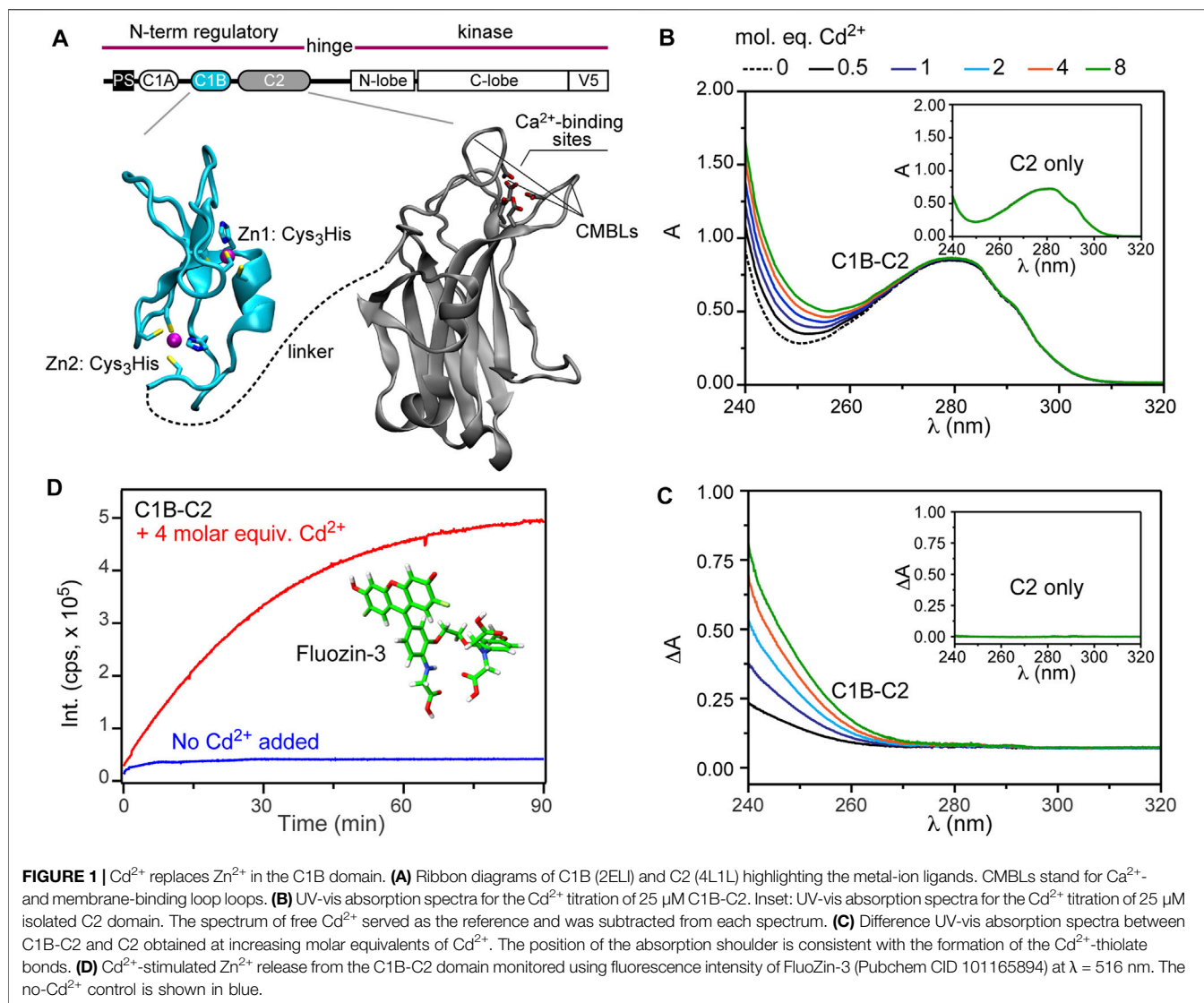
Cole TR and Igumenova TI (2021)
Reactivity of Thiol-Rich Zn Sites in
Diacylglycerol-Sensing PKC C1
Domain Probed by
NMR Spectroscopy.
Front. Mol. Biosci. 8:728711.
doi: 10.3389/fmolb.2021.728711

Conserved homology 1 (C1) domains are peripheral zinc finger domains that are responsible for recruiting their host signaling proteins, including Protein Kinase C (PKC) isoenzymes, to diacylglycerol-containing lipid membranes. In this work, we investigated the reactivity of the C1 structural zinc sites, using the cysteine-rich C1B regulatory region of the PKC α isoform as a paradigm. The choice of Cd $^{2+}$ as a probe was prompted by previous findings that xenobiotic metal ions modulate PKC activity. Using solution NMR and UV-vis spectroscopy, we found that Cd $^{2+}$ spontaneously replaced Zn $^{2+}$ in both structural sites of the C1B domain, with the formation of all-Cd and mixed Zn/Cd protein species. The Cd $^{2+}$ substitution for Zn $^{2+}$ preserved the C1B fold and function, as probed by its ability to interact with a potent tumor-promoting agent. Both Cys $_3$ His metal-ion sites of C1B have higher affinity to Cd $^{2+}$ than Zn $^{2+}$, but are thermodynamically and kinetically inequivalent with respect to the metal ion replacement, despite the identical coordination spheres. We find that even in the presence of the oxygen-rich sites presented by the neighboring peripheral membrane-binding C2 domain, the thiol-rich sites can successfully compete for the available Cd $^{2+}$. Our results indicate that Cd $^{2+}$ can target the entire membrane-binding regulatory region of PKCs, and that the competition between the thiol- and oxygen-rich sites will likely determine the activation pattern of PKCs.

Keywords: protein kinase C, C1 domain, zinc finger, cadmium, thiol-rich sites, cysteine reactivity, NMR spectroscopy, metal ion toxicity

INTRODUCTION

Approximately ~10% of the human proteome uses Zn $^{2+}$ as a cofactor (Andreini et al., 2006). While Zn $^{2+}$ is not redox active, it plays a critical role in many vital cellular processes. Functional annotation of Zn proteome predicts a wide range of biological and enzymatic activities (Bertini et al., 2010), with over 40% of the assigned sequences involved in the regulation of gene expression. One of the key signaling enzymes that require Zn $^{2+}$ is the family of Protein Kinase C isoenzymes (PKCs). By serving as the key node in the phosphoinositide signaling pathway, PKCs regulate cell growth and differentiation (Dempsey et al., 2000; Newton, 2010). Aberrant PKC activity has been implicated in many human diseases including cancer progression (Antal et al., 2015; Rahimova et al., 2020), diabetes (Koya and King, 1998; Mishra and Dey, 2021), as well as neurological (Khan et al., 2009) and cardiovascular dysfunctions (Johnson et al., 1995; Budas et al., 2007; Churchill et al., 2008; Drosatos et al., 2011). Exposure to divalent xenobiotic metal ions, such as Pb $^{2+}$ (Markovac and Goldstein, 1988; Tomsig and Suszkiw, 1995; Sun et al., 1999; Morales et al., 2011) and Cd $^{2+}$ (Beyersmann et al., 1994; Morales et al., 2013b) modulates PKC activity. Specifically, Cd $^{2+}$ can exert both activating and



inhibitory effects on PKCs (Block et al., 1992; Beyersmann et al., 1994; Long, 1997) (Saijoh et al., 1988; Speizer et al., 1989). Cadmium(II) is a known carcinogen (Waalkes and Rehm, 1992; Jarup et al., 1998; Waalkes, 2003; Faroon et al., 2012) with elevated levels in the environment due to human activity. The deleterious effects of cadmium are compounded by its relatively long half-life in the human body (Faroon et al., 2012). The molecular mechanism of how Cd^{2+} modulates PKC activity remains unresolved.

The regulatory domain of conventional (i.e., Ca^{2+} -dependent) PKC isoforms consists of three peripheral membrane binding modules: the tandem C1A and C1B domains that penetrate the membrane in response to binding a signaling lipid, diacylglycerol, and the C2 domain that binds to anionic phospholipids in a Ca^{2+} dependent manner (Figure 1A). The membrane recruitment step, mediated by both C1 and C2, removes the autoinhibition of the enzyme and enables it to phosphorylate its targets. C1 and C2 make use of two metal-ion cofactors: Zn^{2+} and Ca^{2+} ,

respectively. The Zn^{2+} ions, 2 per C1 domain, are coordinated by the Cys₃His motifs each in a tetrahedral geometry (Hubbard et al., 1991; Hommel et al., 1994; Zhang et al., 1995) and are essential for the 3D fold of C1 domains. Ca^{2+} ions are required for the membrane-binding function of C2 but not for its fold (Verdaguer et al., 1999; Morales et al., 2011). Up to three Ca^{2+} ions can bind to the all-oxygen coordination site harbored by the apical loops of C2.

In this work, we applied solution NMR spectroscopy to probe Cd^{2+} interactions with the regulatory region from PKC α , with the primary objective to evaluate the reactivity of the thiol-rich Zn^{2+} -coordinating sites towards Cd^{2+} substitution. For our experiments, we chose the two-domain unit from PKC α (denoted C1B-C2) that comprises C1B and the neighboring C2 connected by the native linker region (Figure 1A). C1B-C2 represents the minimum membrane-binding unit of PKC α capable of coincidence detection of four signaling molecules: diacylglycerol (C1B) and Ca^{2+} /phosphatidylserine/

phosphatidylinositol-4,5-bisphosphate (C2). We found that Cd^{2+} readily displaces Zn^{2+} at both structural sites in C1B, and that this process successfully competes with the Cd^{2+} interactions with the oxygen-rich C2 sites. Furthermore, despite the identical coordination spheres, the two Zn^{2+} sites show different thermodynamics and kinetics of Cd^{2+} binding. C1 and C2 domains are the basic building blocks of more than 100 proteins involved in signal transduction. Hence, the knowledge gained from this study will be applicable to other C1- and C2-containing proteins (Lemmon, 2008), leading to a more complete understanding of how xenobiotic metal ions interfere with the mechanisms of signal transduction and elicit a toxic response.

RESULTS

Cd^{2+} Coordinates Thiol Groups and Ejects Zn^{2+} From C1B

The first step was to determine how Cd^{2+} interacts with the C1B-C2 domain using UV-vis absorption spectroscopy. It is well established that thiolate- Cd^{2+} charge transfer bands have characteristic wavelengths at around ~ 240 nm (Busenlehner et al., 2001; Habjanič et al., 2020). The C1B domain has six cysteine residues, all of which are involved in coordinating the structural Zn^{2+} ions (Figure 1A). C2 is cysteine-free, but can bind Cd^{2+} with high affinity through the vacant oxygen-rich sites formed by the aspartate carboxyl groups and the carbonyl oxygens of W247 and M186 (Morales et al., 2013a). Thus, the presence of thiolate- Cd^{2+} charge transfer bands upon C1B-C2 treatment with Cd^{2+} can only originate from Cd^{2+} coordinating Cys residues of C1B.

Addition of increasing amounts of Cd^{2+} to C1B-C2 resulted in significant spectral changes (Figure 1B). Based on the C2-only control experiment with Cd^{2+} (inset of Figure 1B), these changes can only be attributed to the C1B- Cd^{2+} interactions. The difference UV-Vis spectra, where the protein contribution to the absorbance is subtracted out, clearly shows the build-up of a shoulder near $\lambda = 270$ nm (Figure 1C). The wavelength range is consistent with the position of thiolate- Cd^{2+} charge transfer bands observed in other studies (Busenlehner et al., 2001; Habjanič et al., 2020). Based on this information and previous work on the Zn^{2+} -containing proteins with Cys-rich sites (Wang et al., 2005; Chakraborty et al., 2011; Malgieri et al., 2011), we conclude that Cd^{2+} forms coordination bonds with the cysteine residues of C1B, even in the presence of Cd^{2+} -sequestering C2.

Two scenarios are possible: Cd^{2+} can either eject and substitute for Zn^{2+} , or Cd^{2+} can peripherally coordinate cysteines without displacing Zn^{2+} , forming a binuclear metal cluster similar to that observed in the GAL4 transcription factor (Pan and Coleman, 1990). To distinguish between these two scenarios, we used a highly selective Zn^{2+} fluorophore, FluoZin-3. Four molar equivalents of Cd^{2+} were added to the C1B-C2 domain in the presence of FluoZin-3, and the time-dependent fluorescence intensity was monitored at 516 nm. We observed a steady increase in the fluorescence intensity, indicating that Zn^{2+} is

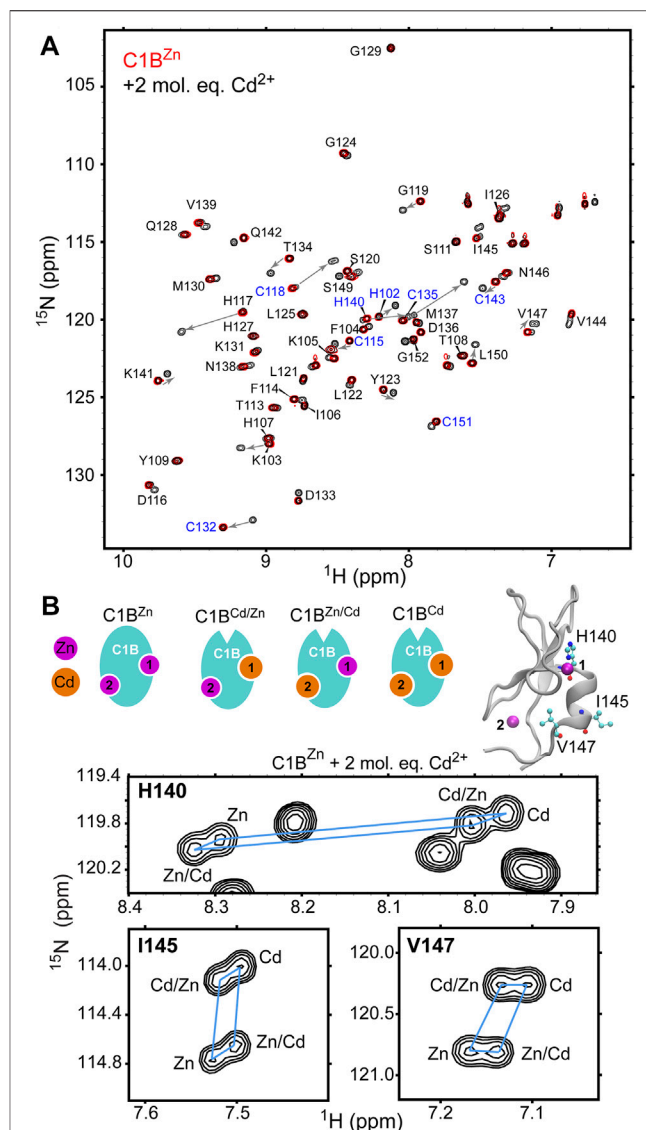


FIGURE 2 | Cd^{2+} treatment results in the formation of fully Cd-bound and Zn/Cd mixed C1B species. **(A)** ^{15}N - ^1H HSQC of 0.1 mM $[\text{U-}^{15}\text{N}]$ C1B $^{\text{Zn}}$ by itself (red) and in the presence of 2 molar-equivalents of Cd^{2+} (black). Addition of Cd^{2+} results in an appearance of a new subset of cross-peaks. Arrows indicate the residue-specific changes in chemical shifts associated with Cd^{2+} binding to C1B. Zn^{2+} -coordinating residues are highlighted in blue. **(B)** Expansions of the ^{15}N - ^1H HSQC spectra for three residues, His140, Ile145, and Val147 that show four distinct cross-peaks upon treatment of C1B $^{\text{Zn}}$ with Cd^{2+} . His140 is a Zn^{2+} -coordinating residue; Ile145 and Val147 reside on the C-terminal α helix. The four Zn/Cd C1B species are shown in cartoon representation.

being displaced from the protein as a result of Cd^{2+} treatment (Figure 1D, red trace). There was no time-dependent increase in fluorescence for an identical experiment conducted in the absence of externally added Cd^{2+} (Figure 1D, blue trace), indicating that FluoZin-3 alone cannot strip Zn^{2+} off C1B. Collectively, these experiments show that Cd^{2+} successfully ejects Zn^{2+} from C1B and forms coordination bonds with cysteines.

TABLE 1 | Relative affinities of Cd^{2+} to the C1B Cys₃His metal ion coordination sites.

Residue	Cys ₃ His, site 1		Cys ₃ His, site 2	
	$\chi[1]^a$	$\mu[1]^a$	$\chi[2]$	$\mu[2]$
H140	2.11	2.02	1.59	1.53
I145	1.74	1.81	1.51	1.57
V147	1.99	1.89	1.60	1.52
Mean ^b	1.94 ± 0.19	1.91 ± 0.11	1.57 ± 0.05	1.54 ± 0.03

^aRelative affinities are calculated for the C1B states where one Cys₃His site is already occupied by either Zn^{2+} (χ) or Cd^{2+} (μ).

^bError is reported as the standard deviation of the χ and μ values for the three residues.

Cd^{2+} Binds to Both Cys₃His Sites With the Formation of All-Cd and Cd/Zn Mixed C1B Species

While the UV-vis data show that Cd^{2+} is displacing Zn^{2+} from C1B they do not contain any site-specific information. We used solution NMR spectroscopy to gain insight into how Cd^{2+} interacts with sites 1 and 2 of C1B (see **Figure 1** for site definitions). The site-specific information was obtained by collecting 2D [¹⁵N, ¹H] HSQC spectra of [¹⁵N] enriched C1B^{Zn} in the absence and presence of Cd^{2+} . Each N-H group in C1B^{Zn} gives rise to a cross-peak in the 2D NMR spectra that we assigned in our previous work (**Figure 2A**, red spectrum) (Stewart et al., 2011). Upon addition of Cd^{2+} , we observed an appearance of a new subset of well-dispersed C1B cross-peaks (**Figure 2A**, black spectrum). We were able to assign this subset to specific Cd/Zn C1B states based on their relative peak intensities and the chemical shifts of the refolded C1B^{Cd} (*vide infra*). From the spectral overlay, it is evident that the N-H resonances of many C1B residues, particularly those coordinating Zn1 and Zn2, experience large chemical shift perturbations upon C1B binding Cd^{2+} .

In addition to native C1B^{Zn}, there are three other possible Cd/Zn protein states: C1B^{Cd}, C1B^{Zn/Cd}, and C1B^{Cd/Zn} that can co-exist in solution. The N-H groups of three residues in C1B: His140, Ile145, and Val147 show four cross-peaks each (**Figure 2B**) and serve as direct evidence for the formation of the all-Cd and Zn/Cd mixed C1B species. Moreover, the distinct chemical shifts of the four cross-peaks enable the calculation of the relative affinities of Cd^{2+} to each metal ion coordination site, using the procedures described in the Materials and Methods section. The relative affinity data presented in **Table 1** show that: (i) Cd^{2+} has a ~2-fold and ~1.6-fold higher affinities than Zn^{2+} for the C1B sites 1 and 2, respectively; and (ii) relative Cd^{2+} affinity for either site does not depend on the chemical identity of the metal ion, Cd^{2+} or Zn^{2+} , that occupies the other site (i.e. for a given site the χ and μ values are essentially identical). We conclude that both thiol-rich coordination sites in C1B are reactive with respect to Cd^{2+} substituting for the native Zn^{2+} ion.

In the PKC α regulatory region, C1B is adjacent to the C2 domain. C2 is metal-ion free in the inactive state of the kinase, but binds Ca^{2+} that is released as a result of the signaling events preceding PKC α activation. The Ca^{2+} binding site is formed by the Ca^{2+} - and membrane-binding loops or CMBLs (**Figure 1A**).

To determine the effect of the C2 domain on the C1B- Cd^{2+} interactions, we compared the [¹⁵N, ¹H] HSQC spectra of C1B-C2 in the absence and presence of 2 molar equivalents of Cd^{2+} . We observed the same signatures of Zn^{2+} replacement as in the isolated C1B domain, including the presence of four cross peaks for Ile145 and Val147 (**Figure 3A**). Overall, there is an excellent correlation between the chemical shift perturbations due to Cd^{2+} binding for isolated C1B and C1B in the context of its neighboring C2 (**Figure 3B**, inset). The full chemical shift perturbation (CSP) plot shows that not only C1B resonances are affected by interactions with Cd^{2+} , but also the CMBLs of C2 (**Figure 3B**). We previously demonstrated that the isolated C2 domain can bind Cd^{2+} with high affinity ($K_d < 1 \mu\text{M}$) through the loop regions (Morales et al., 2013a). Collectively, these data indicate that Cd^{2+} binds simultaneously to both C1B and C2 domains and that the thiol-rich C1B Cys₃His sites can effectively compete for Cd^{2+} with the C2 oxygen-rich sites.

C1B Function Is Preserved Upon Zn^{2+} Replacement With Cd^{2+}

It is evident from the chemical shift dispersion in the 2D spectra that C1B remains folded upon incorporating Cd^{2+} (**Figure 2** and **Figure 3A**). To test if C1B^{Cd} retains its function, we conducted NMR-detected binding experiments between C1B^{Cd} and a tumor-promoting agent, phorbol-12,13-dibutyrate (PDBu, **Figure 4A**). PDBu is an extremely potent exogenous agonist of PKC that binds specifically to C1 domains and drives their membrane insertion as part of the PKC activation sequence. These properties have made PDBu the most commonly used agonist (Katti and Igumenova, 2021) in the PKC field to assess the C1 domain functional competency. To generate C1B^{Cd} as the dominant species in solution, C1B^{Zn} was denatured and refolded in the presence of Cd^{2+} . The 2D [¹⁵N, ¹H] HSQC spectrum of the refolded C1B^{Cd} showed distinct chemical shifts compared to those of C1B^{Zn} (**Figure 4B**), but superimposed exactly onto the spectrum of the Cd^{2+} -bound species that were formed as a result of C1B^{Zn} treatment with Cd^{2+} (**Figure 2A**).

PDBu is an extremely hydrophobic ligand that requires a membrane-mimicking environment to form a soluble complex with C1 domains. To provide such an environment, we used the DPC/DPS mixed micelle system that supports the C1 ligand-binding function (Stewart et al., 2011; Stewart et al., 2014) and faithfully reproduces the outcomes of in-cell experiments. Upon addition of PDBu and mixed micelles to C1B^{Cd}, we observed dramatic changes in the NMR spectrum (**Figure 4B**). Several residues, such as Ser111, Gly124, Leu125, and Ile126 experienced significant chemical shift perturbations upon the formation of the ternary C1B^{Cd}-PDBu-micelle complex. The CSP plot comparing the complex with the apo state showed that the changes are localized to the C1B membrane-binding loop regions, which is responsible for capturing the ligand in the membrane environment (**Figure 4C**). This CSP pattern is essentially identical to that observed for the native C1B^{Zn} protein upon PDBu binding in micelles (Stewart et al., 2011). Because NMR chemical shifts are exquisitely sensitive to the electronic environment of the reporting nuclei, we conclude that C1B^{Cd}

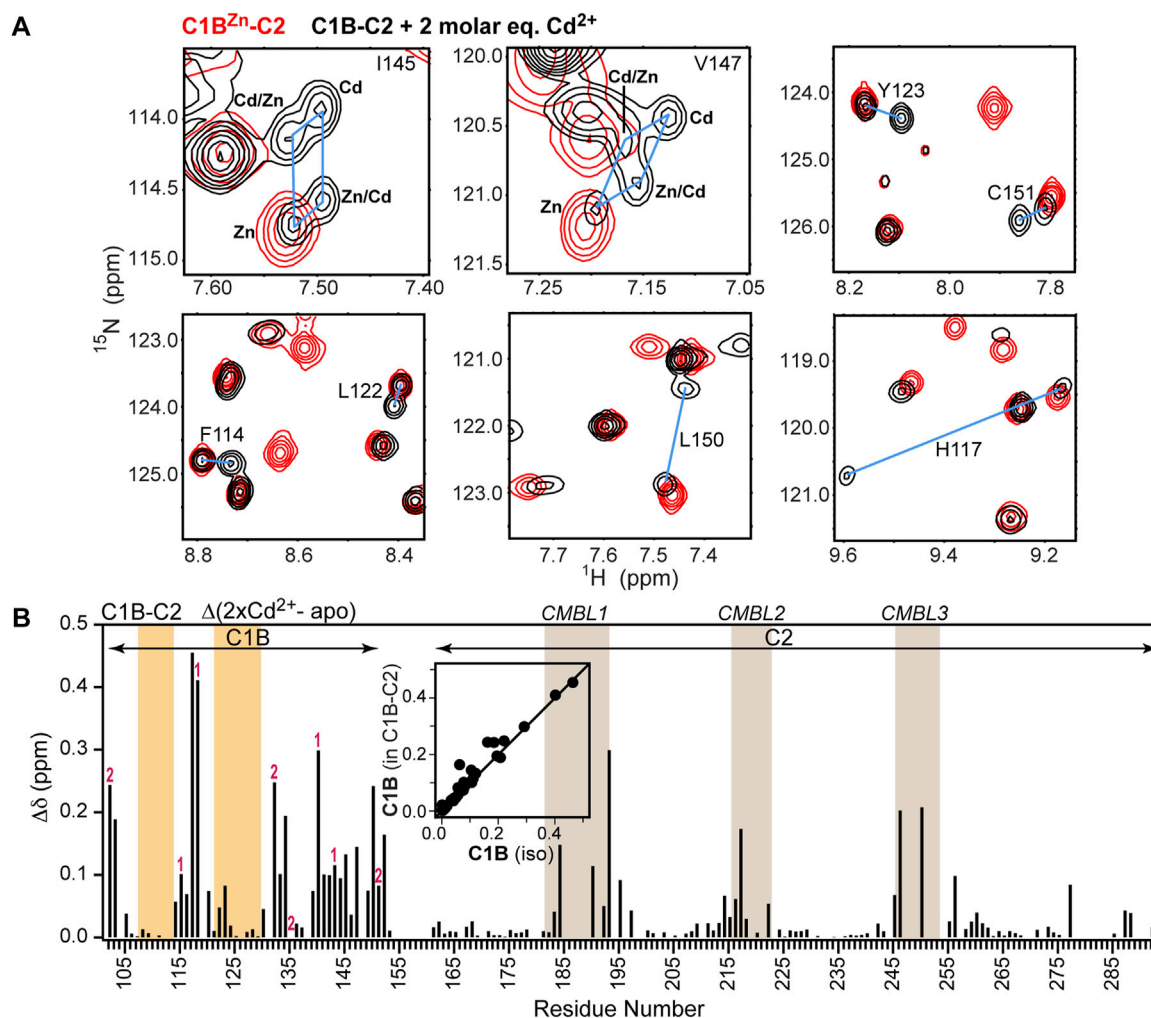


FIGURE 3 | Cd²⁺ simultaneously populates thiol- and oxygen-rich sites in C1B-C2. **(A)** Overlays of the expansions of the [¹⁵N-¹H] HSQC spectra of C1B^{Zn}-C2 (red) and Cd²⁺-treated C1B-C2 (black). The N-H cross-peaks of Cd²⁺ and Zn²⁺ containing species are connected with blue lines. **(B)** Chemical shift perturbations (CSPs) for backbone N-H groups as a function of C1B-C2 primary structure. The CSP values were calculated between the C1B^{Zn}-C2 and Cd²⁺-treated C1B-C2. Cys and His residues that coordinate Zn²⁺ in sites 1 and 2 are labeled accordingly. The C1B and C2 membrane-binding loops are highlighted in orange and tan, respectively. Inset: Correlation of C1B CSPs in the presence of 2 molar equivalents Cd²⁺ in the isolated domain and in C1B-C2.

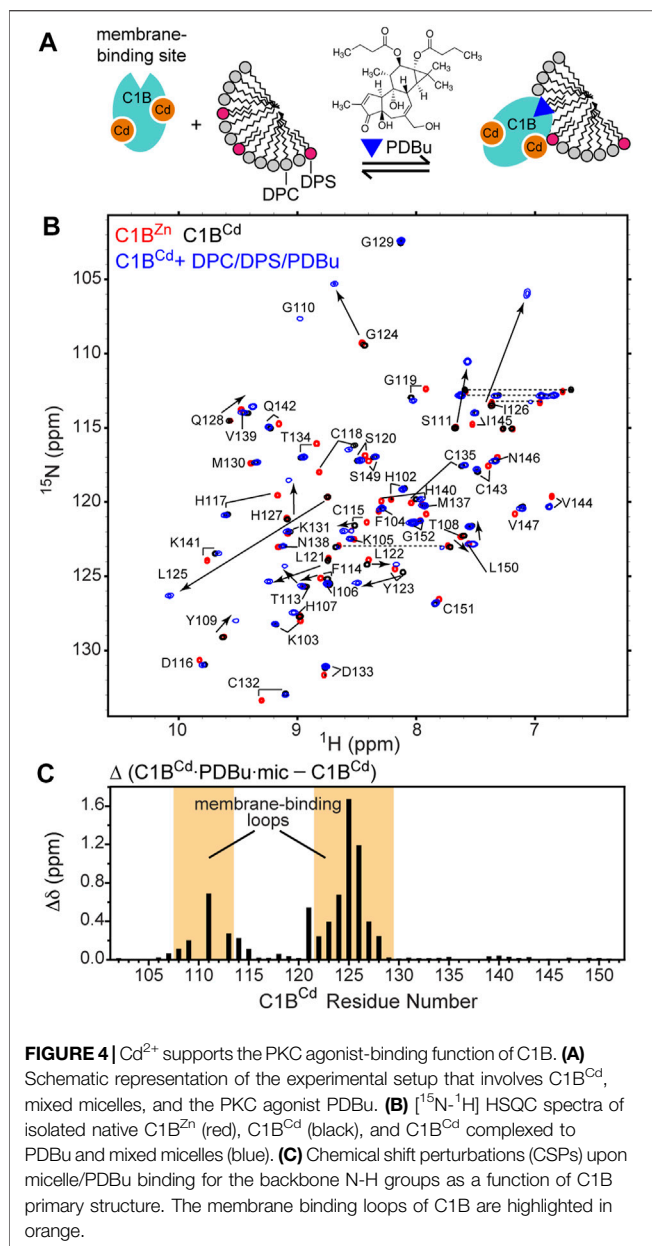
interacts with PDBu and partitions into micelles in a manner identical to that of the native C1B^{Zn}.

Kinetics of Cd²⁺ Binding Reports on the Inequivalency of the Cys₃His Structural Sites

To investigate the site-specific kinetics of Zn²⁺ replacement with Cd²⁺, we used SOFAST HMQC experiments to monitor the build-up of the Cd²⁺-bound C1B species. The population in % was calculated as the ratio of the N-H cross-peak intensities of the Cd²⁺-bound C1B, I_{Cd}, and the combined peak intensities I₀ = I_{Cd} + I_{Zn}. The data were plotted as the mean of the I_{Cd}/I₀ values for a subset of residues (listed in the Methods section) that report on Cd²⁺ binding to either site 1 or site 2. The kinetics data shown in

Figure 5A revealed that sites 1 and 2 differ with respect to their kinetic behavior.

Site 2 is more reactive towards Cd²⁺, reaching the Cd²⁺-bound population of 53% within the first 15 min of the experiment. This exceeds the equilibrium value by ~10%, and the site 1 population by 17%. As shown on the 3D structure of the C1B domain in **Figure 5B**, Zn²⁺ at site 2 brings the termini of C1B together by coordinating His102 at the N-terminus and Cys151 at the C-terminus. This part of the protein has a relatively high degree of solvent exposure and is therefore readily accessible to Cd²⁺. Another distinct feature of site 2 is the presence of a reactive Cys residue, Cys151, which serves as the entry point for the reactive oxygen species that activate PKCα in a process involving Zn²⁺ release. The structural dynamics of site 2, associated with the loss of Cys151 coordination bond with Zn²⁺ (Stewart and Igumenova, 2012), is likely to be another



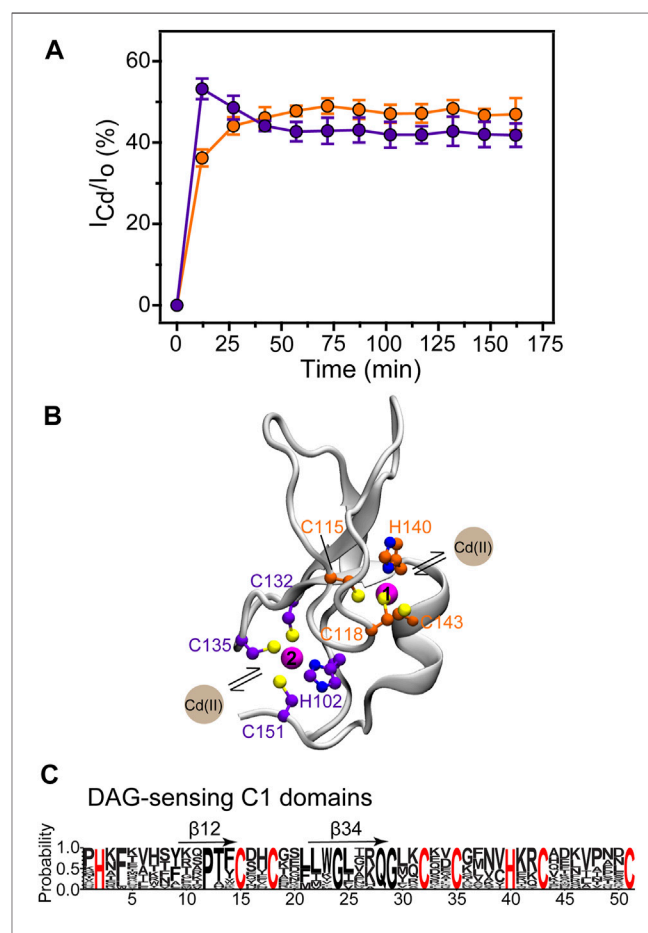
factor that makes site 2 susceptible to Cd^{2+} interactions. Under the conditions of our experiment, the system reached equilibrium within 1 hour. At equilibrium, the Cd^{2+} population of site 1 is higher than that of site 2, fully consistent with the pattern of relative Cd^{2+} affinities (Table 1). Together, the data of Figure 1D, 5(A), and Table 1 show that Cd^{2+} binding accompanied by Zn^{2+} ejection is a slow process, and that sites 1 and 2 are non-equivalent kinetically and thermodynamically.

DISCUSSION

Due to the similarities in charge and ligand preferences, xenobiotic Cd^{2+} ions target proteins that rely on Ca^{2+} and

Zn^{2+} for their function (Choong et al., 2014; Petering, 2017; Duan et al., 2018; Ben Mimouna et al., 2019). Cd^{2+} has high affinity for thiol groups (Krizek et al., 1993) and, just like Zn^{2+} , prefers tetrahedral geometry when coordinated by sulfur and nitrogen ligands. Cd^{2+} interactions with thiol groups in proteins were proposed to drive aggregation of nascent proteins through inhibition of folding *in vitro* (Sharma et al., 2008) and *in vivo* (Jacobson et al., 2017), whereas treatment with Zn^{2+} was shown to have a protective effect. Cd^{2+} can also target protein oxygen-rich sites and engage in either specific interactions in lieu of Ca^{2+} (Morales et al., 2013a; Katti et al., 2017) or opportunistic interactions that result in the formation of well-defined protein aggregates (Cole et al., 2019).

Here, we used Cd^{2+} to probe the reactivity of the structural Zn^{2+} sites in the regulatory membrane-binding region of the Ca^{2+} -activated Protein Kinase Ca. Previous work on Zn^{2+} replacement by Cd^{2+} at protein structural sites suggests that



generally this process can have varying consequences for the protein structure and function (Pan et al., 1990; Huang et al., 2004; Michalek et al., 2012). While in some cases Cd^{2+} was demonstrated to support the protein fold and function (Malgieri et al., 2014), global structural rearrangements due to Cd^{2+} replacing Zn^{2+} and loss of or change in function were also reported (Huang et al., 2004; Malgieri et al., 2011; Michalek et al., 2012). The use of Cd^{2+} in folding the C1 peptides derived from PKC α , β , and γ isoforms revealed isoform-specific differences in the functional behavior (Irie et al., 1998) and highlighted the need to investigate the Cd^{2+} response in the context of the fully folded PKC regulatory region that harbors all potential metal-ion binding sites. To that end, we used the C1B-C2 membrane-binding regulatory region to evaluate the site-specific response and reactivity of the structural thiol-rich Zn^{2+} sites towards Cd^{2+} substitution in the context of the neighboring Ca^{2+} -sensing C2 domain. Despite the presence of competing oxygen-rich C2 metal ion binding sites, Cd^{2+} was able to partially eject Zn^{2+} from C1B-C2 (Figure 1D) with the formation of the all-Cd and Zn/Cd mixed metal ion protein species (Figure 3A). The solution NMR approach was critical here, as it enabled us to follow the Cd^{2+} -binding process in the site-specific manner, starting with the fully folded domains and a native Zn^{2+} ion populating the C1B structural sites.

By specifically focusing on the isolated C1B domain, we were able to identify the spectroscopic signatures of Zn^{2+} replacement with Cd^{2+} (Figure 2) and use them to obtain thermodynamic and kinetic properties of the two Cys₃His sites. The Cd^{2+} replacement occurs spontaneously, due to the ~1.6- (site 2) and ~2-fold (site 1) higher affinity to Cd^{2+} relative to Zn^{2+} (Table 1). The relative affinities can be explained by Cd^{2+} being a softer Lewis acid (larger ionic radius and polarizability) than Zn^{2+} and therefore forming stronger interactions with thiolate ligands (Puljung and Zagotta, 2011). This property confers thermodynamic advantages onto Cd^{2+} interactions with protein sites that are thiol-rich, such as Cys₃His and Cys4 (Kluska et al., 2018). With respect to the $\text{Cd}^{2+}/\text{Zn}^{2+}$ replacement kinetics (Figure 1D and Figure 5), the reaction is slow to reach full equilibrium, likely due to the small Zn^{2+} k_{off} rate constants that are typical for the high-affinity structural sites. Despite the coordination spheres being identical, site 2 is more reactive with respect to Cd^{2+} binding. This is evidenced by the sharp increase in the respective population of Cd^{2+} -bound C1B species that get equilibrated within an hour to form all four possible metal-ion bound states (Figure 5A). We attribute the reactivity of site 2 to Cd^{2+} to its higher solvent exposure and the presence of the reactive Cys residue, Cys 151, in the coordination sphere. We previously demonstrated that in addition to being susceptible to oxidation and alkylation, Cys151 undergoes a dynamic process that slightly opens up site 2 of the C1B structure (Stewart and Igumenova, 2012). Given that this cysteine residue is proposed to be the PKC entry point of reactive oxygen species, we speculate that Cd^{2+} could have a protective effect by forming a stronger bond with the Cys151 residue. The diacylglycerol-sensitive C1 domains share significant sequence homology (Figure 5C), and the two Zn^{2+} -coordinating Cys₃His motifs are strictly conserved. This strongly suggests that our findings on the reactivity of the

Cys₃His sites in C1B from PKC α are broadly applicable to the other C1 domains. It remains to be established if the other C1s show a similar pattern of relative site reactivity, with site 2 being more reactive than site 1.

Our results for the regulatory region of PKC suggest a possible explanation of how Cd^{2+} can modulate PKC activity. Cd^{2+} spontaneously incorporates itself into the C1B structural sites without compromising the fold and PDBu-binding (Figure 4). It is therefore likely that Cd^{2+} -substituted C1 domains will retain at least part of their diacylglycerol-binding function. The membrane-binding function of Ca^{2+} -responsive C2 domains, however, is inhibited by Cd^{2+} —despite its relatively high-affinity to the oxygen-rich sites of the C2 membrane-binding loops (Morales et al., 2013a; Katti et al., 2017). Since the membrane association of both domains is necessary for PKC activation, the inhibitory effect of C2 might be predominant at high Cd^{2+} concentrations. These findings may also have implications for the mechanisms of Cd^{2+} toxicity in the cell, where the identity and occupancy of target protein sites will depend on the concentration of bioavailable Cd^{2+} .

MATERIALS AND METHODS

Buffers and Metal Ion Stock Solutions

The Cd^{2+} stock solutions were prepared by dissolving $\text{Cd}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$ (>99% purity, Sigma-Aldrich) in the appropriate buffer. Unless indicated otherwise, the experiments were conducted in the “MES buffer” comprising 10 mM 2-(N-morpholino)ethanesulfonic acid (MES) at pH 6.0 in HPLC-grade water (Avantor), 150 mM KCl, and 1 mM tris(2-carboxyethyl) phosphine (TCEP). The buffers were passed through the Chelex[®] 100 (Sigma-Aldrich) column to remove residual divalent metal ions.

Protein Expression

The DNA sequences encoding C1B-C2 (residues 100–293), isolated C1B (residues 100–152) or C2 (residues 155–293) of PKC α (*M. musculus* for C1B-C2 and C1B; *R. Norvegicus* for C2) were amplified by PCR using the cDNA clone of PKC α (Open Biosystems) as a template and cloned into the pET-SUMO vector (Invitrogen). Isolated C1B, C2, and C1B-C2 were expressed and purified as described previously (Morales et al., 2011; Stewart et al., 2011; Cole et al., 2019). [^{15}N , 75%- ^2H]-enriched C1B-C2 and [^{15}N]- or [^{13}C , ^{15}N]-enriched C1B were used for the NMR experiments.

UV-Vis Spectroscopy

UV-vis spectra were collected on a Beckman DU 640 spectrophotometer. 25 μM protein (C1B-C2, C2, or C1B) solution or MES buffer (for metal ion-only reference experiments) were placed in the sample cuvette; the reference cuvette always contained metal ion-free MES buffer. Cd^{2+} was added stepwise from the corresponding stock solutions to the sample cuvette. The samples were incubated for 1 h prior to the start of the measurements. The post-acquisition processing included the subtraction of the free Cd^{2+} spectra from the

spectra of protein-containing samples. To eliminate contribution of protein-only absorption bands, the difference spectra were generated by subtracting the spectrum of the apo protein from the spectra of the metal-ion-containing protein. All spectra were corrected for dilution prior to subtraction.

C1B Refolding

[U-¹⁵N]-enriched C1B was dissolved in 6 M guanidine hydrochloride (Acros Organics) and the “refolding buffer” comprising 20 mM MES at pH 6.0 and 1 mM TCEP. The final protein concentration was between 15 and 35 μM during the denaturation step. The refolding was conducted in three dialysis steps, all of them carried out in the refolding buffer: (1) against 8 M urea at room temperature, for 8 h; (2) against 1.5 M urea and 100 μM Cd(II) nitrate at 4°C, overnight; and (3) against urea-free buffer at 4°C, for 3 days to ensure complete removal of urea. The refolded protein was concentrated in a Vivaspin® spin concentrator with a 3 kDa cut-off and subsequently exchanged into an “NMR buffer” (10 mM MES at pH 6.0, 150 mM KCl, 1 mM TCEP, 0.02% NaN₃, and 8% (v/v) D₂O using a Midi-Trap G25 desalting column (GE Healthcare).

NMR Spectroscopy

All proteins were concentrated and buffer exchanged using 10 kDa (C1B-C2), 3 kDa (C1B) and 5 kDa (C2) cut-off Vivaspin® 15R concentrators into an NMR buffer. The experiments were carried out on Avance III HD NMR spectrometers (Bruker Biospin), operating at the ¹H Larmor frequencies of 800 MHz (18.8 Tesla) and 600 MHz (14.1 Tesla) equipped with cryogenically cooled probes, and 500 MHz (11.7 Tesla) equipped with a room temperature probe. The temperature was calibrated using deuterated (D₄, 98%) methanol for cryogenically cooled probes and protonated methanol for the room temperature probe. Spectra were processed using NMRPipe (Delaglio et al., 1995). The cross-peak intensities were obtained using Sparky (Si et al., 2015). Sequence-specific assignments of the ¹H_N and ¹⁵N resonances for apo C1B-C2 were obtained using ²H-decoupled 3D HN(CA)CB, HNCA(CB), HN(COCA)CB, and HN(CO)CA (Yamazaki et al., 1994) experiments on a [U-¹³C, ¹⁵N; 55%-²H] C1B-C2 sample. Resonance assignments for Cd²⁺-substituted C1B (C1B^{Cd}) were transferred from those for the native Zn²⁺-containing protein (C1B^{Zn}) and subsequently verified using 3D CBCA(CO)NH and HNCACB (Muhandiram and Kay, 1994) spectra collected at 14.1 Tesla. Resonance assignments for Cd²⁺-bound C1B-C2 were transferred from those for the isolated C1B^{Cd} and the Cd²⁺-complexed C2 (Morales et al., 2013b) domains. Chemical shift perturbations Δ were calculated between Cd²⁺-free and Cd²⁺-containing C1B-C2 as well as micelle/PDBu bound C1B^{Cd} and apo C1B^{Cd} according to the following equation:

$$\Delta = \sqrt{\Delta\delta_H^2 + (0.152\Delta\delta_N)^2} \quad (1)$$

where Δδ_H and Δδ_N are residue-specific ¹H_N and ¹⁵N chemical shift differences. For the NMR-detected binding experiments, the C1B ligand, phorbol-12,13-dibutyrate (PDBu, Sigma-Aldrich)

was dissolved in [²H₆] DMSO (Cambridge Isotopes) and added to the sample containing 94 μM of [U-¹⁵N] enriched C1B^{Cd} in the presence of 10 mM mixed micelles. Mixed micelles comprising [²H₃₈] dodecylphosphocholine, (DPC, Cambridge Isotopes) and 2-dihexanoyl-sn-glycero-3-[phospho-L-serine] (DPS, Avanti Polar Lipids) at a molar ratio of seven to three were prepared as previously described (Stewart et al., 2011). The final concentration of PDBu in the NMR sample was 100 μM.

Determination of Relative Cd²⁺ and Zn²⁺ Affinities to C1B

The four possible Zn/Cd metallated protein states are identified using the following nomenclature: C1B^{Zn} (native C1B with Zn²⁺ at both structural sites), C1B^{Cd} (Cd²⁺ at both structural sites), C1B^{Zn/Cd} (Zn²⁺ at site 1 and Cd²⁺ at site 2), and C1B^{Cd/Zn} (Cd²⁺ at site 1 and Zn²⁺ at site 2). The fractional populations of those protein species can be defined as:

$$f_{Zn} = \frac{I_{Zn}}{\Sigma}, f_{Zn/Cd} = \frac{I_{Zn/Cd}}{\Sigma}, f_{Cd/Zn} = \frac{I_{Cd/Zn}}{\Sigma}, f_{Cd} = \frac{I_{Cd}}{\Sigma} \quad (2)$$

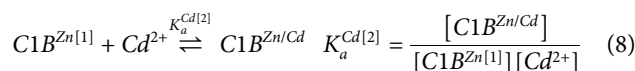
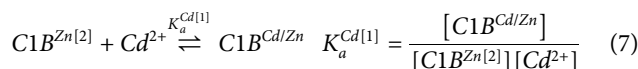
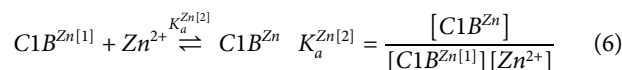
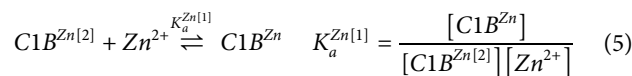
$$\Sigma = I_{Zn} + I_{Zn/Cd} + I_{Cd/Zn} + I_{Cd}$$

where *I* is the intensity of the corresponding N-H cross peaks in the ¹⁵N-¹H HSQC spectra for H140, I145, and V147. The concentrations of free Cd²⁺ ([Cd²⁺]) and Zn²⁺ ([Zn²⁺]) can be calculated from the following mass balance equations:

$$\frac{[Cd^{2+}]_0}{P_0} = \frac{[Cd^{2+}]}{P_0} + 2f_{Cd} + f_{Zn/Cd} + f_{Cd/Zn} \quad (3)$$

$$2 = \frac{[Zn^{2+}]}{P_0} + 2f_{Zn} + f_{Zn/Cd} + f_{Cd/Zn} \quad (4)$$

where *P*₀, [Cd²⁺]₀, and [Zn²⁺]₀ = 2×*P*₀ are the total concentrations of protein, Cd²⁺, and Zn²⁺, respectively. It is convenient to define the affinities of metal ions to C1B in terms of individual sites. For the single metal-ion bound species, we use the *M* [*n*] notation, where *M* = Zn or Cd, and *n* = 1 or 2. For example, C1B^{Zn[2]} defines C1B with site 2 populated by Zn²⁺ and a vacant site 1, and *K*_a^{Zn[1]} defines the association constant for the binding of Zn²⁺ to site 1 when site 2 is already populated by Zn²⁺. The following equilibria describe the binding processes and the associated *K*_a values:



The relative affinity of Cd²⁺ and Zn²⁺ to sites 1 and 2 can then be defined as the ratio of the association constants:

$$\chi[1] = \frac{K_a^{Cd[1]}}{K_a^{Zn[1]}} = \frac{[Zn^{2+}]f_{Cd/Zn}}{[Cd^{2+}]f_{Zn}} \quad (9)$$

$$\chi[2] = \frac{K_a^{Cd[2]}}{K_a^{Zn[2]}} = \frac{[Zn^{2+}]f_{Zn/Cd}}{[Cd^{2+}]f_{Zn}} \quad (10)$$

The $\chi[n]$ ($n = 1$ or 2) values report on the relative affinities of Cd^{2+} and Zn^{2+} to a given site C1B site when Zn^{2+} populates the other. A similar set of equilibria can be constructed to obtain the relative Cd^{2+} and Zn^{2+} affinities when Cd^{2+} populates the other site:

$$\mu[1] = \frac{[Zn^{2+}]f_{Cd}}{[Cd^{2+}]f_{Zn/Cd}} \quad (11)$$

$$\mu[2] = \frac{[Zn^{2+}]f_{Cd}}{[Cd^{2+}]f_{Cd/Zn}} \quad (12)$$

$\chi[n]$ and $\mu[n]$ for sites 1 and 2 (Table 1) were calculated using the NMR cross-peak intensities and the total concentrations of Cd^{2+} , C1B, and Zn^{2+} in the system (see Eqs. 1–3). The NMR cross-peaks intensities were determined using the $[^{15}N\text{-}^1H]$ HSQC spectrum of 0.1 mM $[U\text{-}^{15}N]$ C1B Zn , equilibrated overnight in the presence of 0.1 mM Cd^{2+} .

Site-specific Kinetics of Cd^{2+} Binding to C1B

To monitor the kinetics of Cd^{2+} binding to C1B, 2-fold molar excess of Cd^{2+} was added to 200 μM $[U\text{-}^{13}C, ^{15}N; \sim 75\%\text{-}^2H]$ C1B Zn in 10 mM HEPES buffer at pH 7.2, 75 mM KCl, and 1 mM TCEP. The process of Zn^{2+} replacement with Cd^{2+} was monitored using SOFAST-HMQC NMR experiments that were conducted on a 500 MHz instrument (11.7 Tesla) equipped with a room temperature probe. The first time point started 12 min post Cd^{2+} addition, and each SOFAST HMQC experiment took 15 min. Because the inter-conversion between Zn^{2+} - and Cd^{2+} -

complexed states is in the slow exchange regime, at any time point their fractional population can be determined from the intensities of the corresponding amide cross-peaks in the SOFAST-HMQC spectra. We used the N-H resonances of His102, Cys132, Thr134, Cys135, and Leu150 as the reporters of Cd^{2+} binding to site 1; and Phe114, Cys115, His117, Cys118, Gly119, Ser120, Tyr123, Lys141, and Cys143 as the reporters of Cd^{2+} binding to site 2.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

TI conceived the study and administered the research project. TC and TI designed the study and wrote the article. TC conducted all experimental work and analyzed the data.

FUNDING

This work was supported by the National Science Foundation grant CHE-1905116. TC was supported in part by the National Institutes of Health grant R01 GM108998.

ACKNOWLEDGMENTS

We thank Sachin Katti for generating the sequence alignment shown in Figure 5C and for the critical reading of the article.

REFERENCES

- Andreini, C., Banci, L., Bertini, I., and Rosato, A. (2006). Zinc Through the Three Domains of Life. *J. Proteome Res.* 5 (11), 3173–3178. doi:10.1021/pr0603699
- Antal, C. E., Hudson, A. M., Kang, E., Zanca, C., Wirth, C., Stephenson, N. L., et al. (2015). Cancer-Associated Protein Kinase C Mutations Reveal Kinase's Role as Tumor Suppressor. *Cell.* 160 (3), 489–502. doi:10.1016/j.cell.2015.01.001
- Ben Mimouna, S., Le Charpentier, T., Lebon, S., Steenwinckel, J., Messaoudi, I., and Gressens, P. (2019). Involvement of the Synapse-specific Zinc Transporter ZnT3 in Cadmium-induced Hippocampal Neurotoxicity. *J. Cel Physiol.* 234, 15872–15884. doi:10.1002/jcp.28245
- Bertini, I., Decaria, L., and Rosato, A. (2010). The Annotation of Full Zinc Proteomes. *J. Biol. Inorg. Chem.* 15 (7), 1071–1078. doi:10.1007/s00775-010-0666-6
- Beyersmann, D., Block, C., and Malviya, A. N. (1994). Effects of Cadmium on Nuclear Protein Kinase C. *Environ. Health Perspect.* 102 (Suppl. 3), 177–180. doi:10.2307/3431783
- Block, C., Freyermuth, S., Beyersmann, D., and Malviya, A. N. (1992). Role of Cadmium in Activating Nuclear Protein Kinase C and the Enzyme Binding to Nuclear Protein. *J. Biol. Chem.* 267 (28), 19824–19828. doi:10.1016/s0021-9258(19)88628-3
- Budas, G., Churchill, E., and Mochlyrosen, D. (2007). Cardioprotective Mechanisms of PKC Isozyme-Selective Activators and Inhibitors in the Treatment of Ischemia-Reperfusion Injury. *Pharmacol. Res.* 55 (6), 523–536. doi:10.1016/j.phrs.2007.04.005
- Busenlehner, L. S., Cosper, N. J., Scott, R. A., Rosen, B. P., Wong, M. D., and Giedroc, D. P. (2001). Spectroscopic Properties of the Metalloregulatory Cd(II) and Pb(II) Sites of S. Aureus p1258 CadC. *Biochemistry.* 40 (14), 4426–4436. doi:10.1021/bi010006g
- Chakraborty, S., Yudenfreund Kravitz, J., Thulstrup, P. W., Hemmingsen, L., DeGrado, W. F., and Pecoraro, V. L. (2011). Design of a Three-Helix Bundle Capable of Binding Heavy Metals in a Triscysteine Environment. *Angew. Chem. Int. Ed.* 50 (9), 2049–2053. doi:10.1002/anie.201006413
- Choong, G., Liu, Y., and Templeton, D. M. (2014). Interplay of Calcium and Cadmium in Mediating Cadmium Toxicity. *Chemico-Biological Interactions.* 211, 54–65. doi:10.1016/j.cbi.2014.01.007
- Churchill, E., Budas, G., Vallentin, A., Koyanagi, T., and Mochly-Rosen, D. (2008). PKC Isozymes in Chronic Cardiac Disease: Possible Therapeutic Targets?. *Annu. Rev. Pharmacol. Toxicol.* 48, 569–599. doi:10.1146/annurev.pharmtox.48.121806.154902
- Cole, T. R., Erickson, S. G., Morales, K. A., Sung, M., Holzenburg, A., and Igumenova, T. I. (2019). Cd(II)- and Pb(II)-Induced Self-Assembly of Peripheral Membrane Domains From Protein Kinase C. *Biochemistry.* 58 (6), 509–513. doi:10.1021/acs.biochem.8b01235
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR.* 6 (3), 277–293. doi:10.1007/bf00197809

- Dempsey, E. C., Newton, A. C., Mochly-Rosen, D., Fields, A. P., Reyland, M. E., Insel, P. A., et al. (2000). Protein Kinase C Isozymes and the Regulation of Diverse Cell Responses. *Am. J. Physiology-Lung Cell Mol. Physiol.* 279 (3), L429–L438. doi:10.1152/ajplung.2000.279.3.L429
- Drosatos, K., Bharadwaj, K. G., Lymperopoulos, A., Ikeda, S., Khan, R., Hu, Y., et al. (2011). Cardiomyocyte Lipids Impair β -adrenergic Receptor Function via PKC Activation. *Am. J. Physiology-Endocrinology Metab.* 300 (3), E489–E499. doi:10.1152/ajpendo.00569.2010
- Duan, L., Kong, J.-J., Wang, T.-Q., and Sun, Y. (2018). Binding of Cd(II), Pb(II), and Zn(II) to a Type 1 Metallothionein From maize (*Zea mays*). *Biomaterials*. 31 (4), 539–550. doi:10.1007/s10534-018-0100-z
- Faroon, O., Ashizawa, A., Wright, S., Tucker, P., Jenkins, K., Ingerman, L., et al. (2012). *Toxicological Profile for Cadmium*. Atlanta (GA): Agency for Toxic Substances and Disease Registry.
- Habjanić, J., Chesnov, S., Zerbe, O., and Freisinger, E. (2020). Impact of Naturally Occurring Serine/cysteine Variations on the Structure and Function of Pseudomonas Metallothioneins. *Metallomics*. 12 (1), 23–33. doi:10.1039/C9MT00213H
- Hommel, U., Zurini, M., and Luyten, M. (1994). Solution Structure of a Cysteine Rich Domain of Rat Protein Kinase C. *Nat. Struct. Mol. Biol.* 1 (6), 383–387. doi:10.1038/nsb0694-383
- Huang, M., Krepkiy, D., Hu, W., and Petering, D. H. (2004). Zn-, Cd-, and Pb-Transcription Factor IIIA: Properties, DNA Binding, and Comparison With TFIIIA-finger 3 Metal Complexes. *J. Inorg. Biochem.* 98 (5), 775–785. doi:10.1016/j.jinorgbio.2004.01.014
- Hubbard, S. R., Bishop, W., Kirschmeier, P., George, S., Cramer, S., and Hendrickson, W. (1991). Identification and Characterization of Zinc Binding Sites in Protein Kinase C. *Science*. 254 (5039), 1776–1779. doi:10.1126/science.1763327
- Irie, K., Oie, K., Nakahara, A., Yanai, Y., Ohigashi, H., Wender, P. A., et al. (1998). Molecular Basis for Protein Kinase C Isozyme-Selective Binding: The Synthesis, Folding, and Phorbol Ester Binding of the Cysteine-Rich Domains of All Protein Kinase C Isozymes. *J. Am. Chem. Soc.* 120 (36), 9159–9167. doi:10.1021/ja981087f
- Jacobson, T., Priya, S., Sharma, S. K., Andersson, S., Jakobsson, S., Tanghe, R., et al. (2017). Cadmium Causes Misfolding and Aggregation of Cytosolic Proteins in Yeast. *Mol. Cell Biol.* 37 (17). doi:10.1128/MCB.00490-16
- Järup, L., Berglund, M., Elinder, C. G., Nordberg, G., and Vahter, M. (1998). Health Effects of Cadmium Exposure—a Review of the Literature and a Risk Estimate. *Scand. J. Work Environ. Health*. 24 (Suppl. 1), 1–51. https://www.sjweh.fi/show_abstract.php?abstract_id=281
- Johnson, J. A., Adak, S., and Mochly-Rosen, D. (1995). Prolonged Phorbol Ester Treatment Down-Regulates Protein Kinase C Isozymes and Increases Contraction Rate in Neonatal Cardiac Myocytes. *Life Sci.* 57 (11), 1027–1038. doi:10.1016/0024-3205(95)02048-n
- Katti, S., and Igumenova, T. I. (2021). Structural Insights Into C1-Ligand Interactions: Filling the Gaps by In Silico Methods. *Adv. Biol. Regul.* 79, 100784. doi:10.1016/j.jbior.2020.100784
- Katti, S., Nyenhuis, S. B., Her, B., Srivastava, A. K., Taylor, A. B., Hart, P. J., et al. (2017). Non-Native Metal Ion Reveals the Role of Electrostatics in Synaptotagmin 1-Membrane Interactions. *Biochemistry*. 56 (25), 3283–3295. doi:10.1021/acs.biochem.7b00188
- Khan, T. K., Nelson, T. J., Verma, V. A., Wender, P. A., and Alkon, D. L. (2009). A Cellular Model of Alzheimer's Disease Therapeutic Efficacy: PKC Activation Reverses A β -Induced Biomarker Abnormality on Cultured Fibroblasts. *Neurobiol. Dis.* 34 (2), 332–339. doi:10.1016/j.nbd.2009.02.003
- Kluska, K., Adamczyk, J., and Krężel, A. (2018). Metal Binding Properties, Stability and Reactivity of Zinc Fingers. *Coord. Chem. Rev.* 367, 18–64. doi:10.1016/j.ccr.2018.04.009
- Koya, D., and King, G. L. (1998). Protein Kinase C Activation and the Development of Diabetic Complications. *Diabetes*. 47 (6), 859–866. doi:10.2337/diabetes.47.6.859
- Krizek, B. A., Merkle, D. L., and Berg, J. M. (1993). Ligand Variation and Metal Ion Binding Specificity in Zinc finger Peptides. *Inorg. Chem.* 32(6), 937–940. doi:10.1021/ic00058a030
- Lemmon, M. A. (2008). Membrane recognition by phospholipid-binding domains. *Nat. Rev. Mol. Cell Biol.* 9 (2), 99–111. doi:10.1038/nrm2328
- Long, G. J. (1997). The Effect of Cadmium on Cytosolic Free Calcium, Protein Kinase C, and Collagen Synthesis in Rat Osteosarcoma (ROS 17/2.8) Cells. *Toxicol. Appl. Pharmacol.* 143 (1), 189–195. doi:10.1006/taap.1996.8060
- Malgieri, G., Palmieri, M., Esposito, S., Maione, V., Russo, L., Baglivo, I., et al. (2014). Zinc to Cadmium Replacement in the Prokaryotic Zinc-finger Domain. *Metallomics*. 6 (1), 96–104. doi:10.1039/c3mt00208j
- Malgieri, G., Zaccaro, L., Leone, M., Bucci, E., Esposito, S., Baglivo, I., et al. (2011). Zinc to Cadmium Replacement in the A. Thaliana SUPERMAN Cys2His2 Zinc finger Induces Structural Rearrangements of Typical DNA Base Determinant Positions. *Biopolymers*. 95 (11), a–n. doi:10.1002/bip.21680
- Markovac, J., and Goldstein, G. W. (1988). Picomolar Concentrations of lead Stimulate Brain Protein Kinase C. *Nature* 334 (6177), 71–73. doi:10.1038/334071a0
- Michalek, J. L., Lee, S. J., and Michel, S. L. J. (2012). Cadmium Coordination to the Zinc Binding Domains of the Non-Classical Zinc finger Protein Tristetraprolin Affects RNA Binding Selectivity. *J. Inorg. Biochem.* 112, 32–38. doi:10.1016/j.jinorgbio.2012.02.023
- Mishra, D., and Dey, C. S. (2021). PKC α : Prospects in Regulating Insulin Resistance and AD. *Trends Endocrinol. Metab.* 32 (6), 341–350. doi:10.1016/j.tem.2021.03.006
- Morales, K. A., Lasagna, M., Gribenko, A. V., Yoon, Y., Reinhart, G. D., Lee, J. C., et al. (2011). Pb2+ as a Modulator of Protein-Membrane Interactions. *J. Am. Chem. Soc.* 133 (27), 10599–10611. doi:10.1021/ja2032772
- Morales, K. A., Yang, Y., Long, Z., Li, P., Taylor, A. B., Hart, P. J., et al. (2013a). Cd2+ as a Ca2+ Surrogate in Protein-Membrane Interactions: Isostructural but Not Isofunctional. *J. Am. Chem. Soc.* 135 (35), 12980–12983. doi:10.1021/ja406958k
- Morales, K. A., Yang, Y., Long, Z., Li, P., Taylor, A. B., Hart, P. J., et al. (2013b). Cd2+ as a Ca2+ Surrogate in Protein-Membrane Interactions: Isostructural but Not Isofunctional. *J. Am. Chem. Soc.* 135 (35), 12980–12983. doi:10.1021/ja406958k
- Muhandiram, D. R., and Kay, L. E. (1994). Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments With Improved Sensitivity. *J. Magn. Reson. Ser. B*. 103(3), 203–216. doi:10.1006/jmr.1994.1032
- Newton, A. C. (2010). Protein Kinase C: Poised to Signal. *Am. J. Physiology-Endocrinology Metab.* 298 (3), E395–E402. doi:10.1152/ajpendo.00477.2009
- Pan, T., and Coleman, J. E. (1990). GAL4 Transcription Factor Is Not a "zinc finger" but Forms a Zn(II)2Cys6 Binuclear Cluster. *Proc. Natl. Acad. Sci.* 87 (6), 2077–2081. doi:10.1073/pnas.87.6.2077
- Pan, T., Freedman, L. P., and Coleman, J. E. (1990). Cadmium-113 NMR Studies of the DNA Binding Domain of the Mammalian Glucocorticoid Receptor. *Biochemistry*. 29 (39), 9218–9225. doi:10.1021/bi00491a016
- Petering, D. H. (2017). Reactions of the Zn Proteome with Cd2+ and Other Xenobiotics: Trafficking and Toxicity. *Chem. Res. Toxicol.* 30 (1), 189–202. doi:10.1021/acs.chemrestox.6b00328
- Puljung, M. C., and Zagotta, W. N. (2011). Labeling of Specific Cysteines in Proteins Using Reversible Metal Protection. *Biophysical J.* 100 (10), 2513–2521. doi:10.1016/j.bpj.2011.03.063
- Rahimova, N., Cooke, M., Zhang, S., Baker, M. J., and Kazanietz, M. G. (2020). The PKC Universe Keeps Expanding: From Cancer Initiation to Metastasis. *Adv. Biol. Regul.* 78, 100755. doi:10.1016/j.jbior.2020.100755
- Saijoh, K., Inoue, Y., Katsuyama, H., and Sumino, K. (1988). The Interaction of Cations With Activity of Soluble Protein Kinase C from Mouse Brain. *Pharmacol. Toxicol.* 63 (4), 221–224. doi:10.1111/j.1600-0773.1988.tb00944.x
- Sharma, S. K., Goloubinoff, P., and Christen, P. (2008). Heavy Metal Ions Are Potent Inhibitors of Protein Folding. *Biochem. Biophysical Res. Commun.* 372 (2), 341–345. doi:10.1016/j.bbrc.2008.05.052
- Si, Y.-X., Lee, J., Cai, Y., Yin, S.-J., Yang, J.-M., Park, Y.-D., et al. (2015). Molecular Dynamics Simulations Integrating Kinetics for Pb2+-Induced Arginine Kinase Inactivation and Aggregation. *Process Biochem.* 50 (5), 729–737. doi:10.1016/j.procbio.2015.02.008
- Speizer, L. A., Watson, M. J., Kanter, J. R., and Brunton, L. L. (1989). Inhibition of Phorbol Ester Binding and Protein Kinase C Activity by Heavy Metals. *J. Biol. Chem.* 264 (10), 5581–5585. doi:10.1016/s0021-9258(18)83586-4
- Stewart, M. D., Cole, T. R., and Igumenova, T. I. (2014). Interfacial Partitioning of a Loop Hinge Residue Contributes to Diacylglycerol Affinity of Conserved Region 1 Domains. *J. Biol. Chem.* 289 (40), 27653–27664. doi:10.1074/jbc.M114.585570
- Stewart, M. D., and Igumenova, T. I. (2012). Reactive Cysteine in the Structural Zn2+ Site of the C1B Domain from PKC α . *Biochemistry*. 51 (37), 7263–7277. doi:10.1021/bi300750w
- Stewart, M. D., Morgan, B., Massi, F., and Igumenova, T. I. (2011). Probing the Determinants of Diacylglycerol Binding Affinity in the C1B Domain of Protein Kinase Ca. *J. Mol. Biol.* 408 (5), 949–970. doi:10.1016/j.jmb.2011.03.020

- Sun, X., Tian, X., Tomsig, J. L., and Suszkiw, J. B. (1999). Analysis of Differential Effects of Pb²⁺ on Protein Kinase C Isozymes. *Toxicol. Appl. Pharmacol.* 156 (1), 40–45. doi:10.1006/taap.1999.8622
- Tomsig, J. L., and Suszkiw, J. B. (1995). Multisite Interactions Between Pb²⁺ and Protein Kinase C and its Role in Norepinephrine Release from Bovine Adrenal Chromaffin Cells. *J. Neurochem.* 64 (6), 2667–2673. doi:10.1046/j.1471-4159.1995.64062667.x
- Verdaguer, N., Corbalan-Garcia, S., Ochoa, W. F., Fita, I., and Gómez-Fernández, J. C. (1999). Ca²⁺ Bridges the C2 Membrane-Binding Domain of Protein Kinase Cα Directly to Phosphatidylserine. *EMBO J.* 18 (22), 6329–6338. doi:10.1093/emboj/18.22.6329
- Waalkes, M. P. (2003). Cadmium Carcinogenesis. *Mutat. Res.* 533 (1-2), 107–120. doi:10.1016/j.mrfmmm.2003.07.011
- Waalkes, M. P., and Rehm, S. (1992). Carcinogenicity of Oral Cadmium in the Male Wistar (WF/NCR) Rat: Effect of Chronic Dietary Zinc Deficiency. *Toxicol. Sci.* 19 (4), 512–520. doi:10.1093/toxsci/19.4.512
- Wang, Y., Hemmingsen, L., and Giedroc, D. P. (2005). Structural and Functional Characterization of *Mycobacterium tuberculosis* CmtR, a PbII/CdII-Sensing SmtB/ArsR Metalloregulatory Repressor. *Biochemistry.* 44 (25), 8976–8988. doi:10.1021/bi050094v
- Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., and Kay, L. E. (1994). A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of ¹⁵N, ¹³C, ²H Labeled Proteins with High Sensitivity. *J. Am. Chem. Soc.* 116 (26), 11655–11666. doi:10.1021/ja00105a005
- Zhang, G., Kazanietz, M. G., Blumberg, P. M., and Hurley, J. H. (1995). Crystal Structure of the Cys2 Activator-Binding Domain of Protein Kinase Cδ in Complex with Phorbol Ester. *Cell.* 81 (6), 917–924. doi:10.1016/0092-8674(95)90011-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Cole and Igumenova. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Cholesterol Is a Dose-Dependent Positive Allosteric Modulator of CCR3 Ligand Affinity and G Protein Coupling

Evan van Aalst and Benjamin J. Wylie*

Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, TX, United States

OPEN ACCESS

Edited by:

Vincenzo Venditti,
Iowa State University, United States

Reviewed by:

Oskar Engberg,
University Hospital Leipzig, Germany
Marvin J. Bayro,
University of Puerto Rico, Puerto Rico

*Correspondence:

Benjamin J. Wylie
benjamin.j.wylie@ttu.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 13 June 2021

Accepted: 14 July 2021

Published: 20 August 2021

Citation:

van Aalst E and Wylie BJ (2021)
Cholesterol Is a Dose-Dependent
Positive Allosteric Modulator of CCR3
Ligand Affinity and G Protein Coupling.
Front. Mol. Biosci. 8:724603.
doi: 10.3389/fmolb.2021.724603

Cholesterol as an allosteric modulator of G protein-coupled receptor (GPCR) function is well documented. This quintessential mammalian lipid facilitates receptor–ligand interactions and multimerization states. Functionally, this introduces a complicated mechanism for the homeostatic modulation of GPCR signaling. Chemokine receptors are Class A GPCRs responsible for immune cell trafficking through the binding of endogenous peptide ligands. CCR3 is a CC motif chemokine receptor expressed by eosinophils and basophils. It traffics these cells by transducing the signal stimulated by the CC motif chemokine primary messengers 11, 24, and 26. These behaviors are close to the human immunoresponse. Thus, CCR3 is implicated in cancer metastasis and inflammatory conditions. However, there is a paucity of experimental evidence linking the functional states of CCR3 to the molecular mechanisms of cholesterol–receptor cooperativity. In this vein, we present a means to combine codon harmonization and a maltose-binding protein fusion tag to produce CCR3 from *E. coli*. This technique yields ~2.6 mg of functional GPCR per liter of minimal media. We leveraged this protein production capability to investigate the effects of cholesterol on CCR3 function *in vitro*. We found that affinity for the endogenous ligand CCL11 increases in a dose-dependent manner with cholesterol concentration in both styrene:maleic acid lipid particles (SMALPs) and proteoliposomes. This heightened receptor activation directly translates to increased signal transduction as measured by the GTPase activity of the bound G-protein α inhibitory subunit 3 ($G\alpha_i3$). This work represents a critical step forward in understanding the role of cholesterol-GPCR allostery in regulation of signal transduction.

Keywords: GPCR, CCR3, chemokine receptor, CCL11, cholesterol, G protein, lipid allostery, codon harmonization, GPCR

INTRODUCTION

G protein-coupled receptors (GPCRs) are integral membrane proteins comprising a canonical seven-transmembrane alpha helical architecture (Rosenbaum et al., 2009). In response to external stimuli, this helical bundle undergoes a conformational change that is recognized by an intracellular heterotrimeric G protein (Kim et al., 2013). This molecular recognition event leads to an exchange of bound GDP for GTP in the G protein, triggering dissociation of the α and $\beta\gamma$ subunits (Figure 1) (Sullivan et al., 1987). The α subunit then acts as an effector to influence downstream events such as modulation of adenylate cyclase functionality (Federman et al., 1992), while the $\beta\gamma$ subunit can trigger cleavage of phosphatidylinositol-(4,5)-bisphosphate (PIP_2) (Katz et al., 1992) and ion channel activation (Pegan et al., 2005; Nishida et al., 2007).

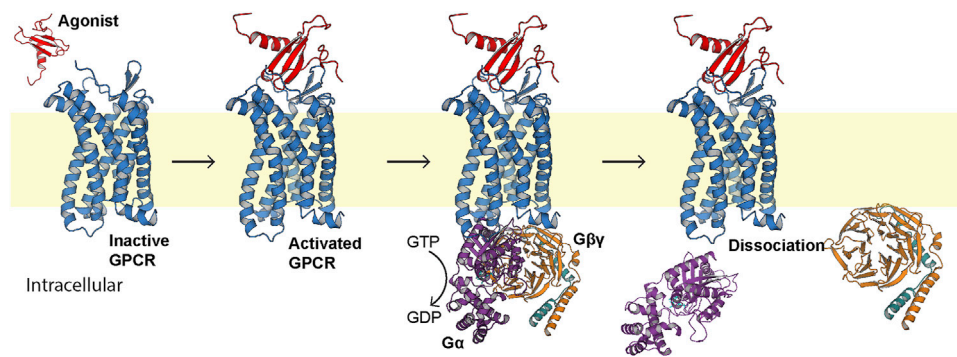


FIGURE 1 | Generalized first step of GPCR signal transduction. The extracellular agonist (red, PDB ID 1EOT) (Crump et al., 1998) binds to the orthosteric pocket of the GPCR (blue) to elicit a conformational change recognized intracellularly by the heterotrimeric G protein (PDB ID 1GP2) (Wall et al., 1995). The G protein binds, exchanges bound GDP for GTP, and the α subunit (purple) dissociates from the $\beta\gamma$ subunits (orange and teal, respectively).

Functional interplay between lipid constituents and membrane proteins is well documented. We previously reported that the bacterial K^+ channel KirBac1.1 orders the membrane (Borcik et al., 2019) and that the membrane activates structural transitions and associated biological functions (Amani et al., 2020; Borcik et al., 2020). Like K^+ channels, GPCRs are regulated by lipids through both direct allosteric interactions and changes to membrane mechanical and thermodynamic properties (Botelho et al., 2006). Perhaps the most widely studied of these functional lipids is cholesterol (Jafurulla et al., 2019), although allosteric effects of phosphoserines (Dawaliby et al., 2016), sphingolipids (Chattopadhyay, 2014), phosphoinositols like PIP_2 (Yen et al., 2018), and the binding synergy between multiple lipid species (Xu et al., 2021) are of increasing interest. A canonical Class A GPCR cholesterol consensus motif (CCM) of $(R,K)^{4.39-4.43}-(W,Y)^{4.50}-(I,V,L)^{4.46}-(F,Y)^{2.41}$ has been identified in the β_2 adrenergic receptor (β_2AR), following the Weinstein-Ballesteros numbering convention (Ballesteros and Weinstein, 1995; Hanson et al., 2008). However, only 21% of Class A receptors contain this sequence (Hanson et al., 2008), and this motif is conspicuously absent in chemokine receptors (Legler et al., 2017). Furthermore, it is observed that such cholesterol binding motifs are not necessarily occupied even when present (Marlow et al., 2021). Cholesterol has, nevertheless, still been implicated as an allosteric modulator of chemokine receptor function (Zhukovsky et al., 2013; Pluhackova et al., 2016; Legler et al., 2017; Gahbauer et al., 2018; Calmet et al., 2020).

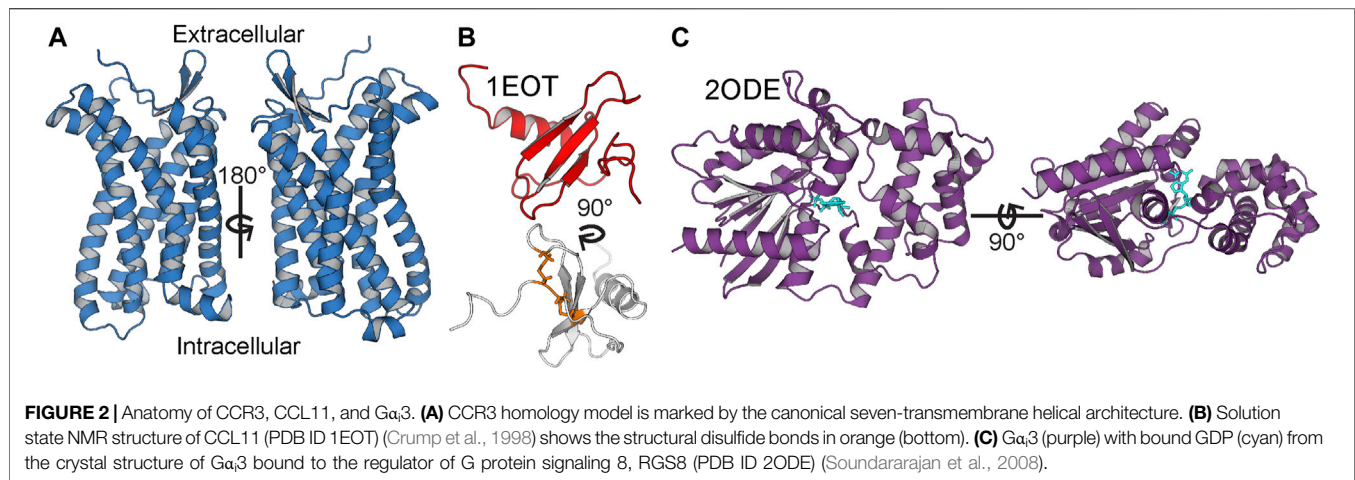
Chemokine receptors are a subclass of Class A GPCRs. They trigger cellular trafficking of immune cells in response to chemotactic cytokine (chemokine) ligands (Stone et al., 2017). Perhaps the most well-known chemokine receptors are CC motif chemokine receptor 5 (CCR5) and CXC motif chemokine receptor 4 (CXCR4), which act as coreceptors for HIV infection (Deng et al., 1996; Ma et al., 1998). Although the CCM identified in β_2AR is absent from chemokine receptors, experimental evidence indicates receptor–lipid interactions, often driving receptor dimerization (Pluhackova et al., 2016; Legler et al., 2017; Gahbauer et al., 2018; Calmet et al., 2020). However,

there are many outstanding questions regarding the direct and indirect influence of cholesterol on function *in vivo*. Much of what is known is derived from crystal structure co-crystallization and molecular dynamics (MD) simulations. While these studies are foundational, the conclusions are not definitive and experimental functional data are needed.

CC motif chemokine receptor 3 (CCR3, **Figure 2A**) is a Class A GPCR mainly expressed by eosinophils. Like all chemokine receptors, its primary messengers are endogenous peptide ligands. Specifically, CCL11, CCL24, and CCL26 (eotaxins 1, 2, and 3, respectively; CCL11 is depicted in **Figure 2B**) activate CCR3 and trigger chemotaxis of the expressing cell (Ge et al., 2015). This occurs through the G protein inhibitory α subunit ($G\alpha_i$, **Figure 2C**), which triggers downstream inhibition of adenylate cyclase (Kitaura et al., 1996). It is, however, unclear as to which of the 3 $G\alpha_i$ subunits is primarily involved in the signaling cascade or if there is significant promiscuity *in vivo* between CCR3 and $G\alpha_i$ s 1, 2, and 3. Moreover, the influence of direct lipid allostery on CCR3-G protein coupling and signal transduction is undocumented.

Given its role in leukocyte trafficking, CCR3 provokes inflammatory conditions such as asthma (Gauvreau et al., 2018), rheumatoid arthritis (Katschke et al., 2001), and eosinophilic esophagitis (Dunn et al., 2020). Furthermore, CCR3 is correlated with heightened invasive potential of metastatic liver (Jin et al., 2017), prostate (Ishida et al., 2018), and kidney cancers (Johrer et al., 2005). It is also a coreceptor for some strains of HIV (He et al., 1997). As a result, CCR3 is an attractive therapeutic target. However, comparatively little is known about the structural biology of CCR3 with respect to other chemokine receptors such as CCR5 and CXCR4. CCR3's natural agonists help regulate the relative monomer–dimer higher-order oligomer populations *in vivo* (Song et al., 2018), but the influence of cholesterol on this interaction is unknown. Therefore, investigation of the structure–function relationship and the lipid agency is an attractive and necessary long-term goal.

In pursuit of this goal, we implemented a codon harmonization scheme we previously reported to positively influence heterologous membrane protein yield (van Aalst



et al., 2020). This technique is called DNA codon usage for measured base optimization, or DUMB optimization (DO). Codon harmonization aims to site-specifically modify the codons comprising the heterologous construct to match the codon usage frequency of the native organism more closely in order to improve cotranslational folding (Angov et al., 2008; Rodriguez et al., 2018). This increases the fraction of protein that is properly folded, membrane-inserted, and functional by slowing translation through the introduction of targeted rare codons. Through implementation of DUMB optimization and incorporation of a maltose-binding protein (MBP) solubility tag (Ge et al., 2015) in our construct, we report the heterologous production of CCR3 after tag removal at yields of ~2.5–2.8 mg/L from M9 minimal media (Bhate et al., 2013). CCL11 and Gα₃ were DUMB optimized as a matter of course in expression optimization, attaining 2.6 ± 0.3 and 15.1 ± 0.3 mg/L, respectively, for each, from M9 minimal media. We describe the positive cooperativity between membrane cholesterol and CCR3 binding affinity to its endogenous ligand CCL11, quantified *via* a fluorescence polarization assay (Rossi and Taylor, 2011) in lipid environments of increasing cholesterol content. We then confirm that this cholesterol-induced modulation of ligand affinity translates to increased signal transduction, measured *via* coupling to and activation of Gα₃ GTPase. This is, to our knowledge, the first experimental evidence of cholesterol-receptor interactions and their effect on ligand affinity and the efficacy of signal transduction catalyzed by CCR3.

MATERIALS AND METHODS

Construct Design and Gene Insertion

The human CCR3 amino acid sequence was obtained from the UniProt database (P51677). The human CCL11 amino acid sequence was obtained from the UniProt database (P51671) and truncated to residues 24–97 to remove the propeptide. The human Gα₃ inhibitory 3 (Gα₃) amino acid sequence was obtained from the UniProt database (P08754). The amino acid sequences were reverse engineered into fully optimized (FO)

DNA sequences according to *E. coli* codon usage, serving as a platform for the application of DUMB optimization. Here, the FO constructs use codons to transcribe the proteins that correspond to only the most abundantly found tRNAs in the expression vector.

The native human CCR3 DNA sequence was obtained from the European Nucleotide Archive (ENA, Sequence: U51241) and codon-harmonized according to DUMB optimization (DO) for expression in *E. coli*, as previously described (van Aalst et al., 2020). Briefly, codons were replaced in the expression sequence to match the codon usage frequencies found within the native sequence as *E. coli* codon usage permitted. Substitution was performed such that no alternative codons in the host system were chosen with a relative usage below 5% of the native usage frequency. Human and *E. coli* codon usage frequencies from the Graphical Codon Usage Analyzer (<http://gcu.schoedl.de/>) were used in designing the DO gene sequence (Fuhrmann et al., 2004). For more information on the codon harmonization process, see reference van Aalst et al., 2020. %MinMax analysis (Rodriguez et al., 2018) of the native, FO, and DO sequences visualizes the extent of deoptimization (Figure 3). The native human Gα₃ (Sequence: J03005.1) and human CCL11 (D49372.1) gene sequences were also obtained from the ENA and codon-harmonized in the same way as a matter of course in the optimization process.

CCR3, CCL11, and Gα₃ gene sequences were ordered from GeneArt (Thermo Fisher). The expression vector pMAL-p4x was ordered from Addgene. 5' EcoRI and 3' Hind III restriction enzymes (Thermo Scientific) were used to clone each sequence into the vector downstream of MalE. This resulted in an N-maltose-binding protein (MBP)-8x His tag-TEV site-CCL11-C construct, in the case of CCL11. For CCR3, the initial FO construct that was described (Ge et al., 2015) was DUMB optimized, and a GGGGS 4x repeat between the TEV site and the N terminus of CCR3 was added to promote cleavage (Chen et al., 2013). This resulted in an N-maltose-binding protein (MBP)-8x His tag-TEV site-(GGGGS)₄-CCR3-C construct. The restriction enzymes NcoI and BamHI (Thermo Scientific) were used to insert the Gα₃ sequence into the

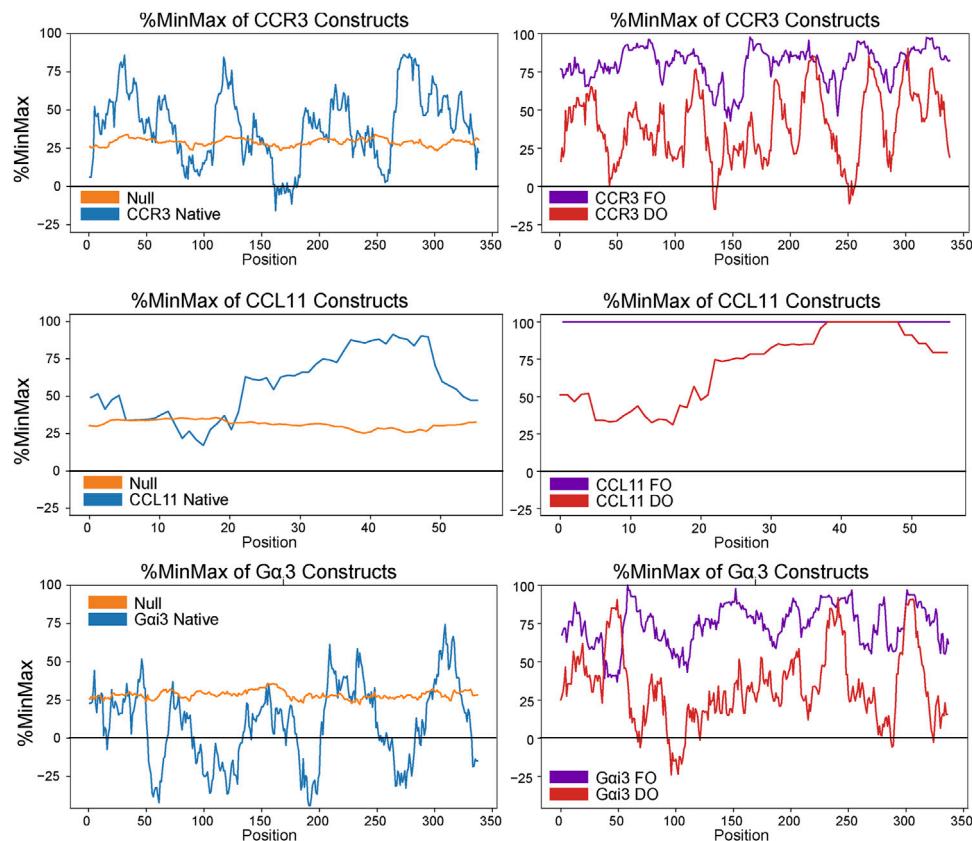


FIGURE 3 | %MinMax (Rodriguez et al., 2018) of native sequences compared to optimized heterologous sequences (FO) and constructs codon-harmonized through DUMB optimization (DO). In general, %MinMax of the DO sequences (red) more closely resemble the native values (blue) than the FO values (purple).

expression vector pET-28a (+) (Novus Biologicals). This resulted in an N-Met-Gly-8x His-TEV site-Gα₃-C construct. In all cases, gene insertion and sequence conservation were verified by sequencing.

CCR3 Expression and Purification

E. coli C43 (DE3) cells harboring the pMAL-p4x-CCR3 plasmid were grown in M9 minimal media containing 2 mM MgSO₄, 0.1 mM CaCl₂, 100 μg/ml ampicillin, 3 ml of 100x minimum essential vitamin stock, 96.22 mM Na₂HPO₄, 44.1 mM K₂HPO₄, 17.1 mM NaCl, 5 g glucose per L (0.5% w/v), 3.75 g NH₄Cl per L (0.375% w/v), and 20 ml of Solution C (**Supplementary Table S1**) per L. Cultures were grown at 220 rpm and 37°C until an OD₆₀₀ of ~1.0 was reached. The cultures were cooled to 18°C and induced with 1 mM IPTG for 24 h. Cell cultures were then pelleted at 5,500 rpm for 10 min and stored at -80°C for future use.

Cell pellets were removed from storage at -80°C and resuspended in 5 ml of lysis cocktail per g of cells (20 mM HEPES, pH 8.0, 150 mM KCl, 0.02% NaN₃, 10 mM MgSO₄, 0.2 mg/ml lysozyme, 0.2 mg/ml RNase A, 1 mM phenylmethylsulfonyl fluoride (PMSF), and 1 mM benzamide). PIERCE EDTA-free protease inhibitor tablets (Thermo Fisher) were added at one tablet per 6 g of cells. Cells were lysed *via* homogenization, and protein was

extracted from membranes *via* addition of n-Dodecyl-β-D-Maltoside (DDM, Anatrace) and cholesteryl hemisuccinate Tris-salt (CHS, Anatrace) to final concentrations of 20 and 2 mM, respectively. Extraction took place overnight at 4°C with rocking. The solution was centrifuged at 125,000 g for 40 min at 4°C to remove cell debris. The supernatant was filtered through a 0.22-μm PES bottle top filter and loaded onto a 5-ml nickel affinity column (NAC, GE Healthcare Life Sciences) equilibrated in wash buffer (20 mM HEPES, pH 8.0, 150 mM KCl, 10 mM imidazole, 0.02% NaN₃, 2.5 mM DDM, and 0.25 mM CHS). The column was then treated with five column volumes of wash buffer before elution with five column volumes of elution buffer (20 mM HEPES, pH 8.0, 150 mM KCl, 250 mM imidazole, 0.02% NaN₃, 2.5 mM DDM, and 0.25 mM CHS).

Eluted protein was transferred into cleavage buffer (20 mM HEPES, pH 8.0, 150 mM KCl, 1 mM EDTA, 1 mM DDM, 0.1 mM CHS, and 0.5 mM DTT) using a desalting column (GE Healthcare Life Sciences) equilibrated with cleavage buffer. Tobacco etch virus (TEV) protease (Kapust et al., 2001) was added from glycerol stocks at a typical ratio of 1 mg TEV to 3 mg MBP-CCR3 with 1 mM DTT and set to rock overnight at 4°C. Samples were then transferred into wash buffer to remove EDTA and subjected to the NAC. Flow-through containing cleaved CCR3 was collected and the

elution peak containing TEV, the MBP tag, and uncleaved MBP-CCR3 was discarded. The flow-through was then concentrated to ~2 mg/ml using an Amicon Stirred Cell with Ultracel 30 kDa Ultrafiltration Discs (Millipore) before loading onto a HiLoad 16/600 Superdex Prep grade 200 column (GE Healthcare Life Sciences) equilibrated in exchange buffer. Cleaved CCR3 fractions were pooled, concentration was determined using optical density at 280 nm, and samples were stored at 4°C for future use.

CCL11 Expression and Purification

pMAL-p4x harboring CCL11 was cultured in M9 minimal media at 220 rpm and 37°C until an OD₆₀₀ of ~0.8 was reached. The cultures were cooled to 18°C and induced with 0.5 mM IPTG for 24 h. Cell aliquots were centrifuged at 5.5 k rpm for 10 min, and the resulting pellets were stored at -80°C for future use.

Cell pellets were resuspended in CCL11 wash buffer (20 mM HEPES, pH 7.5, 300 mM KCl, 10 mM Imidazole, and 0.02% NaN₃) supplemented with 0.2 mg/ml lysozyme, 0.2 mg/ml RNase A, 1 mM PMSF, 1 mM benzamidine, and 1 PIERCE EDTA-free protease inhibitor tablet per 6 g of cells. Cells were lysed *via* triplicate passage through a homogenizer. Cell debris was centrifuged at 17,000 rpm for 30 min followed by clarification of the lysate using a 0.22-μm PES bottle top filter. The clarified lysate was loaded onto an NAC preequilibrated in CCL11 wash buffer, washed five times with the same, and eluted with five column volumes of CCL11 elution buffer (20 mM HEPES, pH 7.5, 300 mM KCl, 250 mM imidazole, and 0.02% NaN₃). The elution peak was then transferred back into CCL11 wash buffer using a desalting column for the reverse NAC.

The TEV was added in a typical ratio of 1 mg TEV per 4 mg MBP-CCL11 and set to cleave at 4°C with rocking for 1 h. The cleavage mixture was then passed through the NAC, and the flow-through containing cleaved CCL11 was collected. This was concentrated to ~5 ml using an Amicon Stirred Cell with Ultracel 3 kDa Ultrafiltration Discs (Millipore) and loaded onto a HiLoad 16/600 Superdex Prep grade 75 column equilibrated in CCL11 exchange buffer (20 mM HEPES, pH 7.5, 300 mM KCl, 1 mM EDTA, and 0.02% NaN₃). The CCL11 elution fractions were collected, and concentration was determined *via* optical density at 280 nm.

Gα_i3 Expression and Purification

E. coli BL21 codon + (DE3) cells harboring the pET-28a (+) Gα_i3 plasmid were grown in M9 minimal media at 220 rpm and 37°C until an OD₆₀₀ of ~1.0 was reached. The cultures were cooled to 20°C and induced with 0.5 mM IPTG for 24 h. Cell cultures were then pelleted at 5,500 rpm for 10 min and stored at -80°C for future use.

Cell pellets were resuspended in Gα_i3 wash buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 10 mM Imidazole, 10 μM GDP, 2.5 mM MgCl₂, and 0.02% NaN₃) supplemented with 0.2 mg/ml lysozyme, 0.2 mg/ml RNase A, 1 mM PMSF, 1 mM benzamidine, and 1 PIERCE EDTA-free protease inhibitor tablet per 6 g of cells. Cells were lysed *via* triplicate passage through a homogenizer. Cell debris was centrifuged at 17,000 rpm for 30 min, followed by clarification of the lysate using a 0.22-μm PES bottle top filter. The

clarified lysate was loaded onto an NAC preequilibrated in wash buffer, washed five times with the same, and eluted with five column volumes of elution buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 250 mM Imidazole, 10 μM GDP, 2.5 mM MgCl₂, and 0.02% NaN₃). The elution peak was then loaded onto a HiLoad 16/600 Superdex Prep grade 75 column equilibrated in Gα_i3 exchange buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 10 μM GDP, 2.5 mM MgCl₂, and 0.02% NaN₃). The elution fractions were collected, and concentration was determined *via* optical density at 280 nm.

SDS-PAGE Analysis

Samples were combined at a ratio of 1:1 with 2x Laemmli buffer (20% glycerol, 125 mM Tris HCl pH 6.8, 4% sodium dodecyl sulfate (SDS), 0.02% bromophenol blue) for denaturation (Laemmli, 1970). Samples were then loaded into a Mini-PROTEAN TGX precast any-kD 10-well gel (Bio-Rad) with a Precision Plus Dual Standard protein ladder (Bio-Rad). The gel was run in running buffer (192 mM glycine, 25 mM, 0.1% SDS) for 53 min to remove loading dye at 400 mA and 150 V on a PowerPac Basic module (Bio-Rad). The gel was removed from the casing and stained in staining buffer (20% methanol and 10% acetic acid with 1 mg/ml Coomassie R250) with orbital rotation at 69 rpm until the gel was no longer visible. The gel was then destained in destaining buffer (20% methanol and 10% acetic acid).

Circular Dichroism Analysis

CCR3 was buffer-exchanged into CCR3 CD buffer (5 mM Tris-HCl, pH 8.0, 50 mM sodium acetate, 1 mM DDM, and 0.1 mM CHS) and concentrated to 0.3 mg/ml (7.2 μM). Immediately prior to analysis, buffer and CCR3-containing samples were diluted 4x to allow for data acquisition. Spectra were acquired using a J-815 CD spectrophotometer (JASCO Co., Easton, MD, United States) at 22°C in the spectral range of 180–260 nm at a rate of 1 nm/sec and a path length of 0.1 cm. 10 spectra each of the diluted buffer blank and CCR3 samples were recorded and averaged. Background spectra of the buffer were acquired identically and subtracted from the experimental data. Spectral fitting and secondary structure analysis for CCR3 were carried out using the DichroWeb (Whitmore and Wallace, 2004; Whitmore and Wallace, 2008) server using the K2D algorithm (Andrade et al., 1993). Presented secondary structure percentages from experimental data were calculated using DichroWeb analysis and compared to a model of CCR3. This model was generated by submitting the full-length CCR3 sequence to the Baker laboratory ROBETTA comparative modeling server (Song et al., 2013), using the CCR5 crystal structure PDB 4MBS (Tan et al., 2013a) as the template because of the high sequence similarity. The model was truncated to residues 23–317, and all truncated residues and the GGGGS4x linker were assumed to be random coils for percent secondary structure calculations. In addition, the CCL11 structure 1EOT (Crump et al., 1998) was docked to the CCR3 model using HADDOCK (van Zundert et al., 2016) to visualize the CCL11-bound CCR3 model observed throughout this work. Residue–residue interaction restraints upon drive docking were derived from

information available at the GPCR Database (Duchesnes et al., 2006; Millard et al., 2014; Pandey-Szekeres et al., 2018). CCL11 was buffer-exchanged into 10 mM sodium phosphate buffer, pH 7.5. Data were acquired at a concentration of 18.14 μ M and processed in the same manner as CCR3. Bestsel was used to fit CCL11 spectral data and predict the percent secondary structure from the experimental data to compare to the published structure 1EOT (Crump et al., 1998; Micsonai et al., 2015; Micsonai et al., 2018; Micsonai et al., 2021). $G\alpha_3$ was buffer-exchanged into 10 mM sodium phosphate buffer, pH 7.5, and 10 μ M GDP. Spectra were acquired in the same way as before at a protein concentration of 1.85 μ M. Data were fit using DichroWeb with the Contin-LL algorithm (Provencher and Glöckner, 1981) and reference set 4 (Sreerama and Woody, 2000). The extrapolated secondary structure from the experimental fit was compared to the crystal structure of activated $G\alpha_3$ in complex with RGS10 (PDB ID 2IHB) (Soundararajan et al., 2008). All non-crystallizing residues were assumed to be random coils for percent secondary structure calculations.

CCR3 Reconstitution and Formation of Polymer Discs

1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC or PC, Avanti Polar Lipids) and cholesterol (Sigma) were solvated in chloroform at 10 mg/ml, and then PC-only and 10, 20, 30, 40, and 60% cholesterol (mol%) mixtures were formed, blown down under a N_2 stream, and dried overnight *in vacuo* to produce a lipid film. Dried films were evenly divided (one aliquot for a protein-free control) and then solvated in non-detergent buffer (NDB, 20 mM HEPES pH 8.0, 150 mM KCl, 0.02% $NaNO_3$, and 1 mM EDTA) supplemented with 25 mM 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS, Anatrace) using mild sonication at 5 mg lipid/ml. Solvated films were then set on the benchtop for 3–5 h before the addition of protein, added at a ratio of 1 mg of protein per 4 mg of lipids. An equal volume of exchange buffer was added to protein-free (PF) samples. For samples reconstituted in the presence of CCL11, the agonist was added at a molar ratio of 5:1 CCL11:CCR3, or an equal volume of CCL11 exchange buffer was added to control samples. All samples were set to anneal for 3 h on the benchtop, during which Bio-Bead SM-2 Resin (Bio-Rad) was prepared by 3x degassing washes with methanol, followed by 3x washes with DI water and resuspension in NDB. A double portion of Bio-Beads (~60 mg) was added to each sample before nutation at room temperature. Samples were nutated in this way for 48–72 h, with ~30 mg Bio-Beads being added twice daily until the detergent was completely removed, evidenced by increased turbidity and loss of detergent bubbles upon manual agitation. Bio-Beads were removed by centrifugation at 500 rpm using PIERCE columns to collect fully formed proteoliposomes. To form lipid particles (SMALPs) from proteoliposomal samples, 3:1 pre-hydrolyzed styrene:maleic acid (SMA) was added at 3 mg SMA per 1 mg of lipids dropwise with inversion to each sample to facilitate polymer disc formation (Lipodisq, Thermo Fisher). Proteoliposomal samples typically turned clear within

moments of SMA addition. All samples were nutated overnight to ensure SMALP formation.

Fluorescence Polarization Assays

CCL11 in CCL11 exchange buffer was concentrated to >2 mg/ml and incubated with a 4x molar excess of fluorescein isothiocyanate (FITC, Invitrogen) for 1 h at room temperature. This was achieved in the dark, using an orbital shaker at 150 rpm and a pH value of 7.5 to facilitate labeling of the N terminus with the fluorophore. Following this, the sample was diluted to 1–2 ml with the same buffer and exchanged back into CCL11 exchange buffer sans excess FITC using a HiPrep 26/10 desalting column. A Lowry assay (DC protein assay, Bio-Rad) was performed to gauge protein concentrations. Fluorescence polarization assays were performed at room temperature using a Biotek Synergy NEO2 fluorescent plate reader equipped with a fluorescein filter (Biotek Instruments, Winooski, VT, United States). Fluorescence polarization was calculated automatically using the instrument as follows (Rossi and Taylor, 2011):

$$P = \frac{I_{||} - I_{\perp}}{I_{||} + I_{\perp}}$$

where $I_{||}$ is the observed parallel intensity, I_{\perp} is the observed perpendicular intensity, and P is polarization. Static concentrations of 100 nM FITC-CCL11 and 0.1 μ g/ μ l bovine serum albumin (BSA, Thermo Scientific) for nonspecific binding were added to each assay well. CCR3 was added to the desired concentration. Protein-free SMALPs or proteoliposomes were then added to balance out the lipid/SMA material such that the concentrations were equivalent across all wells. NDB was then added to fill to 30 μ l. The concentration of FITC-CCL11, BSA, and lipids/SMA were static across all conditions and replicates. Data were normalized by subtracting the lowest zero-point (no CCR3) value in a curve from each read to bring all curves within the same reference frame.

GTP Hydrolysis Assays

$G\alpha_3$ hydrolyzes GTP when bound to and activated by CCR3. Unhydrolyzed GTP is enzymatically converted to ATP and then to luminescence *via* luciferase. GTP turnover was thus quantified using a modified protocol of the GTPase-Glo™ assay (Promega) (Mondal et al., 2015) at room temperature for all steps, with a reaction incubation time of 2 h in all cases. After incubation, reconstituted GTPase-Glo™ reagent was added and incubated for 30 min at room temperature. Detection reagent was added, followed by an additional 5–10 min of incubation. Luminescence was read using a Cytation 3 multimode reader (Biotek Instruments, Winooski, VT, United States). Intrinsic GTPase activity of $G\alpha_3$ was analyzed using 2.5 μ M apo- $G\alpha_3$ in 20 mM HEPES (pH 7.5), 150 mM KCl, 5 mM $MgCl_2$, 20 mM EDTA, 0.1 mM TCEP, 10 μ M GDP, and 1 or 4 μ M GTP. Preliminary CCL11-induced CCR3 activation of $G\alpha_3$ was analyzed in assay buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 1 mM EDTA, 0.1 mM TCEP, 10 μ M GDP, and 4 μ M GTP) containing 5 mM $MgCl_2$, 1 mM DDM, and 0.1 mM CHS. Agonist-driven GTPase activity was analyzed in SMALPs formed from POPC with 0–30% cholesterol in assay buffer with

no additives. CCR3 samples reconstituted in the presence of CCL11 were analyzed in assay buffer plus 5 mM MgCl₂. Assays in detergent, SMALPs, and proteoliposomes were performed with 5 μM CCL11, 1 μM CCR3, and 1 μM Gα_i3. Relative light units (RLUs) of all assay runs were blank corrected by subtracting the average of three blank replicates (buffer with no GTP and background luminescence) from each replicate. % Hydrolysis was calculated from raw data as follows:

$$\% \text{ GTP Hydrolysis} = \frac{\text{RLU}_{\text{std}} - \text{RLU}_{\text{replicate}}}{\text{RLU}_{\text{std}}} \times 100$$

RESULTS

Codon Harmonization and Maltose-Binding Protein Facilitate Heterologous Protein Production

The total yield of functional, folded protein is the main bottleneck in the study of GPCRs. Here, we introduced a maltose-binding protein (MBP) fusion tag to aid in protein folding and solubility, as previously described (Ge et al., 2015). Furthermore, we employed codon harmonization, a method to optimize heterologous plasmid DNA sequences (Angov et al., 2008). We previously showed our codon harmonization strategy, dubbed DNA codon usage for measured base optimization, or DUMB optimization (DO), which dramatically increased both the yield and activity of an exogenously expressed chimeric membrane protein (van Aalst et al., 2020). Results of % MinMax analysis of proteins in this work show the codon usage of the fully optimized and codon-harmonized constructs as compared to native usage (Figure 3). Codon usage is presented as a sliding window of 21 codons to visualize the extent of optimization or deoptimization of each gene sequence to compare native usage (blue) to a random reverse transcription control (orange), FO constructs (purple), and DO constructs (red). These techniques facilitate the production of 2.6 ± 0.2 mg/L ($n = 3, \pm \sigma$) of full length and functional WT CCR3 (Figure 2A), after proteolytic cleavage of the MBP tag, from M9 minimal media. It is expected that an even greater yield would be observed if the expression cultures were grown in rich media. CCL11 and Gα_i3 were DUMB optimized as a matter of course in workflow optimization and produced yields of $\sim 2.6 \pm 0.3$ and 15.1 ± 0.3 mg/L, respectively, from minimal media.

Characterization of CCR3, CCL11, and Gα_i3 Secondary Structure

An apparent band slightly above the 37-kDa marker is visible in the SDS-PAGE analysis of CCR3, consistent with the predicted molecular weight of 42.4 kDa with the linker sequence in our construct (Figure 4A) after MBP cleavage and size exclusion chromatography (SEC) elution (Supplementary Figure S1A). This band conforms to previously reported SDS-PAGE analysis of similar CCR3 constructs in which CCR3 was observed to form

SDS-resistant dimers (Wang et al., 2013; Ge et al., 2015). This could explain the faint bands observed at roughly 80, 120, and 160 kDa. Full gel images are available in the **Supplementary Material (Supplementary Figure S2)**. Circular dichroism (CD) analysis reveals deep wells at 208 and 220 nm, consistent with a highly α-helical protein (Figure 4B). Experimental CD data were fit using DichroWeb (Whitmore and Wallace, 2004; Whitmore and Wallace, 2008). The K2D algorithm (Andrade et al., 1993) was selected over the Contin-LL algorithm (Provencher and Glöckner, 1981) paired with the SMP180 reference set (Abdul-Gader et al., 2011) due to differences in the normalized RMSD (NRMSD, 0.122 vs. 0.365, respectively). NRMSD values between 0.1 and 0.2 generally suggest similarity between experimental and fit secondary structures, whereas values greater than 0.2 indicate little resemblance (Mao et al., 1982). Predicted secondary structures derived from DichroWeb also show good agreement between experimental and model data when model truncated residues are assumed to be random coils (Figure 4C). Interestingly, the small observed β-sheet amount in the model (Figure 2A), derived from the CCR5 crystal structure template (Tan et al., 2013b), is conserved in the experimental CD spectrum according to the fitting.

CCL11 (Figure 2B) is also produced as an MBP fusion construct in *E. coli* to facilitate formation of the structural disulfide bonds in the more oxidizing periplasmic space. While this has a detrimental effect on yield, it consistently produces properly folded protein. An apparent band of ~ 14 kDa is observed for our CCL11 construct after MBP tag removal and subsequent SEC elution (Figure 4D; Supplementary Figure S1B). This is larger than the expected 8.4 kDa but is consistent with previously reported CCL11 constructs (Mingqing Wang, 2014). CD analysis of our construct confirms the conservation of the typical chemokine fold (Figure 4E, NRMSD 0.01858). The secondary structure distributions predicted from the fit using Bestsel (Micsonai et al., 2015; Micsonai et al., 2018; Micsonai et al., 2021) are consistent with the solution state NMR structure 1EOT (Crump et al., 1998) (Figure 4F).

The Gα_i3 subunit (Figure 2C) is expressed as an N terminal His-tagged construct. The theoretical molecular weight of our construct after SEC elution (Supplementary Figure S1C) is ~ 42.7 kDa, consistent with the apparent SDS-PAGE band (Figure 4G). The CD spectrum was fit using DichroWeb using the Contin-LL algorithm (Provencher and Glöckner, 1981) and spectral reference set 4 (Sreerama and Woody, 2000) (Figure 4H, NRMSD 0.045). Comparison of the DichroWeb fit-derived secondary structure to the crystal structure 2IHB (Soundararajan et al., 2008) shows relatively good agreement when accounting for possible minute differences between the inactive, GDP-bound state and the active state found in the crystal structure (Figure 4I).

Fluorescence Polarization Reveals Cholesterol-Induced Modulation of CCR3–CCL11 Affinity

GPCR structures often contain cholesterol binding motifs, but it is unclear how bilayer cholesterol impacts chemokine receptor

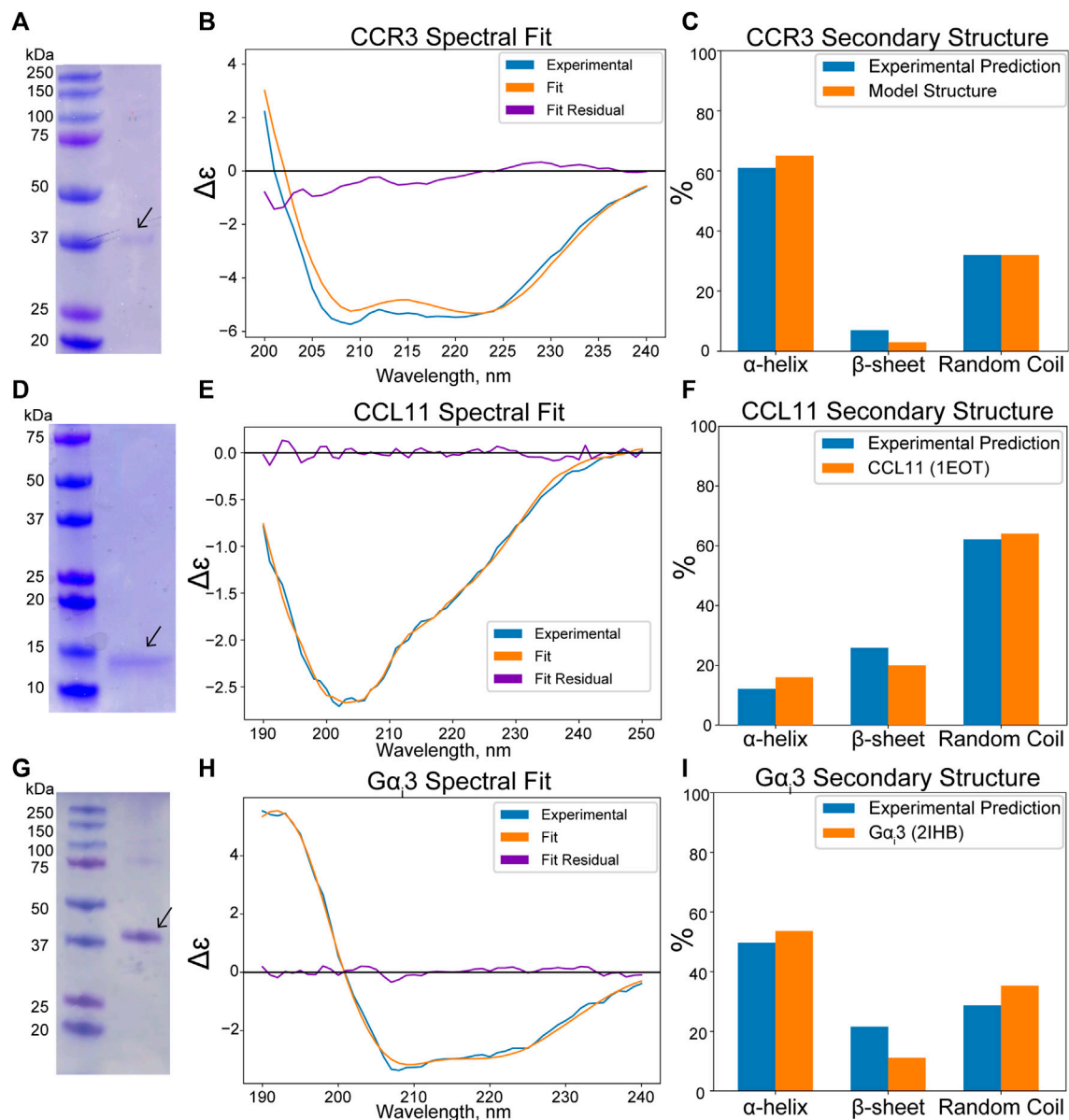


FIGURE 4 | (Wavelength, nm x axis label in panel H) was cut off. **(A)** SDS-PAGE analysis of CCR3 reveals a ~37-kDa band, consistent with the predicted molecular weight of 42.4 kDa. **(B)** Experimental CD spectra of CCR3 (blue) compared to the DichroWeb (Whitmore and Wallace, 2004) experimental fit (orange) and the residual fit (purple). **(C)** Quantitative assessment compared to the CCR3 model suggests a high degree of helical content. **(D)** SDS-PAGE analysis of CCL11 reveals an apparent molecular weight of ~14 kDa, consistent with previous observations. **(E)** CD spectra of CCL11 (blue) compared to the fit (orange) and the residual fit (purple). The spectrum was fit using Bestsel (Micsonai et al., 2015; Micsonai et al., 2018; Micsonai et al., 2021). **(F)** Bestsel-predicted percent secondary structure for the CCL11 construct compared to the NMR structure (PDB ID 1EOT) (Crump et al., 1998). **(G)** SDS-PAGE analysis of Gα3 shows good agreement with the predicted construct molecular weight of 42.7 kDa. **(H)** CD spectra of Gα3 (blue) compared to the DichroWeb experimental fit (orange) and the residual fit (purple). **(I)** Comparison of the experimentally predicted % secondary structure to that found in the crystal structure (PDB ID 2IHB) (Soundararajan et al., 2008).

function. Thus, we employed fluorescence polarization to measure the affinity of CCR3 for CCL11 as a function of bilayer cholesterol concentration. We attached the fluorophore fluorescein isothiocyanate (FITC) to the *N* terminus of CCL11, which unbound is rapidly tumbling in solution, polarizing light to a lower extent (**Figure 5A**). After CCL11 is bound, the decreased

molecular tumbling increases polarization. Titration of cholesterol from 0 to 30% (mol%) into 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC and PC) membrane styrene:maleic acid lipid particles (SMALPs) shows a drastic decrease in K_d (**Figure 5B**). A K_d of 30 ± 10 nM at 30% cholesterol, or a ~5-fold decrease from pure POPC, is quite significant.

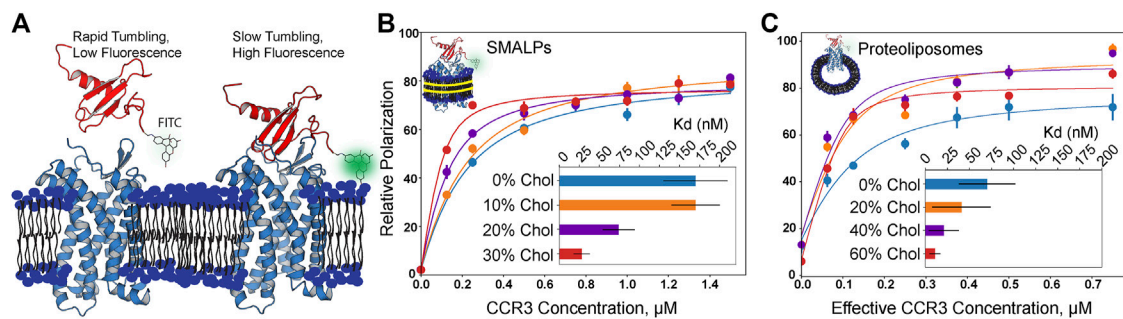


FIGURE 5 | Fluorescence polarization assay results in increasing cholesterol content (mol% in PC). **(A)** Visualization of apo- and CCR3-bound CCL11. FITC-labeled CCL11 is rapidly tumbling in solution, and thus, the probe polarizes light to a lesser degree. Upon binding CCR3, tumbling is slowed, polarizing a greater fraction of light and giving rise to a larger polarization value. CCL11 (PDB ID 1EOT) (Crump et al., 1998) was docked to CCR3 using HADDOCK (van Zundert et al., 2016) to visualize the bound state leading to slower tumbling. **(B)** Membrane cholesterol content is positively correlated with increased ligand affinity in SMALPs. **(C)** Increased observed affinity is conserved in proteoliposomes. Concentration is halved to correct for receptor orientation (**Supplementary Figure S3**). Points indicate the mean \pm SEM for three replicates, each read three times. Error bars may fall within the size of the points. Data were normalized by subtracting the lowest zero-point (no CCR3) value in a curve from each read to bring all curves within the same reference frame.

In order to gauge the effects of cholesterol at higher concentrations, we turned to proteoliposomes, as the membrane rigidity imparted by higher cholesterol content inhibits SMALP formation (Scheidelaar et al., 2015; Dörr et al., 2016) (**Figure 5C**). It is clear that the same trend is observed; however, the experimental error increased for the measurements in proteoliposomes. We corrected the effective CCR3 concentration to account for receptor orientation intractability by halving the total concentration to remove the statistical average of the receptor with the orthosteric site that is facing into the proteoliposome and thus inaccessible to the ligand. The listed receptor concentrations in proteoliposomes reflect the input concentrations that were halved during curve fitting (**Figure 5C**; **Supplementary Figure S3**). Measured K_d values were 20 ± 20 nM in 40% cholesterol and 11 ± 6 nM in 60% cholesterol, following the same inverse trend between K_d and cholesterol content as that seen in SMALPs. However, the K_d values at the same cholesterol concentrations were also lower than the corresponding measurement in SMALPs. Membrane curvature and lateral pressure in the proteoliposomal samples may play a role in GPCR function (Jones et al., 2020). We hypothesize both curvature and lateral pressure are lost in SMALPs, which could contribute to the observed K_d discrepancies between the two conditions. Increasing cholesterol presence did still lead to decreasing K_d in SMALPs, suggesting that curvature and lateral pressure are not the only phenomena responsible for the observed modulation of ligand affinity. Thus, we conclude that cholesterol is a direct positive allosteric effector of CCR3–ligand affinity.

The Extent of $G\alpha_3$ Activation Is Cholesterol Dose Dependent

We next investigated how the observed relationship between cholesterol and CCL11 affinity impacts G protein coupling to CCR3. Heightened activation of CCR3 should lead to increased receptor–G protein coupling. This, in turn, should drive

nucleotide exchange in the $G\alpha_3$ binding pocket with CCR3 functioning as the guanine nucleotide exchange factor (GEF). The remaining GTP is then converted to a luminescent signal after conversion to ATP (**Figure 6A**) (Ford et al., 1994; Ford and Leach, 1998a; Ford and Leach, 1998b). In the absence of a suitable GEF, 20 mM EDTA can be used to stimulate intrinsic activity, verifying construct activity (**Supplementary Figure S4**). We next verified that our CCR3 and $G\alpha_3$ constructs are able to couple in detergent, using Mg^{2+} to stabilize the nucleotide-bound state of $G\alpha_3$ (**Supplementary Figure S5**). We considered this an important step as little information is available concerning which $G\alpha$ subunits CCR3 activates. Upon showing the ability for our constructs to couple, we investigated the effects of cholesterol on this interaction, hypothesizing that the dose-dependent modulation of agonist affinity would directly translate to $G\alpha_3$ coupling, activation, and GTP hydrolysis.

We first attempted to analyze CCL11-driven GTPase activity of $G\alpha_3$ in a SMALP environment as both the extracellular agonist orthosteric pocket and the intracellular G protein docking site would be solvent exposed (**Figure 6B**; **Supplementary Figure S6A**). While it is clear that cholesterol increases the extent of hydrolysis, dose dependence could not be verified, and the formation and integrity of the cholesterol-containing SMALPs is in question. This is likely due to the rigidity imparted by increasing the cholesterol concentration, which is documented to inhibit copolymer intercalation within the membrane (Scheidelaar et al., 2015; Dörr et al., 2016). Unlike in the ligand-binding assay, CCR3 immobilized in SMA-resistant cholesterol proteoliposomes is unable to function in this assay regardless of orientation. This may then explain the puzzling decrease in GTPase activity between 20 and 30% datasets. Furthermore, SMA has a relatively narrow range of compatibility with common biochemical assay components such as divalent cations. Thus, CCL11-driven GTPase activity of $G\alpha_3$ in SMALPs could not be verified in the presence of Mg^{2+} , which also adversely influences $G\alpha_3$ GTPase activity (Mondal et al., 2015).

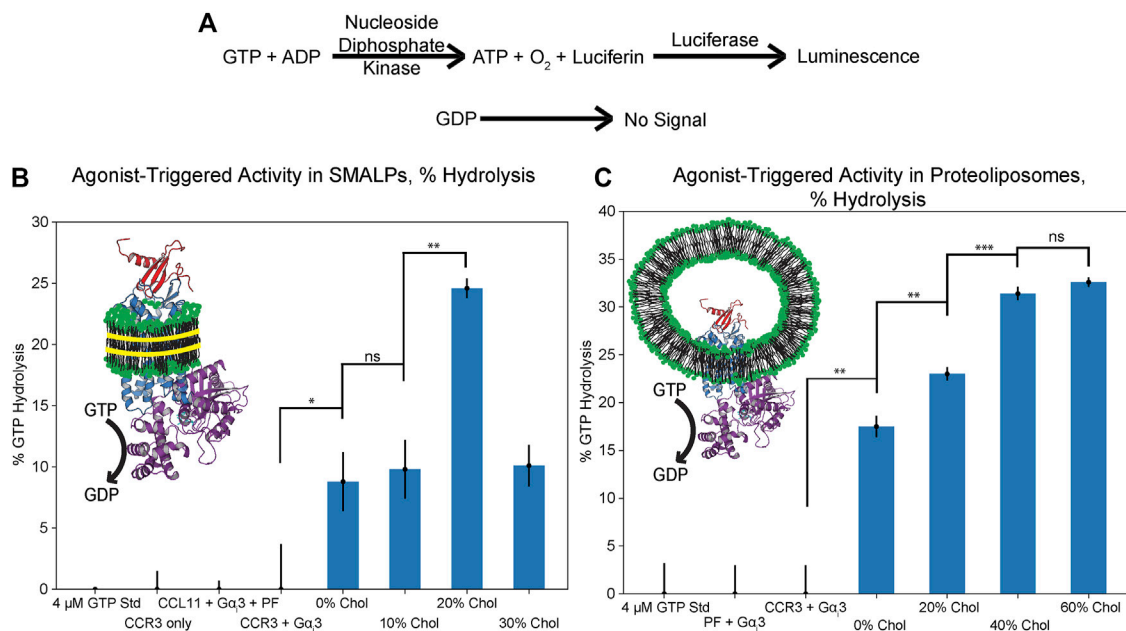


FIGURE 6 | GTPase assay results in SMALPs and proteoliposomes of increasing cholesterol content (mol% in PC), presented as % GTP hydrolyzed over the course of 2 h. **(A)** Schematic representation of GTP hydrolysis translation to the luminescent signal (Mondal et al., 2015). **(B)** Cholesterol enhances GTP hydrolysis in SMALPs but dose dependence could not be verified, likely due to inhibition of disc formation. **(C)** Cholesterol dose dependence of GTP hydrolysis is verified through coreconstitution of CCL11 with CCR3 into proteoliposomes. Cartoons depict CCL11 (red, 1EOT) (Crump et al., 1998) bound to CCR3 (blue) reconstituted in lipids (green headgroups) formed into either SMALPs with SMA (left, yellow) or proteoliposomes (right) with Gα₃ (purple, 2ODE) (Soundararajan et al., 2008) bound to CCR3. Points indicate the mean ± standard deviation for three replicates. Bars indicate statistical significance based on the Student's *t*-test: **p* < 0.05, ***p* < 0.01, ****p* < 0.001, ns is not statistically significant.

To overcome this challenge, we reconstituted CCR3 in the presence of the peptide agonist. CCR3 was reconstituted into proteoliposomes with 0, 20, 40, and 60% cholesterol content with CCL11 present in the buffer at a molar ratio of 1:5 CCR3:CCL11. The GTPase assay was repeated in the presence of 5 mM Mg²⁺ (Figure 6C; Supplementary Figure S6B). Statistical significance between the GTP hydrolysis signal in 0, 20, and 40% cholesterol is clear evidence of cholesterol dose-dependent modulation of function. Furthermore, GTPase activity in 40 vs. 60% cholesterol proteoliposomes is approximately equivalent, within error, which we also observed in the ligand-binding assay. Together, this corroborates our hypothesis that the dose-dependent modulation of agonist affinity directly translates to receptor activation, Gα₃ coupling, GTP hydrolysis, and thus signal transduction.

DISCUSSION

Typically, the total exogenous yield of functional GPCRs confounds functional studies. Thus, GPCR production continues to be an area of innovation (Abiko et al., 2021; Mulry et al., 2021). Here, we implemented two techniques, codon harmonization and an MBP fusion tag, to facilitate the production of a functional human GPCR from *E. coli*. Codon harmonization is underutilized ostensibly due to varying degrees of experimental success. This is likely a result of the sheer number

of variables that contribute to protein expression and folding (Quax et al., 2015). The use of an MBP fusion tag has also gained popularity, and successful implementation has been observed in a few cases (Bertin et al., 1992; Yeliseev et al., 2007; Serrano-Vega et al., 2008; Egloff et al., 2014; Beckner et al., 2020). Lack of widespread adoption of MBP tags for heterologous GPCR expression is, in our opinion, due to the difficulty of tag removal, as the TEV is known to be inhibited by detergents (Mohanty et al., 2003). This study represents an important step in the successful implementation of a dual heterologous expression strategy that we hypothesize will be of use both in our own future experiments and the wider GPCR structural biology community.

It was shown that the titration of membrane cholesterol increases the affinity of CCR3 for its endogenous ligand CCL11 and that this observation directly translates to agonist-driven Gα₃ GTP hydrolysis as a proxy for signal transduction. We hypothesize that the functional interplay between cholesterol, ligand affinity, and G protein docking is due to cholesterol-driven conformational sampling inhibition. Such observations were made previously, where cholesterol was hypothesized to constrain GPCR conformational selection to states with higher ligand affinity (Weis and Kobilka, 2018). Cholesterol's influence on chemokine receptor ligand affinity has, however, been shown sparingly (Babcock et al., 2003; Calmet et al., 2020) and typically not dose-dependently. Such observations are likely conserved in chemokine receptors; therefore, we conclude that cholesterol enhances CCR3 ligand binding through a direct allosteric

mechanism that is directly translated to $G\alpha_i3$ coupling and GTP hydrolysis. Thus, our data indicate that cholesterol is a positive allosteric modulator of CCR3 signal transduction.

Although similar fluorescence experiments have been reported (Kawamura et al., 2014; Purvanov et al., 2018; Matti et al., 2020), the use of a recombinant, fluorescently labeled endogenous ligand is nontrivial and may function as a useful tool in future experiments involving *in vitro* mimicry of native biological interactions. Cholesterol as a modulator of agonist-driven GTP hydrolysis of a chemokine receptor is novel to this work. Implementation of SMALP technology is also nontrivial, although the buffer component and lipid incompatibilities still leave room for improvement. Zwitterionic SMALP polymers have broader compatibility with common biochemical reagents and may serve as a starting place in future works (Fiori et al., 2020). This may be the reason this GTPase assay is not more widely adopted in lipid environments, although some experimental evidence is adopted in high-density lipoprotein particles (nanodiscs) (Staus et al., 2019; Huang et al., 2021), and mixed micelles (Gregorio et al., 2017) are evident.

We have shown that $G\alpha_i3$ -CCR3 coupling is possible *in vitro*, but what role this might play *in vivo* is open to speculation. The main cellular function of $G\alpha_i$ subunits is as intracellular Ca^{2+} ion concentration and adenylate cyclase effectors (Peleg et al., 2002). However, an additional function as modulators of G protein-gated inwardly rectifying K^+ channels (GIRKs) is well documented (Peleg et al., 2002; Rubinstein et al., 2007). $G\alpha_i3$, specifically, was originally named G_k due to its role in the stimulation of GIRK function (Codina et al., 1987; Codina et al., 1988). A mounting body of evidence now suggests some interplay between chemokine expression and nervous system physiology, with CXCL12, the predominant endogenous ligand of CXCR4, strongly implicated in this phenomenon (Guyon, 2014). Further evidence implicates CXCL12 and CCL5 in the activation of GIRK functionality *in vivo* (Picciocchi et al., 2014), the latter of which is a ligand for CCR3 (Zhang et al., 2015). Thus, there is precedence for chemokine receptor-driven modulation of GIRK functionality. Although predominantly expressed by eosinophils, studies have identified expression of CCR3 in neurons, astrocytes, and microglia (Cho and Miller, 2002; Banisadr et al., 2005). It has been implicated in the pathogenesis of neuroinflammatory conditions such as Parkinson's disease (Moghadam-Ahmadi et al., 2020) and Alzheimer's disease (Sui et al., 2019), as well as multiple sclerosis and HIV-associated dementia (Banisadr et al., 2005). Of note, GIRK channels play a direct role in Parkinson's pathophysiology (Mayfield et al., 2015), and upregulation of

serum CCL5 is correlated with disease severity (Tang et al., 2014). Typically, upregulation of CCL5 is related to CCR5 function, but there may be a role for CCL5 and other chemokines triggering CCR3- $G\alpha_i3$ coupling in neurodegenerative disorders that is certainly worth exploring.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

EA designed and performed all experiments in coordination with BW. EA drafted the article. EA and BW contributed to subsequent revisions.

FUNDING

This work was funded by the NIH grant 1R35GM124979 (Maximizing Investigators' Research Award (MIRA) R35) awarded to BW.

ACKNOWLEDGMENTS

We would like to thank Matthew Dominguez, Isaac Scott, and Bryan Sutton (TTU Health Sciences Center) for aid in the use of a spectrometer to acquire the CD spectra. We would like to thank Michael P. Latham (Texas Tech University) for the generous gift of the TEV protease plasmid and for the use of a fluorescence plate reader in the fluorescence polarization ligand-binding assay. We would like to thank Shane Scoggin and Naima Moustaid-Moussa (Texas Tech University) for providing the use of a luminescent plate reader in the GTP hydrolysis assay.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.724603/full#supplementary-material>

REFERENCES

- Abdul-Gader, A., Miles, A. J., and Wallace, B. A. (2011). A Reference Dataset for the Analyses of Membrane Protein Secondary Structures and Transmembrane Residues Using Circular Dichroism Spectroscopy. *Bioinformatics* 27, 1630–1636. doi:10.1093/bioinformatics/btr234
- Abiko, L. A., Rogowski, M., Gautier, A., Schertler, G., and Grzesiek, S. (2021). Efficient Production of a Functional G Protein-Coupled Receptor in *E. coli* for Structural Studies. *J. Biomol. NMR* 75, 25–38. doi:10.1007/s10858-020-00354-6
- Amani, R., Borcik, C. G., Khan, N. H., Versteeg, D. B., Yekefallah, M., Do, H. Q., et al. (2020). Conformational Changes upon Gating of KirBac1.1 into an Open-Activated State Revealed by Solid-State NMR and Functional Assays. *Proc. Natl. Acad. Sci. USA* 117, 2938–2947. doi:10.1073/pnas.1915010117
- Andrade, M. A., Chacón, P., Merelo, J. J., and Morán, F. (1993). Evaluation of Secondary Structure of Proteins from UV Circular Dichroism Spectra Using an Unsupervised Learning Neural Network. *Protein Eng. Des. Sel* 6, 383–390. doi:10.1093/protein/6.4.383
- Angov, E., Hillier, C. J., Kincaid, R. L., and Lyon, J. A. (2008). Heterologous Protein Expression Is Enhanced by Harmonizing the Codon Usage Frequencies of the

- Target Gene with Those of the Expression Host. *PLoS One* 3, e2189. doi:10.1371/journal.pone.0002189
- Babcock, G. J., Farzan, M., and Sodroski, J. (2003). Ligand-independent Dimerization of CXCR4, a Principal HIV-1 Coreceptor. *J. Biol. Chem.* 278, 3378–3385. doi:10.1074/jbc.M210140200
- Ballesteros, J. A., and Weinstein, H. (1995). [19] Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors. *Receptor Mol. Biol.*, 366–428. doi:10.1016/s1043-9471(05)80049-7
- Banisadr, G., Rostène, W., Kitabgi, P., and Parsadaniantz, S. (2005). Chemokines and Brain Functions. *Cdtia* 4, 387–399. doi:10.2174/1568010054022097
- Beckner, R. L., Zoubak, L., Hines, K. G., Gawrisch, K., and Yeliseev, A. A. (2020). Probing Thermostability of Detergent-Solubilized CB2 Receptor by Parallel G Protein-Activation and Ligand-Binding Assays. *J. Biol. Chem.* 295, 181–190. doi:10.1074/jbc.RA119.010696
- Bertin, B., Freissmuth, M., Breyer, R. M., Schütz, W., Strosberg, A. D., and Marullo, S. (1992). Functional Expression of the Human Serotonin 5-HT1A Receptor in *Escherichia coli*. Ligand Binding Properties and Interaction with Recombinant G Protein Alpha-Subunits. *J. Biol. Chem.* 267, 8200–8206. doi:10.1016/s0021-9258(18)42427-1
- Bhate, M. P., Wylie, B. J., Thompson, A., Tian, L., Nimigeon, C., and McDermott, A. E. (2013). Preparation of Uniformly Isotope Labeled KcsA for Solid State NMR: Expression, Purification, Reconstitution into Liposomes and Functional Assay. *Protein Expr. Purif.* 91, 119–124. doi:10.1016/j.pep.2013.07.013
- Borcik, C. G., Versteeg, D. B., Amani, R., Yekefallah, M., Khan, N. H., and Wylie, B. J. (2020). The Lipid Activation Mechanism of a Transmembrane Potassium Channel. *J. Am. Chem. Soc.* 142, 14102–14116. doi:10.1021/jacs.0c01991
- Borcik, C. G., Versteeg, D. B., and Wylie, B. J. (2019). An Inward-Rectifier Potassium Channel Coordinates the Properties of Biologically Derived Membranes. *Biophysical J.* 116, 1701–1718. doi:10.1016/j.bpj.2019.03.023
- Botelhol, A. V., Huber, T., Sakmar, T. P., and Brown, M. F. (2006). Curvature and Hydrophobic Forces Drive Oligomerization and Modulate Activity of Rhodopsin in Membranes. *Biophysical J.* 91, 4464–4477. doi:10.1529/biophysj.106.082776
- Calmet, P., Cullin, C., Cortès, S., Vang, M., Caudy, N., Baccouch, R., et al. (2020). Cholesterol Impacts Chemokine CCR5 Receptor Ligand-binding Activity. *FEBS J.* 287, 2367–2385. doi:10.1111/febs.15145
- Chattopadhyay, A. (2014). GPCRs: Lipid-dependent Membrane Receptors that Act as Drug Targets. *Adv. Biol.* 2014, 1–12. doi:10.1155/2014/143023
- Chen, X., Zaro, J. L., and Shen, W.-C. (2013). Fusion Protein Linkers: Property, Design and Functionality. *Adv. Drug Deliv. Rev.* 65, 1357–1369. doi:10.1016/j.addr.2012.09.039
- Cho, C., and Miller, R. J. (2002). Chemokine Receptors and Neural Function. *J. Neurovirol.* 8, 573–584. doi:10.1080/13550280290101003
- Codina, J., Olate, J., Abramowitz, J., Mattera, R., Cook, R. G., and Birnbaumer, L. (1988). Alpha I-3 cDNA Encodes the Alpha Subunit of G_k, the Stimulatory G Protein of Receptor-Regulated K⁺ Channels. *J. Biol. Chem.* 263, 6746–6750. doi:10.1016/s0021-9258(18)68706-x
- Codina, J., Yatani, A., Grenet, D., Brown, A., and Birnbaumer, L. (1987). The Alpha Subunit of the GTP Binding Protein G_k Opens Atrial Potassium Channels. *Science* 236, 442–445. doi:10.1126/science.2436299
- Crump, M. P., Rajarathnam, K., Kim, K.-S., Clark-Lewis, I., and Sykes, B. D. (1998). Solution Structure of Eotaxin, a Chemokine that Selectively Recruits Eosinophils in Allergic Inflammation. *J. Biol. Chem.* 273, 22471–22479. doi:10.1074/jbc.273.35.22471
- Dawaliby, R., Trubbia, C., Delporte, C., Masureel, M., Van Antwerpen, P., Kobilka, B. K., et al. (2016). Allosteric Regulation of G Protein-Coupled Receptor Activity by Phospholipids. *Nat. Chem. Biol.* 12, 35–39. doi:10.1038/nchembio.1960
- Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhardt, M., et al. (1996). Identification of a Major Co-receptor for Primary Isolates of HIV-1. *Nature* 381, 661–666. doi:10.1038/381661a0
- Dörr, J. M., Scheidehaar, S., Koorengevel, M. C., Dominguez, J. J., Schäfer, M., van Walree, C. A., et al. (2016). The Styrene-Maleic Acid Copolymer: a Versatile Tool in Membrane Research. *Eur. Biophys. J.* 45, 3–21. doi:10.1007/s00249-015-1093-y
- Duchesnes, C. E., Murphy, P. M., Williams, T. J., and Pease, J. E. (2006). Alanine Scanning Mutagenesis of the Chemokine Receptor CCR3 Reveals Distinct Extracellular Residues Involved in Recognition of the Eotaxin Family of Chemokines. *Mol. Immunol.* 43, 1221–1231. doi:10.1016/j.molimm.2005.07.015
- Dunn, J. L. M., Shoda, T., Caldwell, J. M., Wen, T., Aceves, S. S., Collins, M. H., et al. (2020). Esophageal Type 2 Cytokine Expression Heterogeneity in Eosinophilic Esophagitis in a Multisite Cohort. *J. Allergy Clin. Immunol.* 145, 1629–1640.e1624. doi:10.1016/j.jaci.2020.01.051
- Egloff, P., Hillenbrand, M., Klenk, C., Batyuk, A., Heine, P., Balada, S., et al. (2014). Structure of Signaling-Competent Neurotensin Receptor 1 Obtained by Directed Evolution in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 111, E655–E662. doi:10.1073/pnas.1317903111
- Federman, A. D., Conklin, B. R., Schrader, K. A., Reed, R. R., and Bourne, H. R. (1992). Hormonal Stimulation of Adenylyl Cyclase through Gi-Protein $\beta\gamma$ Subunits. *Nature* 356, 159–161. doi:10.1038/356159a0
- Fiori, M. C., Zheng, W., Kamilar, E., Simiyu, G., Altenberg, G. A., and Liang, H. (2020). Extraction and Reconstitution of Membrane Proteins into Lipid Nanodiscs Encased by Zwitterionic Styrene-Maleic Amide Copolymers. *Sci. Rep.* 10, 9940. doi:10.1038/s41598-020-66852-7
- Ford, S. R., and Leach, F. R. (1998a). Bioluminescent Assay of the Guanylates. *Methods Mol. Biol.* 102, 55–68. doi:10.1385/0-89603-520-4:55
- Ford, S. R., and Leach, F. R. (1998b). Improvements in the Application of Firefly Luciferase Assays. *Methods Mol. Biol.* 102, 3–20. doi:10.1385/0-89603-520-4:3
- Ford, S. R., Vaden, V. R., Booth, J. L., Hall, M. S., Webster, J. J., and Leach, F. R. (1994). Bioluminescent Determination of 0.1 Picomole Amounts of Guanine Nucleotides. *J. Biolumin. Chemilumin.* 9, 251–265. doi:10.1002/bio.1170090403
- Fuhrmann, M., Hausherr, A., Ferbitz, L., Schödl, T., Heitzer, M., and Hegemann, P. (2004). Monitoring Dynamic Expression of Nuclear Genes in *Chlamydomonas reinhardtii* by Using a Synthetic Luciferase Reporter Gene. *Plant Mol. Biol.* 55, 869–881. doi:10.1007/s11103-004-2150-610.1007/s11103-005-2150-1
- Gahbauer, S., Pluhackova, K., and Böckmann, R. A. (2018). Closely Related, yet Unique: Distinct Homo- and Heterodimerization Patterns of G Protein Coupled Chemokine Receptors and Their fine-tuning by Cholesterol. *Plos Comput. Biol.* 14, e1006062. doi:10.1371/journal.pcbi.1006062
- Gauvreau, G. M., FitzGerald, J. M., Boulet, L. P., Watson, R. M., Hui, L., Villeneuve, H., et al. (2018). The Effects of a CCR3 Inhibitor, AXP1275, on Allergen-Induced Airway Responses in Adults with Mild-To-Moderate Atopic Asthma. *Clin. Exp. Allergy* 48, 445–451. doi:10.1111/cea.13114
- Ge, B., Wang, M., Li, J., Liu, J., and Huang, F. (2015). Maltose Binding Protein Facilitates Functional Production of Engineered Human Chemokine Receptor 3 in *Escherichia coli*. *Process Biochem.* 50, 285–293. doi:10.1016/j.procbio.2014.12.001
- Gregorio, G. G., Masureel, M., Hilger, D., Terry, D. S., Juetter, M., Zhao, H., et al. (2017). Single-molecule Analysis of Ligand Efficacy in β 2AR-G-protein Activation. *Nature* 547, 68–73. doi:10.1038/nature22354
- Guyon, A. (2014). CXCL12 Chemokine and its Receptors as Major Players in the Interactions between Immune and Nervous Systems. *Front. Cel. Neurosci.* 8, 65. doi:10.3389/fncel.2014.00065
- Hanson, M. A., Cherezov, V., Griffith, M. T., Roth, C. B., Jaakola, V.-P., Chien, E. Y. T., et al. (2008). A Specific Cholesterol Binding Site Is Established by the 2.8 Å Structure of the Human β 2-Adrenergic Receptor. *Structure* 16, 897–905. doi:10.1016/j.str.2008.05.001
- He, J., Chen, Y., Farzan, M., Choe, H., Ohagen, A., Gartner, S., et al. (1997). CCR3 and CCR5 Are Co-receptors for HIV-1 Infection of Microglia. *Nature* 385, 645–649. doi:10.1038/385645a0
- Huang, S. K., Pandey, A., Tran, D. P., Villanueva, N. L., Kitao, A., Sunahara, R. K., et al. (2021). Delineating the Conformational Landscape of the Adenosine A2A Receptor during G Protein Coupling. *Cell* 184, 1884–1894. e1814. doi:10.1016/j.cell.2021.02.041
- Ishida, Y., Kido, A., Akahane, M., Kishi, S., Tsukamoto, S., Fujii, H., et al. (2018). Mesenchymal Stem Cells Up-Regulate the Invasive Potential of Prostate Cancer Cells via the eotaxin-3/CCR3 axis. *Pathol. - Res. Pract.* 214, 1297–1302. doi:10.1016/j.prp.2018.06.012
- Jafurulla, M., Aditya Kumar, G., Rao, B. D., and Chattopadhyay, A. (2019). A Critical Analysis of Molecular Mechanisms Underlying Membrane Cholesterol Sensitivity of GPCRs. *Cholesterol Modulation Protein Funct. Sterol Specificity Indirect Mech.* 1115, 21–52. doi:10.1007/978-3-030-04278-3_2
- Jin, L., Liu, W.-R., Tian, M.-X., Jiang, X.-F., Wang, H., Zhou, P.-Y., et al. (2017). CCL24 Contributes to HCC Malignancy via RhoB- VEGFA-VEGFR2

- Angiogenesis Pathway and Indicates Poor Prognosis. *Oncotarget* 8, 5135–5148. doi:10.18632/oncotarget.14095
- Jöhner, K., Zelle-Rieser, C., Perathoner, A., Moser, P., Hager, M., Ramoner, R., et al. (2005). Up-regulation of Functional Chemokine Receptor CCR3 in Human Renal Cell Carcinoma. *Clin. Cancer Res.* 11, 2459–2465. doi:10.1158/1078-0432.CCR-04-0405
- Jones, A. J. Y., Gabriel, F., Tandale, A., and Nietlispach, D. (2020). Structure and Dynamics of GPCRs in Lipid Membranes: Physical Principles and Experimental Approaches. *Molecules* 25, 4729. doi:10.3390/molecules25204729
- Kapust, R. B., Tözsér, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D., et al. (2001). Tobacco Etch Virus Protease: Mechanism of Autolysis and Rational Design of Stable Mutants with Wild-type Catalytic Proficiency. *Protein Eng.* 14, 993–1000. doi:10.1093/protein/14.12.993
- Katschke, K. J., Jr., Rottman, J. B., Ruth, J. H., Qin, S., Wu, L., LaRosa, G., et al. (2001). Differential Expression of Chemokine Receptors on Peripheral Blood, Synovial Fluid, and Synovial Tissue Monocytes/macrophages in Rheumatoid Arthritis. *Arthritis Rheum.*, 44. 2-N: CO, 1022–1032. doi:10.1002/1529-0131(200105)44:5<1022::AID-ANR181>3.0.CO;2
- Katz, A., Wu, D., and Simon, M. I. (1992). Subunits $\beta\gamma$ of Heterotrimeric G Protein Activate β_2 Isoform of Phospholipase C. *Nature* 360, 686–689. doi:10.1038/360686a0
- Kawamura, T., Stephens, B., Qin, L., Yin, X., Dores, M. R., Smith, T. H., et al. (2014). A General Method for Site Specific Fluorescent Labeling of Recombinant Chemokines. *PLoS One* 9, e81454. doi:10.1371/journal.pone.0081454
- Kim, T. H., Chung, K. Y., Manglik, A., Hansen, A. L., Dror, R. O., Mildorf, T. J., et al. (2013). The Role of Ligands on the Equilibria between Functional States of a G Protein-Coupled Receptor. *J. Am. Chem. Soc.* 135, 9465–9474. doi:10.1021/ja404305k
- Kitaura, M., Nakajima, T., Imai, T., Harada, S., Combadiere, C., Tiffany, H. L., et al. (1996). Molecular Cloning of Human Eotaxin, an Eosinophil-Selective CC Chemokine, and Identification of a Specific Eosinophil Eotaxin Receptor, CC Chemokine Receptor 3. *J. Biol. Chem.* 271, 7725–7730. doi:10.1074/jbc.271.13.7725
- Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* 227, 680–685. doi:10.1038/227680a0
- Legler, D. F., Matti, C., Laufer, J. M., Jakobs, B. D., Purvanov, V., Uetz-von Allmen, E., et al. (2017). Modulation of Chemokine Receptor Function by Cholesterol: New Prospects for Pharmacological Intervention. *Mol. Pharmacol.* 91, 331–338. doi:10.1124/mol.116.107151
- Ma, Q., Jones, D., Borghesani, P. R., Segal, R. A., Nagasawa, T., Kishimoto, T., et al. (1998). Impaired B-Lymphopoiesis, Myelopoiesis, and Derailed Cerebellar Neuron Migration in CXCR4- and SDF-1-Deficient Mice. *Pnas* 95, 9448–9453. doi:10.1073/pnas.95.16.9448
- Mao, D., Wachter, E., and Wallace, B. A. (1982). Folding of the Mitochondrial Proton Adenosine Triphosphatase Proteolipid Channel in Phospholipid Vesicles. *Biochemistry* 21, 4960–4968. doi:10.1021/bi00263a020
- Marlow, B., Kuenze, G., Li, B., Sanders, C. R., and Meiler, J. (2021). Structural Determinants of Cholesterol Recognition in Helical Integral Membrane Proteins. *Biophysical J.* 120, 1592–1604. doi:10.1016/j.bpj.2021.02.028
- Matti, C., D'Uonno, G., Artinger, M., Melgrati, S., Salnikov, A., Thelen, S., et al. (2020). CCL20 Is a Novel Ligand for the Scavenging Atypical Chemokine Receptor 4. *J. Leukoc. Biol.* 107, 1137–1154. doi:10.1002/JLB.2MA0420-295RRR
- Mayfield, J., Blednov, Y. A., and Harris, R. A. (2015). Behavioral and Genetic Evidence for GIRK Channels in the CNS. *Int. Rev. Neurobiol.* 123, 279–313. doi:10.1016/bs.irn.2015.05.016
- Micsonai, A., Bulyáki, É., and Kardos, J. (2021). BeStSel: From Secondary Structure Analysis to Protein Fold Prediction by Circular Dichroism Spectroscopy. *Methods Mol. Biol.* 2199, 175–189. doi:10.1007/978-1-0716-0892-0_11
- Micsonai, A., Wien, F., Bulyáki, É., Kun, J., Moussong, É., Lee, Y.-H., et al. (2018). BeStSel: a Web Server for Accurate Protein Secondary Structure Prediction and Fold Recognition from the Circular Dichroism Spectra. *Nucleic Acids Res.* 46, W315–W322. doi:10.1093/nar/gky497
- Micsonai, A., Wien, F., Kerynia, L., Lee, Y.-H., Goto, Y., Réfrégiers, M., et al. (2015). Accurate Secondary Structure Prediction and Fold Recognition for Circular Dichroism Spectroscopy. *Proc. Natl. Acad. Sci. USA* 112, E3095–E3103. doi:10.1073/pnas.1500851112
- Millard, C. J., Ludeman, J. P., Canals, M., Bridgford, J. L., Hinds, M. G., Clayton, D. J., et al. (2014). Structural Basis of Receptor Sulfotyrosine Recognition by a CC Chemokine: the N-Terminal Region of CCR3 Bound to CCL11/eotaxin-1. *Structure* 22, 1571–1581. doi:10.1016/j.str.2014.08.023
- Mingqing Wang, B. G., Yang, Qiuxia, Xiaoyong, Jiang, and Huang, Fang. (2014). High-level Production of Biologically Active Chemokines in Escherichia coli. 4 ed.
- Moghadam-Ahmadi, A., Khorramdelazad, H., Hassanshahi, G., Shahsavari, S., Moadab, A., and Vakilian, A. (2020). Eotaxins and C-C Chemokine Receptor Type 3 in Parkinson's Disease. *Acta Neurol. Belg.* 120, 589–594. doi:10.1007/s13760-018-01061-8
- Mohanty, A. K., Simmons, C. R., and Wiener, M. C. (2003). Inhibition of Tobacco Etch Virus Protease Activity by Detergents. *Protein Expr. Purif.* 27, 109–114. doi:10.1016/S1046-5928(02)00589-2
- Mondal, S., Hsiao, K., and Goueli, S. A. (2015). A Homogenous Bioluminescent System for Measuring GTPase, GTPase Activating Protein, and Guanine Nucleotide Exchange Factor Activities. *ASSAY Drug Dev. Tech.* 13, 444–455. doi:10.1089/adt.2015.643
- Mulry, E., Ray, A. P., and Eddy, M. T. (2021). Production of a Human Histamine Receptor for NMR Spectroscopy in Aqueous Solutions. *Biomolecules* 11, 632. doi:10.3390/biom11050632
- Nishida, M., Cadene, M., Chait, B. T., and MacKinnon, R. (2007). Crystal Structure of a Kir3.1-prokaryotic Kir Channel Chimera. *Embo J.* 26, 4005–4015. doi:10.1038/sj.emboj.7601828
- Pándy-Szekeres, G., Munk, C., Tsonkov, T. M., Mordalski, S., Harpsøe, K., Hauser, A. S., et al. (2018). GPCRdb in 2018: Adding GPCR Structure Models and Ligands. *Nucleic Acids Res.* 46, D440–D446. doi:10.1093/nar/gkx1109
- Pegan, S., Arrabit, C., Zhou, W., Kwiatkowski, W., Collins, A., Slesinger, P. A., et al. (2005). Cytoplasmic Domain Structures of Kir2.1 and Kir3.1 Show Sites for Modulating Gating and Rectification. *Nat. Neurosci.* 8, 279–287. doi:10.1038/nn1411
- Peleg, S., Varon, D., Ivanina, T., Dessauer, C. W., and Dascal, N. (2002). Gai Controls the Gating of the G Protein-Activated K⁺ Channel, GIRK. *Neuron* 33, 87–99. doi:10.1016/S0896-6273(01)00567-0
- Picciochi, A., Šiaučūnaitė-Gaubard, L., Petit-Hartlein, I., Sadir, R., Revilloud, J., Caro, L., et al. (2014). C-terminal Engineering of CXCL12 and CCL5 Chemokines: Functional Characterization by Electrophysiological Recordings. *PLoS One* 9, e87394. doi:10.1371/journal.pone.0087394
- Pluhackova, K., Gahbauer, S., Kranz, F., Wassenaar, T. A., and Böckmann, R. A. (2016). Dynamic Cholesterol-Conditioned Dimerization of the G Protein Coupled Chemokine Receptor Type 4. *Plos Comput. Biol.* 12, e1005169. doi:10.1371/journal.pcbi.1005169
- Provencher, S. W., and Gloeckner, J. (1981). Estimation of Globular Protein Secondary Structure from Circular Dichroism. *Biochemistry* 20, 33–37. doi:10.1021/bi00504a006
- Purvanov, V., Matti, C., Samson, G. P. B., Kindinger, I., and Legler, D. F. (2018). Fluorescently Tagged CCL19 and CCL21 to Monitor CCR7 and ACKR4 Functions. *Ijms* 19, 3876. doi:10.3390/ijms19123876
- Quax, T. E. F., Claessens, N. J., Söll, D., and van der Oost, J. (2015). Codon Bias as a Means to Fine-Tune Gene Expression. *Mol. Cel* 59, 149–161. doi:10.1016/j.molcel.2015.05.035
- Rodriguez, A., Wright, G., Emrich, S., and Clark, P. L. (2018). %MinMax: A Versatile Tool for Calculating and Comparing Synonymous Codon Usage and its Impact on Protein Folding. *Protein Sci.* 27, 356–362. doi:10.1002/pro.3336
- Rosenbaum, D. M., Rasmussen, S. G. F., and Kobilka, B. K. (2009). The Structure and Function of G-Protein-Coupled Receptors. *Nature* 459, 356–363. doi:10.1038/nature08144
- Rossi, A. M., and Taylor, C. W. (2011). Analysis of Protein-Ligand Interactions by Fluorescence Polarization. *Nat. Protoc.* 6, 365–387. doi:10.1038/nprot.2011.305
- Rubinstein, M., Peleg, S., Berlin, S., Brass, D., and Dascal, N. (2007). Gai3 primes the G Protein-Activated K⁺ channels for Activation by Coexpressed G $\beta\gamma$ in intact Xenopus oocytes. *J. Physiol.* 581, 17–32. doi:10.1113/jphysiol.2006.125864
- Scheidelaar, S., Koorengel, M. C., Pardo, J. D., Meeldijk, J. D., Breukink, E., and Killian, J. A. (2015). Molecular Model for the Solubilization of Membranes into Nanodisks by Styrene Maleic Acid Copolymers. *Biophysical J.* 108, 279–290. doi:10.1016/j.bpj.2014.11.3464

- Serrano-Vega, M. J., Magnani, F., Shibata, Y., and Tate, C. G. (2008). Conformational Thermostabilization of the 1-adrenergic Receptor in a Detergent-Resistant Form. *Proc. Natl. Acad. Sci.* 105, 877–882. doi:10.1073/pnas.0711253105
- Song, Y., DiMaio, F., Wang, R. Y.-R., Kim, D., Miles, C., Brunette, T., et al. (2013). High-resolution Comparative Modeling with RosettaCM. *Structure* 21, 1735–1742. doi:10.1016/j.str.2013.08.005
- Song, Y., Ge, B., Lao, J., Wang, Z., Yang, B., Wang, X., et al. (2018). Regulation of the Oligomeric Status of CCR3 with Binding Ligands Revealed by Single-Molecule Fluorescence Imaging. *Biochemistry* 57, 852–860. doi:10.1021/acs.biochem.7b00676
- Soundararajan, M., Willard, F. S., Kimple, A. J., Turnbull, A. P., Ball, L. J., Schoch, G. A., et al. (2008). Structural Diversity in the RGS Domain and its Interaction with Heterotrimeric G Protein γ -subunits. *Proc. Natl. Acad. Sci.* 105, 6457–6462. doi:10.1073/pnas.0801508105
- Sreerama, N., and Woody, R. W. (2000). Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set. *Anal. Biochem.* 287, 252–260. doi:10.1006/abio.2000.4880
- Staus, D. P., Winkler, L. M., Pichugin, D., Prosser, R. S., and Lefkowitz, R. J. (2019). Detergent- and Phospholipid-Based Reconstitution Systems Have Differential Effects on Constitutive Activity of G-Protein-Coupled Receptors. *J. Biol. Chem.* 294, 13218–13223. doi:10.1074/jbc.AC119.009848
- Stone, M., Hayward, J., Huang, C. Z. E. H., and E. Huma, Z. (2017). Mechanisms of Regulation of the Chemokine-Receptor Network. *Ijms* 18, 342. doi:10.3390/ijms18020342
- Sui, Y., Zhang, Y., Dong, C., Xu, B., and Sun, X. (2019). The Small Molecular CCR3 Antagonist YM344031 Attenuates Neurodegenerative Pathologies and Improves Learning and Memory Performance in a Mouse Model of Alzheimer's Disease. *Brain Res.* 1719, 1–10. doi:10.1016/j.brainres.2019.05.022
- Sullivan, K. A., Miller, R. T., Masters, S. B., Beiderman, B., Heideman, W., and Bourne, H. R. (1987). Identification of Receptor Contact Site Involved in Receptor-G Protein Coupling. *Nature* 330, 758–760. doi:10.1038/330758a0
- Tan, Q., Zhu, Y., Li, J., Chen, Z., Han, G. W., Kufareva, I., et al. (2013a). Structure of the CCR5 Chemokine Receptor-HIV Entry Inhibitor Maraviroc Complex. *Science* 341, 1387–1390. doi:10.1126/science.1241475
- Tan, Q., Zhu, Y., Li, J., Chen, Z., Han, G. W., Kufareva, I., et al. (2013b). Structure of the CCR5 Chemokine Receptor-HIV Entry Inhibitor Maraviroc Complex. *Science* 341, 1387–1390. doi:10.1126/science.1241475
- Tang, P., Chong, L., Li, X., Liu, Y., Liu, P., Hou, C., et al. (2014). Correlation between Serum RANTES Levels and the Severity of Parkinson's Disease. *Oxidative Med. Cell Longevity* 2014, 1–4. doi:10.1155/2014/208408
- van Aalst, E., Yekefallah, M., Mehta, A. K., Eason, I., and Wylie, B. (2020). Codon Harmonization of a Kir3.1-KirBac1.3 Chimera for Structural Study Optimization. *Biomolecules* 10, 430. doi:10.3390/biom10030430
- van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastiris, P. L., Karaca, E., et al. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* 428, 720–725. doi:10.1016/j.jmb.2015.09.014
- Wall, M. A., Coleman, D. E., Lee, E., Iñiguez-Lluhi, J. A., Posner, B. A., Gilman, A. G., et al. (1995). The Structure of the G Protein Heterotrimer Gial β 1 γ 2. *Cell* 83, 1047–1058. doi:10.1016/0092-8674(95)90220-1
- Wang, M., Ge, B., Li, R., Wang, X., Lao, J., and Huang, F. (2013). Milligram Production and Biological Activity Characterization of the Human Chemokine Receptor CCR3. *PLoS One* 8, e65500. doi:10.1371/journal.pone.0065500
- Weis, W. I., and Kobilka, B. K. (2018). The Molecular Basis of G Protein-Coupled Receptor Activation. *Annu. Rev. Biochem.* 87, 897–919. doi:10.1146/annurev-biochem-060614-033910
- Whitmore, L., and Wallace, B. A. (2004). DICHROWEB, an Online Server for Protein Secondary Structure Analyses from Circular Dichroism Spectroscopic Data. *Nucleic Acids Res.* 32, W668–W673. doi:10.1093/nar/gkh371
- Whitmore, L., and Wallace, B. A. (2008). Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases. *Biopolymers* 89, 392–400. doi:10.1002/bip.20853
- Xu, P., Huang, S., Zhang, H., Mao, C., Zhou, X. E., Cheng, X., et al. (2021). Structural Insights into the Lipid and Ligand Regulation of Serotonin Receptors. *Nature* 592, 469–473. doi:10.1038/s41586-021-03376-8
- Yeliseev, A., Zoubak, L., and Gawrisch, K. (2007). Use of Dual Affinity Tags for Expression and Purification of Functional Peripheral Cannabinoid Receptor. *Protein Expr. Purif.* 53, 153–163. doi:10.1016/j.pep.2006.12.003
- Yen, H.-Y., Hoi, K. K., Liko, I., Hedger, G., Horrell, M. R., Song, W., et al. (2018). PtdIns(4,5)P2 Stabilizes Active States of GPCRs and Enhances Selectivity of G-Protein Coupling. *Nature* 559, 423–427. doi:10.1038/s41586-018-0325-6
- Zhang, Q., Qin, J., Zhong, L., Gong, L., Zhang, B., Zhang, Y., et al. (2015). CCL5-Mediated Th2 Immune Polarization Promotes Metastasis in Luminal Breast Cancer. *Cancer Res.* 75, 4312–4321. doi:10.1158/0008-5472.CAN-14-3590
- Zhukovsky, M. A., Lee, P.-H., Ott, A., and Helms, V. (2013). Putative Cholesterol-Binding Sites in Human Immunodeficiency Virus (HIV) Coreceptors CXCR4 and CCR5. *Proteins* 81, 555–567. doi:10.1002/prot.24211

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 van Aalst and Wylie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Ligand Binding Introduces Significant Allosteric Shifts in the Locations of Protein Fluctuations

Ambuj Kumar and Robert L. Jernigan *

Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA, United States

Allostery is usually considered to be a mechanism for transmission of signals associated with physical or dynamic changes in some part of a protein. Here, we investigate the changes in fluctuations across the protein upon ligand binding based on the fluctuations computed with elastic network models. These results suggest that binding reduces the fluctuations at the binding site but increases fluctuations at remote sites, but not to fully compensating extents. If there were complete conservation of entropy, then only the enthalpies of binding would matter and not the entropies; however this does not appear to be the case. Experimental evidence also suggests that energies and entropies of binding can compensate but that the extent of compensation varies widely from case to case. Our results do however always show transmission of an allosteric signal to distant locations where the fluctuations are increased. These fluctuations could be used to compute entropies to improve evaluations of the thermodynamics of binding. We also show the allosteric relationship between peptide binding in the GroEL trans-ring that leads directly to the release of GroES from the GroEL-GroES cis-ring. This finding provides an example of how calculating these changes to protein dynamics induced by the binding of an allosteric ligand can regulate protein function and mechanism.

Keywords: elastic network model, Gaussian network model, structural fluctuations, GroEL, allostery, ligand binding, thermodynamics of protein binding

OPEN ACCESS

Edited by:

Ivan Rivalta,
University of Bologna, Italy

Reviewed by:

Xiaoqin Zou,
University of Missouri, United States
Ilya Vakser,
University of Kansas, United States

*Correspondence:

Robert L. Jernigan
jernigan@iastate.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 29 June 2021

Accepted: 09 August 2021

Published: 01 September 2021

Citation:

Kumar A and Jernigan RL (2021)
Ligand Binding Introduces Significant
Allosteric Shifts in the Locations of
Protein Fluctuations.
Front. Mol. Biosci. 8:733148.
doi: 10.3389/fmolb.2021.733148

INTRODUCTION

Many proteins have intrinsic dynamics that relates to their function, with a specific dynamics that enables them to undergo large conformational transitions in response to external stimuli, such as ligand binding (Skjærven et al., 2011). This binding can cause changes to the protein structure, and because the proteins are densely packed, these can show allosteric changes at sites distant from the binding site. Such allosteric communication within a protein is essential for the progression of biochemical processes. Understanding protein allosteric behavior induced by ligand binding is important for understanding the thermodynamics of binding, since these distant flexibilities are representative of an entropy that has rarely been considered in treating protein binding. This is likely important for protein assembly into machine-like structures as well as for investigations of drug binding to target proteins. It has been shown that the binding of small molecules to a protein is associated with increase in protein thermostability (Fukada et al., 1983; Brandts and Lin, 1990; Shrake and Ross, 1990, 1992) and allosterically associated with large conformational transitions in protein (Pan et al., 2000). A systematic study of such conformational transitions induced by ligand binding can also yield mechanistic insights, as will be shown here for the chaperonin GroEL.

It is well accepted that binding a ligand to a protein causes a gain in enthalpy because of the new interactions formed, which can then be cancelled, at least in part, by a loss in entropy at the binding site, leading to a smaller net change in the free energy of the system. Binding of a ligand is normally favored by the change in residue fluctuations at the protein binding site. This is a type of enthalpy-entropy compensation (EEC), and it has been used to understand the thermodynamic changes involved in ligand binding (Gallicchio et al., 1998; Talhout et al., 2003; Lafont et al., 2007; DeLorbe et al., 2009; Ryde, 2014). It has also been suggested that a stronger interaction between a ligand and a protein will cause a relatively larger loss in entropy in comparison with resulting weaker ligand-protein interactions (Gallicchio et al., 1998). Detailed studies on EEC in ligand recognition and binding has been extensively discussed and reviewed (Dunitz, 1995; Brownowska, 2011; Reynolds and Holloway, 2011; Chodera and Mobley, 2013). EEC calculations for binding of a large number of related ligand molecules show clear evidence of EEC, where no change in the binding affinity was observed despite the large change in ΔH as well as $-T\Delta S$ of the system, where ΔH is the change in enthalpy, T is the temperature and ΔS is the change in entropy (Talhout et al., 2003). Another study showed a linear relationship between ΔH and $T\Delta S$ with evidence of EEC upon binding of Ca^{2+} ion to calmodulin (Kuroki et al., 1992), which has been linked to a folding-like process. Such evidence suggests that EEC changes upon ligand binding can influence protein folding and reorganization of residue contacts.

The conformational changes upon ligand binding are accompanied by changes in the protein packing, which affects its stability, as manifested in changes to the fluctuation spectrum. These changes in fluctuations in a protein structure are changes correlated with the binding process itself. It has been proposed that the residual motions of the protein-ligand system after binding can be estimated with vibrational entropy (Dunitz, 1995). Such vibrational motions can enable the transfer of motion across the distant parts of a protein. Therefore, it is evident that local EEC changes induced by ligand binding can cause other changes to be transmitted across the protein, leading to the losses as well as gains in fluctuations, even away from the binding site. The Cooper-Dryden model (Cooper and Dryden, 1984) considers the importance of entropy in the allosteric regulation of protein mechanism, proposing a role for the shift in vibrational motions in mediating allosteric responses. It has been shown that the changes due to ligand binding can be transmitted to the distant parts of the protein through changes in atomic fluctuations (Cooper and Dryden, 1984). The idea of entropy compensation has been further validated where the change in fluctuation at the ligand binding site has been shown to induce increased fluctuations in physically distant parts of a protein (Müller et al., 1996). Moreover, binding of a ligand leads to some redistribution of conformational entropy, which plays an important role in regulating protein mechanism (Capdevila et al., 2017). Nevertheless, the role of such allosteric changes associated with fluctuations and conformational entropy remain poorly understood. Below, we demonstrate how entropy can almost be viewed to be somewhat conserved before and after ligand binding and that this can have a significant impact on assessment of the

thermodynamics of the binding process. Our results demonstrate that ligand-induced increases in fluctuations in distant parts of the protein are important and may be critical for a better understanding of protein binding.

METHOD

Data Collection

Ligand bound structures were downloaded from the PDB database (Berman et al., 2002). The unbound structures are generated by removing the ligands from these structures since there are few unbound structures. The GroEL-GroES-peptide bound structure was generated by placing peptides into the GroEL-GroES-ADP structure (PDB ID: 1PF9) at their locations in the peptide bound GroEL structure (PDB ID: 1MNF), to minimize the local RMSD fit. Visualization uses Pymol (Vinet and Zhedanov, 2011).

Gaussian Network Model

Elastic Network Models are a simple way to obtain the important protein dynamics. While they may not provide all of the details of atomic molecular dynamics, it has been demonstrated as a reliable way to characterize the overall dynamics of proteins, particularly for the slowest motions that are the most important for function, which are difficult to obtain with molecular dynamics. The earliest such model is still the best type in agreeing best with crystallographic B factors and is used here to characterize the locations of the part of the structure showing the largest fluctuations and is clearly sufficient to demonstrate the large shifts in the location of the most conformationally labile part of the structure. This Elastic Network Model used here is the Gaussian network model (GNM) (Bahar et al., 1997; Rader et al., 2005) that was developed to study the collective motions in terms of the scalar fluctuations for individual amino acids. This model is distinguishable from the commonly used ANM that provides information about the directions of motion that are essential for understanding dynamics. Here we implement this GNM in the customary way on coarse-grained proteins by using only the C^α atoms to represent each amino acid. All C^α atoms and ligand atoms within a distance of 7.5 Å are connected with identical Hookean springs having the same spring constant. The energy for GNM is given as

$$V = \frac{1}{2} \gamma \sum_{i,j}^N \Gamma \left[(\Delta R_i - \Delta R_j)^2 \right] \quad (1)$$

where ΔR_i and ΔR_j are the fluctuation vector of amino acids i and j , γ is the spring constant, Γ is an $N \times N$ symmetric matrix, which specifies the node connectivity, where the springs are placed in the protein, and is defined as

$$\Gamma = \begin{cases} -1 & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum_{i, i \neq j} \Gamma_{ij} & \text{if } i = j \end{cases} \quad (2)$$

Singular value decomposition yields the eigenvector (v) and eigenvalues (λ) of Γ . Fluctuations of each residue i (f_i) are calculated as

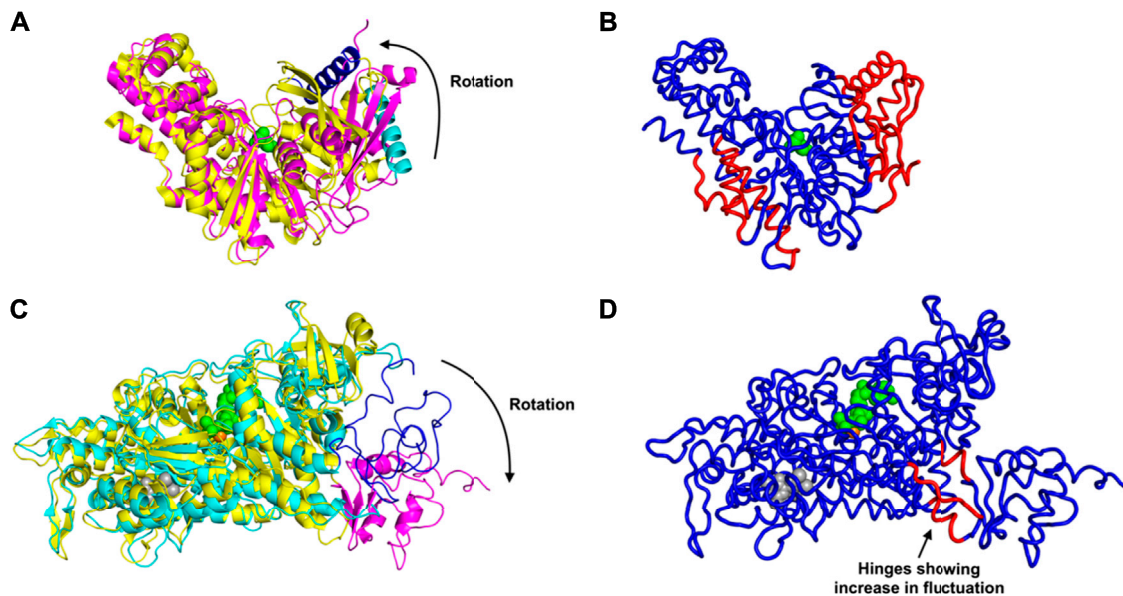


FIGURE 1 | Conformational Transition and Changes in Fluctuations Induced by Ligand Binding, for glucokinase shown in **(A)** as superimposed unbound and bound structures and **(B)** bound. And similarly for the myosin motor protein in **(C)** unbound and bound structures superimposed and **(D)** bound. **(A)** Rotation and domain reorganization upon binding and release of glucose (shown as green spheres). Glucose bound closed form (PDB ID: 1V4S) is shown in magenta and the open form (PDB ID: 1V4T) after glucose is released in yellow. Glucokinase small domain helix residues 117–137 are shown in blue in closed form and in cyan in open form to highlight the domain rotation. **(B)** Increased fluctuations (red) and decreased fluctuations (blue) upon binding of glucose to glucokinase. **(C)** Rotation of the C-terminus (residues 692–776) of the myosin motor protein upon ligand binding. C-terminus residues 692–776 are shown in magenta in the ligand bound form (2JHR) and in blue in the apo structure (PDB: 2Y0R). The ligand ADP metavanadate is shown in green spheres, magnesium ion as an orange sphere, and pentabromopseudilin as grey spheres. **(D)** Sites having increased (red) and decreased (blue) fluctuations in the myosin motor protein upon ligand binding. In the case of the central binding of glucose to glucokinase, these increased fluctuations are on the two outside parts of the structure but in the myosin motor protein, they are instead at the central hinge, indicating a kind of dynamic flexibility between the two domains of this protein.

$$f_i = \sum_m \frac{v_{mi}^2}{\lambda_m} \quad (3)$$

where, v_{mi} is the eigenvector of residue i mode m and λ_m is the eigenvalue of mode m . Differences in residue fluctuations due to ligand binding are calculated from

$$\Delta f_i = f_i^{(perturbed)} - f_i^{(normal)} \quad (4)$$

where, $f_i^{(perturbed)}$ is the fluctuation of residue i when ligand is bound to the protein and $f_i^{(normal)}$ is the fluctuation of residue i for the protein with no ligand bound. Note that Δf_i can have either positive or negative values, indicating either an increase in fluctuations in residue i or its reduced mobility. As will be seen, the localization of fluctuations is drastically impacted by the binding of ligand. While it is readily apparent that fluctuations at the location of the bound ligand are reduced by its specific interactions, it has not been so widely accepted that ligand binding induces specific remote increases in fluctuations. As we show below in several examples, the locations for these changes are readily discerned with the simple Gaussian Network Model. As we will see, the repositioned fluctuations have positions that depend strongly on the structures themselves.

RESULTS AND DISCUSSION

The resulting changes in the fluctuations (and the corresponding entropies) accompanying binding can be quite different and

depend on the details of the protein structure. This work also points up the importance of estimating entropies of proteins, which has always been difficult. The results here suggest that the computed fluctuations could be used directly to estimate the entropies.

First we consider the case of glucokinase that regulates the metabolism of glucose; it undergoes a large conformational transition and reorganization within its small domain upon binding of glucose (Kamata et al., 2004) (**Figure 1A**). The glucose binding induced conformational transition mediates the catalytic cycle of glucokinase (Kamata et al., 2004). This involves multiple steps of conformational changes as well as an intra-domain sliding motion that breaks and forms new contacts within the domain (Kamata et al., 2004). Our results indicate a significant increase in the fluctuations within the glucokinase small domain (on the right side of the structures shown in **Figure 1B**), which undergoes significant atomic displacement upon ligand binding (Kamata et al., 2004), with a relatively large conformational change as can be seen in **Figure 1B**, in comparison with the structure of that domain from **Figure 1A**. Increased fluctuations within this small domain likely assist in the repositioning of these residues and the relatively large conformational changes observed.

Even more striking are the changes observed in the myosin motor protein shown in the unbound form in **Figure 1C** and in the bound form in **Figure 1D**. In this case what had been a relatively extended and loosely pack small domain on the right

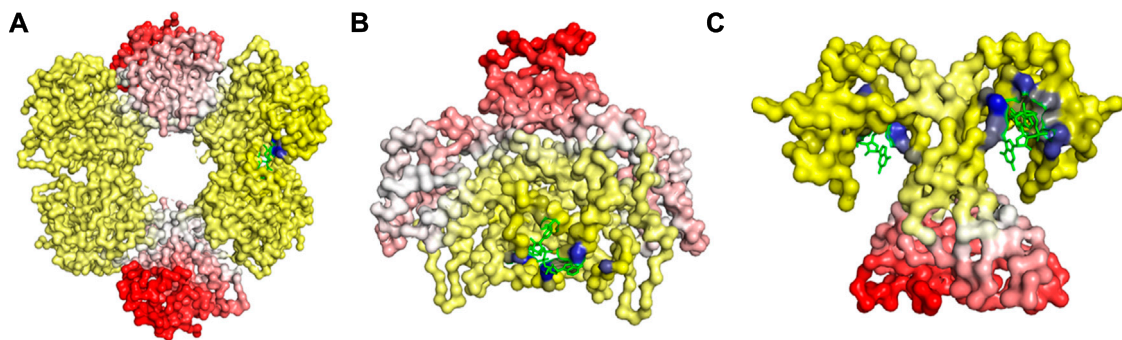


FIGURE 2 | Changes in Fluctuations upon Ligand Binding in (A) UDP-glucose 6-dehydrogenase (PDB ID: 5TJH) bound with uridine-5'-diphosphate-glucose, (B) Citrate synthase (PDB ID: 4JAF) bound with 1, 4-Dihyronicotinamide adenine dinucleotide, and (C) uncharacterized protein VCA0042 (PDB ID: 2RDE) bound with 9,9'-[(2R,3R,3aS,5S,7aR,9R,10R,10aS,12S,14aR)-3,5,10,12-tetrahydroxy-5,12-dioxidoctahydro-2H,7H-difuro[3,2-d:3',2'-j][1,3,7,9,2,8]tetraoxadiphosphacyclododecine-2,9-diyl]bis(2-amino-1,9-dihydro-6H-purin-6-one). The gain in fluctuations is scaled between 0 (shown in white) and 1 (red), and the loss in fluctuations is scaled between 0 (yellow) and -1 (blue). Here, the ligand is in green.

side has become more densely packed and more structured, but the increases in fluctuations after binding are seen in the hinge region between the large and small domains (in red in Figure 1D).

An alternate mechanism of conformational changes induced by ligand binding can be associated with changes in fluctuations within inter-domain hinges. The myosin motor protein has 2,116 residues with the myosin motor at its N-terminal region. It contains a reactive thiol region in the C-terminus of the motor domain comprised of the SH1 (residues 681–689) and SH2 (residues 669–678) helices (Preller et al., 2011). The binding of an allosteric ligand to the motor protein introduces rotation in the C-terminal residues 692–776 (Figure 1C). Our results indicate that binding of an allosteric ligand and ATP together increases the fluctuations in the hinge region at residues 490–496, 499–506, and 686–691, which also contains a part of the SH2 domain (Figure 1D). Increased fluctuations in the hinge region may initiate the rotation of myosin motor domain C-terminal residues following ligand binding.

A trend of decay in change of fluctuation was observed as we move farther from the ligand binding pocket (Figure 2), which was then followed by an increase in fluctuation beyond a certain distance from the binding pocket (Figure 2). These changes indicate that the EEC follows a progressive transmission of fluctuation changes, where the ligand binding pocket experiences a large loss in fluctuations, first leading to relatively lower fluctuation losses in the residues close to the binding site, which is then followed by slight increases in fluctuations as we move farther from the binding site. This trend in fluctuation changes may help us locate the group of residues that are susceptible to conformational transitions upon ligand binding.

We have further investigated the allosteric impact of ligand binding for a variety of proteins (Table 1) showing less detail in the values of the fluctuations, such as Mitogen-activated protein kinase 8 (Figure 3A), Citrate synthase (Figure 3B), Uncharacterized protein VCA0042 (Figure 3C), ATP sulfurylase (Figure 3D), Glutamate dehydrogenase (Figure 3E), Acetyl-Coenzyme A carboxylase (Figure 3F),

TABLE 1 | Proteins used in this study together with their corresponding PDB IDs and their ligands used to calculate residue fluctuations in the bound state. Here ligand 46A represents N-butyl-4,6-dimethyl-N-[[2'-(2H-tetrazol-5-yl) biphenyl-4-yl]methyl]pyrimidin-2-amine, NAI represents 1, 4-Dihyronicotinamide adenine dinucleotide, C2E represents 9,9'-[(2R,3R,3aS,5S,7aR,9R,10R,10aS,12S,14aR)-3,5,10,12-tetrahydroxy-5,12-dioxidoctahydro-2H,7H-difuro[3,2-d:3',2'-j][1,3,7,9,2,8]tetraoxadiphosphacyclododecine-2,9-diyl]bis(2-amino-1,9-dihydro-6H-purin-6-one), PPS represents 3'-Phosphate-adenosine-5'-phosphate sulphate, GTP represents Guanosine-5'-triphosphate, S1A represents Soraphen A, ATP represents Adenosine-5'-triphosphate, AMP represents Adenosine monophosphate, MG represents magnesium ion, RFZ represents 5,6-dichloro-1-beta-D-ribofuranosyl-1H-benzimidazole, and FBP represents 1,6-di-O-phosphono-beta-D-fructofuranose.

Protein	PDB ID	Ligand used
Mitogen-activated protein kinase 8	3O2M	46A
Citrate synthase	4JAF	NAI
Uncharacterized protein VCA0042	2RDE	C2E
ATP sulfurylase	1M8P	PPS
Glutamate dehydrogenase	3ETE	GTP
Acetyl-Coenzyme A carboxylase	1W96	S1A
Isocitrate dehydrogenase kinase/phosphatase	3EPS	ATP, AMP, MG
Casein kinase II subunit alpha	3H30	RFZ
L-lactate dehydrogenase	2LDB	FBP

Isocitrate dehydrogenase kinase/phosphatase (Figure 3G), Casein kinase II subunit alpha (Figure 3H), and L-lactate dehydrogenase (Figure 3I) and colored in a simple binary way to indicate only the changes in blue showing all residues with losses fluctuations and red for all residues showing gains in fluctuation. Results consistently indicate increased fluctuations in the distant regions of each protein far from the location where the ligand is bound. In ATP Sulfurylase, binding of the ligand to the C-terminal region shows increases in the fluctuations in the N-terminal region for all three chains (A, B, and C), indicating a compensatory mechanism of entropy conservation across the protein assembly through changes in fluctuations induced by the ligand binding. A similar trend of fluctuation changes in the distant region of the proteins was observed for all the nine structures (Figure 3).

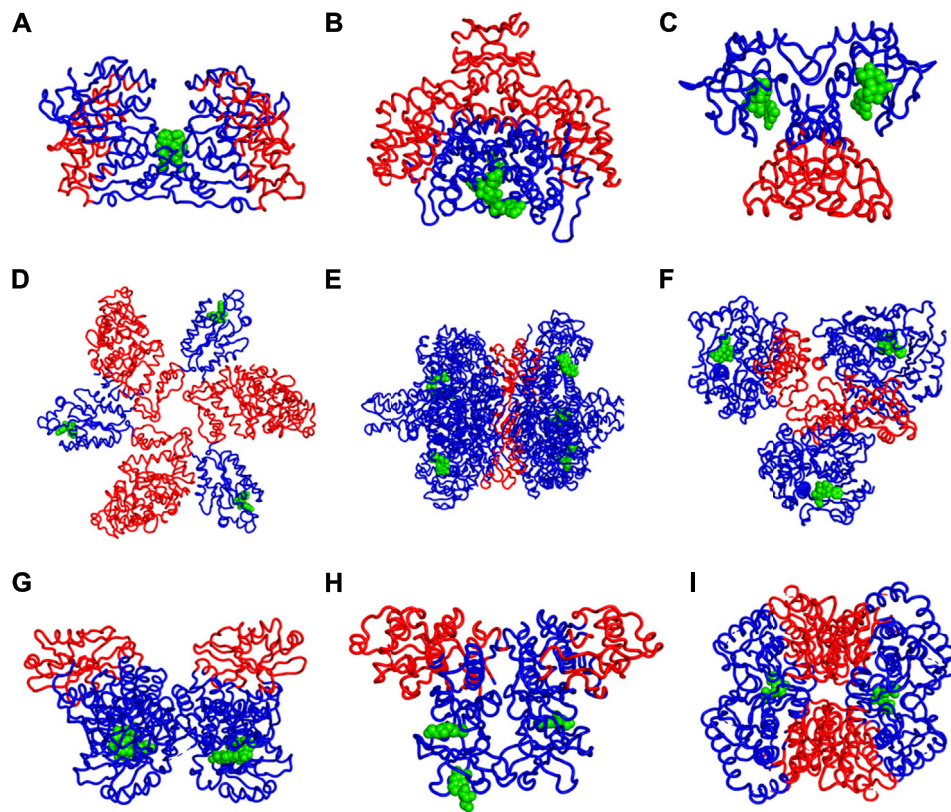


FIGURE 3 | Increases (red) and Decreases (blue) in Fluctuations upon Lgand (green spheres) Binding in **(A)** Mitogen-activated protein kinase 8, **(B)** Citrate synthase, **(C)** Uncharacterized protein VCA0042, **(D)** ATP sulfurylase, **(E)** Glutamate dehydrogenase, **(F)** Acetyl-Coenzyme A carboxylase, **(G)** Isocitrate dehydrogenase kinase/phosphatase, **(H)** Casein kinase II subunit alpha, **(I)** L-lactate dehydrogenase. While there appears to be a large variety of positions that are affected, the universal rule appears to be that the part of the structure furthest away from the binding site has the largest increases in fluctuations. If the binding site is near the outside of the protein, then the largest increases in fluctuations will be at the center of the protein. One particularly interesting such case is shown in part d where there are three occupied binding sites, which can be presumed, each to increase the fluctuations in the most distant part so with the three occupied binding sites all three domains have increased fluctuations near the center of the overall structure.

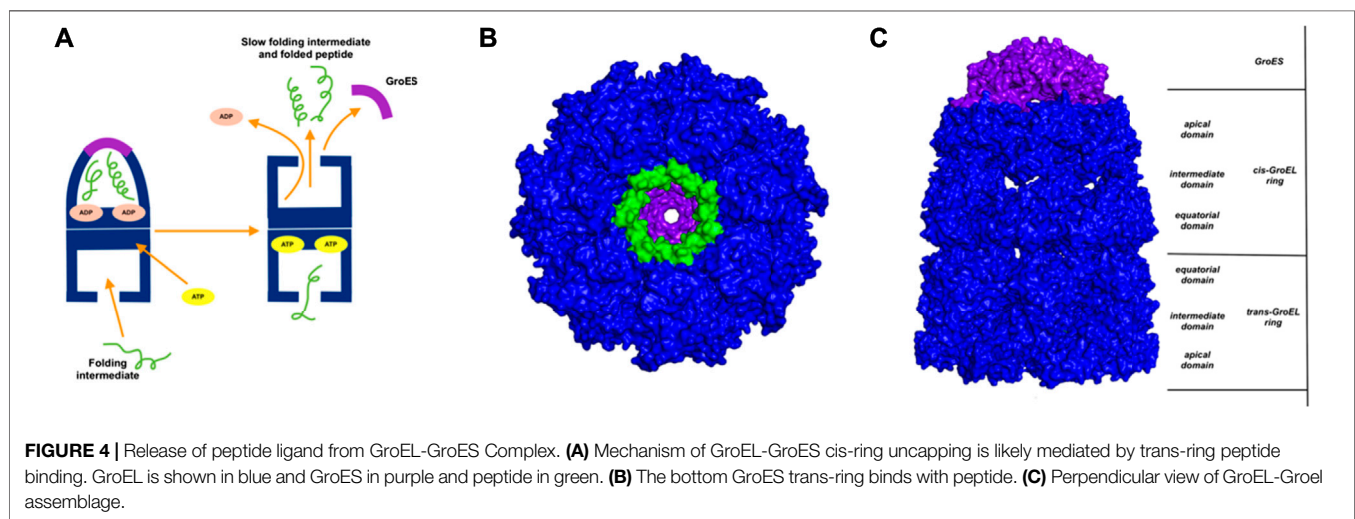
To address the issue of possible artifacts, we have replaced all protein bonded ligand atoms with a decoy ligand modelled by picking one random atom from all the bonded ligands. The increases in fluctuation of the protein residues for the bound known ligand and this artificial decoy ligand binding have been compared to check whether a similar trend of increased fluctuations are also observed for the decoy ligand. The result indicates that decoy ligand induced only a significantly smaller increase in fluctuations in the same regions where increases in fluctuations were observed in response to the known allosteric ligand binding (Table 2). Moreover, in a few cases such as isocitrate dehydrogenase kinase/phosphatase and casein kinase II subunit alpha, binding of this decoy ligand caused larger reductions in the fluctuation gains at the distant region of interest (Table 2). This trend of relatively lower impact of the decoy ligand indicates that the enthalpy-entropic changes observed in our calculations are specific to the actual ligand and are not a general result for any possible ligand. Furthermore, the ratio of mean of increase in fluctuation to the mean of decrease in fluctuation across all residues in the protein upon

the actual known ligand binding shows a diverse trend across 12 proteins (Table 2), ranging from a very low value of 0.59% to a very high value of 98.57%, indicating that enthalpy-entropy compensation can occur across a broad range in different proteins, which all suggests that the extent energy-entropy compensation depends on the details of each case, and that it would almost never be exact compensation. Nonetheless, usually the net result of loss of fluctuations (entropy loss) is normally partially compensated by fluctuations (and entropy gains) at distant regions of the protein.

GroEL is a chaperone protein essential for the proper folding of various misfolded proteins. GroEL contains two rings composed of seven subunit each, connected back-to-back to form a central substrate binding chamber (Figure 4). One GroEL ring can also bind to the GroES cap, a co-chaperone protein closing over the GroEL, to create a cis-ring complex. The GroEL-GroES cis-ring binds to ATP and undergoes biomechanical motions that pull folded proteins apart to enable them to refold. Many experiments have been conducted to understand the mechanism involved in GroEL-GroES based chaperone folding under conditions where

TABLE 2 | Ratios of increases in fluctuations to decreased in fluctuations upon ligand binding, mean increases in fluctuations in the residues where fluctuation increases upon ligand binding, and mean decrease in fluctuation in the residues where fluctuation decreases upon ligand binding.

PDB ID	Ratio of fluctuation change (%)	Mean increases in fluctuations	
		Known ligand	Artificial decoy
4JAF	8.27	0.0014	4.43×10^{-7}
3O2M	10.60	0.0023	3.01×10^{-5}
5TJH	98.57	0.0219	0.0009
1V4S	20.61	0.0033	0.0003
2RDE	52.79	0.0326	5.93×10^{-5}
1M8P	10.19	0.0004	6.50×10^{-5}
3ETE	0.59	4.23×10^{-5}	9.78×10^{-7}
1W96	7.48	0.0009	1.61×10^{-5}
3EPS	6.63	0.0023	-7.12×10^{-6}
3H30	9.13	0.0022	-6.64×10^{-6}
2LDB	52.20	0.0075	0.0001
2JHR	3.6	0.0005	4.66×10^{-6}



the spontaneous folding mechanism fails to proceed (Weissman et al., 1994; Fenton and Horwich, 1997; Sigler et al., 1998; Ishino et al., 2015).

With the gain of broader insights into the GroEL molecular mechanism, the detailed understanding of its allosteric regulation has seen recent gains. It has been shown that there is direct allosteric coupling between the assembly of the GroEL-GroES cis-ring on one end with the disassembly of the cis-ring on the other end of the protein (Lin and Rye, 2006). Binding of ATP and the peptide to the open trans-ring on a GroEL complex triggers the disassembly of the cis-ring (Lin and Rye, 2006) (Figure 4A). In this work, we are investigating the role of binding of intermediately folded peptide binding to the trans-ring of GroEL to trigger the disassembly of the cis-ring (Figure 4B). Our results show that binding of peptide chains to the trans-ring allosterically increases the fluctuations in the GroEL cis-ring apical domain, in the GroES protein as well as in the equatorial domain of GroEL cis and trans-ring (Figure 5). The increased fluctuations in the GroEL cis-ring apical domain and GroES are likely to affect the binding affinity of GroES cap,

facilitating the release of the GroES. The increased fluctuations with the GroES cap itself are likely to play an important role in its unbinding. Moreover, the increase in flexibility in equatorial domain of the cis-ring may also facilitate the release of the bound ADP molecules.

Detailed insights into how ligand binding relates to EEC can help to understand large conformational changes, molecular motions associated with protein function, structure-function relationships, as well as how to better evaluate docked protein poses, which are usually based on enthalpies while ignoring entropies. The extent of energy-entropy compensation appears to be highly variable for different cases as determined by isothermal titration calorimetry (Olsson et al., 2011). It is difficult to predict the effects of small changes; for example, an added hydrogen acceptor group on a HIV-protease inhibitor yielded a gain in enthalpy, but the corresponding loss in entropy compensated and resulted in an insignificant change in the binding affinity of the ligand (Lafont et al., 2007). A similar phenomena was observed where chemical modifications to a thrombin ligand failed to cause significant changes in ligand

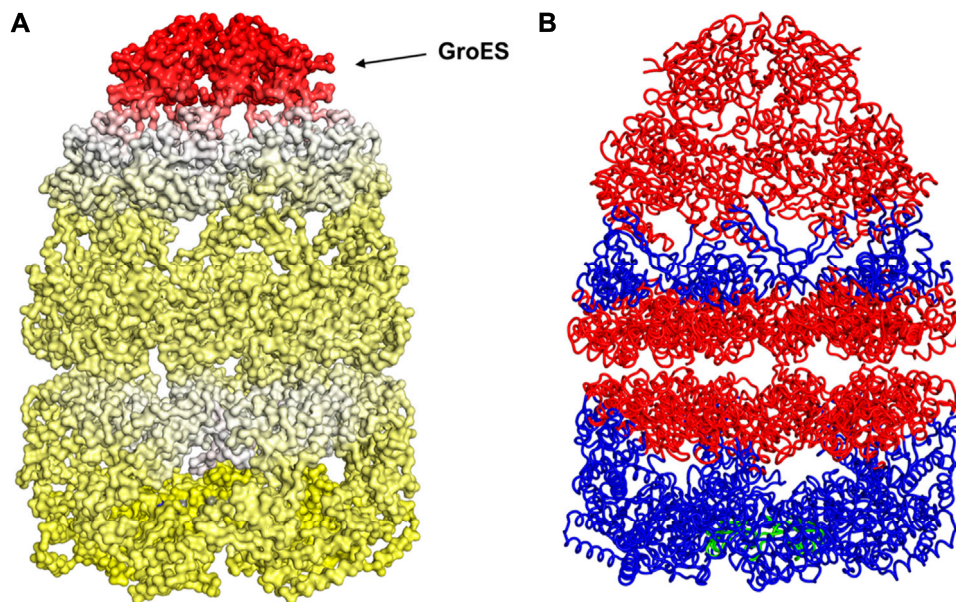


FIGURE 5 | Mechanism for release of GroES (and internal peptide). Change in fluctuations in the GroEL-GroES assemblage upon peptide binding. **(A)** Gains of fluctuations are scaled between 0 (white) and 1 (red), and loss of fluctuations shown also but separately between 0 (yellow) and -1 (blue), **(B)** Increased (red) and decreased (blue) fluctuations upon peptide binding is shown. This suggests that peptide binding to the lower ring causes release of GroES and internal peptide from the upper ring.

binding affinity due to competing changes between the enthalpic and entropic contributions (Baum et al., 2010). Such EEC is commonly observed for ligand-protein interactions (Whitesides and Krishnamurthy, 2005). Therefore, a better understanding of the entropic changes induced by ligand binding is needed for more reliable computations of the effects of binding.

CONCLUSION

We have demonstrated the allosteric transfer of fluctuations to distant regions of a variety of proteins as differential changes in fluctuations, driven by some level of conservation of entropy throughout the whole protein. These changes in fluctuations may induce a conformational rearrangement within protein domains or cause allosteric conformational transitions. The pattern of ligand binding induced fluctuation changes observed in this study indicates that changes to the local packing density can influence the dynamics of the most distant regions within even the largest protein assemblages such as the example shown here for GroEL

and may play a significant role in regulating and controlling their molecular mechanisms.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication

FUNDING

Support provided by NIH grant R01GM127701 and NSF grant DBI 1661391.

REFERENCES

- Bahar, I., Atilgan, A. R., and Erman, B. (1997). Direct Evaluation of thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential. *Folding Des.* 2, 173–181. doi:10.1016/S1359-0278(97)00024-2
- Baum, B., Muley, L., Smolinski, M., Heine, A., Hangauer, D., and Klebe, G. (2010). Non-additivity of Functional Group Contributions in Protein-Ligand Binding: A Comprehensive Study by Crystallography and Isothermal Titration Calorimetry. *J. Mol. Biol.* 397, 1042–1054. doi:10.1016/j.jmb.2010.02.007
- Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., et al. (2002). The Protein Data Bank. *Acta Crystallogr. D Biol. Cryst.* 58, 899–907. doi:10.1107/S0907444902003451
- Brandts, J. F., and Lin, L. N. (1990). Study of Strong to Ultratight Protein Interactions Using Differential Scanning Calorimetry. *Biochemistry* 29, 6927–6940. doi:10.1021/bi00481a024
- Bronowska, A. K. (2011). “Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design,” In *Thermodynamics - Interaction Stud. - Sol. Liquids Gases*. Editor Pirajan, J. C. M. IntechOpen 130, 735–736. doi:10.5772/19447

- Capdevila, D. A., Braymer, J. J., Edmonds, K. A., Wu, H., and Giedroc, D. P. (2017). Entropy Redistribution Controls Allostery in a Metalloregulatory Protein. *Proc. Natl. Acad. Sci. USA* 114, 4424–4429. doi:10.1073/pnas.1620665114
- Chodera, J. D., and Mobley, D. L. (2013). Entropy-enthalpy Compensation: Role and Ramifications in Biomolecular Ligand Recognition and Design. *Annu. Rev. Biophys.* 42, 121–142. doi:10.1146/annurev-biophys-083012-130318
- Cooper, A., and Dryden, D. T. F. (1984). Allostery without Conformational Change. *Eur. Biophys. J.* 11, 103–109. doi:10.1007/BF00276625
- DeLorbe, J. E., Clements, J. H., Teresk, M. G., Benfield, A. P., Plake, H. R., Millsbaugh, L. E., et al. (2009). Thermodynamic and Structural Effects of Conformational Constraints in Protein–Ligand Interactions. Entropic Paradox Associated with Ligand Preorganization. *J. Am. Chem. Soc.* 131, 16758–16770. doi:10.1021/ja904698q
- Dunitz, J. D. (1995). Win Some, Lose Some: Enthalpy-Entropy Compensation in Weak Intermolecular Interactions. *Chem. Biol.* 2, 709–712. doi:10.1016/1074-5521(95)90097-7
- Fenton, W. A., and Horwich, A. L. (1997). GroEL-mediated Protein Folding. *Protein Sci.* 6, 743–760. doi:10.1002/pro.5560060401
- Fukada, H., Sturtevant, J. M., and Quioco, F. A. (1983). Thermodynamics of the Binding of L-Arabinose and of D-Galactose to the L-Arabinose-Binding Protein of *Escherichia coli*. *J. Biol. Chem.* 258, 13193–13198. doi:10.1016/s0021-9258(17)44100-7
- Gallicchio, E., Kubo, M. M., and Levy, R. M. (1998). Entropy–Enthalpy Compensation in Solvation and Ligand Binding Revisited. *J. Am. Chem. Soc.* 120, 4526–4527. doi:10.1021/ja974061h
- Ishino, S., Kawata, Y., Taguchi, H., Kajimura, N., Matsuzaki, K., and Hoshino, M. (2015). Effects of C-Terminal Truncation of Chaperonin GroEL on the Yield of In-Cage Folding of the green Fluorescent Protein. *J. Biol. Chem.* 290, 15042–15051. doi:10.1074/jbc.M114.633636
- Kamata, K., Mitsuya, M., Nishimura, T., Eiki, J.-i., and Nagata, Y. (2004). Structural Basis for Allosteric Regulation of the Monomeric Allosteric Enzyme Human Glucokinase. *Structure* 12, 429–438. doi:10.1016/j.str.2004.02.005
- Kuroki, R., Nitta, K., and Yutani, K. (1992). Thermodynamic Changes in the Binding of Ca²⁺ to a Mutant Human Lysozyme (D86/92). Enthalpy-Entropy Compensation Observed upon Ca²⁺ Binding to proteins Enthalpy-Entropy Compensation Observed upon Ca²⁺ Binding to Proteins. *J. Biol. Chem.* 267, 24297–24301. doi:10.1016/s0021-9258(18)35764-8
- Lafont, V., Armstrong, A. A., Ohtaka, H., Kiso, Y., Mario Amzel, L., and Freire, E. (2007). Compensating Enthalpic and Entropic Changes Hinder Binding Affinity Optimization. *Chem. Biol. Drug Des.* 69, 413–422. doi:10.1111/j.1747-0285.2007.00519.x
- Lin, Z., and Rye, H. S. (2006). GroEL-mediated Protein Folding: Making the Impossible, Possible. *Crit. Rev. Biochem. Mol. Biol.* 41, 211–239. doi:10.1080/10409230600760382
- Müller, C., Schlauderer, G., Reinstein, J., and Schulz, G. (1996). Adenylate Kinase Motions during Catalysis: An Energetic Counterweight Balancing Substrate Binding. *Structure* 4, 147–156. doi:10.1016/S0969-2126(96)00018-4
- Olsson, T. S. G., Ladbury, J. E., Pitt, W. R., and Williams, M. A. (2011). Extent of Enthalpy-Entropy Compensation in Protein-Ligand Interactions. *Protein Sci.* 20, 1607–1618. doi:10.1002/pro.692
- Pan, H., Lee, J. C., and Hilser, V. J. (2000). Binding Sites in *Escherichia coli* Dihydrofolate Reductase Communicate by Modulating the Conformational Ensemble. *Proc. Natl. Acad. Sci.* 97, 12020–12025. doi:10.1073/pnas.220240297
- Preller, M., Bauer, S., Adamek, N., Fujita-Becker, S., Fedorov, R., Geeves, M. A., et al. (2011). Structural Basis for the Allosteric Interference of Myosin Function by Reactive Thiol Region Mutations G680A and G680V. *J. Biol. Chem.* 286, 35051–35060. doi:10.1074/jbc.M111.265298
- Rader, A. J., Chennubhotla, C., Yang, L. W., and Bahar, I. (2005). The Gaussian Network Model: Theory and Applications. *Normal Mode Anal. Theor. Appl. Biol. Chem. Syst.*, 65–88. doi:10.1201/9781420035070-9
- Reynolds, C. H., and Holloway, M. K. (2011). Thermodynamics of Ligand Binding and Efficiency. *ACS Med. Chem. Lett.* 2, 433–437. doi:10.1021/ml200010k
- Ryde, U. (2014). A Fundamental View of Enthalpy-Entropy Compensation. *Med. Chem. Commun.* 5, 1324–1336. doi:10.1039/c4md00057a
- Shrake, A., and Ross, P. D. (1990). Ligand-induced Biphasic Protein Denaturation. *J. Biol. Chem.* 265, 5055–5059. doi:10.1016/s0021-9258(19)34083-9
- Shrake, A., and Ross, P. D. (1992). Origins and Consequences of Ligand-Induced Multiphasic thermal Protein Denaturation. *Biopolymers* 32, 925–940. doi:10.1002/bip.360320804
- Sigler, P. B., Xu, Z., Rye, H. S., Burston, S. G., Fenton, W. A., and Horwich, A. L. (1998). Structure and Function in GroEL-Mediated Protein Folding. *Annu. Rev. Biochem.* 67, 581–608. doi:10.1146/annurev.biochem.67.1.581
- Skjærven, L., Reuter, N., and Martinez, A. (2011). Dynamics, Flexibility and Ligand-Induced Conformational Changes in Biological Macromolecules: A Computational Approach. *Future Med. Chem.* 3, 2079–2100. doi:10.4155/fmc.11.159
- Talhout, R., Villa, A., Mark, A. E., and Engberts, J. B. F. N. (2003). Understanding Binding Affinity: A Combined Isothermal Titration Calorimetry/molecular Dynamics Study of the Binding of a Series of Hydrophobically Modified Benzamidinium Chloride Inhibitors to Trypsin. *J. Am. Chem. Soc.* 125, 10570–10579. doi:10.1021/ja034676g
- Vinet, L., and Zhedanov, A. (2011). A 'missing' Family of Classical Orthogonal Polynomials. *J. Phys. A: Math. Theor.* 44, 085201. doi:10.1088/1751-8113/44/8/085201
- Weissman, J. S., Kashi, Y., Fenton, W. A., and Horwich, A. L. (1994). GroEL-mediated Protein Folding Proceeds by Multiple Rounds of Binding and Release of Nonnative Forms. *Cell* 78, 693–702. doi:10.1016/0092-8674(94)90533-9
- Whitesides, G. M., and Krishnamurthy, V. M. (2005). Designing Ligands to Bind Proteins. *Quart. Rev. Biophys.* 38, 385–395. doi:10.1017/S0033583506004240

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Kumar and Jernigan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Dissecting Monomer-Dimer Equilibrium of an RNase P Protein Provides Insight Into the Synergistic Flexibility of 5' Leader Pre-tRNA Recognition

Danyun Zeng¹, Ainur Abzhanova¹, Benjamin P. Brown^{2,3} and Nicholas J. Reiter^{1*}

¹Department of Chemistry, Marquette University, Milwaukee, WI, United States, ²Chemical and Physical Biology Program, Medical Scientist Training Program, Vanderbilt University, Nashville, TN, United States, ³Center for Structural Biology, Vanderbilt University School of Medicine, Nashville, TN, United States

OPEN ACCESS

Edited by:

Vincenzo Venditti,
Iowa State University, United States

Reviewed by:

Ishwar Radhakrishnan,
Northwestern University,
United States
Dipali Sashital,
Iowa State University, United States

*Correspondence:

Nicholas J. Reiter
nicholas.reiter@marquette.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 24 June 2021

Accepted: 18 August 2021

Published: 03 September 2021

Citation:

Zeng D, Abzhanova A, Brown BP and Reiter NJ (2021) Dissecting Monomer-Dimer Equilibrium of an RNase P Protein Provides Insight Into the Synergistic Flexibility of 5' Leader Pre-tRNA Recognition. *Front. Mol. Biosci.* 8:730274. doi: 10.3389/fmolb.2021.730274

Ribonuclease P (RNase P) is a universal RNA-protein endonuclease that catalyzes 5' precursor-tRNA (ptRNA) processing. The RNase P RNA plays the catalytic role in ptRNA processing; however, the RNase P protein is required for catalysis *in vivo* and interacts with the 5' leader sequence. A single P RNA and a P protein form the functional RNase P holoenzyme yet dimeric forms of bacterial RNase P can interact with non-tRNA substrates and influence bacterial cell growth. Oligomeric forms of the P protein can also occur *in vitro* and occlude the 5' leader ptRNA binding interface, presenting a challenge in accurately defining the substrate recognition properties. To overcome this, concentration and temperature dependent NMR studies were performed on a thermostable RNase P protein from *Thermatoga maritima*. NMR relaxation (R_1 , R_2), heteronuclear NOE, and diffusion ordered spectroscopy (DOSY) experiments were analyzed, identifying a monomeric species through the determination of the diffusion coefficients (D) and rotational correlation times (τ_c). Experimental diffusion coefficients and τ_c values for the predominant monomer ($2.17 \pm 0.36 \times 10^{-10} \text{ m}^2/\text{s}$, $\tau_c = 5.3 \text{ ns}$) or dimer ($1.87 \pm 0.40 \times 10^{-10} \text{ m}^2/\text{s}$, $\tau_c = 9.7 \text{ ns}$) protein assemblies at 45°C correlate well with calculated diffusion coefficients derived from the crystallographic P protein structure (PDB 1NZ0). The identification of a monomeric P protein conformer from relaxation data and chemical shift information enabled us to gain novel insight into the structure of the P protein, highlighting a lack of structural convergence of the N-terminus (residues 1–14) in solution. We propose that the N-terminus of the bacterial P protein is partially disordered and adopts a stable conformation in the presence of RNA. In addition, we have determined the location of the 5' leader RNA in solution and measured the affinity of the 5' leader RNA–P protein interaction. We show that the monomer P protein interacts with RNA at the 5' leader binding cleft that was previously identified using X-ray crystallography. Data support a model where N-terminal protein flexibility is stabilized by holoenzyme formation and helps to accommodate the 5' leader region of ptRNA. Taken together, local structural changes of the P protein and the 5' leader RNA provide a means to obtain optimal substrate alignment and activation of the RNase P holoenzyme.

Keywords: ribonuclease P, solution NMR, tRNA processing, substrate recognition, diffusion coefficient, folding, binding, RNA

INTRODUCTION

Ribonuclease (RNase) P is an essential RNA processing enzyme involved in the 5' endonucleolytic cleavage of precursor transfer RNA (ptRNA). It was one of the first identified examples of an RNA-catalytic reaction and exists as a multi-turnover RNA-based enzyme in bacteria, archaea, and eukaryotes (Guerrier-Takada et al., 1983). In bacteria, a large catalytic RNA (P RNA) and a small protein (P protein) component assemble as a holoenzyme complex to recognize and cleave ptRNA. Synergistic molecular recognition of ptRNA substrates by RNase P requires RNA-RNA shape complementarity, intermolecular base pairing, and stabilization of the 5' leader single stranded RNA (ssRNA) region by the P protein component (Kazantsev and Pace 2006; Torres-Larios et al., 2006; Esakova and Krasilnikov 2010; Chan et al., 2013; Mondragon 2013; Klemm et al., 2016; Zhang and Ferré-Damaré 2016; Gray and Gopalan 2020).

Structural studies of the RNase P holoenzyme with tRNA in bacteria and eukaryotes show how the RNA-protein complex functions as a single, monomeric assembly to selectively recognize ptRNA targets (Reiter et al., 2010; Lan et al., 2018; Wu et al., 2018). Interestingly, a cryo-EM structure of *Methanocaldococcus jannaschii* (Mja) RNase P reveals the components can be organized as a dimeric conformation for efficient catalysis (Wan et al., 2019). In all cases, the structures reveal conserved RNase P ribozyme features that are universal, such as a dual T-loop P RNA tertiary motif that recognizes the TΨC-D elbow region of tRNA and conserved P RNA nucleotides that comprise the ribozyme active site (Krasilnikov et al., 2003; Krasilnikov et al., 2004; Kazantsev et al., 2005; Torres-Larios et al., 2005). The collective X-ray and Cryo-EM determined structures suggest a largely pre-organized RNA active site, though a conserved and bulged uridine has been proposed to undergo a conserved dynamic motion that helps to position the substrate and trigger catalytic activation (Kaye et al., 2002; Christian et al., 2006; Reiter et al., 2010; Chen et al., 2011; Martin and Reiter 2017; Lan et al., 2018). These conserved structural and dynamic motions of the P RNA appear to be essential for accurate substrate recognition and the formation of the metallo-ribozyme active site.

In contrast to the P RNA, the RNase P protein has little or no structural similarity across all three domains but its structure is highly conserved in bacteria. The bacterial P protein is required for *in vivo* activity, binds the catalytic P RNA with nanomolar affinity utilizing its arginine-rich "RNR" motif, forms extensive interactions with the 5' leader ptRNA region, increases catalytic efficiency by over two orders of magnitude, and facilitates product release (Peck-Miller and Altman 1991; Tallsjo and Kirsebom 1993; Talbot and Altman 1994; Crary et al., 1998; Kurz et al., 1998; Kurz and Fierke 2002; Buck et al., 2005a; Buck et al., 2005b; Sun et al., 2006; Koutmou et al., 2010b; Reiter et al., 2010; Sun et al., 2010; Koutmou et al., 2011; Reiter et al., 2012; Lin et al., 2016; Niland et al., 2017). The bacterial P protein is a highly stable, rigid scaffold that bind and stabilizes the P RNA and serves as a unique binding cleft for ssRNA. However, the bacterial P protein also contains some intrinsically disordered regions, especially within its N-terminus (Kirsebom 2007). This

N-terminus (residues 1–21) plays a critical role in binding the P RNA as part of the holoenzyme complex and in optimally aligning the 5' leader ptRNA substrate. We sought to better understand the structure and flexibility of the N-terminus of the P protein as well as define the 5' leader RNA binding interaction using solution NMR spectroscopy.

High-resolution crystal structures and biochemical studies of the bacterial RNase P protein by itself have provided some insight into the ssRNA 5' leader binding site, though no structure of an isolated P protein-5' leader RNA complex has been determined to date (Stams et al., 1998; Spitzfaden et al., 2000; Kazantsev et al., 2003; Henkels et al., 2007; Ha et al., 2018). In addition, much less is known about how P protein flexibility contributes to 5' leader binding and catalytic activation. We chose to explore the solution structure, protein flexibility, and 5' leader binding properties of the RNase P protein from *Thermotoga maritima* because extensive structure, biochemical analyses, and small molecule screening studies have been performed on this system (Paul et al., 2001; Krivenko et al., 2002; Buck et al., 2005a; Torres-Larios et al., 2005; Reiter et al., 2010; Reiter et al., 2012; Madrigal-Carrillo et al., 2019). The high-resolution structure of the *Thermotoga maritima* RNase P protein, hence termed P protein, crystallized as a tetramer and its oligomeric state occluded the 5' leader binding through lattice contacts at the P protein's N-terminus (Kazantsev et al., 2003). This propensity towards oligomerization makes structural analysis of the 5' leader- P protein interface recalcitrant and intractable to crystallographic methods.

Here, we describe an NMR-based strategy to overcome oligomerization of the P protein from *Thermotoga maritima* and directly monitor two protein conformers that coexist in solution. The identification of an NMR-derived monomeric model emphasizes the lack of structure within the N-terminus and allowed us to define the 5' leader RNA binding properties of the bacterial P protein.

MATERIALS AND METHODS

Sample Preparation

The ¹⁵N, ¹³C-labeled *T. maritima* RNase P protein sample was prepared following the previous protocol (Zeng et al., 2018). A pGEX4Ta vector containing the *rnpA* gene from *T. maritima* and an N-terminal glutathione S-transferase (GST) fusion protein was transformed and expressed in the BL21 Gold *Escherichia coli* cell strain and cultured in M9 minimal media at 303 K supplemented with ¹⁵N ammonium chloride and/or ¹³C glucose. After lysis sonication in the presence of cOmplete™ protease inhibitor cocktail (Roche), lysate was separated by centrifugation (28,000 g), filtered, and treated with 800 NIH units of thrombin per 40 ml to remove the GST tag from RNase P protein. A denaturation-renaturation purification strategy was applied (Paul et al., 2001; Krivenko et al., 2002; Buck et al., 2005a; Buck et al., 2005b; Reiter et al., 2012). The thrombin-treated lysate was combined with denaturation buffer (50 mM Tris-HCl pH7.5, 4 mM EDTA, 8 M urea) to a final concentration of 5 M urea, and was subject to 15S cation exchange chromatography. Fractions containing the denatured protein were subsequently dialyzed

against refolding buffer (50 mM Tris-HCl pH 7.5, 0.2 mM EDTA, 1 M NaCl) for 1–2 days, upon which the sample was diluted and subject to a second 15S cation exchange column under the identical condition excluding urea. Purified ^{15}N , ^{13}C -labeled RNase P protein fractions were collected, concentrated, and dialyzed against a buffer of 50 mM Tris-HCl pH 7.5, 0.2 mM EDTA.

NMR Spectroscopy

All NMR experiments were conducted using either low (153 μM) or high (466 μM) ^{15}N , ^{13}C RNase P protein concentrations in 20 mM sodium phosphate pH 6.0, 80 mM NaCl, 50 μM 4, 4-dimethyl-4-silapentane-sulfonate (DSS) at 318 K with 10% (v/v) D_2O in a 3 mm Norell® select series NMR tube. NMR spectra were acquired on Bruker Avance-III 600 and 800 MHz and 900 MHz spectrometers, equipped with cryogenic probes, as well as Varian VNMRS 600 and 800 MHz spectrometer equipped with a cryogenic probe. NMR spectra were processed with Topspin 3.5.7 (Bruker Inc.) and NMRPipe (Delaglio et al., 1995), and analyzed with NMRViewJ (One Moon Scientific, Inc.) and CARRA (Keller 2004). ^1H chemical shifts were referenced with respect to internal DSS, and ^{13}C and ^{15}N chemical shifts were referenced indirectly using nuclei-specific gyromagnetic ratios (Wishart et al., 1995).

NMR Assignments of the P Protein Backbone According to Different Concentrations and Temperatures

NMR chemical shift assignments of the protein backbone were determined on separate samples with protein concentrations of 466 and 153 μM , respectively. At the higher protein concentration, backbone assignments were derived from ^1H , ^{15}N -heteronuclear single quantum coherence (HSQC) and a set of traditional triple-resonance experiments, including HNCO, HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, and HN(CA)CO. In addition, backbone and side-chain assignments were carried out using a three-dimensional ^{15}N -NOESY-HSQC experiment (120 ms mixing time). At the lower protein concentration, backbone assignments were confirmed with ^1H , ^{15}N -HSQC, HNCO, HNCA, and HNCACB, comparing and referencing to the assignments of the higher concentration sample. The chemical shift assignments for the 466 μM and the 153 μM sample were deposited to the biological magnetic resonance bank (BMRB) as accession numbers 27307 and 51021, respectively. To investigate the temperature dependence of the backbone assignments of P protein (153 μM), a set of ^1H , ^{15}N -HSQC spectra were acquired with identical parameters at 35 (308), 45 (318), 55 (328), and 65°C (338 K). Additional parameters and data processing, including window function, base-line correction, and linear prediction were identical for all ^1H , ^{15}N -HSQC spectra.

Diffusion Measurements of RNase P Protein

Two $^{13}\text{C}/^{15}\text{N}$ -labeled protein samples with concentrations of 153 and 466 μM were prepared for high-resolution DOSY (diffusion-ordered NMR spectroscopy) experiments. 2D DOSY- ^1H , ^{15}N -HSQC experiments were performed on a

Varian VNMRS 800 MHz spectrometer with a diffusion delay (Δ) of 70 ms and gradient width (δ) of 3 ms at 318 K. Six different gradient strengths (0.9, 9.7, 18.6, 27.4, 36.2, and 45.1 Gauss/cm) were used to measure diffusion rate of the sample with higher concentration; while 8 gradient strengths (0.88, 7.24, 13.6, 20.0, 26.3, 32.7, 39.1, and 45.4 Gauss/cm) were applied to the low concentration sample. Each experiment was acquired with a data matrix of 168 (t_1 , ^{15}N) \times 2048 (t_2 , ^1H) complex points and 64 scans. NMRPipe and NMRViewJ were used for data processing and analysis, respectively. The exact centers of the cross peaks for the selected resonances were identified and the fitting of the results and error were propagated in the analysis of signal decay curves. Reference diffusion coefficients for the RNase P monomer, dimer, and tetramer species were calculated from the high-resolution crystal structure (PDB: 1NZ0) using HYDROPRO and HYDRONMR (García de la Torre et al., 2000; Ortega et al., 2011). Molecule A, A/C, and A-D of PDB 1NZ0 were used for simulations of the monomer, dimer, and tetramer, respectively.

Measurement of Relaxation Parameters

Relaxation parameters were measured at 318 K of the 466 μM P protein on a Bruker Avance-III 800 MHz spectrometer. For R_1 , five relaxation time points were collected at 100, 200, 400, 600, and 1,000 ms. For R_2 , five relaxation time points were collected at 17, 34, 51, 85, and 119 ms. For R_1 and R_2 measurements, a recycle delay of 1 s was used between transitions. Errors were estimated by duplicate measurements using the shortest and longest relaxation time. For heteronuclear NOE measurements, the steady-state ^1H saturation time was 5 s, a recycle delay of 5 s was implemented in the reference experiment, and both reference and NOE measurements were repeated in duplicate. All spectra of relaxation measurements were collected with the data matrix of 300 (t_1 , ^{15}N) \times 1,024 (t_2 , ^1H) points and 32 scans. NMRPipe was used for data processing and data were evaluated using the rate analysis module of NMRViewJ (Johnson and Blevins 1994; Delaglio et al., 1995). To estimate rotational correlation time from R_1 and R_2 values, calculation were made by performing Eq. 1 (Kay et al., 1989):

$$\tau_c \approx \frac{1}{4\pi\nu_N} \sqrt{6 \frac{R_2}{R_1} - 7} \quad (1)$$

where ν_N is the ^{15}N resonance frequency in Hz.

T. maritima RNase P Model Prediction Based on NMR Chemical Shifts

Two sets of NMR chemical shifts were extracted from the high and low protein concentration samples (BMRB #27307 and #51021) and models were generated using chemical-shift (CS-) Rosetta predictive modeling as described elsewhere (Shen et al., 2008; Shen and Bax 2013). The identical number of chemical shift assignments from the “low” and “high” concentration samples were imported to the CS-Rosetta web server to assess predictive modeling results in a direct, comparative manner. NMR chemical shift information was the only experimental data incorporated

into model prediction calculations. A total of 3,000 structural models were generated for each set (low and high P protein concentration) and the top 10 best scoring models were selected. Visualization and analysis of predictive models were performed in Molmol (Koradi et al., 1996).

The Interaction Between RNase P Protein and 5' Leader RNA

NMR titrations investigated the binding process between the RNase P protein and the ptRNA 5' leader. A 7-mer 5' leader sequence of ptRNA (AAGGCGU) was purchased from Dharmacon Co. with purity >95%. Both protein and ptRNA leader were pre-equilibrated in the identical NMR buffer as described above. During titrations, the leader RNA was gradually added with increased molar ratios of 0.2:1, 0.4:1, 1:1, and 2:1 to the 153 μ M protein sample. At each molar ratio, a ^1H , ^{15}N -HSQC spectrum was acquired at 318 K. All data were processed by NMRpipe and analyzed by NMRviewJ. The analysis of amide chemical shift changes ($\Delta\delta$) were calculated following Eq. 2 (Schumann et al., 2007; Collier et al., 2014):

$$\Delta\delta(\text{ppm}) = \sqrt{\Delta\delta(^1\text{H})^2 + (0.152\Delta\delta(^{15}\text{N}))^2} \quad (2)$$

$\Delta\delta$ s of the peaks at the highest leader RNA concentration (5' leader/protein molar ratio = 2:1) were used for chemical shift perturbation (CSP) analysis. Dissociation constants (K_D) for each shifting peak were calculated by fitting the data to Eq. 3:

$$\Delta\delta(\text{ppm}) = \Delta_{\text{max}} \frac{(K_D + [L] + [P]) - \sqrt{(K_D + [L] + [P])^2 - 4[P][L]}}{2[P]} \quad (3)$$

where $\Delta\delta(\text{ppm})$ is the amide chemical shift changes, Δ_{max} the maximum amide chemical shift changes, $[P]$ and $[L]$ the protein and ligand concentrations, respectively. Only defined monomeric chemical shift assignments were used for the K_D calculation. However, the CSP analysis included monomer-only peaks and identified monomer-dimer assigned peaks. In the 153 μ M P protein sample, dimer-only peaks undergo chemical shift changes upon RNA binding but were too weak and unreliable to be included in the analysis.

Molecular Docking of RNase P Protein and 5' tRNA Leader

HADDOCK was used to predict the structure of the RNase P protein-5' ptRNA leader complex, based on the CSP results of NMR titrations (Dominguez et al., 2003). Certain residues which have $\Delta\delta$ s larger than $\Delta\delta_{\text{average}} + \text{STD}_{\Delta\delta}$ were set as active residues, while amides having $\Delta\delta$ s between $\Delta\delta_{\text{average}}$ and $\Delta\delta_{\text{average}} + \text{STD}_{\Delta\delta}$ were set as passive residues. The initial structure of the P protein was the representative structure from the CS-Rosetta calculated ensemble that incorporated the NMR peak assignments at low protein concentration. The initial model of the 7-mer ptRNA leader was generated in Coot (Emsley and Cowtan 2004). The

docking calculations were performed on the HADDOCK web server; generating 127 predicted models using standard HADDOCK settings.

RESULTS

Concentration Dependent Equilibrium of *T. maritima* RNase P Protein in Solution

The functional *T. maritima* ribonuclease P holoenzyme structure contains single RNA (ribozyme) and protein components, yet the large 110 kDa P RNA-only crystallizes as a dimer and the small 14 kDa P protein-only crystal structure (PDB-1NZ0) contains two dimers within its asymmetric unit (Figure 1) (Paul et al., 2001; Kazantsev et al., 2003; Torres-Larios et al., 2005; Reiter et al., 2012). The P protein dimer interface in the crystal structure obscures the 5' leader RNA binding track (Kazantsev et al., 2003) and this is problematic in characterizing the 5' leader pre-tRNA substrate-protein interaction. Preliminary NMR experiments reveal that highly purified *T. maritima* RNase P protein, hence termed P protein, persists as a dimer at 25°C but is a heterogeneous species in solution. Thus, oligomerization is not merely an artifact of crystallization. To define the NMR assignments of an RNA binding competent P protein, ^1H , ^{15}N -HSQC spectra at various temperatures and on samples containing protein concentrations in the range of 153–466 μ M were collected and analyzed. A series of triple-resonance experiments at both high (466 μ M) and low (153 μ M) protein concentrations were also collected and analyzed at 45°C. At P protein concentrations of 153 and 466 μ M, there are always 2 conformations in solution (Figure 2, Supplementary Figure S1). The analysis of various concentration and temperature spectra enabled the identification and assignment of two sets of amide peaks that correspond to two conformations in solution (BMRB #27307 and #51021).

When comparing P protein low (153 μ M) and high (466 μ M) concentrations at 45°C, we observed overall peak intensity changes between the two conformations depending upon the protein concentrations (Figure 2). This suggests a concentration-dependent shift in the equilibrium of the two species in solution. Between 35 to 65°C, similar temperature dependent chemical shifts were observed for both samples but no significant changes occurred in the relative peak intensities for the two conformations (Supplementary Figure S2). This suggests that both conformations are stable over a wide temperature range and that the oligomeric equilibrium of the P protein is concentration dependent and temperature-independent in the range of 35–65°C.

We report high confidence backbone assignments at high (96.5% amides, 99.1% for all C_α and C_β , and 97.4% C' resonances) and low P protein concentrations (93.0% amides, 90.6% for C_α , 70.7% for C_β , and 87.2% C' resonances). Nearly 50% of all residues have different chemical shift assignments in two conformations of the ^1H , ^{15}N -HSQC, thus facilitating the accurate analysis of the distinct conformations of the P protein (Supplementary Table S1, BMRB #27307 and #51021).

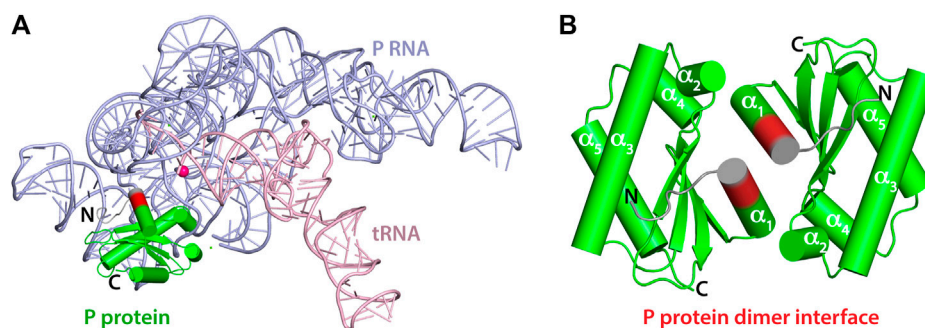


FIGURE 1 | Crystal structures of *T. maritima* RNase P components. **(A)** Crystal structure of the holoenzyme in complex with tRNA (PDB 3Q1Q). RNase P RNA component is shown in light blue, the P protein component in green, and tRNA in light pink. The enzyme active site is denoted by the location of a catalytic magnesium ion (magenta sphere). RNAs are represented as loops (backbones) and sticks (nucleobases) and the P protein is represented as cylinders (α -helix) and arrows (β -sheets). Protein residues 14–17 are highlighted in red to indicate the location of the dimerization interface of the P protein alone (PDB 1NZ0). Residues 8–21 play a critical role in binding the P RNA as part of the holoenzyme complex and in optimally aligning the 5' leader ptRNA substrate. **(B)** 1.2 Å resolution crystal structure of the P protein shown as a dimer, with molecules A and C (PDB 1NZ0). Protein residues 14–17 are highlighted in red and the N-termini in colored grey. Residues 1–15 are oriented differently between **A** and **B**.

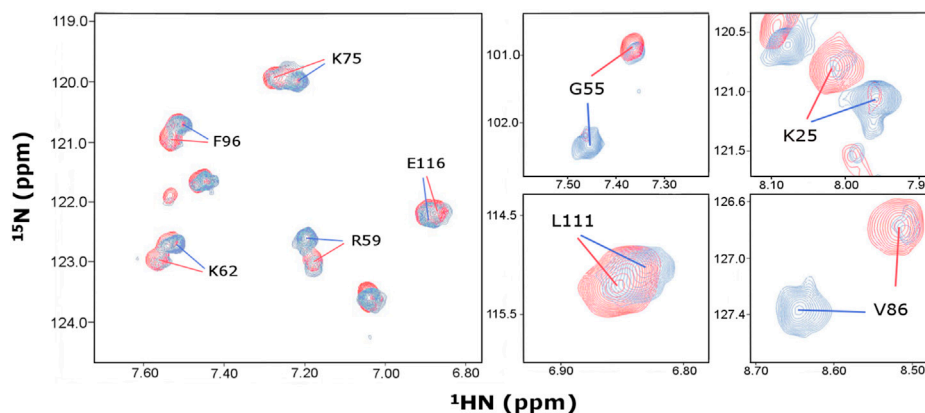


FIGURE 2 | Identification of distinct chemical shift assignments at 466 and 153 μ M P protein concentrations. Overlay of ^1H , ^{15}N -HSQC spectra acquired on samples containing 466 μ M (red) and 153 μ M (blue) P protein concentrations under identical buffer conditions at 318K. Representative regions of the spectra are shown, revealing slow exchange between monomer and dimer conformers in solution. Amino acid single letter code and *T. maritima* RNase P protein numbering corresponds to UniProtKB–Q9X1H4.

Identification of a Monomer and Dimer P Protein

To define oligomerization status of the distinct conformations, 2D-DOSY spectra were acquired at low (153 μ M) and high (466 μ M) P protein concentrations (Table 1). Analysis was performed on 48 residues that contained distinct chemical shifts for the two conformations in the ^1H , ^{15}N -HSQC spectra. Peak intensities of the dominant signal at low or high P protein concentrations were identified and selected for signal decay curve analyses from DOSY experiments. The average diffusion coefficients at 45°C were determined from the analysis of the selected peak intensities against gradient field strengths (Table 1). Experimental diffusion coefficient values of the low (153 μ M) concentration conformer ($D_{\text{low}} = (2.17 \pm 0.36) \times 10^{-10} \text{ m}^2/\text{s}$) and the high (466 μ M) concentration conformer ($D_{\text{high}} = (1.87 \pm 0.40)$

$\times 10^{-10} \text{ m}^2/\text{s}$) are consistent with HYDROPRO and HYDRONMR software-simulated diffusion coefficients of a monomer ($D \sim 1.96 \times 10^{-10} \text{ m}^2/\text{s}$) and dimer ($D \sim 1.54 \times 10^{-10} \text{ m}^2/\text{s}$) P protein conformation (Table 1). The calculated, reference diffusion coefficients from HYDROPRO and HYDRONMR were derived from the high-resolution crystal structure PDB-1NZ0. A value difference of 0.3 for the experimental diffusion data ($D \sim 2.17$ vs. 1.87) compared to the 0.42 value difference in the calculated values ($D \sim 1.96$ vs. 1.54) reflect the fact that both the low and high concentration samples contain mixed monomer/dimer populations. While there is no evidence of higher order oligomerization, the monomeric form predominates at low protein concentration and the dimeric form primarily occurs at higher concentrations. The resolution and calculated error in the NMR-derived diffusion coefficients likely reflect the mixed populations present at both concentrations. Nonetheless, this comparative

TABLE 1 | Experimental and calculated diffusion coefficients (D , $\cdot 10^{-10}$ m²/s) of *T. maritima* RNase P protein.

Method	Conformation	
	153 μ M concentration	466 μ M concentration
DOSY	2.17 \pm 0.36	1.87 \pm 0.40
	Monomer	Dimer Tetramer
HYDROPRO	1.96	1.55 1.08
HYDRONMR	1.93	1.53 1.10

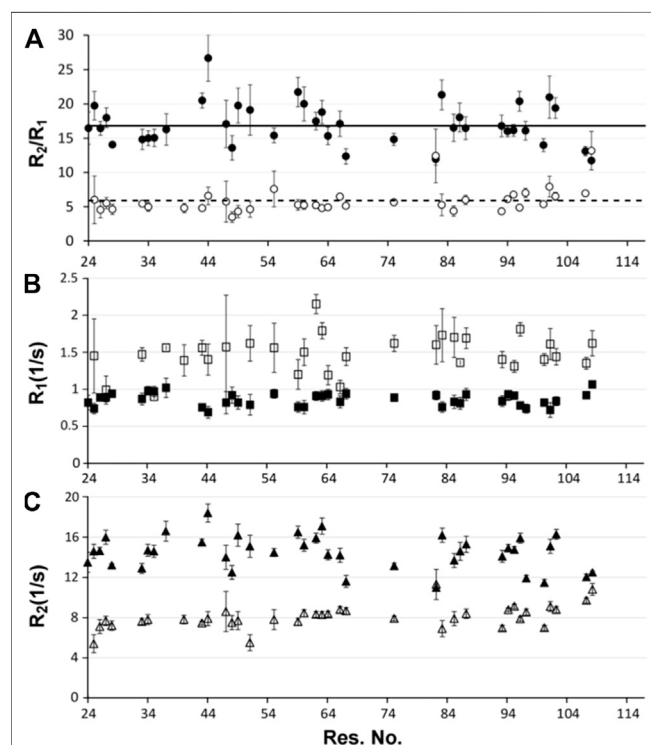


FIGURE 3 | NMR relaxation analysis (A) R_2/R_1 ratios, (B) R_1 and (C) R_2 of distinct *T. maritima* RNase P protein conformers. The dominant P protein conformer at 466 μ M is shown in filled circles, rectangles, and triangles in the three plots, respectively; while the minor conformer is presented as the open/white filled symbols. The R_2/R_1 ratio (A), as well as individual R_1 (B) and R_2 (C) relaxation experiments were performed on a 466 μ M P protein sample at 318K. For each data set, errors represent multiple ^{15}N T_2 and T_1 relaxation experiments in addition to the uncertainty of the exponential fit. The average R_1 , R_2 , and R_2/R_1 ratios of the dominant and minor P protein conformations are indicated as solid and dashed lines, respectively.

diffusion coefficient analysis identifies that the assigned resonances in the low (153 μ M) and high (466 μ M) concentration P protein samples correspond primarily to the monomeric and dimeric species, respectively.

Verification of the identified monomer/dimer species was also determined through the analysis of the R_2/R_1 relaxation data. Despite heterogeneous sample conditions, 48 chemical shifts were extracted from two sets of residues in the spectra. One set of analyzed peaks was representative of the dominant conformer at the high concentration (466 μ M) and the other set of peaks was representative of a minor

conformer that is also present at 466 μ M P protein. The observed minor conformer chemical shifts at 466 μ M are identical to the dominant conformer at 153 μ M P protein concentration. R_1 and R_2 rates were determined for the selected residues of the minor and dominant P protein conformers and their average R_2/R_1 ratios identified as 5.95 and 17.04, respectively, (Figures 3A–C). R_2 values were considerably higher for peaks arising from the dominant conformer in comparison to peaks from the minor conformer (Figures 3A,C). In addition, rotational correlation times (τ_c) (Kay et al., 1989) were determined from R_2/R_1 ratio values, revealing $\tau_{c,low}$ and $\tau_{c,high}$ values of approximately 5.3 and 9.7 ns, respectively. These parameters support a monomer–dimer equilibrium model, where the monomer conformer ($\tau_{c, monomer} \sim 5.3$ ns) predominates at ~ 150 μ M and the dimer species ($\tau_{c, dimer} \sim 9.7$ ns) predominates at higher P protein concentrations. The estimated τ_c values are substantially smaller than a typical 14 or 28 kDa idealized spherical protein at 25°C, but are consistent with the estimated empirical range of correlation times at 45°C (<http://nickanthis.com/tools/tau>). Thus, analysis of two independent NMR methodologies (DOSY and NMR relaxation) both validate that monomer and dimer P protein assemblies coexist in solution. The monomeric conformation predominates at 153 μ M P protein concentrations whereas the dimer species predominates at higher (466 μ M P protein) concentrations.

Predictive Modeling of the 2 Conformations

The two sets of backbone NMR assignments (HN, C_α , and C') based upon BMRB entries #27307 and #51021 were input into CS-Rosetta web server to generate models of the monomer and dimer forms of the P protein in solution. To inspect the structural differences caused only by chemical shifts, the same numbers of chemical shifts from exactly the same assigned resonances were implemented for CS-Rosetta calculations. The ensemble models converge well and reflect the dimeric and monomeric forms of the P protein in solution (Figures 4A,B, respectively). All ensembles include the 10 lowest-energy models and both sets of predicted structures contain near identical backbone folds that are consistent with the crystal structure (PDB: 1NZ0). However, whereas the backbone RMSD of residues 24–117 of the monomer and dimer species compared to PDB-1NZ0 are 0.97 Å and 1.19 Å, respectively, both models show a higher degree of flexibility at the N-terminus (residues 1–23). Both monomeric and dimeric conformers exhibit a lack of structural convergence within the first 23 residues at N terminal of the P protein, with a higher degree of structural variability observed in the monomeric-derived ensemble (Figure 4B).

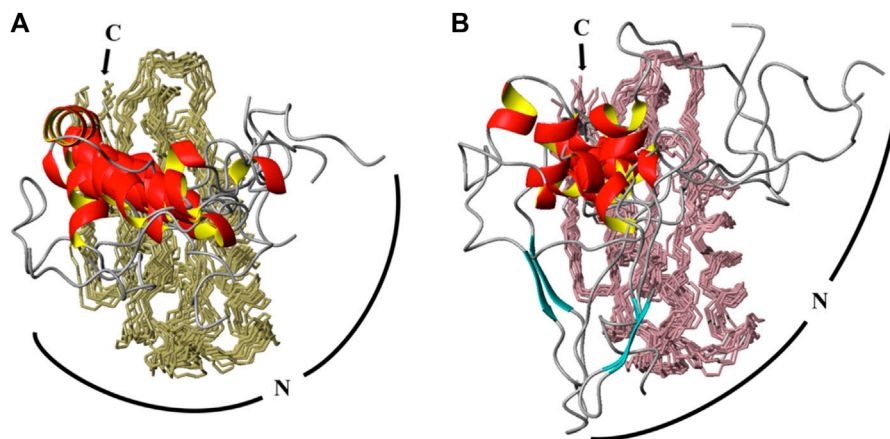


FIGURE 4 | CS-Rosetta calculated ensembles of *T. maritima* P protein at high (466 μ M) and low (153 μ M) concentrations. The same number of chemical shifts arising from the same residues (distributed equally) for identified monomer and dimer conformations were incorporated into CS-Rosetta calculations. The top 10 scored models of each conformation are overlaid in **A** (at 466 μ M) and **B** (at 153 μ M). In both ensembles, 10 structures models are superposed from residue 24 to 117, with backbone regions colored olive (**A**) and light pink (**B**). The N-terminus region (residues 1–23) are presented as loop (gray coil-coiled) and ribbons (red helix) segments. The C-terminus (C) is labeled and a curved line of the N-terminus (N) is included to highlight the predicted amplitudes and flexibility observed in the dimer (**A**) and monomer (**B**) derived ensembles.

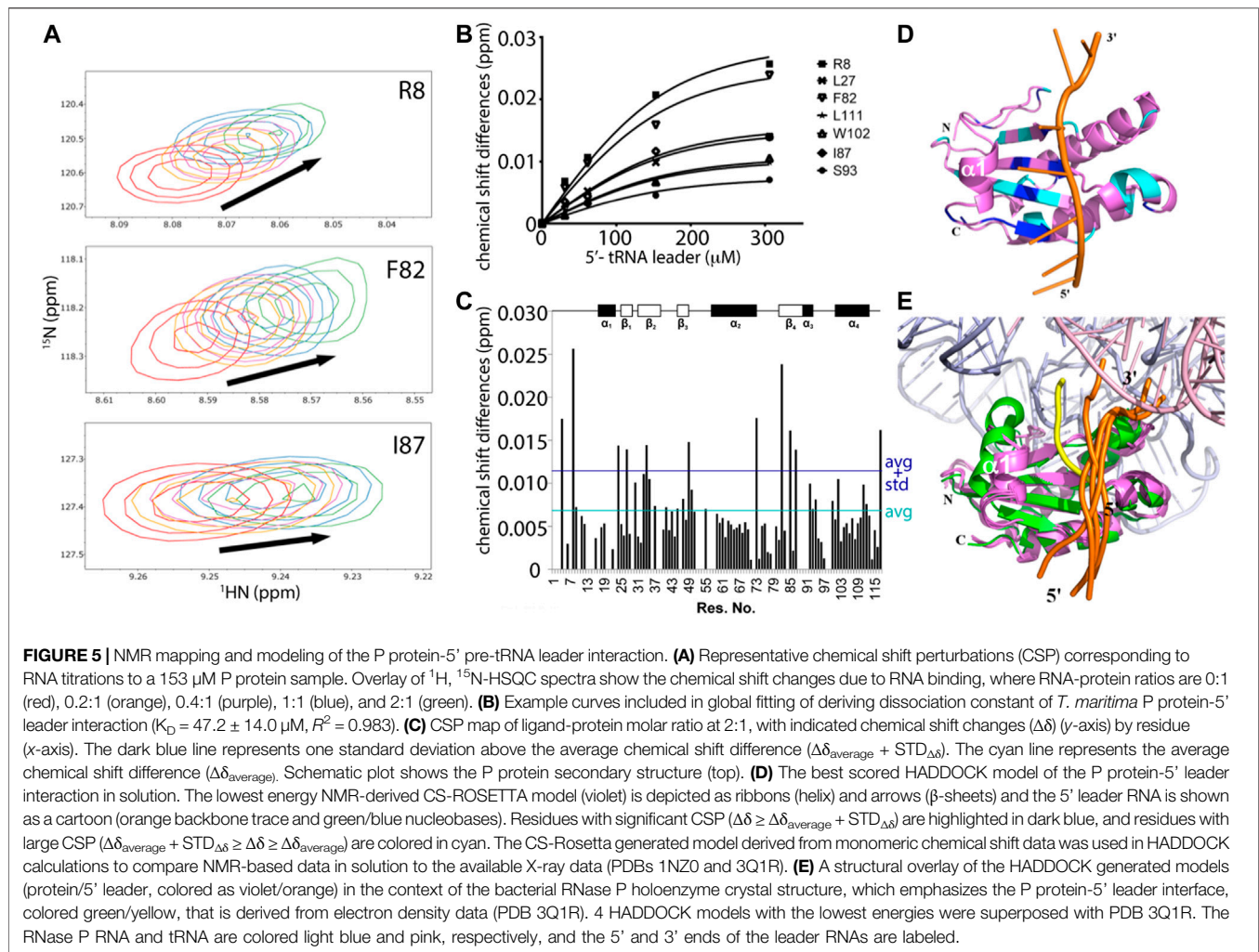
A comparison of NOEs within the N-terminus (residues 8–12, 18–22) and a defined α -helical region (residues 59–68) reflect the extent or lack of structural convergence observed in CS-Rosetta models (**Supplementary Figure S3**). Poorly ordered regions at the N-terminus (residues 8–12) reveal only a few HN-HN ($i, i+1$) NOEs, a partially structured region that contains α -helical character (residues 18–22, α_1 helix) reveals sequential ($i, i+1$) and some medium range NOEs, and a highly ordered region (residues 59–68, α_3 helix) shows HN-HN ($i, i+2$), H_α -HN ($i, i+3$), and H_α -HN ($i, i+4$) medium and long range NOE connectivities. The representative NOEs identified in ^{15}N NOESY-HSQC spectrum highlight the lack of structural convergence in the N-terminus observed in the CS-Rosetta model.

Thus, comparative modeling and analysis of NOE data reveal that the N-terminal P protein is more flexible than other regions of the P protein. The CS-Rosetta data also suggest that the N-terminal region of the monomer conformation is more flexible than the dimer conformer despite using the exact same number of chemical shifts from the same assigned resonances. The α_1 helix (α_1) extends from residue 14–23 in the dimeric assigned CS-Rosetta ensemble with a backbone RMSD of 0.38 Å relative to PDB 1NZ0 (**Figure 4B**). This implies that the dimerization interface in solution is similar to the X-ray structure (PDB 1NZ0) (**Figure 1B**, **Supplementary Figures S4, S5**). In contrast, the monomer assigned CS-Rosetta models reveals a truncated α_1 helix (residues 17–23) that appears intrinsically disordered (RMSD \sim 1.56 Å) relative to 1NZ0. Taken together, this NMR-based data supports a structural mechanism whereby the N-terminal region (residues 1–12) becomes stabilized by the P RNA in the RNase P holoenzyme crystal structure (Reiter et al., 2010) and the transient structure of the N-terminal helix (residues 13–23) helps to optimally align the 5' leader ssRNA substrate region (Niranjanakumari et al., 1998; Zahler et al., 2003; Buck et al., 2005a; Buck et al., 2005b; Sun et al., 2006;

Niranjanakumari et al., 2007; Koutmou et al., 2010b; Reiter et al., 2012; Lin et al., 2016; Niland et al., 2017).

The Binding Mode of P Protein and 5' Pre-tRNA Leader

To understand the mechanism of RNase P holoenzyme activation and ptRNA cleavage, it is essential to study the structure and binding interaction between RNase P protein and 5' ptRNA leader. We chose to work with RNase P from *Thermotoga maritima* because extensive structure, biochemical analyses, and small molecule screening studies have been performed on the RNase P holoenzyme (Paul et al., 2001; Buck et al., 2005a; Torres-Larios et al., 2005; Reiter et al., 2010; Reiter et al., 2012; Madrigal-Carrillo et al., 2019). We determined that the functionally relevant P protein predominates at low (153 μ M) concentrations, enabling NMR titrations to probe the binding interactions of the P protein and 5' leader RNA. Chemical shift perturbation (CSP) analysis was performed when the 5' leader RNA (5'- AAGGCGU-3') was titrated into the P protein to a ratio of 2:1. Higher RNA:protein titration ratios were attempted >4:1 yet were followed by rapid sample precipitation, prohibiting the accurate collection of reliable chemical shift information. Nonetheless, CSP analysis of the protein-RNA titration was monitored via ^1H , ^{15}N -HSQC experiments, focusing on 57 assignments attributed to the monomer, 28 peaks corresponding to both monomer and dimer conformers, and 8 resonances attributed solely to the dimer species (**Supplementary Figure S6**). Dimer-identified peaks are minor at the 153 μ M P protein concentrations (**Figure 2**); yet detectable minor chemical shift differences ($\Delta\delta$) upon ligand binding were observed for both monomer and dimer species (**Supplementary Figure S7**). Specifically, residues Q28 and F82 represent dimer peaks that were shifted



upon RNA titration at 153 μM P protein concentrations. Both Q28 and F82 HN resonances are structurally accessible to bind RNA in the monomer and dimer conformers yet only Q28 exists at a functionally relevant RNA binding interface. While it is possible that further RNA addition may shift the equilibrium from a dimer species towards the monomeric conformation, it is unclear from the available NMR data whether RNA binding substantially alters the dimer-monomer conformational equilibrium.

Nonetheless, the magnitude of changes in peak positions induced by RNA binding correlate to weak protein-RNA interactions that occur in fast exchange on the NMR timescale (Figure 5A). The dissociation constant (K_D) was determined by examining the individual change in peak positions ($\Delta\delta$ s) against ligand concentrations for residues attributed to only the monomer P protein. 16 data sets were fitted and then combined for a global fitting K_D determination, yielding an overall apparent K_D^{global} of $47.2 \pm 14.0 \mu\text{M}$ (Figure 5B). Significant chemical shifts were observed exclusively in the identified monomer assignments (11 residues, S4, R8, G24, L27, V34, V49, R73, ^{82}F , V85, I87, and G117), where significant shifts imply $\Delta\delta \geq \Delta\delta_{\text{average}} + \text{STD}_{\Delta\delta}$ (Figure 5C).

Nearly all mapped protein residues undergo fast exchange kinetics upon RNA binding except for amide residues of S26 and V99, which are selectively line broadened and undergo millisecond exchange kinetics. The observed different binding modes may be due to the 5' leader RNA interactions with the monomer and a heterogeneous monomer-dimer species, thus complicating interpretation. For this reason, only residues that have defined monomeric chemical shift assignments and undergoing fast exchange were included in the dissociation constant calculation. The distributed RNA binding P protein residues were mapped on one of the CS-Rosetta derived model and, as expected, residues with the largest changes in peak positions are located across the β -sheet binding cleft (Figure 5D). These NMR titration experiments validate the 5' leader-binding interface previously identified in the *T. maritima* RNase P holoenzyme-tRNA complex (Reiter et al., 2010).

Based on CSP results, HADDOCK was performed to generate an ensemble of the P protein monomer-5' pre-tRNA leader interaction in solution (Figures 5D,E). Chemical shift perturbation results were included as ambiguous restraints between the individual P protein residues of the 5' leader RNA. Monomeric P protein residues exhibiting peak changes

located near the central β -sheet cleft were selected for docking, consistent with *T. maritima* RNase P structure and biochemical studies (Reiter et al., 2010; Reiter et al., 2012). Models generated were solely dependent upon CSP data and emphasize the flexibility of the P protein in accommodating a broad array of 5' leader RNA nucleotides.

DISCUSSION

Sample heterogeneity, oligomerization, and the identification of a functionally relevant conformer represent long-standing challenges in the analysis of biochemical data. Dimerization and higher order oligomers of RNA binding proteins can function as essential features for splicing regulation, post-transcriptional processing, and RNA biogenesis, or they can represent aberrant pathways prone to aggregation that can dominate the pathology of a disease (Lagier-Tourenne et al., 2010; Couthouis et al., 2011; Prusty et al., 2017; Montalbano et al., 2020; Montemayor et al., 2020). In bacterial RNase P, the issue of a functional dimeric holoenzyme has been extensively discussed (Fang et al., 2001; Barrera et al., 2002; Barrera and Pan 2004; Buck et al., 2005a; Niland et al., 2017), though it has been structurally validated that the RNA-protein holoenzyme complex in bacteria and eukaryotes function as single, monomeric assemblies to perform ptRNA recognition and catalysis. The propensity for dimer formation is apparent in some bacteria P RNA and P protein crystal structures, potentially masking a functionally relevant conformation in solution (Kazantsev et al., 2003; Torres-Larios et al., 2005). Specifically, the oligomeric state of the P protein crystal structures can block the 5' leader RNA binding interface, making it difficult to obtain atomic level insight into the RNA binding interface of the bacterial RNase P holoenzyme.

Here, we developed an NMR-based method to overcome oligomerization of the P protein from *Thermotoga maritima* and investigated the RNA binding interface in solution. Diffusion coefficient and NMR relaxation experiments independently validate a concentration-dependent monomeric P protein in solution. In addition, the determination of distinct chemical shift assignments for the monomer and the dimer conformations indicate that monomer-dimer equilibrium undergoes slow exchange kinetics on the NMR timescale. A comparison of the CS-Rosetta predicted structures from chemical shifts from the monomer show that the N-terminus of the P protein has increased flexibility in the absence of RNA ligand or dimer formation, with residues 14–17 no longer part of a stable helix and lacking structural convergence. This conformational flexibility within the N-terminus is consistent with previous NMR relaxation studies of the P protein (Spitzfaden et al., 2000; Henkels et al., 2007). This suggests that the N-terminus (residues 1–17) represents a disordered region of the *T. maritima* P protein in the absence of P RNA or the ptRNA substrate.

In addition, NMR-monitored titration of a short 7-mer 5' leader RNA show that the largest changes in peak positions due to RNA binding occur across the β -sheet binding cleft that correspond to the 5' leader binding site previously identified via x-ray crystallography. Structure prediction of the P protein-5' leader RNA complex based on experimental CSP-NMR analysis

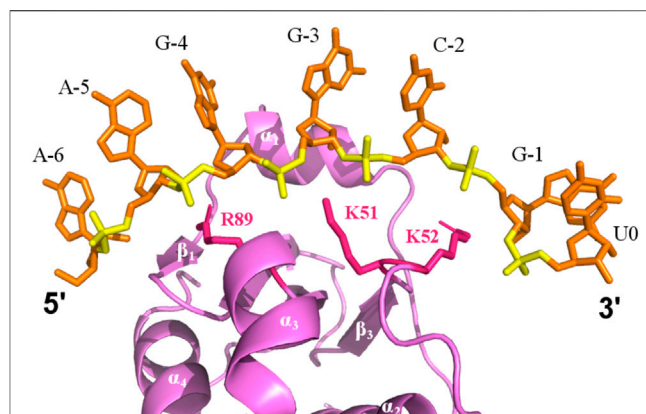


FIGURE 6 | Electrostatic binding interface of a P protein-5' leader complex in solution. The predicted structure is the best-scored model calculated by HADDOCK. The 5' leader (orange sticks) lies along the β -sheet cleft of P protein (violet). Electrostatic interactions between positive charged residues of the P protein (sticks, pink) are labeled and are positioned close to the phosphate groups of the 5' leader (yellow).

indicate that complex formation is largely stabilized through electrostatic interactions between electropositive amino acid side chains and the RNA phosphate backbone (Figure 6). Taken together, and in conjunction with previous structural studies, these results support a model where flexibility of the N-terminus and $\alpha 1$ helix of the P protein can become stabilized through formation of a holoenzyme complex (Henkels et al., 2007; Reiter et al., 2010; Reiter et al., 2012). This P protein N-terminus flexibility may not only be important during holoenzyme assembly and promoting optimal alignment of the 5' leader of the pre-tRNA substrate, but may also contribute to product release of the cleaved tRNA product.

Due to the large size of the functional enzyme and potential sample heterogeneity in some RNase P systems, few solution NMR studies exist of RNase P components (Schmitz and Tinoco 2000; Spitzfaden et al., 2000; Getz et al., 2007; Henkels et al., 2007; Koutmou et al., 2010a; Zeng et al., 2018). Our NMR-based study provides an avenue for dissecting the RNA interactions of a distinguishable monomeric P protein in solution; however, there are some limitations of this approach. Additional insight into the conformational changes of the structure and P protein-5' leader interface could be gleaned through pH-titration experiments, the addition of osmolytes, the alteration of salt concentrations, and the application of high pressure NMR. All of these sample optimization strategies could be effective in controlling the oligomeric state of the P protein or RNase P holoenzyme, and it is possible that the observed dimer interactions are unique to the *T. maritima* P protein (Supplementary Figure S4, S5). Nonetheless, the ability to distinguish between monomer and dimer conformations demonstrates the utility of NMR and provides an excellent starting point for additional sample optimization, helping to further shift the equilibrium completely towards a monomeric P protein conformation. Another limitation of the present NMR study is that only amide chemical shift changes were monitored in NMR relaxation studies and to map the RNA binding site, yet it appears that protein side chain flexibility is

critical to understand RNA binding specificity, with side chain electrostatic stabilization occurring along the β -sheet binding cleft and potential hydrophobic interactions with residues within the $\alpha 1$ helix. Protein side chain-RNA interactions through CSP analysis or through intermolecular protein-RNA NOE experiments would provide atomic-level insight into the potential nucleobase binding specificity within the 5' leader RNA region.

Through combined chemical shift perturbation analysis and NMR-based structure prediction studies, we have illuminated the N-terminal structural flexibility of a bacterial P protein in solution and validated the 5' leader RNA binding interface. Consistent with previous structural and biochemical studies, we confirm that a series of weak, electrostatic-based interactions along the β -sheet binding cleft help to explain how the P protein accommodates different 5' leader single stranded RNAs and can optimally align a variety of RNA substrates. Dissecting conformational heterogeneity within the N-terminus and monitoring RNA-side chain interactions of the *T. maritima* P protein serve as key next steps towards understanding how intrinsically disordered regions contribute to RNA binding specificity and the activation of the RNase P holoenzyme.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the BMRB online repository (accession numbers 27307 and 51021).

REFERENCES

- Barrera, A., Fang, X., Jacob, J., Casey, E., Thiyagarajan, P., and Pan, T. (2002). Dimeric and Monomeric *Bacillus Subtilis* RNase P Holoenzyme in the Absence and Presence of Pre-tRNA Substrates. *Biochemistry* 41, 12986–12994. doi:10.1021/bi020416k
- Barrera, A., and Pan, T. (2004). Interaction of the *Bacillus Subtilis* RNase P with the 30S Ribosomal Subunit. *RNA* 10, 482–492. doi:10.1261/rna.5163104
- Buck, A. H., Dalby, A. B., Poole, A. W., Kazantsev, A. V., and Pace, N. R. (2005a). Protein Activation of a Ribozyme: the Role of Bacterial RNase P Protein. *EMBO J.* 24, 3360–3368. doi:10.1038/sj.emboj.7600805
- Buck, A. H., Kazantsev, A. V., Dalby, A. B., and Pace, N. R. (2005b). Structural Perspective on the Activation of RNase P RNA by Protein. *Nat. Struct. Mol. Biol.* 12, 958–964. doi:10.1038/nsmb1004
- Chan, C. W., Chetani, B., and Mondragón, A. (2013). Structure and Function of the T-loop Structural Motif in Noncoding RNAs WIREs *RNA* 4. Wiley Interdiscip Rev *RNA*, 507–522. doi:10.1002/wrna.1175
- Chen, W.-Y., Xu, Y., Cho, I.-M., Oruganti, S. V., Foster, M. P., and Gopalan, V. (2011). Cooperative RNP Assembly: Complementary rescue of Structural Defects by Protein and RNA Subunits of Archaeal RNase P. *J. Mol. Biol.* 411, 368–383. doi:10.1016/j.jmb.2011.05.012
- Christian, E. L., Smith, K. M., Perera, N., and Harris, M. E. (2006). The P4 Metal Binding Site in RNase P RNA Affects Active Site Metal Affinity through Substrate Positioning. *RNA* 12, 1463–1467. doi:10.1261/rna.158606
- Collier, S. E., Voehler, M., Peng, D., Ohi, R., Gould, K. L., Reiter, N. J., et al. (2014). Structural and Functional Insights into the N-Terminus of *Schizosaccharomyces pombe* Cdc5. *Biochemistry* 53, 6439–6451. doi:10.1021/bi5008639
- Couthouis, J., Hart, M. P., Shorter, J., DeJesus-Hernandez, M., Erion, R., Oristano, R., et al. (2011). A Yeast Functional Screen Predicts New Candidate ALS Disease Genes. *Proc. Natl. Acad. Sci.* 108, 20881–20890. doi:10.1073/pnas.1109434108
- Crary, S. M., Niranjankumari, S., and Fierke, C. A. (1998). The Protein Component of *Bacillus Subtilis* Ribonuclease P Increases Catalytic Efficiency

AUTHOR CONTRIBUTIONS

DZ: Conceptualization, Data curation, Formal analysis, Writing-original draft, Writing-reviewing and editing, AA: Data curation, BB: Conceptualization, Writing-reviewing and editing, NR Conceptualization, Data curation, Formal analysis, Funding acquisition, Writing-original draft, Writing-review and editing.

FUNDING

We thank S. Cai, M. Voehler, and M. Tonelli for NMR assistance, and Q. Timerghazin for molecular modeling discussions. NR acknowledges the NIH (GM120572) and start-up funds provided by Marquette University's Department of Chemistry. AA acknowledges fellowship support from Marquette University's Department of Chemistry. BB is supported through the NIH by a Ruth L. Kirschstein NRSA fellowship (F30DK118774).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.730274/full#supplementary-material>

- by Enhancing Interactions with the 5' Leader Sequence of Pre-tRNA^{asp}. *Biochemistry* 37, 9409–9416. doi:10.1021/bi980613c
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* 6, 277–293. doi:10.1007/BF00197809
- Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003). HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* 125, 1731–1737. doi:10.1021/ja026939x
- Emsley, P., and Cowtan, K. (2004). Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. D Biol. Cryst.* 60, 2126–2132. doi:10.1107/s0907444904019158
- Esakova, O., and Krasilnikov, A. S. (2010). Of Proteins and RNA: the RNase P/MRP Family. *RNA* 16, 1725–1747. doi:10.1261/rna.2214510
- Fang, X.-W., Yang, X.-J., Littrell, K., Niranjankumari, S., Thiyagarajan, P., Fierke, C. A., et al. (2001). The *Bacillus Subtilis* RNase P Holoenzyme Contains Two RNase P RNA and Two RNase P Protein Subunits. *RNA* 7, 233–241. doi:10.1017/s1355838201001352
- García de la Torre, J., Huertas, M. L., and Carrasco, B. (2000). HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations. *J. Magn. Reson.* 147, 138–146. doi:10.1006/jmre.2000.2170
- Getz, M. M., Andrews, A. J., Fierke, C. A., and Al-Hashimi, H. M. (2007). Structural Plasticity and Mg²⁺ Binding Properties of RNase P P4 from Combined Analysis of NMR Residual Dipolar Couplings and Motionally Decoupled Spin Relaxation. *RNA* 13, 251–266. doi:10.1261/rna.264207
- Gray, M. W., and Gopalan, V. (2020). Piece by Piece: Building a Ribozyme. *J. Biol. Chem.* 295, 2313–2323. doi:10.1074/jbc.rev119.009929
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA Moiety of Ribonuclease P Is the Catalytic Subunit of the Enzyme. *Cell* 35, 849–857. doi:10.1016/0092-8674(83)90117-4
- Ha, L., Colquhoun, J., Noinaj, N., Das, C., Dunman, P. M., and Flaherty, D. P. (2018). Crystal Structure of the Ribonuclease-P-Protein Subunit from *Staphylococcus aureus*. *Acta Cryst. Sect F* 74, 632–637. doi:10.1107/s2053230x18011512

- Henkels, C. H., Chang, Y.-C., Chamberlin, S. I., and Oas, T. G. (2007). Dynamics of Backbone Conformational Heterogeneity in *Bacillus Subtilis* Ribonuclease P Protein. *Biochemistry* 46, 15062–15075. doi:10.1021/bi701425n
- Johnson, B. A., and Blevins, R. A. (1994). NMR View: A Computer Program for the Visualization and Analysis of NMR Data. *J. Biomol. NMR* 4, 603–614. doi:10.1007/bf00404272
- Kay, L. E., Torchia, D. A., and Bax, A. (1989). Backbone Dynamics of Proteins as Studied by Nitrogen-15 Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* 28, 8972–8979. doi:10.1021/bi00449a003
- Kaye, N. M., Zahler, N. H., Christian, E. L., and Harris, M. E. (2002). Conservation of Helical Structure Contributes to Functional Metal Ion Interactions in the Catalytic Domain of Ribonuclease P RNA. *J. Mol. Biol.* 324, 429–442. doi:10.1016/s0022-2836(02)01094-x
- Kazantsev, A. V., Krivenko, A. A., Harrington, D. J., Carter, R. J., Holbrook, S. R., Adams, P. D., et al. (2003). High-resolution Structure of RNase P Protein from *Thermotoga Maritima*. *Proc. Natl. Acad. Sci.* 100, 7497–7502. doi:10.1073/pnas.0932597100
- Kazantsev, A. V., Krivenko, A. A., Harrington, D. J., Holbrook, S. R., Adams, P. D., and Pace, N. R. (2005). Crystal Structure of a Bacterial Ribonuclease P RNA. *Proc. Natl. Acad. Sci.* 102, 13392–13397. doi:10.1073/pnas.0506662102
- Kazantsev, A. V., and Pace, N. R. (2006). Bacterial RNase P: a New View of an Ancient Enzyme. *Nat. Rev. Microbiol.* 4, 729–740. doi:10.1038/nrmicro1491
- Keller, R. L. J. (2004). *The Computer Aided Resonance Assignment Tutorial*. Zürich, Switzerland: Cantina Verlag, The Swiss Federal Institute of Technology.
- Kirsebom, L. (2007). RNase P RNA Mediated Cleavage: Substrate Recognition and Catalysis. *Biochimie* 89, 1183–1194. doi:10.1016/j.biochi.2007.05.009
- Klemm, B. P., Wu, N., Chen, Y., Liu, X., Kaitany, K. J., Howard, M. J., et al. (2016). The Diversity of Ribonuclease P: Protein and RNA Catalysts with Analogous Biological Functions. *Biomolecules* 6, 26. doi:10.3390/biom6020027
- Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: a Program for Display and Analysis of Macromolecular Structures. *J. Mol. Graphics* 14, 51–55. doi:10.1016/0263-7855(96)00009-4
- Koutmou, K. S., Casiano-Negroni, A., Getz, M. M., Pazicni, S., Andrews, A. J., Penner-Hahn, J. E., et al. (2010a). NMR and XAS Reveal an Inner-Sphere Metal Binding Site in the P4 helix of the Metallo-Ribozyme Ribonuclease P. *Proc. Natl. Acad. Sci.* 107, 2479–2484. doi:10.1073/pnas.0906319107
- Koutmou, K. S., Day-Storms, J. J., and Fierke, C. A. (2011). The RNR Motif of B. Subtilis RNase P Protein Interacts with Both PRNA and Pre-tRNA to Stabilize an Active Conformer. *RNA* 17, 1225–1235. doi:10.1261/rna.2742511
- Koutmou, K. S., Zahler, N. H., Kurz, J. C., Campbell, F. E., Harris, M. E., and Fierke, C. A. (2010b). Protein-Precursor tRNA Contact Leads to Sequence-specific Recognition of 5' Leaders by Bacterial Ribonuclease P. *J. Mol. Biol.* 396, 195–208. doi:10.1016/j.jmb.2009.11.039
- Krasilnikov, A. S., Xiao, Y., Pan, T., and Mondragón, A. (2004). Basis for Structural Diversity in Homologous RNAs. *Science* 306, 104–107. doi:10.1126/science.1101489
- Krasilnikov, A. S., Yang, X., Pan, T., and Mondragón, A. (2003). Crystal Structure of the Specificity Domain of Ribonuclease P. *Nature* 421, 760–764. doi:10.1038/nature01386
- Krivenko, A. A., Kazantsev, A. V., Adamidi, C., Harrington, D. J., and Pace, N. R. (2002). Expression, Purification, Crystallization and Preliminary Diffraction Analysis of RNase P Protein from *Thermotoga Maritima*. *Acta Crystallogr. D Biol. Cryst.* 58, 1234–1236. doi:10.1107/s0907444902007965
- Kurz, J. C., and Fierke, C. A. (2002). The Affinity of Magnesium Binding Sites in the *Bacillus subtilis* RNase P-Pre-tRNA Complex Is Enhanced by the Protein Subunit†. *Biochemistry* 41, 9545–9558. doi:10.1021/bi025553w
- Kurz, J. C., Niranjanakumari, S., and Fierke, C. A. (1998). Protein Component of *Bacillus subtilis* RNase P Specifically Enhances the Affinity for Precursor-tRNA^{Asp}†. *Biochemistry* 37, 2393–2400. doi:10.1021/bi972530m
- Lagier-Tourenne, C., Polymenidou, M., and Cleveland, D. W. (2010). TDP-43 and FUS/TLS: Emerging Roles in RNA Processing and Neurodegeneration. *Hum. Mol. Genet.* 19, R46–R64. doi:10.1093/hmg/ddq137
- Lan, P., Tan, M., Zhang, Y., Niu, S., Chen, J., Shi, S., et al. (2018). Structural Insight into Precursor tRNA Processing by Yeast Ribonuclease P. *Science* 362, 657–667. doi:10.1126/science.aat6678
- Lin, H.-C., Zhao, J., Niland, C. N., Tran, B., Jankowsky, E., and Harris, M. E. (2016). Analysis of the RNA Binding Specificity Landscape of C5 Protein Reveals Structure and Sequence Preferences that Direct RNase P Specificity. *Cel Chem. Biol.* 23, 1271–1281. doi:10.1016/j.chembiol.2016.09.002
- Madrigal-Carrillo, E.-A., Díaz-Tufinio, C.-A., Santamaría-Suárez, H.-A., Arciniega, M., and Torres-Larios, A. (2019). A Screening Platform to Monitor RNA Processing and Protein-RNA Interactions in Ribonuclease P Uncovers a Small Molecule Inhibitor. *Nucleic Acids Res.* 47, 6425–6438. doi:10.1093/nar/gkz285
- Martin, W. J., and Reiter, N. J. (2017). Structural Roles of Noncoding RNAs in the Heart of Enzymatic Complexes. *Biochemistry* 56, 3–13. doi:10.1021/acs.biochem.6b01106
- Mondragón, A. (2013). Structural Studies of RNase P. *Annu. Rev. Biophys.* 42, 537–557. doi:10.1146/annurev-biophys-083012-130406
- Montalbano, M., McAllen, S., Puangmalai, N., Sengupta, U., Bhatt, N., Johnson, O. D., et al. (2020). RNA-binding Proteins Musashi and Tau Soluble Aggregates Initiate Nuclear Dysfunction. *Nat. Commun.* 11, 4305. doi:10.1038/s41467-020-18022-6
- Montemayor, E. J., Virta, J. M., Hayes, S. M., Nomura, Y., Brow, D. A., and Butcher, S. E. (2020). Molecular Basis for the Distinct Cellular Functions of the Lsm1-7 and Lsm2-8 Complexes. *RNA* 26, 1400–1413. doi:10.1261/rna.075879.120
- Niland, C. N., Anderson, D. R., Jankowsky, E., and Harris, M. E. (2017). The Contribution of the C5 Protein Subunit of *Escherichia coli* Ribonuclease P to Specificity for Precursor tRNA Is Modulated by Proximal 5' Leader Sequences. *RNA* 23, 1502–1511. doi:10.1261/rna.056408.116
- Niranjanakumari, S., Day-Storms, J. J., Ahmed, M., Hsieh, J., Zahler, N. H., Venters, R. A., et al. (2007). Probing the Architecture of the B. Subtilis RNase P Holoenzyme Active Site by Cross-Linking and Affinity Cleavage. *RNA* 13, 521–535. doi:10.1261/rna.308707
- Niranjanakumari, S., Stams, T., Crary, S. M., Christianson, D. W., and Fierke, C. A. (1998). Protein Component of the Ribozyme Ribonuclease P Alters Substrate Recognition by Directly Contacting Precursor tRNA. *Proc. Natl. Acad. Sci.* 95, 15212–15217. doi:10.1073/pnas.95.26.15212
- Ortega, A., Amorós, D., and García de la Torre, J. (2011). Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophysical J.* 101, 892–898. doi:10.1016/j.bpj.2011.06.046
- Paul, R., Lazarev, D., and Altman, S. (2001). Characterization of RNase P from *Thermotoga Maritima*. *Nucleic Acids Res.* 29, 880–885. doi:10.1093/nar/29.4.880
- Peck-Miller, K. A., and Altman, S. (1991). Kinetics of the Processing of the Precursor to 4-5 S RNA, a Naturally Occurring Substrate for RNase P from *Escherichia coli*. *J. Mol. Biol.* 221, 1–5. doi:10.1016/0022-2836(91)80194-y
- Prusty, A. B., Meduri, R., Prusty, B. K., Vanselow, J., Schlosser, A., and Fischer, U. (2017). Impaired Spliceosomal UsnRNP Assembly Leads to Sm mRNA Down-Regulation and Sm Protein Degradation. *J. Cel Biol* 216, 2391–2407. doi:10.1083/jcb.201611108
- Reiter, N. J., Osterman, A. K., and Mondragón, A. (2012). The Bacterial Ribonuclease P Holoenzyme Requires Specific, Conserved Residues for Efficient Catalysis and Substrate Positioning. *Nucleic Acids Res.* 40, 10384–10393. doi:10.1093/nar/gks744
- Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragón, A. (2010). Structure of a Bacterial Ribonuclease P Holoenzyme in Complex with tRNA. *Nature* 468, 784–789. doi:10.1038/nature09516
- Schmitz, M., and Tinoco, I., Jr (2000). Solution Structure and Metal-Ion Binding of the P4 Element from Bacterial RNase P RNA. *RNA* 6, 1212–1225. doi:10.1017/s1355838200000881
- Schumann, F. H., Riepl, H., Maurer, T., Gronwald, W., Neidig, K.-P., and Kalbitzer, H. R. (2007). Combined Chemical Shift Changes and Amino Acid Specific Chemical Shift Mapping of Protein-Protein Interactions. *J. Biomol. NMR* 39, 275–289. doi:10.1007/s10858-007-9197-z
- Shen, Y., and Bax, A. (2013). Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *J. Biomol. NMR* 56, 227–241. doi:10.1007/s10858-013-9741-y
- Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G., et al. (2008). Consistent Blind Protein Structure Generation from NMR Chemical Shift Data. *Proc. Natl. Acad. Sci.* 105, 4685–4690. doi:10.1073/pnas.0800256105
- Spitzfaden, C., Nicholson, N., Jones, J. J., Guth, S., Lehr, R., Prescott, C. D., et al. (2000). The Structure of Ribonuclease P Protein from *Staphylococcus aureus*

- Reveals a Unique Binding Site for Single-Stranded RNA. *J. Mol. Biol.* 295, 105–115. doi:10.1006/jmbi.1999.3341
- Stams, T., Niranjanakumari, S., Fierke, C. A., and Christianson, D. W. (1998). Ribonuclease P Protein Structure: Evolutionary Origins in the Translational Apparatus. *Science* 280, 752–755. doi:10.1126/science.280.5364.752
- Sun, L., Campbell, F. E., Yandek, L. E., and Harris, M. E. (2010). Binding of C5 Protein to P RNA Enhances the Rate Constant for Catalysis for P RNA Processing of Pre-tRNAs Lacking a Consensus G(+1)/C(+72) Pair. *J. Mol. Biol.* 395, 1019–1037. doi:10.1016/j.jmb.2009.11.027
- Sun, L., Campbell, F. E., Zahler, N. H., and Harris, M. E. (2006). Evidence that Substrate-specific Effects of C5 Protein lead to Uniformity in Binding and Catalysis by RNase P. *EMBO J.* 25, 3998–4007. doi:10.1038/sj.emboj.7601290
- Talbot, S. J., and Altman, S. (1994). Gel Retardation Analysis of the Interaction between C5 Protein and M1 RNA in the Formation of the Ribonuclease P Holoenzyme from *Escherichia coli*. *Biochemistry* 33, 1399–1405. doi:10.1021/bi00172a016
- Tallsjö, A., and Kirsebom, L. A. (1993). Product Release Is a Rate-Limiting Step during Cleavage by the Catalytic RNA Subunit of *Escherichia coli* RNase P. *Nucl. Acids Res.* 21, 51–57. doi:10.1093/nar/21.1.51
- Torres-Larios, A., Swinger, K. K., Krasilnikov, A. S., Pan, T., and Mondragón, A. (2005). Crystal Structure of the RNA Component of Bacterial Ribonuclease P. *Nature* 437, 584–587. doi:10.1038/nature04074
- Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragón, A. (2006). Structure of Ribonuclease P - a Universal Ribozyme. *Curr. Opin. Struct. Biol.* 16, 327–335. doi:10.1016/j.sbi.2006.04.002
- Wan, F., Wang, Q., Tan, J., Tan, M., Chen, J., Shi, S., et al. (2019). Cryo-electron Microscopy Structure of an Archaeal Ribonuclease P Holoenzyme. *Nat. Commun.* 10, 2617. doi:10.1038/s41467-019-10496-3
- Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., et al. (1995). ¹H, ¹³C and ¹⁵N Chemical Shift Referencing in Biomolecular NMR. *J. Biomol. NMR* 6, 135–140. doi:10.1007/bf00211777
- Wu, J., Niu, S., Tan, M., Huang, C., Li, M., Song, Y., et al. (2018). Cryo-EM Structure of the Human Ribonuclease P Holoenzyme. *Cell* 175, 1393–1404. e1311. doi:10.1016/j.cell.2018.10.003
- Zahler, N. H., Christian, E. L., and Harris, M. E. (2003). Recognition of the 5' Leader of Pre-tRNA Substrates by the Active Site of Ribonuclease P. *RNA* 9, 734–745. doi:10.1261/rna.5220703
- Zeng, D., Brown, B. P., Voehler, M. W., Cai, S., and Reiter, N. J. (2018). NMR Resonance Assignments of RNase P Protein from *Thermotoga Maritima*. *Biomol. NMR Assign* 12, 183–187. doi:10.1007/s12104-018-9806-7
- Zhang, J., and Ferré-Damaré, A. (2016). Trying on tRNA for Size: RNase P and the T-Box Riboswitch as Molecular Rulers. *Biomolecules* 6, 18. doi:10.3390/biom6020018

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zeng, Abzhanova, Brown and Reiter. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Role of Substrate Mediated Allostery in the Catalytic Competency of the Bacterial Oligosaccharyltransferase PglB

Brittany R. Morgan and Francesca Massi*

Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, United States

OPEN ACCESS

Edited by:

Vincenzo Venditti,
Iowa State University, United States

Reviewed by:

Davit Potoyan,
Iowa State University, United States
Nicholas Charles Fitzkee,
Mississippi State University,
United States

*Correspondence:

Francesca Massi
francesca.massi@umassmed.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 13 July 2021

Accepted: 24 August 2021

Published: 15 September 2021

Citation:

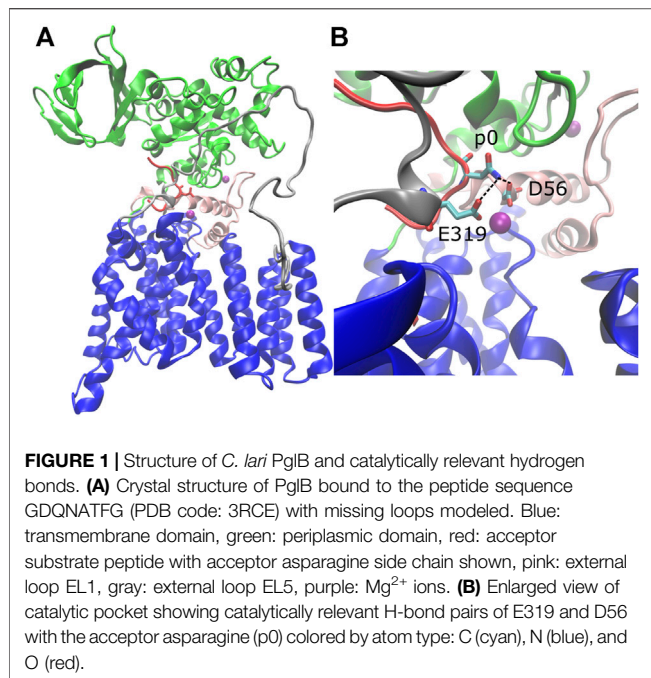
Morgan BR and Massi F (2021) The
Role of Substrate Mediated Allostery in
the Catalytic Competency of the
Bacterial
Oligosaccharyltransferase PglB.
Front. Mol. Biosci. 8:740904.
doi: 10.3389/fmolb.2021.740904

The oligosaccharyltransferase of *Campylobacter lari* (PglB) catalyzes the glycosylation of asparagine in the consensus sequence N-X-S/T, where X is any residue except proline. Molecular dynamics simulations of PglB bound to two different substrates were used to characterize the differences in the structure and dynamics of the substrate-enzyme complexes that can explain the higher catalytic efficiency observed for substrates containing threonine at the +2 position rather than serine. We observed that a threonine-containing substrate is more tightly bound than a serine-containing substrate. Because serine lacks a methyl group relative to threonine, the serine-containing peptide cannot stably form simultaneous van der Waals interactions with T316 and I572 as the threonine-containing substrate can. As a result, the peptide-PglB interaction is destabilized and the allosteric communication between the periplasmic domain and external loop EL5 is disrupted. These changes ultimately lead to the reorientation of the periplasmic domain relative to the transmembrane domain such that the two domains are further apart compared to PglB bound to the threonine-containing peptide. The crystal structure of PglB bound to the peptide and a lipid-linked oligosaccharide analog shows a pronounced closing of the periplasmic domain over the transmembrane domain in comparison to structures of PglB with peptide only, indicating that a closed conformation of the domains is needed for catalysis. The results of our studies suggest that lower enzymatic activity observed for serine versus threonine results from a combination of less stable binding and structural changes in PglB that influence the ability to form a catalytically competent state. This study illustrates a mechanism for substrate specificity via modulation of dynamic allosteric pathways.

Keywords: molecular dynamics, substrate binding, consensus sequence specificity, N-glycosylation efficiency, allosteric communication

INTRODUCTION

Asparagine-linked glycosylation (N-glycosylation) is an essential function in eukaryotes (Helenius and Aebi, 2004; Mohorko et al., 2011; Aebi, 2013). The attachment of a hydrophilic glycan changes the chemical properties of the protein, conferring new interactions and potentially affecting folding (Mohorko et al., 2011). N-glycosylation is also involved in quality control: it provides information



about folding status and thus can target misfolded proteins for degradation (Helenius and Aebi, 2004; Aebi, 2013). In prokaryotes, N-glycosylation is non-essential but plays important roles such as in the extremophile abilities of archaea (Eichler, 2013) and in the pathogenesis of bacteria (Szymanski et al., 2002; Nothhaft and Szymanski, 2013).

N-glycosylation is catalyzed by oligosaccharyltransferase (OST), which attaches a lipid-linked oligosaccharide (LLO) to the consensus sequence N-X-S/T of the acceptor substrate (where X is any amino acid except proline) (Aebi, 2013). OST is usually an oligomeric membrane protein complex in eukaryotes (Kelleher and Gilmore, 2006). In mammals, OST consists of seven subunits: ribophorin I, DAD1, N33/IAP, OST4, STT3A/B, Ost48, and ribophorin II (Kelleher and Gilmore, 2006); in prokaryotes, OST is a single subunit homologous to the catalytic STT3 eukaryotic subunit (Szymanski and Wren, 2005; Eichler, 2013). The structural homology with the eukaryotic STT3 is strongest for *Campylobacter jejuni* OST and it is thought to have a similar catalytic mechanism (Wacker et al., 2002; Kelleher and Gilmore, 2006; Li et al., 2010). The bacterial OST from *C. jejuni* and its homolog from *C. lari* have been studied to understand the fundamentals of N-glycosylation because of the strong homology and their relative simplicity (Wacker et al., 2002; Li et al., 2010; Jaffee and Imperiali, 2011; Ihssen et al., 2012; Gerber et al., 2013; Lizak et al., 2014; Barre et al., 2017). These studies have identified conserved motifs and catalytically important residues. A critical step forward in understanding how these features work together was provided by the crystal structure of the full-length *C. lari* OST (commonly referred to as PglB) bound to an acceptor substrate peptide, which provided structural insights into the catalytic mechanism and the role of conserved motifs (Lizak et al., 2011). Recent structures of eukaryotic OST complexes, which elucidate the interactions

between different subunits, further support the structural homology between the catalytic STT3 subunit and *C. lari* PglB (Bai et al., 2018; Braunger et al., 2018; Ramírez et al., 2019). Due to the size and complexity of the eukaryotic OSTs, the homologous bacterial OSTs continue to be useful model systems to understand common features of N-glycosylation.

The structure of *C. lari* PglB is composed of a transmembrane domain and a periplasmic domain (Lizak et al., 2011). The transmembrane domain (residues 1–432) consists of thirteen transmembrane helices (Figure 1A, blue) connected by short cytoplasmic or external loops and two long external loops (EL): EL1 (Figure 1A, pink) and EL5 (Figure 1A, gray). The N-terminus of EL5 is unstructured when only the acceptor substrate is bound (Lizak et al., 2011) and becomes ordered upon binding of LLO (Napiórkowska et al., 2017). The periplasmic domain (residues 433–711; Figure 1A, green) exhibits a mixed α/β fold with two major structural elements: the insertion domain and the conserved central core (Maenaka et al., 2008; Maita et al., 2010). The insertion domain is the beta-barrel-like structure of the periplasmic domain and its function is currently unknown; the central core contains residues important for recognition and binding of the acceptor substrate (Maita et al., 2010). The acceptor substrate is bound in a large cavity at the interface between the transmembrane and periplasmic domains and is pinned by the structured C-terminus of external loop EL5 (Lizak et al., 2011) (Figure 1A, red). LLO binds in the second large cavity located opposite the acceptor substrate binding cavity; the two cavities are linked by the acceptor asparagine (Napiórkowska et al., 2017). The crystal structures suggest a glycosylation mechanism where the catalytically essential and highly conserved magnesium coordinating residues D56 (EL1) and E319 (EL5) form hydrogen bonds with the acceptor asparagine and prime it for nucleophilic attack on LLO (Lizak et al., 2011; Napiórkowska et al., 2017) (Figure 1B and Supplementary Figures S1A,B, conserved residues in red).

While the typical consensus sequence of OSTs consists of N-X-S/T (X \neq P), a threonine residue at the +2 position (where the acceptor asparagine is defined as position 0) allows for glycosylation 40 times more efficiently in eukaryotes than a serine in this position (Bause, 1984). In *C. lari* PglB, N-X-S peptides have been reported to have 4-fold reduced peptide binding affinity and turnover is 1.2-fold faster for N-X-T sequences (Gerber et al., 2013). In the crystal structure of PglB, a threonine at position p+2 of the peptide forms hydrogen bonds with the strictly conserved WWD motif (residues 463–465; Supplementary Figure S1C, red) and a van der Waals contact with I572 (3.6 Å distance between gamma-methyl groups) (Lizak et al., 2011). One obvious explanation for the difference in catalytic activity between N-X-S and N-X-T is that serine cannot form a van der Waals contact with I572; supporting this is that an I572 V mutant with N-X-T peptides had a similar 4-fold reduction in peptide binding affinity (Gerber et al., 2013). However, the lost van der Waals interaction is a small contribution compared to the multiple hydrogen bonds with the WWD motif (which are not predicted to be affected), making it difficult to explain the significant effect on glycosylation. For this reason, we performed molecular dynamics simulations to

understand the change in substrate-enzyme interactions between the two amino acids at position p+2 of the substrate.

In this paper we focus on the substrate binding phase of the catalytic cycle and assess the differences that could affect the catalytic competency of PglB when bound to acceptor substrates containing threonine or serine at the +2 position using molecular dynamics (MD) simulations. The catalytic activity of PglB requires binding of both the acceptor substrate and LLO, although the order is thought to be unimportant (Lizak et al., 2011; Napiórkowska et al., 2017). While LLO may bind before the acceptor substrate, to our knowledge there is no evidence that LLO binding is the rate limiting step. Thus, understanding PglB-substrate interactions and the propensity of the complex to form a catalytically competent state suitable for LLO attachment can shed light on the observed differences in catalytic efficiency.

MD simulations of *C. lari* PglB with the optimal acceptor sequence, GDQNATFG (PglB-NAT), and the peptide variant GDQNASFG (PglB-NAS) were used to discover the detailed differences between threonine and serine at the +2 position in the substrate-bound state of the catalytic cycle. We observed an increased propensity for substrate disassociation in the trajectories of PglB-NAS compared to PglB-NAT. Our results show that threonine at the +2 position forms a network of interactions with the periplasmic domain and EL5 that promotes a more closed orientation of the periplasmic domain. The loss of the methyl group in the serine disrupts the substrate-enzyme interaction and the allosteric communication between the periplasmic domain and external loop EL5, allowing for increased motion of the periplasmic domain and leading to a smaller population of stably bound, catalytically competent states.

MATERIALS AND METHODS

Simulation Protocol

Molecular dynamics simulations of the bacterial OST PglB from *C. lari* with bound optimal peptide (GDQNATFG) were started from the crystal structure (PDB code 3RCE). The missing residues of external loop EL5 (residues 283–306) and missing loop residues in the periplasmic domain (residues 605–607) were built with modeler (Fiser et al., 2000) using a protocol that optimizes the positions of all non-hydrogen atoms of the loop without altering the coordinates of the existing residues. The optimization protocol used conjugate gradient minimization followed by molecular dynamics with simulated annealing (Fiser et al., 2000). The missing residues of the EL5 loop (residues 283–306) remain mostly disordered during the simulations, but they sample α -helical conformations with a probability of ~5%. Residues 284–289 are known to become structured and form a α -helix upon binding LLO (Napiórkowska et al., 2017). Using VMD (Humphrey et al., 1996), hydrogens were added to the crystal structure, the protein and peptide were inserted into a lipid bilayer, and the system was solvated with TIP3P (Jorgensen et al., 1983; MacKerell et al., 1998) water (the PglB/peptide system was electrostatically neutral). To evaluate the effect of the lipid

content of the bilayer on the dynamics of PglB, we used either POPC or POPE lipid bilayers.

The simulations were performed with NAMD 2.13 (Phillips et al., 2005) using the CHARMM27 force field for proteins and lipids (MacKerell et al., 1998). In the first phase of the equilibration, all atoms except the hydrocarbon tails of the POPC or POPE lipids were fixed. Conjugate gradient minimization followed by equilibration with constant temperature and volume (NVT ensemble) allowed the lipid tails to melt while the water, lipid head groups, Mg^{2+} ions, PglB, and peptide positions remained fixed. The following steps were then performed thirteen times to generate independent trajectories (ten with POPC lipids, three with POPE lipids). In the second phase of equilibration with constant temperature and pressure (NPT ensemble), restraints were removed and a force was applied to any water molecules that entered the lipid bilayer to prevent infiltration and disruption of the bilayer. At the beginning of the second phase, conjugate gradient minimization was performed on all atoms, followed by velocity reinitialization and temperature and pressure equilibration. The final phase of equilibration was performed with no constraints or applied forces using constant temperature and pressure to allow for final density equilibration of the system. Multiple restraint conditions were tried in the final phase of equilibration and the unrestrained equilibration protocol was determined to be the most appropriate to reproduce the coordination of the Mg^{2+} ion given the uncertainty in the Mg^{2+} coordination due to resolution of the crystal structure (3.4 Å) (Lizak et al., 2011). The coordination obtained by unrestrained equilibration is consistent with previous molecular dynamics simulations of PglB (Pedebos et al., 2015; Lee and Im, 2017).

The production phase of the data generation used constant temperature, pressure, and constant area in the plane of the bilayer ($NP_{xy}P_zT$ ensemble). Pressure and temperature were maintained at 1 atm and 310 K, respectively, using Langevin dynamics and the Nosé-Hoover Langevin piston method. The SHAKE constraint algorithm (Ryckaert et al., 1997) was used to allow a 2 fs time step and the particle mesh Ewald method (Essmann et al., 1995) was used for electrostatic interactions with periodic boundary conditions. Thirteen independent trajectories (ten with POPC lipids, three with POPE lipids) of 500 ns each were collected for PglB with the GDQNATFG peptide (PglB-NAT). The consensus sequence variant (T→S) was created with the Mutator plugin of VMD (Humphrey et al., 1996) and thirteen independent trajectories of up to 500 ns each (ten with POPC lipids, three with POPE lipids) of PglB with the GDQNASFG peptide (PglB-NAS) were collected using the same simulation protocol as above. Three of the trajectories (1, 5, and 12) were terminated after observing peptide unbinding (at 400 ns, 100 and 300 ns, respectively, see *Definition of Peptide Binding*).

The analysis presented in this paper focuses on differences in the bound states of the two peptides. The results of the simulations that used POPC and POPE lipids were found to be the same within the error of the measured observables (interaction energies, H-bond probabilities, distances between

TABLE 1 | Probability of the two peptides (NAT and NAS) forming the H-bonds identified as important in binding and the catalytically relevant H-bonds (mean and standard deviation for bound states).

	p (PglB-NAT)	p (PglB-NAS)
p+1 O-M318 N	0.83 ± 0.24	0.61 ± 0.35
p+2 N-D465 Oδ ^a	0.91 ± 0.10	0.68 ± 0.20
p+2 Oγ-W463 Nε	0.78 ± 0.20	0.38 ± 0.27
p+2 Oγ-W464 Nε	0.32 ± 0.17	0.16 ± 0.10
p+2 Oγ-D465 Oδ ^a	0.92 ± 0.11	0.58 ± 0.19
p+3 N-T316 O	0.81 ± 0.31	0.36 ± 0.31
p0 Nδ-D56 Oδ ^a	0.35 ± 0.35	0.41 ± 0.28
p0 Nδ-E319 Oε ^a	0.25 ± 0.27	0.21 ± 0.25

^aCombined probability of hydrogen bond formation with either Oδ₁/Oδ₂ or Oε₁/Oε₂.

PglB and key peptide residues, etc.), indicating that the dynamics of PglB is not influenced by the lipid composition of the bilayer. For example, the difference in the peptide-PglB interaction energy for the two lipid types was less than the standard deviation: PglB-NAT POPC = −228.7 ± 24.8 kcal/mol, PglB-NAT POPE = −240.0 ± 26.7 kcal/mol, Δ(NAT POPC-POPE) = 11.3 kcal/mol. Therefore, we combined all the trajectories for better sampling. PglB-NAT spent 5,063 ns in the bound state and PglB-NAS spent 2,780 ns in the bound state and all histograms were normalized to account for the difference in the number of data points. To ensure that we have collected enough of the relevant states to be able to make meaningful comparisons between the two peptides, we assessed the quality of sampling using jack-knife analysis (Lohr, 1999) of the peptide-PglB interaction energy. The standard deviation of the jack-knife means was calculated and found to be approximately 1/10 of the standard deviation of the interaction energy, indicating that we have good sampling (Supplementary Table S1). Furthermore, a jack-knife style analysis of the first two principal components (where each trajectory is excluded in turn) of the bound frames shows that the trajectories sample similar regions of conformational space independent of the lipid type (POPC: trajectories 1–10, POPE: trajectories 11–13), another indicator of adequate sampling (see Supplementary Figures S2, S3).

Analysis of Trajectories

Except where otherwise specified, trajectory analysis was performed with VMD (Humphrey et al., 1996) on bound frames only (see *Definition of Peptide Binding*) excluding the first 50 ns to allow for structural equilibration, figures were prepared with Tableau Software v10.4 (Seattle, WA), and structures were visualized with VMD using the STRIDE algorithm for secondary structure identification (Frishman and Argos, 1995).

Hydrogen Bonds

A hydrogen bond is defined by a donor-acceptor distance of less than 3.5 Å and a donor-hydrogen-acceptor angle of 130° < θ < 180°.

Definition of Peptide Binding

During the simulations we observed that the peptide could be in different states: bound, partially bound, or unbound from the

enzyme. We identified six non-catalytic H-bonds between the enzyme and the peptide as important indicators of peptide binding (see Table 1). The presence or absence of these H-bonds are used as the criteria for the bound, partially bound, and unbound states (Table 2). In addition, trajectories were terminated when the peptide had moved far away from the binding pocket of the enzyme. This was quantified using the sum of the donor-acceptor distances of three H-bonds formed by the peptide with the binding pocket (p+1 O-M318 N, p+2 Oγ-W463 Nε, and p+2 N-D465 Oδ) plus the distance between p+2 Cβ-I572 Cδ (a proxy for the p+2-I572 van der Waals interaction). Trajectories were terminated when this sum of distances exceeded 80 Å (see Supplementary Figures S4, S5). In the bound state, the average of this sum is 13.9 ± 1.8 Å for PglB-NAT and 17.8 ± 4.4 Å for PglB-NAS (the average individual distances are listed in Supplementary Table S2).

Average and Most Populated Structures

The average bound structure was generated by averaging over all bound frames for each system. The average structure representing the unbound state was generated by averaging over all frames of PglB-NAS after the loss of the six non-catalytic H-bonds identified as important indicators of binding (see Table 1). To validate the use of average structures, the Ramachandran plot of the average structures was compared to the ProCheck (Laskowski et al., 1993) Ramachandran plot for the crystal structure. Only residues in solvent exposed turn/coil regions of the average structures were found to be outside of the allowed area predicted for the crystal structure: E258, G281, K296, and L303. E258 is also outside of the Ramachandran plot in the crystal structure. G281, K296, and L303 are in the highly flexible unstructured EL5 N-terminal region absent from the crystal structure.

To identify the most populated structure we used two parameters: 1) the largest difference in the Ca-Ca distance between the periplasmic and transmembrane domains (residues 535–537 of the periplasmic domain and residue 279 of the transmembrane domain) and 2) the backbone RMSD (root mean square displacement). From the trajectories of PglB-NAT and PglB-NAS we identified the structures corresponding to the most populated inter-domain distance (PglB-NAT: 44.1 Å, PglB-NAS: 48.4 Å) and RMSD values. These structures were then validated as representative of the most populated conformation by locating them in the plot of the first two principal components and were found to be located in the regions of highest density (Supplementary Figure S6).

RMSD between average or most populated structures was calculated as

$$RMSD(NAT_{ave} - NAS_{ave}) = \sqrt{\left(\frac{1}{T} \sum_t NAT - \frac{1}{T} \sum_t NAS \right)^2}, \quad (1)$$

where NAT and NAS are the subset of backbone atoms of PglB-NAT and PglB-NAS, respectively, for which the RMSD is calculated.

Structural Alignment (Identification of the Stable Core)

The choice of residues to use for structural alignment is important in cases where domain motion occurs. A poor choice of alignment

TABLE 2 | Definition of peptide state relative to the PglB binding pocket, based on peptide-PglB interactions.

State	Criteria
Fully bound	Presence of at least one of the following H-bonds: p+1 O-M318 N p+2 O γ -W463 Ne p+2 N-D465 O δ AND presence of at least one of the following H-bonds: p+2 O γ -W464 Ne p+2 O γ -D465 O δ p+3 N-T316 O
Partially bound	All the following H-bonds are lost for >5 ns: p+1 O-M318 N p+2 O γ -W463 Ne p+2 N-D465 O δ AND presence of at least one of the following H-bonds: p+2 O γ -W464 Ne p+2 O γ -D465 O δ p+3 N-T316 O
Unbound	All the following H-bonds are lost for >5 ns: p+1 O-M318 N p+2 O γ -W463 Ne p+2 N-D465 O δ p+2 O γ -W464 Ne p+2 O γ -D465 O δ p+3 N-T316 O
Trajectory termination	After the sum of the distances (p+1 O-M318 N) + (p+2 O γ -W463 Ne) + (p+2 N-D465 O δ) + (p+2 C β -I572 C δ) exceeds 80 Å at any point in the trajectory

can obscure differences while a good choice will highlight them. Alignment on the transmembrane domain (excluding EL1 and EL5) was found to perform better than an all-atom alignment. However, previous simulations of PglB have found that the transmembrane domain does not behave as a single rigid domain (Lee and Im, 2017). To further refine the choice of alignment residues, the trajectories were analyzed with the *bio3d* geostas algorithm (Grant et al., 2006; Romanowska et al., 2012) in the R environment (R development core team, 2006; <http://www.R-project.org>) to determine which residues have the least variation in position over the trajectory (the stable core). The stable core residues were determined from the combined trajectories of PglB-NAT and PglB-NAS and consist of transmembrane domain residues 118–124, 126–130, 132, 342–347. Alignments on all atoms of PglB, the transmembrane domain, and the stable core were compared for structural analysis, cross-correlation calculations, and principal component analysis. In all cases, alignment on the stable core identified by the geostas algorithm (Romanowska et al., 2012) performed the best at highlighting relevant differences between the systems. The stable core was used for structural alignment to remove rotational and translational motion in all analyses reported here.

Principal Component Analysis

Cartesian principal component analysis was performed for bound frames only excluding the unstructured N-terminus of EL5 (residues 282–306) using the *bio3d* package (Grant et al., 2006) in the R environment (R development core team, 2006; <http://www.R-project.org>). The structured C-terminus of EL5 (residues

307–324), which is present in the crystal structure (Lizak et al., 2011), is included in the calculations. The calculation used a single combined trajectory for both systems aligned on the stable core identified by the *bio3d* geostas algorithm (Romanowska et al., 2012). Initial analysis of the trajectories showed a high similarity in the principal components of PglB-NAT and PglB-NAS (with a root mean square inner product (Grant et al., 2006) of the principal components of 0.84); because of the similarity in the principal components, the trajectories were combined in the final analysis to facilitate the comparison of the conformational space sampled by PglB-NAT and PglB-NAS. The first two principal components (PC1 and PC2) are representative of the dynamics of the combined system, capturing 84 and 9% of the motion, respectively. Density heat maps of the first two principal components were generated using the *mclust* R package (Fraley and Raftery, 2002).

Interaction Energy

The interaction energy was calculated using the namdenergy plugin of VMD (Humphrey et al., 1996). For PglB/p+2 interaction energies by residue, backbone and sidechain distances were used to determine which residues of PglB were close enough to interact for computational efficiency. The interaction energy for all residues with a center of mass distance within 15 Å of p+2 was found to be in excellent agreement with the total p+2/PglB interaction energy (see **Supplementary Table S3**). The interaction energy between the peptide and the catalytic Mg²⁺ ion was also calculated and the results did not alter the conclusions based on the PglB/peptide interaction energy (**Supplementary Table S4**).

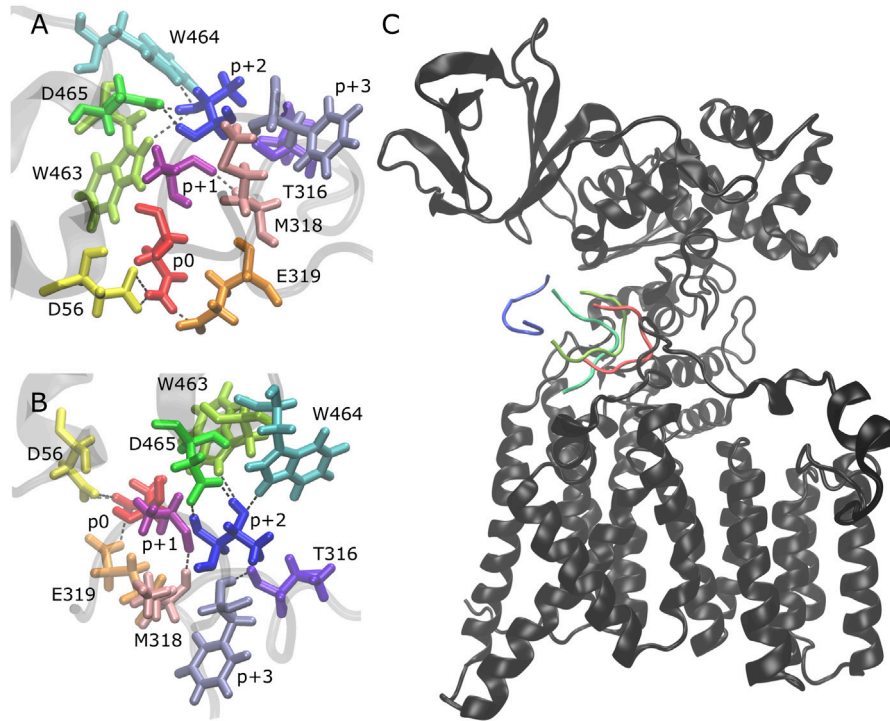


FIGURE 2 | Peptide unbinding and key hydrogen bonds. **(A,B)** Front and side view of key hydrogen bonding pairs between peptide and PglB (NAT peptide shown, backbone of peptide and adjacent PglB residues in transparent gray); catalytically relevant H-bond pairs between p0 (red) and D56 (yellow)/E319 (orange); backbone H-bond pairs of EL5 C-terminus residues T316 (violet) and M318 (light pink) with p+3 (lavender) and p+1 (dark pink), respectively; H-bond pairs between p+2 (blue) and periplasmic domain residues W463 (lime), W464 (cyan), and D465 (green). **(C)** Example of peptide unbinding from PglB-NAS, trajectory 1. PglB (black) is shown in the fully bound conformation after equilibration ($t = 50$ ns) and during unbinding (light green, $t = 276$ ns; dark green, $t = 291$ ns; blue, $t = 307$ ns).

Network Analysis

Network analysis was performed by the *bio3d* package (Grant et al., 2006) in the R environment (R development core team, 2006; <http://www.R-project.org>) using cross-correlations calculated over all bound frames aligned on the stable core. Cross-correlations between PglB, the peptide, and Mg^{2+} ions were included in the calculations with a threshold of 0.7 (corresponding to an angle between motion vectors of $<45^\circ$). Community membership was determined using the greedy algorithm (Clauset et al., 2004) and simplified network graphs were pruned of unconnected communities.

Network graphs overlaid on the structure represent each community by a sphere located at the geometric center of each community. The community radius (R_i) and edge weight between two communities (E_{ij}) are defined as

$$R_i = \frac{\sqrt{N_i}}{3} \quad (2)$$

$$E_{ij} = \frac{-\ln|\max(C_{ij})|}{4}, \quad (3)$$

where N_i is the number of residues in each community and $\max(C_{ij})$ is the maximum correlation coefficient between residues in each connected pair of communities. The scaling was chosen to illustrate relative differences between communities without

obscuring the underlying structure. Residues of the structure are colored according to community membership with unconnected communities in white in **Supplementary Figure S7A**. Simplified network graphs generated in R (**Supplementary Figures S7B,C**) follow the same community coloring (unconnected communities not shown) with community radii and edge weights as defined below.

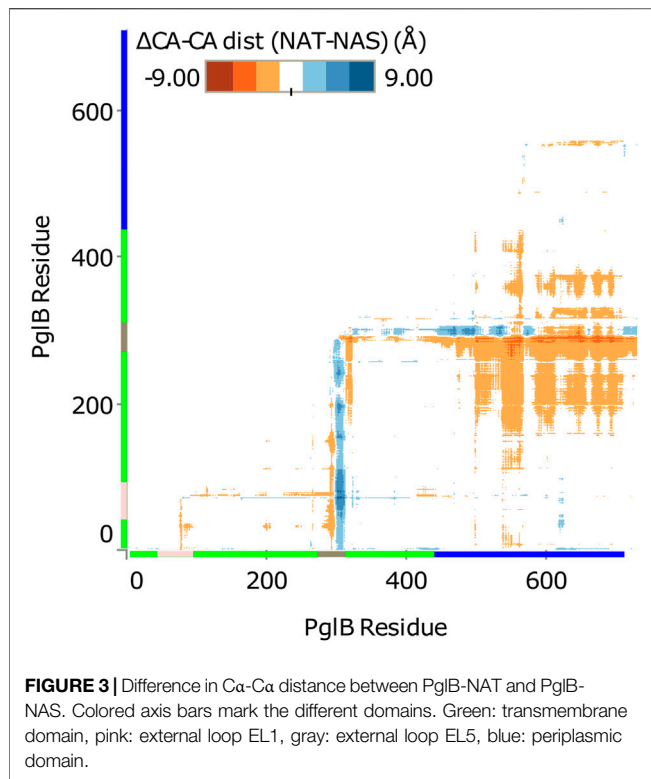
$$R_i = \frac{N_i}{3} \quad (4)$$

$$E_{ij} = -\ln|\max(C_{ij})|. \quad (5)$$

The number of shortest paths that pass through each node (betweenness), the number of edges of each node (degree), and suboptimal path analysis are calculated from the full network defined by the underlying cross-correlation matrix (where each residue is a node). The nodes contained in the suboptimal paths of PglB-NAT were identified using 500 paths and visualized using the first 100 paths for clarity.

Sequence Alignment

Sequence alignments were performed using with the Clustal algorithm (Madeira et al., 2019) on the reviewed sequences of InterPro 83.0 (December 2, 2020), family IPR003674 (29 sequences) (Blum et al., 2021). A subset of these sequences



was chosen to highlight conservation across the three families (prokaryotes, archaea, and eukaryotes, **Supplementary Figures S1A–C**) or the presence of a longer hydrophobic motif after the DXNK motif among eukaryotes (**Supplementary Figure S1D**).

RESULTS

Differences in the Bound State of NAT and NAS Peptides

Molecular dynamics simulations of *C. lari* PglB with the optimal acceptor sequence, GDQNATFG (PglB-NAT), and the peptide variant GDQNASFG (PglB-NAS) were used to discover the detailed differences between threonine and serine at the +2 position in the substrate-bound state of the catalytic cycle. Thirteen replicas with a cumulative simulation time of 6.5 and 5.8 μ s were collected for PglB-NAT and PglB-NAS, respectively. We observed a striking difference in the behavior of NAT and NAS peptides: out of the thirteen independent trajectories collected for each system, four trajectories of PglB-NAT exhibit partial peptide unbinding while ten trajectories of PglB-NAS exhibit partial peptide unbinding. Since a stably bound substrate is prerequisite to forming a catalytically competent state, we wanted to understand the critical interactions in the bound state that lead to this behavior. To do so, we needed to quantitatively determine when the peptide was in the bound state versus a partially bound state or fully unbound state in our simulations. We then analyzed the frames where the peptide is bound to identify differences in the bound states of the two peptides.

We identified six non-catalytic H-bonds whose absence corresponds to loss of the catalytically relevant H-bonds and disassociation from the binding pocket. These H-bonds were used to define the bound, partially bound, and fully unbound states (see **Table 2**). The H-bonds p+1 O-M318 N, p+2 N-D465 O δ , and p+2 O γ -W463 N ϵ (see **Supplementary Figures S4, S5** for donor-acceptor distances) provide a good indicator of partial unbinding because once all three are lost, with no specific order, the peptide no longer simultaneously forms the catalytically relevant H-bonds with D56 and E319 (the definition of the partially bound state as used in this work) and stable rebinding is never observed.

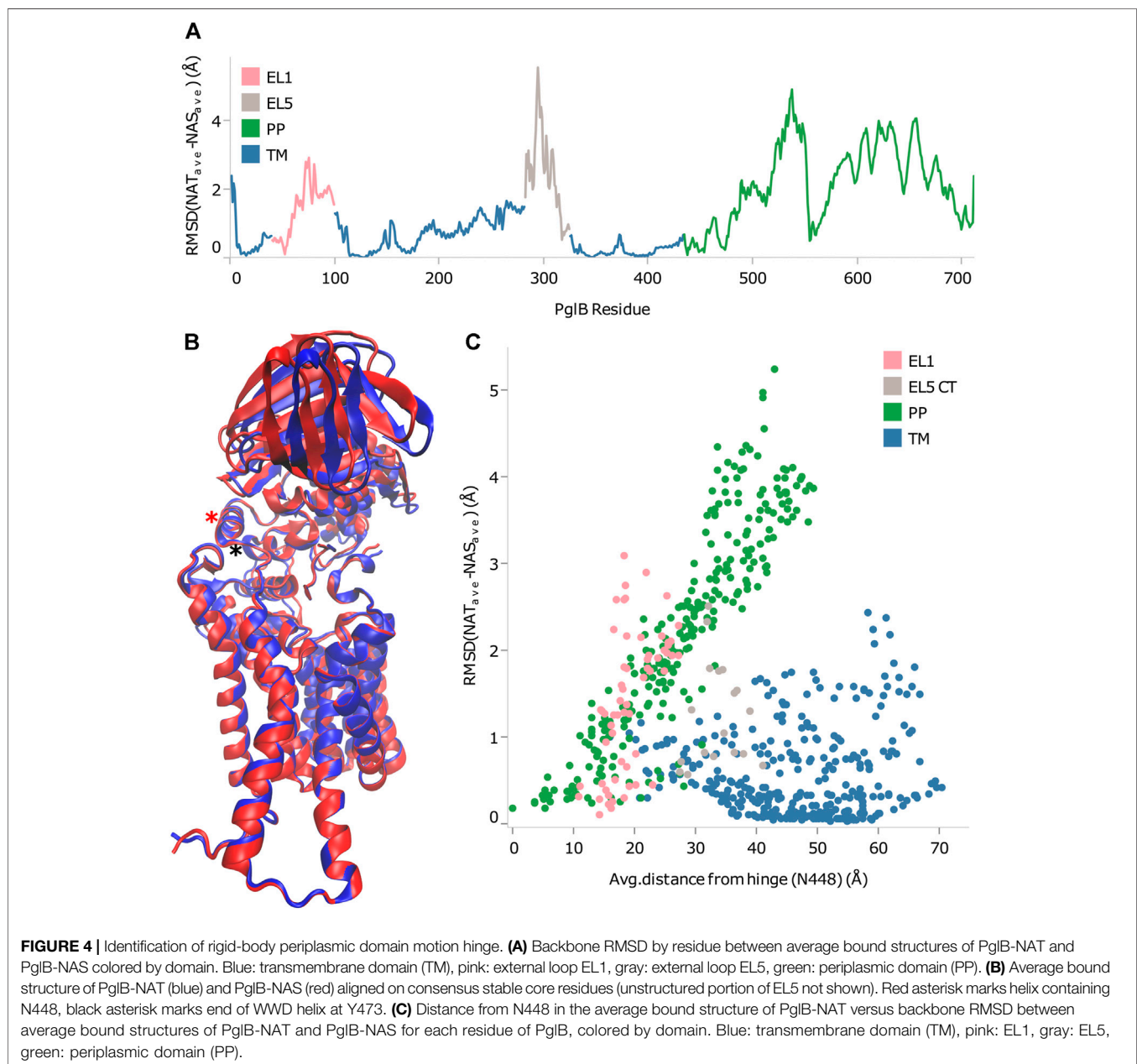
We analyzed the behavior of the peptides after becoming partially bound and observed that there is no clear pathway that leads to the full unbinding and dissociation from the binding pocket. After the loss of the H-bonds p+1 O-M318 N (**Figures 2A,B, Supplementary Figure S8**: dark pink/light pink), p+2 N-D465 O δ (**Figures 2A,B, Supplementary Figure S8**: blue/green), and p+2 O γ -W463 N ϵ (**Figures 2A,B, Supplementary Figure S8**: blue/lime), the peptide is still in the binding pocket but is more loosely associated with the enzyme (**Figure 2C**, light green). The peptide can remain in this state for the remainder of the simulation time or it can become fully dissociated (for example, **Supplementary Figure S5**: trajectory 3 vs. trajectory 1). The presence of three additional H-bonds keeps the peptide in this partially bound state: p+2 O γ -W464 N ϵ (**Figures 2A,B, Supplementary Figure S8**: blue/cyan), p+2 O γ -D465 O δ (**Figures 2A,B, Supplementary Figure S8**: blue/green), and p+3 N-T316 O (**Figures 2A,B, Supplementary Figure S8**: violet/lavender). Upon losing these H-bonds (which are never observed to reform once all three are lost), the peptide moves further away and becomes fully dissociated from the binding pocket (**Figure 2C**, dark green and blue). Thus, our results suggest that multiple interactions are important for the stability of the bound substrate: three key H-bonds are essential to keep the peptide tightly bound in the binding pocket of the enzyme (p+1 O-M318 N, p+2 N-D465 O δ , and p+2 O γ -W463 N ϵ) and three additional H-bonds keep the peptide in a partially bound state and prevent the full dissociation from the binding pocket (p+2 O γ -W464 N ϵ , p+2 O γ -D465 O δ , and p+3 N-T316 O).

The probabilities of all six of the hydrogen bonds that stabilize the bound and partially bound states are decreased in PglB-NAS compared to PglB-NAT (**Table 1**). To evaluate the stability of the bound substrate peptides, we calculated the probability of maintaining or transiently losing the first three and all six of these bonds simultaneously. We find that the NAS peptide is much less stably bound than the NAT peptide. All three of the H-bonds whose presence defines the bound state are simultaneously maintained in 68% of PglB-NAT bound states while only 24% of PglB-NAS bound states meet this criterion. Similarly, the probability of maintaining all six of these H-bonds is decreased and the probability of transiently losing all three or six of these H-bonds is increased by factors of 5–10 for PglB-NAS (**Supplementary Table S5**). This difference is seen over all trajectories and when comparing only those trajectories where the peptide eventually becomes partially bound (**Supplementary Table S5**, starred probabilities). Surprisingly, when comparing

TABLE 3 | Difference in PglB-peptide interaction energy between PglB-NAT and PglB-NAS by peptide residue.

Peptide residue	Δ PglB Int. Energy (NAT-NAS) (kcal/mol)
p-3 (Gly)	-3.1
p-2 (Asp)	10.5
p-1 (Gln)	2.0
p0 (Asn)	2.4
p+1 (Ala)	-3.0
p+2 (Thr/Ser)	-11.1
p+3 (Phe)	-2.7
p+4 (Gly)	-1.8
Total	-6.8

the peptides in the bound state, they have the same probability of forming the catalytically relevant H-bonds p0 N δ -D56 O δ (Figures 3A,B, Supplementary Figure S8: red/yellow) and p0 N δ -E319 O ϵ (Figures 3A,B, Supplementary Figure S8: red/orange) (Table 1), which are formed transiently in the absence of LLO [as seen previously (Pedebos et al., 2015)]. However, when considering the full trajectories (bound, partially bound, and unbound frames) of PglB-NAT and PglB-NAS, the population of catalytically competent states (where both catalytically relevant H-bonds are present) is lower for the NAS peptides by half compared to NAT peptides (Supplementary Table S6) because the NAS peptides populate the bound state



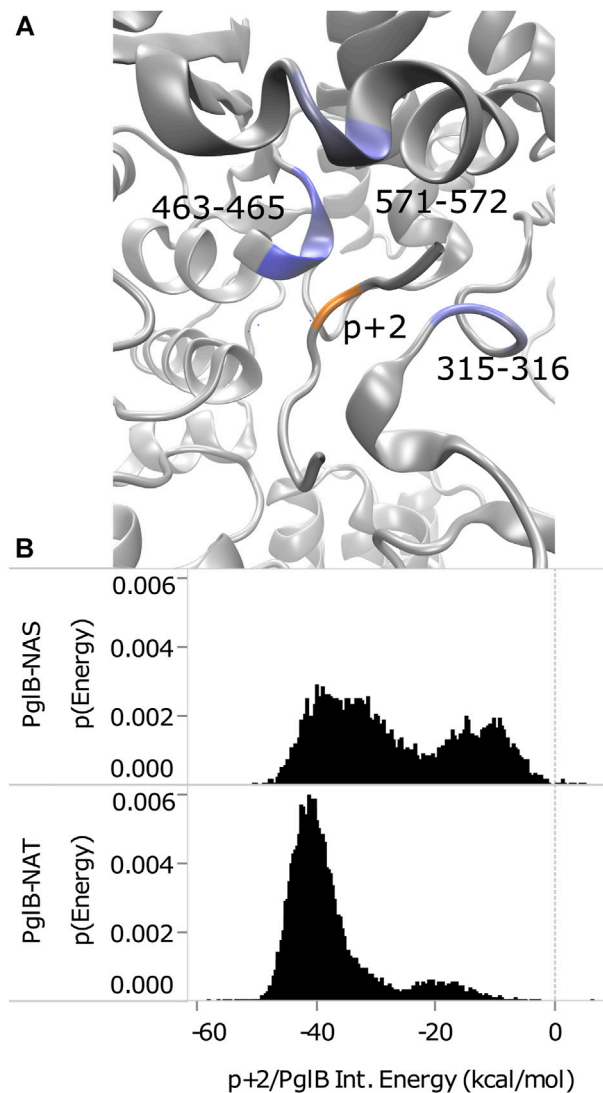


FIGURE 5 | Comparison of p+2/PglB interaction energy for PglB-NAT and PglB-NAS. **(A)** Differences in interaction energy between PglB and p+2 (NAT-NAS, blue = more favorable in PglB-NAT). Backbone of peptide residue p+2 shown in orange. **(B)** Comparison of total p+2/PglB interaction energy for PglB-NAT and PglB-NAS.

with lower probability (**Supplementary Table S6**). From this we conclude that the differences between NAT and NAS peptides that lead to increased unbinding in NAS are crucial to understanding the difference in catalytic competency.

None of the six H-bonds identified as important for stable binding should be impacted by the T/S variation in the consensus sequence at position p+2 and two out of six (p+1-M318 and p+3-T316) do not involve p+2 at all. To understand the difference in stability of the bound NAT and NAS peptides, we calculated the interaction energy between the peptide residues and PglB. The largest change is observed for p+2, however, the interaction energy while bound is affected beyond this residue and shows a large degree of compensation at peptide residue p-2 (**Table 3**).

The interaction energy between PglB and the NAS peptide is overall less favorable (**Table 3**), consistent with increased

disassociation. To understand the changes in conformation underlying the altered interaction energy, we compared the average bound structures of PglB-NAT and PglB-NAS and found differences in both the conformations of the peptide and of the periplasmic domain. The backbone RMSD between average structures (see Materials and methods, **Eq. 1**) is 1.46 Å for the peptide and 2.48 Å for the periplasmic domain. In contrast, the transmembrane domain shows minimal change between the two structures (backbone RMSD of 0.73 Å).

The compensation observed in the interactions with p-2 can be explained by the changes observed in the NAS peptide. The aspartic acid at position p-2 shifts relative to its transmembrane domain hydrogen-bonding partners R147 and R331, decreasing the favorable interaction energy with R147 (PglB-NAT interaction more favorable by -4.1 kcal/mol) and increasing the favorable interaction energy with R331 (PglB-NAS

TABLE 4 | Interaction energy between p+2:T and residues of PglB and the difference between p+2:T and p+2:S interaction energies for residues where the absolute value of the difference in interaction energy is ≥ 0.5 kcal/mol.

PglB residue	p+2/PglB energy (NAT) (kcal/mol)	Δ p+2/PglB energy (NAT-NAS) (kcal/mol)
E315	-0.9 ± 0.4	-0.6
T316	-3.4 ± 1.3	-1.2
W463	-3.0 ± 1.4	-2.0
W464	-3.0 ± 1.5	-1.3
D465	-20.7 ± 5.7	-4.4
R571	-1.5 ± 0.9	-0.5
I572	-1.8 ± 0.5	-1.0
Total	-34.4 ± 6.9	-11.0

interaction more favorable by -12.6 kcal/mol), accounting for 80% of the difference in interaction energy at p-2. The remaining difference comes from numerous long-range electrostatic interactions of the charged aspartic acid side chain.

For p+2, the key interacting residues of the WWD motif and I572 from the crystal structure (Lizak et al., 2011) show deviations in average position between PglB-NAS and PglB-NAT. The backbone RMSD between average structures is $0.87\text{--}1.18$ Å for the WWD motif, 1.67 Å for I572, and 0.93 Å for p+2, indicating that conformational change of both the peptide and the periplasmic domain contribute to the differences in the interaction energy.

The Substrate Influences the Periplasmic Domain Orientation of PglB

Comparison of the average structures suggests a difference in the conformation of PglB when bound to NAT versus NAS peptides. The difference between PglB Ca-Ca distances was calculated for PglB-NAT and PglB-NAS to identify regions of PglB that differ in conformation when bound to the two peptides. We found that the distance between the periplasmic domain and the transmembrane domain is increased in PglB-NAS compared to PglB-NAT (Figure 3). The average Ca-Ca distance between the periplasmic domain and the transmembrane domain (excluding the large external loops EL1 and EL5) is increased by 0.8 Å per residue pair in PglB-NAS compared to PglB-NAT, indicating that on average, the region between the two domains (containing the catalytic pocket) is more open in PglB-NAS (designated the open state) than in PglB-NAT (designated the closed state).

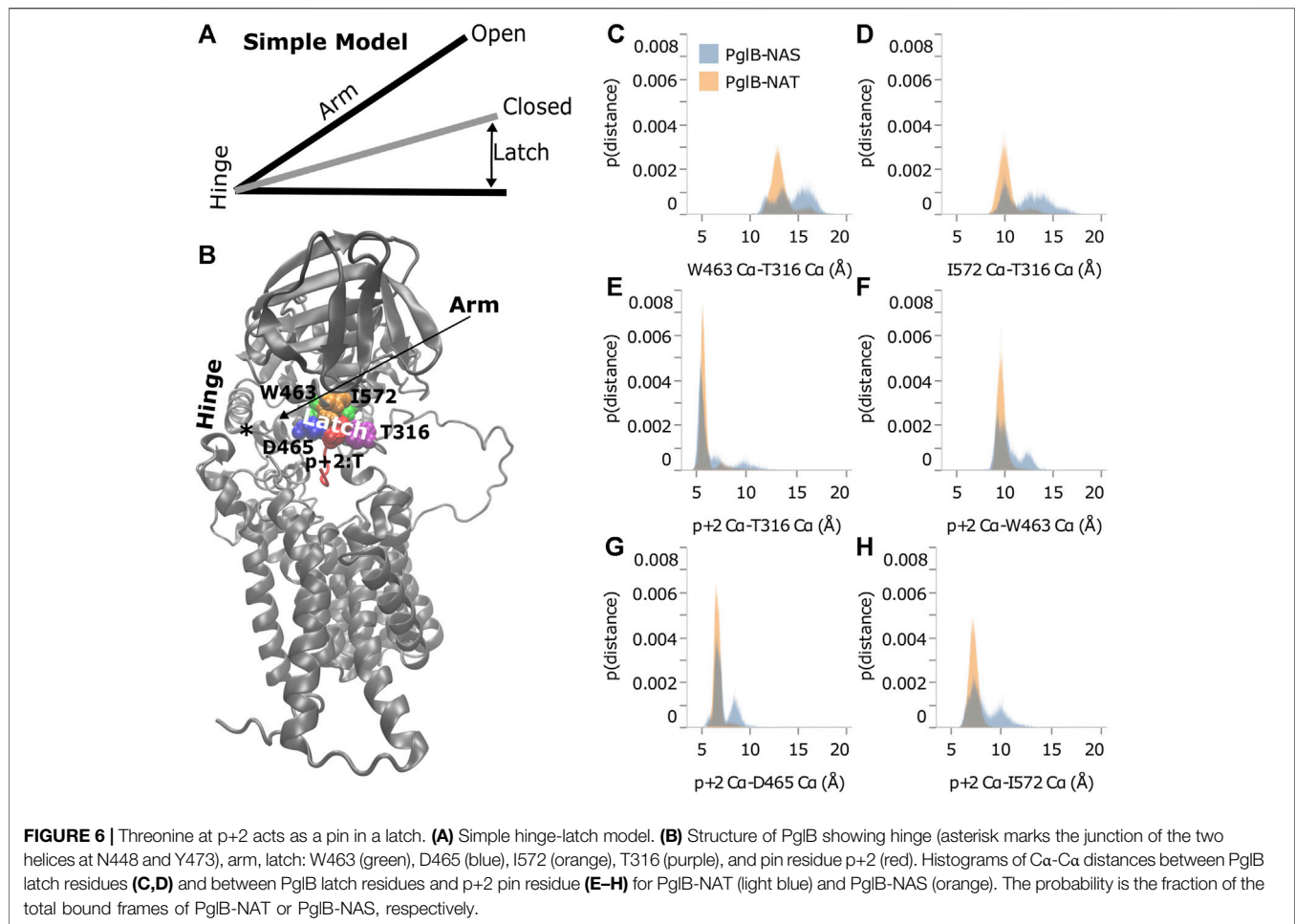
In contrast, the intradomain Ca-Ca distances of the periplasmic domain and the transmembrane domain differ remarkably little between PglB-NAT and PglB-NAS, with average changes of 0.01 Å and 0.04 Å per residue pair, respectively. The minimal change in Ca-Ca distances within the periplasmic and transmembrane domains indicates that the overall structure of both the periplasmic and transmembrane domains do not change between PglB-NAS and PglB-NAT. These results suggest that while the structure of the two domains is conserved, their relative positions change and that there should be a “hinge” that allows for rigid domain motion. If true, we would expect the hinge to meet the following criteria: 1) the hinge should have low backbone RMSD between

average structures, 2) the hinge should be located in an interfacial region between transmembrane and periplasmic domains, and 3) the magnitude of the displacement observed between the open and closed positions should correlate with the distance from the hinge. In our system, this means we expect that the further a residue is from the hinge, the more different its position will be between the open state and the closed state for the mobile portion of the structure (the periplasmic domain). Two pairs of structures were used to identify the differences between PglB-NAS and PglB-NAT: the average structure and as an additional validation, the most populated structure (calculated as described in *Analysis of Trajectories*). The results obtained using either the average or the most populated structure are the same.

From the RMSD between of PglB-NAT and PglB-NAS, we chose residues with low RMSD in regions of the structure appropriate for the hinge. We then calculated the correlation of the backbone RMSD and the distance from each of the chosen residues to see which of these residues agreed best with the criteria for the hinge residue (Supplementary Table S7). From this, we found that N448 and Y473 best meet the criteria for the hinge between the closed state (PglB-NAT) and the open state (PglB-NAS). Both residues are located in regions of low RMSD (Figure 4A) and have the highest correlation coefficient (0.95) between distance and backbone RMSD of the periplasmic domain residues (Supplementary Table S7). These two residues are adjacent in the structure of PglB at the interface between the two domains (Figure 4B): N448 (red asterisk) is located on a helix in between the transmembrane domain and the periplasmic domain with Y473 directly below N448 at the terminus of the helix containing the WWD motif (black asterisk), indicating that the hinge is located at the junction of these two helices. Additionally, the backbone RMSD between the periplasmic domain residues (the mobile portion of the structure) increases linearly with distance from N448 and Y473 but the correlation is weak for other regions of PglB (see Figure 4C for example). Lastly, we chose other residues distant from the hinge region with low RMSD across the sequence of PglB and these residues show low correlation between distance and RMSD (Supplementary Table S8), validating that the strong linear relationship is unique to the hinge region.

Origin of Observed Structural Differences

The interaction energy between p+2 and the residues of PglB was calculated for both peptides (Table 4) to understand how the T/S



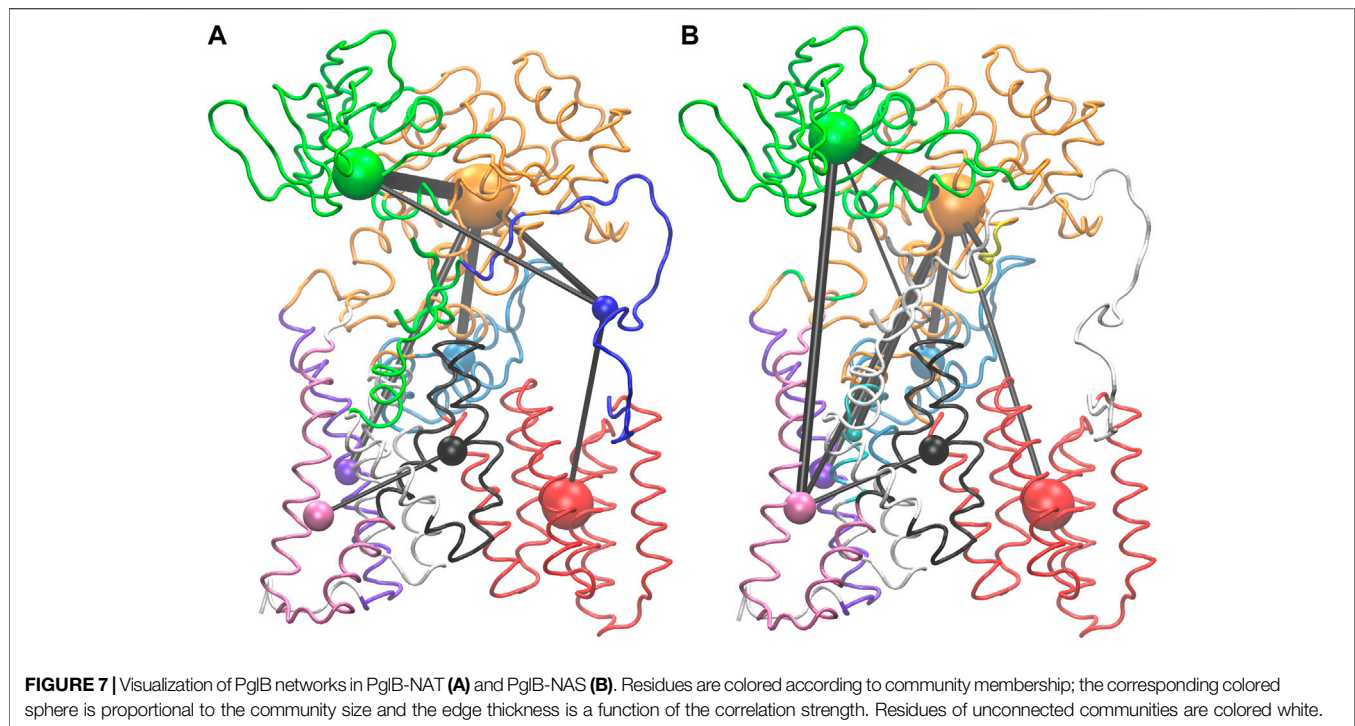
difference results in widespread changes in the peptide-PglB interaction (**Table 3**) and how it affects the orientation of the periplasmic domain (**Figure 3**). The location of residues that differ in interaction energy between PglB-NAT and PglB-NAS by $\geq |\pm 0.5|$ kcal/mol (absolute value of the difference ≥ 0.5) are shown in **Figure 5A**. For each of these residues, the interaction energy is less favorable in PglB-NAS (**Table 4**). As expected, the loss of the methyl group in p+2:S decreases the interaction energy with I572 but also its neighbor R571. The WWD motif (463–465) shows the largest differences in interaction energy and surprisingly, residues in the structured C-terminal portion of EL5 (315–316) show significant differences as well.

Although the average differs by $\geq |\pm 0.5|$ kcal/mol for the residues listed in **Table 4**, a histogram of the interaction energy shows that the two peptides sample states with similar p+2 interaction energies (**Supplementary Figure S9**). The measured difference in the average interaction energy of p+2 is due to the higher population of states with near-zero interaction energies by the NAS peptide (**Supplementary Figure S9**), decreasing the average interaction energy of the residues highlighted in **Figure 5A**. This suggests that p+2:S interacts in a similar way with residues of PglB where the

interaction energy differs by $\geq |\pm 0.5|$ kcal/mol, but that the interactions are less stable, consistent with the decreased stability of H-bonds identified as important for peptide binding involving many of these same residues (**Table 1** and **Supplementary Table S5**).

The total effect of sampling higher energy conformations with higher probability in PglB-NAS can be seen in the total interaction energy between p+2 and PglB (**Figure 5B**). Both PglB-NAT and PglB-NAS have bimodal distributions indicating the sampling of two states, one more favorable than the other. However, the more favorable (primary) mode and the less favorable (secondary) mode are more favorable in PglB-NAT by approximately 6 and 7 kcal/mol, respectively. Additionally, the more favorable state is more populated in both PglB-NAT and PglB-NAS but the difference in the population of the two states is significantly larger for PglB-NAT (**Figure 5B**).

These data indicate that the loss of the methyl group in p+2:S results in fewer interactions between the peptide and PglB. The result of these changes is less favorable total interaction energy in PglB-NAS (**Figure 5B**) due to fewer simultaneous interaction partners. The interactions of the residue at p+2 with the periplasmic domain and the EL5 loop provide a mechanism for the T/S variation at the +2 position to affect the



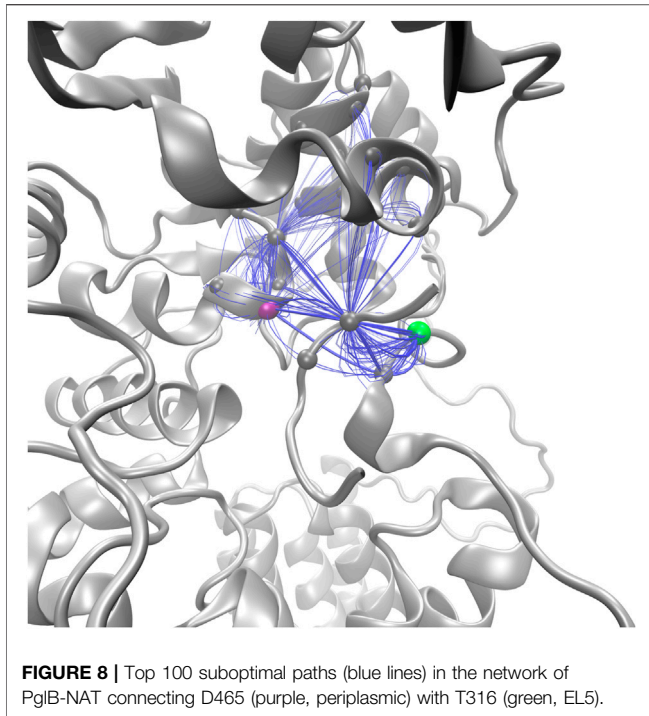
periplasmic domain orientation: threonine may act as a pin to close a latch, preventing opening of the periplasmic domain hinge (near N448/Y473). We define residues of PglB with an absolute difference in the p+2 interaction energy ≥ 1.0 kcal/mol as the latch residues (W463 and D465 of the WWD motif, I572, and EL5 residue T316). Our hypothesis is that p+2 interacting with these latch residues acts as a pin to close the latch which limits the hinge motion when p+2 is threonine, but that this ability is impaired when p+2 is serine. The two sides of the latch are formed by the interacting residues of the periplasmic and EL5 domains (**Figure 5A**) and by linking the two regions which contain the latch, threonine at p+2 closes the latch and restricts the motion of the hinge. The ability of serine at p+2 to pin the latch closed is impaired because serine is incapable of forming stable simultaneous interactions with the periplasmic domain and the EL5 loop due to the loss of the methyl group.

As shown in **Figure 4C**, distance from the hinge (N448/Y473) is strongly correlated with backbone displacement of the periplasmic domain, consistent with a hinge-like motion. We propose that the difference between the two peptides when bound to PglB is consistent with a simple hinge-latch model (**Figure 6A**). In this model, the pin connects the two sides of the latch when closed. An arm connects the hinge and latch and mediates the restriction of the domain motion when the pin closes the latch. In PglB, N448/Y473 is the hinge, residues W463, D465, I572, and T316 form the two sides of the latch, p+2 acts as the pin to hold the latch closed when threonine, and the helix connecting the hinge and latch is the arm (residues 463–472) (**Figure 6B**).

If the hinge-latch model is accurate for PglB, the motion should be different when latched (PglB-NAT) versus unlatched

(PglB-NAS). The motion should be more restrained in PglB-NAT and the motion should be freer in PglB-NAS. Principal component analysis of PglB (excluding the unstructured EL5 N-terminus) shows that the motion of PglB-NAT and PglB-NAS is similar (**Supplementary Figures S2, S3**). However, the frequency of sampling different regions of the conformational space is different (**Supplementary Figure S6**). PglB-NAT is more concentrated in the center of the conformational space, which is consistent with restrained motion of the hinge when the latch is pinned (**Supplementary Figure S6A**). PglB-NAS is more broadly distributed with higher sampling in the tails, which is indicative of increased motion consistent with an unlatched hinge (**Supplementary Figure S6B**).

The motion of PglB-NAT versus PglB-NAS supports the existence of a latch. To verify that our choice of latch and pin residues is correct, we looked at the distance between these residues in the latched (PglB-NAT) and unlatched (PglB-NAS) conditions. For the latch and pin to restrict the motion of the hinge (as we propose occurs in PglB-NAT), the latch and pin complex must be stable (the distances between these residues should not fluctuate significantly). When the pin is unable to close the latch (as we propose occurs in PglB-NAS), we expect to see greater fluctuations in the distance between the pin and latch residues. The histograms of the distances between the PglB residues on opposite sides of the latch (W463 and I572 of the periplasmic domain with T316 of the EL5 loop) show that the latch residues in PglB-NAT predominantly sample a narrow range of distances (**Figures 6C,D**, gray). The distances between the pin (p+2) and the latch residues show a similar preference for sampling a restricted range of distances (**Figures 6E–H**, gray). In PglB-NAS, the distances between the latch and pin residues sample larger distances more often than in PglB-NAT (**Figures 6C–H**, red). These results show that when p+2 is



threonine, the latch/pin complex is stable, as required to restrict the motion of the hinge in the hinge-latch model. The increased sampling of larger distances between the pin and latch residues in PglB-NAS is instead consistent with a pin that is unable to keep the latch closed.

The behavior of the pin and latch residues explains the interplay between peptide binding and periplasmic domain orientation. The distances between the latch residues in PglB-NAS increase because serine lacks the methyl group and is unable to form the stable simultaneous interactions necessary to pin the latch closed (keeping the two sides of the latch in closer proximity). The NAS peptide therefore shifts its position to interact with the latch residues on opposite sides (periplasmic domain and EL5) because the latch residues are further apart. As the NAS peptide moves towards one side of the latch, it loses the interaction with the other side of the latch (reflected in the increased sampling of near-zero interaction energies (**Supplementary Figure S9**). The increased distance between the latch residues and the subsequent motion of the NAS peptide explain the weakened H-bonds important for substrate binding because these bonds between the enzyme and substrate overlap with the latch/pin complex (**Table 1** and **Supplementary Table S5**). Of the six H-bonds identified as important for a stably bound peptide, four involve the PglB latch residues, two involve neighboring residues, and both sides of the latch are important (four H-bonds with the periplasmic domain and two H-bonds with EL5). Similarly for the substrate, of these six H-bonds, four involve the pin residue (p+2) and two involve adjacent substrate residues (p+1/p+3). For these reasons, the state of the latch (pinned or unpinned) is inextricably linked to the partial unbinding of the substrate.

Substrate Mediated Allosteric Networks

Network analysis was used to look for allosteric communication mediated by the peptide as suggested by

the hinge and latch model. The network of PglB-NAT consists of eight communities while the PglB-NAS network contains nine communities (**Figure 7, Supplementary Figure S7**). Seven of these communities are analogous in PglB-NAT and PglB-NAS with 60–100% similarity containing 626 residues of PglB-NAT and 603 residues of PglB-NAS (out of 711 PglB residues, 8 peptide residues, and 2 Mg^{2+} ions). The two largest communities contain most of the periplasmic domain residues (**Figure 7, Supplementary Figure S7**: orange, green) and are 60–70% analogous. The third largest community (transmembrane helices 4–9; **Figure 7, Supplementary Figure S7**: red) is 98% analogous in the two networks and is consistent with previous results that showed these helices move as a single unit to accommodate the presence of LLO (Lee and Im, 2017). The last four analogous communities contain 30–50 residues each and are comprised of the EL1 residues (**Figure 7, Supplementary Figure S7**: light blue) and subsets of the remaining transmembrane helices (**Figure 7, Supplementary Figure S7**: black, purple, pink). Of the residues not found in the analogous communities, 32 PglB-NAT residues are part of a community not connected in the network of PglB-NAS (**Figure 7, Supplementary Figure S7**: blue, EL5) and 11 PglB-NAS residues are part of a community not connected in the network of PglB-NAT (**Figure 7, Supplementary Figure S7**: cyan, transmembrane helix). An additional five residues (483–487) containing part of the bacterial DGGK motif (residues 481–484; **Supplementary Figure S1C**, cyan), which is suggested to be important for promoting proper conformation of the p0 side chain for LLO attachment (Pedebos et al., 2015; Napiórkowska et al., 2017), is part of a separate community (**Figure 7, Supplementary Figure S7**: yellow) in PglB-NAS instead of the large community containing the WWD motif as in PglB-NAT (**Figure 7, Supplementary Figure S7**: orange).

Lastly, 63 PglB-NAT residues and 102 PglB-NAS residues are unconnected (**Figure 7, Supplementary Figure S7**: white). In PglB-NAT, the residues in unconnected communities are part of the transmembrane domain. These residues include the N-terminal tail and residues around the stable core (see *Analysis of Trajectories*). Half of the unconnected residues in PglB-NAS are similarly located in the transmembrane domain. Significantly, the remaining unconnected residues in PglB-NAS are the entire EL5 loop (43 residues) and the peptide (8 residues) (**Figure 7, Supplementary Figure S7**: white). In PglB-NAT, the EL5 loop is part of three communities: blue (residues 283–309, 311–315), orange (residue 310) and green (residues 316–324) (**Figure 7, Supplementary Figure S7**). The NAT peptide is also part of the green community, which contains most of the residues important for both peptide binding and the pin and latch behavior demonstrated for p+2:T (see **Figure 5A** and **Table 4**), excluding only the WWD motif which is located in the adjacent, strongly connected orange community in both PglB-NAT and PglB-NAS (**Figure 7, Supplementary Figure S7**). The lack of edges connecting

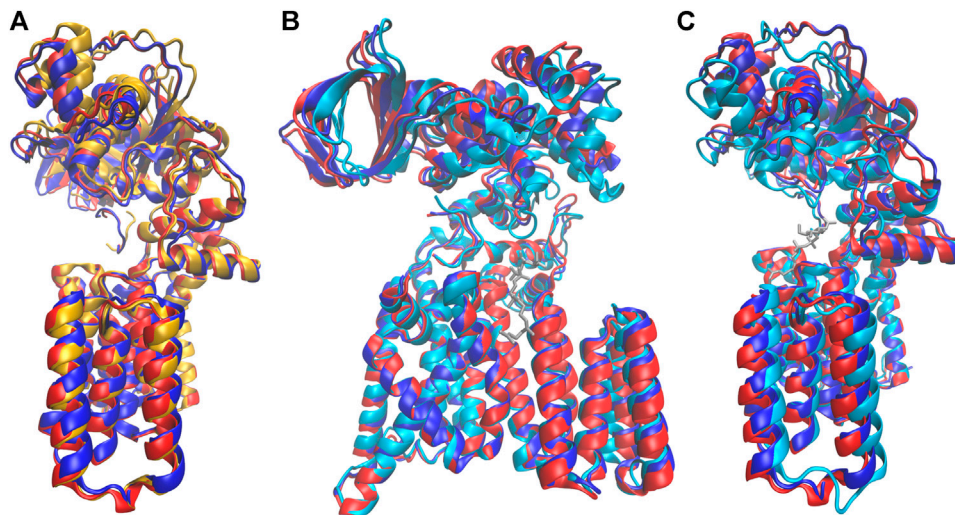


FIGURE 9 | Comparison of structures at different states in the catalytic cycle. Average bound structure of PglB-NAT (blue) and PglB-NAS (red) aligned on consensus stable core residues compared with **(A)** the average unbound structure from PglB-NAS (yellow) and **(B,C)** the crystal structure of PglB in complex with the acceptor peptide and an LLO analog (cyan, LLO analog in gray), front **(B)** and side **(C)** views (PDB code: 5OGL). EL5 not shown for clarity.

the communities containing the peptide and EL5 in PglB-NAS suggests that allosteric communication between the periplasmic domain and EL5, mediated by the peptide, is lost in the PglB-NAS complex.

Suboptimal path analysis can show paths of allosteric communication between different parts of the network (Scarabelli and Grant, 2014; Yao et al., 2016). The periplasmic latch residues (W463, D465, and I572) are connected to the EL5 latch residue (T316) by a diverse network of paths passing through the pin (p+2, 99.6% of paths) (**Figure 8**). Network attributes such as the number of edges per node (degree) and the number of shortest paths that pass through each node (betweenness) further illustrate the high connectivity of the NAT peptide and that adjacent peptide residues p+1 and p+3 play a supporting role in the network of PglB-NAT (**Supplementary Table S9**). In contrast, at the threshold used for the network analysis (cross-correlation >0.7), no paths connect the periplasmic domain and EL5 in PglB-NAS (**Figure 7**). Since a threshold imposes an arbitrary cutoff, we examined the cross-correlation values between the three most important peptide residues in the network of PglB-NAT (p+1, p+2, and p+3) and the latch residues (W463, D465, I572, and T316). Only three residue pairs have cross-correlation values near the threshold in PglB-NAS (cross correlation >0.68): p+1 with W463 and W464 and p+2 with the adjacent I317 (**Supplementary Table S10**). In PglB-NAT, 12 out of 18 pairs meet the 0.7 threshold and 15 out of 18 pairs meet the 0.68 threshold (**Supplementary Table S10**); the cross-correlation cutoff would have to be lowered to 0.51 in PglB-NAS to match the level of connectivity in the network of PglB-NAT (**Supplementary Table S10**). These results support that the allosteric communication between

the periplasmic domain and EL5 is disrupted in PglB-NAS because the NAS peptide does not mediate the interaction effectively.

Comparison of the Substrate Binding Pocket of PglB and Human OST

In the cryo-EM structure of the human OST complex with acceptor peptide (OST-B, PDB code 6S7T (Ramírez et al., 2019)), the p+2 binding pocket is structurally similar to the binding pocket in PglB (**Supplementary Figure S10**), thus our findings are likely generalizable to the eukaryotic OSTs. However, a lysine residue (K674) is present in the analogous position of I572 in the structure of PglB (**Supplementary Figure S10**, orange). The cryo-EM structure suggests that K674 in the human OST may form a hydrogen bond with D606 of the WWD motif (**Supplementary Figure S9**, blue). This H-bond would stabilize the nearby tryptophan (W677), which is structurally analogous to V575 in the extended bacterial MIV motif (Maita et al., 2010) (**Supplementary Figure S10**, yellow). The methyl group of p+2 is rotated towards W677 in the human structure (**Supplementary Figure S10**, red/yellow), suggesting W677 replaces the stabilizing van der Waals interaction formed with I572 in the bacterial structure. W677 is conserved in eukaryotes (**Supplementary Figure S1D**) as part of an extended hydrophobic motif after the previously identified DXCK motif (Maita et al., 2010). W677 also stacks with S402 (of the SVSE motif, **Supplementary Figure S1B**) in the structurally analogous position to T316 (**Supplementary Figure S10**, purple), possibly explaining the conservation of the smaller serine in this position for eukaryotes (Jaffee and Imperiali, 2011). Overall, the structural similarity between the binding pockets suggests that a similar

hinge/latch mechanism is present in eukaryotic OSTs with variations that modulate the catalytic efficiency for serine-containing substrates.

DISCUSSION AND CONCLUSION

The crystal structure of *C. lari* PglB shed light on the catalytic mechanism of asparagine-linked glycosylation and provided a structural explanation for the difference in catalytic efficiency between N-X-T and N-X-S substrates (where X can be any residue but proline in the consensus sequence): the loss of the methyl group would eliminate the van der Waals interaction between the substrate residue at position p+2 and I572 of PglB (Lizak et al., 2011). However, there remained four hydrogen bonds between the p+2 substrate residue and the WWD motif, making the van der Waals interaction with I572 seemingly a relatively small contribution. In this paper, we focused on the substrate bound phase of the glycosylation mechanism of PglB to explore how the two different residues at position p+2 (threonine and serine) could affect the catalytic competency. We found that the two substrates exhibited a striking difference: the substrate peptide containing serine was 2.5 times more likely to experience partial unbinding over 500 ns than the substrate peptide containing threonine. While striking, this is supported by the different binding affinities measured for the NAT and NAS peptides. The peptide binding affinity for PglB is 4 times lower for NAS peptides compared to NAT peptides (Gerber et al., 2013), indicative of a higher off-rate for NAS over NAT peptides.

Examining the PglB-peptide interactions while bound, we identified multiple hydrogen bonds that are important for stable binding—previously identified interactions with the WWD motif and newly identified interactions with the structured EL5 C-terminus. PglB when bound to the NAS peptide sampled a greater range of conformations and exhibited decreased formation of these key hydrogen bonds compared to PglB when bound to the NAT peptide. Despite this, the catalytically relevant H-bonds between the acceptor asparagine and D56/E319 had similar transient formation for the two peptides in the absence of LLO (necessary for a catalytically competent state). However, since being bound in the catalytic pocket is a prerequisite to the formation of these H-bonds, the increased partial substrate disassociation observed for PglB-NAS decreased the population of catalytically competent states by roughly half.

More remarkably, the identity of the amino acid at the p+2 position of the substrate peptide also affected the periplasmic domain orientation of PglB. The two bound structures are similar, but they differ in ways that suggest either the unbound or the peptide/LLO bound states. Comparison of the average bound structures of PglB-NAT and PglB-NAS with the average unbound structure (Figure 9A) shows that the average structure of bound PglB-NAS more closely resembles the more open unbound structure than does the average structure of bound PglB-NAT. More open periplasmic domain conformations have been observed in the absence of substrates and were hypothesized to play a role in product release (Pedebos et al., 2015). In contrast,

comparison of the average peptide-bound structures with the crystal structure of the PglB-peptide-LLO complex (Napiórkowska et al., 2017) (Figures 9B,C) shows that the periplasmic domain is oriented such that the binding pocket between the periplasmic and transmembrane domains is more compact when the LLO analog is present. This suggests that a more closed conformation of the periplasmic domain is important for a catalytically competent state, and thus the more closed periplasmic domain orientation of PglB-NAT compared to PglB-NAS may play a role in their different catalytic competencies.

Rigid body domain motion was found to describe the difference between the conformations of the periplasmic domain in PglB-NAT and PglB-NAS with a hinge near the junction of the two helices containing N448 and Y473, near the interface of the periplasmic and transmembrane domains. Threonine at position +2 acts as a pin in a latch to inhibit the hinge motion by forming interactions with residues in the periplasmic domain and EL5. Loss of the methyl group in serine results in the disruption of these interactions, allowing the latch to open. The pin/latch complex and the H-bonds necessary for stable substrate binding involve many of the same residues. We found that the opening of the latch necessarily disrupts the binding, and thus combined with the more open periplasmic domain orientation produces the fundamental difference observed between the simulations of PglB-NAT and PglB-NAS, namely, the 2.5 times increase in the frequency of partial NAS peptide unbinding. The importance of the newly identified substrate-EL5 interaction is supported by, and may help explain, the conservation of T316 as part of the bacterial TIXE motif (Jaffee and Imperiali, 2011). Network analysis confirmed the existence of allosteric communication mediated by the peptide between the periplasmic domain and EL5 for NAT peptides and showed that it is greatly weakened for NAS peptides.

Neighboring substrate residues at positions p+1 and p+3 play supporting roles in the binding and allosteric communication. Both p+1 and p+3 form strong H-bonds in PglB-NAT ($p > 0.8$) that we identified as important for substrate binding stability. Substrate residues p+1 and p+3 form backbone H-bonds with the structured portion of EL5; EL5 pins the substrate in the binding pocket and p+1/p+3 play a role in stabilizing the peptide-EL5 interaction. We also found a robust network of interactions between the periplasmic domain and EL5 mediated by the NAT peptide with high connectivity of multiple substrate residues (including p+1 and p+3), facilitating allosteric communication.

These results offer potential explanations for the observation that the catalytic competency of N-X-S substrates is more sensitive to nearby residues p+1 and p+3 than N-X-T substrates (Mellquist et al., 1998; Malaby and Kobertz, 2014). With the disruption of the H-bonds with the WWD motif, the p+1 and p+3 H-bonds with EL5 may play a larger role in keeping serine-containing substrates bound in the catalytic pocket. Intriguingly, the degree to which the p+1 and p+3 interactions are affected in our simulations mirrors their observed effect on the catalytic competency. In our substrate

peptides, p+1 is an alanine [found in better substrates (Malaby and Kobertz, 2014)] and p+3 is a phenylalanine [found in worse substrates (Mellquist et al., 1998)]. Despite having more than three times the number of edges (p+1 has 15 vs. 52 edges for p+3), the number of shortest paths that pass through p+3 is a fraction of the shortest paths that pass through p+1 (46 vs. 2,268 for p+1 and p+3, respectively), indicating that p+1 is far more robustly connected in the network of PglB-NAT. While the p+1 and p+3 H-bonds with EL5 form with similar probability in PglB-NAT, this difference in substrate residue connectivity appears to have a more functional affect in PglB-NAS, where the p+1-M318 H-bond probability is decreased by 0.22 but the p+3-T316 H-bond probability is decreased by 0.45, more than twice the decrease of the p+1 H-bond. Similarly, in PglB-NAS the cross-correlations between the peptide and key PglB residues decrease on average compared to PglB-NAT by half as much for p+1 compared to p+2 and p+3 (a decrease of 0.1 for p+1 vs. 0.2 for p+2/p+3).

Experimental data have shown that an interplay exists between the catalytic competency and the substrate residues at positions +1, +2, and +3—serine at p+2 increases the sensitivity to p+1 and p+3, and a less optimal residue at p+1 was found to make differences at p+3 more pronounced (Mellquist et al., 1998; Malaby and Kobertz, 2014). This observation is consistent with our simulation results that show a dynamic network between substrate and enzyme: the strong network of interactions with p+2 when threonine is present leads to more stable substrate-enzyme binding, facilitating the formation of interactions between PglB and p+1/p+3, even with less optimal residues; with serine, this dominant interaction is weakened and the supporting roles of p+1/p+3 become more important to the catalytic competency and thus deficiencies are more consequential.

Certain features that play a role in the behavior observed in our simulations are known to differ between bacterial and eukaryotic OSTs and may contribute to greater sensitivity to serine-containing substrates in eukaryotes (Bause, 1984; Breuer et al., 2001; Chen et al., 2007; Gerber et al., 2013). We found that the conserved D at the p-2 position in the longer bacterial consensus sequence [D-X₋₁-N-X₊₁-S/T, where X₋₁ and X₊₁ can be any residue except proline (Young et al., 2006)] compensates partially for the lost favorable interactions in serine substrates; without the compensation offered by the aspartic acid at p-2 the difference in the interaction energy with PglB between NAT and NAS peptides would be almost double. In eukaryotic OSTs, the interaction at position p-2 with the asparagine residues of the transmembrane domain is not conserved, which may explain why eukaryotic OSTs have greatly reduced catalytic activity for N-X-S peptides (Bause, 1984; Breuer et al., 2001) while bacterial OSTs have a more modest difference in comparison (Chen et al., 2007; Gerber et al., 2013).

Van der Waals interactions between p+2 and T316 are part of the network of interactions that facilitate allosteric communication between the periplasmic domain and EL5. Although this interaction is directly affected by the lost methyl group, the NAS peptide was able to still form transient

interactions with T316. T316 is part of the conserved TIXE motif containing the catalytic E319 in bacteria; in eukaryotes, a different sequence is conserved: SVSE (Jaffee and Imperiali, 2011) (**Supplementary Figure S1B**). The replacement of threonine by the smaller serine in this sequence would further compromise the ability of serine-containing substrates to form stable interactions with EL5 in eukaryotes.

From our results we conclude that the loss of the methyl group in serine has repercussions far beyond the van der Waals interaction with I572 observed in threonine substrates. The loss of the methyl group destabilizes the entire constellation of interactions present with threonine and disrupts neighboring interactions between substrate and enzyme. This interrupts the allosteric communication between the periplasmic domain and EL5 mediated by acceptor substrates with threonine, allowing for opening of the periplasmic domain. While threonine at the +2 position acts as a pin in a latch to favor more closed periplasmic domain orientations reminiscent of the crystal structure of PglB with acceptor substrate and LLO analog (Napiórkowska et al., 2017), the more open periplasmic domain orientations seen with serine likely combine with the weakened substrate-enzyme interactions to increase substrate release and thereby decrease the formation of catalytically competent states.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

BM and FM conceived the study. BM performed the simulations and analyzed the corresponding results. BM wrote the paper with FM.

FUNDING

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM139316 and R01GM137529.

ACKNOWLEDGMENTS

The authors thank William R. Kobertz for helpful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.740904/full#supplementary-material>

REFERENCES

- Aebi, M. (2013). N-linked Protein Glycosylation in the ER. *Biochim. Biophys. Acta Mol. Cell Res.* 1833, 2430–2437. doi:10.1016/j.bbamcr.2013.04.001
- Bai, L., Wang, T., Zhao, G., Kovach, A., and Li, H. (2018). The Atomic Structure of a Eukaryotic Oligosaccharyltransferase Complex. *Nature* 555, 328–333. doi:10.1038/nature25755
- Barre, Y., Nothhaft, H., Thomas, C., Liu, X., Li, J., Ng, K. K., et al. (2017). A Conserved DGGK Motif Is Essential for the Function of the PglB Oligosaccharyltransferase from *Campylobacter* Jejuni. *Glycobiology* 27, 978–989. doi:10.1093/glycob/cwx067
- Bause, E. (1984). Model Studies on N-Glycosylation of Proteins. *Biochem. Soc. Trans.* 12, 514–517. doi:10.1042/bst0120514
- Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., et al. (2021). The InterPro Protein Families and Domains Database: 20 Years on. *Nucleic Acids Res.* 49, D344–D354. doi:10.1093/nar/gkaa977
- Braunger, K., Pfeffer, S., Shrimall, S., Gilmore, R., Berninghausen, O., Mandon, E. C., et al. (2018). Structural Basis for Coupling Protein Transport and N-Glycosylation at the Mammalian Endoplasmic Reticulum. *Science* 360, 215–219. doi:10.1126/science.aar7899
- Breuer, W., Klein, R. A., Hardt, B., Bartoschek, A., and Bause, E. (2001). Oligosaccharyltransferase Is Highly Specific for the Hydroxy Amino Acid in Asn-Xaa-Thr/Ser. *FEBS Lett.* 501, 106–110. doi:10.1016/S0014-5793(01)02641-2
- Chen, M. M., Glover, K. J., and Imperiali, B. (2007). From Peptide to Protein: Comparative Analysis of the Substrate Specificity of N-Linked Glycosylation in *C. Jejuni*. *Biochemistry* 46, 5579–5585. doi:10.1021/bi602633n
- Clauset, A., Newman, M. E. J., and Moore, C. (2004). Finding Community Structure in Very Large Networks. *Phys. Rev. E* 70, 6–2. doi:10.1103/PhysRevE.70.066111
- Eichler, J. (2013). Extreme Sweetness: Protein Glycosylation in Archaea. *Nat. Rev. Microbiol.* 11, 151–156. doi:10.1038/nrmicro2957
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* 103, 8577–8593. doi:10.1063/1.470117
- Fiser, A., Do, R. K. G., and Šali, A. (2000). Modeling of Loops in Protein Structures. *Protein Sci.* 9, 1753–1773. doi:10.1110/ps.9.9.1753
- Fraley, C., and Raftery, A. E. (2002). Model-based Clustering, Discriminant Analysis and Density Estimation. *J. Am. Stat. Assoc.* 97, 611. doi:10.1198/016214502760047131
- Frishman, D., and Argos, P. (1995). Knowledge-based Protein Secondary Structure Assignment. *Proteins* 23, 566–579. doi:10.1002/prot.340230412
- Gerber, S., Lizak, C., Michaud, G., Bucher, M., Darbre, T., Aebi, M., et al. (2013). Mechanism of Bacterial Oligosaccharyltransferase. *J. Biol. Chem.* 288, 8849–8861. doi:10.1074/jbc.M112.445940
- Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., and Caves, L. S. D. (2006). Bio3d: an R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* 22, 2695–2696. doi:10.1093/bioinformatics/btl461
- Helenius, A., and Aebi, M. (2004). Roles of N-Linked Glycans in the Endoplasmic Reticulum. *Annu. Rev. Biochem.* 73, 1019–1049. doi:10.1146/annurev.biochem.73.011303.073752
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual Molecular Dynamics. *J. Mol. Graph.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5
- Igura, M., Maita, N., Kamishikiry, J., Yamada, M., Obita, T., Maenaka, K., et al. (2008). Structure-guided Identification of a New Catalytic Motif of Oligosaccharyltransferase. *EMBO J.* 27, 234–243. doi:10.1038/sj.emboj.7601940
- Ihsen, J., Kowarik, M., Wiesli, L., Reiss, R., Wacker, M., and Thöny-Meyer, L. (2012). Structural Insights from Random Mutagenesis of *Campylobacter* Jejuni Oligosaccharyltransferase PglB. *BMC Biotechnol.* 12, 67. doi:10.1186/1472-6750-12-67
- Jaffee, M. B., and Imperiali, B. (2011). Exploiting Topological Constraints to Reveal Buried Sequence Motifs in the Membrane-Bound N-Linked Oligosaccharyl Transferases. *Biochemistry* 50, 7557–7567. doi:10.1021/bi201018d
- Jorgensen, W. L., Chandrasekhar, J., and Madura, J. D. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79, 926. doi:10.1063/1.445869
- Kelleher, D. J., and Gilmore, R. (2006). An Evolving View of the Eukaryotic Oligosaccharyltransferase. *Glycobiology* 16, 47R–62R. doi:10.1093/glycob/cwj066
- Kowarik, M., Young, N. M., Numao, S., Schulz, B. L., Hug, I., Callewaert, N., et al. (2006). Definition of the Bacterial N-Glycosylation Site Consensus Sequence. *EMBO J.* 25, 1957–1966. doi:10.1038/sj.emboj.7601087
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Cryst.* 26, 283–291. doi:10.1107/S0021889892009944
- Lee, H. S., and Im, W. (2017). Transmembrane Motions of PglB Induced by LLO Are Coupled with EL5 Loop Conformational Changes Necessary for OST Activity. *Glycobiology* 27, 787. doi:10.1093/glycob/cwx059
- Li, L., Woodward, R., Ding, Y., Liu, X.-w., Yi, W., Bhatt, V. S., et al. (2010). Overexpression and Topology of Bacterial Oligosaccharyltransferase PglB. *Biochem. Biophys. Res. Commun.* 394, 1069–1074. doi:10.1016/j.bbrc.2010.03.126
- Lizak, C., Gerber, S., Numao, S., Aebi, M., and Locher, K. P. (2011). X-ray Structure of a Bacterial Oligosaccharyltransferase. *Nature* 474, 350–355. doi:10.1038/nature10151
- Lizak, C., Gerber, S., Zinne, D., Michaud, G., Schubert, M., Chen, F., et al. (2014). A Catalytically Essential Motif in External Loop 5 of the Bacterial Oligosaccharyltransferase PglB. *J. Biol. Chem.* 289, 735–746. doi:10.1074/jbc.M113.524751
- Lohr, S. L. (1999). *Sampling: Design and Analysis*. California: Duxbury Press.
- MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., et al. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. *J. Phys. Chem. B* 102 (18), 3586–3616. doi:10.1021/jp973084f
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641. doi:10.1093/nar/gkz268
- Maita, N., Nyirenda, J., Igura, M., Kamishikiry, J., and Kohda, D. (2010). Comparative Structural Biology of Eubacterial and Archaeal Oligosaccharyltransferases. *J. Biol. Chem.* 285, 4941–4950. doi:10.1074/jbc.M109.081752
- Malaby, H. L. H., and Kobertz, W. R. (2014). The Middle X Residue Influences Cotranslational N-Glycosylation Consensus Site Skipping. *Biochemistry* 53, 4884–4893. doi:10.1021/bi500681p
- Mellquist, J. L., Kasturi, L., Spitalnik, S. L., and Shakin-Eshleman, S. H. (1998). The Amino Acid Following an Asn-X-Ser/Thr Sequon Is an Important Determinant of N-Linked Core Glycosylation Efficiency. *Biochemistry* 37, 6833–6837. doi:10.1021/bi972217k
- Mohorko, E., Glockshuber, R., and Aebi, M. (2011). Oligosaccharyltransferase: the central Enzyme of N-Linked Protein Glycosylation. *J. Inherit. Metab. Dis.* 34, 869–878. doi:10.1007/s10545-011-9337-1
- Napiórkowska, M., Boilevin, J., Sovdat, T., Darbre, T., Reymond, J.-L., Aebi, M., et al. (2017). Molecular Basis of Lipid-Linked Oligosaccharide Recognition and Processing by Bacterial Oligosaccharyltransferase. *Nat. Struct. Mol. Biol.* 24, 1100–1106. doi:10.1038/nsmb.3491
- Nothhaft, H., and Szymanski, C. M. (2013). Bacterial Protein N-Glycosylation: New Perspectives and Applications. *J. Biol. Chem.* 288, 6912–6920. doi:10.1074/jbc.R112.417857
- Pedebos, C., Arantes, P. R., Giesel, G. M., and Verli, H. (2015). In silico Investigation of the PglB Active Site Reveals Transient Catalytic States and Octahedral Metal Ion Coordination. *Glycobiology* 25, 1183–1195. doi:10.1093/glycob/cwv053
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* 26 (16), 1781–1802. doi:10.1002/jcc.20289
- Ramírez, A. S., Kowal, J., and Locher, K. P. (2019). Cryo-electron Microscopy Structures of Human Oligosaccharyltransferase Complexes OST-A and OST-B. *Science* 366, 1372–1375. doi:10.1126/science.aaz3505
- Romanowska, J., Nowiński, K. S., and Trylska, J. (2012). Determining Geometrically Stable Domains in Molecular Conformation Sets. *J. Chem. Theor. Comput.* 8, 2588–2599. doi:10.1021/ct300206j
- Ryckaert, J. P., Cicciotti, G., and Berendsen, H. J. C. (1997). Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* 23, 327–341.
- Scarabelli, G., and Grant, B. J. (2014). Kinesin-5 Allosteric Inhibitors Uncouple the Dynamics of Nucleotide, Microtubule, and

- Neck-Linker Binding Sites. *Biophys. J.* 107, 2204–2213. doi:10.1016/j.bpj.2014.09.019
- Szymanski, C. M., Burr, D. H., and Guerry, P. (2002). Campylobacter Protein Glycosylation Affects Host Cell Interactions. *Infect. Immun.* 70, 2242–2244. doi:10.1128/IAI.70.4.2242-2244.2002
- Szymanski, C. M., and Wren, B. W. (2005). Protein Glycosylation in Bacterial Mucosal Pathogens. *Nat. Rev. Microbiol.* 3, 225–237. doi:10.1038/nrmicro1100
- Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., et al. (2002). N-Linked Glycosylation in Campylobacter Jejuni and its Functional Transfer into *E. coli*. *Science* 298, 1790–1793. doi:10.1126/science.298.5599.1790
- Yao, X.-Q., Malik, R. U., Griggs, N. W., Skjærven, L., Traynor, J. R., Sivaramakrishnan, S., et al. (2016). Dynamic Coupling and Allosteric Networks in the α Subunit of Heterotrimeric G Proteins. *J. Biol. Chem.* 291, 4742–4753. doi:10.1074/jbc.M115.702605

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Morgan and Massi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Structural Characterization of the RNA-Binding Protein SERBP1 Reveals Intrinsic Disorder and Atypical RNA Binding Modes

OPEN ACCESS

Edited by:

Vincenzo Venditti,
Iowa State University, United States

Reviewed by:

Lalit Deshmukh,
University of California, San Diego,
United States
Jun Zhang,
University of Alabama at Birmingham,
United States
Steven. T Whitten,
Texas State University, United States

*Correspondence:

David S. Libich
libich@uthscsa.edu

†ORCID:

Antoine Baudin
orcid.org/0000-0002-1390-1196
Emily. E. Selig
orcid.org/0000-0002-3012-6814
David. S. Libich
orcid.org/0000-0001-6492-2803

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 20 July 2021

Accepted: 07 September 2021

Published: 24 September 2021

Citation:

Baudin A, Moreno-Romero AK, Xu X, Selig EE, Penalva LOF and Libich DS (2021) Structural Characterization of the RNA-Binding Protein SERBP1 Reveals Intrinsic Disorder and Atypical RNA Binding Modes. *Front. Mol. Biosci.* 8:744707. doi: 10.3389/fmolb.2021.744707

Antoine Baudin^{1,2†}, Alma K. Moreno-Romero^{1,2}, Xiaoping Xu^{1,2}, Emily E. Selig^{1,2†}, Luiz O. F. Penalva^{1,3} and David S. Libich^{1,2*†}

¹Greehey Children's Cancer Research Institute, The University of Texas Health Science Center at San Antonio, San Antonio, TX, United States, ²Department of Biochemistry and Structural Biology, The University of Texas Health Science Center at San Antonio, San Antonio, TX, United States, ³Department of Cell Systems and Anatomy, The University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

RNA binding proteins (RBPs) are essential for critical biological processes such as translation regulation and mRNA processing, and misfunctions of these proteins are associated with diseases such as cancer and neurodegeneration. SERBP1 (SERPINE1 mRNA Binding Protein 1) is an RBP that comprises two RG/RGG repeat regions yet lacks other recognizable RNA-binding motifs. It is involved in mRNA maturation, and translational regulation. It was initially identified as a hyaluronic acid binding protein, but recent studies have identified central roles for SERBP1 in brain function and development, especially neurogenesis and synaptogenesis. SERBP1 regulates One-carbon metabolism and epigenetic modification of histones, and increased SERBP1 expression in cancers such as leukemia, ovarian, prostate, liver and glioblastoma is correlated with poor patient outcomes. Despite these important regulatory roles for SERBP1, little is known about its structural and dynamic properties, nor about the molecular mechanisms governing its interaction with mRNA. Here, we define SERBP1 as an intrinsically disordered protein, containing highly conserved elements that were shown to be functionally important. The RNA binding activity of SERBP1 was explored using solution NMR and other biophysical techniques. The outcome of these experiments revealed that SERBP1 preferentially samples compact conformations including a central, stable α -helix and show that SERBP1 recognizes G-rich RNA sequences at the C-terminus involving the RGG box and neighboring residues. Despite the role in RNA recognition, the RGG boxes do not seem to stabilize the central helix and the central helix does not participate in RNA binding. Further, SERBP1 undergoes liquid-liquid phase separation, mediated by salt and RNA, and both RGG boxes are necessary for the efficient formation of condensed phases. Together, these results provide a foundation for understanding the molecular mechanisms of SERBP1 functions in physiological and pathological processes.

Keywords: SERBP1, RNA binding protein, mRNA binding, intrinsically disordered protein, NMR

INTRODUCTION

SERBP1 (SERPINE1 mRNA binding protein 1) is a highly conserved RNA binding protein (RBP) containing two RG/RGG repeat regions yet lacks other readily recognizable, canonical or structured RNA binding motifs. RBPs containing RG/RGG repeats are essential for normal brain function and have been implicated in neurological and neuromuscular diseases as well as certain cancers (Järvelin et al., 2016; Hentze et al., 2018). High SERBP1 expression in glioblastoma multiform (GBM) is linked to poor patient outcome and response to therapy while *in vitro* and *in vivo* studies showed that expression levels of SERBP1 affect several related cancer phenotypes, stemness, neuronal differentiation and tumor growth (Koensgen et al., 2007; Serce et al., 2012; Costa et al., 2014; Wang et al., 2017; Kosti et al., 2020). We recently established that SERBP1 functions as a novel oncogenic factor in GBM through regulation of One-carbon metabolism, methionine production, and histone methylation (Kosti et al., 2020). Moreover, knockdown of SERBP1 affected the expression of genes linked to neurogenesis and synaptogenesis. A strong negative expression correlation was observed between genes in these categories and SERBP1 both in brain and patient-derived GBM samples, implicating SERBP1 in brain function and development (Kosti et al., 2020). Two proximity-dependent biotinylating screening studies identified SERBP1 as an interaction partner of RBPs known to regulate synaptic plasticity such as FMR1, FXR1, FXR2, CAPRIN1, and SYNCRIP (Youn et al., 2018; Go et al., 2021). Additionally, SERBP1 has a role in the SUMOylation of certain proteins (Lemos and Kobarg, 2006), and is itself SUMOylated (Hendriks et al., 2014) on a lysine-rich sequence between the two RGG boxes. SUMOylation of SERBP1 has been suggested as a factor in the development of GBM, as aberrations in SUMOylation pathways can lead to the development of cancer (Fox et al., 2019).

SERBP1 structure and its RNA recognition and binding activity are poorly characterized. SERBP1 was reported to bind preferentially to GC-rich motifs (Kosti et al., 2020), subsequent studies revealed that these motifs could include G-quadruplexes (Su et al., 2021). SERBP1 was identified in the structures of non-translating 80 S ribosomes, blocking the mRNA entrance channel suggesting that it serves to regulate mRNA translation (Ahn et al., 2015; Brown et al., 2018; Muto et al., 2018). Like SERBP1, the SARS-CoV-2 non-structural protein 1 (Nsp1) mediates translation inhibition of mRNA by binding and blocking the ribosomal mRNA channel through interactions with its disordered C-terminal domain (Schubert et al., 2020). Translational regulation does not account for the RNA binding activity of SERBP1 and thus it likely participates in additional regulatory processes. For instance, *vig* and *vig2*, SERBP1 *Drosophila* homologues, have been identified in RNAi complexes and heterochromatin, (Gracheva et al., 2009), and as regulators of histone genes (Tsui et al., 2018). As noted above, SERBP1 also interacts with arginine-methylated and stress granule-associated proteins (Youn et al., 2018), and selective methylation of either RGG repeat region modulates its subcellular distribution between the nucleus and cytoplasm

(Lee et al., 2012a). To better understand the structural determinants guiding SERBP1 roles in these diverse physiological and pathological processes, a structural and biophysical analysis was undertaken to characterize its structural and functional properties. A combination of solution NMR spectroscopy and biophysical assays were used to define the structural and dynamic properties of SERBP1. The data reveal that SERBP1 is primarily an intrinsically disordered protein (IDP) yet it adopts compact, and partially structured conformations. Moreover, the SERBP1 binding site of a GC-rich RNA oligonucleotide was identified, and it was determined that RNA binding negatively affects the liquid-liquid phase separation (LLPS) propensity of SERBP1. Together, these results are a step toward understanding the multifunctional nature of SERBP1 and determining the structural underpinnings of its diverse physiological roles in healthy cells as well as its aberrant function in GBM and other tumor types.

METHODS

Phylogenetics

COBALT was used to perform multiple sequence alignment to identify conserved regions among SERBP1 homologues from eleven different species (nine different phyla) (Papadopoulos and Agarwala, 2007). Amino-acid identity was prioritized, and an amino-acid was defined as conserved if it was present in at least seven species. Protein sequences were obtained from the NCBI (<https://www.ncbi.nlm.nih.gov/>) with the following accession numbers: *H. sapiens* (NP_001018077.1), *Saccoglossus kowalevskii* (XP_002740974.1), *Acanthaster planci* (XP_022107686.1), *Drosophila melanogaster* (NP_523572.1), *Stegodyphus mimosarum* (KFM69747.1), *Centruroides sculpturatus* (XP_023216060.1), *Octopus sinensis* (XP_029640355.1), *Teladorsagia circumcincta* (PIO76243.1), *Echinococcus granulosus* (XP_024349465.1), *Nematostella vectensis* (XP_001637231.1), *A. queenslandica* (XP_011406826.2).

Protein Expression and Purification

The SERBP1 189–400 construct (Addgene accession number 172315) was expressed in *E. coli* as previously described (Baudin et al., 2021). The genes for full-length SERBP1 and a SERBP1 149–400 construct were amplified by PCR (primer are shown in **Supplementary Table S1**), digested by *KasI* and *BamHI* restriction enzymes and cloned into a custom pAG8Ha-His vector that introduced an 8x histidine tag followed by a TEV cleavage site N-terminal to the coding sequence. Plasmids were transformed by heat shock into *E. coli* BL21 Star™ (DE3) (Invitrogen, MA) and a colony from the resultant agar plate was used to inoculate a 5 ml LB starter culture that was grown at 37°C for 6–8 h. The LB starter culture was used to inoculate a 100 ml M9 preculture that was grown overnight at 37°C with shaking. The overnight preculture was then used to inoculate an expression culture of 900 ml M9 minimal media. Both the 100 ml starter and 900 ml expression cultures were supplemented with ¹⁵NH₄Cl (¹³C-glucose) for isotopic enrichment. For non-

isotopically enriched protein, the overnight culture was 10 ml LB and 4 ml was used to inoculate a 1 L LB expression culture. All cultures were supplemented with 100 µg/ml of ampicillin. Expression cultures were grown at 37°C in baffled Fernbach flasks with shaking, and protein expression was induced at OD₆₀₀ ~0.6–0.8 with 1 mM IPTG and continued for 3 or 6 h, for LB or M9 cultures, respectively. The cells were harvested by centrifugation at 4000 g for 20 min and the resulting pellets were stored at –80°C.

SERBP1 constructs were purified as previously described for SERBP1 189–400 (Baudin et al., 2021). Briefly, frozen *E. coli* pellets were thawed and resuspended in 8 M urea, 50 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole, lysed by sonication (6 cycles of 10 s on, 30 s off), and the lysate was cleared by centrifugation for 30 min at 45000 g at 4°C. The supernatant was applied to a 5 ml HisTrap HP column (Cytiva, MA, United States), equilibrated with the lysis buffer, washed with 25 column volumes of the same buffer, and eluted with 8 M urea, 50 mM Tris pH 8.0, 150 mM NaCl, 500 mM imidazole. Eluted protein was concentrated with a 3 kDa (149–400, 189–400) or 10 kDa (full-length) cutoff Amicon centrifugal concentrator (Merck, NJ, United States) to ~4 ml and diluted into 40 ml of 100 mM sodium phosphate buffer, pH 7, 1 mM PMSF, 1 mM EDTA buffer. Precipitated protein was removed by centrifugation at 3,000 g for 15 min, and the protein was then dialyzed against 100 mM sodium phosphate buffer pH 7 (3 x 2 L), on the third change 500 µL of 1.6 mg/ml Tobacco Etch Virus (TEV) protease was added and the mixture was incubated at room temperature overnight. The 8x His tag was not removed from full-length SERBP1. Post-dialysis, the sample was centrifuged at 3,000 g for 15 min to remove any precipitate and concentrated to approximately 4 ml with an Amicon centrifugal concentrator. A final cation exchange chromatography polishing step was used for SERBP1 149–400 and 189–400. The 4 ml sample was diluted into 60 ml with 50 mM sodium acetate buffer pH 4.5, 0.5 mM PMSF, 0.5 mM EDTA, loaded on a 5 ml SP Sepharose fast-flow column (Cytiva, MA, United States), washed with 20 column volumes of the same buffer and eluted with a 0.05–1 M NaCl gradient. Fractions containing SERBP1 149–400 were concentrated, and buffer exchanged into NMR buffer (20 mM sodium phosphate buffer pH 6.9, 60 mM NaCl, 1 mM PMSF, 0.2 mM EDTA, 10 % D₂O).

NMR Spectroscopy

All experiments were recorded on a Bruker Avance NEO spectrometer operating at a proton Larmor frequency of 700.13 MHz, at a temperature of 5°C using a 5 mm TCI z-axis gradient cryogenic probe. Data were processed with the NMRPipe software suite (Delaglio et al., 1995) and analyzed with CCPNMR Analysis 2.5 software (Skinner et al., 2016). SERBP1 149–400 ¹H, ¹³C_α, ¹³C_β, ¹³C' and ¹⁵N backbone resonances were assigned through the analysis of a set of 2D and 3D experiments, namely ¹H,¹⁵N-HSQC, HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO and HCC(CO)NH, recorded on a ¹³C,¹⁵N-labeled sample at a concentration of 300 µM, in NMR buffer. The ¹H,¹⁵N-HSQC was recorded with 128* x 1024* complex points in the indirect (¹⁵N) and direct (¹H)

dimensions, corresponding to acquisition times of 75.2 and 112.6 ms, respectively. Acquisition parameters for the HNCO and HN(CA)CO consisted of 32* x 64* x 1024* complex points in the indirect (F1, ¹³C), (F2, ¹⁵N) and direct (F3 ¹H) dimensions, corresponding to acquisition times of 16.5, 37.6, 112.6 ms, respectively; acquisition parameters for the HNCACB, CBCA(CO)NH and HCC(CO)NH consisted of 128* x 64* x 1024* complex points in the indirect (F1, ¹³C), (F2, ¹⁵N) and direct (F3 ¹H) dimensions, corresponding to acquisition times of 11, 37.6, 112.6 ms, respectively. All 3D experiments were recorded in non-uniform sampling (NUS) mode (Delaglio et al., 2017) with a sampling density of 20%, and the spectra were reconstructed using the SMILE algorithm implemented in NMRPipe (Ying et al., 2017).

¹⁵N R₁ and R₂ relaxation rates were calculated from T₁ and T_{1ρ} experiments, recorded on 50 and 100 µM samples for SERBP1 189–400 and SERBP1 149–400, respectively, in NMR buffer using 64* x 1024* complex data points in the indirect (¹⁵N) and direct (¹H) dimensions corresponding to acquisition times of 37.6 and 112.6 ms, respectively. The ¹⁵N T₁ experiment consisted of eight interleaved spectra with the following relaxation delays: 40, 80, 200, 280, 300, 400, 600, and 800 ms. The T_{1ρ} experiment was recorded using a B₁ field of 1400 Hz and eight interleaved spectra with the following relaxation delays: 1, 21, 31, 41, 61, 81, 121 and 161 ms. ¹⁵N R₂ rates were calculated using the following equation (Massi et al., 2004):

$$R_{1\rho} = R_1 \cos^2 \theta + R_2 \sin^2 \theta \quad (1)$$

with $\theta = \arctan(\omega_1/\Omega)$, where ω_1 is the B₁ field strength (here 1400 Hz) and Ω is the offset from the spinlock carrier frequency. ¹H-¹⁵N heteronuclear NOE experiments were recorded on the same samples and consisted of two interleaved experiments, with and without proton saturation, using a recycle delay of 4 s. Spectra were acquired with 64* x 1024* complex data points in the indirect (¹⁵N) and direct (¹H) dimensions corresponding to acquisition times of 37.6 and 112.6 ms, respectively.

For RNA binding experiments, SERBP1 189–400 or 149–400 samples were diluted to 50 µM in 500 µL in NMR buffer and added to a 5 mm NMR tube. The RNA sequence 5'-GCGCGGG-3', representing a G-quartet, was synthesized (IDT, IA), desalted, dried, and resuspended in RNAase-free water (Qiagen, MD) to a concentration of 3.125 mM. The RNA stock was titrated into the SERBP1 sample to final RNA:SERBP1 ratios of 2:5, 4:5, 6:5, and 8:5. This required the addition of a maximum of 12 µL of the RNA stock, thus the effect of dilution is negligible. ¹H,¹⁵N-HSQC spectra were recorded at 4°C for each titration point with 64* x 1024* complex data points in the indirect (¹⁵N) and direct (¹H) dimensions corresponding to acquisition times of 37.6 and 112.6 ms, respectively. Spectra were apodized with a sine bell function and zero filled to twice the number of acquired points for data analysis. Chemical shift perturbations (CSP) were calculated by weighting the ¹H and ¹⁵N chemical shifts with respect to their gyromagnetic ratio using the following equation (Williamson, 2013):

$$\Delta\delta = \sqrt{(\delta^1\text{H})^2 + 0.1(\delta^{15}\text{N})^2} \quad (2)$$

CSPs were considered significant when they were higher than the standard deviation of $\Delta\delta_{\max}$ for all residues (Williamson, 2013).

CD Spectroscopy

Circular dichroic spectra were recorded on 10 μM samples of SERBP1 149–400 or 189–400 dissolved in 20 mM Na_2HPO_4 pH 6.9, 60 mM NaCl in a 2 mm pathlength circular cuvette using a Jasco 810 spectropolarimeter (Jasco, OK) at a scan speed of 50 nm/min with 0.5 nm. Temperature was controlled with a recirculating external water bath and was allowed to equilibrate for 20 min before recording data after each ramp. Each temperature point (20–80°C in 10°C steps) was recorded in triplicate, averaged and converted to mean residual ellipticity using previously described relationships (Chemes et al., 2012).

Size Exclusion Chromatography

Size exclusion chromatography (SEC) experiments were performed using a BioRad NGC fast-performance liquid chromatography system (BioRad, CA, United States) equipped with a Superdex 200 10 x 300 mm analytical size exclusion column (Cytiva, MA, United States), equilibrated with NMR buffer (20 mM Na_2HPO_4 pH 6.9, 60 mM NaCl, 1 mM PMSF, 0.2 mM EDTA) in the presence or absence of 2 M guanidinium hydrochloride (GdnHCl). 200 μL of a 15 μM sample of SERBP1 189–400 or 149–400 was applied to the column. The flow rate was 0.5 ml/min, and elution was monitored at three wavelengths: 215, 280, and 340 nm.

Analytical Ultracentrifugation

SERBP1 189–400 and 149–400 stock samples were diluted to a concentration of 25 μM into 20 mM Na_2HPO_4 pH 7.0, 60 mM NaCl buffer, in presence or absence of 2 M GdnHCl. 360 μL samples were loaded into one sector of a 12 mm double-sector epon-filled centerpiece, and 400 μL of a reference solution of the sample buffer was loaded into the other sector. Radial absorbance scans were collected at 4 min intervals at 280 nm, using an XL-1 analytical centrifuge (Beckman Coulter, CA, United States) equipped with a Ti60 rotor at 40000 rpm, 20°C. Total experiment time was ~10 h. Data was fit using a continuous $c(s)$ distribution model using the SEDFIT software (Schuck, 2000), using adjusted buffer density (1.04861) and buffer relative viscosity (1.09833) values to account for GdnHCl.

Dynamic Light Scattering

A SERBP1 189–400 stock sample was diluted to a concentration of 125 μM in 60 μL of 20 mM Na_2HPO_4 pH 7.0, 60 mM NaCl buffer, in presence or absence of 2 M GdnHCl, and transferred to a low-volume quartz cuvette. Measurements were conducted on a DynaPro NanoStar dynamic light scattering (DLS) instrument (Wyatt, CA, United States) at 20°C. For each sample 150 10 s acquisitions were collected and averaged. Data were analyzed using the algorithms available in the instrument control software and output as size distributions. Experimentally measured R_h values were compared with calculated R_h values using the following equation proposed by Marsh and Forman-Kay to account for intrinsic disorder (Marsh and Forman-Kay, 2010):

$$R_h^{\text{IDP}} = 2.49N^{0.509} \quad (3)$$

where N is the number of residues in the protein polymer chain.

Phase Separation

SERBP1 149–400 was labeled with the DyLight™ 650 fluorophore (Thermo Fisher Scientific, MO, United States) at the N-terminus using *S. aureus* sortase A following established protocols (Theile et al., 2013; Antos et al., 2017). The sortase-recognition peptide KLPETGG was synthesized, HPLC purified, lyophilized (Genscript, NJ) and reacted with the fluorophore following the manufacturer's instructions. The fluorescently labeled peptide and SERBP1 were mixed at a 1:1 molar ratio. Recombinant sortase A was added to a final concentration of 2.5 μM and the reaction was allowed to proceed overnight at room temperature. Labeled SERBP1 was separated from unconjugated peptide and free fluorophore by passing over a Superdex 75 HiLoad 16/60 size exclusion column (Cytiva, MA, United States).

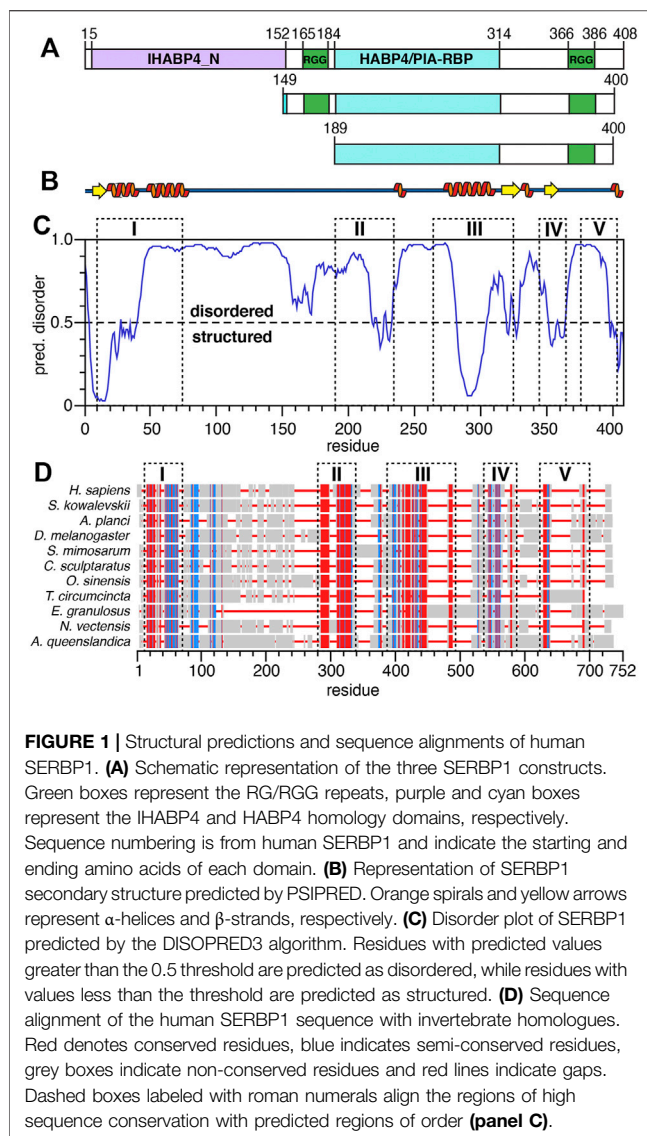
Torula yeast RNA (MilliporeSigma, MA, United States) was dissolved at 10 mg/ml in 20 mM Tris pH 7.4, 150 mM NaCl, centrifuged at 13000 g for 10 min, desalted using a PD-10 desalting column (Cytiva, MA, United States), and diluted to a final stock concentration of 3.6 mg/ml determined by A_{260} . Unlabeled SERBP1 149–400 was mixed with DyLight 650-labeled SERBP1 149–400 at a molar ratio of 0.03% and dissolved to 6 or 12 μM in 154 mM NaCl, 64 mM Tris pH 7.5, 12.8 % (v/v) glycerol, 1.28 mM DTT, 12.8 % (w/v) PEG3000. Stock RNA was added to a final concentration of 0.4, 0.2, or 0.05 mg/ml. Samples of SERBP1 189–400 were prepared identically except without fluorescent labeling. Control samples consisting of buffer only or buffer plus RNA only remained clear. SERBP1 samples used in the salt series were prepared similarly except the buffer the protein was dissolved in contained only 20 mM Tris pH 7.5 and 0, 0.05, 0.15, 0.3, or 1 M NaCl.

Phase separation was assayed by transferring 4.5 μL samples to chambered glass coverslips (Grace Biolabs, OR). Chambers were sealed with a second coverslip to reduce evaporation and incubated for 10 min at ambient temperature before imaging with an Olympus FV3000 inverted confocal microscope (Olympus, PA, United States) operating at 1% laser power on the 640 nm channel. Images were acquired simultaneously in differential interference contrast (DIC) and fluorescent modes. Fluorescence recovery after photobleaching (FRAP) measurements were conducted on 12 μM SERBP1 149–400 samples with low (0.05 mg/ml) or no RNA, and generally recovered to ~80% of the initial fluorescent intensity. Image contrast was adjusted globally, and droplet area was measured using the appropriate subroutines in Fiji (ImageJ) (Schindelin et al., 2012).

RESULTS

SERBP1 is an Atypical RBP Lacking Canonical RNA Binding Motifs

The largest and predominant SERBP1 isoform contains 408 amino-acids encompassing distinct domains including two



hyaluronic acid binding protein (HABP) homology domains, IHABP4 (intracellular HABP4, spanning residues 5–152) and HABP4 (residues 189–314) as indicated by UniProtKB entry Q8NC51 (**Figure 1A**). Two distinct RG/RGG repeats or boxes comprising residues 165–184 and 366–386 are important for RNA-binding and LLPS (Chong et al., 2018). Aside from the RGG boxes, SERBP1 does not contain any other identifiable RNA-binding motif such as an RNA-recognition motif (RRM), zinc-finger, or K-homology (KH) domain (Maris et al., 2005; Valverde et al., 2008; Cook et al., 2015). We used the PSIPRED (Buchan and Jones, 2019) and DISOPRED3 (Ward et al., 2004) webserver to predict regions of secondary structure and intrinsic disorder respectively in SERBP1 (**Figures 1B,C**). The results of the PSIPRED algorithm, shown as a cartoon representation (**Figure 1B**), predicts two regions of secondary structure roughly spanning residues 5–80 and 270–360. These results align well with the DISOPRED3 algorithm predictions which indicate SERBP1 is predominantly disordered, except for the first

~40 N-terminal residues and residues 285–300 (**Figure 1C**). Several other segments of SERBP1 approach the cutoff threshold for order (e.g., 230–240, 320–330, 350–365, 395–408) and align well with the PSIPRED secondary structure predictions indicating the potential for transiently-formed structure.

SERBP1 is highly conserved among vertebrates, thus to gain insight into the functional relevance of the different domains of SERBP1, we aligned human SERBP1 to several homologous proteins from invertebrates (**Figure 1D**). The alignments are shown with gaps, and dashed boxes denote regions with conserved (red) or semi-conserved (blue) sequence homology. Full alignments are presented in **Supplementary Figure S1**. Notably, the highly homologous regions align well with the predicted secondary structure (**Figure 1B**) and ordered regions (**Figure 1C**), in particular residues 285–300, which are predicted to be helical, are highly conserved (region “III” in **Figures 1C,D**), indicating a potentially crucial role for SERBP1 function. Moreover, it appears that although the two RGG boxes show high conservation among vertebrates, these regions are highly heterogeneous among non-vertebrates (**Figure 1D**; **Supplementary Figure S1**). To expand on the role of the RGG boxes for SERBP1 function, we designed two truncated constructs: 149–400 or 189–400 which contain either both (149–400) or only the C-terminal (189–400) RGG box(es) (**Figure 1A**). In both constructs, we also removed eight C-terminal residues (401–408) that are hydrophobic and contributed to instability and were consistently degraded from SERBP1. To confirm that the truncated proteins faithfully reproduce the structural features of the full-length protein, ^1H , ^{15}N -HSQC spectra were recorded. Overlays of the spectra from full-length SERBP1 with spectra from each of the truncation proteins are almost identical indicating that the truncations did not lead to major structural changes (**Supplementary Figure S2**). The N-terminal region, that was predicted as structured, was not detected as such in the full-length spectrum, and thus we decided to focus our study on the truncated proteins.

Structural Characterization of SERBP1 Indicates the Presence of a Stable α -Helix

We previously reported the backbone resonance assignments for SERBP1 189–400 (BMRB accession number 50953) and here we report the backbone resonance assignments for SERBP1 149–400 (**Supplementary Figure S3**). Over 77% of the backbone resonances from the 252 residue SERBP1 149–400 were assigned, excluding the 11 proline residues. Of the 56 residues not assigned, 32 are arginines or glycines that belong to the two RG/RGG repeats and thus, because of the inherent sequence degeneracy in these regions, could not be unambiguously assigned. The remaining 24 residues were either ambiguous or were severely overlapped in the spectrum. Overall, assignments of SERBP1 149–400 Ca , $\text{C}\beta$, and C' chemical shifts were 81, 84.6, and 77.8% complete respectively. Analysis of the Ca , $\text{C}\beta$, C' , H_N , N , and Ha chemical shifts using the Secondary Structure Propensity (SSP) algorithm (Marsh et al., 2006) indicate that the majority of SERBP1 189–400 is disordered, except for residues

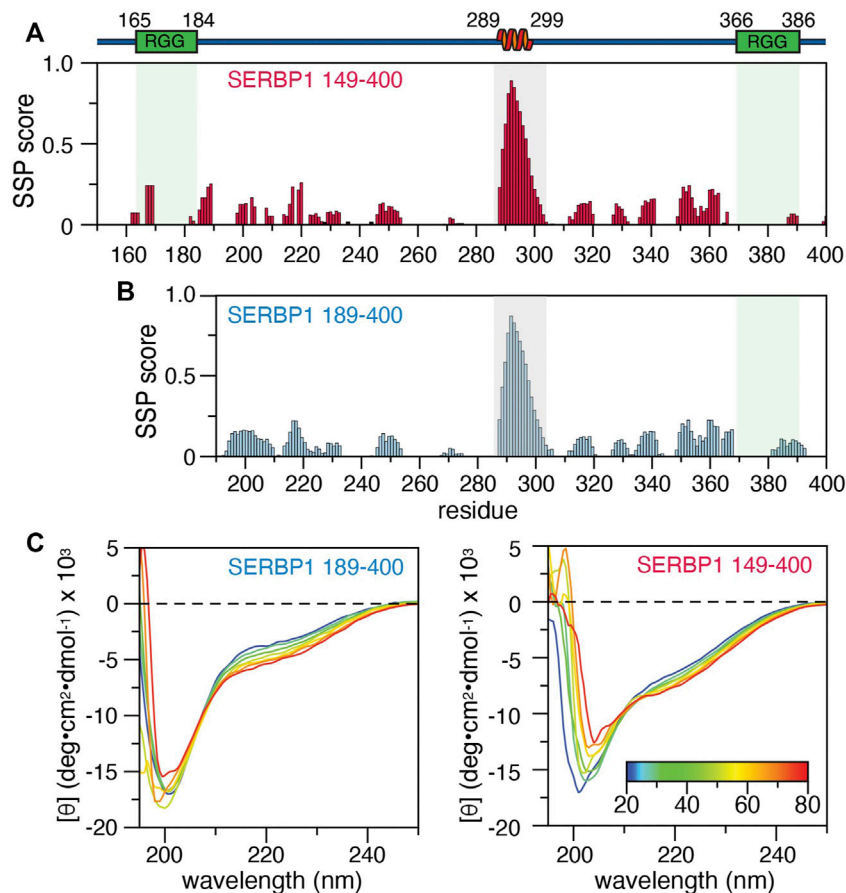


FIGURE 2 | Secondary structure of SERBP1 149–400 and 189–400. Secondary structure propensity calculated from backbone chemical shifts by the SSP algorithm for **(A)** SERBP1 149–400 and **(B)** 189–400. Gray boxes outline residues 289–299, identified as having a high α -helical character. A cartoon representation of the α -helix and relative position of the RGG boxes (green) is shown above the plots. **(C)** CD spectra of SERBP1 189–400 (**right**) and SERBP1 149–400 (**left**) acquired at 10°C intervals from 20°C to 80°C.

289–299, which have significant α -helical character (**Figure 2A**). A similar analysis for the 149–400 construct revealed α -helical propensity for the same residues (**Figure 2B**), indicating that the first RGG domain does not influence the formation or stability of the helix. Similar results were obtained with the δ 2D (Camilloni et al., 2012) and TALOS-N (Shen and Bax, 2013) algorithms which indicate substantial (>80 %) α -helical propensity for residues 289–299 in both the SERBP1 149–400 and 189–400 constructs (**Supplementary Figure S4**). Circular dichroic (CD) spectra of SERBP1 149–400 and 189–400 recorded to measure thermal denaturation have a similar appearance and contain features indicative of IDPs such as the strong negative transition at 200 nm (**Figure 2C**). For both constructs, a negative transition at approximately 222 nm becomes more negative as the temperature is increased from 20 to 80°C. The temperature gradient also reveals an isodichroic point at 212 nm indicative of a two-state transition that could translate to exchange between partially ordered and completely disordered conformations (Greenfield, 2006). Deconvolution of these CD spectra to estimate helical content is complicated by the inherent

contributions from polyproline II conformations that contribute to the observed transitions at 200 and 212 nm (Kjaergaard and Poulsen, 2011). Taken together these data are consistent with the chemical shift analysis (**Figures 2A,B**), secondary structure, and disorder predictions **Figures 1B,C** and indicate that SERBP1 is primarily disordered with some α -helical character.

To further define the structural properties of SERBP1, we investigated the fast timescale dynamics of the truncation mutants by measuring the ^{15}N R_1 and R_2 relaxation rates, as well as the ^1H - ^{15}N heteronuclear NOE (**Figure 3; Supplementary Figure S5**). For the 189–400 construct, the average R_1 and R_2 values are 1.5 s^{-1} and 9.6 s^{-1} , respectively, except for residues 289–299 in the region identified as helical by SSP, which have average R_1 and R_2 of 1.2 s^{-1} and 22.5 s^{-1} respectively. We observed similar average R_1 and R_2 (1.5 s^{-1} and 9.5 s^{-1} respectively) along with variations in the same region (289–299) for SERBP1 149–400 (**Supplementary Figure S5**). Average heteronuclear NOE values for both constructs were 0.4 with the residues in the 289–299 stretch displaying values approaching 0.75, indicating the motions of these amino acids are

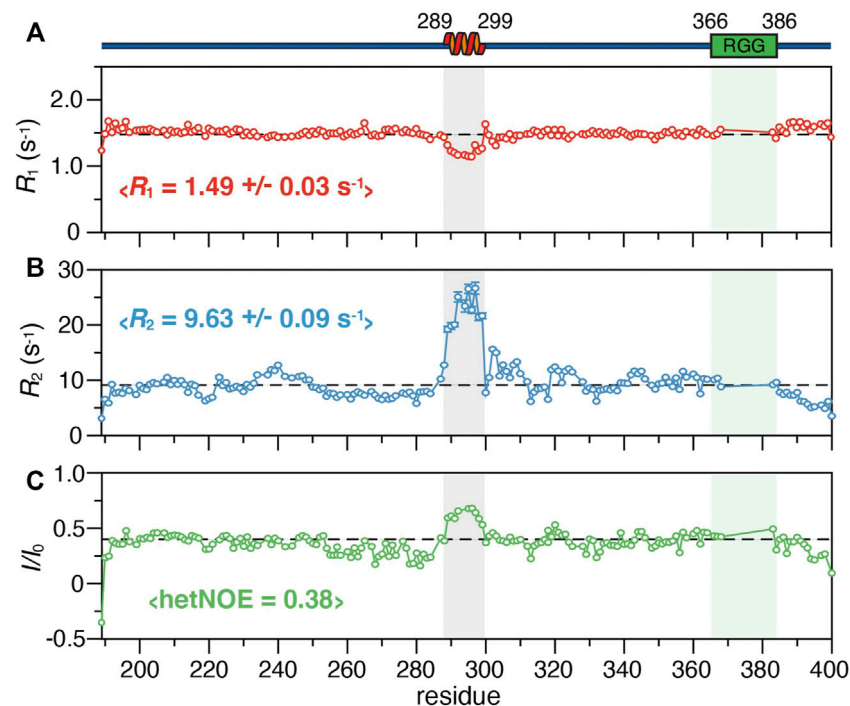


FIGURE 3 | ^{15}N relaxation parameters of SERBP1 189–400 including (A) R_1 and (B) R_2 rates, and (C) the ^1H - ^{15}N heteronuclear NOE plotted against the protein sequence. The dashed lines represent the average value for each experiment, also indicated in brackets. A cartoon representation of the α -helix identified from chemical shift information and relative position of the RGG boxes (green) is shown above the plots. The shaded boxes align the helical region and RGG box over all panels.

more restricted than the rest of the chain. These data reveal that the two truncated proteins retain virtually identical dynamics and contain a stable α -helix comprising residues 289–299, the stability of which seems to be independent of the RGG boxes.

SERBP1 Behaves as a Compact, Monomeric IDP

Size exclusion chromatography (SEC), dynamic light scattering (DLS) and analytical ultracentrifugation (AUC) were employed to assess the oligomeric state of SERBP1 and to determine if the stable α -helix acts as a dimerization interface, as has previously been described for TAR DNA-binding protein (TDP-43) (Conicella et al., 2016). At physiological conditions, SERBP1 189–400 elutes at 14.7 ml (Figure 4A), which, according to column calibration (data not shown) corresponds to a molecular weight of 51.6 kDa, suggesting a possible dimerization of SERBP1 (the molecular weight of monomers is 23.7 kDa) if SERBP1 were a globular, folded protein. However, due to their extended conformations, IDPs are expected to elute earlier in SEC corresponding to larger apparent molecular weights (Uversky, 2012). The SEC experiment was repeated in the presence of 2 M of GdnHCl, resulting in an elution volume of 13.3 ml, corresponding to a molecular weight of 95.9 kDa (Figure 4A). Similar behavior was observed for SERBP1 149–400 which displayed elution volumes of 14.8 and 12.8 ml in absence and in presence of 2 M GdnHCl, respectively (Figure 4B). If SERBP1 formed oligomers in the absence of

GdnHCl, the elution volume should increase upon the addition GdnHCl, contrary to the observed decrease. These results suggest that SERBP1 becomes more extended, occupying a much larger steric volume in 2 M GdnHCl than in physiological conditions.

To further assess the oligomeric state of SERBP1, DLS was used to measure the apparent hydrodynamic radius (R_h) at physiological conditions and in the presence of 2 M GdnHCl (Figure 4C). A mean R_h estimation of 3.9 and 4.4 nm under physiological and denaturing conditions, respectively, was obtained for SERBP1 189–400. Notably, the polydispersity in the presence of GdnHCl was ~30%, higher than the ~13% that was observed under non-denaturing conditions. The small relative difference in R_h in the presence and absence of GdnHCl indicates that SERBP1 is a monomer at physiological conditions, consistent with the SEC results, adopting a conformation that is more compact than a fully disordered chain. The theoretical R_h value, calculated by the method proposed by Marsh and Forman-Kay (Marsh and Forman-Kay, 2010) was 3.8 nm, in very good agreement with the experimentally measured value of 3.9 nm, further evidence that SERBP1 is monomeric at these conditions (Figure 4C). Finally, AUC experiments were conducted on SERBP1 189–400 in absence and presence of 2 M GdnHCl (Figure 4D). The fitted sedimentation coefficient was lower (1.3 vs. 1.9) and the frictional ratio was higher (2.2 vs. 1.8, not shown) in presence of GdnHCl, indicating SERBP1 189–400 is more compact under physiological conditions than a fully extended polypeptide chain, and that it

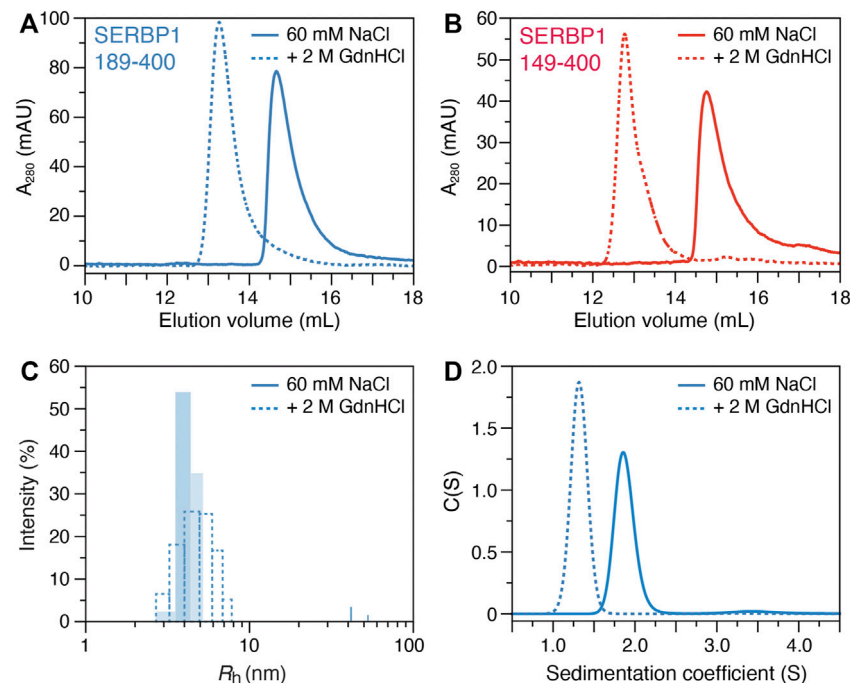


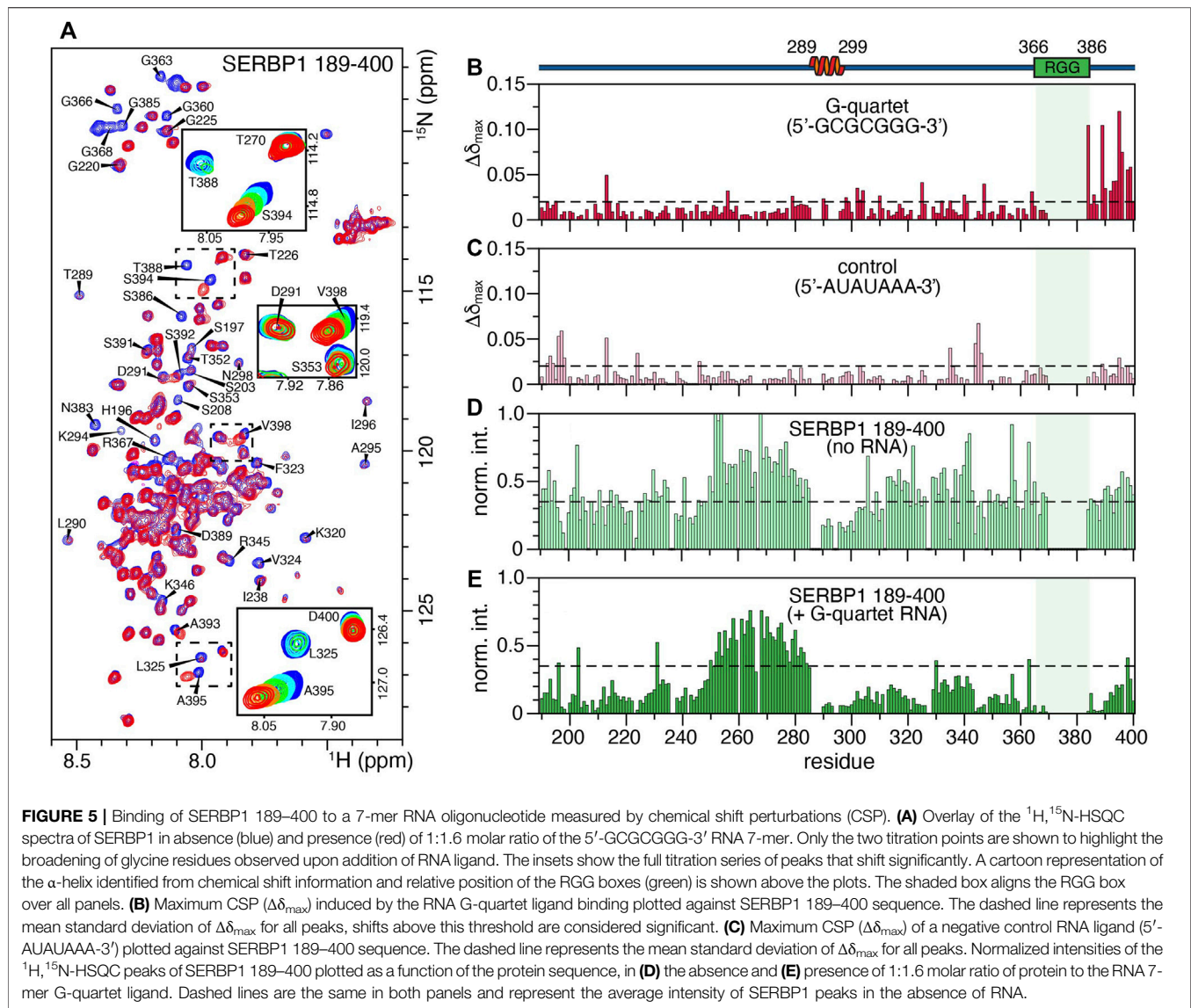
FIGURE 4 | Biophysical characterization of SERBP1 oligomer state in solution. Size exclusion chromatography profiles of **(A)** SERBP1 189–400 and **(B)** SERBP1 149–400, respectively, in the absence (solid) and presence (red) of 2 M GdnHCl recorded at A_{280} . **(C)** Size distribution histograms of the apparent hydrodynamic radius (R_h) derived from dynamic light scattering experiments of SERBP1 189–400 in the absence (solid) and presence (dashed) of 2 M GdnHCl. **(D)** Analytical ultracentrifugation sedimentation velocity coefficient profiles of SERBP1 189–400 in the absence (solid) and presence (dashed) of 2 M GdnHCl.

becomes less compact in denaturing conditions. Similar results were obtained for SERBP1 149–400 with a lower sedimentation coefficient (1.5 vs. 2.2) and a higher frictional ratio (2.1 vs. 1.7, not shown) in the presence vs. absence of 2 M GdnHCl respectively (**Supplementary Figure S6**). These observations are consistent with the average R_2 rates recorded for both constructs, ($\sim 9.5 \text{ s}^{-1}$) which are faster than expected for proteins that would preferentially exist in extended conformations (**Figure 3B**; **Supplementary Figure S5B**). Together, these data are highly consistent and reveal that SERBP1 is monomeric under the experimental conditions used here, adopting conformations that are more compact than a fully extended polymer.

SERBP1 Interacts With Guanine-Rich RNA

SERBP1 is thought to be an important regulator of many different mRNAs although the structural details of RNA recognition remain unclear. Previously, the RNAcompete assay defined 5'-GCGCGGG-3' as the SERBP1 consensus binding motif and subsequently it was shown that an RNA oligonucleotide containing this sequence binds full-length SERBP1 with sub-micromolar affinity (Kosti et al., 2020). Here, heteronuclear NMR was used to study the interaction between SERBP1 189–400 and the 7-mer RNA oligonucleotide. A series of ^1H , ^{15}N heteronuclear single quantum correlation (HSQC) spectra of SERBP1 189–400 were acquired in the presence of increasing concentrations of the 7-mer RNA (**Figure 5A**). Overall, the spectra remain similar in response to increasing concentrations of the 7-mer indicating the RNA

ligand does not induce folding of SERBP1, a phenomenon that has been observed for other IDPs (Sugase et al., 2007; Tompa and Fuxreiter, 2008; Shammas et al., 2016; Bonetti et al., 2018; Robustelli et al., 2020). However, some peaks display strong chemical shift perturbations (CSP), such as A395, S394, and V398, indicating these residues interact strongly with the RNA (**Figure 5A**, insets). Additionally, glycine peaks (e.g., G363, G366) belonging to the RGG box broaden significantly as the ligand concentration increases, likely due to exchange between the free and the bound forms at an intermediate rate on the NMR time scale. These observations lead to the proposal that electrostatic interactions occur between the positively charged arginine residues of the RGG box and the negatively charged phosphates of the RNA backbone. This encounter complex could then be stabilized by interactions with residues N383–V398, especially with T388 and S394 since serine and threonine are excellent hydrogen bond donors. Plotting the maximal CSP ($\Delta\delta_{\text{max}}$) against the protein sequence reveals that the RNA interacts with residues spanning from the second RGG box to the C-terminus of SERBP1, with some $\Delta\delta_{\text{max}}$ higher than 0.1 ppm (e.g., N383, T388, S394) (**Figure 5B**). If an initial SERBP1:RNA encounter complex forms through charge-charge interactions, the complementary RNA sequence (5'-AUAUAAA-3') may induce similar broadening in the RGG-specific glycine residues. No specific CSPs were observed in the C-terminal region of the SERBP1 189–400 (**Figure 5C**), and broadening of resonances belonging to the RGG box was not detected (**Supplementary Figure S7**). Some CSPs higher than



0.05 ppm are randomly dispersed along the chain, but these are more likely indicative of non-specific protein-RNA interactions. Taken together, these data reveal that further experimentation is required to fully elucidate the nature of the SERBP1 RNA recognition and complex formation.

The relative peak intensities of SERBP1 residues before and after titration were plotted as a function of the amino acid sequence to gain insight into the global effect of the RNA binding on the protein. (Figures 5D,E). Uniform broadening of the peaks along the entire protein sequence was observed upon the addition of stoichiometric quantities of the 7-mer RNA, with the exception of residues 250–285, that precede the α -helix. Notably, these residues retain approximately equal intensities in the absence or presence of the RNA (compare Figures 5D,E), suggesting they remain unaffected by interactions with RNA. This region has lower than average heteronuclear NOE (~ 0.25) and R_2 rates ($\sim 7 \text{ s}^{-1}$) indicating that these residues are more

dynamic compared to the rest of the protein and remain mobile in the presence of RNA (Figures 3B,C). Finally, no CSPs or peak broadening beyond the uniform broadening described above were observed for residues 289–299 that comprise the α -helix indicating that this region is not involved in recognition or binding of the RNA 7-mer.

SERBP1 Phase Separation is Inhibited by the Presence of Salt and Modulated by RNA

Intrinsically disordered RBPs are overrepresented in the group of proteins that are known to participate in biomolecular condensates and undergo LLPS (Burke et al., 2015; Elbaum-Garfinkle et al., 2015; Boeynaems et al., 2017). The phase separation properties of the SERBP1 149–400 and 189–400 constructs were assayed using droplet formation assays and fluorescence recovery after photobleaching (FRAP). At neutral

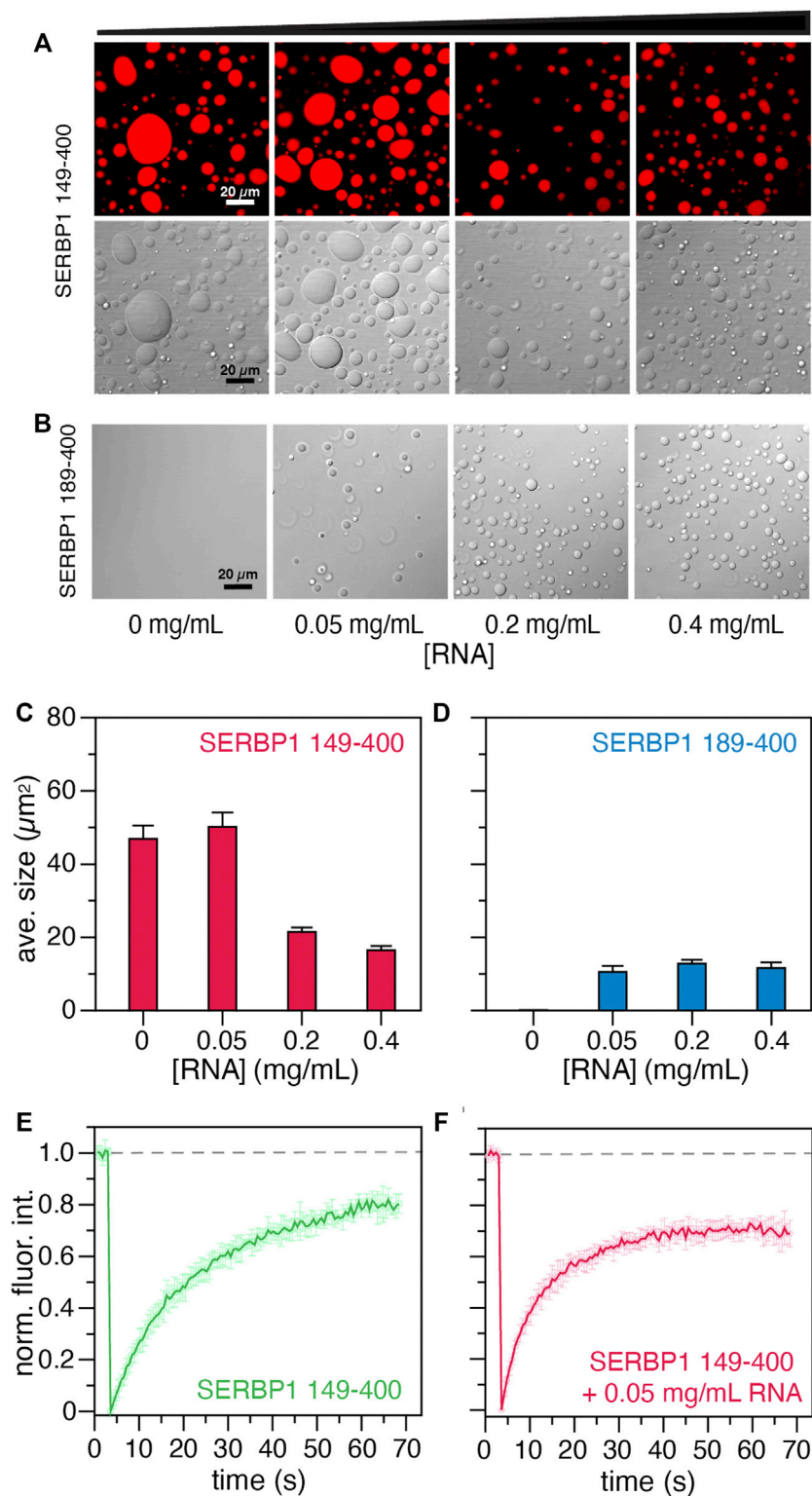


FIGURE 6 | Phase separation assays of SERBP1 149–400 and SERBP1 189–400. Effect of RNA on the LLPS propensity of 12 μM (A) SERBP1 149–400 or (B) SERBP1 189–400 in buffer containing 13 % (v/v) glycerol and 13 % (w/v) PEG3000. Torula yeast RNA was titrated into SERBP1 protein solutions to final concentrations ranging from 0 to 0.4 mg/ ml. The average droplet sizes (μm^2) shown for (C) SERBP1 149–400 and (D) SERBP1 189–400 were derived from three independent images. Normalized fluorescence recovery after photobleaching (FRAP) profiles of 12 μM SERBP1 149–400 in the (E) absence or (F) presence of 0.05 mg/ ml torula yeast RNA. Dashed line indicates average the fluorescence before photobleaching.

pH, SERBP1 149–400 readily phase separates at a concentration of 10 μM in absence of NaCl. The addition of NaCl abrogates SERBP1 phase separation with noticeably smaller and fewer droplets forming in the presence of 50 mM NaCl and complete dissipation at 150 mM NaCl (**Supplementary Figure S8A**). Conversely, SERBP1 189–400 does not phase separate at all under the same physiochemical conditions (**Supplementary Figure S8B**). Increasing NaCl concentrations (50–1000 mM) had no effect on the LLPS characteristics of SERBP1 189–400. Since the main difference between the two constructs is the N-terminal RGG box, it is reasonable to conclude that the RGG boxes promote phase separation of SERBP1, consistent with what has been described for other RGG-containing RBPs (Chong et al., 2018).

Given the observed effects of NaCl on phase separation of SERBP1, the effect of RNA was also tested. In this assay, 150 mM NaCl, 13 % glycerol, and 13% PEG3000 were included in the droplet formation buffer with 12 μM SERBP1 149–400 or SERBP1 189–400 and increasing concentrations of torula yeast RNA (**Figures 6A,B**). Contrary to the effect of NaCl, RNA promoted phase separation of SERBP1 189–400 at concentrations as low as 0.05 mg/ml RNA (**Figure 6B**). Under the three RNA concentrations tested (0.05, 0.2, and 0.4 mg/ml), the observed average droplet size for SERBP1 189–400 remained consistent at $\sim 12 \mu\text{m}^2$ (**Figure 6D**). At the same protein concentration, phase separation of SERBP1 149–400 was promoted by the presence of glycerol and PEG3000 (compare the 0 mg/ml panels in **Figures 6A,B**) while droplet size strongly correlated to the RNA concentration (**Figure 6A**). At no added RNA or 0.05 mg/ml RNA, the average observed droplet size for SERBP1 149–400 was $\sim 50 \mu\text{m}^2$. The average droplet size significantly decreased to $\sim 20 \mu\text{m}^2$ as the RNA concentration increased to 0.2 and 0.4 mg/ml (**Figure 6C**). This trend held when the experiment was repeated with 6 μM SERBP1 149–400 albeit the average observed droplet size was lower, ranging from 10 μm^2 without RNA to 2 μm^2 at an RNA concentration of 0.4 mg/ml (**Supplementary Figure S9**). RNA modulates SERBP1 phase separation, promoting phase separation for SERBP1 189–400 under non-permissive conditions, as well as altering the droplet characteristics formed from SERBP1 149–400. While the influence of crowding agents cannot be separated from the contribution of the RGG boxes to phase separation in these experiments, taken together these results show that the RGG boxes are important for SERBP1 LLPS, which likely proceeds through a charge-mediated mechanism.

The liquid-like nature of the condensates formed by SERBP1 149–400 in the presence and absence of RNA were probed using FRAP experiments (**Figure 6D**). In the absence of RNA, the recovery of fluorescence plateaus at ~ 70 s, while in the presence of 0.05 mg/ml of RNA, the fluorescence intensity recovers faster, reaching a plateau in ~ 40 s. The faster recovery of fluorescence in the presence of RNA is indicative of faster diffusion and hence more dynamic droplets. These results support the droplet formation assays suggesting that the two RGG boxes might interact synergistically with the RNA to modulate the condensate behavior of SERBP1. Further investigation will be

necessary to decipher the molecular mechanisms of SERBP1 LLPS.

DISCUSSION

SERBP1 was initially identified as a hyaluronic acid binding protein in biochemical pull-down studies using hyaluronic acid as bait (Heaton et al., 2001). The pattern of expression in normal tissue of the two known human hyaluronic acid binding proteins, HABP2 and HABP4 are different than SERBP1. HABP2 is predominantly found in the liver, while expression of HABP4 is the opposite of what was found for SERBP1 with high expression in all regions of the brain and very low expression in immortalized cells (Kosti et al., 2020). HABP4 binds mRNA and is known to interact with Receptor of activated protein C kinase (RACK1), a protein implicated in mRNA splicing and translation, although the significance of these associations is unknown (Huang et al., 2000; Nery et al., 2004). Similarly, SERBP1 was recently identified as a binding partner of RACK1 (Bolger, 2017), an interaction that was proposed to play a role in ribosomal composition and translation regulation. Thus, it is possible that although SERBP1 and HABP4 share significant sequence homology, the similarities are coincidental and represent an example of convergent or parallel evolution (Storz, 2016).

Sequence alignments with non-vertebrate SERBP1 homologues revealed five broadly defined regions of homology that roughly correlate with predicted SERBP1 ordered structure (compare dashed boxes in **Figures 1C,D**). In particular, chemical shift-derived secondary structure propensity, and CD measurements support predictions that SERBP1 is not entirely disordered but rather contains stable native structure in the form of an α -helix (residues 289–299), a structural feature common to many IDPs (Lee et al., 2012b), in a stretch of residues that are highly conserved among vertebrate and invertebrate sequences. The α -helix appears stable in both the 189–400 and 149–400 constructs and displays faster than expected transverse (R_2) relaxation rates on the order of 20–25 s^{-1} . These observations suggested the α -helix might be stabilized by intra- or intermolecular associations, or possibly mediate homodimerization. Such behavior has been described for TDP-43, a protein involved amyotrophic lateral sclerosis, whose disordered C-terminus domain contains an α -helix involved in protein-protein homodimerization and liquid-liquid phase separation (Conicella et al., 2016; Lim et al., 2016; Conicella et al., 2020). Therefore, intramolecular interactions of SERBP1 may contribute to the stability of the helix and promote preferential sampling of compact conformations, as observed for other IDPs (Marsh and Forman-Kay, 2010). Alternatively, the α -helix, or the other conserved domains might act as molecular recognition motifs, mediating associations with various biomolecular partners, through folding-upon-binding mechanisms (Sugase et al., 2007). Relaxation dispersion experiments did not reveal exchange with alternate SERBP1 conformers, yet these experiments do not rule out this possibility since the exchange timescale may be inaccessible for

CPMG-based experiments. Paramagnetic relaxation enhancements, filtered NOE, or off-resonance $R_{1\rho}$ experiments maybe more suitable to probe SERBP1 conformational exchange and are an active area of investigation.

SERBP1 is a multifunctional mRNA binding protein that has roles in regulating the expression of several mRNAs yet despite clear evidence of RNA binding (Heaton et al., 2001; Anger et al., 2013; Muto et al., 2018) it does not seem to possess any canonical RNA binding motifs, such as RRM, zinc-fingers or KH domains, which are predominant in other RNA-binding proteins (Maris et al., 2005; Valverde et al., 2008; Cook et al., 2015). The recognizable RNA-binding features of SERBP1 are two RG/RGG repeat regions comprising residues 165–184 and 366–386. CSPs and differential peak broadening suggest that the G-quartet RNA interaction spans the C-terminal RGG box and incorporates adjacent residues. Mutagenesis experiments will be useful to further define the role of the RGG box as well as uncover the source of the specificity of guanines over the other nucleotides (**Figures 5B,C**). Furthermore, these results are consistent with recent observations that some RBPs, particularly intrinsically disordered RBPs, can interact with RNA without the presence of specific folded RNA-binding domains (Castello et al., 2016; Järvelin et al., 2016; Hentze et al., 2018). As described here for SERBP1, RGG repeats of this class of RBP seem to play an important role in mediating interactions with RNA (Chong et al., 2018). For example, the RGG box of fragile X mental retardation protein (FMRP) was shown to interact with G-quadruplex forming RNA sequences (Darnell et al., 2001; Phan et al., 2011). Since SERBP1 was recently shown to bind G-quadruplexes (Su et al., 2021), it is important to elucidate the binding mode, the relative contribution of both RGG boxes, and any involvement of other conserved sequence motifs. A particularly important question to investigate is if SERBP1 actively stabilizes or destabilizes G-quadruplexes.

Proximity labeling approaches have identified SERBP1 interactions with myriad RBPs involved in mRNA regulation, stabilization, and splicing, translation, neurogenesis, synaptogenesis, and ribosome binding such as CAPRIN1, EIF4B, FXR1, LARP1, PABPC4, SYNCRIP, and YTHDF3 among others (Youn et al., 2018; Pedley et al., 2020). Similar to SERBP1, the subcellular localization of many of these proteins is controlled by arginine methylation, enabling them to shuttle between the nucleus and cytoplasm (Lee et al., 2012a). Indeed, many of these proteins localize to stress granules or other subcellular biomolecular condensates (Lee et al., 2014), and likewise, SERBP1 was shown here to undergo LLPS mediated by its RGG boxes. SERBP1 149–400 readily phase separates in low ionic strength buffers at neutral pH while the 189–400 construct does not, highlighting the importance of the presence of both RGG boxes to promote LLPS (**Figure 6**). The response to increasing ionic strength indicate that electrostatic interactions are an important mechanism mediating the transient self-associative contacts responsible for condensate formation as has been described for the BRD4-IDR (Sabari et al., 2018; Han et al., 2020) or Ddx4 (Brady et al., 2017). Indeed, LLPS of protein-RNA complexes

has recently been described as a way of regulating biological processes such as transcriptional or translational events (Han et al., 2012; Berry et al., 2015; Smith et al., 2016), and in some cases, RNA has been proposed to slow the formation of droplets of aggregation-prone prion-like proteins (Maharana et al., 2018). While RNA was shown to impact the LLPS properties of SERBP1, the effect of specific RNA structures like stem-loops or G-quadruplexes on SERBP1 condensate formation will require further investigation. Additionally, the α -helix in SERBP1 may contribute to electrostatic contacts important for LLPS since several residues comprising the α -helix are charged (E287, D291, E292, K294, D300).

Genomic studies indicate SERBP1 has a significant role in the one-carbon metabolism cycle and was implicated in synaptogenesis and neuronal development (Kosti et al., 2020). These pathways seem to be important in the development and progression of GBM. Indeed, SERBP1 is overexpressed in the brains of GBM patients and is negatively correlated with a favorable prognosis (Kunkle et al., 2013). Involvement of SERBP1 in these pathways may be through regulation of mRNA translation, transcription or splicing or through a combination of these functions. For example, SERBP1 was identified bound to the mRNA tunnel in the structures of inactive 80 S ribosomes reported by different groups (Anger et al., 2013; Brown et al., 2018; Muto et al., 2018), and was hypothesized to aid in translational control by negatively regulating ribosomal activity. The significance of these discoveries is not fully understood but may be related to a feedback cycle linking mRNA splicing to transcription mediated by RBPs like SERBP1. Recent work revealed a similar function for the SARS-CoV-2 protein Nsp1, which inserts its C-terminal domain into the ribosome mRNA channel, interrupting host transcriptional regulation (Schubert et al., 2020). Further, Zhang et al., found that SARS-CoV-2 also dysregulates One-carbon metabolism by increasing *de novo* purine synthesis and glycolysis (Zhang et al., 2021), a mechanism that was recently described for SERBP1 in GBM metabolism regulation (Kosti et al., 2020). These findings require further investigation to determine whether this is a specific strategy involving host mimicry of SERBP1 or simply a consequence of the general viral infection strategy of disrupting host translational regulation.

In summary, SERBP1 is an atypical RBP lacking known RNA recognition motifs and has multifunctional roles in translation and mRNA regulation and modulates One-carbon metabolism, neuronal differentiation and synaptogenesis. We identified the G-quartet RNA binding site on SERBP1 using NMR and present several lines of evidence such as conserved secondary structure and semi-compact conformations that suggest SERBP1 also recognizes higher-order RNA structures including stem-loops and G-quadruplexes. Current efforts are focused on identifying SERBP1 interactions with G-quadruplexes and assaying for folding or destabilizing activity. Additionally, the partially folded and compact nature of SERBP1 may be indicative of

pre-encounter conformations important for recognition and binding to other biomolecular targets. Future studies will examine the conformational dynamics of the α -helix and neighboring regions as potential binding sites for the myriad identified binding partners.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. The expression plasmid encoding SERBP1 149–400 was deposited with Addgene (176516) and the backbone resonance assignments are available from the Biological Magnetic Resonance Bank (51080).

AUTHOR CONTRIBUTIONS

AB made NMR samples, collected, processed, and interpreted data, wrote the manuscript, AM-R preformed the droplet formation assays, XX made expression constructs, and purified protein, ES collected, processed and analyzed the AUC experiments, LP performed phylogenetic analysis, wrote the manuscript, conceptualized the study and obtained funding, DL collected and processed data, wrote the manuscript, conceptualized the study and obtained funding, all authors contributed to editing manuscript. All authors have read and approved the manuscript for publication.

REFERENCES

- Ahn, J.-W., Kim, S., Na, W., Baek, S.-J., Kim, J.-H., Min, K., et al. (2015). SERBP1 Affects Homologous Recombination-Mediated DNA Repair by Regulation of CtIP Translation during S Phase. *Nucleic Acids Res.* 43 (13), 6321–6333. doi:10.1093/nar/gkv592
- Anger, A. M., Armache, J.-P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D. N., et al. (2013). Structures of the Human and Drosophila 80S Ribosome. *Nature* 497 (7447), 80–85. doi:10.1038/nature12104
- Antos, J. M., Ingram, J., Fang, T., Pishesha, N., Truttmann, M. C., and Ploegh, H. L. (2017). Site-Specific Protein Labeling via Sortase-Mediated Transpeptidation. *Curr. Protoc. Protein Sci.* 89, 1–15. doi:10.1002/cpps.38
- Baudin, A., Xu, X., and Libich, D. S. (2021). The ¹H, ¹⁵N and ¹³C Resonance Assignments of the C-Terminal Domain of Serpine mRNA Binding Protein 1 (SERBP1). *J. Biomol. NMR Assignments*. In Press. doi:10.1007/s12104-021-10046-3
- Berry, J., Weber, S. C., Vaidya, N., Haataja, M., and Brangwynne, C. P. (2015). RNA Transcription Modulates Phase Transition-Driven Nuclear Body Assembly. *Proc. Natl. Acad. Sci. USA*. 112 (38), E5237–E5245. doi:10.1073/pnas.1509317112
- Boeynaems, S., Bogaert, E., Kovacs, D., Konijnenberg, A., Timmerman, E., Volkov, A., et al. (2017). Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Mol. Cell*. 65 (6), 1044–1055. doi:10.1016/j.molcel.2017.02.013
- Bolger, G. B. (2017). The RNA-Binding Protein SERBP1 Interacts Selectively with the Signaling Protein RACK1. *Cell Signal*. 35, 256–263. doi:10.1016/j.cellsig.2017.03.001
- Bonetti, D., Troilo, F., Brunori, M., Longhi, S., and Gianni, S. (2018). How Robust Is the Mechanism of Folding-Upon-Binding for an Intrinsically Disordered Protein? *Biophysical J.* 114 (8), 1889–1894. doi:10.1016/j.bpj.2018.03.017

FUNDING

DL is a St. Baldrick's Scholar and acknowledges the support of the St. Baldrick's Foundation (634706). This study was funded in part by the Max and Minnie Tomerlin Voelcker Fund (Voelcker Foundation Young Investigator Grant to DL), the Welch Foundation (AQ-2001–20190330 to DL), the Cancer Prevention and Research Institute of Texas (RP200595 to LP), and a Mays Cancer Center Pilot Grant (to LP). EES was supported by a CPRIT Research Training Award (RP170345). This work is based upon research conducted in the Structural Biology Core Facilities, a part of the Institutional Research Cores at the University of Texas Health Science Center at San Antonio supported by the Office of the Vice President for Research and the Mays Cancer Center Drug Discovery and Structural Biology Shared Resource (NIH P30 CA05417X4).

ACKNOWLEDGMENTS

The authors thank Dr. Kristin Cano for NMR technical assistance and members of the DL and LP laboratories for valuable discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.744707/full#supplementary-material>

- Brady, J. P., Farber, P. J., Sekhar, A., Lin, Y.-H., Huang, R., Bah, A., et al. (2017). Structural and Hydrodynamic Properties of an Intrinsically Disordered Region of a Germ Cell-Specific Protein on Phase Separation. *Proc. Natl. Acad. Sci. USA*. 114 (39), E8194–E8203. doi:10.1073/pnas.1706197114
- Brown, A., Baird, M. R., Yip, M. C., Murray, J., and Shao, S. (2018). Structures of Translationally Inactive Mammalian Ribosomes. *eLife* 7, e40486. doi:10.7554/eLife.40486
- Buchan, D. W. A., and Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 Years on. *Nucleic Acids Res.* 47 (W1), W402–W407. doi:10.1093/nar/gkz297
- Burke, K. A., Janke, A. M., Rhine, C. L., and Fawzi, N. L. (2015). Residue-by-Residue View of *In Vitro* FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell*. 60 (2), 231–241. doi:10.1016/j.molcel.2015.09.006
- Camilloni, C., De Simone, A., Vranken, W. F., and Vendruscolo, M. (2012). Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry* 51 (11), 2224–2231. doi:10.1021/bi3001825
- Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A.-M., Foehr, S., et al. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell*. 63 (4), 696–710. doi:10.1016/j.molcel.2016.06.029
- Chemes, L. B., Alonso, L. G., Noval, M. G., and de Prat-Gay, G. (2012). Circular Dichroism Techniques for the Analysis of Intrinsically Disordered Proteins and Domains. *Methods Mol. Biol.* 895, 387–404. doi:10.1007/978-1-61779-927-3_22
- Chong, P. A., Vernon, R. M., and Forman-Kay, J. D. (2018). RGG/RG Motif Regions in RNA Binding and Phase Separation. *J. Mol. Biol.* 430 (23), 4650–4665. doi:10.1016/j.jmb.2018.06.014
- Conicella, A. E., Dignon, G. L., Zerze, G. H., Schmidt, H. B., D'Ordine, A. M., Kim, Y. C., et al. (2020). TDP-43 α -helical Structure Tunes Liquid-Liquid Phase Separation and Function. *Proc. Natl. Acad. Sci. USA*. 117 (11), 5883–5894. doi:10.1073/pnas.1912055117

- Conicella, A. E., Zerze, G. H., Mittal, J., and Fawzi, N. L. (2016). ALS Mutations Disrupt Phase Separation Mediated by α -Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* 24 (9), 1537–1549. doi:10.1016/j.str.2016.07.007
- Cook, K. B., Hughes, T. R., and Morris, Q. D. (2015). High-Throughput Characterization of Protein-RNA Interactions. *Brief. Funct. Genomics* 14 (1), 74–89. doi:10.1093/bfpg/elu047
- Costa, F. C., Saito, A., Gonçalves, K. A., Vidigal, P. M., Meirelles, G. V., Bressan, G. C., et al. (2014). Ki-1/57 and CGI-55 Ectopic Expression Impact Cellular Pathways Involved in Proliferation and Stress Response Regulation. *Biochim. Biophys. Acta (Bba) - Mol. Cel Res.* 1843 (12), 2944–2956. doi:10.1016/j.bbamer.2014.08.016
- Darnell, J. C., Jensen, K. B., Jin, P., Brown, V., Warren, S. T., and Darnell, R. B. (2001). Fragile X Mental Retardation Protein Targets G Quartet mRNAs Important for Neuronal Function. *Cell* 107 (4), 489–499. doi:10.1016/s0092-8674(01)00566-9
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* 6 (3), 277–293. doi:10.1007/BF00197809
- Delaglio, F., Walker, G. S., Farley, K. A., Sharma, R., Hoch, J. C., Arbogast, L. W., et al. (2017). Non-Uniform Sampling for All: More NMR Spectral Quality, Less Measurement Time. *Am. Pharm. Rev.* 20 (4), 339681.
- Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C. C.-H., Eckmann, C. R., Myong, S., et al. (2015). The Disordered P Granule Protein LAF-1 Drives Phase Separation into Droplets with Tunable Viscosity and Dynamics. *Proc. Natl. Acad. Sci. USA* 112 (23), 7189–7194. doi:10.1073/pnas.1504822112
- Fox, B. M., Janssen, A., Estevez-Ordóñez, D., Gessler, F., Vicario, N., Chagoya, G., et al. (2019). SUMOylation in Glioblastoma: A Novel Therapeutic Target. *Int. J. Mol. Sci.* 20 (8), 1853. doi:10.3390/ijms20081853
- Go, C. D., Knight, J. D. R., Rajasekharan, A., Rathod, B., Hesketh, G. G., Abe, K. T., et al. (2021). A Proximity-Dependent Biotinylation Map of a Human Cell. *Nature* 595 (7865), 120–124. doi:10.1038/s41586-021-03592-2
- Gracheva, E., Dus, M., and Elgin, S. C. R. (2009). Drosophila RISC Component VIG and its Homolog Vig2 Impact Heterochromatin Formation. *PLoS One* 4 (7), e6182. doi:10.1371/journal.pone.0006182
- Greenfield, N. J. (2006). Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat. Protoc.* 1 (6), 2876–2890. doi:10.1038/nprot.2006.202
- Han, T. W., Kato, M., Xie, S., Wu, L. C., Mirzaei, H., Pei, J., et al. (2012). Cell-Free Formation of RNA Granules: Bound RNAs Identify Features and Components of Cellular Assemblies. *Cell* 149 (4), 768–779. doi:10.1016/j.cell.2012.04.016
- Han, X., Yu, D., Gu, R., Jia, Y., Wang, Q., Jaganathan, A., et al. (2020). Roles of the BRD4 Short Isoform in Phase Separation and Active Gene Transcription. *Nat. Struct. Mol. Biol.* 27 (4), 333–341. doi:10.1038/s41594-020-0394-8
- Heaton, J. H., Dlakic, W. M., Dlakic, M., and Gelehrter, T. D. (2001). Identification and cDNA Cloning of a Novel RNA-Binding Protein that Interacts with the Cyclic Nucleotide-Responsive Sequence in the Type-1 Plasminogen Activator Inhibitor mRNA. *J. Biol. Chem.* 276 (5), 3341–3347. doi:10.1074/jbc.M006538200
- Hendriks, I. A., D'Souza, R. C. J., Yang, B., Verlaan-de Vries, M., Mann, M., and Vertegaal, A. C. O. (2014). Uncovering Global SUMOylation Signaling Networks in a Site-Specific Manner. *Nat. Struct. Mol. Biol.* 21 (10), 927–936. doi:10.1038/nsmb.2890
- Hentze, M. W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A Brave New World of RNA-Binding Proteins. *Nat. Rev. Mol. Cel Biol.* 19 (5), 327–341. doi:10.1038/nrm.2017.130
- Huang, L., Grammatikakis, N., Yoneda, M., Banerjee, S. D., and Toole, B. P. (2000). Molecular Characterization of a Novel Intracellular Hyaluronan-Binding Protein. *J. Biol. Chem.* 275 (38), 29829–29839. doi:10.1074/jbc.M002737200
- Järvelin, A. I., Noerberg, M., Davis, I., and Castello, A. (2016). The New (Dis) order in RNA Regulation. *Cell Commun Signal* 14, 9. doi:10.1186/s12964-016-0132-3
- Kjaergaard, M., and Poulsen, F. M. (2011). Sequence Correction of Random Coil Chemical Shifts: Correlation between Neighbor Correction Factors and Changes in the Ramachandran Distribution. *J. Biomol. NMR* 50 (2), 157–165. doi:10.1007/s10858-011-9508-2
- Koensgen, D., Mustea, A., Klamann, I., Sun, P., Zafarakas, M., Lichtenegger, W., et al. (2007). Expression Analysis and RNA Localization of PAI-RBP1 (SERBP1) in Epithelial Ovarian Cancer: Association with Tumor Progression. *Gynecol. Oncol.* 107 (2), 266–273. doi:10.1016/j.ygyno.2007.06.023
- Kosti, A., de Araujo, P. R., Li, W.-Q., Guardia, G. D. A., Chiou, J., Yi, C., et al. (2020). The RNA-Binding Protein SERBP1 Functions as a Novel Oncogenic Factor in Glioblastoma by Bridging Cancer Metabolism and Epigenetic Regulation. *Genome Biol.* 21, 195. doi:10.1186/s13059-020-02115-y
- Kunkle, B. W., Yoo, C., and Roy, D. (2013). Reverse Engineering of Modified Genes by Bayesian Network Analysis Defines Molecular Determinants Critical to the Development of Glioblastoma. *PLoS One* 8 (5), e64140. doi:10.1371/journal.pone.0064140
- Lee, S.-H., Kim, D.-H., J. Han, J., Cha, E.-J., Lim, J.-E., Cho, Y.-J., et al. (2012). Understanding Pre-Structured Motifs (PreSMos) in Intrinsically Unfolded Proteins. *Curr. Protein Pept. Sci.* 13 (1), 34–54. doi:10.2174/138920312799277974
- Lee, Y.-J., Hsieh, W.-Y., Chen, L.-Y., and Li, C. (2012). Protein Arginine Methylation of SERBP1 by Protein Arginine Methyltransferase 1 Affects Cytoplasmic/nuclear Distribution. *J. Cel. Biochem.* 113 (8), 2721–2728. doi:10.1002/jcb.24151
- Lee, Y.-J., Wei, H.-M., Chen, L.-Y., and Li, C. (2014). Localization of SERBP1 in Stress Granules and Nucleoli. *FEBS J.* 281 (1), 352–364. doi:10.1111/febs.12606
- Lemos, T. A., and Kobarg, J. (2006). CGI-55 Interacts with Nuclear Proteins and Co-Localizes to P80-Coilin Positive-Coiled Bodies in the Nucleus. *Cell Biochem Biophys.* 44 (3), 463–474. doi:10.1385/CBB:44:3:463
- Lim, L., Wei, Y., Lu, Y., and Song, J. (2016). ALS-Causing Mutations Significantly Perturb the Self-Assembly and Interaction with Nucleic Acid of the Intrinsically Disordered Prion-Like Domain of TDP-43. *Plos Biol.* 14 (1), e1002338. doi:10.1371/journal.pbio.1002338
- Maharana, S., Wang, J., Papadopoulos, D. K., Richter, D., Pozniakovsky, A., Poser, I., et al. (2018). RNA Buffers the Phase Separation Behavior of Prion-Like RNA Binding Proteins. *Science* 360 (6391), 918–921. doi:10.1126/science.aar7366
- Maris, C., Dominguez, C., and Allain, F. H.-T. (2005). The RNA Recognition Motif, a Plastic RNA-Binding Platform to Regulate Post-Transcriptional Gene Expression. *FEBS J.* 272 (9), 2118–2131. doi:10.1111/j.1742-4658.2005.04653.x
- Marsh, J. A., and Forman-Kay, J. D. (2010). Sequence Determinants of Compaction in Intrinsically Disordered Proteins. *Biophysical J.* 98 (10), 2383–2390. doi:10.1016/j.bpj.2010.02.006
- Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006). Sensitivity of Secondary Structure Propensities to Sequence Differences between α - and γ -Synuclein: Implications for Fibrillation. *Protein Sci.* 15 (12), 2795–2804. doi:10.1110/ps.062465306
- Massi, F., Johnson, E., Wang, C., Rance, M., and Palmer, A. G., 3rd (2004). NMR R1 ρ Rotating-Frame Relaxation with Weak Radio Frequency Fields. *J. Am. Chem. Soc.* 126 (7), 2247–2256. doi:10.1021/ja038721w
- Muto, A., Sugihara, Y., Shibakawa, M., Oshima, K., Matsuda, T., and Nadano, D. (2018). The mRNA-Binding Protein Serbp1 as an Auxiliary Protein Associated with Mammalian Cytoplasmic Ribosomes. *Cell Biochem Funct* 36 (6), 312–322. doi:10.1002/cbf.3350
- Nery, F. C., Passos, D. O., Garcia, V. S., and Kobarg, J. (2004). Ki-1/57 Interacts with RACK1 and Is a Substrate for the Phosphorylation by Phorbol 12-myristate 13-Acetate-Activated Protein Kinase C. *J. Biol. Chem.* 279 (12), 11444–11455. doi:10.1074/jbc.M306672200
- Papadopoulos, J. S., and Agarwala, R. (2007). COBALT: Constraint-Based Alignment Tool for Multiple Protein Sequences. *Bioinformatics* 23 (9), 1073–1079. doi:10.1093/bioinformatics/btm076
- Pedley, R., King, L. E., Mallikarjun, V., Wang, P., Swift, J., Brennan, K., et al. (2020). BioID-Based Proteomic Analysis of the Bid Interactome Identifies Novel Proteins Involved in Cell-Cycle-Dependent Apoptotic Priming. *Cell Death Dis.* 11 (10), 872. doi:10.1038/s41419-020-03091-8
- Phan, A. T., Kuryavyi, V., Darnell, J. C., Serganov, A., Majumdar, A., Ilin, S., et al. (2011). Structure-Function Studies of FMRP RGG Peptide Recognition of an RNA Duplex-Quadruplex Junction. *Nat. Struct. Mol. Biol.* 18 (7), 796–804. doi:10.1038/nsmb.2064
- Robustelli, P., Piana, S., and Shaw, D. E. (2020). Mechanism of Coupled Folding-Upon-Binding of an Intrinsically Disordered Protein. *J. Am. Chem. Soc.* 142 (25), 11092–11101. doi:10.1021/jacs.0c03217
- Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., et al. (2018). Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control. *Science* 361 (6400), eaar3958. doi:10.1126/science.aar3958

- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: An Open-Source Platform for Biological-Image Analysis. *Nat. Methods* 9 (7), 676–682. doi:10.1038/nmeth.2019
- Schubert, K., Karousis, E. D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L.-A., et al. (2020). SARS-CoV-2 Nsp1 Binds the Ribosomal mRNA Channel to Inhibit Translation. *Nat. Struct. Mol. Biol.* 27 (10), 959–966. doi:10.1038/s41594-020-0511-8
- Schuck, P. (2000). Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophysical J.* 78 (3), 1606–1619. doi:10.1016/S0006-3495(00)76713-0
- Serce, N. B., Boesl, A., Klamann, I., von Serényi, S., Noetzel, E., Press, M. F., et al. (2012). Overexpression of SERBP1 (Plasminogen Activator Inhibitor 1 RNA Binding Protein) in Human Breast Cancer Is Correlated with Favourable Prognosis. *BMC cancer* 12, 597. doi:10.1186/1471-2407-12-597
- Shammas, S. L., Crabtree, M. D., Dahal, L., Wicky, B. I. M., and Clarke, J. (2016). Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies. *J. Biol. Chem.* 291 (13), 6689–6695. doi:10.1074/jbc.R115.692715
- Shen, Y., and Bax, A. (2013). Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *J. Biomol. NMR* 56 (3), 227–241. doi:10.1007/s10858-013-9741-y
- Skinner, S. P., Fogh, R. H., Boucher, W., Ragan, T. J., Mureddu, L. G., and Vuister, G. W. (2016). CcpNmr AnalysisAssign: A Flexible Platform for Integrated NMR Analysis. *J. Biomol. NMR* 66 (2), 111–124. doi:10.1007/s10858-016-0060-y
- Smith, J., Calidas, D., Schmidt, H., Lu, T., Rasoloson, D., and Seydoux, G. (2016). Spatial Patterning of P Granules by RNA-Induced Phase Separation of the Intrinsically-Disordered Protein MEG-3. *Elife* 5, e21337. doi:10.7554/eLife.21337
- Storz, J. F. (2016). Causes of Molecular Convergence and Parallelism in Protein Evolution. *Nat. Rev. Genet.* 17 (4), 239–250. doi:10.1038/nrg.2016.11
- Su, H., Xu, J., Chen, Y., Wang, Q., Lu, Z., Chen, Y., et al. (2021). Photoactive G-Quadruplex Ligand Identifies Multiple G-Quadruplex-Related Proteins with Extensive Sequence Tolerance in the Cellular Environment. *J. Am. Chem. Soc.* 143 (4), 1917–1923. doi:10.1021/jacs.0c10792
- Sugase, K., Dyson, H. J., and Wright, P. E. (2007). Mechanism of Coupled Folding and Binding of an Intrinsically Disordered Protein. *Nature* 447 (7147), 1021–1025. doi:10.1038/nature05858
- Theile, C. S., Witte, M. D., Blom, A. E. M., Kundrat, L., Ploegh, H. L., and Guimaraes, C. P. (2013). Site-Specific N-Terminal Labeling of Proteins Using Sortase-Mediated Reactions. *Nat. Protoc.* 8 (9), 1800–1807. doi:10.1038/nprot.2013.102
- Tomba, P., and Fuxreiter, M. (2008). Fuzzy Complexes: Polymorphism and Structural Disorder in Protein-Protein Interactions. *Trends Biochem. Sci.* 33 (1), 2–8. doi:10.1016/j.tibs.2007.10.003
- Tsui, C., Inouye, C., Levy, M., Lu, A., Florens, L., Washburn, M. P., et al. (2018). dCas9-Targeted Locus-Specific Protein Isolation Method Identifies Histone Gene Regulators. *Proc. Natl. Acad. Sci. USA.* 115 (12), E2734–E2741. doi:10.1073/pnas.1718844115
- Uversky, V. N. (2012). Size-Exclusion Chromatography in Structural Analysis of Intrinsically Disordered Proteins. *Methods Mol. Biol.* 896, 179–194. doi:10.1007/978-1-4614-3704-8_11
- Valverde, R., Edwards, L., and Regan, L. (2008). Structure and Function of KH Domains. *FEBS J.* 275 (11), 2712–2726. doi:10.1111/j.1742-4658.2008.06411.x
- Wang, T., Xu, L., Jia, R., and Wei, J. (2017). MiR-218 Suppresses the Metastasis and EMT of HCC Cells via Targeting SERBP1. *Acta Biochim. Biophys. Sinica* 49 (5), 383–391. doi:10.1093/abbs/gmx017
- Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004). Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* 337 (3), 635–645. doi:10.1016/j.jmb.2004.02.002
- Williamson, M. P. (2013). Using Chemical Shift Perturbation to Characterise Ligand Binding. *Prog. Nucl. Magn. Reson. Spectrosc.* 73, 1–16. doi:10.1016/j.pnmrs.2013.02.001
- Ying, J., Delaglio, F., Torchia, D. A., and Bax, A. (2017). Sparse Multidimensional Iterative Lineshape-Enhanced (SMILE) Reconstruction of Both Non-Uniformly Sampled and Conventional NMR Data. *J. Biomol. NMR* 68 (2), 101–118. doi:10.1007/s10858-016-0072-7
- Youn, J.-Y., Dunham, W. H., Hong, S. J., Knight, J. D. R., Bashkurov, M., Chen, G. I., et al. (2018). High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Mol. Cell* 69 (3), 517–532. e11. doi:10.1016/j.molcel.2017.12.020
- Zhang, Y., Guo, R., Kim, S. H., Shah, H., Zhang, S., Liang, J. H., et al. (2021). SARS-CoV-2 Hijacks Folate and One-Carbon Metabolism for Viral Replication. *Nat. Commun.* 12 (1), 1676. doi:10.1038/s41467-021-21903-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Baudin, Moreno-Romero, Xu, Selig, Penalva and Libich. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Conserved Structural Role for the Walker-A Lysine in P-Loop Containing Kinases

Fatum Hajredini^{1,2} and Ranajeet Ghose^{1,2,3,4*}

¹Department of Chemistry and Biochemistry, The City College of New York, New York, NY, United States, ²PhD Program in Biochemistry, The Graduate Center of CUNY, New York, NY, United States, ³PhD Program in Chemistry, The Graduate Center of CUNY, New York, NY, United States, ⁴PhD Program in Physics, The Graduate Center of CUNY, New York, NY, United States

OPEN ACCESS

Edited by:

George Lisi,
Brown University, United States

Reviewed by:

Xiakun Chu,
Stony Brook University, United States
Julien Roche,
Iowa State University, United States

*Correspondence:

Ranajeet Ghose
rgghose@ccny.cuny.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 25 July 2021

Accepted: 20 September 2021

Published: 01 October 2021

Citation:

Hajredini F and Ghose R (2021) A
Conserved Structural Role for the
Walker-A Lysine in P-Loop
Containing Kinases.
Front. Mol. Biosci. 8:747206.
doi: 10.3389/fmolb.2021.747206

Bacterial tyrosine kinases (BY-kinases) and shikimate kinases (SKs) comprise two structurally divergent P-loop containing enzyme families that share similar catalytic site geometries, most notably with respect to their Walker-A, Walker-B, and DxD motifs. We had previously demonstrated that in BY-kinases, a specific interaction between the Walker-A and Walker-B motifs, driven by the conserved “catalytic” lysine housed on the former, leads to a conformation that is unable to efficiently coordinate Mg^{2+} •ATP and is therefore incapable of chemistry. Here, using enhanced sampling molecular dynamics simulations, we demonstrate that structurally similar interactions between the Walker-A and Walker-B motifs, also mediated by the catalytic lysine, stabilize a state in SKs that deviates significantly from one that is necessary for the optimal coordination of Mg^{2+} •ATP. This structural role of the Walker-A lysine is a general feature in SKs and is found to be present in members that encode a Walker-B sequence characteristic of the family (*Coxiella burnetii* SK), and in those that do not (*Mycobacterium tuberculosis* SK). Thus, the structural role of the Walker-A lysine in stabilizing an inactive state, distinct from its catalytic function, is conserved between two distantly related P-loop containing kinase families, the SKs and the BY-kinases. The universal conservation of this element, and of the key characteristics of its associated interaction partners within the Walker motifs of P-loop containing enzymes, suggests that this structural role of the Walker-A lysine is perhaps a widely deployed regulatory mechanism within this ancient family.

Keywords: P-loop, protein kinases, shikimate kinase, molecular dynamics (MD), bacterial tyrosine kinase

INTRODUCTION

P-loop containing proteins are widespread in all three branches of the tree of life suggesting that they emerged prior to the last universal common ancestor (LUCA) (Aravind et al., 2002; Edwards et al., 2013; Longo et al., 2020). Nucleotide triphosphatases (NTPases) (Leipe et al., 2002) and P-loop kinases constitute two of several divergent families of P-loop containing enzymes that encode sequences known as Walker-A (GxxxGK[S/T], x is any residue) and Walker-B ($\phi\phi\phi\phi$ [D/E]xxG, ϕ is a hydrophobic residue) motifs, or variations thereon (Walker et al., 1982). The BY-kinases (for Bacterial tYrosine kinases), that are widely conserved within the prokaryotic kingdom, but are without archaeal or eukaryotic counterparts (Shi et al., 2014), constitute a unique family of protein tyrosine kinases within the SIMIBI (named for the signal-recognition GTPases, MinD and BioD superfamilies) class of NTPases (Leipe et al., 2002). Instead of the

dual-lobed fold characteristic of the eukaryotic protein kinases (Taylor and Radzio-Andzelm, 1994; Scheeff and Bourne, 2005), the catalytic domains (CDs) of BY-kinases (Lee and Jia, 2009; Grangeasse et al., 2012) utilize a fold that closely resembles ATPases of the MinD family. While BY-kinase CDs may be classified as P-loop kinases based on their function, they deviate significantly from the P-loop kinase superfamily (comprising kinases that phosphorylate small molecules e.g., nucleosides, shikimate, 6-phosphofructose etc.) as defined by Leipe *et al.* (Leipe et al., 2003) (we will adhere to that definition here), notably through the absence of a “LID” domain present in the latter (**Supplementary Figure S1A**). In addition to modified Walker-A and Walker-B motifs, BY-kinase CDs also encode a sequence, $\phi\phi\phi\phi\text{DxDxR}$, that has been termed the Walker-A' motif (Grangeasse et al., 2012). The DxD sequence present in the Walker-A' motif is also found in many members of the SIMBI class of P-loop enzymes including MinD (**Supplementary Figure S1B**). Interestingly, the DxD motif, is characteristic of many P-loop kinases including those that utilize shikimate, gluconate or adenosine 5'-phosphosulfate (APS) as substrates (Leipe et al., 2003). A key feature of these “DxD-group” P-loop kinases is a Walker-B motif that is degraded at its C-terminal end and marked by the conspicuous absence of the acidic Asp (or Glu) residue (**Supplementary Figure S1C**) (Leipe et al., 2003). While the BY-kinase CD and P-loop kinases deviate significantly in their three-dimensional structures, they display similar catalytic site geometries formed by closely related structural motifs (**Supplementary Figure S1C**). Given the close similarities in the catalytic site architecture despite being embedded in divergent structural scaffolds, and the fact that both families perform similar phospho-transfer chemistry, we wondered whether these two distantly related families retain common structural/dynamic features that characterize their functional states.

We have previously demonstrated the unique conformational dynamics of the catalytic core of the archetypal BY-kinase, *Escherichia coli* (K12) Wzc (Wzc_{CDAC}); these dynamics appear to have significant functional consequences. Our studies revealed the presence of two major conformations with distinct long- and short-range structural features (Hajredini et al., 2020) in Wzc_{CDAC}. We designated these conformations as open (open state, OS) and closed (closed state, CS) based on their global compactness. The OS has high affinity for ADP but is unable to efficiently engage ATP•Mg²⁺. The CS, that can optimally engage ATP•Mg²⁺, is stabilized upon formation of a closed octameric ring seen both in the structure of the isolated CD (Bechet et al., 2010) as well as in the membrane-anchored full-length protein (Yang et al., 2021). Oligomer formation is necessary to enable intermolecular autophosphorylation. Thus, the differential nucleotide preferences of the OS and CS together with the coupling of their relative stabilities to oligomerization appear to play a critical role in maintaining optimal Wzc function through modulation of the substrate affinity and the prevention of futile ATP hydrolysis (Hajredini et al., 2020). A key structural feature that was found to distinguish the OS from the CS involved an alteration in the mode of interaction between

conserved Walker-A and Walker-B residues (Hajredini et al., 2020; Hajredini and Ghose, 2021). In the OS, the conserved Walker-A Lys (the so-called “catalytic lysine”, K540) establishes a salt bridge with the conserved Walker-B Asp (D642) generating a conformation that cannot effectively coordinate Mg²⁺ (and consequently, ATP•Mg²⁺). In the CS, the Walker-A Lys disengages from the Walker-B Asp and the latter then contacts the conserved Walker-A Thr (T541) generating a conformation required to optimally coordinate Mg²⁺. To assess whether similar conformational states stabilized by interactions between key conserved residues also populate the landscape of the P-loop kinases, we performed enhanced sampling molecular dynamics simulations (REST2) (Wang et al., 2011) using *Mycobacterium tuberculosis* SK (SK_{Mtu} from hereon) in its unliganded, and its ATP•Mg²⁺-bound states. Our results suggest that SK_{Mtu} samples distinct global conformations that, as in the case of the Wzc_{CDAC}, are distinguished by local interactions of the Walker-A and Walker-B motifs, most notably those involving the catalytic Lys (K15). This structural feature of the Walker-A Lys in SKs is also preserved despite sequence variations in the Walker-B motif (as in *Coxiella burnetii* SK, SK_{Cbu}) suggesting that this particular role of this residue in parsing functional conformations may be a general feature in P-loop containing enzymes.

MATERIALS AND METHODS

Preparation of Starting Structures

The starting conformation of the SK_{Mtu}•ATP•Mg²⁺ complex was generated using the coordinates of SK_{Mtu} from the structure of the SK_{Mtu}•ADP•Mg²⁺•shikimate complex (PDB: 1WE2) (Pereira et al., 2004) after removal of the bound ADP and shikimate. The Mg²⁺ ion and all solvent molecules were retained. The coordinates of ATP were derived from the structure of the SK_{Mtu}•AMPPCP•shikimate complex (PDB: 1ZYU) (Gan et al., 2006) by aligning the 1WE2 and 1ZYU structures using the Ca atoms of their respective protein components and converting the bound AMPPCP to ATP through the replacement of the relevant carbon atom by oxygen. Subsequently, the protein and shikimate coordinates of the 1ZYU structure were removed to generate the SK_{Mtu}•ATP•Mg²⁺ complex. For consistency, the corresponding apo simulations utilized the same starting structure after removal of ATP•Mg²⁺. For unliganded SK_{Cbu}, the starting structure was obtained directly from the coordinates found in PDB: 3TRF (Franklin et al., 2015) after removal of all ligands except solvent.

Molecular Dynamics Simulations

The starting structures, generated as discussed above, were equilibrated using previously described protocols (Hajredini et al., 2020). Briefly, the systems were parameterized within the CHARMM36m force-field (Huang et al., 2017), solvated with TIP3P water molecules, energy minimized and then equilibrated, first in an NVT ensemble, and then in an NPT

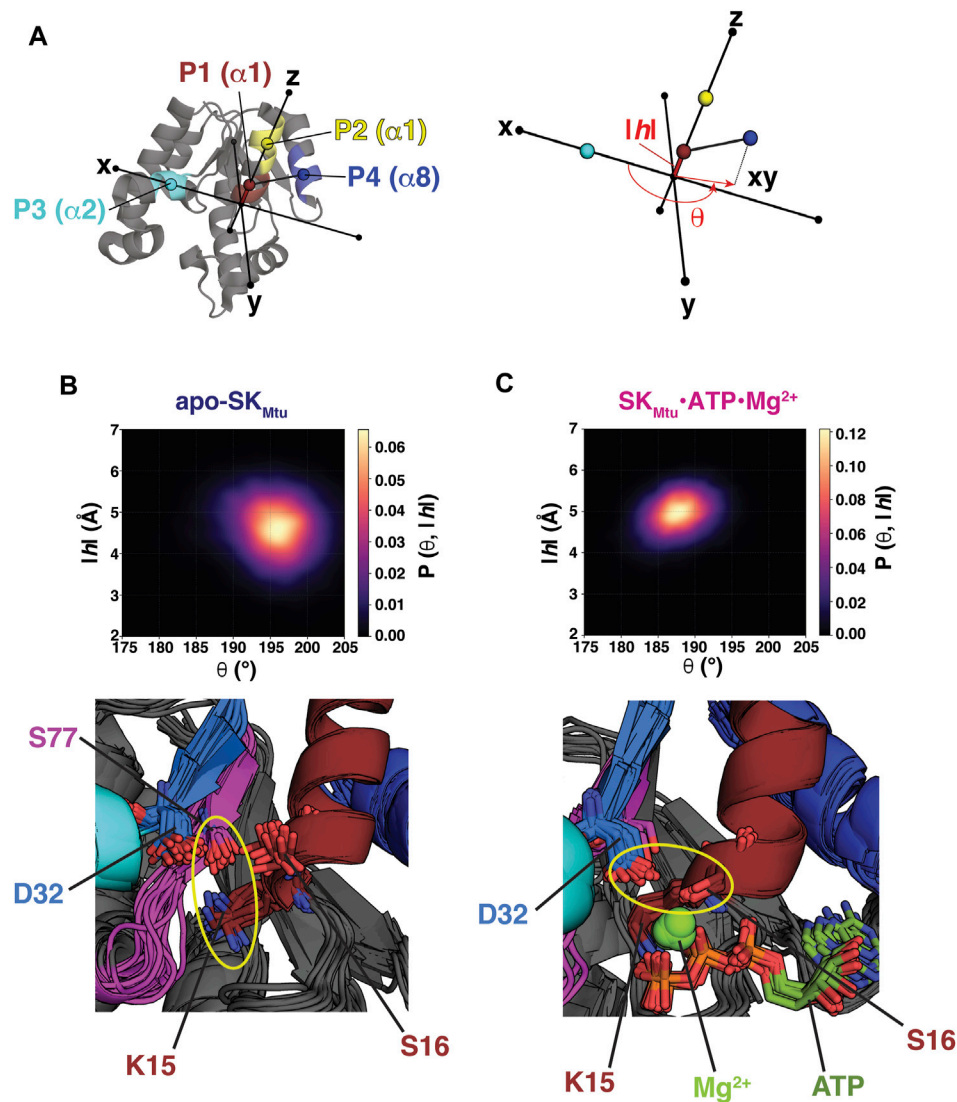


FIGURE 1 | (A) Points P1 (brown), P2 (yellow), P3 (cyan) and P4 (dark blue) used to define the cylindrical coordinate frame for analyses of the REST2-generated ensembles of SK_{Mtu} are indicated on its structure. The definitions of the angle θ and the rise $|h|$ are shown on the right panel. The REST2-generated structural ensembles of unliganded SK_{Mtu} (B) and the SK_{Mtu}•ATP•Mg²⁺ complex (C) are shown projected onto θ - $|h|$ space and plotted using kernel density estimation. The bottom panels show key interactions that stabilize each state using 20 representative structures drawn from the corresponding regions of highest probability density (194–200° and 4–5 Å in the $|h|$ and θ dimensions, respectively for apo-SK_{Mtu} and 185–190° and 4.5–5.5 Å for the corresponding dimensions of the SK_{Mtu}•ATP•Mg²⁺ complex). A key feature that distinguishes the Walker Open State (WOS) from the Walker Closed State (WCS) is the orientation of the Walker-A Lys (K15 in SK_{Mtu}). For the WOS seen for apo-SK_{Mtu}, K15 hydrogen bonds with S77 of the Walker-B motif in the bulk of the structures (K15 and S77 are proximal in all of them). This interaction is broken in the SK_{Mtu}•ATP•Mg²⁺ complex. The yellow ellipses highlight the altered interactions of D32 in the two states.

ensemble to generate starting structures for the enhanced sampling, REST2, (Wang et al., 2011) simulations. Subsequently, REST2 production runs were carried out for 200 ns using 14 replicas spanning a 300–400 K temperature range. The final exchange probabilities for the SK_{Mtu}•ATP•Mg²⁺, SK_{Mtu}, and SK_{Cbu} simulations were 23.3 ± 8.2 , 24.2 ± 8.4 , and $29.0 \pm 0.6\%$, respectively. The first 40 ns in each of the simulations were used as “burn-in” periods and discarded, and all subsequent analyses were carried out utilizing the remaining 160 ns. As shown in **Supplementary Figure S2**, 40 ns was more than sufficient

to ensure equilibration of the key observables (θ and $|h|$, see below) probed in the simulations.

Definition of the Cylindrical Coordinate Frame

To describe the global conformational landscape of SK_{Mtu}, a cylindrical coordinate frame, comprising of an angle θ and a rise $|h|$ (Figure 1A), was defined as in the case of WzC_{CDAC}, described previously (**Supplementary Figure S3A**) (Hajredini et al., 2020). The centers of mass of the Ca atoms of two

segments on helix $\alpha 1$ (that contains the Walker-A motif), comprising residues 15–19 (P1, brown) and 21–25 (P2, yellow), were used to define a vector that was aligned to the z -axis of the molecular frame. Each structure within the REST2-generated ensembles was rotated and translated such that point P3 (cyan), defined by the center of mass of the Ca atoms of the first turn of helix $\alpha 2$ (33–37), was placed on the x -axis along the positive direction. θ is then defined as the angle between the xy -projection of a vector extending from point P1 to P4 (blue; defined by the center of mass of the Ca atoms of residues 155–159 that form the first turn of helix $\alpha 8$). The rise $|h|$ is defined as the absolute value of the distance between points P1 and P3 along the z -axis. Thus, the angle θ and the rise $|h|$ provide measures of the outward rotation and the up/down movement, respectively, of the Walker-A motif relative to the Walker-B and DxD motifs. The crystal structures of SK_{Mtu} (listed in **Supplementary Table S1**) were projected onto the cylindrical coordinate frame using the same protocol. For the structures of the SK enzymes from other organisms (listed in **Supplementary Table S1**), a similar procedure was utilized to define the points P1–P4 (described in **Supplementary Table S2**) and subsequently, θ and $|h|$.

RESULTS AND DISCUSSION

Unliganded and Liganded Conformational Ensembles of SK_{Mtu} Populate Distinct Structural States

We performed two sets of enhanced sampling simulations utilizing the replica exchange with solute scaling (REST2) approach. REST2 represents a modification of the Hamiltonian scaling protocol of the replica exchange with solute tempering (REST) scheme (Liu et al., 2005) that greatly enhances its efficiency by decoupling the acceptance rate from the solvent. For the REST2 simulations, protocols similar to those described before (Hajredini et al., 2020) were utilized with either unliganded (apo) or ATP•Mg²⁺-bound SK_{Mtu}. To analyze the global conformational states populated by these two ensembles we utilized the cylindrical coordinate frame defined in **Figure 1A**. Note that this frame is similar, but not identical, to that previously defined for WzC_{CDAC} (Hajredini et al., 2020) (compare **Supplementary Figure S3A**) given the significant structural differences between the two proteins. The orientations of the $\alpha 8$ helix (carrying P4) and $\alpha 1$ (carrying P1 and P2) in SK_{Mtu} are similar to the $\alpha 2$ - $\alpha 3$ pair in WzC_{CDAC} (the latter carries the Walker-A sequence) (Hajredini et al., 2020). *In lieu* of a missing WzC_{CDAC} $\alpha 4$ equivalent in SK_{Mtu}, the $\alpha 2$ helix (that hosts the second Asp, D34, of the DxD motif) was utilized to define P3. Thus, while this definition allows for a qualitative comparison of the global conformational states sampled by SK_{Mtu} and WzC_{CDAC} with respect to their key catalytic elements, a quantitative equivalence should not be expected.

Projection of the conformational ensembles of apo-SK_{Mtu} (**Figure 1B**, top panel) and the SK_{Mtu}•ATP•Mg²⁺ complex

(**Figure 1C**, top panel) onto the 2-dimensional θ - $|h|$ space suggests differences in global conformations between the two cases. The presence of ATP•Mg²⁺ induces a more closed conformation, characterized by smaller θ values, compared to a more open conformation (larger θ values) seen for apo-SK_{Mtu}. We designate these global states as a Walker Closed State (WCS, seen in the SK_{Mtu}•ATP•Mg²⁺ complex), and a Walker Open State (WOS, seen in apo-SK_{Mtu}) to distinguish them from the closed and open classifications of the relative conformation of the SB and the LID domains traditionally used in the SK literature (Hartmann et al., 2006). Contrasts between the θ - $|h|$ projections of SK_{Mtu} and WzC_{CDAC} are evident upon comparing **Figures 1B,C** with **Supplementary Figures S3B,C**. In the case of WzC_{CDAC}, the OS is characterized by high values of both θ and $|h|$, while the CS exhibits low values for both variables (**Supplementary Figures S3B,C**), suggesting that these variables are positively correlated (Hajredini et al., 2020). In contrast, the regions of maximal density for the WOS (**Figure 1B**, top panel) and the WCS (**Figure 1C**, top panel) suggest a weak anti-correlation between the θ (WOS: $196 \pm 4^\circ$, WCS: $188 \pm 3^\circ$) and $|h|$ dimensions (WOS: $4.6 \pm 0.7 \text{ \AA}$, WCS: $4.9 \pm 0.4 \text{ \AA}$), with lower θ and higher $|h|$ for the WCS, and the opposite for the WOS. Further, the apparent separation between the WOS and the WCS seems less prominent than that between the OS and CS of WzC_{CDAC} (compare **Figures 1B,C** with **Supplementary Figures S3B,C**; also see **Supplementary Figure S4A**). As an independent validation of our simulations, we note that the regions in θ - $|h|$ space spanned by the WOS and the WCS in our REST2 ensembles are well represented (**Supplementary Figure S4A**) in the several available crystal structures of SKs (and their complexes) from a variety of organisms (listed in **Supplementary Table S1**).

Representative structures of apo- (**Figure 1B**, bottom panel) and ATP•Mg²⁺-bound (**Figure 1C**, bottom panel) SK_{Mtu} extracted from the regions of maximal density in θ - $|h|$ space of the corresponding ensembles show altered conformations of key catalytic site elements. In the presence of ATP•Mg²⁺ i.e., in the WCS, the Walker-A Lys (K15) assumes a downwards orientation and contacts the β - and γ -phosphates of ATP; the first Asp (D32) of the DxD motif forms a hydrogen bond with the Walker-A S16, thus functionally mimicking the absent Walker-B Asp (Leipe et al., 2003). The D32-S16 interaction is necessary for the optimal coordination of Mg²⁺ (**Figure 1C**, bottom panel). Indeed, as shown in **Supplementary Figure S4B**, the D32-C γ -S16-O γ distance is well defined and shows a narrow distribution ($3.7 \pm 0.2 \text{ \AA}$) within the ensemble. In the absence of ligands, i.e., in the WOS, K15 rotates outwards to establish a hydrogen-bond with S77 of the Walker-B motif (**Figure 1B**, bottom panel) in a significant fraction of the structures. In this configuration, the D32-S16 distance shows a broad distribution ($6.0 \pm 1.2 \text{ \AA}$), suggestive of increased disorder, with a maximum reflective of significantly increased separation (**Supplementary Figure S4B**) suggesting a conformation that is not conducive for Mg²⁺ coordination. Conformations displaying the K15-S77 hydrogen bond have been observed in several crystal structures of SK_{Mtu}, that are, not surprisingly, free of Mg²⁺. The structures (representative examples shown in **Supplementary Figure S4C**) (Hartmann et al., 2006) that

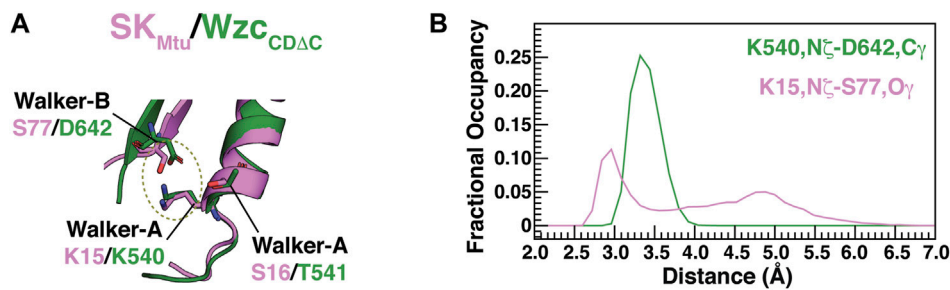


FIGURE 2 | (A) Interactions that are characteristic of the open states of BY-kinases (OS) and SKs (WOS). S77 from the degraded Walker-B motif from SK_{Mtu} (pink) occupies approximately the same spatial position as the Walker-B D642 from Wzc_{CDAC} (green) allowing the formation of a S77-K15 hydrogen bond reminiscent of the D642-K540 interaction in Wzc_{CDAC}. **(B)** Distribution of the K540, Nε-D642, Cγ (green) and K15, Nε-S77, Oγ (pink) distances from the REST2-generated ensembles of unliganded Wzc_{CDAC} and SK_{Mtu}, respectively.

display the K15-S77 interaction have the highest values of θ , and correspondingly, the lowest values of $|h|$, and exist at the outer edges of the WOS. Taken together, these results suggest that the global conformations of SK_{Mtu} are coupled to the conformational state of the Walker-A Lys as was observed for Wzc_{CDAC} (Hajredini et al., 2020; Hajredini and Ghose, 2021).

As mentioned above, in a significant fraction of the structures that comprise the WOS of SK_{Mtu}, the Walker-A K15 forms a hydrogen bond with the Walker-B S77 (Figure 1B), generating a conformation that mirrors that seen in the Wzc_{CDAC} OS. In the latter case, the equivalent Walker-A Lys (K540) forms a salt bridge with the Walker-B D642. The β -strand housing the Walker-B motif in SK_{Mtu} is displaced relative to that in Wzc_{CDAC}; this enables S77 from the degraded Walker-B in the former to assume a spatial position that is approximately equivalent to the Walker-B D642 in the latter to facilitate an interaction with K15 (Figure 2A). It would therefore appear from our current results, that small local structural rearrangements allow the polar Ser to substitute for the now missing Walker-B Asp in preserving the interaction with the Walker-A K15 to maintain the open conformation (WOS) in SK_{Mtu}. However, despite these similarities, there are differences in stability between the open conformations involving the Walker-A Lys of SKs and BY-kinases. In Wzc_{CDAC}, this interaction mode, as assessed through the K540, Nε-D642, Cγ distance distribution, is well defined over the entire OS ensemble, showing a single peak with a maximum at ~ 3.3 Å (green trace in Figure 2B). The corresponding distance, K15, Nε-S77, Oγ in SK_{Mtu}, shows a broad distribution with the dominant peak centered at ~ 2.8 Å but containing only approximately 50% of the population in the WOS (pink trace in Figure 2B; this disorder is also evident from the bottom panel of Figure 1B). This suggests that while the orientation of the Walker-A Lys is well-defined in Wzc_{CDAC}, it is somewhat more dynamic in SK_{Mtu}. The reasons for this difference are analyzed in detail below.

As mentioned earlier, the conformational dynamics of the LID and SB domains have been suggested to be critical for function in the SKs in serving to stabilize the reaction-compatible state, and in facilitating product release (Blanco et al., 2013). The dynamics of these entities have been suggested as suitable targets for drug discovery (Prado et al., 2016). For

the sake of completeness, we also analyzed the local conformational variability in the REST2-generated ensembles of SK_{Mtu}. There is, not unexpectedly, a slight decrease in overall flexibility upon binding ATP•Mg²⁺ (Supplementary Figure S5A), and transition from the WOS to the WCS. The LID domain that is the most dynamic region of the protein in the unliganded state remains so in the presence of ATP•Mg²⁺, though there is some decrease in its overall flexibility. The SB domain is also dynamic in both states of SK_{Mtu}, albeit less so than the LID. To probe the presence of opening/closing motions of the SB domain with respect to the protein core, we defined a distance (d_{47-109} , Supplementary Figure S5B, top left panel) between the Ca atoms of D47 (one of the more dynamic parts of the SB domain and therefore likely to sense these motions, if any) and R109 (near start of the LID domain; one of the more rigid parts of the protein in its ATP•Mg²⁺-bound state). In the absence of ATP•Mg²⁺ a somewhat bimodal distribution of distances is seen (Supplementary Figure S5B top right panel). Analyses of structures corresponding to the two maxima indicate an open conformation with greater separation between the LID and SB domains (Supplementary Figure S5B bottom right panel), and a more closed conformation where this distance is reduced (Supplementary Figure S5B bottom middle panel). Only the latter conformation is seen in the ATP•Mg²⁺-bound state (Supplementary Figure S4B bottom left panel).

Conserved Features Define the Global Conformational States of P-Loop Enzymes

While S77 of SK_{Mtu} appears to substitute for the Asp/Glu residue of a canonical Walker-B motif in maintaining the WOS through its interaction with the Walker-A Lys, this position is poorly conserved in SKs. In contrast to the $\phi\phi$ SLGGG sequence in SK_{Mtu}, the consensus sequence for Walker-B motifs of SKs is $\phi\phi\phi$ TGGG (Leipe et al., 2003); the corresponding consensus in BY-kinases is $\phi\phi\phi\phi$ DT (Leipe et al., 2002). Notably, in Wzc_{CDAC}, the conserved Walker-B Thr (T643) participates in multiple interactions that also serve to stabilize the OS (Figure 3A, top panel). T643 forms a hydrogen bond through its backbone carbonyl

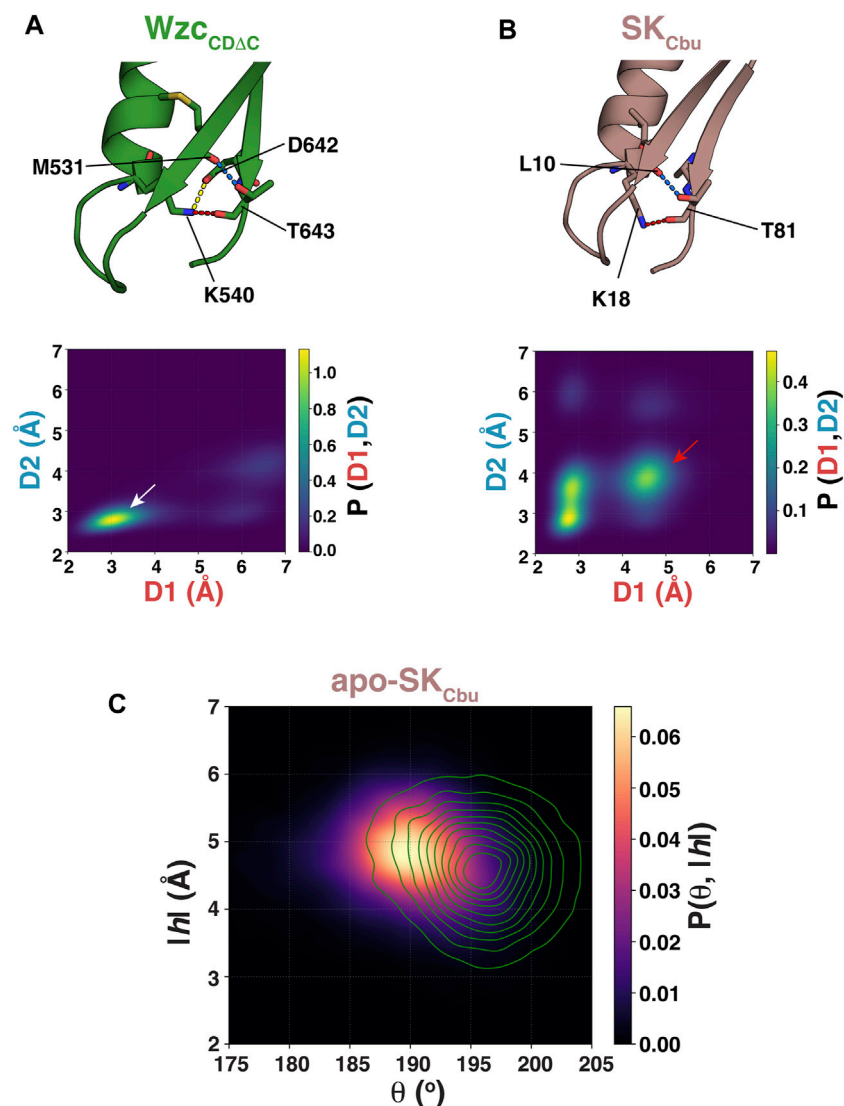


FIGURE 3 | (A) Interactions of the Walker-A K540 with the Walker-B D642 (yellow dashed line), of K540 with T643 (red dashed line), and of T643 with M531 (blue dashed line) for Wzc_{CDAC} in the OS are shown. **(B)** Interactions of the Walker-A K18 with Walker-B T81 (red dashed line) and of T81 with L10 (blue dashed line) for SK_{Cbu} in the WOS are shown. The lower panels, in each case, show the structural ensembles of unliganded Wzc_{CDAC} and SK_{Cbu} projected onto the 2-dimensional space spanned by the {K540-T643 (D1: K540,N ξ -T643,O), T643-M531 (D2: T643,O γ 1-M531,O)} and {K18-T81 (D1: K18,N ξ -T81,O), T81-L10 (D2: T81,O γ 1-L10,O)} distances, respectively. **(C)** The structural ensemble of unliganded SK_{Cbu} is projected onto θ - $|h|$ space and plotted using kernel density estimation. Also shown as green contours in the corresponding projection of the unliganded SK_{Mtu} ensemble.

with the K540 sidechain (supplementing the interaction of the latter with D642), and an additional hydrogen bond through its sidechain hydroxyl with the backbone carbonyl of M531 on the adjacent β -sheet. Projection of the unliganded Wzc_{CDAC} ensemble (Hajredini et al., 2020) onto the 2-dimensional space spanned by the K540-T643 (D1: K540,N ξ -T643,O) and T643-M531 (D2: T643,O γ 1-M531,O) distances, indicates a single major state (white arrow, **Figure 3A**, bottom panel). This robust network of interactions (K540-D642, K540-T643 and T643-M531) is the likely reason for the well-defined orientation of K540 in the OS (as is evident from the green trace in **Figure 2B**). This set of interactions is

drastically altered in the CS (T643 now interacts T532, D642 interacts with T541, the latter interaction is necessary to optimally engage Mg²⁺) and this is the likely cause of the well-defined separation between in the OS and the CS upon comparing the apo- and ATP•Mg²⁺ complexes of Wzc_{CDAC} (**Supplementary Figures S3B,C**).

To test whether a similar interaction mode involving the Walker-B Thr also stabilizes the WOS in SKs that carry the more conventional $\phi\phi\phi$ TGGG sequence, we performed an additional set of REST2 simulations on the unliganded state of *Coxiella burnetii* SK (SK_{Cbu}). Inspection of the structural ensemble of apo-SK_{Cbu} reveals the presence of hydrogen bonds between the

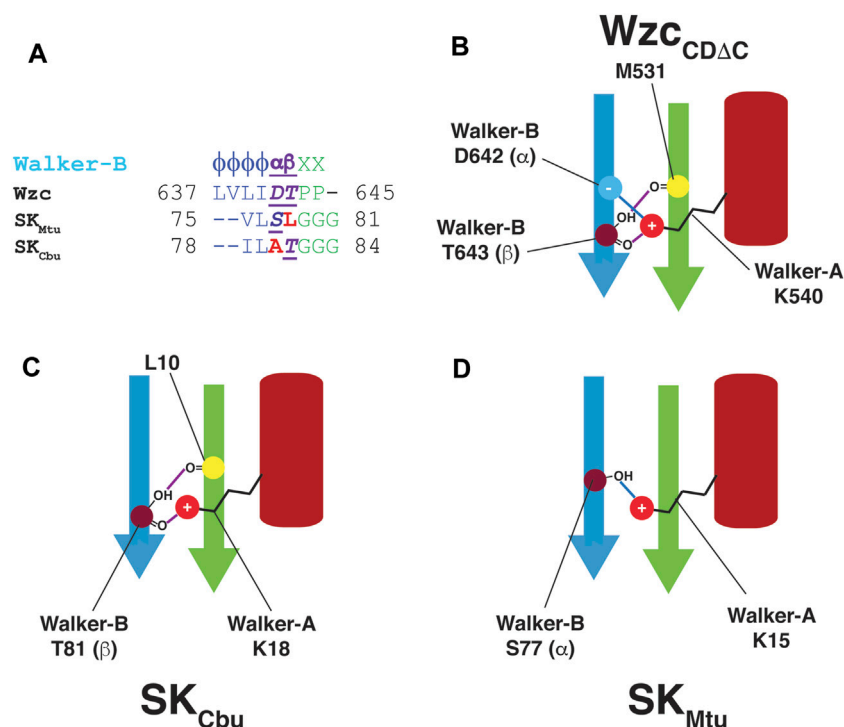


FIGURE 4 | (A) Comparison of the Walker-B sequences in Wzc, SK_{Mtu} and SK_{Cbu}. Wzc contains polar residues at both the α - and β -positions while only one of these positions is polar in SK_{Cbu} (β) and SK_{Mtu} (α). **(B)** The presence of polar residues at both the α - (D642) and β -positions (T643) in the Walker-B motif of Wzc_{CDAC} allows the intricate network of interactions to appropriately align the Walker-A Lys (K540) leading to its well-defined orientation in the OS. These interactions are depicted schematically; hydrogen-bonds are indicated using purple and blue lines. Since only one of the α - or β -positions contain a polar residue in **(C)** SK_{Cbu} (β :T81) and **(D)** SK_{Mtu} (α :S77), only one of these sets of interactions (blue or purple) is preserved allowing a greater degree of orientational freedom in the WOS for the corresponding Walker-A Lys (K18 in SK_{Cbu}, K15 in SK_{Mtu}).

backbone carbonyl and sidechain hydroxyl of the Walker-B Thr (T81) with the sidechain of the Walker-A K18 and the backbone carbonyl of L10 (that lies on the adjacent β -sheet), respectively (**Figure 3B**, top panel), mirroring the interaction mode seen in Wzc_{CDAC}. However, unlike in Wzc_{CDAC}, projection of the resultant ensemble onto the 2-dimensional space spanned by the K18-T81 (D1: K18,N ϵ -T81,O) and T81-L10 (D2: T81,O γ 1-L10,O) distances reveals a more diffuse set of states (**Figure 3B**, bottom panel). Nevertheless, a disruption in the T81-K18 interaction leads to a corresponding disruption of the T81-L10 interaction (red arrow in **Figure 3B**, bottom panel) suggesting that these two sets of interactions are coupled as in Wzc_{CDAC}. It is of note that the canonical Walker-B Thr in SK_{Cbu} forms two sets of the hydrogen bonds involving both its backbone and its sidechain to stabilize the WOS. This contrasts a single hydrogen bond involving the sidechain of the unique Walker-B Ser (S77) in SK_{Mtu}. One can therefore expect the WOS in SK_{Cbu} to be somewhat more closed than in SK_{Mtu}. Inspection of **Figure 3C** shows that this is indeed the case with reduced θ values (SK_{Cbu}: $190 \pm 5^\circ$, SK_{Mtu}: $196 \pm 4^\circ$) and a corresponding increase in the $|h|$ values (SK_{Cbu}: $4.8 \pm 0.6 \text{ \AA}$, SK_{Mtu}: $4.6 \pm 0.7 \text{ \AA}$).

Based on the discussion above, an SK Walker-B motif may be generalized by the following sequence: $\phi\phi\alpha\beta\text{GGG}$; the α - and β -positions represent structural equivalents of the BY-kinase

Walker-B Asp and Thr, respectively (**Figure 4A**). In unliganded Wzc_{CDAC}, both the α - and β -positions are optimal in that they contain polar residues. Thus, the Walker-A Lys (K540) can form hydrogen bonds with both the Walker-B Asp (D642 at the α -position) and with the Walker-B Thr (T643); the latter being a polar residue at the β -position is then able to utilize its sidechain to hydrogen bond with M531. This generates a mesh of interactions (shown schematically in **Figure 4B**) leading to a well-defined orientation of the Walker-A Lys (K540) in a stabilized OS. In contrast, the fact that only one of these two positions contains a polar residue for the two SKs analyzed here, with T81 (β) in SK_{Cbu} (**Figure 4C**) or S77 (α) in SK_{Mtu} (**Figure 4D**). Thus, only a subset of these interactions is possible in the SKs resulting in less well-defined orientation for the Walker-A Lys and a more diffuse WOS.

CONCLUSION

Through enhanced sampling MD simulations on unliganded SK_{Mtu} and its ATP•Mg²⁺ complex, we have shown that as in the case of BY-kinases, the conformational landscapes of SKs contain open (WOS) and closed (WCS) states. The former is found in the unliganded state, while the latter is induced by the presence of nucleotide and Mg²⁺. In the WOS, the Walker-A

K15 of SK_{Mtu} populates an extended conformation allowing the formation of a hydrogen bond with S77 of the degraded Walker-B motif that mimics the conserved Asp of the intact Walker-B motif in BY-kinases. In SK_{Cbu}, that contains a Walker-B sequence that is more characteristic of P-loop kinases of the Dx/D group (Leipe et al., 2003), the conserved T81 forms a hydrogen-bond with the Walker-A K18 through its backbone while simultaneously using its polar sidechain to interact with the neighboring L10 backbone to stabilize the WOS. Thus, it appears that specific contacts between the Walker-A and Walker-B motifs, more specifically those involving the so-called catalytic Lys on the former, are necessary in P-loop-containing kinases to maintain the open conformation that cannot efficiently co-ordinate Mg²⁺ (and therefore, ATP•Mg²⁺) and is not conducive for chemistry. The importance of the hydrogen-bond involving the Walker-A Lys is evident from the fact that the absence of a Walker-B Asp as in the SKs, is compensated by the presence of a polar residue at a similar position. While the functional role the OS and CS and the interactions encoded within each state for the catalytic domain of Wzc has been extensively tested experimentally (Hajredini et al., 2020; Hajredini and Ghose, 2021), we expect that the computational analyses presented here will provide guidelines for similar experimental validation of key functional features in SKs in particular, and in P-loop kinases, in general.

REFERENCES

- Aravind, L., Mazumder, R., Vasudevan, S., and Koonin, E. V. (2002). Trends in Protein Evolution Inferred from Sequence and Structure Analysis. *Curr. Opin. Struct. Biol.* 12, 392–399. doi:10.1016/S0959-440X(02)00334-2
- Bechet, E., Gruszczyk, J., Terreux, R., Gueguen-Chaignon, V., Vigouroux, A., Obadia, B., et al. (2010). Identification of Structural and Molecular Determinants of the Tyrosine-Kinase Wzc and Implications in Capsular Polysaccharide export. *Mol. Microbiol.* 77, 1315–1325. doi:10.1111/j.1365-2958.2010.07291.x
- Blanco, B., Prado, V., Lence, E., Otero, J. M., Garcia-Doval, C., Van Raaij, M. J., et al. (2013). Mycobacterium tuberculosis Shikimate Kinase Inhibitors: Design and Simulation Studies of the Catalytic Turnover. *J. Am. Chem. Soc.* 135, 12366–12376. doi:10.1021/ja405853p
- Edwards, H., Abeln, S., and Deane, C. M. (2013). Exploring Fold Space Preferences of New-Born and Ancient Protein Superfamilies. *Plos Comput. Biol.* 9, e1003325. doi:10.1371/journal.pcbi.1003325
- Franklin, M. C., Cheung, J., Rudolph, M. J., Burshteyn, F., Cassidy, M., Gary, E., et al. (2015). Structural Genomics for Drug Design against the Pathogen Coxiella Burnetii. *Proteins* 83, 2124–2136. doi:10.1002/prot.24841
- Gan, J., Gu, Y., Li, Y., Yan, H., and Ji, X. (2006). Crystal Structure of Mycobacterium tuberculosis Shikimate Kinase in Complex with Shikimic Acid and an ATP Analogue†,‡. *Biochemistry* 45, 8539–8545. doi:10.1021/bi0606290
- Grangeasse, C., Nessler, S., and Mijakovic, I. (2012). Bacterial Tyrosine Kinases: Evolution, Biological Function and Structural Insights. *Phil. Trans. R. Soc. B* 367, 2640–2655. doi:10.1098/rstb.2011.0424
- Hajredini, F., Piserchio, A., and Ghose, R. (2020). Long-range Dynamic Correlations Regulate the Catalytic Activity of the Bacterial Tyrosine Kinase Wzc. *Sci. Adv.* 6, eabd3718. doi:10.1126/sciadv.abd3718
- Hajredini, F., and Ghose, R. (2021). An ATPase with a Twist: a Unique Mechanism Underlies the Activity of the Bacterial Tyrosine Kinase. *Wzc. Sci. Adv.* 7, eabj5836. doi:10.1126/sciadv.abj5836
- Hartmann, M. D., Bourenkov, G. P., Oberschall, A., Strizhov, N., and Bartunik, H. D. (2006). Mechanism of Phosphoryl Transfer Catalyzed by Shikimate Kinase from *Mycobacterium tuberculosis*. *J. Mol. Biol.* 364, 411–423. doi:10.1016/j.jmb.2006.09.001
- Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., et al. (2017). CHARMM36m: an Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* 14, 71–73. doi:10.1038/nmeth.4067
- Lee, D. C., and Jia, Z. (2009). Emerging Structural Insights into Bacterial Tyrosine Kinases. *Trends Biochem. Sci.* 34, 351–357. doi:10.1016/j.tibs.2009.03.003
- Leipe, D. D., Koonin, E. V., and Aravind, L. (2003). Evolution and Classification of P-Loop Kinases and Related Proteins. *J. Mol. Biol.* 333, 781–815. doi:10.1016/j.jmb.2003.08.040
- Leipe, D. D., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002). Classification and Evolution of P-Loop GTPases and Related ATPases. *J. Mol. Biol.* 317, 41–72. doi:10.1006/jmbi.2001.5378
- Liu, P., Kim, B., Friesner, R. A., and Berne, B. J. (2005). Replica Exchange with Solute Tempering: a Method for Sampling Biological Systems in Explicit Water. *Proc. Natl. Acad. Sci.* 102, 13749–13754. doi:10.1073/pnas.0506346102
- Longo, L. M., Jabłońska, J., Vyas, P., Kanade, M., Kolodny, R., Ben-Tal, N., et al. (2020). On the Emergence of P-Loop NTPase and Rossmann Enzymes from a Beta-Alpha-Beta Ancestral Fragment. *Elife* 9, e64415. doi:10.7554/eLife.64415
- Pereira, J. H., De Oliveira, J. S., Canduri, F., Dias, M. V. B., Palma, M. S., Basso, L. A., et al. (2004). Structure of Shikimate Kinase from Mycobacterium Tuberculosis reveals the Binding of Shikimic Acid. *Acta Crystallogr. D Biol. Cryst.* 60, 2310–2319. doi:10.1107/S090744490402517x
- Prado, V., Lence, E., Maneiro, M., Vázquez-Ucha, J. C., Beceiro, A., Thompson, P., et al. (2016). Targeting the Motion of Shikimate Kinase: Development of Competitive Inhibitors that Stabilize an Inactive Open Conformation of the Enzyme. *J. Med. Chem.* 59, 5471–5487. doi:10.1021/acs.jmedchem.6b00483
- Scheeff, E. D., and Bourne, P. E. (2005). Structural Evolution of the Protein Kinase-like Superfamily. *Plos Comput. Biol.* 1, e49. doi:10.1371/journal.pcbi.0010049
- Shi, L., Ji, B., Kolar-Znika, L., Boskovic, A., Jadeau, F., Combet, C., et al. (2014). Evolution of Bacterial Protein-Tyrosine Kinases and Their Relaxed Specificity toward Substrates. *Genome Biol. Evol.* 6, 800–817. doi:10.1093/gbe/evu056
- Taylor, S. S., and Radzio-Andzelm, E. (1994). Three Protein Kinase Structures Define a Common Motif. *Structure* 2, 345–355. doi:10.1016/S0969-2126(00)00036-8
- Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly Related Sequences in the Alpha- and Beta-Subunits of ATP Synthase, Myosin, Kinases

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

RG conceived the project; FH performed and analyzed all simulations. FH prepared a first draft of the paper and figures that were refined by RG with input from FH.

FUNDING

Support from NSF grants MCB1937937 and PHY1811770 is acknowledged. FH is partially supported by United States Department of Education GAANN award P200A150068.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.747206/full#supplementary-material>

- and Other ATP-Requiring Enzymes and a Common Nucleotide Binding Fold. *EMBO J.* 1, 945–951. doi:10.1002/j.1460-2075.1982.tb01276.x
- Wang, L., Friesner, R. A., and Berne, B. J. (2011). Replica Exchange with Solute Scaling: a More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* 115, 9431–9438. doi:10.1021/jp204407d
- Yang, Y., Liu, J., Clarke, B. R., Seidel, L., Bolla, J. R., Ward, P. N., et al. (2021). The Molecular Basis of Regulation of Bacterial Capsule Assembly by Wzc. *Nat. Commun.* 12, 4349. doi:10.1038/s41467-021-24652-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Hajredini and Ghose. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Mapping Function from Dynamics: Future Challenges for Network-Based Models of Protein Structures

Lorenza Pacini^{1,2}, Rodrigo Dorantes-Gilardi^{2,3}, Laurent Vuillon³ and Claire Lesieur^{1,2*}

¹Ecole Centrale de Lyon, Ampère, UMR5005, Univ. Lyon, CNRS, INSA Lyon, Université Claude Bernard Lyon 1, Villeurbanne, France, ²Institut Rhônealpin des Systèmes Complexes, IXXI-ENS-Lyon, Lyon, France, ³USMB, CNRS, LAMA UMR5127, Le Bourget du Lac, France

OPEN ACCESS

Edited by:

George Lisi,
Brown University, United States

Reviewed by:

Guang Hu,
Soochow University, China
Luis Diambra,
National University of La Plata,
Argentina

*Correspondence:

Claire Lesieur
claire.lesieur@ens-lyon.fr

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 20 July 2021

Accepted: 19 August 2021

Published: 11 October 2021

Citation:

Pacini L, Dorantes-Gilardi R, Vuillon L
and Lesieur C (2021) Mapping
Function from Dynamics: Future
Challenges for Network-Based Models
of Protein Structures.
Front. Mol. Biosci. 8:744646.
doi: 10.3389/fmolb.2021.744646

Proteins fulfill complex and diverse biological functions through the controlled atomic motions of their structures (functional dynamics). The protein composition is given by its amino-acid sequence, which was assumed to encode the function. However, the discovery of functional sequence variants proved that the functional encoding does not come down to the sequence, otherwise a change in the sequence would mean a change of function. Likewise, the discovery that function is fulfilled by a set of structures and not by a unique structure showed that the functional encoding does not come down to the structure either. That leaves us with the possibility that a set of atomic motions, achievable by different sequences and different structures, encodes a specific function. Thanks to the exponential growth in annual depositions in the Protein Data Bank of protein tridimensional structures at atomic resolutions, network models using the Cartesian coordinates of atoms of a protein structure as input have been used over 20 years to investigate protein features. Combining networks with experimental measures or with Molecular Dynamics (MD) simulations and using typical or ad-hoc network measures is well suited to decipher the link between protein dynamics and function. One perspective is to consider static structures alone as alternatives to address the question and find network measures relevant to dynamics that can be subsequently used for mining and classification of dynamic sequence changes functionally robust, adaptable or faulty. This way the set of dynamics that fulfill a function over a diversity of sequences and structures will be determined.

Keywords: protein structure, protein dynamics, protein function, network-based models, space occupancy

INTRODUCTION

Proteins fulfill complex and diverse biological functions through the controlled atomic motions of their structures (Wingert et al., 2021). The protein composition is given by its amino-acid sequence, which was assumed to encode the function. However, the discovery of functional sequence variants proved that the functional encoding does not come down to the sequence, otherwise a change in the sequence would mean a change of function. Likewise, the discovery that function is fulfilled by a set of structures and not by a unique structure showed that the functional encoding does not come down to the structure either (Jaffe, 2005; Jaffe, 2020; Parisi et al., 2015).

The next alternative is that sets of atomic motions, achievable by different sequences and different structures, encode a specific function (Bahar et al., 2010). This is consistent with the multiple

dynamic paths that fulfill allostery (Buchenberg et al., 2017). The challenge lies in distinguishing the set of dynamics associated with sequence variants functionally robust or functionally adapted (change of function) from the set of dynamics associated with functional failure. Inferring function from the protein dynamics is also important because pathological variants impacting the protein dynamics but not the protein structure, limit traditional structure-based drug discovery methods (Demir et al., 2021).

Network science is appropriate to study system dynamics from protein structures because it offers multiple avenues to study the complex spatiotemporal relationships between interacting entities (Barrat et al., 2004; Barabási, 2013; Unicomb et al., 2017). Integrative approaches combining experimental data or Molecular Dynamics (MD) simulation with network-based models enable to link protein structures to protein dynamics and function Demir et al. (2011), Leitner and Yamato (2018), Liang et al. (2018), Ponzoni and Bahar (2018), Bourgeat et al. (2019), Gheeraert et al. (2019), Melo et al. (2020), Bourgeat et al. (2021), Di Paola and Leitner (2021) and for a review see (Liang et al., 2020). On one hand, global mode analysis, elastic network models (ENM), dynamics network models (DNM) and protein energy networks (PEN) are used to track multiple scale dynamics in proteins, identify allosteric pathways and residues involved in biological activities. On the other, perturbation response scanning (PRS) and evolutionary network models are used to investigate the impact of mutations on protein features in particular for disease mutations. ELM applied on different protein family members also allows associating scale of motions to various types of activities (Wingert et al., 2021).

The advantages of network based models in probing protein dynamics come from the inference of amino acid and atomic links from the structure. Now, one on-going question is to clarify why network measures pinpoint functional residues (e.g., allostery) or distinguish disease mutations from the rest of the residues in order to better understand what properties amino acids have in a structure that make them functionally tolerant to mutations or not. The comparison of network measures and network models over proteins spanning large dynamics scales from enzyme to pore-forming toxins and over their sequence variants will help validating network measures as hallmarks of functional dynamics and of functional dynamic perturbations related to diseases.

One alternative perspective to network integrative approaches is to find network measures that are relevant to functional dynamics simply from protein structures. This implies a network measure probing collective slow motions and therefore shared across proteins and independent of amino acid features as observed from global modes (Bahar et al., 2010). In addition, a network measure with amino acid specific characteristics is expected if it embed the dynamics of a specific function. An allosteric enzyme and a pore-forming toxin have 3D-structures that share multiple scale collective dynamics but yet they have very different motions to fulfill their functions.

We consider the neighborhoods of each amino acid of a protein as potentially relevant to the problem. This is because

on average over its neighbors every amino acid makes moderate and similar number of atomic interactions, a property shared by many different proteins (Dorantes-Gilardi et al., 2018). In addition, each neighborhood is different in terms of number of neighbors and type of neighbors (Dorantes-Gilardi et al., 2018). Thus, neighborhoods satisfy the two conditions to embed protein dynamics. Moreover, the neighborhoods describe the spatial position of the amino acids in the structure, which carves the space occupied by the amino acid atoms and hence uncovers the space left available between amino acids where atomic motions can take place. The relation between the space occupied by entities and the system dynamics is a broad topic from granular material to urban and protein systems (Liang and Dill, 2001; Majmudar and Behringer, 2005; Henzler-Wildman and Kern, 2007; Majmudar et al., 2007; Barthelemy, 2011; Dorantes-Gilardi et al., 2018; Gheeraert et al., 2019; Naganathan, 2019).

To analyze neighborhoods, a protein structure is modeled by an amino acid network (AAN), where the nodes are amino acids and the links are atomic interactions between amino acids, inferred from atomic proximity (Dorantes-Gilardi et al., 2018). The space occupied by neighborhoods is described in terms of amino acids with the node degree and in terms of atoms with the node weight (Methods). Classically, a unique cutoff around the threshold for chemical interactions (5 Å) is used to investigate protein features but here the neighborhoods are computed at different cutoff distances to probe the space occupancy at different spatial scales (Vuillon and Lesieur, 2015; Vilorio et al., 2017). This condition is necessary to track the multiple dynamics scales associated with functional dynamics (Henzler-Wildman and Kern, 2007; Munoz and Cerminara, 2016).

Our case study is the third PDZ domain of the synaptic protein PDS-95 (PDB 1BE9) chosen because the functional impact of most of its single amino-acid mutations is known from experiments as well as some double mutations (McLaughlin et al., 2012; Salinas and Ranganathan, 2018). Thus, this case study is appropriate for future validation of the network measures to link dynamics and function.

METHODS

Starting from the Protein Data Bank (PDB) data, protein structures are modeled using the Amino Acid Network (AAN), an established model in Computational Biology (Dorantes-Gilardi et al., 2018). The AAN is a graph $G = (V; E)$, with V is the set of the N nodes of the network (vertices of the graph) and E the set of links of the network (edges of the graph).

Nodes of the AAN: Each node in the AAN corresponds to one amino acid of the protein's structure named according to the protein sequence:

$$V = \{i | i \text{ is an amino acid}\}. \quad (1)$$

Links of the AAN: A link is an atomic interaction defined by atomic proximity: two amino acids i and j are connected if there exists at least one couple of atoms, one belonging to i and one

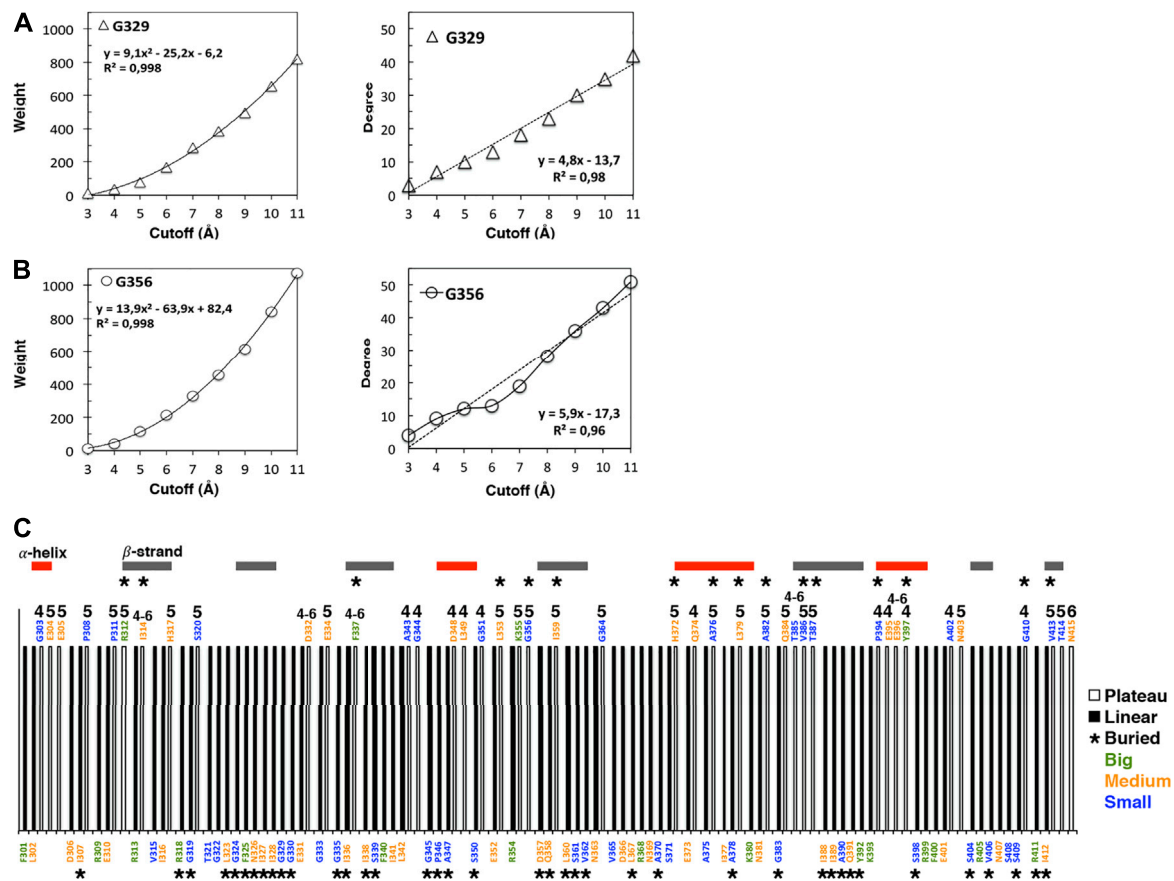


FIGURE 1 | Cutoff dependencies of the weight and the degree of the AAN nodes of the 1BE9 structure. **(A)** Examples of the weight cutoff dependencies for glycine, the smallest amino acid. The increase is quadratic ($R^2 \sim 1$). **(B)** Example of the degree cutoff dependencies for the same amino acids. One is linear ($R^2 \sim 0,98$, top panel), and the other has a plateau between cutoffs 5 and 6 Å (bottom panel). **(C)** 1D-barcode: cutoff dependencies of the 1BE9 sequence. The amino acids are colored according to the length of their extended side chains (blue: length < 3 Å, orange: $3 \text{ Å} \leq \text{length} < 5$ Å, and green: length ≥ 5 Å). The colored horizontal bars represent the secondary structures along the sequence. The star is for buried amino acids, and the numbers indicate the cutoff range of the plateau: four is for a plateau at cutoffs 4 to 5 Å, 5 is for plateau at cutoffs five to 6 Å, six is for a plateau at cutoffs 6 to 7 Å and four-six is for a plateau at cutoffs 4 to 6 Å.

belonging to j , at a distance lower or equal to a given threshold (Cutoff distance c),

$$E = \{(i, j) | i, j \in V \text{ with } i \neq j \text{ and } \exists (\text{atom}_i \in i, \text{atom}_j \in j) \text{ with } \text{dist}(\text{atom}_i, \text{atom}_j) \leq c\}. \quad (2)$$

Link weights of the AAN: Each link is weighted according to the number of atomic couples that respect the cutoff condition:

$$w_{ij} = \left| \left\{ (\text{atom}_i \in i, \text{atom}_j \in j) \text{ with } \text{dist}(\text{atom}_i, \text{atom}_j) \leq c \text{ and } i \neq j \right\} \right| \quad (3)$$

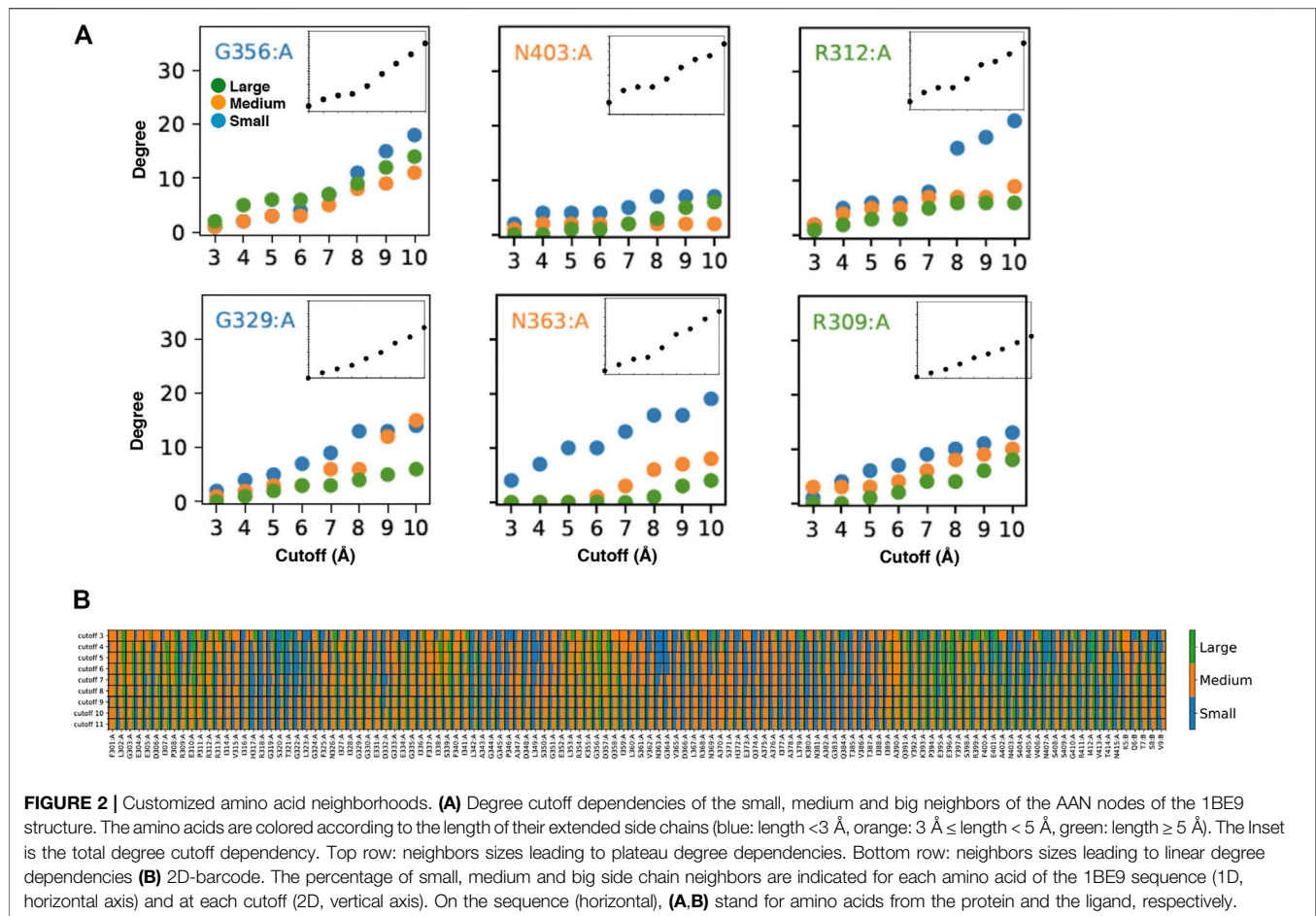
where the pipe symbol $|$ denotes the cardinality of the set (i.e. the number of elements of the set). When $c = 5$ Å, that is a threshold for chemical interactions, the link weights measure the number of atomic interactions between two amino acids.

Packing around amino acids: In the AAN, the node degree k_i , defined as the number of amino acid neighbors of a node i , measures the amino-acid packing around the amino acid i ,

referred to as the amino-acid neighborhood. The node weight w_i is defined as the sum over all the weights of the links that connect the node i to its neighbors ($w_i = \sum_{j \in N(i)} w_{ij}$ with $N(i)$ the set of neighbors of node i). The node weight measures the atomic packing around the amino acid i , referred to as the atomic neighborhood.

Cutoff distance: Different cutoff distances are used in this study such that the packing around each amino acid is described at different length-scale via the neighborhoods at variable cutoffs. The cutoffs are integers and range from 3 to 11 Å such that the packing within chemical reach (≤ 5 Å) and above chemical reach (> 5 Å) are monitored. The rationale is to distinguish amino acids by their 'chemical' neighborhoods and above chemical reach neighborhoods to probe space occupancy involved in multiple spatiotemporal scales.

Plateau versus linear degree dependencies: The degree dependency with the cutoff is plotted. Some amino acids show a linear dependency of the degree with the cutoff and are referred to as linear amino acids. Some amino acids exhibit a plateau over



some cutoffs determined with the derivative (here just Δk because the cutoffs are consecutive integers) equals to zero or equals to one if and only if the derivative at cutoffs before the plateau is equal to four or more. We refer to these amino acids as plateau amino acids.

Amino acid side chain length classification: Amino acids are classified by side-chain lengths as follows. Side chain length < 3 Å are small amino acids (G, A, P, S, V, I, T, C), side chain length between $3 \text{ Å} \leq \text{length} < 5 \text{ Å}$ are medium amino acids (L, E, D, H, N, Q, M) and side chain length $\geq 5 \text{ Å}$ are big amino acids (F, K, R, Y, W).

1D-barcode: The 1D-barcode represents the degree dependency with the cutoff (plateau- or linear-) of the amino acids of the 1BE9 sequence (1D).

2D-barcode: The 2D-barcode represents the degree dependency with the cutoff of the amino acids of the 1BE9 sequence (1D) at variable cutoffs (2D).

RESULTS AND DISCUSSION

The AAN of the PDZ domain of the synaptic protein PDS-95 is generated using the PDB 1BE9 and the weights and degrees of every amino acid nodes are computed at different cutoff distances

(Methods). We recall that the weight of a node describes its atomic packing, referred to as its atomic neighborhood while the degree of a node describes its amino-acid packing referred to as its amino-acid neighborhood. The atomic neighborhood takes into account the features of the amino acids. The weight and degree cutoff dependencies are plotted to investigate how atoms and amino acid neighbors occupy the space around each amino acid at different scales in the protein structure. This is a proxy of the dynamics as the more space occupied the less space left available for atomic motions.

The weight cutoff dependencies are quadratic indicating that the weight (the number of atomic interactions) increases with the square of the cutoff distance, i.e., with a surface of contact between atoms rather than a volume of contacts (**Figure 1A; Supplementary Figure S1A**). The degree cutoff dependencies are not quadratic but exhibit two distinct behaviors: neighborhoods with linear dependency with the cutoff (**Figure 1B**, top in the protein structure) and neighborhoods with a plateau over some cutoffs (**Figure 1B**, bottom and **Supplementary Figure S1B**). We call linear (plateau) amino acids, the amino acids whose degree follows a linear (plateau) dependency with the cutoff. The plateau is due to a lack of amino acid neighbors at higher cutoffs and not to a lack of atoms since the weight cutoff dependencies have no plateau (**Supplementary Figure S1A**). The plateau can result from having big amino acid neighbors

that occupy the space over more than one cutoff versus having small amino acids that lead to a linear increase over the same cutoffs (**Figure 2A**).

Ten nodes have a plateau between 4 and 5 Å, twenty between 5 and 6 Å, one between 6 and 7 Å and six over 4 to 6 Å (**Figure 1B**).

All amino-acid types observed in 1BE9 except histidine which appears only twice, adopt plateau and linear neighborhoods (**Figure 1C**). Methionine, cysteine and tryptophan are not present in 1BE9. Twenty-six surface-exposed amino acids out of fifty-eight are plateau (~half) and sixteen buried amino acids out of fifty-seven are plateau (~a third). 27% of the β -strand amino acids are plateau against 35 and 48% for loops and α -helix amino acids, respectively. Thus the plateau and linear neighborhoods are achieved regardless amino-acid type, position in the structure and secondary structure, which makes the degree cutoff dependencies insensitive to amino acid features as global modes. Together with the plateau built at chemical-interaction threshold, it is consistent with the degree cutoff dependencies probing collective motions. Accordingly, a 1D-barcode representing the plateau- or linear-neighborhood of each amino acid of the sequence could be used to characterize the multiple scale dynamic features of a protein (**Figure 1C**).

To have a plateau or a linear degree cutoff dependency for one amino acid type implies customizing neighborhoods as illustrated by the higher number of big amino acid neighbors for plateau amino acids than linear amino acids (**Figure 2A**, compare top and bottom). To have it for any amino acid type also implies neighborhood customized to the central amino acids as seen on **Figure 2A** (compare across amino acid types). Thus, the linear and plateau neighborhoods accessible to all amino acids are nevertheless built from specific amino-acid neighborhoods (**Figure 2A**; **Supplementary Figure S2**). This means the protein dynamic-functional specificity could be embedded in the collective motions through specific space occupancy arising from neighborhood diversities. A 2D-barcode representing the amino acid neighborhood size specificities of each amino acid of the protein sequence (**Figure 2B**, horizontal axis) at different cutoffs (**Figure 2B**, vertical axis) could be used to characterize the protein specific embedded dynamics (**Figure 2B**).

We can see from this single case 2D barcode classes of neighborhoods in terms of neighbor sizes which anticipate classes of dynamics and of responses upon mutations supporting the possibility to use the data for mining dynamics and its relation to function (**Figure 2B**). Some positions are composed of a majority of one-size neighbors over the cutoffs (e.g. small neighbors: R318, T321, G322, L349, N363; medium: F301, G303, R312, G333, G351, A390) while others are a mixture of neighbor sizes (e.g. A308, I341, K355, A383, Q391, E401). In addition, some positions change neighbor sizes over the cutoffs (e.g., E396, E401). However, these sole data do not have the

statistics to make hypothesis on which neighborhood classes lead to which dynamic classes or draw conclusion between the neighborhood specificity, dynamics and function.

CONCLUSION

The study shows that amino acid neighborhoods and not only amino acids and amino acid pairs, contain information relevant to protein dynamics, opening a new perspective to explore the link between dynamics and function. The 1D barcode of, for example, an enzyme and a pore-forming toxin can be compared to determine common and distinct features which can in turn be analyzed with the 2D barcode to survey both the neighborhood diversity of the common 1D barcode features and of the distinct 1D barcode features assuming the former identifies positions functionally insensitive (protein aspecific) and the later positions functionally sensitive (protein specific). The analysis of each protein variants can be used to validate the assumption as well as database analysis.

This will contribute to diagnosing dynamic functional faults and dynamic functional diversity.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

LP performed research, analyzed the data, and contributed to writing the manuscript; RD-G performed the research; LV analyzed the data and contributed to writing the manuscript and CL designed the research, analyzed the data, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank the CNRS for funding.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.744646/full#supplementary-material>

REFERENCES

- Bahar, I., Lezon, T. R., Bakan, A., and Shrivastava, I. H. (2010). Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chem. Rev.* 110 (3), 1463–1497. doi:10.1021/cr900095e
- Barabási, A.-L. (2013). Network Science. *Phil. Trans. R. Soc. A.* 371 (1987), 20120375. doi:10.1098/rsta.2012.0375
- Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A. (2004). The Architecture of Complex Weighted Networks. *Proc. Natl. Acad. Sci.* 101 (11), 3747–3752. doi:10.1073/pnas.0400087101
- Barthelemy, M. (2011). Spatial Networks. *Phys. Rep.* 499 (1–3), 1–101. doi:10.1016/j.physrep.2010.11.002

- Bourgeat, L., Pacini, L., Serghei, A., and Lesieur, C. (2021). Experimental Diagnostic of Sequence-Variant Dynamic Perturbations Revealed by Broadband Dielectric Spectroscopy. *Structure*. doi:10.1016/j.str.2021.05.005
- Bourgeat, L., Serghei, A., and Lesieur, C. (2019). Experimental Protein Molecular Dynamics: Broadband Dielectric Spectroscopy Coupled with Nanoconfinement. *Sci. Rep.* 9 (1), 17988. doi:10.1038/s41598-019-54562-8
- Buchenberg, S., Sittel, F., and Stock, G. (2017). Time-resolved Observation of Protein Allosteric Communication. *Proc. Natl. Acad. Sci. USA* 114 (33), E6804–E6811. doi:10.1073/pnas.1707694114
- Demir, Ö., Baronio, R., Salehi, F., Wassman, C. D., Hall, L., Hatfield, G. W., et al. (2011). Ensemble-based Computational Approach Discriminates Functional Activity of P53 Cancer and rescue Mutants. *Plos Comput. Biol.* 7 (10), e1002238. doi:10.1371/journal.pcbi.1002238
- Demir, Ö., Barros, E. P., Offutt, T. L., Rosenfeld, M., and Amaro, R. E. (2021). An Integrated View of P53 Dynamics, Function, and Reactivation. *Curr. Opin. Struct. Biol.* 67, 187–194. doi:10.1016/j.sbi.2020.11.005
- Di Paola, L., and Leitner, D. M. (2021). Network models of biological adaptation at the molecular scale: Comment on "Dynamic and thermodynamic models of adaptation" by AN Gorban et al. *Phys. Life Rev.* 38, 124–126. doi:10.1016/j.plrev.2021.05.008
- Dorantes-Gilardi, R., Bourgeat, L., Pacini, L., Vuillon, L., and Lesieur, C. (2018). In Proteins, the Structural Responses of a Position to Mutation Rely on the Goldilocks Principle: Not Too many Links, Not Too Few. *Phys. Chem. Chem. Phys.* 20 (39), 25399–25410. doi:10.1039/c8cp04530e
- Gheeraert, A., Pacini, L., Batista, V. S., Vuillon, L., Lesieur, C., and Rivalta, I. (2019). Exploring Allosteric Pathways of a V-type Enzyme with Dynamical Perturbation Networks. *The J. Phys. Chem. B* 123, 3452. doi:10.1021/acs.jpcb.9b01294
- Henzler-Wildman, K., and Kern, D. (2007). Dynamic Personalities of Proteins. *Nature* 450 (7172), 964–972. doi:10.1038/nature06522
- Jaffe, E. K. (2020). Wrangling Shape-Shifting Morphees to Tackle Disease and Approach Drug Discovery. *Front. Mol. Biosciences* 7, 582966. doi:10.3389/fmolb.2020.582966
- Jaffe, E. K. (2005). Morphees - a New Structural Paradigm for Allosteric Regulation. *Trends Biochemical Sciences* 30 (9), 490–497. doi:10.1016/j.tibs.2005.07.003
- Leitner, D. M., and Yamato, T. (2018). *Mapping Energy Transport Networks in Proteins*. Hoboken, New Jersey, US: Wiley Online Library.
- Liang, J., and Dill, K. A. (2001). Are Proteins Well-Packed? *Biophysical J.* 81 (2), 751–766. doi:10.1016/s0006-3495(01)75739-6
- Liang, Z., Hu, J., Yan, W., Jiang, H., Hu, G., and Luo, C. (2018). Deciphering the Role of Dimer Interface in Intrinsic Dynamics and Allosteric Pathways Underlying the Functional Transformation of DNMT3A. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1862 (7), 1667–1679. doi:10.1016/j.bbagen.2018.04.015
- Liang, Z., Verkhivker, G. M., and Hu, G. (2020). Integration of Network Models and Evolutionary Analysis into High-Throughput Modeling of Protein Dynamics and Allosteric Regulation: Theory, Tools and Applications. *Brief. Bioinformatics* 21 (3), 815–835. doi:10.1093/bib/bbz029
- Majmudar, T. S., Sperl, M., Luding, S., and Behringer, R. P. (2007). Jamming Transition in Granular Systems. *Phys. Rev. Lett.* 98 (5), 058001. doi:10.1103/PhysRevLett.98.058001
- Majmudar, T. S., and Behringer, R. P. (2005). Contact Force Measurements and Stress-Induced Anisotropy in Granular Materials. *Nature* 435 (7045), 1079–1082. doi:10.1038/nature03805
- McLaughlin, R. N., Jr, Poelwijk, F. J., Raman, A., Gosal, W. S., and Ranganathan, R. (2012). The Spatial Architecture of Protein Function and Adaptation. *Nature* 491 (7422), 138–142. doi:10.1038/nature11500
- Melo, M. C. R., Bernardi, R. C., De La Fuente-nunez, C., and Luthey-Schulten, Z. (2020). Generalized Correlation-Based Dynamical Network Analysis: a New High-Performance Approach for Identifying Allosteric Communications in Molecular Dynamics Trajectories. *J. Chem. Phys.* 153 (13), 134104. doi:10.1063/5.0018980
- Muñoz, V., and Cerminara, M. (2016). When Fast Is Better: Protein Folding Fundamentals and Mechanisms from Ultrafast Approaches. *Biochem. J.* 473 (17), 2545–2559. doi:10.1042/bcj20160107
- Naganathan, A. N. (2019). Modulation of Allosteric Coupling by Mutations: from Protein Dynamics and Packing to Altered Native Ensembles and Function. *Curr. Opin. Struct. Biol.* 54, 1–9. doi:10.1016/j.sbi.2018.09.004
- Parisi, G., Zea, D. J., Monzon, A. M., and Marino-Buslje, C. (2015). Conformational Diversity and the Emergence of Sequence Signatures during Evolution. *Curr. Opin. Struct. Biol.* 32, 58–65. doi:10.1016/j.sbi.2015.02.005
- Ponzoni, L., and Bahar, I. (2018). Structural Dynamics Is a Determinant of the Functional Significance of Missense Variants. *Proc. Natl. Acad. Sci. U S A* 115, 4164–4169. doi:10.1073/pnas.1715896115
- Salamanca Vilorio, J., Allega, M. F., Lambugh, M., and Papaleo, E. (2017). An Optimal Distance Cutoff for Contact-Based Protein Structure Networks Using Side-Chain Centers of Mass. *Sci. Rep.* 7 (1), 2838–2911. doi:10.1038/s41598-017-01498-6
- Salinas, V. H., and Ranganathan, R. (2018). Coevolution-based Inference of Amino Acid Interactions Underlying Protein Function. *eLife* 7, e34300. doi:10.7554/eLife.34300
- Unicomb, S., Iniguez, G., and Karsai, M. (2017). Threshold Driven Contagion on Weighted Networks. *arXiv preprint arXiv:1707.02185*.
- Vuillon, L., and Lesieur, C. (2015). From Local to Global Changes in Proteins: a Network View. *Curr. Opin. Struct. Biol.* 31, 1–8. doi:10.1016/j.sbi.2015.02.015
- Wingert, B., Krieger, J., Li, H., and Bahar, I. (2021). Adaptability and Specificity: How Do Proteins Balance Opposing Needs to Achieve Function? *Curr. Opin. Struct. Biol.* 67, 25–32. doi:10.1016/j.sbi.2020.08.009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Pacini, Dorantes-Gilardi, Vuillon and Lesieur. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Molecular Level Insights Into the Structural and Dynamic Factors Driving Cytokine Function

Jennifer Y. Cui and George P. Lisi*

Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, United States

OPEN ACCESS

Edited by:

Yong Wang,
Zhejiang University, China

Reviewed by:

Hongbin Wan,
Novartis Institutes for BioMedical
Research, United States
Chao-Yie Yang,
University of Tennessee Health
Science Center (UTHSC),
United States
Hao-Jen Hsu,
Tzu Chi University, Taiwan

*Correspondence:

George P. Lisi
george_lisi@brown.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 09 September 2021

Accepted: 06 October 2021

Published: 25 October 2021

Citation:

Cui JY and Lisi GP (2021) Molecular
Level Insights Into the Structural and
Dynamic Factors Driving
Cytokine Function.
Front. Mol. Biosci. 8:773252.
doi: 10.3389/fmolb.2021.773252

Cytokines are key mediators of cellular communication and regulators of biological advents. The timing, quantity and localization of cytokines are key features in producing specific biological outcomes, and thus have been thoroughly studied and reviewed while continuing to be a focus of the cytokine biology community. Due to the complexity of cellular signaling and multitude of factors that can affect signaling outcomes, systemic level studies of cytokines are ongoing. Despite their small size, cytokines can exhibit structurally promiscuous and dynamic behavior that plays an equally important role in biological activity. In this review using case studies, we highlight the recent insight gained from observing cytokines through a molecular lens and how this may complement a system-level understanding of cytokine biology, explain diversity of downstream signaling events, and inform therapeutic and experimental development.

Keywords: cytokine, protein dynamics, immunology, spectroscopy, allostery

INTRODUCTION

Cytokines are small signaling proteins that serve as effectors of immunity and require spatial or temporal expression to regulate biological outcomes (Altan-Bonnet and Mukherjee, 2019). Cytokines are especially prevalent in disease states where their presence can indicate pathological progression or conversely, a protective effect. For example, signaling cascades that maintain homeostatic balance use cytokines to coordinate appropriate responses that initiate or terminate inflammation. The lens through which researchers often view cytokine biochemistry involves their localization and quantity, using phenomenological models to derive system-level explanations of function (Altan-Bonnet and Mukherjee, 2019). Despite the utility of this approach, the molecular aspects of cytokine function remain unclear, hindering efforts to intuitively modulate their mechanisms and promiscuous interactions. Understanding the structural and environmental factors that stimulate transformations of cytokines is critical to disentangling their role in biological events. Such insight can inform 1) the role played by cytokines as agents of cellular communication, 2) the improvement of tools leveraged for research, and 3) the development of cytokines as therapeutics or targets in clinical treatment plans. In the past decade, many enzymes and metabolic proteins that accommodate multiple non-overlapping functions by “moonlighting” have been shown to utilize intrinsic protein dynamics as a mechanistic driving force (Mani et al., 2015; Boukouris et al., 2016; Chen et al., 2018; Chen et al., 2021). However, structural plasticity in cytokines with broadly demonstrated abilities to harbor multiple functions within compact scaffolds has not been as thoroughly investigated as a moonlighting mechanism.

Researchers use cytokines to replicate immunological conditions in the laboratory in order to study behaviors of specific cell types (Schluns and Lefrançois, 2003; Arango Duque and Descoteaux,

2014; Roan et al., 2019), to replicate extracellular environments, to observe biological responses *in vitro* (Kämpfer et al., 2017; Berghaus et al., 2018; Barnes and Somerville, 2020), and as a measurable biomarker of disease. Cytokines are also used as therapeutics in clinical settings to specifically stimulate immune responses, including interleukins (ILs) in cancer and colony-stimulating factors (CSFs) in general neutropenia (Mehta et al., 2015; Berraondo et al., 2019; Conlon et al., 2019). Given the breadth of the roles played by cytokines intrinsically and as research tools or in clinical settings (i.e. from bench to bedside), it is clear that having thorough understanding of the mechanisms driving cytokine function is crucial. One factor that complicates structure-function investigations of cytokines, and the establishment of general rules that govern their function, is the fact that these proteins exert a great deal of their influence under conditions of disease, which are more difficult to mimic in structural characterizations. In this Review, we discuss current efforts to define the mechanisms of cytokines at the “residue level,” where changes in atomic structure or conformational dynamics of the protein yields information relevant to physiological scenarios. We highlight case studies of cytokines with emerging diverse roles, while also showcasing experimental and theoretical approaches for understanding cytokine biochemistry. Though we cannot comprehensively discuss every avenue of investigation in this Review (Künze et al., 2014; Herring et al., 2015; Joseph et al., 2015; Künze et al., 2016; Park et al., 2017; Azimzadeh Irani and Ejtehadi, 2019; Penk et al., 2019; Gangele et al., 2020; Roy, 2020; Yang, 2020), each of the case studies below offers a useful piece insight into an extremely complex class of biomolecules that may aid in the development of broadly applicable strategies to study molecular interactions.

CASE STUDY I: MACROPHAGE MIGRATION INHIBITORY FACTOR

Macrophage migration inhibitory factor (MIF) is implicated in a wide range of disease states, where elevated MIF levels are markers for tumorigenesis (Bach et al., 2008; Noe and Mitchell, 2020), sepsis (Toldi et al., 2021), acute respiratory distress syndrome (ARDS) (Donnelly et al., 1997; Lai et al., 2003), arthritis (Onodera et al., 1999; Kim et al., 2007; Llamas-Covarrubias et al., 2012), colitis (de Jong et al., 2001), and Malaria (Han et al., 2010; Sun et al., 2012; Baeza Garcia et al., 2018). MIF is highly conserved in diverse organisms and is heavily involved in inflammatory responses, cell cycle control, the sensing of pathogenic stimuli, activation of innate defense responses, recruitment of immune cells, and prevention of p53-mediated apoptosis (Hudson et al., 1999; Mitchell et al., 2002). Despite a compact structural scaffold and a lack of clear modular domains, MIF contains several sequence motifs, including two enzymatic sites for disparate tautomerase and oxidoreductase activities (a third nuclease activity was recently proposed), that provide MIF with a surprising functional complexity (Kleemann et al., 1998; Lubetsky et al., 1999; Wang et al., 2021). MIF is also remarkably amenable to interactions with a number of target proteins, despite

its overall architectural rigidity (Mühlhahn et al., 1996), and is proposed to be involved, directly or indirectly, in 49 signaling events. (Subbannayya et al., 2016). This functional promiscuity necessitates an intimate understanding of the mechanism(s) that govern its cellular interactions, but also complicate efforts to assign any singular function to MIF with a high level of mechanistic detail.

Efforts to understand MIF at the molecular level have focused on 1) studies of mutations that affect its structure and function, 2) studies of signaling within its homotrimeric structure, and 3) studies of interactions between MIF and drug-like molecules (Lubetsky et al., 1999; Mühlhahn et al., 1996; Pantouris et al., 2015; Pantouris et al., 2020; Fan et al., 2013; Trivedi-Parmar and Jorgensen, 2018; Cournia et al., 2009; Cho et al., 2010; Takahashi et al., 2009). Site-directed mutagenesis has been employed by several laboratories to examine the structural and biological impact of perturbations to the MIF catalytic and receptor binding sites (Pantouris et al., 2015; Pantouris et al., 2020; Fan et al., 2013; El-Turk et al., 2012; Oda et al., 2008; El-Turk et al., 2008; Bendrat et al., 1997). X-ray crystallographic reports of >100 MIF structures in the Protein Data Bank show essentially no observable changes in response to mutations, substrates or inhibitors, suggesting MIF is highly rigid and resistant to architectural deformation. However, identical solution NMR experiments show that atomic level fingerprints of MIF variants are unique, indicating that the local structure of MIF is in fact quite flexible in solution and may play a significant role in modulating its biochemistry (Pantouris et al., 2018; Schinagl et al., 2018). More recent work has leveraged these findings to explore the existence of signaling pathways that propagate chemical information between distant regions of the MIF structure. Pantouris, et al characterized two such pathways in successive reports, first demonstrating through molecular dynamics (MD) simulations that nanosecond timescale conformational fluctuations were critical to the organization of the binding interface for the MIF-induced activation of CD74 (Pantouris et al., 2015). Further analysis of correlated dynamics in MIF revealed the CD74 activation signal to originate from a previously uncharacterized residue, Tyr99, at the solvent channel formed by the three-fold symmetric MIF structure (**Figure 1A**). Interestingly, this allosteric site modulates the dynamics of the MIF protein on multiple timescales when mutated, and not only affects CD74 activation *in vivo*, but also regulates tautomerase enzymatic activity (**Figure 1B**) through a different signaling pathway originating from Tyr99 (Pantouris et al., 2018; Pantouris et al., 2020; Parkins et al., 2021). The factors that influence the ramping up or down of MIF enzymatic activity include protein dynamics, intramolecular hydrogen bonds and hydrogen bonds with solvent molecules, and pi-stacking interactions from aromatic side chains lining the signaling pathway. At present, it is still unclear how, or if, CD74 activation or tautomerase activity are preferentially activated in the cell. Nonetheless, recent solution studies of MIF revealed two of its biochemical functions to be connected by a region of its structure previously uncharacterized at the molecular level. Numerous studies of MIF in complex with drugs and other modulators have never targeted the Tyr99 regulatory site, but

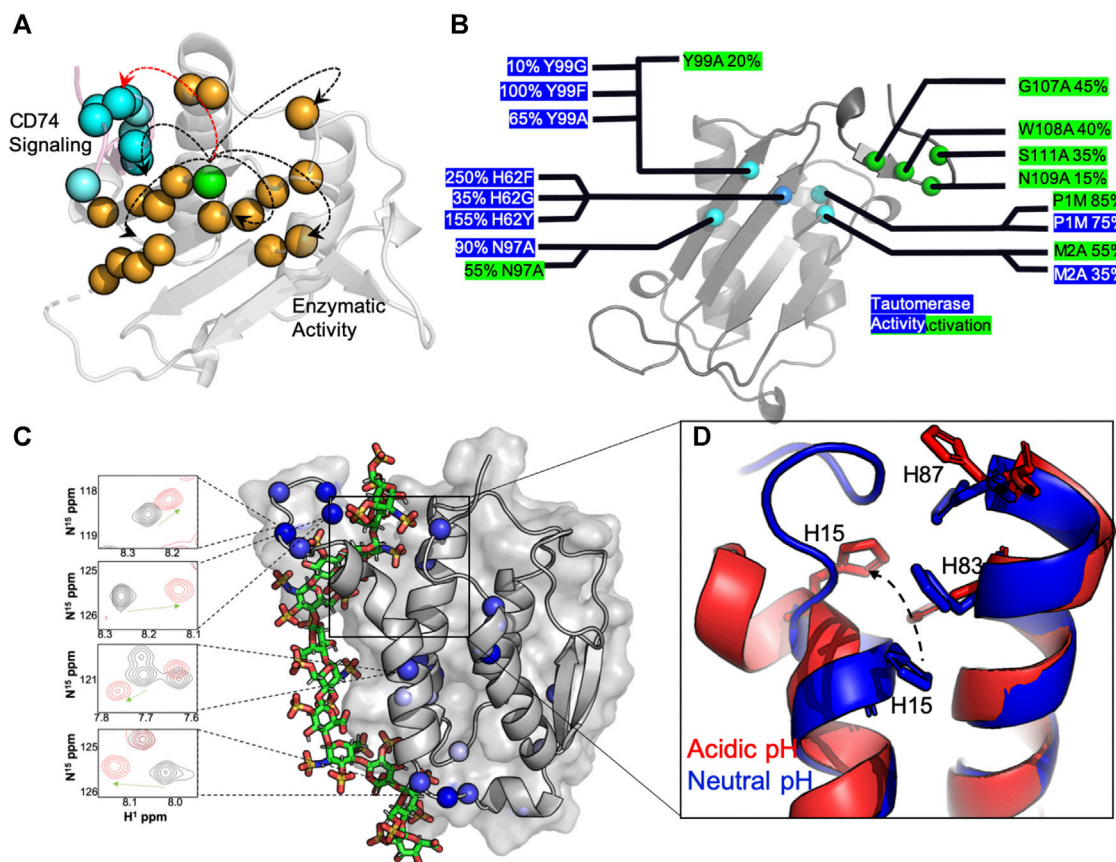


FIGURE 1 | Molecular insights from structure-function studies of MIF and GMCSF. **(A)** Tyr99 mutations (green) induce widespread NMR chemical shifts and line broadening (orange spheres), suggesting that a large portion of the MIF structure “senses” short and long-range correlations (arrows) that influence MIF signaling. The C-terminus of the adjacent MIF monomer, the site of CD74 activation, is shown in pink with blue spheres. **(B)** Experimental measurements of MIF activities determined from mutagenesis. Activity levels are presented as a percentage of wild-type (wt) MIF activity, with enzymatic catalysis (blue) determined from the enol-keto tautomerization of 4-hydroxyphenyl pyruvate (4-HPP) (Lubetsky et al., 1999) and CD74 activation (green) determined from the recruitment of neutrophils to murine lungs *in vivo* (Takahashi et al., 2009). **(C)** A computationally-derived GMCSF structure in the presence of a heparin oligo at pH 5.5. Experimental NMR chemical shift perturbations (blue spheres) are mapped onto GMCSF, and NMR resonances in the presence (red) and absence (black) of heparin are shown to correlate with the proposed binding site. **(D)** An overlay of GMCSF structures at pH 7.4 (blue, PDB: 2GMF) and pH 5.5 (red, from simulations), where disruption of H15, H83 and H87 pi-stacking alignment is clear at acidic pH and in the presence of heparin.

have modulated the same enzymatic and CD74 activities from different locations within MIF, (Cournia et al., 2009; Cho et al., 2010; Trivedi-Parmar and Jorgensen, 2018), presenting a new avenue for structure-function evaluations.

CASE STUDY II: GRANULOCYTE MACROPHAGE COLONY-STIMULATING FACTOR

Granulocyte macrophage colony-stimulating factor (GMCSF) is involved in myelopoiesis (Becher et al., 2016; Ushach and Zlotnik, 2016), immunomodulation (Lotfi et al., 2019), and pro-inflammatory signaling (Bhattacharya et al., 2015; Hamilton, 2019), specifically the promotion of alveolar macrophage activity leading to surfactant accumulation in the lungs (Antoniou, 2010; Kumar et al., 2018). GMCSF has been shown

to aggravate conditions such as rheumatoid arthritis and multiple sclerosis (van Nieuwenhuijze et al., 2013; Shiomi et al., 2016), but it can also ameliorate diseases such as type-I diabetes (Frydrych et al., 2019), and is used clinically to combat neutropenia (Mehta et al., 2015). As such, it is a clear example of a cytokine that performs both beneficial and pathological functions. Interestingly, several studies have identified alternate receptor conformations of stoichiometric assembly that can lead to receptor activation (i.e. cell survival, activation, or differentiation) (Hansen et al., 2008; Broughton et al., 2016) and beyond its receptor structure, GMCSF itself has been shown to adopt different structural and dynamic properties dependent on the pH of its chemical environment, which may contribute to its modular signaling or suggest additional non-overlapping roles for GMCSF (Cui et al., 2020).

The GMCSF structure has been studied by X-ray crystallography (Rozwarski et al., 1996), and more recently

NMR spectroscopy (Aubin et al., 2008), providing unique insight into its overall fold and plasticity. One particular aspect of GMCSF function that has benefitted substantially from atomic level studies has been the knowledge of its interactions with clinically relevant molecules. In these cases, there has been particular interest in understanding how GMCSF, a clinically administered protein with a promiscuous interactome, could interfere with other therapeutics under conditions of disease. For example, Wettreich and coworkers showed GMCSF to form pH-dependent complexes with heparin, another widely used therapeutic, via analytical chromatography and light scattering (Wettreich et al., 1999). These experiments proposed a binding mechanism and laid the groundwork for understanding how signaling proteins like GMCSF can be tuned by environmental stimuli to form selective ligand complexes. The GMCSF-heparin interaction has since been resolved at the atomic level with NMR spectroscopy under varied solution pH conditions and in the presence of precisely controlled heparin oligosaccharide chain lengths (Cui et al., 2020). The GMCSF protein was found to undergo pH-dependent structural and dynamic changes that modulated its affinity for heparin. Such a pH shift would occur at air-lung interfaces or in tumor microenvironments and was proposed to be essential for the ionization of three histidine residues that nucleate the heparin and the stimulation of multi-timescale dynamics that shift the conformation of N-terminal α -helices to expose the binding pocket. NMR chemical shifts mapped the heparin binding site and relaxation experiments, supported by molecular simulations, confirmed a high level of conformational flexibility along the binding interface and proposed a model for the complex structure (**Figure 1C**). This binding interaction is dependent not only on the accessibility of the binding pocket, but the size of the heparin as well, where heparins able to span the anchor points at the histidine triad (**Figure 1D**) and the adjacent flexible linker with several lysine residues bind with greater affinity than smaller chain heparins.

Atomic level studies of GMCSF mapped a clinically-relevant interaction to reveal the binding site and the biophysical factors that influence its accessibility to ligands. This work also reinforced an important aspect of cytokine biochemistry that is intimately tied to the hypothesized dynamic mechanisms of action; the utility of conducting biophysical studies in regimes beyond those of “ideal” conditions. In this case, if GMCSF were only characterized based on its presence in a certain niche (i.e. a microenvironment at neutral pH mimicking only circulating GMCSF), the resulting biochemical outcomes would not have captured its high affinity state for heparin and a gap in our understanding of its mechanism(s) would remain. Put another way, if GMCSF is expressed in high quantities in acidic environments and in the presence of high heparin concentrations, a measurable amount of biological activity presumed to be related to the GMCSF concentration would likely not correlate with its expected activity-to-quantity ratio.

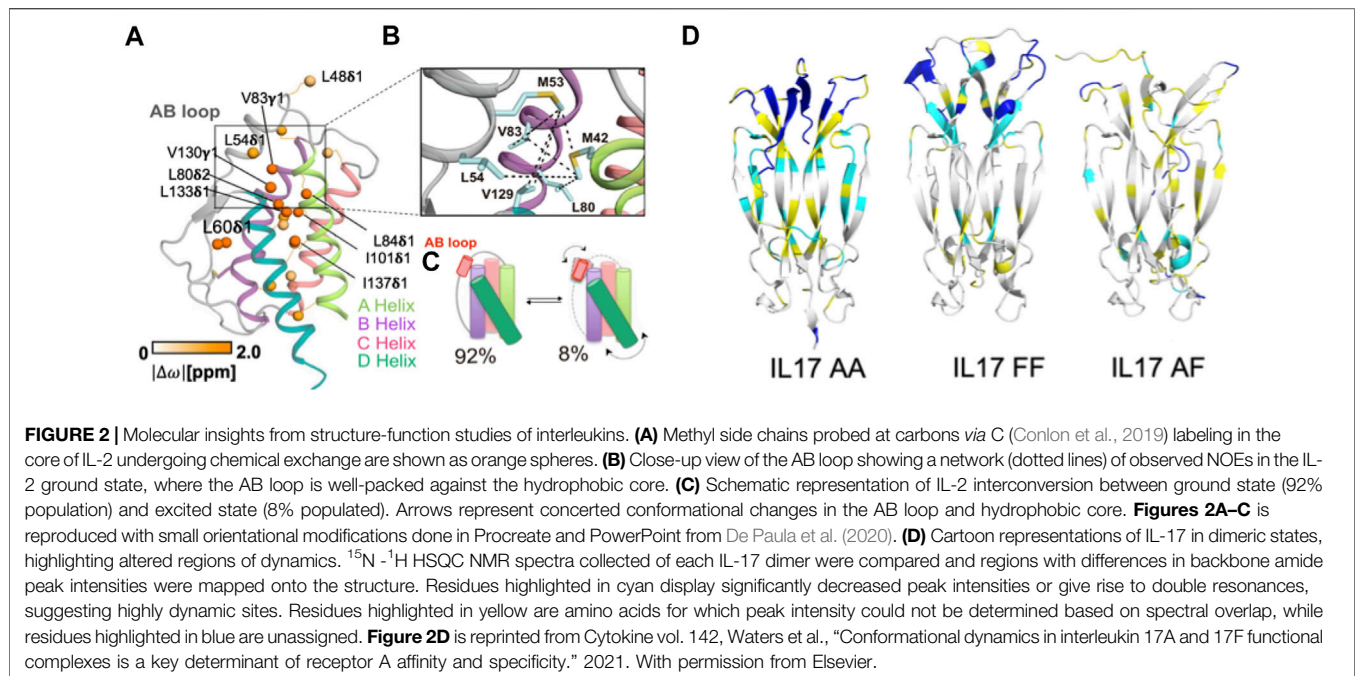
CASE STUDY III: INTERLEUKIN-1

Interleukins are a diverse class of cytokines with low sequence identity that are secreted by several cell types (beyond the originally presumed leukocyte-specific secretion that spawned their name). One subclass, the interleukin (IL)-1 family, comprises 11 pro- and anti-inflammatory cytokines that have abnormal expression profiles in many autoimmune diseases (Dinarello, 2018), which presumably affects their signal transduction by the IL-1 receptor (IL-1R). Interestingly, Ge et al recently demonstrated that IL-1R regulates its signaling activity via structural dynamics (Ge et al., 2019). Most IL-1 family members are commonly expressed as full-length precursors that require proteolytic processing for biologically mature forms, leading to significant structural variability. IL-1 α is cleaved by the cysteine protease calpain, whereas IL-1 β and IL-18 require proteolytic cleavage by the inflammasome (Black et al., 1988; van de Veerdonk et al., 2011). IL-33 and IL-36 utilize neutrophil proteinases such as elastase and proteinase-3 (Clancy et al., 2018), while IL-37 is cleaved by caspase-1 (Kumar et al., 2002) and IL-38 is bioactive as a full-length molecule. Mechanisms between members of the IL-1 family differ in binding interactions with the receptor machinery as Günther et al. have shown that IL-1 β and IL-33 use different molecular mechanisms for engaging with the IL-1 receptor accessory protein (IL-1Ra) (Günther et al., 2017). While Hou et al. took an alternate approach by leveraging different binding characteristics of IL-1 β and IL-1Ra to engineer chimeric receptor antagonists of IL-1R to alleviate the chronically inflammatory dry eye disease (Hou et al., 2013).

Several groups have explored the biophysical aspects of these structures in order to understand how IL-1s transduce signals throughout the cell. One example is the computational approach taken by Ozbabacan et al to map the structural postures of IL-1, IL-1R1 and IL-1RAP as well as downstream signaling proteins (such as MYD88 and TOLLIP) when previously identified oncogenic SNPs cause changes in amino acid residues (Acuner Ozbabacan et al., 2014). This study highlights the change in protein dynamics and its subsequent effect on an entire downstream signaling pathway. The authors present a computationally derived structural pathway detailing the conformational changes related to IL-1 signaling, revealing the mechanisms behind mutations that lead to oncogenic consequences. This work is one example that underscores the importance of atomistic shifts in cytokine structure (rather than a global ensemble) as valuable in the pursuit of targeted therapies and explorations of regulatory mechanisms.

CASE STUDY IV: INTERLEUKIN-2

IL-2 is produced predominantly by activated T-cells in secondary lymphoid organs, where it is consumed by these and other CD25⁺ cells, including regulatory T-cells (T_{Reg}) and lymphocytes. IL-2 has been used in both agonistic and antagonistic contexts for immunotherapies, including 1) low-dose efforts to increase T_{Reg} cell counts in autoimmunity, chronic inflammatory conditions and graft rejection (Tahvildari and Dana, 2019) and 2) high-dose



administration to expand cytotoxic lymphocyte populations for the treatment of metastatic cancer (Jiang et al., 2016; Buchbinder et al., 2019). Efforts to understand the signaling mechanism and therapeutic potential of IL-2 have focused primarily on cellular and live animal studies, though others have tackled questions relating to its molecular structure and conformational landscape.

Critically, it has been established that IL-2 adopts discrete structures in solution that are populated *via* allosteric activation as shown by Sgourakis and coworkers (De Paula et al., 2020). These results build upon a previous body of literature that clearly demonstrate divergent cell fates based on signaling related to the structural plasticity of IL-2. Specifically, methyl-based chemical exchange NMR spectroscopy interrogated IL-2 dynamics on a timescale (μs - ms) relevant to enzyme catalysis, ligand binding, and allostery (Figures 2A,B) (Lisi and Loria, 2016). The authors reported that a flexible loop between two N-terminal α -helices acts as a switch (Figure 2C), toggling between an autoinhibited state that has greater affinity for antibody binding, leading to proliferation of T_{reg} cell populations, and a productive state, where IL-2 is primed to interact with IL-2 receptor- α , leading to activation of T_{eff} (effector) cell populations (De Paula et al., 2020). De Paula et al. go on to propose an allosteric network of residues serving as the link between the hydrophobic core of IL-2 and the AB loop (Figure 2B), highlighting how a focus on localized molecular motions can lead to novel and detailed mechanistic insight about a cytokine that can be both leveraged and accounted for in future therapeutic efforts.

CASE STUDY V: INTERLEUKIN-17

IL-17 promotes inflammation and plays a protective role in antimicrobial immunity (McGeachy et al., 2019). In the latter

case, IL-17 was shown to mediate protection against extracellular pathogens (Valeri and Raffatellu, 2016; Eyerich et al., 2017) by working with IL-22 (a related cytokine also produced by IL-17-expressing cells) to stimulate production of antimicrobial peptides (Wozniak et al., 2014; Mulcahy et al., 2016; Moyat et al., 2017). IL-17 is therefore a double-edged sword with biological properties that make it difficult to predict its role in inflammatory diseases with a polymicrobial etiology. It is possible that IL-17 exerts both protective and destructive effects, as suggested in distinct mouse models (Borkner et al., 2021; Milovanovic et al., 2020; Cruz et al., 2010; Dallenbach et al., 2015; Righetti et al., 2018; O'Connor Jr et al., 2009), although chronic IL-17 receptor signaling can turn a potentially protective acute inflammatory response into chronic immunopathology (Zenobia and Hajishengallis, 2015).

Despite its status as one of best studied cytokines in immunology, recent structure-function relationships and dynamic studies have provided new insight into its regulatory mechanism. Specifically, IL-17 conformational dynamics were shown to dictate signaling efficiency *via* a strong influence over binding affinity for this system. The structural interconversions of IL-17A and IL-17F were studied with NMR spectroscopy by Waters et al, where the IL-17A homodimer was demonstrated to be far more dynamic on slow timescales than the IL-17F homodimer or IL-17A/IL-17F heterodimer (Figure 2D). The aforementioned dynamics aid the IL-17A homodimer in toggling between at least two distinct conformations and the structural perturbations during this process were mapped to the receptor binding region (Waters et al., 2021). The authors propose that these structural fluctuations are likely to affect binding affinity to receptor A (but not other IL-17 receptors), providing insight into the activation of receptor signaling based on atomic level dynamics that cannot be resolved from systems

level observations examining only the presence and quantity of IL-17.

CONCLUDING REMARKS

The mechanisms by which cytokines influence a large number of biochemical pathways and molecular interactions have been mysterious. The fact that most proteins in this class have very small structures with simple folds only adds to this paradox. It is difficult to link all of cytokine biochemistry with a common thread, primarily because the majority of information regarding cytokine functions comes from cellular or live animal studies, with comparatively few biophysical investigations. However, the current literature describing cytokine structure has generally shown that 1) protein dynamics that toggle active conformations of cytokines are a driver of functional

promiscuity, 2) the cellular or chemical environment modulates cytokine structure and interactions with specific binding partners, and 3) despite their small size, mechanisms of cytokine regulation can be driven by principles of allostery and concerted motion.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by Rhode Island Foundation Grant GR5290658.

REFERENCES

- Acuner Ozbabacan, S. E., Gursoy, A., Nussinov, R., and Keskin, O. (2014). The Structural Pathway of Interleukin 1 (IL-1) Initiated Signaling Reveals Mechanisms of Oncogenic Mutations and SNPs in Inflammation and Cancer. *Plos Comput. Biol.* 10. e1003470 doi:10.1371/journal.pcbi.1003470
- Altan-Bonnet, G., and Mukherjee, R. (2019). Cytokine-mediated Communication: a Quantitative Appraisal of Immune Complexity. *Nat. Rev. Immunol.* 19, 205–217. doi:10.1038/s41577-019-0131-x
- Antoniu, S. A. (2010). GM-CSF Pathway Correction in Pulmonary Alveolar Proteinosis. *Expert Opin. Biol. Ther.* 10, 1357–1365. doi:10.1517/14712598.2010.510507
- Arango Duque, G., and Descoteaux, A. (2014). Macrophage Cytokines: Involvement in Immunity and Infectious Diseases. *Front. Immunol.* 5, 491. doi:10.3389/fimmu.2014.00491
- Aubin, Y., Gingras, G., and Sauvé, S. (2008). Assessment of the Three-Dimensional Structure of Recombinant Protein Therapeutics by NMR Fingerprinting: Demonstration on Recombinant Human Granulocyte Macrophage-colony Stimulation Factor. *Anal. Chem.* 80, 2623–2627. doi:10.1021/ac7026222
- Azimzadeh Irani, M., and Ejtehadi, M. R. (2019). GAG Positioning on IL-1RI; A Mechanism Regulated by Dual Effect of Glycosylation. *Glycobiology* 29, 803–812. doi:10.1093/glycob/cwz055
- Bach, J.-P., Rinn, B., Meyer, B., Dodel, R., and Bacher, M. (2008). Role of MIF in Inflammation and Tumorigenesis. *Oncology* 75, 127–133. doi:10.1159/000155223
- Baeza Garcia, A., Siu, E., Sun, T., Exler, V., Brito, L., Hekele, A., et al. (2018). Neutralization of the Plasmodium-Encoded MIF Ortholog Confers Protective Immunity against Malaria Infection. *Nat. Commun.* 9, 2714. doi:10.1038/s41467-018-05041-7
- Barnes, B. J., and Somerville, C. C. (2020). Modulating Cytokine Production via Select Packaging and Secretion from Extracellular Vesicles. *Front. Immunol.* 11, 1040. doi:10.3389/fimmu.2020.01040
- Becher, B., Tugues, S., and Greter, M. (2016). GM-CSF: From Growth Factor to Central Mediator of Tissue Inflammation. *Immunity* 45, 963–973. doi:10.1016/j.immuni.2016.10.026
- Bendrat, K., Al-Abed, Y., Callaway, D. J. E., Peng, T., Calandra, T., Metz, C. N., et al. (1997). Biochemical and Mutational Investigations of the Enzymatic Activity of Macrophage Migration Inhibitory Factor. *Biochemistry* 36, 15356–15362. doi:10.1021/bi971153a
- Berghaus, L. J., Giguère, S., Bordin, A. I., and Cohen, N. D. (2018). Effects of Priming with Cytokines on Intracellular Survival and Replication of *Rhodococcus Equi* in Equine Macrophages. *Cytokine* 102, 7–11. doi:10.1016/j.cyt.2017.12.011
- Berraondo, P., Sanmamed, M. F., Ochoa, M. C., Etxeberria, I., Aznar, M. A., Pérez-Gracia, J. L., et al. (2019). Cytokines in Clinical Cancer Immunotherapy. *Br. J. Cancer* 120, 6–15. doi:10.1038/s41416-018-0328-y
- Bhattacharya, P., Budnick, I., Singh, M., Thirupathi, M., Alharshaw, K., Elshabrawy, H., et al. (2015). Dual Role of GM-CSF as a Pro-inflammatory and a Regulatory Cytokine: Implications for Immune Therapy. *J. Interferon Cytokine Res.* 35, 585–599. doi:10.1089/jir.2014.0149
- Black, R. A., Kronheim, S. R. K., Cantrell, M., Deeley, M. C., March, C. J., Prickett, K. S., et al. (1988). Generation of Biologically Active Interleukin-1 Beta by Proteolytic Cleavage of the Inactive Precursor. *J. Biol. Chem.* 263, 5. doi:10.1016/s0021-9258(19)76559-4
- Borkner, L., Curham, L. M., Wilk, M. M., Moran, B., and Mills, K. H. G. (2021). IL-17 Mediates Protective Immunity against Nasal Infection with *Bordetella Pertussis* by Mobilizing Neutrophils, Especially Siglec-F+ Neutrophils. *Mucosal Immunol.* 14, 1183–1202. doi:10.1038/s41385-021-00407-5
- Boukouris, A. E., Zervopoulos, S. D., and Michelakis, E. D. (2016). Metabolic Enzymes Moonlighting in the Nucleus: Metabolic Regulation of Gene Transcription. *Trends in Biochemical Sciences* 41, 712–730. doi:10.1016/j.tibs.2016.05.013
- Broughton, S. E., Hercus, T. R., Nero, T. L., Dottore, M., McClure, B. J., Dhagat, U., Taing, H., Gorman, M. A., King-Scott, J., Lopez, A. F., and Parker, M. W. (2016). Conformational Changes in the GM-CSF Receptor Suggest a Molecular Mechanism for Affinity Conversion and Receptor Signaling. *Structure* 24, 1271–1281. doi:10.1016/j.str.2016.05.017
- Buchbinder, E. I., Dutcher, J. P., Daniels, G. A., Curti, B. D., Patel, S. P., Holtan, S. G., Miletello, G. P., Fishman, M. N., Gonzalez, R., Clark, J. I., Richart, J. M., Lao, C. D., Tykodi, S. S., Silk, A. W., and McDermott, D. F. (2019). Therapy with high-dose Interleukin-2 (HD IL-2) in metastatic melanoma and renal cell carcinoma following PD1 or PDL1 inhibition. *J. immunotherapy cancer* 7, 49. doi:10.1186/s40425-019-0522-3
- Chen, C., Liu, H., Zabad, S., Rivera, N., Rowin, E., Hassan, M., Gomez De Jesus, S. M., Llinás Santos, P. S., Kravchenko, K., Mikhova, M., Ketterer, S., Shen, A., Shen, S., Navas, E., Horan, B., Raudsepp, J., and Jeffery, C. (2021). MoonProt 3.0: an update of the moonlighting proteins database. *Nucleic Acids Res* 49, D368–D372. doi:10.1093/nar/gkaa1101
- Chen, C., Zabad, S., Liu, H., Wang, W., and Jeffery, C. (2018). MoonProt 2.0: an Expansion and Update of the Moonlighting Proteins Database. *Nucleic Acids Res.* 46, D640–D644. doi:10.1093/nar/gkx1043
- Cho, Y., Crichlow, G. V., Vermeire, J. J., Leng, L., Du, X., Hodsdon, M. E., et al. (2010). Allosteric Inhibition of Macrophage Migration Inhibitory Factor Revealed by Ibudilast. *Proc. Natl. Acad. Sci.* 107, 11313–11318. doi:10.1073/pnas.1002716107
- Clancy, D. M., Sullivan, G. P., Moran, H. B. T., Henry, C. M., Reeves, E. P., McElvaney, N. G., et al. (2018). Extracellular Neutrophil Proteases Are Efficient Regulators of IL-1, IL-33, and IL-36 Cytokine Activity but Poor Effectors of Microbial Killing. *Cel Rep.* 22, 2937–2950. doi:10.1016/j.celrep.2018.02.062
- Conlon, K. C., Miljkovic, M. D., and Waldmann, T. A. (2019). Cytokines in the Treatment of Cancer. *J. Interferon Cytokine Res.* 39, 6–21. doi:10.1089/jir.2018.0019

- Cournia, Z., Leng, L., Gandavadi, S., Du, X., Bucala, R., and Jorgensen, W. L. (2009). Discovery of Human Macrophage Migration Inhibitory Factor (MIF)-CD74 Antagonists via Virtual Screening. *J. Med. Chem.* 52, 416–424. doi:10.1021/jm801100v
- Cruz, A., Fraga, A. G., Fountain, J. J., Rangel-Moreno, J., Torrado, E., Saraiva, M., et al. (2010). Pathological Role of Interleukin 17 in Mice Subjected to Repeated BCG Vaccination after Infection with *Mycobacterium tuberculosis*. *J. Exp. Med.* 207, 1609–1616. doi:10.1084/jem.20100265
- Cui, J. Y., Zhang, F., Nierzwicki, L., Palermo, G., Linhardt, R. J., and Lisi, G. P. (2020). Mapping the Structural and Dynamic Determinants of pH-Sensitive Heparin Binding to Granulocyte Macrophage Colony Stimulating Factor. *Biochemistry* 59, 3541–3553. doi:10.1021/acs.biochem.0c00538
- Dallenbach, K., Maurer, P., Röhn, T., Zabel, F., Kopf, M., and Bachmann, M. F. (2015). Protective Effect of a Germline, IL-17-neutralizing Antibody in Murine Models of Autoimmune Inflammatory Disease. *Eur. J. Immunol.* 45, 1238–1247. doi:10.1002/eji.201445017
- de Jong, Y. P., Abadia-Molina, A. C., Satoskar, A. R., Clarke, K., Rietdijk, S. T., Faubion, W. A., et al. (2001). Development of Chronic Colitis Is Dependent on the Cytokine MIF. *Nat. Immunol.* 2, 1061–1066. doi:10.1038/ni720
- De Paula, V. S., Jude, K. M., Nerli, S., Glassman, C. R., Garcia, K. C., and Sgourakis, N. G. (2020). Interleukin-2 Druggability Is Modulated by Global Conformational Transitions Controlled by a Helical Capping Switch. *Proc. Natl. Acad. Sci. USA* 117, 7183–7192. doi:10.1073/pnas.2000419117
- Dinarello, C. A. (2018). Overview of the IL-1 Family in Innate Inflammation and Acquired Immunity. *Immunol. Rev.* 281, 8–27. doi:10.1111/immr.12621
- Donnelly, S. C., Haslett, C., Reid, P. T., Grant, I. S., Wallace, W. A. H., Metz, C. N., et al. (1997). Regulatory Role for Macrophage Migration Inhibitory Factor in Acute Respiratory Distress Syndrome. *Nat. Med.* 3, 320–323. doi:10.1038/nm0397-320
- El-Turk, F., Cascella, M., Ouertatani-Sakouhi, H., Narayanan, R. L., Leng, L., Bucala, R., et al. (2008). The Conformational Flexibility of the Carboxy Terminal Residues 105–114 Is a Key Modulator of the Catalytic Activity and Stability of Macrophage Migration Inhibitory Factor†. *Biochemistry* 47, 10740–10756. doi:10.1021/bi800603x
- El-Turk, F., Fauvet, B., Ashrafi, A., Ouertatani-Sakouhi, H., Cho, M.-K., Neri, M., et al. (2012). Characterization of Molecular Determinants of the Conformational Stability of Macrophage Migration Inhibitory Factor: Leucine 46 Hydrophobic Pocket. *PLoS One* 7, e45024. doi:10.1371/journal.pone.0045024
- Eyerich, K., Dimartino, V., and Cavani, A. (2017). IL-17 and IL-22 in Immunity: Driving protection and Pathology. *Eur. J. Immunol.* 47, 607–614. doi:10.1002/eji.201646723
- Fan, C., Rajasekaran, D., Syed, M. A., Leng, L., Loria, J. P., Bhandari, V., et al. (2013). MIF Intersubunit Disulfide Mutant Antagonist Supports Activation of CD74 by Endogenous MIF Trimer at Physiologic Concentrations. *Proc. Natl. Acad. Sci.* 110, 10994–10999. doi:10.1073/pnas.1221817110
- Frydrych, L. M., Bian, G., Fattahi, F., Morris, S. B., O'Rourke, R. W., Lumeng, C. N., et al. (2019). GM-CSF Administration Improves Defects in Innate Immunity and Sepsis Survival in Obese Diabetic Mice. *J. I.* 202, 931–942. doi:10.4049/jimmunol.1800713
- Gangele, K., Gulati, K., Joshi, N., Kumar, D., and Poluri, K. M. (2020). Molecular Insights into the Differential Structure-Dynamics-Stability Features of Interleukin-8 Orthologs: Implications to Functional Specificity. *Int. J. Biol. Macromolecules* 164, 3221–3234. doi:10.1016/j.ijbiomac.2020.08.176
- Ge, J., Remesh, S. G., Hammel, M., Pan, S., Mahan, A. D., Wang, S., et al. (2019). Functional Relevance of Interleukin-1 Receptor Inter-domain Flexibility for Cytokine Binding and Signaling. *Structure* 27, 1296–1307. e1295 doi:10.1016/j.str.2019.05.011
- Günther, S., Deredge, D., Bowers, A. L., Luchini, A., Bonsor, D. A., Beadenkopf, R., et al. (2017). IL-1 Family Cytokines Use Distinct Molecular Mechanisms to Signal through Their Shared Co-receptor. *Immunity* 47, 510–523. e514 doi:10.1016/j.immuni.2017.08.004
- Hamilton, J. A. (2019). GM-CSF-Dependent Inflammatory Pathways. *Front. Immunol.* 10, 2055. doi:10.3389/fimmu.2019.02055
- Han, C., Lin, Y., Shan, G., Zhang, Z., Sun, X., Wang, Z., et al. (2010). Plasma Concentration of Malaria Parasite-Derived Macrophage Migration Inhibitory Factor in Uncomplicated Malaria Patients Correlates with Parasitemia and Disease Severity. *Clin. Vaccin. Immunol.* 17, 1524–1532. doi:10.1128/CVI.00149-10
- Hansen, G., Hercus, T. R., McClure, B. J., Stomski, F. C., Dottore, M., Powell, J., et al. (2008). The Structure of the GM-CSF Receptor Complex Reveals a Distinct Mode of Cytokine Receptor Activation. *Cell* 134, 496–507. doi:10.1016/j.cell.2008.05.053
- Herring, C. A., Singer, C. M., Ermakova, E. A., Khairutdinov, B. I., Zuev, Y. F., Jacobs, D. J., et al. (2015). Dynamics and Thermodynamic Properties of CXCL7 Chemokine. *Proteins* 83, 1987–2007. doi:10.1002/prot.24913
- Hou, J., Townson, S. A., Kovalchin, J. T., Masci, A., Kiner, O., Shu, Y., et al. (2013). Design of a superior Cytokine Antagonist for Topical Ophthalmic Use. *Proc. Natl. Acad. Sci.* 110, 3913–3918. doi:10.1073/pnas.1217996110
- Hudson, J. D., Shoaibi, M. A., Maestro, R., Carnero, A., Hannon, G. J., and Beach, D. H. (1999). A Proinflammatory Cytokine Inhibits P53 Tumor Suppressor Activity. *J. Exp. Med.* 190, 1375–1382. doi:10.1084/jem.190.10.1375
- Jiang, T., Zhou, C., and Ren, S. (2016). Role of IL-2 in Cancer Immunotherapy. *Oncimmunology* 5, e1163462. doi:10.1080/2162402X.2016.1163462
- Joseph, P. R. B., Mosier, P. D., Desai, U. R., and Rajarathnam, K. (2015). Solution NMR Characterization of Chemokine CXCL8/IL-8 Monomer and Dimer Binding to Glycosaminoglycans: Structural Plasticity Mediates Differential Binding Interactions. *Biochem. J.* 472, 121–133. doi:10.1042/BJ20150059
- Kämpfer, A. A. M., Urbán, P., Gioria, S., Kanase, N., Stone, V., and Kinsner-Ovaskainen, A. (2017). Development of an *In Vitro* Co-culture Model to Mimic the Human Intestine in Healthy and Diseased State. *Toxicol. Vitro* 45, 31–43. doi:10.1016/j.tiv.2017.08.011
- Kim, H. R. P. M., Park, M. K., Cho, M. L., Yoon, C. H., Lee, S. H., Park, S. H., et al. (2007). Macrophage Migration Inhibitory Factor Upregulates Angiogenic Factors and Correlates with Clinical Measures in Rheumatoid Arthritis. *J. Rheumatol.* 34, 927–936.
- Kleemann, R., Kapurniotu, A., Frank, R. W., Gessner, A., Mischke, R., Flieger, O., et al. (1998). Disulfide Analysis Reveals a Role for Macrophage Migration Inhibitory Factor (MIF) as Thiol-Protein Oxidoreductase. *J. Mol. Biol.* 280, 85–102. doi:10.1006/jmbi.1998.1864
- Kumar, A., Abdelmalak, B., Inoue, Y., and Culver, D. A. (2018). Pulmonary Alveolar Proteinosis in Adults: Pathophysiology and Clinical Approach. *Lancet Respir. Med.* 6, 554–565. doi:10.1016/s2213-2600(18)30043-2
- Kumar, S., Hanning, C. R., Brigham-Burke, M. R., Rieman, D. J., Lehr, R., Khandekar, S., et al. (2002). Interleukin-1f7b (Il-1h4/Il-1f7) Is Processed By Caspase-1 And Mature Il-1f7b Binds To The Il-18 Receptor But Does Not Induce Ifn-γ Production. *Cytokine* 18, 61–71. doi:10.1006/cyto.2002.0873
- Künze, G., Gehrcke, J.-P., Pisabarro, M. T., and Huster, D. (2014). NMR Characterization of the Binding Properties and Conformation of Glycosaminoglycans Interacting with Interleukin-10. *Glycobiology* 24, 1036–1049. doi:10.1093/glycob/cwu069
- Künze, G., Köhling, S., Vogel, A., Rademann, J., and Huster, D. (2016). Identification of the Glycosaminoglycan Binding Site of Interleukin-10 by NMR Spectroscopy. *J. Biol. Chem.* 291, 3100–3113. doi:10.1074/jbc.M115.681759
- Lai, K. N., Leung, J. C., Metz, C. N., Lai, F. M., Bucala, R., and Lan, H. Y. (2003). Role for Macrophage Migration Inhibitory Factor in Acute Respiratory Distress Syndrome. *J. Pathol.* 199, 496–508. doi:10.1002/path.1291
- Lisi, G. P., and Loria, J. P. (2016). Solution NMR Spectroscopy for the Study of Enzyme Allostery. *Chem. Rev.* 116, 6323–6369. doi:10.1021/acs.chemrev.5b00541
- Llamas-Covarrubias, M. A., Valle, Y., Navarro-Hernández, R. E., Guzmán-Guzmán, I. P., Ramírez-Dueñas, M. G., Rangel-Villalobos, H., et al. (2012). Serum Levels of Macrophage Migration Inhibitory Factor Are Associated with Rheumatoid Arthritis Course. *Rheumatol. Int.* 32, 2307–2311. doi:10.1007/s00296-011-1951-6
- Lotfi, N., Thome, R., Rezaei, N., Zhang, G.-X., Rezaei, A., Rostami, A., et al. (2019). Roles of GM-CSF in the Pathogenesis of Autoimmune Diseases: An Update. *Front. Immunol.* 10, 1265. doi:10.3389/fimmu.2019.01265
- Lubetsky, J. B., Swope, M., Dealwis, C., Blake, P., and Lolis, E. (1999). Pro-1 of Macrophage Migration Inhibitory Factor Functions as a Catalytic Base in the Phenylpyruvate Tautomerase Activity. *Biochemistry* 38, 7346–7354. doi:10.1021/bi990306m
- Mani, M., Chen, C., Ambler, V., Liu, H., Mathur, T., Zwicke, G., et al. (2015). MoonProt: a Database for Proteins that Are Known to Moonlight. *Nucleic Acids Res.* 43, D277–D282. doi:10.1093/nar/gku954

- McGeachy, M. J., Cua, D. J., and Gaffen, S. L. (2019). The IL-17 Family of Cytokines in Health and Disease. *Immunity* 50, 892–906. doi:10.1016/j.immuni.2019.03.021
- Mehta, H. M., Malandra, M., and Corey, S. J. (2015). G-CSF and GM-CSF in Neutropenia. *J. Immunol.* 195, 1341–1349. doi:10.4049/jimmunol.1500861
- Milovanovic, J., Arsenijevic, A., Stojanovic, B., Kanjevac, T., Arsenijevic, D., Radosavljevic, G., et al. (2020). Interleukin-17 in Chronic Inflammatory Neurological Diseases. *Front. Immunol.* 11, 947. doi:10.3389/fimmu.2020.00947
- Mitchell, R. A., Liao, H., Chesney, J., Fingerle-Rowson, G., Baugh, J., David, J., et al. (2002). Macrophage Migration Inhibitory Factor (MIF) Sustains Macrophage Proinflammatory Function by Inhibiting P53: Regulatory Role in the Innate Immune Response. *Proc. Natl. Acad. Sci.* 99, 345–350. doi:10.1073/pnas.012511599
- Moyat, M., Bouzourene, H., Ouyang, W., Iovanna, J., Renaud, J.-C., and Velin, D. (2017). IL-22-induced Antimicrobial Peptides Are Key Determinants of Mucosal Vaccine-Induced protection against *H. pylori* in Mice. *Mucosal Immunol.* 10, 271–281. doi:10.1038/mi.2016.38
- Mühlhahn, P., Czisch, M., Georgescu, J., Renner, C., Ross, A., Holak, T. A., et al. (1996). NMR Characterization of Structure, Backbone Dynamics, and Glutathione Binding of the Human Macrophage Migration Inhibitory Factor (MIF). *Protein Sci.* 5, 2095–2103. doi:10.1002/pro.5560051016
- Mulcahy, M. E., Leech, J. M., Renaud, J.-C., Mills, K. H., and McLoughlin, R. M. (2016). Interleukin-22 Regulates Antimicrobial Peptide Expression and Keratinocyte Differentiation to Control *Staphylococcus aureus* Colonization of the Nasal Mucosa. *Mucosal Immunol.* 9, 1429–1441. doi:10.1038/mi.2016.24
- Noe, J. T., and Mitchell, R. A. (2020). MIF-dependent Control of Tumor Immunity. *Front. Immunol.* 11, 609948 doi:10.3389/fimmu.2020.609948
- O'Connor Jr, W., Jr., Kamanaka, M., Booth, C. J., Town, T., Nakae, S., Iwakura, Y., et al. (2009). A Protective Function for Interleukin 17A in T Cell-Mediated Intestinal Inflammation. *Nat. Immunol.* 10, 603–609. doi:10.1038/ni.1736
- Oda, S., Oda, T., Nishi, K., Takabuchi, S., Wakamatsu, T., Tanaka, T., et al. (2008). Macrophage Migration Inhibitory Factor Activates Hypoxia-Inducible Factor in a P53-dependent Manner. *PLoS One* 3, e2215. doi:10.1371/journal.pone.0002215
- Onodera, S., Tanji, H., Suzuki, K., Kaneda, K., Mizue, Y., Sagawa, A., et al. (1999). High Expression of Macrophage Migration Inhibitory Factor in the Synovial Tissues of Rheumatoid Joints. *Cytokine* 11, 163–167. doi:10.1006/cyto.1998.0402
- Pantouris, G., Ho, J., Shah, D., Syed, M. A., Leng, L., Bhandari, V., et al. (2018). Nanosecond Dynamics Regulate the MIF-Induced Activity of CD74. *Angew. Chem. Int. Ed.* 57, 7116–7119. doi:10.1002/anie.201803191
- Pantouris, G., Khurana, L., Ma, A., Skeens, E., Reiss, K., Batista, V. S., et al. (2020). Regulation of MIF Enzymatic Activity by an Allosteric Site at the Central Solvent Channel. *Cel Chem. Biol.* 27, 740–750. e745 doi:10.1016/j.chembiol.2020.05.001
- Pantouris, G., Syed, M. A., Fan, C., Rajasekaran, D., Cho, T. Y., Rosenberg, E. M., et al. (2015). An Analysis of MIF Structural Features that Control Functional Activation of CD74. *Chem. Biol.* 22, 1197–1205. doi:10.1016/j.chembiol.2015.08.006
- Park, S. H., Berkamp, S., Radoicic, J., De Angelis, A. A., and Opella, S. J. (2017). Interaction of Monomeric Interleukin-8 with CXCR1 Mapped by Proton-Detected Fast MAS Solid-State NMR. *Biophysical J.* 113, 2695–2705. doi:10.1016/j.bpj.2017.09.041
- Parkins, A., Skeens, E., McCallum, C. M., Lisi, G. P., and Pantouris, G. (2021). The N-Terminus of MIF Regulates the Dynamic Profile of Residues Involved in CD74 Activation. *Biophysical J.* 120, 3893–3900. doi:10.1016/j.bpj.2021.08.025
- Penk, A., Baumann, L., Huster, D., and Samsonov, S. A. (2019). NMR and Molecular Modeling Reveal Specificity of the Interactions between CXCL14 and Glycosaminoglycans. *Glycobiology* 29, 715–725. doi:10.1093/glycob/cwz047
- Righetti, R. F., Santos, T. M. d., Camargo, L. d. N., Aristóteles, L. R. C. R. B., Fukuzaki, S., Souza, F. C. R. d., et al. (2018). Protective Effects of Anti-IL17 on Acute Lung Injury Induced by LPS in Mice. *Front. Pharmacol.* 9, 1021. doi:10.3389/fphar.2018.01021
- Roan, F., Obata-Ninomiya, K., and Ziegler, S. F. (2019). Epithelial Cell-Derived Cytokines: More Than Just Signaling the Alarm. *J. Clin. Invest.* 129, 1441–1451. doi:10.1172/JCI124606
- Roy, U. (2020). Insight into the Structures of Interleukin-18 Systems. *Comput. Biol. Chem.* 88, 107353. doi:10.1016/j.compbiolchem.2020.107353
- Rozwarski, D. A., Diederichs, K., Hecht, R., Boone, T., and Karplus, P. A. (1996). Refined crystal Structure and Mutagenesis of Human Granulocyte-Macrophage colony-stimulating Factor. *Proteins* 26, 304–313. doi:10.1002/(sici)1097-0134(199611)26:3<304:Aid-prot6>3.0.Co;2-d
- Schinagl, A., Kerschbaumer, R. J., Sabarth, N., Douillard, P., Scholz, P., Voelkel, D., et al. (2018). Role of the Cysteine 81 Residue of Macrophage Migration Inhibitory Factor as a Molecular Redox Switch. *Biochemistry* 57, 1523–1532. doi:10.1021/acs.biochem.7b01156
- Schluns, K. S., and Lefrançois, L. (2003). Cytokine Control of Memory T-Cell Development and Survival. *Nat. Rev. Immunol.* 3, 269–279. doi:10.1038/nri1052
- Shiomi, A., Usui, T., and Mimori, T. (2016). GM-CSF as a Therapeutic Target in Autoimmune Diseases. *Inflamm. Regener* 36, 8. doi:10.1186/s41232-016-0014-5
- Subbannayya, T., Variar, P., Advani, J., Nair, B., Shankar, S., Gowda, H., et al. (2016). An Integrated Signal Transduction Network of Macrophage Migration Inhibitory Factor. *J. Cel Commun. Signal.* 10, 165–170. doi:10.1007/s12079-016-0326-x
- Sun, T., Holowka, T., Song, Y., Zierow, S., Leng, L., Chen, Y., et al. (2012). A Plasmodium-Encoded Cytokine Suppresses T-Cell Immunity during Malaria. *Proc. Natl. Acad. Sci.* 109, E2117–E2126. doi:10.1073/pnas.1206573109
- Tahvildari, M., and Dana, R. (2019). Low-Dose IL-2 Therapy in Transplantation, Autoimmunity, and Inflammatory Diseases. *J.I.* 203, 2749–2755. doi:10.4049/jimmunol.1900733
- Takahashi, K., Koga, K., Linge, H. M., Zhang, Y., Lin, X., Metz, C. N., et al. (2009). Macrophage CD74 Contributes to MIF-Induced Pulmonary Inflammation. *Respir. Res.* 10, 33. doi:10.1186/1465-9921-10-33
- Toldi, J., Nemeth, D., Hegyi, P., Molnar, Z., Solymar, M., Farkas, N., et al. (2021). Macrophage Migration Inhibitory Factor as a Diagnostic and Predictive Biomarker in Sepsis: Meta-Analysis of Clinical Trials. *Sci. Rep.* 11, 8051. doi:10.1038/s41598-021-87613-0
- Trivedi-Parmar, V., and Jorgensen, W. L. (2018). Advances and Insights for Small Molecule Inhibition of Macrophage Migration Inhibitory Factor. *J. Med. Chem.* 61, 8104–8119. doi:10.1021/acs.jmedchem.8b00589
- Ushach, I., and Zlotnik, A. (2016). Biological Role of Granulocyte Macrophage colony-stimulating Factor (GM-CSF) and Macrophage colony-stimulating Factor (M-CSF) on Cells of the Myeloid Lineage. *J. Leukoc. Biol.* 100, 481–489. doi:10.1189/jlb.3RU0316-144R
- Valeri, M., and Raffatellu, M. (2016). Cytokines IL-17 and IL-22 in the Host Response to Infection. *Pathog. Dis.* 74, ftw111 doi:10.1093/femspd/ftw111
- van de Veerdonk, F. L., Netea, M. G., Dinarello, C. A., and Joosten, L. A. B. (2011). Inflammasome Activation and IL-1 β and IL-18 Processing during Infection. *Trends Immunol.* 32, 110–116. doi:10.1016/j.it.2011.01.003
- van Nieuwenhuijze, A., Koenders, M., Roeleveld, D., Sleeman, M. A., van den Berg, W., and Wicks, I. P. (2013). GM-CSF as a Therapeutic Target in Inflammatory Diseases. *Mol. Immunol.* 56, 675–682. doi:10.1016/j.molimm.2013.05.002
- Wang, Y., Chen, Y., Wang, C., Yang, M., Wang, Y., Bao, L., Wang, J. E., Kim, B., Chan, K. Y., Xu, W., Capota, E., Ortega, J., Nijhawan, D., Li, G. M., Luo, W., and Wang, Y. (2021). MIF is a 3' flap nuclease that facilitates DNA replication and promotes tumor growth. *Nat Commun* 12, 2954. doi:10.1038/s41467-021-23264-z
- Waters, L. C., Veverka, V., Strong, S. L., Muskett, F. W., Dedi, N., Lawson, A. D. G., Prosser, C. E., Taylor, R. J., Henry, A. J., and Carr, M. D. (2021). Conformational dynamics in interleukin 17A and 17F functional complexes

- is a key determinant of receptor A affinity and specificity. *Cytokine* 142, 155476 doi:10.1016/j.cyto.2021.155476
- Wettrich, A., Sebollela, A., Carvalho, M. A., Azevedo, S. P., Borojevic, R., Ferreira, S. T., and Coelho-Sampaio, T. (1999). Acidic pH modulates the interaction between human granulocyte-macrophage colony-stimulating factor and glycosaminoglycans. *Journal of Biological Chemistry* 274, 31468–31475. doi:10.1074/jbc.274.44.31468
- Wozniak, K. L., Hole, C. R., Yano, J., Fidel, P. L., and Wormley, F. L. (2014). Characterization of IL-22 and antimicrobial peptide production in mice protected against pulmonary *Cryptococcus neoformans* infection. *Microbiology (Reading)* 160, 1440–1452. doi:10.1099/mic.0.073445-0
- Yang, C.-Y. (2020). Comparative Analyses of the Conformational Dynamics Between the Soluble and Membrane-Bound Cytokine Receptors. *Sci Rep* 10, 7399. doi:10.1038/s41598-020-64034-z
- Zenobia, C., and Hajishengallis, G. (2015). Basic biology and role of interleukin-17 in immunity and inflammation. *Periodontol* 2000 69, 142–159. doi:10.1111/prd.12083

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Cui and Lisi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Elucidating the Activation Mechanism of AMPK by Direct Pan-Activator PF-739

Elnaz Aledavood¹, Aria Gheeraert^{2,3}, Alessia Forte¹, Laurent Vuillon³, Ivan Rivalta^{2,4}, F. Javier Luque^{1,5} and Carolina Estarellas^{1*}

¹Department of Nutrition, Food Science and Gastronomy, Faculty of Pharmacy and Food Sciences, and Institute of Theoretical and Computational Chemistry (IQTCUB), University of Barcelona, Barcelona, Spain, ²Dipartimento di Chimica Industriale "Toso Montanari" Università di Bologna, Bologna, Italy, ³LAMA, University of Savoie Mont Blanc, CNRS, LAMA, Le Bourget du Lac, France, ⁴Université de Lyon, École Normale Supérieure de Lyon, CNRS UMR 5182, Laboratoire de Chimie, Lyon, France, ⁵Institute of Biomedicine (IBUB), University of Barcelona, Barcelona, Spain

OPEN ACCESS

Edited by:

Yong Wang,
Zhejiang University, China

Reviewed by:

Jon Oakhill,
University of Melbourne, Australia
Zhaoxi Sun,
East China Normal University, China

*Correspondence:

Carolina Estarellas
cestarellas@ub.edu

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 17 August 2021

Accepted: 08 October 2021

Published: 05 November 2021

Citation:

Aledavood E, Gheeraert A, Forte A, Vuillon L, Rivalta I, Luque FJ and Estarellas C (2021) Elucidating the Activation Mechanism of AMPK by Direct Pan-Activator PF-739. *Front. Mol. Biosci.* 8:760026. doi: 10.3389/fmolb.2021.760026

Adenosine monophosphate-activated protein kinase (AMPK) is a key energy sensor regulating the cell metabolism in response to energy supply and demand. The evolutionary adaptation of AMPK to different tissues is accomplished through the expression of distinct isoforms that can form up to 12 heterotrimeric complexes, which exhibit notable differences in the sensitivity to direct activators. To comprehend the molecular factors of the activation mechanism of AMPK, we have assessed the changes in the structural and dynamical properties of β 1- and β 2-containing AMPK complexes formed upon binding to the pan-activator PF-739. The analysis revealed the molecular basis of the PF-739-mediated activation of AMPK and enabled us to identify distinctive features that may justify the slightly higher affinity towards the β 1-isoform, such as the β 1-Asn111 to β 2-Asp111 substitution, which seems to be critical for modulating the dynamical sensitivity of β 1- and β 2 isoforms. The results are valuable in the design of selective activators to improve the tissue specificity of therapeutic treatment.

Keywords: AMPK, protein dynamic, protein activation mechanism, pan-activator, isoform selectivity, molecular dynamics simulation

INTRODUCTION

AMP-activated protein kinase (AMPK) is a Ser/Thr protein kinase with a key role as a sensor in cellular energy homeostasis (Xiao et al., 2011). Upon activation, AMPK increases the levels of ATP, favoring the reduction of anabolic pathways and up-regulation of catabolic pathways. Due to its critical role in cell metabolism, AMPK is implicated in numerous metabolic disorders such as type 2 diabetes, cardiovascular diseases, and obesity (Carling, 2017). However, one of the most interesting aspects of this enzyme comes from the different tissue distribution that is directly related to its structural complexity. AMPK is a heterotrimeric complex consisting of a catalytic α -subunit and two regulatory subunits, namely β and γ . Each subunit can be found in different isoforms, involving two for α (α 1, α 2), two for β (β 1, β 2), and three for γ (γ 1, γ 2, γ 3) (Calabrese et al., 2014). The N-terminus of the α catalytic subunit contains a kinase domain, while its C-terminus is needed for the formation of the complex with the other subunits. The β -subunit has a central carbohydrate-binding module (CBM) that mediates AMPK interaction with glycogen, and the C-terminal region acts as a scaffold for the heterotrimeric assembly. Finally, the γ -subunit has four tandem repeats of the cystathionine

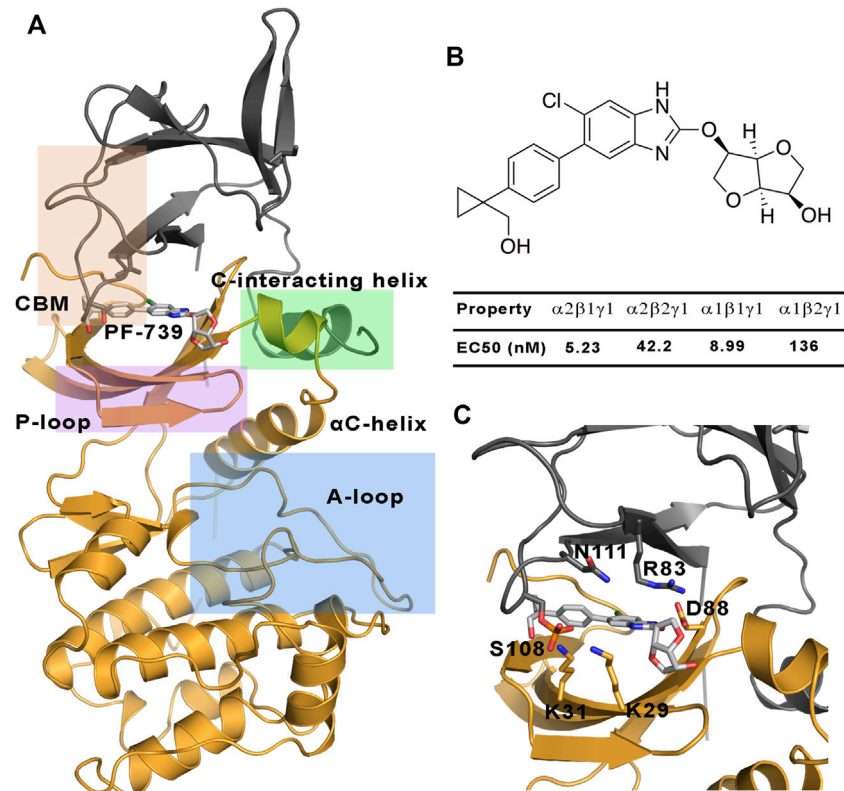


FIGURE 1 | (A) Representation of the system selected for the study of the direct activation mechanism, which is composed of $\alpha 2$ (yellow cartoon) and $\beta 1$ -/ $\beta 2$ -subunits (grey cartoons). The most important regions of these subunits are highlighted: P-loop (purple), activation loop (A-loop, cyan), CBM domain (orange) and C-interacting helix (green). **(B)** Chemical structure of PF-739, together with the experimental data obtained in Cokorinos et al. (2017) for the half maximal effective concentration (EC₅₀). **(C)** The activator bound to the ADaM site, with selected residues of the ADaM site shown as sticks.

β -synthase (CBS) domain, forming up to four potential nucleotide binding sites although only sites 1, 3 and 4 can really bind them (Scott et al., 2004; Scott et al., 2008; Carling et al., 2012).

AMPK is finely regulated by different mechanisms (Mahlapuu et al., 2004). An allosteric activation involves the phosphorylation of $\alpha 2$ -Thr172 in the activation loop of the kinase domain by upstream kinases such as LKB1 and CaMKKb, together with the binding of AMP to the CBS domain in the γ -subunit. The active AMPK complex can thus respond to subtle fluctuations in the AMP/ATP ratio, it being several thousand-fold more active (Carling et al., 2012; Chen et al., 2012; Willows et al., 2017). On the other side, AMPK can also be indirectly activated by compounds such as metformin, phenformin and oligomycin (Vazquez-Martin et al., 2012), which are able to increase the intracellular levels of AMP. However, much interest is focused on the understanding of the direct activation mechanism of AMPK by small organic molecules. The first reported direct activator was the thienopyridone drug A-769662 (Cool et al., 2006), which is bound to a cavity located at the interface between the CBM domain of the β -subunit and the kinase domain of the α -subunit, namely the allosteric drug and metabolite (ADaM) site (Langendorf and Kemp, 2015). One of the main features of the direct activation is that this kind of activation is

independent of the Thr172 phosphorylation, while it is enhanced by phosphorylation of Ser108 in the CBM domain of the β -subunit, increasing the AMPK activity by >90-fold (Hardie, 2014). Since then, a lot of efforts have been invested in obtaining direct AMPK activators, which in some cases exhibit a marked isoform selectivity (Olivier et al., 2018), while in other cases no significant selectivity is observed towards specific subunit isoforms. The isoform selectivity is relevant for the tissue distribution of the AMPK complexes. While $\alpha 1$, $\beta 1$ and $\gamma 1$ have low tissue specificity, $\alpha 2$ is basically found in the heart and skeletal muscle, $\beta 2$ in the skeletal muscle and $\gamma 2$ is mainly found in the heart muscle, and $\gamma 3$ is found in the skeletal muscle (Uhlén et al., 2015; Human Protein Atlas (2021, 2021)). The tissue specificity is related to the specific function of AMPK in these tissues, and therefore all the isoforms in the skeletal muscle have an important role in the glucose uptake, making AMPK a promising target for diabetes type 2 disease. In the last years an increasing effort has been devoted to design tissue-specific direct AMPK activators. As an example, the SC4 small-molecule, which was designed to increase the selectivity towards the α -subunit (being more selective for the $\alpha 2$ -isoform) (Ngoei et al., 2018), can activate both $\beta 1$ - and $\beta 2$ -containing AMPK complexes, although a slightly higher activation is observed for the $\beta 1$ -isoform. Other interesting examples are the pan-activators

PF-739, which is able to activate both $\alpha 2\beta 1\gamma 1$ and $\alpha 2\beta 2\gamma 1$ (**Figure 1**), and MK-8722 which can activate the 12 heterotrimeric AMPK complexes (Myers et al., 2017). Regarding the selectivity of β -isoform, although the half maximal effective concentration (EC_{50}) determined for PF-739 and the binding affinity measurements for MK-8722 shows that they still exhibit a larger affinity for the $\beta 1$ -containing isoforms, they are the most potent activators of $\beta 2$ complexes reported up to date (Cokorinos et al., 2017). However, it is still necessary to achieve a higher specificity to avoid off-tissue target effects. Accordingly, understanding of the molecular factors that favor the binding to specific isoforms is an outstanding issue.

In our previous works (Aledavood et al., 2019; Aledavood et al., 2021), we have studied the molecular factors that determine the selective activation of $\beta 1$ - and $\beta 2$ -containing AMPK complexes formed with A-769662 and SC4. We have hypothesized that the change of $\beta 1$ -Asn111 by $\beta 2$ -Asp111 could be a key factor in mediating the distinctive “mechanical” sensitivity of AMPK complexes to these activators. Here, we extend this analysis to the pan-activator PF-739 with the aim to examine how the binding of this compound affects the dynamical response of AMPK considering the trends disclosed for A-769662 and SC4. At this point, it is worth noting that while A-769662 is selective for $\beta 1$ -containing complexes, SC4 exhibits a mild preference for this isoform, a trend which was attributed to the presence of the carboxylate group present in the chemical structure of this activator. In contrast, PF-739 is a neutral compound, which suggests that other chemical features might also regulate the mild preference for binding to $\beta 1$ -containing AMPK complexes. Understanding the role of the factors that regulate the mechanical response of AMPK could thus be valuable for the tailored design of isoform-adapted pharmacophores useful in the search of selective direct activators. With this aim in mind, we have carried out extensive molecular dynamic simulations (MD) and network analysis to examine the differential trends in structural, dynamical and interaction patterns emerging for AMPK complexes with PF-739.

RESULTS AND DISCUSSION

MD simulations were run to assess the structural and dynamics properties of the AMPK complexes formed by the $\alpha 2$ -isoform bound to either $\beta 1$ - or $\beta 2$ - subunits. The neglect of the γ subunit in the simulated systems obeys two main motivations. First, following a divide-and-conquer strategy, this permits to focus the conformational sampling of the activator-induced changes on the ADaM site, which is shaped by residues in α and β subunits. Second, the adoption of these systems permits a direct comparison with the results obtained previously for the complexes formed with A-769662 and SC4 (Human Protein Atlas (2021, 2021; Ngoei et al., 2018). Accordingly, this study is focused on the conformational ensemble collected for the apo species of $\alpha 2\beta 1$ and $\alpha 2\beta 2$ systems, the corresponding complexes formed with PF-739 (holo species), and finally the complexes formed with both PF-739 and ATP molecule (holo+ATP), the latter being located in the ATP-binding site within the kinase

domain of the α -subunit. For each system (apo, holo, and holo+ATP), the analysis involves the conformational ensemble explored in three independent replicas (1 μ s/replica), leading to a total simulation time of 6 μ s for the apo species and 12 μ s for the ligand-bound complexes.

Structural Analysis of AMPK Complexes

We have examined the effect of PF-739 binding to the ADaM site (holo structures), and the simultaneous presence of PF-739 and ATP in both ADaM and ATP-binding sites (holo+ATP structures) on the global structural conformation of apo $\alpha 2\beta 1$ and $\alpha 2\beta 2$ by means of the root mean square deviation (RMSD) of the protein backbone along the corresponding 1 μ s simulations (**Figure 2**). The RMSD was determined using the average structure of the holo+ATP species sampled in the last 200 ns of the three independent replicas run for either $\alpha 2\beta 1$ or $\alpha 2\beta 2$ species as reference. For the holo+ATP systems there is a high structural resemblance for all the replicas, as noted in the small fluctuations of the RMSD profiles (**Figure 2C**), which agrees with the preservation of the overall protein fold upon binding of both PF-739 and ATP. In particular, the RMSD values for the holo+ATP species range from 2.0 to 2.5 Å for $\alpha 2\beta 1$ and from 2.7 to 3.0 Å for $\alpha 2\beta 2$ (**Table 1**). These values are lower than the RMSD values obtained for the apo species ($\alpha 2\beta 1$: 2.5–2.7 Å; $\alpha 2\beta 2$: 2.9–3.4 Å).

Binding of PF-739 to the $\alpha 2\beta 1$ species has no significant effect on the RMSD of the holo species (from 2.5 to 2.9 Å), which is close to the values obtained for the apo form. Only the presence of both the ligand and ATP (holo+ATP) gives rise to a reduction in the RMSD. This effect is even more remarkable in the $\alpha 2\beta 2$ species, as the RMSD of the protein backbone is generally larger than the RMSD value determined for the $\alpha 2\beta 1$ complex in all the states (apo, holo and holo+ATP; see **Figure 2** and **Table 2**). These findings suggest that PF-739 exerts a weak structural stabilization upon binding to both $\alpha 2\beta 1$ and $\alpha 2\beta 2$ species.

Regarding the per-residue mean square fluctuation (RMSF) profile, similar results are observed for both $\alpha 2\beta 1$ and $\alpha 2\beta 2$ species, as noted in the resemblance of the fluctuation patterns obtained by averaging the RMSF of the three replicas run for every system (**Figure 3**). The highest fluctuations in the α -subunit correspond to residues in the activation loop (residues 165–185, highlighted in blue in **Figure 3**) and the α -helix formed by residues 210–230. It is worth noting the higher fluctuation of the P-loop (residues 15–35; purple in **Figure 3**) in the holo state in comparison to both apo and holo+ATP systems. Thus, binding of PF-739 significantly affects the flexibility of the P-loop in both $\alpha 2\beta 1$ and $\alpha 2\beta 2$ species, which may have functional relevance since the P-loop contributes to shape both the ADaM and ATP-binding sites. Regarding the β -subunit, the largest fluctuations are in the CBM domain, which contains Ser108 (highlighted in orange in **Figure 3**; phosphorylated in both holo and holo+ATP states), and the regions near the C-interacting helix (residues 162–172, highlighted in green in **Figure 3**). It is worth noting that the binding of PF-739 (holo) and ATP (holo+ATP) increases the fluctuations of the α -subunit elements mentioned above, while reduces the fluctuations in the β -subunit, independently of the β -isoform. These findings are in

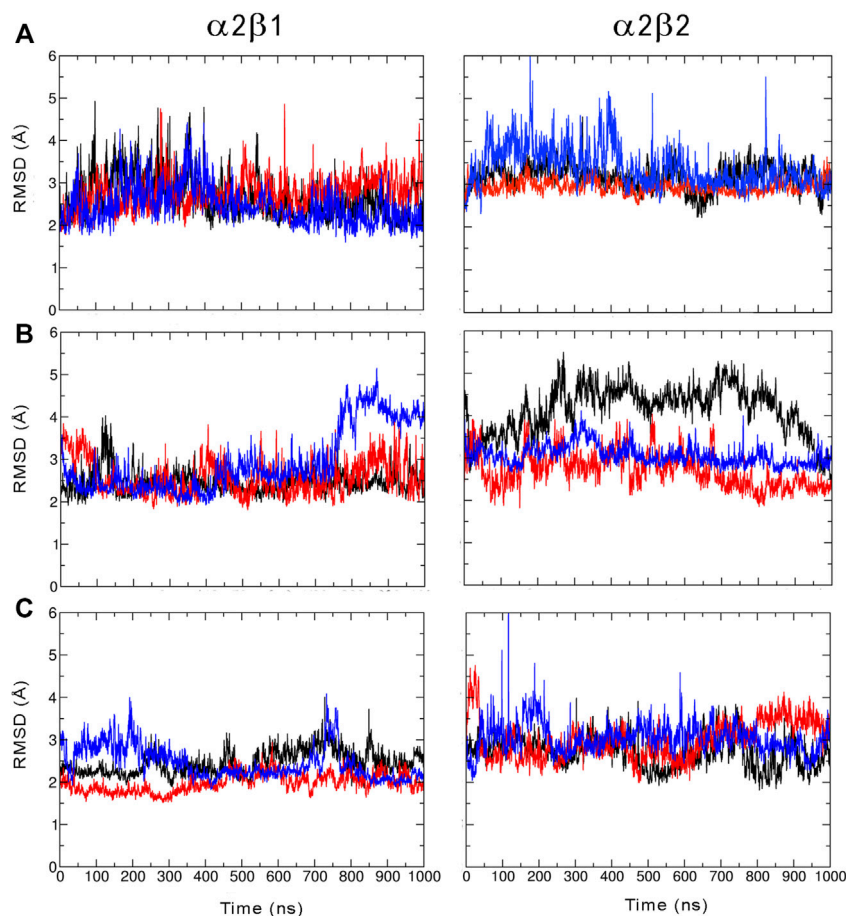


FIGURE 2 | Root mean squared deviation (RMSD, Å) determined for the protein backbone along the three 1 μ s MD simulations run for the (A) apo, (B) holo and (C) holo+ATP species of AMPK isoforms $\alpha 2\beta 1$ and $\alpha 2\beta 2$ bound to PF-739 (each replica is shown in black, blue and red, respectively). For each analysis the reference structure used corresponds to the energy-minimized average structure of the holo+ATP sampled in the last 200 ns of the three independent MD simulations.

TABLE 1 | RMSD and standard deviation (Å) determined for the protein backbone of the snapshots sampled along the last 500 ns of MD simulations performed for all systems (apo, holo and holo+ATP states) of AMPK isoforms $\alpha 2\beta 1$ and $\alpha 2\beta 2$. Values were determined using the energy-minimized holo+ATP species averaged for the last 200 ns of each simulation system as reference structure.

System	Replica 1	Replica 2	Replica 3	Average
$\alpha 2\beta 1$ Apo	2.6 \pm 0.6	2.7 \pm 0.4	2.5 \pm 0.5	2.6
$\alpha 2\beta 1$ Holo	2.5 \pm 0.3	2.6 \pm 0.5	2.9 \pm 0.8	2.6
$\alpha 2\beta 1$ holo+ATP	2.5 \pm 0.3	2.0 \pm 0.2	2.4 \pm 0.4	2.3
$\alpha 2\beta 2$ Apo	3.2 \pm 0.3	2.9 \pm 0.2	3.4 \pm 0.5	3.2
$\alpha 2\beta 2$ Holo	4.1 \pm 0.6	2.7 \pm 0.4	3.1 \pm 0.3	3.3
$\alpha 2\beta 2$ holo+ATP	2.7 \pm 0.4	3.0 \pm 0.5	3.0 \pm 0.4	2.9

agreement with the higher RMSD fluctuations observed in some replicas of the holo states for both $\alpha 2\beta 1$ and $\alpha 2\beta 2$ species.

Dynamic Properties of AMPK Complexes

In order to examine the effect of the activator on the conformational behavior of AMPK complexes, we have

TABLE 2 | Contribution of the essential motion (%) to the structural variance of different AMPK systems and the total contribution of the first four projections.

Systems	Proj. 1	Proj. 2	Proj. 3	Proj. 4	Total _(P1-P4)
$\alpha 2\beta 1$ apo	41.2	12.0	8.1	4.6	66.0
$\alpha 2\beta 1$ holo	38.6	12.1	7.6	4.0	62.3
$\alpha 2\beta 1$ holo+ATP	30.7	12.6	7.0	5.1	55.4
$\alpha 2\beta 2$ apo	30.9	12.6	8.8	4.8	57.1
$\alpha 2\beta 2$ holo	33.1	12.7	8.6	5.3	59.7
$\alpha 2\beta 2$ holo+ATP	29.0	13.2	7.2	5.3	54.7

analyzed both the essential dynamics (ED) of the protein backbone and the dynamic correlation between residues.

The ED provides information about the essential motions of the protein and can be used to examine the effect of activator on the major motions of the protein skeleton. The results for the first essential motion for the apo ($\alpha 2\beta 1$ and $\alpha 2\beta 2$) states show a concerted bending that brings α - and β -subunits closer and then moves them apart (Figure 4). The most interesting feature is that the P-loop seems to act as a hinge, assisting the concerted bending

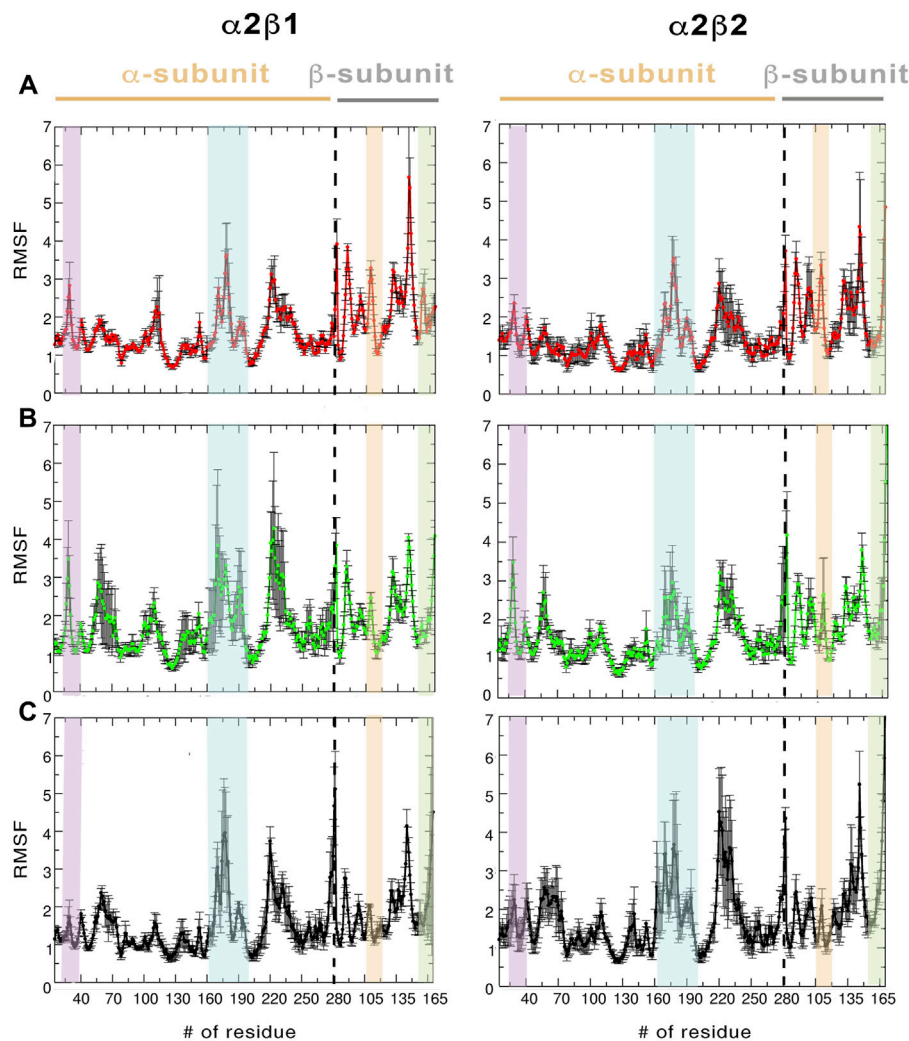


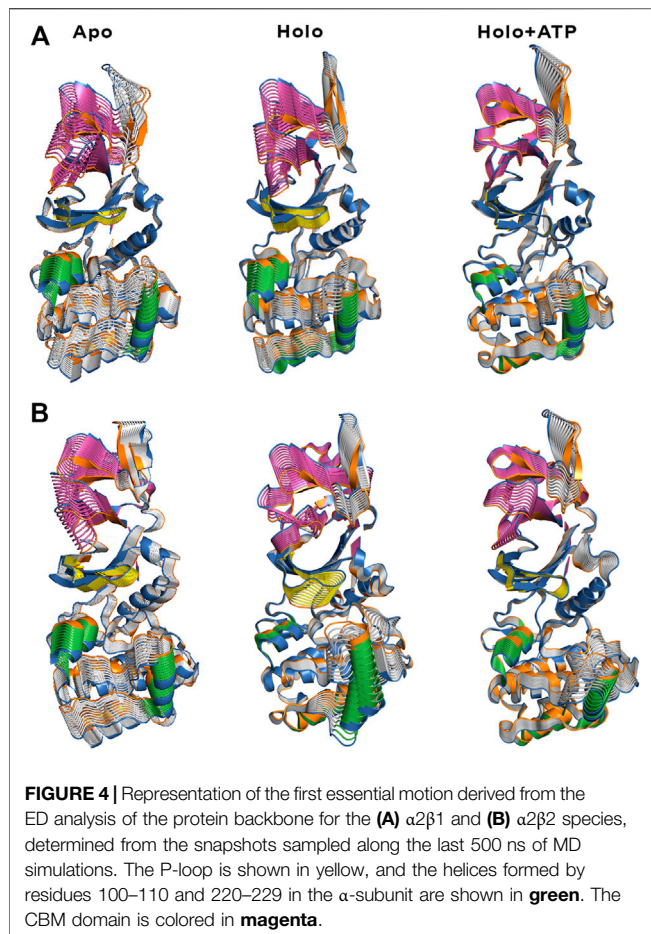
FIGURE 3 | RMSF (Å) average of the residues determined for the protein backbone along the last 500 ns of the three independent replicas runs for (A) apo, (B) holo and (C) holo+ATP. The standard deviation for each residue is shown as an error bar. The highlighted regions denote the moieties corresponding to P-loop (purple), activation loop (cyan), CBM domain (orange), and C-interacting helix (green).

between the subunits. Indeed, the first motion accounts on average for 41/31% of the structural variance in $\alpha 2\beta 1/\alpha 2\beta 2$ species, and the contribution of the first four motions accounts for 66/57% of the total structural variance (Table 2). This emphasizes the importance of the first essential motion to the conformational flexibility of the AMPK complexes.

Comparison of the ED results obtained for apo, holo and holo+ATP states reveals that binding of the activator has a mild effect on the conformational variance, which is reduced from 66% (apo) to 62% (holo) and 55% (holo+ATP) for the $\alpha 2\beta 1$ species (Table 2). However, for the $\alpha 2\beta 2$ species the activator triggers a slight increase in the conformational variance relative to the apo species, while subsequent binding of ATP results in a reduction of the structural variance (apo: 57.1%; holo: 59.7%; holo+ATP: 54.7%). These results are also reflected in the contribution of the first essential motion (Figure 4 and Table 2). In the holo ($\alpha 2\beta 1$ and $\alpha 2\beta 2$) states, this motion reflects a synchronous

motion of the P-loop and the CBM domain, which is in contrast with the increased stiffness observed in the holo+ATP state, especially regarding the P-loop, the helical domain in the α -subunit, as well as the region of the CBM domain nearest to the ADaM site. However, although the movements of the CBM domain are very similar between $\alpha 2\beta 1$ and $\alpha 2\beta 2$ species, the P-loop and the helices at the C-terminal region of the α -subunit exhibit higher fluctuations in $\alpha 2\beta 2$ with respect to $\alpha 2\beta 1$ (Supplementary Figure S1). Finally, it is worth noting that the enhanced stiffness achieved upon ATP binding to holo is again more remarkable in the case of the $\alpha 2\beta 1$ complex (Figure 4 and Table 2).

Besides the qualitative inspection of the overall dynamics of the systems shown in Figure 4, we have determined the similarity indices for the first essential motions (Supplementary Table S1). The similarity index for the apo species (i.e., the most flexible one) is close to 0.70 and 0.60 for $\alpha 2\beta 1$ and $\alpha 2\beta 2$, respectively, reflecting



the preservation of the major deformation of the protein skeleton in the three replicas. These results also agree with the higher conformational flexibility observed for $\alpha 2\beta 2$ systems. In the holo species, the similarity indices are 0.75 for $\alpha 2\beta 1$ and 0.47 for $\alpha 2\beta 2$ systems. These results agree with the essential motion observed for the holo state of $\alpha 2\beta 1$ (Figure 4A, middle panel) and $\alpha 2\beta 2$ (Figure 4B, middle panel). In the former, the variance of the system is more balanced between certain regions, i.e., CBM domain, P-loop, A-loop and helices P220-G229 and E100-R110 (colored in green, Figure 4). However, higher fluctuations account for the structural elements in the α -subunit in $\alpha 2\beta 2$. These findings are in agreement with the previous RMSD and RMSF results. Finally, for the holo+ATP systems the similarity index is close to 0.35 for $\alpha 2\beta 1$ and $\alpha 2\beta 2$, respectively. However, this simply means that binding of both activator and ATP rigidifies the protein skeleton, annihilating the large-scale deformations observed in the apo species as observed in Figure 4. The ED, shown in Figure 4, as well as the similarity indexes calculated, in Supplementary Table S1, have been obtained considering the last 500 ns of the simulation time of the three replicas. However, in order to check the statistical value of our simulations, we have also calculated the similarity indexes for the first three essential motions of the apo $\alpha 2\beta 1$ and $\alpha 2\beta 2$ derived from the ED analysis in time windows 200–600 and

600–1,000 ns for the three replicas (Supplementary Table S2). The similarity index amounts in general to 0.8. For the first replica of $\alpha 2\beta 1$ system a lower similarity is observed, suggesting a slower structural relaxation, as noted in the similarity obtained for more advanced time windows (Supplementary Table S3). Overall, these results suggest that selection of the last 500 ns to perform the statistical analysis of the simulations is well suited for the comparison between replicas, although these results also suggest that shorter time periods might be also usable. For this reason this 500 ns time window has been used in further analysis.

To complement the results of ED analysis, we have performed two additional analyses with the aim to assess the dynamic correlation between residues and disclose specific relationships between the α - and β -subunits: a dynamical perturbation network (DPN, Figure 5) and a dynamic cross-correlation (DCC, Figure 6; see Methods and Materials for technical details) analysis.

The dynamical perturbation network (DPN) was calculated for apo and holo species as an average of the three independent replicas. Figure 5 shows the changes in the correlation of residues between apo and holo states, where blue/red edges stand for contacts weakened/strengthened in holo relative to apo state. Thus, these networks provide information of how the interaction of the activator with the enzyme affects the contact network between residues. For the sake of comparison, this analysis was performed not only for PF-739, but also for A-769662, which exhibits a marked selectivity for $\beta 1$ -containing AMPK complexes. Our previous studies (Human Protein Atlas (2021, 2021; Ngoei et al., 2018) revealed that A-769662 acts as molecular glue between the $\alpha 2$ - and $\beta 1$ -subunits, while this effect is lost in the $\alpha 2\beta 2$ species due to the higher dynamical resilience of this specie towards the activator. The dynamical contact network for A-769662 (Figure 5A) perfectly agrees with these findings. In fact, the changes between apo and holo in $\alpha 2\beta 1$ mainly reveal a higher number of contacts between the P-loop of the $\alpha 2$ -subunit and the CBM of the $\beta 1$ -subunit as well as between the α C-helix of the $\alpha 2$ -subunit and the C-interacting helix of the $\beta 1$ -subunit. Conversely, the contact network that emerges for the $\alpha 2\beta 2$ complex is more complex, involving regions located far from the ADaM site. This result agrees with the higher flexibility of the $\alpha 2\beta 2$ species, and the lower impact of A-769662 on the dynamical response of this complex.

For the pan-activator PF-739, the $\alpha 2\beta 1$ complex exhibits fewer and more specific contacts, which primarily affect the CBM/P-loop and the α C-helix/C-interacting helix/A-loop, than the $\alpha 2\beta 2$ species, thus resembling the results discussed for A-769662. However, the number of contacts weakened or even lost between the A-loop and the α C-helix in the $\alpha 2\beta 1$ holo state is remarkably higher for PF-739-bound complexes compared to A-769662-bound ones (Figures 5A,B, left side). For $\alpha 2\beta 2$, the number and weights of the edges are larger in this species, and the distribution of contacts involves wider regions from the CBM domain to the A-loop (see highlighted region in cyan, Figure 5B, right panel). Noteworthy, DPN analysis reveals that binding of A-769662 gives rise to a much larger difference in the dynamical network of $\alpha 2\beta 1$ /A-769662 and $\alpha 2\beta 2$ /A-769662 complexes than for $\alpha 2\beta 1$ /PF-739 and $\alpha 2\beta 2$ /PF-739 complexes, as the pattern

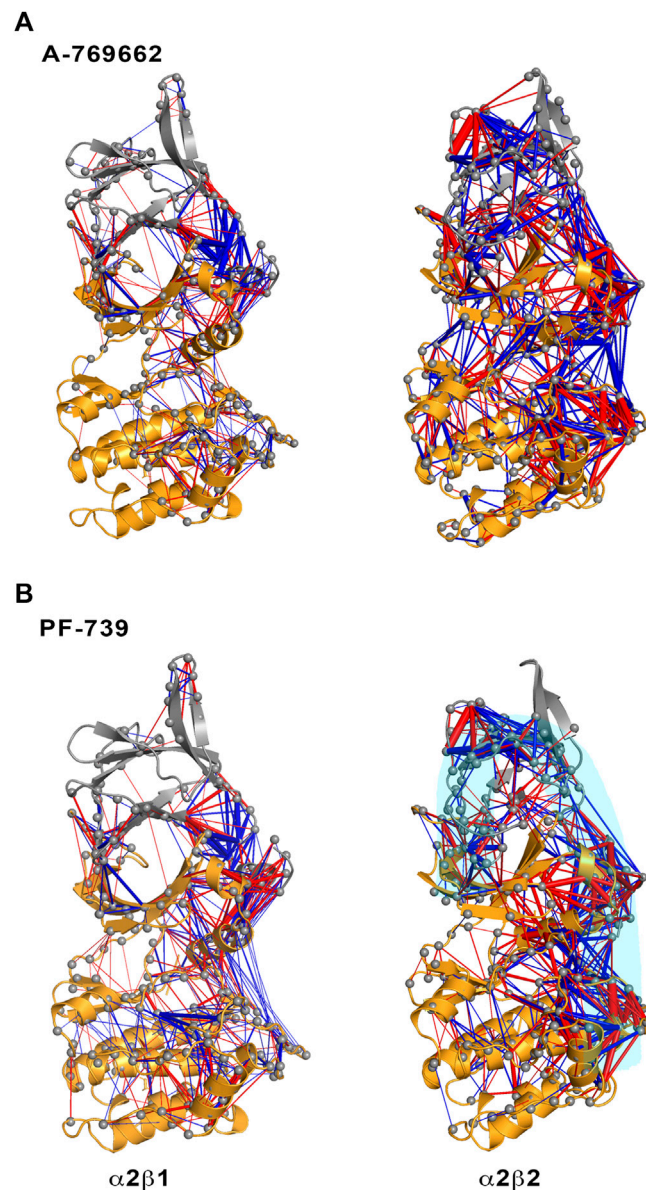


FIGURE 5 | Representation of the contact changes observed in the dynamical perturbation network between apo and holo states for (left) $\alpha 2\beta 1$ and (right) $\alpha 2\beta 2$ species. The holo states were calculated in presence of (A) A-769662 and (B) PF-739 activators. Red/blue edges show interresidue atomic contact increase/decrease in the holo state relative to the apo form. The magnitude of these changes is indicated by the width of the edges.

observed for the last two AMPK complexes exhibit a similar pattern (Figure 5). This is in agreement with the experimental results that indicate that the PF-739 is active against both $\beta 1$ - and $\beta 2$ -containing isoforms, in contrast with the selective activation of $\beta 1$ -containing AMPK complexes reported for A-769662.

Finally, the dynamic cross-correlation (DCC) analysis was performed to examine the correlated motions of residues in $\alpha 2\beta 1$ and $\alpha 2\beta 2$ AMPK complexes. For the apo systems (Figure 6A) one may notice a significant correlation between residues in the P-loop and the α C-helix, both in the α -subunit, and between the α C-helix from the α -subunit and the C-interacting helix from the β -subunit (as noted by the yellow marks). It is worth noting that there is a slight

correlation between the P-loop and the CBM domain (β -subunit), more remarkable in $\alpha 2\beta 1$ than in $\alpha 2\beta 2$, as noted by the similarity indexes of 0.82 for $\alpha 2\beta 1$, which is reduced to 0.75 in $\alpha 2\beta 2$ (Supplementary Table S4). The holo+ATP systems show lower dynamical correlation between residues, as observed by the progressive reduction in the number and intensity of the areas that exhibit a pronounced correlation (shown in yellow and blue for highly correlated and anticorrelated fluctuations between residues, respectively). On the contrary, the correlation between the motion of the P-loop and the CBM domain is reinforced in the holo and holo+ATP states (black square in Figure 6). These effects are more noticeable for the comparison of holo in $\alpha 2\beta 1$ (similarity indices of 0.63 in $\alpha 2\beta 1$ vs 0.55 in

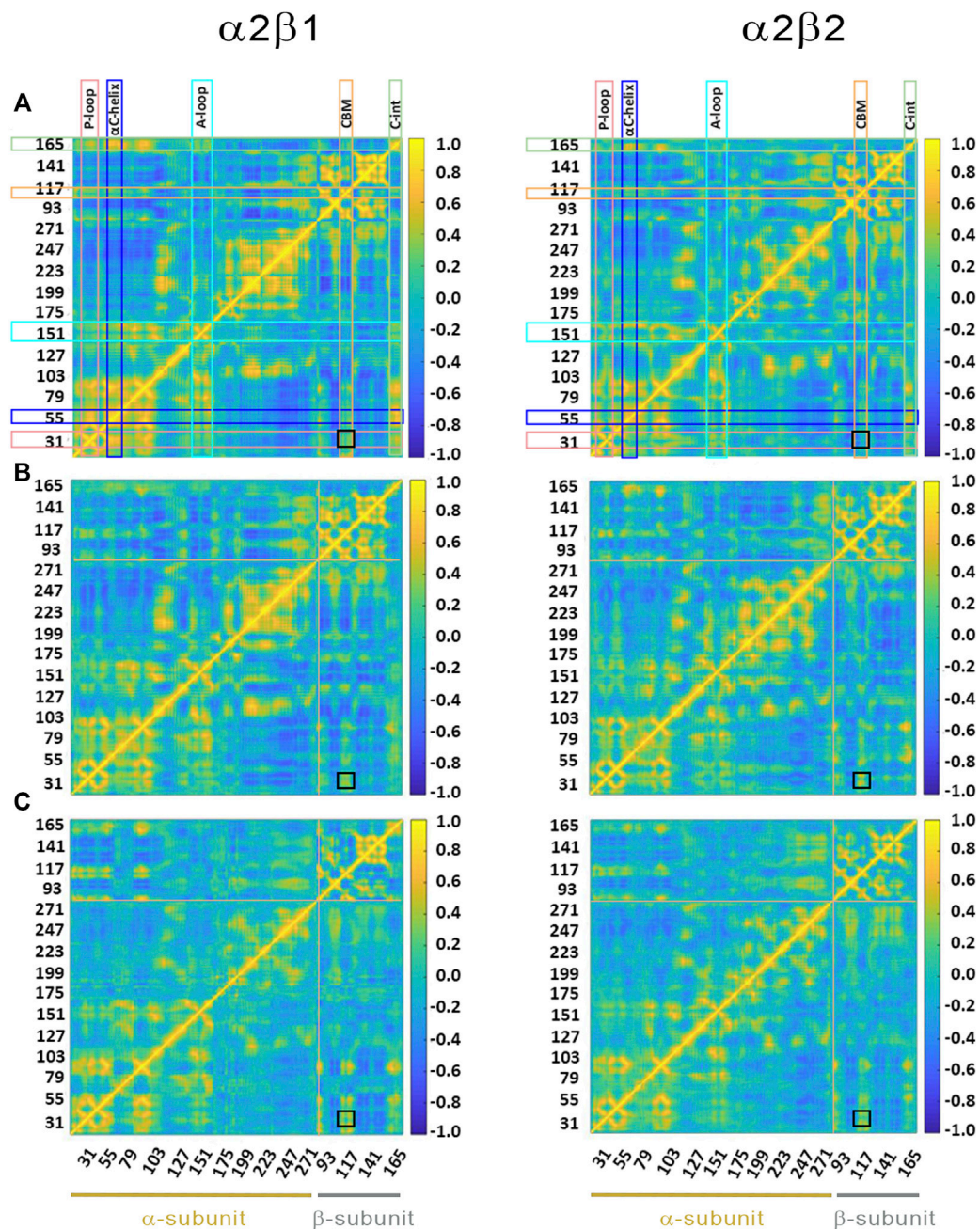


FIGURE 6 | Dynamic cross-correlation (DCC) matrices for (A) apo, (B) holo and (C) holo+ATP complexes of (left) $\alpha 2\beta 1$ and (right) $\alpha 2\beta 2$ with PF-739. The x- and y-axis denote the numbering of residue in the α - and β -subunit, indicated at the bottom of the plot (yellow and gray lines for α - and β -subunits, respectively). Specific regions of AMPK are highlighted with squares at the top of the plot: P-loop (pink), α C-helix (blue), activation loop (A-loop; cyan), CBM unit (orange) and C-interacting helix (C-int; green). Regions colored in yellow/blue show high correlated/anticorrelated fluctuations. The black square in the plots highlights the motion between the CBM domain and the P-loop.

$\alpha 2\beta 2$, Supplementary Table S4), while lower differences exist for holo+ATP systems in $\alpha 2\beta 1$ and $\alpha 2\beta 2$, in agreement with previous analyses.

Although the preceding results show a high similarity in the dynamical behavior of both $\alpha 2\beta 1$ and $\alpha 2\beta 2$ species bound to PF-739 activator, which agrees with the definition of PF-739 as a pan-activator, these analyses still reveal subtle differences between $\beta 1$ -

and $\beta 2$ -containing AMPK complexes. In particular, the results suggest that the $\alpha 2\beta 2$ species have a larger resilience to the structural modulation exerted by the activator, whereas the $\alpha 2\beta 1$ isoform is more sensitive to the conformational adaptation induced upon activator binding to the ADaM site, enhancing the stiffness of protein backbone for the $\beta 1$ -containing complex (Figures 4, 5). These results agree with the fact that PF-

739, which can activate both $\alpha 2\beta 1\gamma 1$ and $\alpha 2\beta 2\gamma 1$ complexes, still exhibits a larger affinity for the $\beta 1$ -isoform (**Figure 1**) (Cokorinos et al., 2017).

Pre-Organization of ATP-Binding Site

To explore how PF-739 could influence the activation of AMPK, we have evaluated the dynamical response of the ATP-binding site due to the binding of the activator in the ADaM site. Specifically, we have assessed the pre-organization of the ATP-binding site in the apo, holo and holo+ATP states, using as a reference the average structure of the holo+ATP complex.

For the holo+ATP states, the residues of the ATP-binding site sample a conformational space with a high peak centered at a positional RMSD of 1.2 Å° and a shoulder at 1.9 Å° for $\alpha 2\beta 1$, while a wider distribution is observed with a peak centered at 1.8 Å° for $\alpha 2\beta 2$ (**Figure 7**, Gaussian distributions colored in yellow). Unexpectedly, the apo state shows a narrower distribution with a unique peak centered at 2.0 Å° for both $\alpha 2\beta 1$ and $\alpha 2\beta 2$ species. In fact, the conformations sampled by the apo state have a notable overlap with the distribution of holo+ATP, this resemblance being more significant for the $\alpha 2\beta 2$ species. In contrast, the holo state exhibits a wider distribution, showing a bimodal RMSD profile, with peak values at 1.7 and 3.2 Å° for $\alpha 2\beta 1$, and at 1.8 and 2.5 Å° for $\alpha 2\beta 2$. These results suggest that the binding of PF-739 enhances the fluctuations of P-loop residues that shape the ATP-binding. Due to this higher conformational flexibility, the ATP-binding site can adopt conformations close to those populated in the holo+ATP state, but also visit more dissimilar conformational regions even in comparison with the apo state (**Figure 7**).

Structural Basis of the AMPK Activation by Pan-Activator PF-739 and Its Comparison With Other Direct Activators

To complement the previous analyses, we have examined the interaction network formed by PF-739 and the residues in both α - and β -subunits. To this end, we have clustered the snapshots sampled along the last 500 ns simulation of each replica for both holo $\alpha 2\beta 1$ and $\alpha 2\beta 2$ species, summing a total of 1.5 μ s. The results for holo- $\alpha 2\beta 1$ system display up to 4 different clusters, which account for 67.5, 11.8, 10.5 and 10.2% of the conformational ensemble, where the main difference is the conformation adopted by the sugar-like mannitol ring appendage of PF-739 (**Figure 8A**). In all cases two regions can be identified in the interaction network. The first one corresponds to the salt bridge formed between $\beta 1$ -Arg83 and $\alpha 2$ -Asp88 (3.0 ± 0.3 Å), which at the same time is hydrogen-bonded to PF-739 (3.5 ± 0.6 Å). For the second cluster (11.8%), an additional interaction between $\beta 1$ -Arg83 and the sugar-like mannitol ring is observed (3.7 ± 0.6 Å; **Figure 8A**). The second region involves salt bridges between pSer108 located at the β -subunit CBM domain and $\alpha 2$ -Lys29 (3.7 ± 0.9 Å) and $\alpha 2$ -Lys31 (4.4 ± 1.3 Å), both from the P-loop of the α -subunit. Moreover, $\alpha 2$ -Lys29 and $\alpha 2$ -Lys31 establish contacts with PF-739, such as a hydrogen bond between the Lys31 and the hydroxymethyl-cyclopropyl group (3.2 ± 0.7 Å),

which is found in all clusters, and an additional interaction between Lys29 and the N of the benzimidazole ring (3.9 ± 0.9 Å, **Figure 8A**) present in clusters 2 and 3. These interactions networks are very similar to those found in our previous study of SC4 (Aledavood et al., 2021), suggesting that the structural differences between these two compounds, mainly regarding the *o*-toluic substitution of SC4 by mannitol-like ring appendage in PF-739, and the 4'-nitrogen of imidazopyridine in SC4 by a carbon atom in PF-739, do not have a dramatic effect over the interaction at the ADaM site (see also **Supplementary Table S5**). Indeed, these findings remark the key role of the $\beta 1$ -Arg83/ $\beta 2$ -Arg82 in the organization of these interactions networks as we explain below.

The cluster analysis performed for the holo- $\alpha 2\beta 2$ system yields four clusters that differ in the orientation of the sugar-like mannitol ring of PF-739, accounting for 76.7, 13.3, 6.0 and 4.0% of the structural ensemble (**Figure 8B**). However, these clusters show higher structural diversity than those determined for the holo- $\alpha 2\beta 1$ system. Thus, two distinct orientations of $\beta 2$ -Arg82 are found in all clusters (**Figure 9**). In one case (**Figure 9A**), $\beta 2$ -Arg82 interacts with $\alpha 2$ -Asp88 (3.9 ± 1.3 Å), which forms a hydrogen bond with PF-739 (2.9 ± 0.2 Å). This arrangement represents 54.4% of all the conformations sampled for the $\alpha 2\beta 2$ holo species. In the second orientation $\beta 2$ -Arg82 interacts with $\beta 2$ -Asp111 (3.8 ± 0.9 Å), accounting for 45.6% of the conformational ensemble (**Figure 9B**). Notably, in the $\alpha 2\beta 1$ holo species this latter interaction is not observed, which can be attributed to the substitution of $\beta 2$ -Asp111 by $\beta 1$ -Asn111. The second orientation found for $\beta 2$ -Arg82 reinforces the interaction network observed through β -pSer108, which maintains its interactions with both α -Lys29 (3.3 ± 0.6) and α -Lys31 (3.8 ± 1.0) from the P-loop. Additionally, the interaction between α -Lys31 and the hydroxymethyl-cyclopropyl group (3.2 ± 0.6 Å) of PF-739 is maintained in all clusters, while the interaction between Lys29 and the N of the benzimidazole ring is less stable and only slightly observed in cluster #3 (4.4 ± 0.7 Å, **Figure 8B**).

These results suggest that the arrangement of the sugar-like mannitol unit structural, which exhibit notable differences between clusters, does not have a significant impact on the interaction network observed along the simulations, since the main interactions are preserved in all cases. Indeed, the arrangement of the sugar-like mannitol ring gives rise to new interactions between $\beta 1$ -Arg83 ($\beta 2$ -Arg82) and PF-739 only in cluster #2 (11.8%) for $\alpha 2\beta 1$ and cluster #4 (4.0%) for $\alpha 2\beta 2$. Furthermore, the conformation of the $\beta 1$ -Arg83/ $\beta 2$ -Arg82 residue emerges as a key structural feature. While in the $\alpha 2\beta 1$ holo specie, $\beta 1$ -Arg83 forms a salt bridge with $\alpha 2$ -Asp88 in all sampled conformations, two orientations are found for $\beta 2$ -Arg82 in the $\alpha 2\beta 2$ holo species (**Figure 9**). This distinctive trait can be attributed to the substitution $\beta 1$ -Asn111 \rightarrow $\beta 2$ -Asp111, since the presence of $\beta 2$ -Asp111 in $\alpha 2\beta 2$ promotes an electrostatic competition with $\alpha 2$ -Asp88 for the interaction with $\beta 2$ -Arg82.

To confirm these results, we have calculated the major interaction pathways identified from WISP analysis for the holo species formed with PF-739. **Figure 10** show the WISP

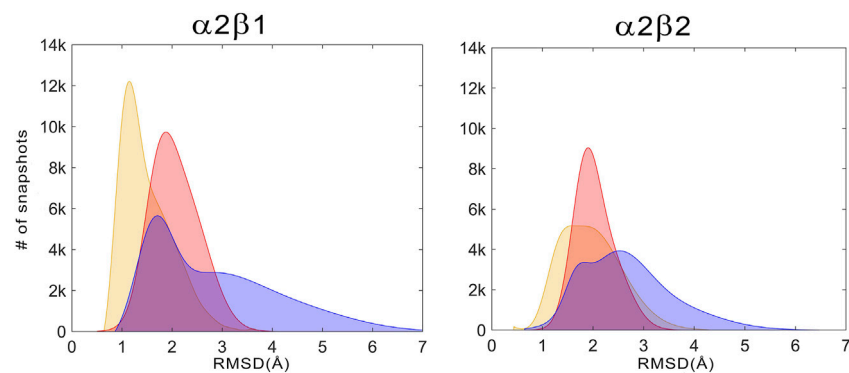


FIGURE 7 | Distribution of the positional deviation (RMSD; Å) of the structures sampled along the trajectories run for apo (red), holo (blue), and holo+ATP (yellow) for the residues that shape the ATP-binding site (residues $\alpha 22$ – $\alpha 32$, $\alpha 42$ – $\alpha 46$, $\alpha 75$ – $\alpha 79$, $\alpha 142$ – $\alpha 147$, and $\alpha 153$ – $\alpha 157$). A total of 60,000 snapshots taken from the last 500 ns of MD simulations were considered for each system in the analysis.

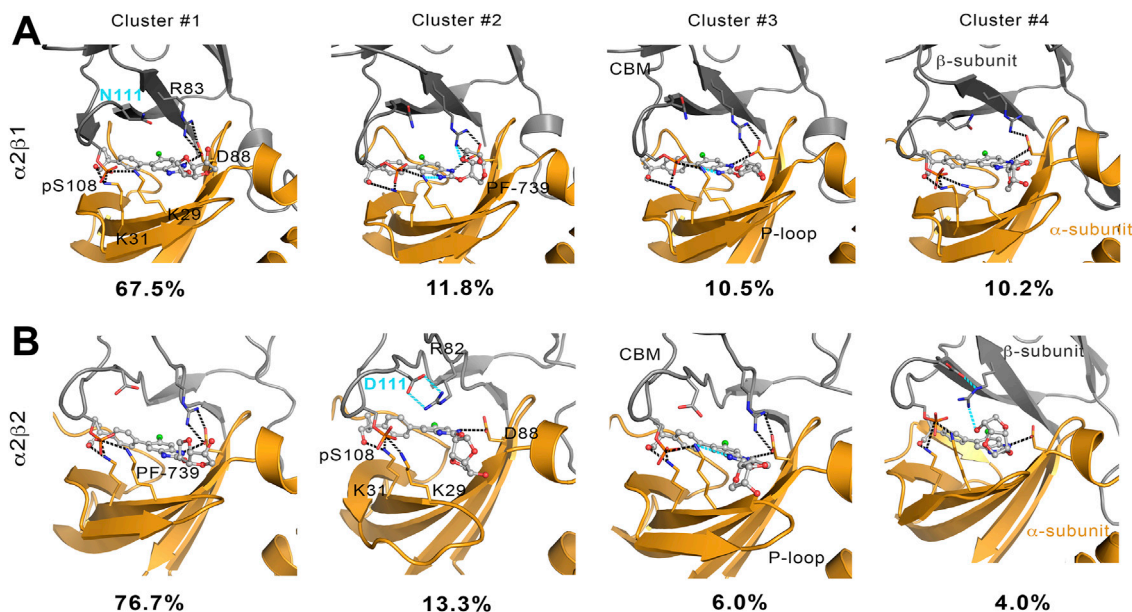


FIGURE 8 | Representation of main interactions between the CBM, P-loop and PF-739 for holo states of (A) $\alpha 2\beta 1$ and (B) $\alpha 2\beta 2$ species for the four clusters obtained along the last 500 ns of simulation of each replica. The α -subunit is shown in orange cartoon, while the β -subunit is shown in grey cartoons. PF-739 is shown in grey ball and sticks in the ADaM site. Selected polar interactions maintained through all the MD simulations and clusters are highlighted in black dashed lines, while those formed in specific clusters are shown in cyan.

results obtained in our previous work (Ngoei et al., 2018) for A-769662 (Figure 10A) and SC4 (Figure 10B), as well as the results obtained for PF-739 (Figure 10C). For the $\alpha 2\beta 1$ /A-769662 complex three major paths are found between the CBM domain and the P-loop, which involve i) pSer108, ii) the hydrophobic core of the ADaM site, and iii) the interaction $\beta 1$ -Arg83– $\alpha 2$ -Asp88. All of them are directly connected with the activator through the residues participating in the path, supporting the role of A-769662 as a molecular glue between $\alpha 2$ - and $\beta 1$ -subunits. However, only the pSer108 path is observed for the $\alpha 2\beta 2$ /A-769662 complex. This can be attributed to the $\beta 1$ -Asn111 \rightarrow $\beta 2$ -Asp111 substitution, weakens the interaction

between $\beta 2$ -Arg82 and $\alpha 2$ -Asp88, and strengthens the path through pSer108. In turn, this agrees with the selective activation observed for AMPK complexes containing the $\beta 1$ -isoform. In contrast, two representative paths are found in the holo states formed with SC4 (Figure 10B), corresponding to the networks through pSer108 and through the pair $\beta 1/2$ -Arg83– $\alpha 2$ -Asp88. Furthermore, SC4 exhibit a similar pattern in both $\alpha 2\beta 1$ and $\alpha 2\beta 2$, which is in agreement with the ability to activate both kinds of AMPK complexes (Hardie, 2014). Interestingly, the $\beta 1$ -Asn111 \rightarrow $\beta 2$ -Asp111 substitution seems to be less sensitive to the presence of SC4, an effect that can be attributed to the negative charge of the activator that can

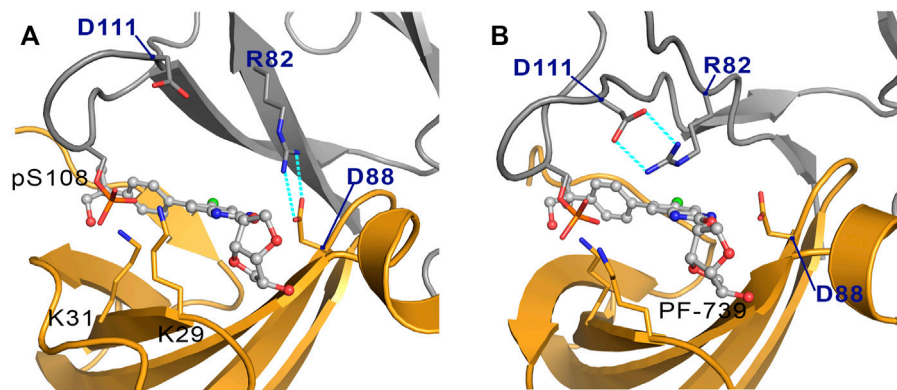


FIGURE 9 | Representation of the two orientations of the β2-Arg82 in α2β2 species, where the interaction with (A) the α2-Asp88 and (B) the β2-Asp111 are highlighted in cyan dashed lines. The α-subunit is shown in orange cartoon, while the β-subunit is shown in grey cartoons. PF-739 is shown in grey ball and sticks in the ADaM site.

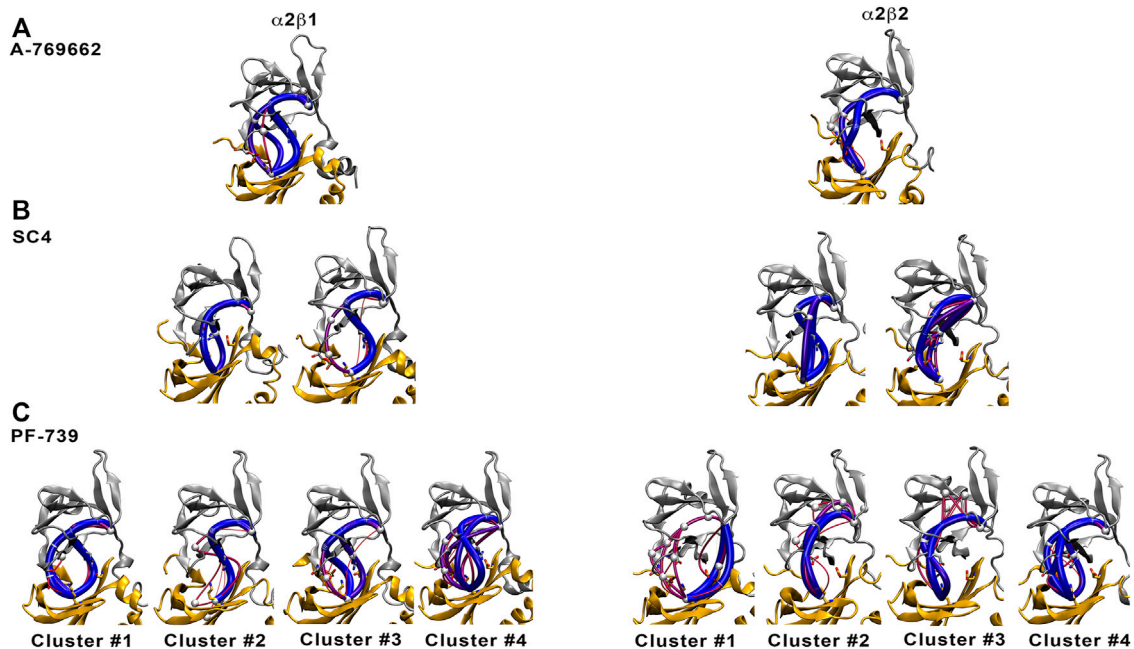


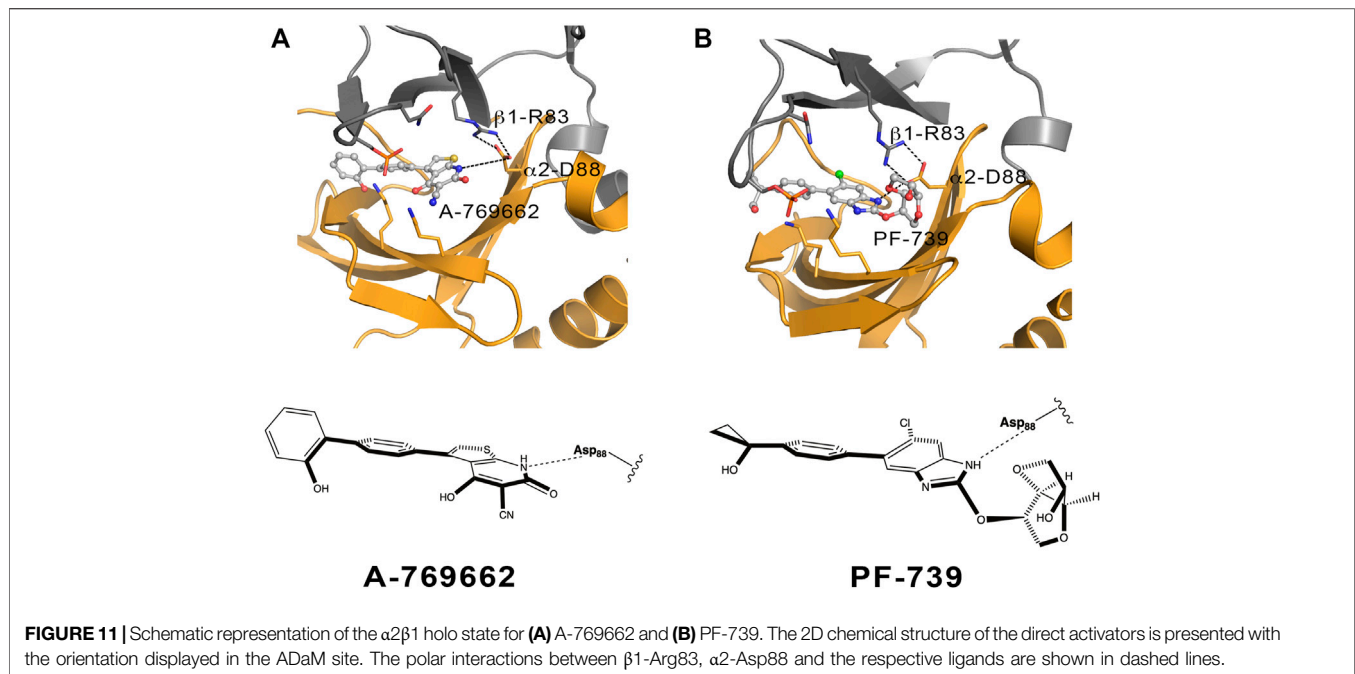
FIGURE 10 | Comparison of major interaction networks obtained from WISP analysis for α2β1 (left panel) and α2β2 (right panel) species of the holo states for (A) A-769662, (B) SC4 and (C) PF-739 direct activators.

modulate the linking role of β2-Arg82 towards a preferential interaction with either β2-Asp111 and α2-Asp88.

In light of these findings, we have performed the WISP analysis separately for the four main clusters obtained for PF-739 (Figure 10C). In the case of the holo-α2β1 state, the three pathways described above for A-769662 can be identified in the whole set of clusters. Although one may notice distinct traits for each cluster, at least two main paths can be observed for clusters #1, #3 and #4. In particular, for the most populated cluster (#1; 67.5%) they correspond to the paths mediated by pSer108 and the pair β1-Arg83–α2-Asp88, respectively. However, the analysis of

the holo-α2β2 state reveals a weaker connectivity, since a single path dominates the interaction network in all clusters. For the most populated cluster #1 (76.7%), the path involves the β2-Arg82–α2-Asp88 pair, with a minor contribution of the pSer108-mediated path. In the other clusters, nevertheless, the pSer108 path is predominant, resembling the behavior found for A-769662 (Figure 10A, right panel).

These results suggest that the β1-Asn111 → β2-Asp111 substitution plays a critical role in defining the mechanical sensitivity of AMPK to the direct activator. Besides the pSer108-mediated path, the presence of β1-Asn111 in α2β1



favors the formation of an additional path that involves the concerted interaction between $\beta 1$ -Arg83, $\alpha 2$ -Asp88, activator and $\alpha 2$ -Lys29/ $\alpha 2$ -Lys31. Nevertheless, the substitution $\beta 1$ -Asn111 \rightarrow $\beta 2$ -Asp111 favors the breaking of the $\beta 2$ -Arg82– $\alpha 2$ -Asp88 interaction and the formation of the salt bridge with $\beta 2$ -Asp111, which reinforces the contribution of the pSer108 path, making the $\alpha 2\beta 2$ complex less sensitive to the modulation by the activator.

The chemical features of the activator also exerts role in assisting the conformational activation of both $\alpha 2\beta 1$ and $\alpha 2\beta 2$ species. The main difference between A-769662 and PF-739 is the replacement of the thienopyridone ring by a benzimidazole derivative with a sugar-like mannitol appendage in PF-739 (Figure 11). The $\beta 1$ -Arg83– $\alpha 2$ -Asp88–A-769662– $\alpha 2$ -Lys29/ $\alpha 2$ -Lys31 network of interactions acts as a transmission band that connect the dynamical motion of the CBM domain with the P-loop, assisting the effective transition toward conformations that resemble the ATP-binding site in the holo+ATP state for the $\alpha 2\beta 1$ species (Supplementary Figure S2, left). However, breakage of this interaction path in the $\alpha 2\beta 2$ holo species prevents the activator to mediate the transmission of the dynamical fluctuations of the CBM domain and the P-loop, which is reflected in a wider conformational distribution of the ATP-binding site (peak centered at 3.0 Å; see Supplementary Figure S2, top). This reflects the inability of A-769662 to pre-organize the ATP-binding site in $\beta 2$ -containing AMPK complexes.

The conformational response caused by PF-739 is more complex, reflecting the structural variability of the clusters regarding the orientation of the sugar-like mannitol appendage for both $\alpha 2\beta 1$ and $\alpha 2\beta 2$ species, and the two arrangements of $\beta 2$ -Arg82 in $\alpha 2\beta 2$ compared to the single conformation of $\beta 1$ -Arg83 in $\alpha 2\beta 1$. The analysis of the pre-organization of ATP-binding site (Figure 7) reveals that the activator is unable to reduce the

conformational sampling to structures well suited for the binding of ATP, which would diminish the activation effect of PF-739. At this point let us remark the bimodal behavior shown in Figure 7A, with only 33.3/45.0% of the sampled structures of ATP-binding site resembling the holo+ATP in $\alpha 2\beta 1/\alpha 2\beta 2$, whereas A-769662 triggers a marked shift in the population distribution in the holo complex of $\alpha 2\beta 1$ (Supplementary Figure S2). On the one side, this agrees with the ability of PF-739 to exert a mild activation in both $\alpha 2\beta 1$ and $\alpha 2\beta 2$. The distribution of holo+ATP-like conformations in $\alpha 2\beta 2$ is wider than in $\alpha 2\beta 1$, which reflects the higher structural plasticity observed in $\alpha 2\beta 2$ species. On the other side, these findings are also in agreement with the WISP results, which show how PF-739 activator has higher gluing effect than A-769662 in $\alpha 2\beta 2$, allowing the transmission of the information between α - and β -subunit through the pSer108 and $\beta 2$ -Arg82– $\alpha 2$ -Asp88 pathways, explaining in this way why PF-739 acts as a pan-activator.

CONCLUSION

Discerning the molecular factors that regulates the structure-function relationships of AMPK isoforms is of utmost importance to rationalize the tissue-dependent expression of AMPK complexes, and thus enabling the design of specific compounds active against specific metabolic disorders. However, the recognition of the differences between isoforms that allow a different ligand behavior (i.e., selective activator, pan-activator or even inhibitor) is very challenging due to the high structural complexity of the enzyme and the highly correlated dynamics observed for both $\alpha 2\beta 1$ and $\alpha 2\beta 2$ species.

Our results confirmed that the subtle difference of $\beta 1$ -Asn111 to $\beta 2$ -Asp111 has great implications in the dynamical response of

AMPK to the binding of activators. This single substitution can change the interaction networks formed surrounded the activator, thus inducing a better mechanical response of the $\alpha 2\beta 1$ specie towards the interaction of PF-739, than in the case of the $\alpha 2\beta 2$ species. So, even in case of a pan-activator like the PF-739, able to activate both β -isoforms, still subtle residue substitutions in the ADaM site are responsible of difference in affinity towards the isoform. Additionally, we hypothesized that the bulkier substitutions in the chemical structure of the ligands located nearest to the $\alpha 2$ -Asp88 residue could involve a higher variability in the conformational space, thus preventing to discern between β -isoforms.

In summary, we were able to characterize the key molecular features that mediate the activation of pan-activator towards $\alpha 2\beta 1$ and $\alpha 2\beta 2$ species. All these findings shed light in the comprehension of the role of specific residues in the ADaM site that can modulate or completely change the direct activation mechanism of $\beta 1$ - and $\beta 2$ -containing AMPK complexes. Future studies will be appreciated to distinguish the structural basis of the different sensitivity of AMPK complexes formed by distinct α -subunits, and which is more important, the study of the full complex to disentangle the full allosteric network connection. This understanding will really enable us the design of tissue-selective modulators of this cellular energy sensor.

MATERIALS AND METHODS

Molecular Dynamics Simulations

Extended molecular dynamics (MD) simulations were utilized to analyze the structural and dynamical characteristics of the simulated system. For this purpose, the $\alpha 2\beta 1\gamma 1$ systems were built up using the complexes with A-769662 (PDB entry 4CFF) (Xiao et al., 2013). On the other hand, the system related to the complex of $\alpha 2\beta 2\gamma 1$ bound to SC4 (PDB entry 6B2E) (Ngoei et al., 2018) was also used as a template to model the complexes with PF-739. Following our previous studies, (Aledavood et al., 2019; Aledavood et al., 2021), the γ -subunit was not considered in MD simulations for several reasons. First, the ADaM site is shaped only by α - and β - isoforms. Furthermore, the lack of precise structural information about stretches of both α - and β -subunits, particularly regarding the C-terminal regions, which are located close to the γ -subunit, would introduce an additional level of uncertainty, opening the way to potential artefacts in the simulations. Finally, inclusion of the γ -subunit would have required a larger computational cost to guarantee a proper sampling of the dynamical motions of the three isoforms. Accordingly, following the “divide-and-conquer” strategy outlined above, the simulated systems comprise only α - and β -subunits. Specifically, simulations were performed for residues 8–278 of the $\alpha 2$ isoform, and residues 78–173 and 77–171 of the $\beta 1$ - and $\beta 2$ -isoforms, which were solved without disruptions in the X-ray structures. Finally, these structures were used to model the apo protein, the complexes of the activators bound to the phosphorylated Ser108 (pSer108)-containing isoforms (holo), and the corresponding holo+ATP complexes with both activator in the ADaM site and ATP in the ATP-binding site.

The Molecular dynamic (MD) simulations were performed using the AMBER18 package (Case et al., 2018) and the Amber ff99SBILDN force field (Lindorff-Larsen et al., 2010) for the protein, whereas the ligand (PF-739) were parameterized using the GAFF force field (Wang et al., 2004) in conjunction with restrained electrostatic potential-fitted (RESP) partial atomic charges derived from B3LYP/6-31G(d) calculations (Bayly et al., 1993). The parameters used for the ATP molecule were obtained from the Amber parameters database from Bryce group at the University of Manchester (AMBER parameter database, 2021; Meagher et al., 2003). The standard protonation state at physiological pH was assigned to ionisable residues, and a capping group (N-methyl) was added to the C-terminus of the α -subunit. The simulated systems were immersed in an octahedral box of TIP3P water molecules considering a solute-edge distance of 12 Å (Jorgensen et al., 1983), and counterions atoms were added to maintain the neutrality of the simulated systems (Joung and Cheatham, 2008). The final systems included the AMPK protein (368 residues for $\alpha 2\beta 1$ and 367 residues for $\alpha 2\beta 2$), around 25,000–26,700 water molecules, and a variable number of Na^+ and Cl^- ions, leading to simulated systems containing between 81,000 and 86,000 atoms (specific values are gathered in **Supplementary Table S6**).

Simulations were performed in the NPT ensemble for equilibration and NVT for MD productions using periodic boundary conditions and Ewald sums (grid spacing 1 Å) for treating long-range electrostatic interactions. Apo, holo and holo+ATP systems were simulated in triplicate. The minimization of the systems was performed refining the position of hydrogen atoms in the protein (2,000 cycles of steepest descent algorithm followed by 8,000 cycles of conjugate gradient), subsequently minimizing the position of water molecules (using again the previous scheme), and finally minimization of the whole system (4,000 cycles for steepest descent and 1,000 cycles of conjugate gradient). Later, the temperature of the system was gradually raised from 100 to 300 K in five steps, 50 ps each using the NVT ensemble and Langevin dynamics for the temperature regulation. In this process, suitable restraints ($5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) were imposed to keep the ligand (activator, ATP) in the binding pocket and prevent artefactual rearrangements along the equilibration stage. In order to equilibrate the density of the system an additional 5 ns step performed in the NPT ensemble using the Berendsen barostat. In addition, the restraints were progressively eliminated in this later step. Production MD simulations were run for 1 μs per replica, leading to a total simulation time of 12 μs for the ligand-bound AMPK complexes, and 6 μs for the two apo species of AMPK.

Essential Dynamics

This method was utilized to specify the most important motions from the structural variance sampled in MD simulations. In essential dynamics (ED) (Amadei et al., 1993), the dynamics along the individual modes can be studied and visualized separately, so we can filter the main collective motions during our simulations. Therefore, the positional covariance matrix is created and diagonalized in order to achieve the collective deformation modes, i.e., the eigenvectors, while the eigenvalues

account for the contribution of each motion to the structural variance of the protein. ED analysis was done for 25,000 snapshots from the last 500 ns of each simulation, taking into account only the backbone atoms and the calculations were performed with PCAsuite program (available at <http://www.mmb.irbbarcelona.org/software/pcasuite/pcasuite.html>), which is integrated in the pyPCCazip program, a suite of tools for compression and analysis of molecular simulations (Shkurti et al., 2016).

Dynamical Perturbation Network

Contact networks represent a protein as a collection of nodes, i.e., the residues that are connected by edges if those residues satisfy a contact condition. Here, in line with previous works (Vuillon and Lesieur, 2015; Dorantes-Gilardi et al., 2018; Gheeraert et al., 2019), the contact condition is satisfied if at least one heavy atom from a residue is at a distance below 5 Å from a heavy atom of another residue. Edges are then weighted by the total number of atomic couples that satisfy this contact condition. Individual contact networks from the frames of one MD simulation are built and averaged (considering the average total number of atomic contacts from various replicas) in order to create a dynamical weighted contact network, which represents a time-averaged contact network associated to the corresponding MD simulations.

To compare MD simulations of a protein in various states (i.e., apo, holo and holo+ATP complexes), we computed perturbation contact networks (Gheeraert et al., 2019) by subtracting two dynamical weighted contact networks associated to each pair of states. To differentiate increases and decreases in contact we assign colors to the edges of the dynamical perturbation network according to the sign of its edges. Finally, for visualization purposes a weight threshold can be applied so that only edges with a weight greater than the threshold are kept for visualization, here set to 5 as in previous work (Gheeraert et al., 2019). Nodes isolated after this process are also pruned to simplify the visualization.

Dynamic Cross-Correlation Analysis

To complement the information gained from the ED analysis, dynamic cross-correlation (DCC) was used to examine the correlation motion of residues along a given trajectory. To this end, all the snapshots were aligned by means of least-square fitting of Cα atoms of the whole protein to the equilibrated starting configuration. Then, the DCC matrix was determined as noted in Eq. 1.

$$C_{ij} = \frac{c_{ij}}{c_{ii}^{1/2} c_{jj}^{1/2}} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\left[(\langle r_i^2 \rangle \langle r_j^2 \rangle) (\langle r_i^2 \rangle - \langle r_i \rangle^2) (\langle r_j^2 \rangle - \langle r_j \rangle^2) \right]^{1/2}} \quad (1)$$

where the position vectors of two Cα atoms *i* and *j* fitted in the structure at time *t* are denoted as *r_i(t)* and *r_j(t)*, respectively.

The cross-correlation coefficients range from −1 to +1, which represent anticorrelated and correlated motions, respectively, whereas values close to zero indicate the absence of correlated motions (Hünenberger et al., 1995). This analysis was performed using the module available in AMBER package. The similarity between the DCC matrices computed for the three replicas run for apo, holo and holo+ATP systems was estimated using the

Tanimoto similarity index. This parameter is a distance metrics used to quantify the degree of similarity between two sets of data. While this index is widely adopted to compare the descriptors that characterize the chemical structure of molecules, in this study it is used to compare the correlated motions determined for pairs of residues in the AMPK complexes.

Cluster Analysis

Cluster analysis is a way of determining structure populations from MD simulations. Clustering results in a partitioning data so that data inside a cluster are more similar to each other than they are outside a cluster. In MD, this is a mean of grouping similar conformations together. Similarity is defined by a distance metric, the smaller the distance, the more similar the structures. We used coordinate RMSD as the distance metric parameter. Additionally, we used K-means algorithm as implemented in cpptraj software (Shao et al., 2007), to perform cluster analysis. The K-means identifies *k* number of centroids, and then allocates every data point to the nearest cluster, while maintaining the centroids as small as possible (Shao et al., 2007). We set the sieve parameter to 10 to reduce the expense of generating the pair-wise distance matrix by using “total/10” frames for initial clustering. The sieved frames are then added to the initial clusters. This analysis was done for 100,000 snapshots from the last 500 ns of each simulation, considering only the backbone atoms.

Interaction Energy Network

Networks of local interactions are intrinsically linked to the structural response of proteins to external factors (O’Rourke et al., 2016). For our purposes, Weighted Implementation of Suboptimal Path (WISP) (Van Wart et al., 2014) was utilized to analyze the allosteric network. This method enabled us to perform a dynamic network analysis to understand how the binding of a ligand in an allosteric cavity could affect another binding site. In particular, WISP relies on the dynamical interdependence among the protein residues. To this end, each amino acid is treated as a node, which was located at the residue center-of-mass, and the interdependence among nodes is represented as a connecting edge with an associated numeric value that reflects its strength. The interdependence is determined from an *N* × *N* matrix *C* (*N* is the number of nodes) with values corresponding to the weights of each edge, reflecting the correlated motion among node-node pairs. Finally, the weight between the edge that connects nodes *i* and *j* is expressed as *w_{ij}* = −log(|*C_{ij}*|), so that highly correlated or anticorrelated motions are characterized by small values of *w_{ij}*. This analysis was performed for the last 500 ns of the MD simulations.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <http://www.wwpdb.org/>, 5UFU; <http://www.wwpdb.org/>, 6B2E; <http://www.wwpdb.org/>, 6B1U.

AUTHOR CONTRIBUTIONS

EA: Formal analysis, Investigation, Visualization, Writing—original draft. AG: Formal analysis, Investigation, Visualization. AF: Formal analysis, Investigation, Visualization. LV: Methodology, Investigation. IR: Methodology, Investigation. CE: Conceptualization, Methodology, Investigation, Supervision, Writing—review and editing. FL: Conceptualization, Methodology, Investigation, Supervision, Writing—review and editing. Funding acquisition.

ACKNOWLEDGMENTS

We thank the Spanish Ministerio de Economía y Competitividad (SAF2017-88107-R, and Maria de Maetzu MDM-2017-0767, AEI/FEDER), and the Generalitat de Catalunya (2017SGR1746) for

financial support and the Barcelona Supercomputing Center (BCV-2019-2-0017 and BCV-2019-1-0009) and the Consorci de Serveis Universitaris de Catalunya (CSUC) for computational resources. EA thanks AGAUR (Generalitat of Catalunya; 2018FI-B1-00001) for a fellowship. IR and AG acknowledge the support of the Institut Rhônealpin des systèmes complexes, IXXI-ENS-Lyon, Lyon, France, and the use of HPC resources of the “Pôle Scientifique de Modélisation Numérique” (PSMN) at the École Normale Supérieure de Lyon, France.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.760026/full#supplementary-material>

REFERENCES

- Aledavood, E., Forte, A., Estarellas, C., and Javier Luque, F. (2021). Structural Basis of the Selective Activation of Enzyme Isoforms: Allosteric Response to Activators of $\beta 1$ - and $\beta 2$ -containing AMPK Complexes. *Comput. Struct. Biotechnol. J.* 19, 3394–3406. doi:10.1016/j.csbj.2021.05.056
- Aledavood, E., Moraes, G., Lameira, J., Castro, A., Luque, F. J., and Estarellas, C. (2019). Understanding the Mechanism of Direct Activation of AMP-Kinase: Toward a Fine Allosteric Tuning of the Kinase Activity. *J. Chem. Inf. Model.* 59, 2859–2870. doi:10.1021/acs.jcim.8b00890
- Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. (1993). Essential Dynamics of Proteins. *Proteins* 17, 412–425. doi:10.1002/prot.340170408
- AMBER parameter database (2021). Bryce Group: Computational Biophysics and Drug Design. Available at: <http://amber.manchester.ac.uk>.
- Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993). A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: the RESP Model. *J. Phys. Chem.* 97, 10269–10280. doi:10.1021/j100142a004
- Calabrese, M. F., Rajamohan, F., Harris, M. S., Caspers, N. L., Magyar, R., Withka, J. M., et al. (2014). Structural Basis for AMPK Activation: Natural and Synthetic Ligands Regulate Kinase Activity from Opposite Poles by Different Molecular Mechanisms. *Structure* 22, 1161–1172. doi:10.1016/j.str.2014.06.009
- Carling, D. (2017). AMPK Signalling in Health and Disease. *Curr. Opin. Cell Biol.* 45, 31–37. doi:10.1016/j.ccb.2017.01.005
- Carling, D., Thornton, C., Woods, A., and Sanders, M. J. (2012). AMP-activated Protein Kinase: New Regulation, New Roles. *Biochem. J.* 445, 11–27. doi:10.1042/bj20120546
- Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., and Cruzeiro, V. W. D. (2018). *AMBER 2018*. San Francisco: University of California.
- Chen, L., Wang, J., Zhang, Y.-Y., Yan, S. F., Neumann, D., Schlattner, U., et al. (2012). AMP-activated Protein Kinase Undergoes Nucleotide-dependent Conformational Changes. *Nat. Struct. Mol. Biol.* 19, 716–718. doi:10.1038/nsmb.2319
- Cokorinos, E. C., Delmore, J., Reyes, A. R., Albuquerque, B., Kjøbsted, R., Jørgensen, N. O., et al. (2017). Activation of Skeletal Muscle AMPK Promotes Glucose Disposal and Glucose Lowering in Non-human Primates and Mice. *Cel Metab.* 25, 1147–1159. doi:10.1016/j.cmet.2017.04.010
- Cool, B., Zinker, B., Chiou, W., Kifle, L., Cao, N., Perham, M., et al. (2006). Identification and Characterization of a Small Molecule AMPK Activator that Treats Key Components of Type 2 Diabetes and the Metabolic Syndrome. *Cel Metab.* 3, 403–416. doi:10.1016/j.cmet.2006.05.005
- Dorantes-Gilardi, R., Bourgeat, L., Pacini, L., Vuillon, L., and Lesieur, C. (2018). In Proteins, the Structural Responses of a Position to Mutation Rely on the Goldilocks Principle: Not Too many Links, Not Too Few. *Phys. Chem. Chem. Phys.* 20, 25399–25410. doi:10.1039/c8cp04530e
- Gheeraert, A., Pacini, L., Batista, V. S., Vuillon, L., Lesieur, C., and Rivalta, I. (2019). Exploring Allosteric Pathways of a V-type Enzyme with Dynamical
- Perturbation Networks. *J. Phys. Chem. B* 123, 3452–3461. doi:10.1021/acs.jpcc.9b01294
- Hardie, D. G. (2014). AMPK-sensing Energy while Talking to Other Signaling Pathways. *Cel Metab.* 20, 939–952. doi:10.1016/j.cmet.2014.09.013
- Human Protein Atlas (2021). Human Protein Atlas. available at: <http://www.proteinatlas.org>.
- Hünenberger, P. H., Mark, A., and Mark, W. F. (1995). Fluctuation and Cross-Correlation Analysis of Protein Motions Observed in Nanosecond Molecular Dynamics Simulations. *J. Mol. Biol.* 252, 492–503. doi:10.1006/jmbi.1995.0514
- Jørgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79, 926–935. doi:10.1063/1.445869
- Joung, I. S., and Cheatham, T. E. (2008). Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* 112, 9020–9041. doi:10.1021/jp8001614
- Langendorf, C. G., and Kemp, B. E. (2015). Choreography of AMPK Activation. *Cell Res* 25, 5–6. doi:10.1038/cr.2014.163
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., et al. (2010). Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins* 78, 1950–1958. doi:10.1002/prot.22711
- Mahlpuu, M., Johansson, C., Lindgren, K., Hjälm, G., Barnes, B. R., Krook, A., et al. (2004). Expression Profiling of the γ -subunit Isoforms of AMP-Activated Protein Kinase Suggests a Major Role for $\gamma 3$ in white Skeletal Muscle. *Am. J. Physiology-Endocrinology Metab.* 286, E194–E200. doi:10.1152/ajpendo.00147.2003
- Meagher, K. L., Redman, L. T., and Carlson, H. A. (2003). Development of Polyphosphate Parameters for Use with the AMBER Force Field. *J. Comput. Chem.* 24, 1016–1025. doi:10.1002/jcc.10262
- Myers, R. W., Guan, H.-P., Ehrhart, J., Petrov, A., Prahalada, S., Tozzo, E., et al. (2017). Systemic Pan-AMPK Activator MK-8722 Improves Glucose Homeostasis but Induces Cardiac Hypertrophy. *Science* 357, 507–511. doi:10.1126/science.aah5582
- Ngoei, K. R. W., Langendorf, C. G., Ling, N. X. Y., Hoque, A., Varghese, S., Camerino, M. A., et al. (2018). Structural Determinants for Small-Molecule Activation of Skeletal Muscle AMPK $\alpha 2\beta 2\gamma 1$ by the Glucose Importer SC4. *Cel Chem. Biol.* 25, 728–737. doi:10.1016/j.chembiol.2018.03.008
- O'Rourke, K. F., Gorman, S. D., and Boehr, D. D. (2016). Biophysical and Computational Methods to Analyze Amino Acid Interaction Networks in Proteins. *Comput. Struct. Biotechnol. J.* 14, 245–251. doi:10.1016/j.csbj.2016.06.002
- Olivier, S., Foretz, M., and Viollet, B. (2018). Promise and Challenges for Direct Small Molecule AMPK Activators. *Biochem. Pharmacol.* 153, 147–158. doi:10.1016/j.bcp.2018.01.049
- Scott, J. W., Hawley, S. A., Green, K. A., Anis, M., Stewart, G., Scullion, G. A., et al. (2004). CBS Domains Form Energy-Sensing Modules Whose Binding of Adenosine Ligands Is Disrupted by Disease Mutations. *J. Clin. Invest.* 113, 274–284. doi:10.1172/jci19874

- Scott, J. W., van Denderen, B. J. W., Jorgensen, S. B., Honeyman, J. E., Steinberg, G. R., Oakhill, J. S., et al. (2008). Thienopyridone Drugs Are Selective Activators of AMP-Activated Protein Kinase β 1-Containing Complexes. *Chem. Biol.* 15, 1220–1230. doi:10.1016/j.chembiol.2008.10.005
- Shao, J., Tanner, S. W., Thompson, N., and Cheatham, T. E. (2007). Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theor. Comput.* 3 (6), 2312–2334. doi:10.1021/ct700119m
- Shkurti, A., Goni, R., Andrio, P., Breitmoser, E., Bethune, I., Orozco, M., et al. (2016). pyPcazip: A PCA-Based Toolkit for Compression and Analysis of Molecular Simulation Data. *SoftwareX* 5, 44–50. doi:10.1016/j.softx.2016.04.002
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Tissue-based Map of the Human Proteome. *Science* 347, 1260419. doi:10.1126/science.1260419
- Van Wart, A. T., Durrant, J., Votapka, L., and Amaro, R. E. (2014). Weighted Implementation of Suboptimal Paths (WISP): an Optimized Algorithm and Tool for Dynamical Network Analysis. *J. Chem. Theor. Comput.* 10, 511–517. doi:10.1021/ct4008603
- Vazquez-Martin, A., Vellon, L., Quirós, P. M., Cufí, S., Ruiz de Galarreta, E., Oliveras-Ferraro, C., et al. (2012). Activation of AMP-Activated Protein Kinase (AMPK) Provides a Metabolic Barrier to Reprogramming Somatic Cells into Stem Cells. *Cell Cycle* 11, 974–989. doi:10.4161/cc.11.5.19450
- Vuillon, L., and Lesieur, C. (2015). From Local to Global Changes in Proteins: a Network View. *Curr. Opin. Struct. Biol.* 31, 1–8. doi:10.1016/j.sbi.2015.02.015
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and Testing of a General AMBER Force Field. *J. Comput. Chem.* 25, 1157–1174. doi:10.1002/jcc.20035
- Willows, R., Sanders, M. J., Xiao, B., Patel, B. R., Martin, S. R., Read, J., et al. (2017). Phosphorylation of AMPK by Upstream Kinases Is Required for Activity in Mammalian Cells. *Biochem. J.* 474, 3059–3073. doi:10.1042/bcj20170458
- Xiao, B., Sanders, M. J., Carmena, D., Bright, N. J., Haire, L. F., Underwood, E., et al. (2013). Structural Basis of AMPK Regulation by Small Molecule Activators. *Nat. Commun.* 4, 3017. doi:10.1038/ncomms4017
- Xiao, B., Sanders, M. J., Underwood, E., Heath, R., Mayer, F. V., Carmena, D., et al. (2011). Structure of Mammalian AMPK and its Regulation by ADP. *Nature* 472, 230–233. doi:10.1038/nature09932

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Aledavood, Gheeraert, Forte, Vuillon, Rivalta, Luque and Estarellas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Improving Small-Molecule Force Field Parameters in Ligand Binding Studies

Stefano Raniolo¹ and Vittorio Limongelli^{1,2*}

¹Faculty of Biomedical Sciences, Euler Institute, Università della Svizzera italiana (USI), Lugano, Switzerland, ²Department of Pharmacy, University of Naples "Federico II", Naples, Italy

Small molecules are major players of many chemical processes in diverse fields, from material science to biology. They are made by a combination of carbon and heteroatoms typically organized in system-specific structures of different complexity. This peculiarity hampers the application of standard force field parameters and their *in silico* study by means of atomistic simulations. Here, we combine quantum-mechanics and atomistic free-energy calculations to achieve an improved parametrization of the ligand torsion angles with respect to the state-of-the-art force fields in the paradigmatic molecular binding system benzamidine/trypsin. Funnel-Metadynamics calculations with the new parameters greatly reproduced the high-resolution crystallographic ligand binding mode and allowed a more accurate description of the binding mechanism, when the ligand might assume specific conformations to cross energy barriers. Our study impacts on future drug design investigations considering that the vast majority of marketed drugs are small-molecules.

OPEN ACCESS

Edited by:

Ivan Rivalta,
University of Bologna, Italy

Reviewed by:

Qinghua Liao,
University of Barcelona, Spain
Francois Sicard,
University College London,
United Kingdom

*Correspondence:

Vittorio Limongelli
vittoriolimongelli@gmail.com

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 17 August 2021

Accepted: 20 October 2021

Published: 13 December 2021

Citation:

Raniolo S and Limongelli V (2021)
Improving Small-Molecule Force Field
Parameters in Ligand Binding Studies.
Front. Mol. Biosci. 8:760283.
doi: 10.3389/fmolb.2021.760283

Keywords: benzamidine, trypsin, CGenFF, GAFF, QM calculations, free-energy calculations, force fields

INTRODUCTION

Small molecules are organic compounds of relatively low molecular weight which are responsible of specific chemical reactions. Their range of application is broad and spans from material science to pharmacology where they historically represent the end product of drug discovery. Understanding the mechanism of action of a drug by elucidating the binding interaction with its molecular target is fundamental to speed up the drug discovery pipeline, however it might be a long and daunting process, generally requiring extensive and costly experiments (Lindsay, 2003; Tian et al., 2013; Tulloch et al., 2018). In addition, this effort is seldom rewarded since most drug candidates end up in failure due to poor pharmacokinetics properties or the interaction with off-targets (Lindsay, 2003; Waring et al., 2015; Mao et al., 2016). Nonetheless, few of them are successful and in even fewer cases the ligand/protein bound poses are experimentally resolved by X-ray crystallography, providing important insights at atomistic level on the possible interaction mechanism and the binding site in the target molecule. Such information is invaluable for a *de novo* drug design campaign on that target or to modify the ligand structure in order to improve its activity and toxicity, but it is available merely for a selection of drugs. Therefore, the development of reliable and relatively fast computational techniques capable of inferring mechanisms of binding in ligand/protein complexes has taken the scene.

In this scenario, molecular dynamics (MD) simulations and the related enhanced sampling techniques are gaining ground in the years, being able to reproduce crystallographic binding poses and compute absolute binding free energies in agreement with experimental data (Limongelli et al., 2013; Troussicot et al., 2015; Broomhead and Soliman, 2016; Yuan et al., 2018; Limongelli, 2020;

Raniolo and Limongelli, 2020; Souza et al., 2020; Monticelli et al., 2021). The possibility to extrapolate atomistic details at relatively low cost makes these computational approaches invaluable tools both to assist drug design and to shed light on important features of biologically relevant processes. In particular, the development of enhanced sampling techniques and the growing computing power are allowing to alleviate the sampling issue with simulation times even closer to the real timescales of ligand/protein binding. However, the precision and accuracy of the physics applied in MD simulations are highly dependent on the reliability of the chosen force fields (FF), which implement properties of atoms - or beads in the case of coarse-grained (CG) MD - that compose the system under investigation. There are plenty of molecular FFs available and all of them are based on data extracted from experimental assays or quantum mechanical (QM) calculations (Mackerell Jr, 2004). Although they approximate the known properties of molecules, FFs yet allow reproducing complicated systems, such as ligand/protein complexes, with a fair accuracy. On the other hand, standalone QM calculations might ideally provide more accurate results, but they are unfeasible due to the enormous amount of atoms to consider and the sheer complexity of the calculation.

The first biologically-related FFs considered mainly proteins, nucleotides, lipids, and carbohydrates, while small drug-like compounds had been neglected in the first stage of the FFs' life. Eventually, we saw in the last decades the birth of a number of small-molecule FFs, such as the Generalized Amber Force Field (GAFF and its more recent version GAFF2), the CHARMM General Force Field (CGenFF), and OPLS3 among the most famous (Wang et al., 2004; Vanommeslaeghe et al., 2010; Harder et al., 2016). Lately, an extended version of the OPLS FF has been published (i.e., OPLS3e and OPLS4), featuring improved torsional angle description and a wider coverage of the chemical space (Roos et al., 2019; Lu et al., 2021). The main difficulty in developing such FFs is the huge variety of scaffolds and functional groups that a ligand can possess, and taking into consideration all of them requires an astonishing amount of experimental information and/or QM calculations. Therefore, most of the FFs only consider a smaller sample of compounds, trying to extrapolate general rules for all the possible cases. This operation might lead to severe approximations in the generation of the ligand parameters and a manual correction from the investigator is often required. This is especially true in torsion angle parameterization of ligands with π electron conjugated systems, where quantum (ligand-specific) effects are particularly important in defining the correct molecular conformation. In this article, we showcase the effect of these problems for benzamidine, a small molecule binding to the protein trypsin, composed of an amidine group conjugated to a benzene ring and used as prototypical model in drug/protein binding studies. Our results show important changes in sampling behaviour of the small molecule depending on different parametrizations and their consequences when free energy calculations are performed to study the binding mechanism with its molecular target.

MATERIALS AND METHODS

In this section we describe the process to obtain the benzamidine parameters using the most widely used and freely available FFs for

ligands (GAFF, GAFF2, and CGenFF); the simulations employed to obtain the benzamidine conformational potential energy; and the binding free-energy calculations between benzamidine and trypsin.

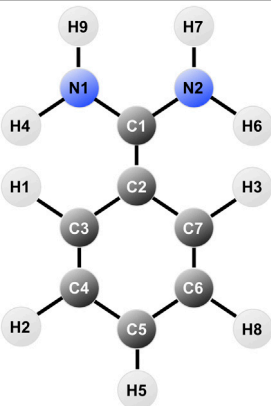
In order to simulate the molecular binding process, it is necessary to set the parameters to describe the properties of both benzamidine and trypsin. Regarding the latter, there are several optimised FFs for proteins, which have reached a satisfactory level of accuracy. On the contrary, small molecules are made by different combinations of diverse atoms that render unfeasible to catalogue ligands as done for the amino acids in a protein. Therefore, for benzamidine new parameters have to be created either automatically (e.g., from software that try to infer molecular properties) or manually. Among the available FFs for trypsin, we opted for Amber14SB, which represents an improvement with respect to the previous Amber99SB version (Maier et al., 2015). Regarding benzamidine, new parameters compatible with Amber14SB had to be generated. This is possible by employing the Amber "generalised" parameters collected inside the GAFF library, which contains several atom types, bonded and non-bonded parameters to describe the sampling behaviour of organic molecules. In 2016, a new version of this library (i.e., GAFF2) was released to account for improved torsional characterisation and molecular properties, such as intermolecular energy, liquid density, heat of vaporisation, and hydration free energy.

The parametrization of trypsin using Amber14SB was quite straightforward, contrarily to benzamidine that requested a "construction protocol", which was inspired by the GAFF reference paper and that will be described in the following section (Wang et al., 2004).

Ligand Parametrization

The first step is to perform a geometry optimization of the benzamidine structure using the QM software Gaussian09 (Frisch et al., 2009). This step was divided in two different optimisation steps with increasing basis set complexity (3-21G and 6-31G*, respectively), where a Hartree-Fock calculation was requested, followed by a Möller-Plesset correlation energy correction truncated at the second order (MP2) (Møller and Plesset, 1934; Gordon et al., 1982; Petersson et al., 1988). Total charge for the system was set to +1, since in solution at physiological pH the amidine group is protonated ($pK_a = 11.6$) (Lam et al., 2003). Once obtained the optimised structure, we performed a population analysis with Hartree-Fock to produce charges considering the electrostatic potential at points following the Merz-Singh-Kollman scheme (Besler et al., 1990). The output file was post-processed by the Restrained Electrostatic Potential (RESP) method to obtain the partial charges per atom (Bayly et al., 1993). Restraints in the charge allocation were applied to account for benzamidine symmetry.

The atomic charges thus obtained were used to create the topology of benzamidine in the case of GAFF and GAFF2, whereas CGenFF already has tabulated charges for the molecule so we applied them for internal consistency with the rest of the parameters. Notably, the CGenFF charges differ from those obtained through RESP, especially for the amidine group, as listed in **Table 1**.

TABLE 1 | Table reporting the atomic charges for benzamidine in CGenFF and RESP.


Atom	CGenFF	RESP
N1	-0.600	-0.901011
H4	0.320	0.447034
H9	0.320	0.474397
C1	0.730	0.798507
N2	-0.600	-0.901011
H6	0.320	0.447034
H7	0.320	0.474397
C2	0.190	-0.102912
C3	-0.115	-0.121702
H1	0.115	0.157313
C4	-0.115	-0.136868
H2	0.115	0.173449
C5	-0.115	-0.045171
H5	0.115	0.164353
C6	-0.115	-0.136868
H8	0.115	0.173449
C7	-0.115	-0.121702
H3	0.115	0.157313

The bonded parameters for GAFF and GAFF2 were created using the Antechamber package of Amber14 (Wang et al., 2004). A new *ad hoc* topology was also produced by taking as a template the parameters from GAFF and modifying the dihedral angle along the bond between amidine and benzene. The role of this peculiar dihedral angle will be better highlighted in the “Results and Discussion” section. The *ad hoc* parameters were obtained by means of a “Scan” calculation using the settings of the second geometry optimisation process. Such procedure involves a number of energy minimisation calculations in which the dihedral angle of interest is restrained in a specific conformation and relaxing the rest of the molecule. Once obtained the energy profile for the dihedral angle under investigation through QM calculations, we resolved the parameters that reproduced the QM potential energy function using the Amber dihedral formula $E_{dih} = k (1 + \cos (n\phi - \psi))$, with k being the spring constant for the dihedral angle, n the period, ϕ the angle, and ψ the phase. The function that best reproduced the QM behaviour was computed with an in-house genetic algorithm and it is a combination of two torsional angles that has the equation $E_{dih} = 2.4 (1 + \cos (2\phi - \pi)) + 1.0 (1 + \cos (4\phi))$. A similar protocol is reported in the GAFF reference manuscript

and has already been employed in literature for other compounds (Pophristic et al., 2006). The new parameters, together with adjustments to angle values taken from the QM geometry optimisation, replaced the original GAFF values and represent our *ad hoc* topology.

Ligand Conformational Analysis

The simulations with the all-atom FFs were performed using well-tempered metadynamics (MetaD) (Barducci et al., 2008), using the torsional angle connecting the amidine and the phenyl group as collective variable (i.e., the degree of freedom that we are going to sample), and compared the obtained energy profiles with that from QM calculations. The MetaD calculations were run using Amber14 patched with Plumed-2.3.3 (Tribello et al., 2014), setting a low value as height of the metadynamics Gaussian functions (i.e., 0.1 kJ/mol for CGenFF and the *ad hoc* topology and 0.05 kJ/mol for GAFF and GAFF2), a sigma of 0.05 radians, a biasfactor of 15, and a deposition rate of 1,000 steps. All simulations were carried out in vacuum and converged within 100 ns (see **Supplementary Figure S1** in the “Supplementary Material” for convergence plots). Post-processing analysis was performed with Plumed-2.3.3 and Visual Molecular Dynamics (VMD) (Humphrey et al., 1996; Tribello et al., 2014).

MD Calculations on the Benzamidine/Trypsin Complex

Classical MD simulations on the benzamidine/trypsin complex in water were performed with the Amber14 software to assess the stability of benzamidine in the binding pocket employing GAFF and *ad hoc* ligand parametrization (Case et al., 2014; Abraham et al., 2015). We produced 500 ns of simulation of benzamidine in the bound pose with the Amber14SB FF and water model TIP3P.

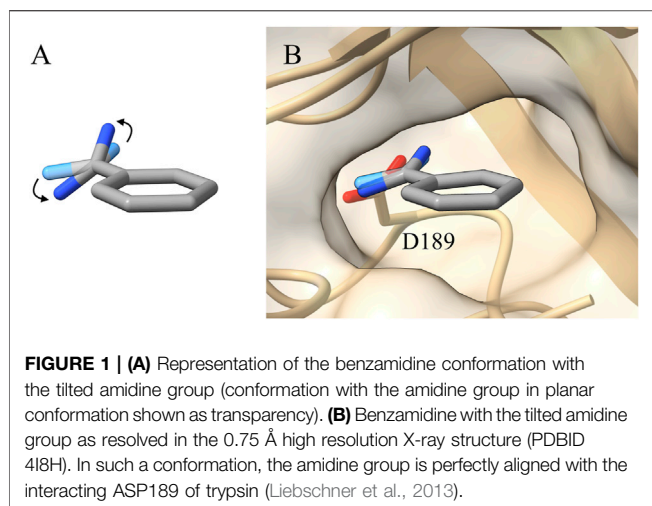
Binding Free-Energy Calculations on the Benzamidine/Trypsin Complex

Funnel-Metadynamics simulations were used to obtain the binding free-energy surface of the benzamidine/trypsin complex (Limongelli et al., 2013). The GAFF free energy data were taken from our previous work (Limongelli et al., 2013), while new calculations were performed to build the *ad hoc* binding free energy surface. In particular, the protocol we have recently reported in Nature Protocols was employed (Raniolo and Limongelli, 2020). More details on the Funnel-Metadynamics calculations can be found in the “Supplementary Material” and the reference manuscript (Raniolo and Limongelli, 2020). As reported in Limongelli et al. (Limongelli et al., 2013), the estimate of the absolute ligand/protein binding free-energy (ΔG_b^0) was obtained using the following formula:

$$K_b = \pi R_{cyl}^2 \int_{site} e^{-\beta[W(z) - W_{ref}]} dz \quad (1)$$

$$\Delta G_b^0 = -k_b T \ln(K_b C^0) \quad (2)$$

where K_b is the binding constant, R_{cyl} is the radius of the cylinder section of the FM potential, β is the inverse of the Boltzmann constant (k_b) multiplied by temperature (T), $W(z)$ is the potential



of mean force (PMF) at given value z for the projection of the ligand over the axis of the funnel potential, W_{ref} is a reference value of the PMF for the unbound state, and C^0 is the standard concentration.

RESULTS AND DISCUSSION

Benzamidine is a small molecule acting as an inhibitor of trypsin and it interacts mainly through an amidine group contacting an aspartate in the binding pocket of the protein (ASP189) (**Figure 1**).

This complex is widely used as a benchmark system for newly developed computational techniques aimed at disclosing ligand/protein binding modes and calculate binding free energies. The results obtained with a new technique can be indeed compared with those achieved with already established methods and with experimental values obtained through isothermal titration calorimetry (ITC) (Katz et al., 2001; Talhout and Engberts, 2001; Doudou et al., 2009; Buch et al., 2011; Söderhjelm et al., 2012; Limongelli et al., 2013; Takahashi et al., 2014).

While setting up the system, we generated the topology for both benzamidine and trypsin, as reported in the “Materials and Methods” section. During the geometrical optimization of benzamidine with Gaussian09, a rotation of the amidine group with respect to the benzene ring was observed, ending up in a $\pi/4$ radians tilt between the two groups (**Figure 1A**). The rotation disaligned the two groups, *de facto* partially disrupting the conjugation between them. This behaviour is likely due to the steric clash formed by the hydrogen atoms of the amidine group and those of the benzene ring. In the attempt to resolve the steric hindrance, the amidine group tilts, resulting in the conformation obtained by Gaussian09 (**Figure 1A**).

It is worth noting that this system has been widely investigated in the past and in several works a planar benzamidine—with the amidine group not tilted—was reported in the benzamidine-trypsin complex (Doudou et al., 2009; Buch et al., 2011; Söderhjelm et al., 2012; Limongelli et al., 2013; Takahashi et al., 2014; Tiwary et al., 2015). However, there are cases in which a thorough analysis has been carried out, obtaining

benzamidine geometries similar to the one defined by our calculation (Li et al., 2009).

In order to gain more insights, we compared the Gaussian09 optimised structure with several crystals deposited in the Protein Databank (PDB) and we found a more recent high-resolution structure (0.75 Å, PDBID 4I8H) where the benzamidine with the tilted amidine group was crystallized inside the binding pocket of trypsin (**Figure 1B**) (Liebschner et al., 2013).

Then, we decided to investigate more deeply the rotation around the torsional angle responsible for this behaviour and we performed a thorough analysis of the dihedral angle with Gaussian09 (see “Materials and Methods” for further details). The results clearly show that the fully planar conformation is highly disfavoured with an energy barrier of around 4.5 kcal/mol, whereas the minima reside at $\pi/4 + n\pi/2$ angles, with n being an integer number that takes into account the periodicity of the angle (**Figure 2**, black line).

This finding contrasts with some of the results previously reported in literature, therefore we proceeded with comparing the conformational energy profile of benzamidine obtained with QM calculations with those computed using the available small-molecule libraries (Doudou et al., 2009; Buch et al., 2011; Söderhjelm et al., 2012; Limongelli et al., 2013; Takahashi et al., 2014; Tiwary et al., 2015). In particular, the topology file of the ligand, necessary to perform the conformational analysis simulations, can be generated using position and point charges of the ligand’s atoms together with the bonded and non-bonded parameters from the FF’s small-molecule library files (i.e., GAFF or GAFF2 using the AmberTools module in the case of the Amber FF). The parameters are assigned based on identity or similarity of the atoms present in the molecule with the atom types reported in the FF library. This procedure might lead to major approximations in the calculation if the ligand’s atoms are not very similar to those already parametrised in the library. Thus, we created the benzamidine topology for a selection of widely used, open-source FFs for small molecules to assess the goodness of the parameters, especially in reproducing the correct conformational behaviour of the dihedral angle connecting the amidine group and the phenyl ring. In particular, we investigated topologies created from the two versions of the Amber GAFF library (GAFF and GAFF2) and the equivalent for the CHARMM-36 FF CGenFF, which includes specific parameters for benzamidine (BAMI). In addition, we created a brand new *ad hoc* topology by modifying the ligand parameters from GAFF in order to replicate as close as possible the torsion potential computed through QM calculations (see the “Materials and Methods” section). Finally, for each topology we performed all-atom metadynamics calculations using as single collective variable (CV) the dihedral angle connecting the amidine group and the benzene ring (see **Supplementary Figure S1** in the “Supplementary Materials”).

The obtained conformational energy profiles of benzamidine are compared with that from the QM calculations in **Figure 2**.

Interestingly, the best energy profile is achieved by means of the *ad hoc* topology with a difference in the estimate of the energy barriers lower than 0.5 kcal/mol with respect to QM (**Figure 2**, blue line). Conversely, all the Amber and CHARMM FFs underestimate the energy barriers. In

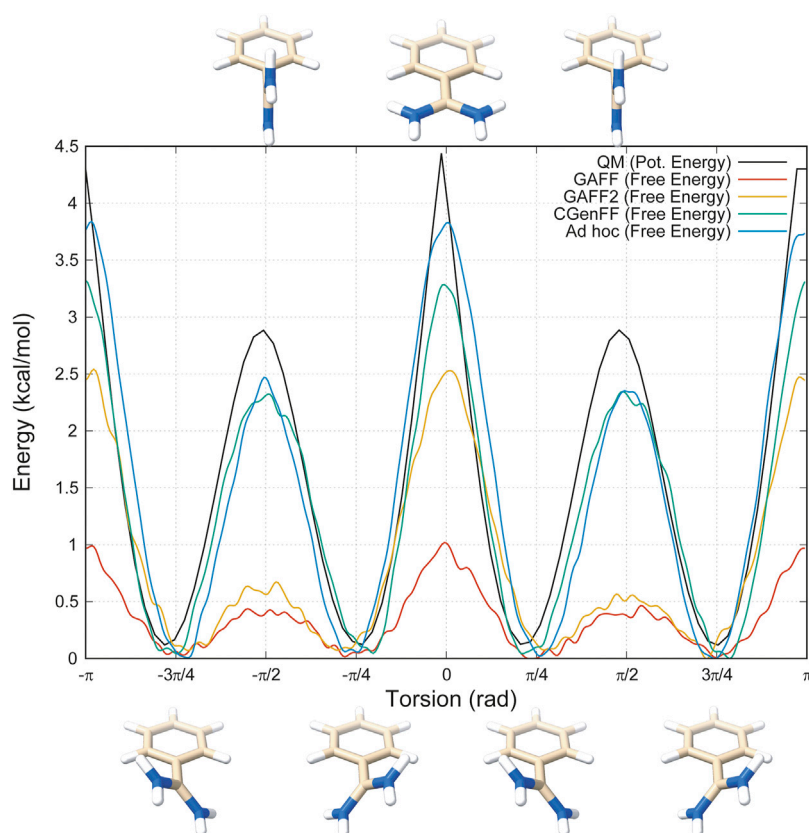


FIGURE 2 | Energy profiles of the dihedral angle between the benzene and the amidine group. In black is represented the one obtained with Gaussian09, in red GAFF, in orange GAFF2, in green CGenFF, and in blue the *ad hoc* parametrization purposely developed for this study.

particular, GAFF delivers the worst profile showing energy barriers of 1 and 0.5 kcal/mol at the planar and perpendicular conformations, respectively. By using GAFF2, the situation slightly improves with an energy barrier at the planar conformation around 2.5 kcal/mol, closer to 4.5 kcal/mol measured at QM level. On the other hand, the barrier at the perpendicular conformation at $\pm \pi/2$ remains largely underestimated compared to QM. In this scenario, CGenFF performs much better, providing around 3.0 and 2.5 kcal/mol for the barriers at the planar and perpendicular conformation, respectively. On the other hand, the atomic partial charges computed for benzamidine by CGenFF are quite different from the one calculated by QM (see “Materials and Methods” for details). This is an important aspect to consider, especially in processes like the benzamidine interaction with trypsin, where the binding is governed mainly by electrostatic contributions between the amidine group of the ligand and the carboxyl group of ASP189. In summary, all the FFs correctly identify the conformation with the tilted amidine group as the lowest energy state, however the Amber FFs severely underestimate the energy barriers that separate the different conformations assumed by benzamidine (**Figure 2**).

Prompted by these results, we decided to investigate the effect of the different ligand parametrization in its binding interaction

to trypsin. To this end, we performed plain all-atom MD calculations on two benzamidine/trypsin systems:

1. using GAFF (replicating the same conditions used in our previous work (Limongelli et al., 2013));
2. using the *ad hoc* topology.

Comparing these simulations, we expected to observe a higher conformational freedom for benzamidine using GAFF, considering the lower energy barriers. The prediction was indeed confirmed (**Figure 3**).

In the case of GAFF, benzamidine’s dihedral fluctuates around the value -0.13 radians (**Figure 3**, orange line) and it frequently populates the planar conformation (0 radian), whereas the latter is rarely visited using our *ad hoc* parametrization, where the dihedral value fluctuates around -0.5 radians (**Figure 3**, red line). In particular, in the latter case visiting conformations close to the barrier at 0 radian is energetically disfavoured, while the ligand assumes a more stable conformation with the tilted amidine group in which the nitrogen atoms of benzamidine are perfectly aligned with the oxygens of the carboxylic group of ASP189 in trypsin (**Figure 1B**). We extracted the structures of the benzamidine/trypsin complex representing the most populated conformational clusters visited during the *ad hoc* simulation at

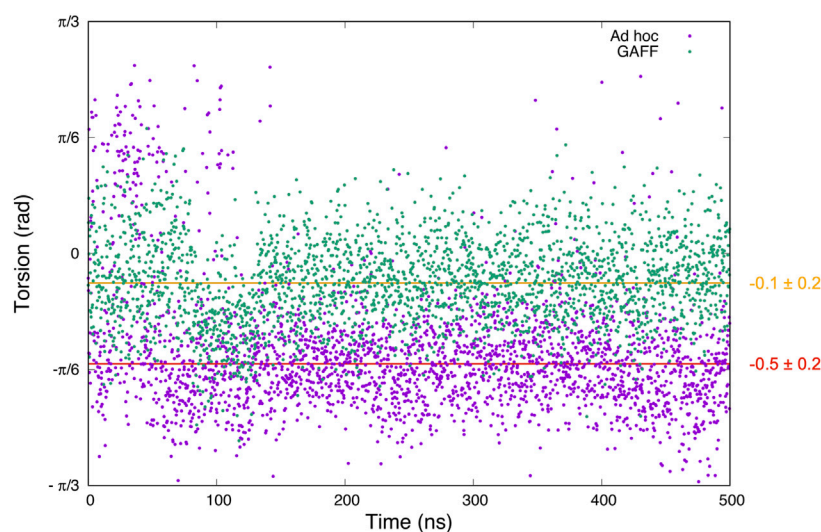


FIGURE 3 | Graph showing the values of the dihedral angle connecting the amidine group and the benzene ring of benzamidine observed along 500 ns of plain MD simulations. In green is the result for the GAFF parameters, while in violet is the one obtained with our *ad hoc* topology. The torsion average values and standard deviation using GAFF and *ad hoc* parameters are shown in orange and red, respectively.

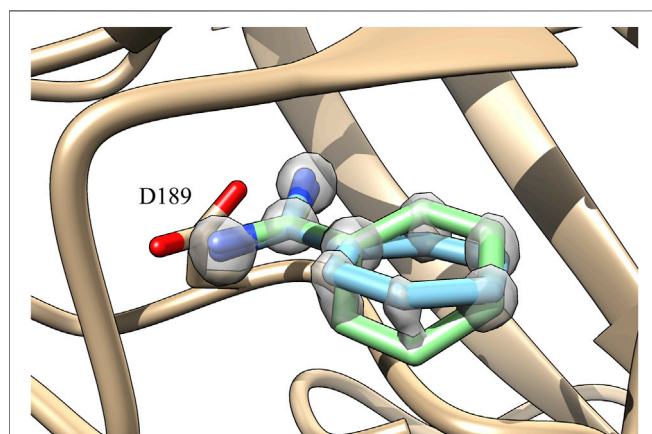


FIGURE 4 | Superimposition with the X-ray electronic density map of the two benzamidine conformations obtained with the *ad hoc* parametrization having the dihedral angle at 0.5 and -0.5 radians shown in green and cyan, respectively. The density map of benzamidine has been represented at a contour level of 2.6 \AA and transparency 0.5. The density map of the protein has been undisplayed for clarity reason (Liebschner et al., 2013).

which the studied dihedral angle assumes around ± 0.5 radians and we fit them in the electronic density map obtained through high-resolution X-ray experiments by Liebschner et al. (2013). Notably, the two in-silico conformations perfectly fit the map evidencing that our *ad hoc* parametrization allows reproducing the experimental ligand binding mode at a very high atomistic resolution of 0.75 \AA (Figure 4).

In order to investigate how the different benzamidine parameters affect the ligand binding mechanism to trypsin - intended as the physical pathway followed by the ligand to reach the binding site from its fully solvated state - we performed

binding free-energy calculations using Funnel-Metadynamics (Limongelli et al., 2013; Raniolo and Limongelli, 2020) and the two different parametrizations (i.e., using GAFF and the *ad hoc* topology). Details on the procedure are reported in the “Materials and Methods” section, while we refer the reader to our recent protocol manuscript for a more extensive description of the methodology, where the *ad hoc* topology has been employed (Raniolo and Limongelli, 2020). It is worth mentioning that we focused on the comparison with GAFF since the latter was employed in our previous work on the same system using the same charge protocol (Limongelli et al., 2013), thus making the comparison straightforward. In Figure 5, it is possible to see the free-energy landscapes computed in the two cases.

In detail, the absolute binding free-energy value remains almost the same (i.e., -8.5 and -8.2 kcal/mol for the GAFF and *ad hoc* simulation, respectively), falling in the range from -9.0 to -5.5 kcal/mol obtained in previous theoretical and experimental studies (Katz et al., 2001; Talhout and Engberts, 2001; Doudou et al., 2009; Buch et al., 2011; Söderhjelm et al., 2012; Limongelli et al., 2013; Takahashi et al., 2014; Tiwary et al., 2015). However, the number of free-energy minima changes with two minima for GAFF and only one for the *ad hoc* system (Figure 5). In particular, the second lowest energy minimum of the GAFF system (minimum B in the left graph in Figure 5) is characterized by a completely planar conformation assumed by benzamidine. Such state does not represent a low energy minimum in the *ad hoc* system and this is likely due to the higher energy barrier corresponding to the planar ligand conformation. Instead, the crystallographic binding mode was extensively explored and represents the absolute minimum in both cases (letter A in Figure 5). Furthermore, when the ligand is bound, its torsional freedom is much reduced with the *ad hoc* topology, exploring

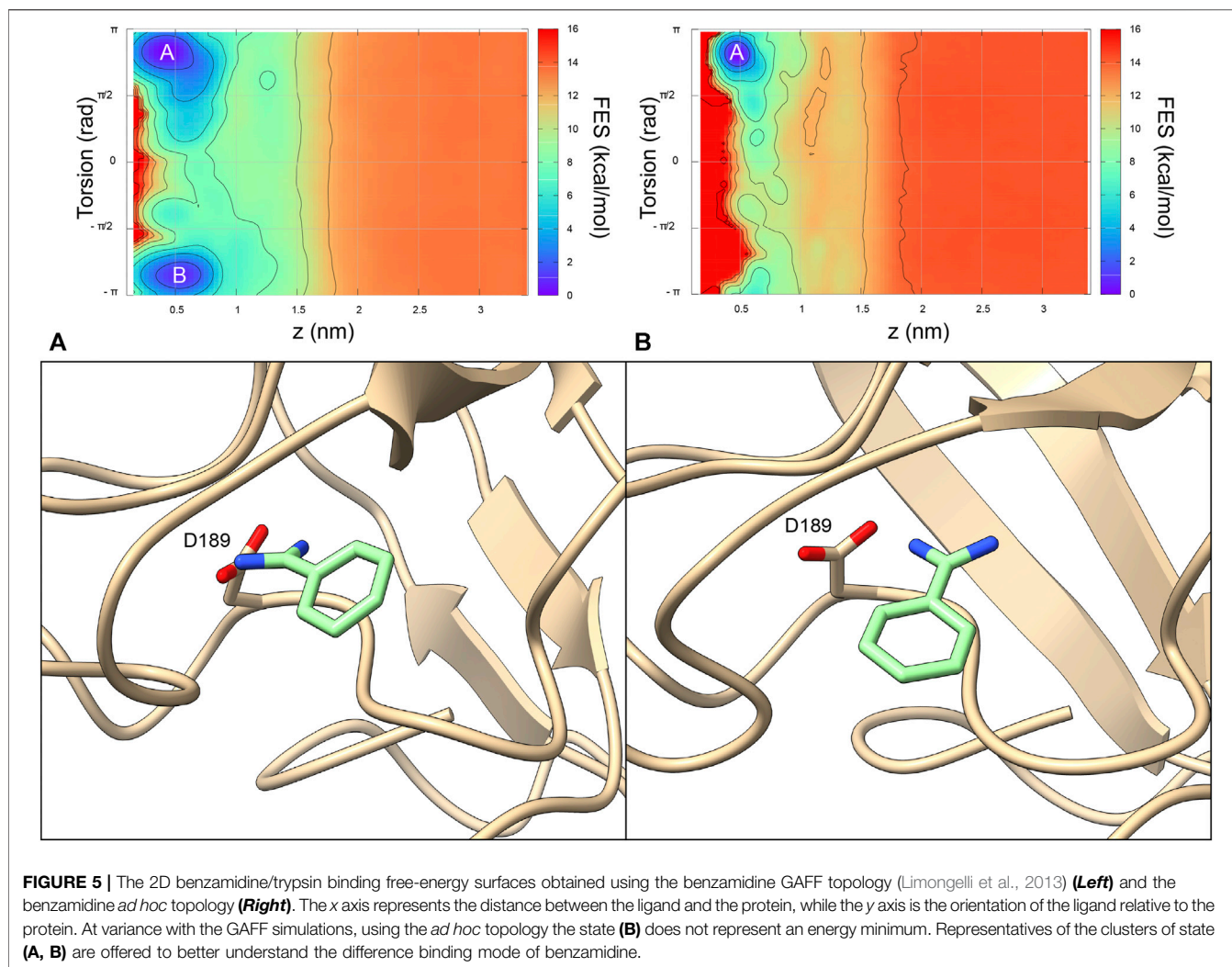


FIGURE 5 | The 2D benzamidine/trypsin binding free-energy surfaces obtained using the benzamidine GAFF topology (Limongelli et al., 2013) (**Left**) and the benzamidine *ad hoc* topology (**Right**). The x axis represents the distance between the ligand and the protein, while the y axis is the orientation of the ligand relative to the protein. At variance with the GAFF simulations, using the *ad hoc* topology the state (B) does not represent an energy minimum. Representatives of the clusters of state (A, B) are offered to better understand the difference binding mode of benzamidine.

much of the time the crystallographic binding mode (further information in the **Supplementary Data** and **Supplementary Figure S2**), which results in a narrower minimum A.

Finally, a different free energy profile for the ligand binding pathway was obtained in GAFF and *ad hoc* system. In particular, in the latter a larger number of states at higher energy values were found (**Figure 6**), indicating that the ligand has to cross multiple high energy barriers during the binding to pass from one state to another and eventually reach the binding site. This results in longer times of binding and unbinding in the *ad hoc* system if compared with the GAFF one. Such an aspect should be carefully considered if the aim of the investigation is to disclose by means of simulations states determining ligand binding kinetics, and compute association and dissociation rates.

CONCLUSION

In the present work, we have carried out a thorough study on ligand parametrization in the prototypical case of the benzamidine/trypsin binding complex. We have demonstrated that having accurate ligand parameters is fundamental to simulate

properly the ligand conformational freedom and more importantly its binding mechanism to the molecular target. In fact, during the physical approach of a drug to its molecular target, the ligand might assume specific conformations corresponding to certain values of ligand torsion angles - that are necessary to cross energy barrier and eventually reach the binding site. We have showed how the optimization of a single torsional angle potential deeply affects the free-energy landscape of the binding process, and in turn the characterization of both the thermodynamically and kinetically-relevant states which allow disclosing the ligand binding mode and the state determining the ligand binding rate, respectively. This process is fundamental to achieve an accurate estimate of the ligand binding constant K_b and ligand binding kinetics rate k_{on} and k_{off} which are useful to develop more potent ligands. It is worth noting that similar conclusions were achieved by Pophristic et al. and Li et al. who showed that an improved description of the ligand binding mode for arylamide and benzamidine-like compounds could be obtained through a more accurate force field parametrization (Pophristic et al., 2006; Li et al., 2009). The evidences of our study are expected to impact on future drug design investigations, especially considering that the majority of

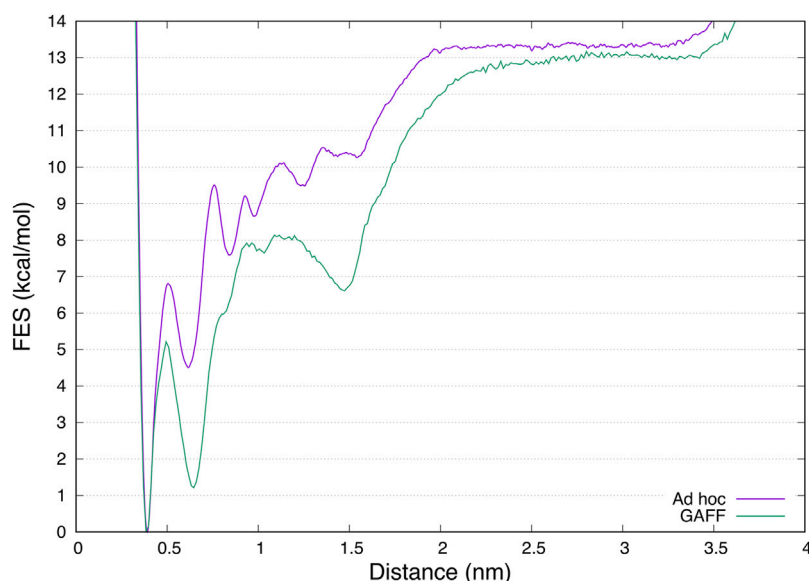


FIGURE 6 | Superimposition of the mono-dimensional free-energy profiles as a function of the distance between benzamidine and trypsin obtained from FM simulations. While the absolute minimum and the unbound region are almost overlapping, the in between region representing the ligand binding mechanism is different, characterized by more barriers at higher energy in the case of the *ad hoc* calculations.

the marketed drugs has much more complex structures than benzamidine, endowed with π electron conjugated systems, which might require a dedicated attention by the investigator and an *ad hoc* parametrization.

DATA AVAILABILITY STATEMENT

The data generated in this work can be found at Google drive. They are organized in folders thematically named as follows:

- QM - output files of the quantum-mechanical calculations employed for geometrical optimization, charge calculation and scanning procedures (input of each job is reported as header of each file);
- RESP - input and output files of the calculations of atomistic punctual charges that are subsequently added to the Amber topology file;
- simulations - inputs and trajectory files useful to launch and reproduce the simulations results reported in the dihedral_analysis, dihedral_sampling, and distance_analysis folders;
- topologies - topology files of all the employed force fields.

All QM calculations were performed with Gaussian 09 (Frisch et al., 2009). Gromacs-5.1.4, Gromacs-2016.5 and Amber14 were used as MD engines for the vacuum simulations to study the dihedral angle with different topologies and to analyse the sampling in the presence of the complex (Case et al., 2014; Abraham et al., 2015). All the metadynamics simulations were performed using Plumed 2.3.3 (Tribello et al., 2014). The results displayed in Figure 5 and 6 are from simulations on the

benzamidine-trypsin complex reported in our previous works. We report the employed methodology in chapter 1 (“**Supplementary Material**”) of the Supplementary Material, however more details on data and software employed can be found in Limongelli et al. (2013) and Raniolo and Limongelli (2020).

AUTHOR CONTRIBUTIONS

VL designed the research. SR performed the QM calculations and MM simulations. Both the authors analysed the results, contributed to manuscript writing, and approved the submitted version.

FUNDING

This work is part of a project that received funding from the European Research Council (ERC) (“CoMMBi” grant agreement no. 101001784). We further acknowledge the support from the Swiss National Science Foundation (Project No. 200021_163281), the Italian MIUR-PRIN 2017 (2017FJZZRC), and the Swiss National Supercomputing Center (CSCS) (project ID u8). Furthermore, we thank NVIDIA Corporation for the donation of the Tesla K40 GPU used in this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.760283/full#supplementary-material>

REFERENCES

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* 1-2, 19–25. doi:10.1016/J.SOFTX.2015.06.001
- Barducci, A., Bussi, G., and Parrinello, M. (2008). Well-tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* 100, 020603. doi:10.1103/PhysRevLett.100.020603
- Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993). A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The Resp Model. *J. Phys. Chem.* 97, 10269–10280. doi:10.1021/j100142a004
- Besler, B. H., Merz, K. M., and Kollman, P. A. (1990). Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* 11, 431–439. doi:10.1002/jcc.540110404
- Broomhead, N. K., and Soliman, M. E. (2016). Can We Rely on Computational Predictions to Correctly Identify Ligand Binding Sites on Novel Protein Drug Targets? Assessment of Binding Site Prediction Methods and a Protocol for Validation of Predicted Binding Sites. *Cell Biochem. Biophys.* 75, 15–23. doi:10.1007/s12013-016-0769-y
- Buch, I., Giorgino, T., and De Fabritiis, G. (2011). Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci.* 108, 10184–10189. doi:10.1073/pnas.1103547108
- Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., et al. (2014). *Amber14*. San Francisco: University of California.
- Doudou, S., Burton, N. A., and Henchman, R. H. (2009). Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. *J. Chem. Theor. Comput.* 5, 909–918. doi:10.1021/ct8002354
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian-09 Revision E.01*. Gaussian Inc. Wallingford CT 2009.
- Gordon, M. S., Binkley, J. S., Pople, J. A., Pietro, W. J., and Hehre, W. J. (1982). Self-consistent Molecular-Orbital Methods. 22. Small Split-Valence Basis Sets for Second-Row Elements. *J. Am. Chem. Soc.* 104, 2797–2803. doi:10.1021/ja00374a017
- Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., et al. (2016). Opls3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theor. Comput.* 12, 281–296. doi:10.1021/acs.jctc.5b00864
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual Molecular Dynamics. *J. Mol. Graphics* 14, 33–38. doi:10.1016/0263-7855(96)00018-5
- Katz, B. A., Elrod, K., Luong, C., Rice, M. J., Mackman, R. L., Sprengeler, P. A., et al. (2001). A Novel Serine Protease Inhibition Motif Involving a Multi-Centered Short Hydrogen Bonding Network at the Active site Edited by D. Rees. *J. Mol. Biol.* 307, 1451–1486. doi:10.1006/jmbi.2001.4516
- Lam, P. Y. S., Clark, C. G., Li, R., Pinto, D. J. P., Orwat, M. J., Gallemmo, R. A., et al. (2003). Structure-based Design of Novel Guanidine/benzamidine Mimics: Potent and Orally Bioavailable Factor Xa Inhibitors as Novel Anticoagulants. *J. Med. Chem.* 46, 4405–4418. doi:10.1021/jm020578e
- Li, X., He, X., Wang, B., and Merz, K., Jr. (2009). Conformational Variability of Benzanidinium-Based Inhibitors. *J. Am. Chem. Soc.* 131, 7742–7754. doi:10.1021/ja9010833
- Liebschner, D., Dauter, M., Brzuszkiewicz, A., and Dauter, Z. (2013). On the Reproducibility of Protein crystal Structures: Five Atomic Resolution Structures of Trypsin. *Acta Crystallogr. D Biol. Cryst.* 69, 1447–1462. doi:10.1107/S0907444913009050
- Limongelli, V., Bonomi, M., and Parrinello, M. (2013). Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci.* 110, 6358–6363. doi:10.1073/pnas.1303186110
- Limongelli, V. (2020). Ligand Binding Free Energy and Kinetics Calculation in 2020. *Wires Comput. Mol. Sci.* 10, 1–32. doi:10.1002/wcms.1455
- Lindsay, M. A. (2003). Target discovery. *Nat. Rev. Drug Discov.* 2, 831–838. doi:10.1038/nrd1202
- Lu, C., Wu, C., Ghoreishi, D., Chen, W., Wang, L., Damm, W., Ross, G. A., and Dahlgren, M. (2021). Opls4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theor. Comput.* 17, 4291–4300. doi:10.1021/acs.jctc.1c00302
- Mackerell, A. D., Jr (2004). Empirical Force fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* 25, 1584–1604. doi:10.1002/jcc.20082
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14sb: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99sb. *J. Chem. Theor. Comput.* 11, 3696–3713. doi:10.1021/acs.jctc.5b00255
- Mao, F., Ni, W., Xu, X., Wang, H., Wang, J., Ji, M., et al. (2016). Chemical Structure-Related Drug-like Criteria of Global Approved Drugs. *Molecules* 21, 75–18. doi:10.3390/molecules21010075
- Møller, C., and Plesset, M. S. (1934). Note on an Approximation Treatment for many-electron Systems. *Phys. Rev.* 46, 618–622. doi:10.1103/PhysRev.46.618
- Petersson, G. A., Bennett, A., Tensfeldt, T. G., Al-Laham, M. A., Shirley, W. A., and Mantzaris, J. (1988). A Complete Basis Set Model Chemistry. I. The Total Energies of Closed-shell Atoms and Hydrides of the First-row Elements. *J. Chem. Phys.* 89, 2193–2218. doi:10.1063/1.455064
- Pophristic, V., Vempala, S., Ivanov, I., Liu, Z., Klein, M. L., and DeGrado, W. F. (2006). Controlling the Shape and Flexibility of Arylamides: A Combined Ab Initio, Ab Initio Molecular Dynamics, and Classical Molecular Dynamics Study†. *J. Phys. Chem. B* 110, 3517–3526. doi:10.1021/jp054306+
- Raniolo, S., and Limongelli, V. (2020). Ligand Binding Free-Energy Calculations with Funnel Metadynamics. *Nat. Protoc.* 15, 2837–2866. doi:10.1038/s41596-020-0342-4
- Roos, K., Wu, C., Damm, W., Reboul, M., Stevenson, J. M., Lu, C., et al. (2019). Opls3e: Extending force field coverage for drug-like small molecules. *J. Chem. Theor. Comput.* 15, 1863–1874. doi:10.1021/acs.jctc.8b01026
- Söderhjelm, P., Tribello, G. A., and Parrinello, M. (2012). Locating Binding Poses in Protein-Ligand Systems Using Reconnaissance Metadynamics. *Proc. Natl. Acad. Sci.* 109, 5170–5175. doi:10.1073/pnas.1201940109
- Souza, P. C. T., Limongelli, V., Wu, S., Marrink, S. J., and Monticelli, L. (2021). Perspectives on High-Throughput Ligand/protein Docking with Martini Md Simulations. *Front. Mol. Biosci.* 8, 657222. doi:10.3389/fmolb.2021.657222
- Souza, P. C. T., Thallmair, S., Conflitti, P., Ramírez-Palacios, C., Alessandri, R., Raniolo, S., et al. (2020). Protein-ligand Binding with the Coarse-Grained Martini Model. *Nat. Commun.* 11, 1–11. doi:10.1038/s41467-020-17437-5
- Takahashi, R., Gil, V. A., and Guallar, V. (2014). Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations. *J. Chem. Theor. Comput.* 10, 282–288. doi:10.1021/ct400678g
- Talhout, R., and Engberts, J. B. F. N. (2001). Thermodynamic Analysis of Binding of P-Substituted Benzamidines to Trypsin. *Eur. J. Biochem.* 268, 1554–1560. doi:10.1046/j.1432-1033.2001.01991.x
- Tian, S., Li, Y., Li, D., Xu, X., Wang, J., Zhang, Q., et al. (2013). Modeling Compound-Target Interaction Network of Traditional Chinese Medicines for Type II Diabetes Mellitus: Insight for Polypharmacology and Drug Design. *J. Chem. Inf. Model.* 53, 1787–1803. doi:10.1021/ci400146u
- Tiwary, P., Limongelli, V., Salvaglio, M., and Parrinello, M. (2015). Kinetics of Protein-Ligand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. *Proc. Natl. Acad. Sci. USA* 112, E386–E391. doi:10.1073/pnas.1424461112
- Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. (2014). Plumed 2: New Feathers for an Old Bird. *Comp. Phys. Commun.* 185, 604–613. doi:10.1016/j.cpc.2013.09.018
- Troussicot, L., Guillière, F., Limongelli, V., Walker, O., and Lancelin, J.-M. (2015). Funnel-metadynamics and Solution Nmr to Estimate Protein-Ligand Affinities. *J. Am. Chem. Soc.* 137, 1273–1281. doi:10.1021/ja51133610.1021/ja511336z
- Tulloch, L. B., Menzies, S. K., Coron, R. P., Roberts, M. D., Florence, G. J., and Smith, T. K. (2018). Direct and Indirect Approaches to Identify Drug Modes of Action. *IUBMB Life* 70, 9–22. doi:10.1002/iub.1697

- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., et al. (2009). Charmm General Force Field: A Force Field for Drug-like Molecules Compatible with the Charmm All-Atom Additive Biological Force fields. *J. Comput. Chem.* 31, NA. doi:10.1002/jcc.21367
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and Testing of a General Amber Force Field. *J. Comput. Chem.* 25, 1157–1174. doi:10.1002/jcc.20035
- Waring, M. J., Arrowsmith, J., Leach, A. R., Leeson, P. D., Mandrell, S., Owen, R. M., et al. (2015). An Analysis of the Attrition of Drug Candidates from Four Major Pharmaceutical Companies. *Nat. Rev. Drug Discov.* 14, 475–486. doi:10.1038/nrd4609
- Yuan, X., Raniolo, S., Limongelli, V., and Xu, Y. (2018). The Molecular Mechanism Underlying Ligand Binding to the Membrane-Embedded Site of a G-Protein-Coupled Receptor. *J. Chem. Theor. Comput.* 14, 2761–2770. doi:10.1021/acs.jctc.8b00046

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Raniolo and Limongelli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Cysteine Variant at an Allosteric Site Alters MIF Dynamics and Biological Function in Homo- and Heterotrimeric Assemblies

Erin Skeens^{1†}, Georgios Pantouris^{2,3†}, Dilip Shah⁴, Ramu Manjula², Michael J. Ombrello⁵, N. Karl Maluf⁶, Vineet Bhandari⁴, George P. Lisi^{1*} and Elias J. Lolis^{2*}

¹Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, RI, United States, ²Department of Pharmacology, Yale University School of Medicine, New Haven, CT, United States, ³Department of Chemistry, University of the Pacific, Stockton, CA, United States, ⁴Section of Neonatology, Department of Pediatrics, Cooper University Hospital, Camden, NJ, United States, ⁵Translational Genetics and Genomic Unit, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, MD, United States, ⁶KBI Biopharma, Louisville, CO, United States

OPEN ACCESS

Edited by:

Ravi Pratap Barnwal,
Panjab University, India

Reviewed by:

Conrado Pedebos,
University of Oxford, United Kingdom
Haiguang Liu,
Beijing Computational Science
Research Center (CSRC), China

*Correspondence:

George P. Lisi
george_lisi@brown.edu
Elias J. Lolis
elias.lolis@yale.edu

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 26 September 2021

Accepted: 12 January 2022

Published: 08 February 2022

Citation:

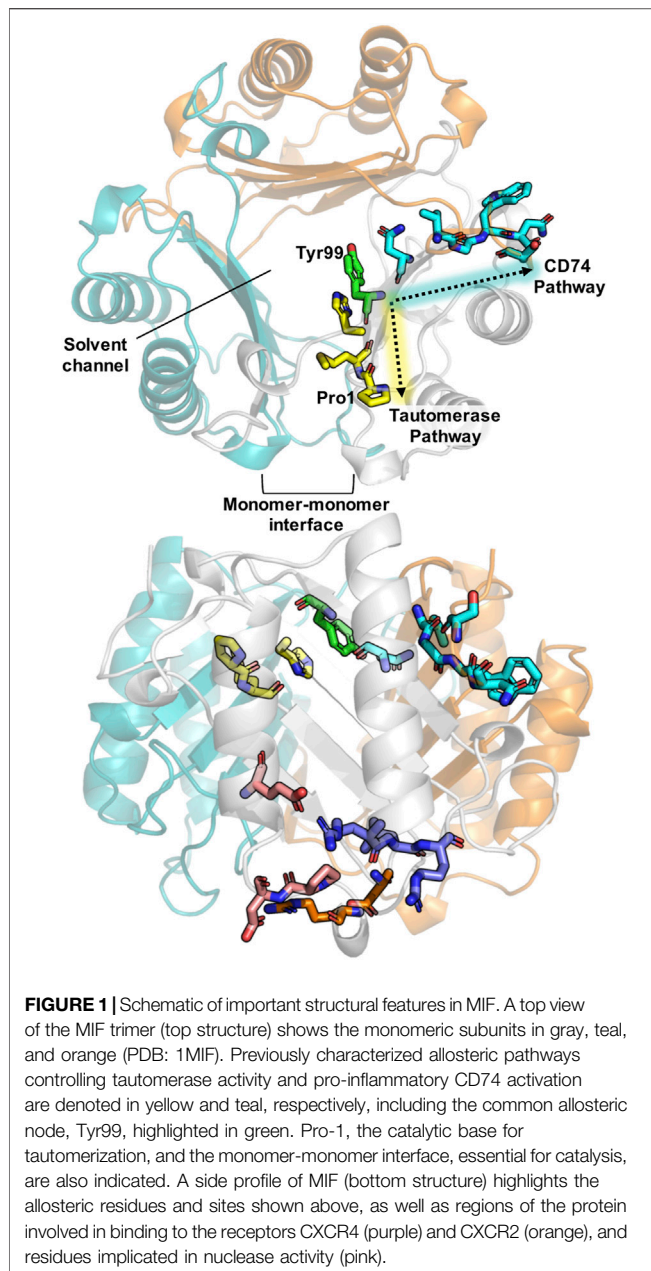
Skeens E, Pantouris G, Shah D,
Manjula R, Ombrello MJ, Maluf NK,
Bhandari V, Lisi GP and Lolis EJ (2022)
A Cysteine Variant at an Allosteric Site
Alters MIF Dynamics and Biological
Function in Homo- and
Heterotrimeric Assemblies.
Front. Mol. Biosci. 9:783669.
doi: 10.3389/fmolb.2022.783669

Macrophage migration inhibitory factor (MIF) is an inflammatory protein with various non-overlapping functions. It is not only conserved in mammals, but it is found in parasites, fish, and plants. Human MIF is a homotrimer with an enzymatic cavity between two subunits with Pro1 as a catalytic base, activates the receptors CD74, CXCR2, and CXCR4, has functional interactions in the cytosol, and is reported to be a nuclease. There is a solvent channel down its 3-fold axis with a recently identified gating residue as an allosteric site important for regulating, to different extents, the enzymatic activity and CD74 binding and signaling. In this study we explore the consequence of converting the allosteric residue Tyr99 to cysteine (Y99C) and characterize its crystallographic structure, NMR dynamics, stability, CD74 function, and enzymatic activity. In addition to the homotrimeric variant, we develop strategies for expressing and purifying a heterotrimeric variant consisting of mixed wild type and Y99C for characterization of the allosteric site to provide more insight.

Keywords: macrophage migration inhibitory factor, allosteric site, mutagenesis, biophysics, biological activities

INTRODUCTION

Macrophage migration inhibitory factor (MIF) is a pro-inflammatory cytokine involved in multiple non-overlapping physiological pathways (Bernhagen et al., 2007). It is expressed in cells from a variety of tissues with pronounced expression in the cells of the immune system. MIF forms a homotrimeric structure with a solvent channel coincident with its 3-fold axis and possesses three putative catalytic cavities formed at the interface between each of two adjacent subunits (Sun et al., 1996; Suzuki et al., 1996). These were first identified by the similarity of the MIF structure to two microbial enzymes that catalyzed isomerase or tautomerase reactions with a conserved and unusual N-terminal proline functioning as a catalytic base (Stivers et al., 1996a; Stivers et al., 1996b; Subramanya et al., 1996). These enzymes and MIF have 5–20% sequence homology but share an N-terminal proline that is evolutionarily invariant for MIF (Swope et al., 1998). The *bona fide* substrate for MIF has yet to be identified, but “pseudosubstrates” D-dopachrome and hydroxyphenylpyruvate have been useful for designing inhibitors or optimizing high throughput screens (Rosengren et al., 1996; Rosengren et al., 1997; Lubetsky et al., 2002; Cho et al., 2011). These inhibitors were used as tools to probe MIF biological activities and signaling, including functional



interactions with p53, ERK-1/2, PI3K/Akt in the cytosol (Hudson et al., 1999; Mitchell et al., 1999; Li et al., 2009), and the extracellular receptors CD74, CXCR2, and CXCR4 (Leng et al., 2003; Bernhagen et al., 2007; Weber et al., 2008; Rajasekaran et al., 2016; Lacy et al., 2018). More recently, a nuclease function was ascribed to MIF in a process known as parthanatos (Wang et al., 2016; Wang et al., 2021).

The mechanisms by which MIF promotes these activities are unclear, though signaling studies, biochemistry, and structures have provided some molecular details of possible mechanisms. A hypothesis that trimeric MIF functions as an enzyme at higher concentrations in the cytosol and subsequently dissociates into monomers when it is diluted in the extracellular milieu to activate

receptors was disproven. Sedimentation velocity and equilibrium ultracentrifugation experiments found MIF has an unusual partial volume for a protein and forms an extremely tight trimer (Philo et al., 2004). Up to four structure-based mutations at the subunit interface also failed to disrupt the homotrimer; MIF with a fifth mutation did not express or was extremely unstable (Fan et al., 2013). Despite the tight homotrimeric association of MIF, it displays remarkable dynamics by NMR (Cho et al., 2010; Crichlow et al., 2012; Pantouris et al., 2018; Pantouris et al., 2020). Recently, we discovered an allosteric regulatory site at the center of the symmetric MIF homotrimer that serves as a hub for two of its biochemical activities: the enzymatic enol-keto tautomerase and CD74 receptor activation (Pantouris et al., 2018; Pantouris et al., 2020) (**Figure 1**). The allosteric residues, Tyr99 from all three subunits, modulate enzymatic and CD74 receptor activation to different extents when mutated, while concurrently altering the structural and dynamic properties of the protein, as well as hydrogen bonding with solvent molecules in the central solvent channel. Notably, the size of the solvent channel opening and dynamics at the allosteric site were found to be critical diagnostics of functional outcomes, though it is still unclear exactly how the chemical properties of the allosteric site at Tyr99 or the solvent channel influence the biophysical signatures of MIF. Further insights would facilitate the rational design of site-specifically modified allosteric pathways to regulate the information flow to-and-from Tyr99, creating enhanced spatial and temporal resolution of MIF function.

To investigate this relationship further, we introduced a cysteine residue at position 99 (Y99C). The sulfhydryl group of Cys99 is the same molecular distance (~ 8 Å) from the enzymatic active site as the C γ of Tyr99. We hypothesized that this mutation would enable disulfide stabilized dimers within the MIF trimers, providing a more comprehensive understanding of the structural, dynamic, and functional impact of Cys99 substitution. This method was used in a previous MIF study, when a cysteine mutation was introduced at N110 and formed an inter subunit disulfide with a Cys80 and induced changes in the oligomeric and functional properties (Fan et al., 2013). We generated a unique Y99C MIF cysteine variant, as well as a method to generate homo- and hetero-trimeric Cys proteins, and characterized their molecular structure, dynamics, and effect on enzymatic and CD74 activities using X-ray crystallography, NMR spectroscopy, biochemical assays, and *in vivo* experiments.

MATERIALS AND METHODS

MIF Mutagenesis and Protein Purification

The Y99C MIF cDNA was produced by site-directed mutagenesis using the PfuTurbo DNA polymerase (Agilent). DpnI was utilized to digest the parental pET-11b plasmid with the WT MIF cDNA that contained methylated and hemimethylated DNA (Pantouris et al., 2018). BL21-Gold (DE3) cells (Agilent) were transformed with the sequenced Y99C MIF plasmid and selected for on LB agar containing 100 μ g/ml Amp (LB/Amp). WT MIF or

Y99C MIF-containing cells were grown in 1L LB/Amp at 37°C, and induced by 1 mM of isopropyl β -D-1-thiogalactopyranoside (IPTG) at an OD₆₀₀ of 0.6. Levels of expression were assessed using a NuPAGE 4–12% Bis-Tris gels 4 h after addition of IPTG. The cells were centrifuged and stored in –80°C until further use. Frozen cells were thawed on ice and dissolved in 20 mM Tris, 20 mM NaCl, pH 7.4 containing a mini EDTA-free protease inhibitor cocktail tablet (Sigma-Aldrich) and lysed by sonication. After centrifugation at 4°C, the filtered supernatant was loaded onto Q-Sepharose and SP ion-exchange chromatography columns connected in series. MIF mutant was in the flow-through with ~95% purity. After concentrating the sample using 10 kDa MWCO Amicon centrifugal filters (Millipore), the protein was loaded onto a 16/60 Superdex 75 size-exclusion column to remove the remaining ~5% of contaminants with 20 mM Tris, pH 7.4, with 20 mM NaCl as the running buffer. Protein concentration was determined using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific).

Y99C/WT MIF Mixed Trimer: Plasmid Design, Expression, and Purification

The Y99C MIF sequence was synthesized and cloned into the pET-24b (+) vector (Kan-resistance) with a C-terminal TEV protease cleavage site, followed by a SNAP-tag (19.4 kDa) and a His₆-tag (GenScript United States, Inc.). The Y99C MIF/SNAP-tag/His₆-tag plasmid was co-transformed with the WT MIF plasmid (pET-11b; Amp-resistance) into BL21-Gold (DE3) competent cells (Agilent) and selected for on LB agar plates containing 1.0 mg/ml Amp and 0.75 mg/ml Kan. 1 L LB with 1.0 mg/ml ampicillin and 0.75 mg/ml kanamycin was inoculated with a colony containing the WT MIF and the Y99C MIF/SNAP/His₆ plasmids, and grown at 37°C to an OD₆₀₀ of 0.6–0.8. Cells were cooled on ice for 30 min, then induced with 0.2 mM IPTG and incubated for 16 h at 18°C. The cells were harvested by centrifugation and stored at –80°C.

Cells were resuspended in 25 ml lysis buffer (20 mM sodium phosphate, 300 mM NaCl, and 5 mM imidazole at pH 7.4) with a mini EDTA-free protease inhibitor cocktail tablet, lysed by sonication, and centrifuged to remove cell debris. Y99C/WT MIF mixed trimers were purified by Ni-NTA affinity chromatography, such that WT MIF homotrimers, containing no His₆-tags, elute from the column in the flow through, while Y99C/WT MIF mixed trimers containing 1–3 monomers of the Y99C MIF/SNAP-tag/His₆-tag construct are bound to the Ni-NTA resin (GoldBio) and subsequently washed with 50 ml of 20 mM sodium phosphate, 300 mM NaCl, and 25 mM imidazole buffer at pH 7.4 and eluted with 15 ml of 20 mM sodium phosphate, 300 mM NaCl, and 500 mM imidazole buffer at pH 7.4. To confirm the presence of Y99C/WT MIF mixed trimers, the protein was purified over a HiLoad 26/600 Superdex 200 size exclusion column (GE Healthcare) in 20 mM sodium phosphate, 150 mM sodium chloride, and 1 mM EDTA at pH 7.4. The SNAP-tag and His₆-tag were then cleaved with TEV protease and subsequently removed via Ni-NTA chromatography.

SDS-PAGE band intensities were analysed using densitometry via ImageJ software (NIH) (Schneider et al., 2012). Briefly, the optical density of each band was determined by the software, with band intensities calculated as the area of the resulting optical density peaks. Each lane was analysed separately due to differing protein concentration per sample. The MIF monomer and MIF monomer + SNAP-tag bands were compared relative to each other and normalized such that the MIF monomer + SNAP-tag band was equal to 1.00.

Analytical Ultracentrifugation (AUC) of the Y99C MIF Homotrimer Using Sedimentation Velocity

Y99C MIF homotrimer samples were loaded into cells with 2-channel charcoal-epon centerpieces with 12 mm optical pathlength. Samples were analyzed in PBS (from Corning). The reference buffer (PBS) was loaded into the reference channel of each cell to serve as an optical reference blank. The cells were then loaded into an AN-60Ti analytical rotor, which was then placed into a Beckman-Coulter ProteomeLab XL-A analytical ultracentrifuge. After thermal equilibrium was established at 20°C, the samples were scanned at 3,000 RPM at 230 nm to confirm proper cell loading. The rotor was stopped, and then accelerated to the final run speed of 50,000 rpm. Scans were recorded at this rotor speed every ~4 min for 110 total scans for each sample (~7.3 h total sedimentation time).

The raw data were analyzed using the SEDFIT software, which implements the *c(s)* method developed by Peter Schuck at the N.I.H. (Schuck, 2000). In this approach the raw data scans are analyzed by non-linear least squares to determine the sedimentation coefficient distribution associated with the sample. This method models the influence of diffusion on the data in order to enhance the resolution of the corresponding size distribution. A maximum entropy regularization probability of 0.683 (1 σ) was used, and time-invariant noise was removed. The resultant size distributions were graphed, and the peaks were integrated using OriginLab Origin[®] version 7.

CD Spectroscopy

CD spectra and thermal unfolding experiments were acquired on a JASCO J-815 spectropolarimeter equipped with a variable temperature Peltier device and using a 2 mm quartz cuvette. CD spectra were collected at 25°C and denaturation curves were recorded at 218 nm over a temperature range of 20–100°C (293K–373K). *T_m* for WT MIF and MIF variants were determined via nonlinear curve fitting using GraphPad Prism.

Crystallization of Y99C MIF, Structure Determination, and Characterization of Solvent Channel

Using the Formulatrix NT8 robot, 400 nl of 18 mg/ml protein was mixed with 200 nl reservoir solution (2 M ammonium sulfate, 3% 2-propanol, 0.1 M Tris-HCl, pH 7.5) in one drop and set to crystallization at 20°C using a hanging drop vapor diffusion method with a reservoir volume of 100 μ L. Crystallization was

observed with a RI-1000 Formulatrix Imager, and crystals grew to their maximum size in 30 days. Crystals were then flash-frozen in liquid N₂ with 15% glycerol which acts as a cryoprotectant during the data collection. X-ray crystallographic data sets were collected at 100 K using a Rigaku-007 Micromax Generator with a PILATUS Dectris 200K Pixel Array Detector and AFC 4-axis goniometer. Diffraction data were integrated and scaled using HKL3000 (Minor et al., 2006). The crystal structure was solved by molecular replacement method using the MIF monomer structure (PDB ID:2OOW) and PHASER (McCoy et al., 2007). The structure solution yielded three monomers of MIF in the asymmetric unit. The refinement of the structure was performed using the module *Phenix.refine* of the *PHENIX* package (Adams et al., 2010). Cycles of refinement and model building were done using REFMAC (Winn et al., 2011) and Coot (Emsley and Cowtan, 2004). The stereochemistry of these crystal structures was assessed with *MOLPROBITY* (Chen et al., 2010). The X-ray data collection, scaling, and refinement statistics are summarized in **Supplementary Table S1**. Figures of the solvent channels and the minimal radius for WT, Y99C, and Y99A were calculated with MOLE 2.0 (Sehna et al., 2013) and MoleOnline (Pravda et al., 2018), respectively.

NMR Spectroscopy

Isotopically labeled MIF samples for NMR studies were grown in 1 L M9 minimal media supplemented with glucose (¹²C₆H₁₂O₆, natural abundance) and ammonium chloride (¹⁵NH₄Cl, Cambridge Isotope Labs) as the sole carbon and nitrogen sources, respectively. The ¹⁵N-MIF samples were expressed and purified as described above, then dialyzed into 20 mM sodium phosphate, 1 mM EDTA, and 7.5% D₂O at pH 7.0 and concentrated to 0.5–1.0 mM with a 10 kDa MWCO Amicon centrifugal filter (Millipore).

NMR experiments were carried out on a Bruker Avance NEO 600 MHz spectrometer at 30°C. ¹H-¹⁵N TROSY HSQC and relaxation data were collected with the ¹H and ¹⁵N carriers set to the water resonance and 120 ppm, respectively. NMR spin relaxation experiments were performed using TROSY-based pulse sequences adapted from Palmer and coworkers (Loria et al., 1999). Longitudinal relaxation rates (*R*₁) were measured with *T*₁ delays of 20, 60, 100, 200, 600, 800, and 1,200 ms. Transverse relaxation rates (*R*₂) were collected with 1.0 ms spacing between 180° CPMG pulses at total relaxation delays of 1, 2, 4, 8, 10, and 12 ms. The recycle delay in these experiments was 2.5 s. Longitudinal and transverse relaxation rates were extracted by non-linear least squares fitting of the peak heights (major peaks in cases of slow exchange) to a single exponential decay using in-house software. Uncertainties in these rates were determined from replicate spectra. The heteronuclear cross-relaxation rate (¹H-¹⁵N) NOE) was obtained by interleaving pulse sequences with and without proton saturation and calculated from the ratio of peak heights from these experiments. NMR data were processed using NMRPipe (Delaglio et al., 1995) and analysed in Sparky (Lee et al., 2015) along with in-house scripts.

In vivo Neutrophil Recruitment Assays

Wild-type mice of genetic background strain (C57BL/6J) were purchased from the Jackson Laboratory (Bar Harbor, ME) and housed in a pathogen-free animal facility at Cooper University Healthcare. All experiments were done in 10–12 weeks old male mice. Mice were administered a one-time intratracheal instillation of 100 µl of normal saline alone (vehicle) or 10 µg/ml of WT-MIF or Y99C MIF in a normal saline solution. Mice were sacrificed after 6 h of vehicle only or vehicle + experimental agent administration and bronchoalveolar lavage (BAL) fluid was collected. BAL collection was performed by cannulating the trachea with a blunt 22-gauge needle and lavaging the lungs with 800 µl of pre-cooled sterile PBS solution. Total protein concentration in the BAL fluid was measured using the Pierce™ BCA assay kit (Thermo Scientific, Rockford, IL), as previously described (Shah et al., 2019). The animal study protocol was approved by the Institutional Animal Care and Use Committee of Cooper University Healthcare, Camden, NJ.

Steady-State Kinetics

Michaelis-Menten kinetics were performed as previously described (Pantouris et al., 2020). Briefly, a 100 mM stock of 4-hydroxyphenyl pyruvate (4-HPP) in 0.5 M ammonium acetate, pH 6.2 was prepared by overnight equilibration to generate the predominant keto form of 4-HPP. Enzymatic assays used 96-well plates containing a mixture of 10 µL of the keto form of HPP at concentrations ranging from 0–2 mM and 130 µL borate at a final concentration 0.420 M per well. The reaction was initiated with 50 nM protein. The concentration of the enol, which forms an enol-borate complex, was to measure at 306 nm ($\epsilon_{306} = 11,400 \text{ M}^{-1} \text{ cm}^{-1}$) every 10 seconds for a total of 1.5 min using a Tecan Infinite M200. Each reaction was conducted in triplicate per protein sample.

RESULTS

Biophysical Characterization of Y99C and Y99C/WT MIF Trimers

Since Tyr99 is located the end of a β -strand of the MIF structure, we speculated it would take local unfolding to move the Cys residues in proximity to form disulfide linkages. We also set out to generate the first mixed trimer species of MIF, with the rationale of using the steric bulk of a Tyr99 side chain to constrain other Cys99 residues into forming disulfide bonds. This approach is also a proof-of-principle demonstration for the engineering of novel MIF variants with structural scaffolds and activities distinct from WT MIF and homotrimeric variants. While the Y99C MIF homotrimer displays a very similar size exclusion chromatographic profile to that of WT MIF, purification of the Y99C/MIF mixed trimers, facilitated by a SNAP-tag and a His₆-tag on Y99C monomeric subunits, shows distinct MIF species with differing molecular weights that can be followed by SDS-PAGE (**Figures 2A,B**). The MIF monomer band (12.4 kDa) increases in intensity, relative to the MIF monomer/SNAP-tag band (31.8 kDa), from peak 1 (P1) through peak 3 (P3) on the size exclusion chromatogram,

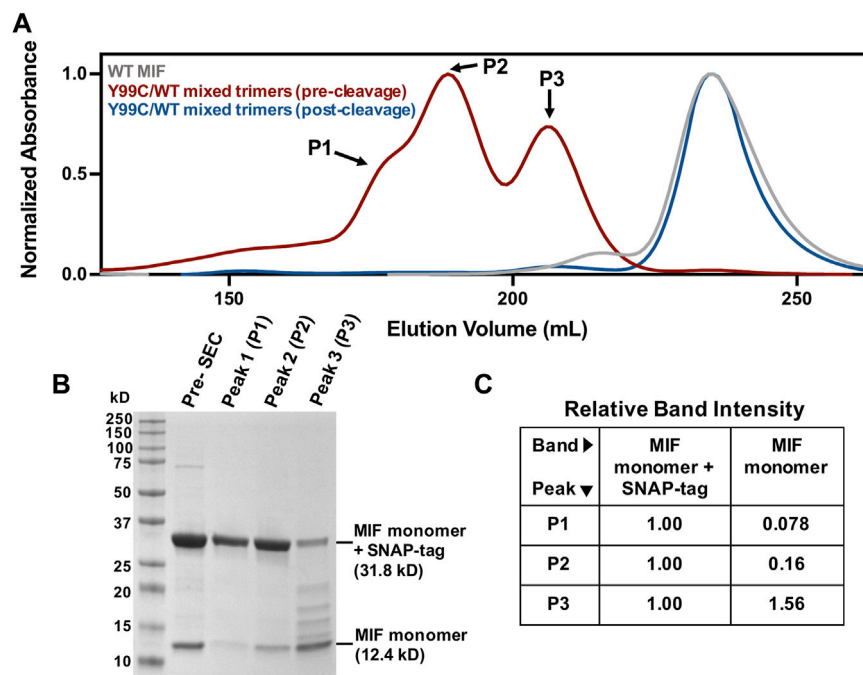


FIGURE 2 | Verification of novel Y99C/WT mixed trimers **(A)** Chromatographic overlay of WT MIF and the Y99C/WT mixed trimers pre- and post-tag cleavage as purified by size exclusion chromatography (SEC) **(B)** SDS-PAGE analysis of the Y99C/WT mixed trimer species showing the presence of the MIF monomer (12.4 kDa) and the MIF monomer + SNAP-tag (31.8 kDa) in peaks 1-3 **(C)** Relative SDS-PAGE band intensities of the MIF monomer and the MIF monomer + SNAP-tag bands from **(B)** via ImageJ software (NIH). Each lane was analyzed separately and normalized such that the MIF monomer + SNAP-tag band is equal to 1.00.

representative of MIF species with increasing numbers of WT MIF monomers within the MIF trimer (**Figure 2C**). Thus, P1 is likely the Y99C homotrimer (96.1 kDa with SNAP-tags), P2 is the Y99C/WT heterotrimer with two Y99C monomers and one WT monomer (76.7 kDa with SNAP-tags), and P3 is the Y99C/WT heterotrimer with one Y99C monomer and two WT monomers (57.3 kDa with SNAP-tag). When P1-P3 are pooled and cleaved of the SNAP-tags and His₆-tags, the Y99C/WT mixed trimers elute at the same volume as WT MIF at 37.9 kDa.

The Y99C MIF homotrimer was probed by analytical ultracentrifugation at three loading concentrations (**Supplementary Figure S1**). The sedimentation coefficient of the Y99C homotrimer is in good agreement with published values for WT MIF (Philo et al., 2004). The molar mass of 37.8 ± 0.8 kDa calculated directly from the primary data agrees very well that of WT MIF (37.9 ± 2.8 g/mol per trimer). Thus, the Y99C variant is a trimer and also has the same elevated partial specific volume seen for WT MIF (within error). With these data, we see no differences between WT MIF and the Y99C MIF homotrimer in terms of oligomeric state, shape, or partial specific volume.

X-Ray Crystallography of the Y99C MIF Homotrimer

The structure of homotrimeric Y99C MIF was solved to 1.60 Å to probe whether there were sufficient protein dynamics for the Cys99 side chain to form a disulfide and to detect any unexpected effects. The structural alignment of WT MIF, Y99A, and Y99C

were calculated and viewed in PyMol (**Figure 3A**). From this alignment, the residue 99 sidechains indicate the C β carbons overlap very well (**Figure 3A**, right panel). Interestingly, the thiol group in all three monomers has two conformations in the channel as shown in **Figure 3A** where Y99 and Y99A are also shown. The 2Fo-Fc and omit map of Y99C is also shown (**Figure 3B**, **Supplementary Figure S2**). The closest distances among the thiol groups of adjacent subunits are ~ 6.5 Å (**Figure 3B**). The two conformations of the thiol groups in this study do not interact with other protein atoms nor do they induce large conformational changes (**Figure 3C**). The RMSD of Y99C to WT MIF (3DJH) and the Y99A (5EIZ) structures are 0.134 Å and 0.077 Å, respectively, and support a lack of sufficient movement of Y99C to induce conformational changes. Though the shared C β atom for WT, Y99A, Y99C MIF align very well, the radius of the solvent channel opening where the three Y99C residues appear is 3.4 Å and similar to Y99A (3.6 Å), in contrast to WT MIF (1.4 Å) as calculated by MOLEonline (Pravda et al., 2018). In a previous study involving three Tyr99 variants and four His62 mutations, there was an inverse correlation between the size of the opening and catalytic efficiency (Pantouris et al., 2020).

The MIF Structure Is Destabilized by Cys99 Residues

To assess the impact of the Cys99 mutation on the secondary structure and stability of MIF, far-UV circular dichroism (CD)

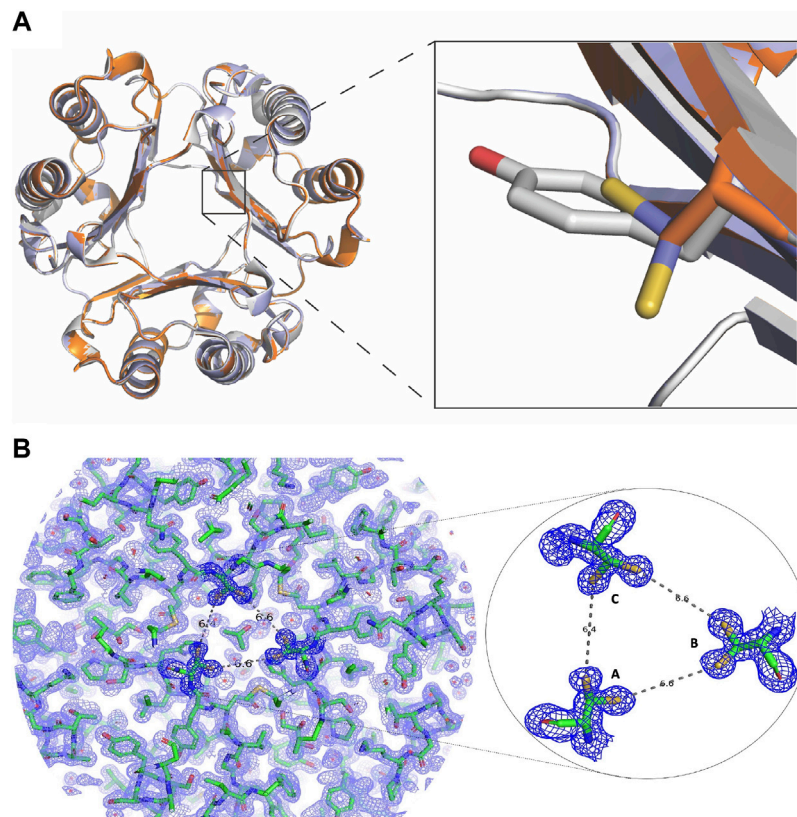


FIGURE 3 | Changes in the MIF structure with mutations at Tyr99 **(A)** The crystal structure of Y99C MIF has two thiol group conformations per monomer in the solvent channel (PDB: 7KQX). The inset shows overlapping side chains for residue 99 in WT MIF (3DJH; gray) and the Y99C (7KQX; blue) and Y99A (5EIZ; orange) MIF variants **(B)** Electron density map for MIF Y99C homotrimer. On the left, the 2Fo-Fc electron density map for the Y99C MIF homotrimer is shown. On the right, the 2Fo-Fc map and the two conformations for the thiol groups of Y99C are enlarged. The distances among the closest sulfur atoms from each subunit are over 6 Å apart and too far from each other to form disulfides.

experiments were conducted. CD spectra of the Y99C homotrimer and Y99C/WT mixed trimers display α - β secondary structure consistent with that of WT MIF, with only subtle differences observed (**Supplementary Figure S3A**), which is in good agreement with the crystallographic analysis of the Y99C homotrimer. Similarly, thermal unfolding experiments reveal that the Y99C mutation has a destabilizing effect on the MIF structure, with T_m values for the Y99C homotrimer and Y99C/MIF mixed trimers of 76.9°C ($\Delta T_m = -1.6^\circ\text{C}$) and 77.5°C ($\Delta T_m = -1.0^\circ\text{C}$), respectively, relative to WT MIF ($T_m = 78.5^\circ\text{C}$). (**Supplementary Figure S3B**). Together, these data suggest that MIF becomes increasingly destabilized as the number of Cys99 residues increases per trimer.

NMR Spectroscopy Highlights Altered Solution Structures and Dynamics of Y99C Variants

We recently employed NMR spectroscopy to characterize significant solution structural and dynamic changes associated with mutations at the Tyr99 allosteric site that are not obvious through crystallographic analysis (Pantouris et al., 2020). Here, we find that the NMR spectrum of the Y99C homotrimer is

distinct from that of WT MIF, with significant chemical shift perturbations observed in residues at or near sites previously implicated in the allosteric signaling that controls MIF function (**Figure 4**). Chemical shift perturbations of residues at the N-terminus and proximal to His62 are consistent with prior studies of Tyr99 variants that confirmed the allosteric pathway (Tyr99-His62-Met2-Pro1) that modulates MIF enzymatic activity (Pantouris et al., 2020). We also observed significant perturbations at the C-terminus, where the CD74 receptor activation site has been mapped (**Figure 4**) (Pantouris et al., 2015; Pantouris et al., 2018). Comparatively, the NMR spectrum of the Y99C/WT mixed trimers reveals chemical shift perturbations in the same regions of the MIF structure as the Y99C homotrimer, though to a lesser extent. The highly symmetric WT MIF appears as a 114-residue monomer without the cleaved initiating methionine with a molecular weight of 12,388 Da protein by NMR, although it is actually a 37,164 Da trimer. While the heterozygous incorporation of mutations into MIF alters its local chemical environment, it does not perturb the global symmetry of MIF to an extent that renders its NMR spectrum more difficult to interpret (**Figure 4A**). Although the similarity of the NMR spectra indicate that mixed trimers also do not facilitate disulfide

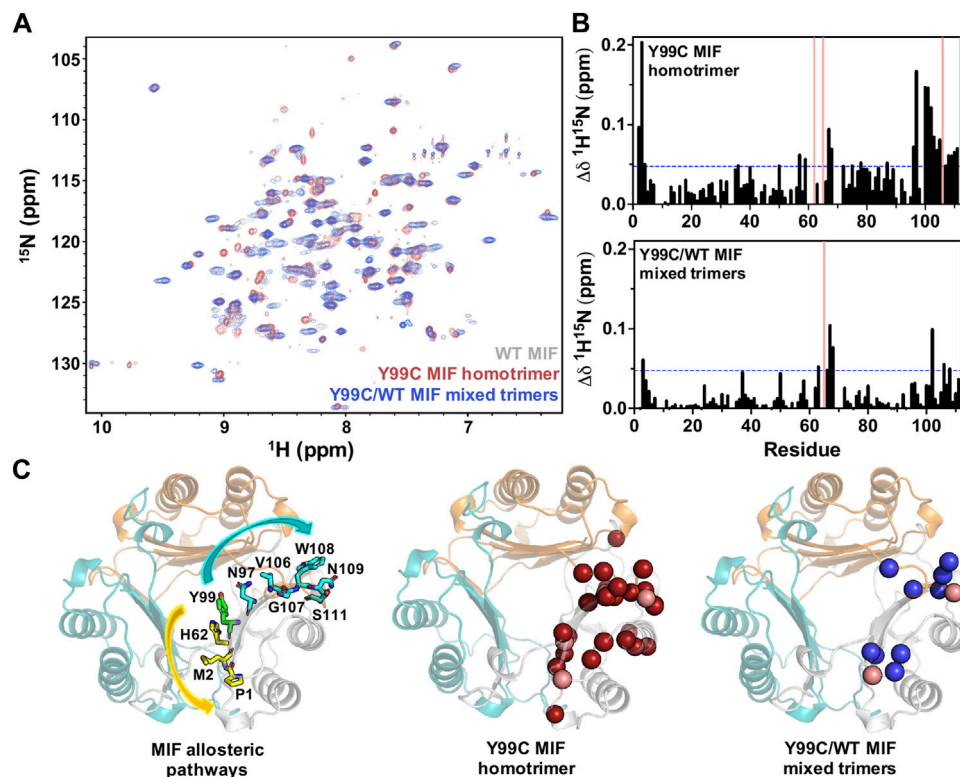


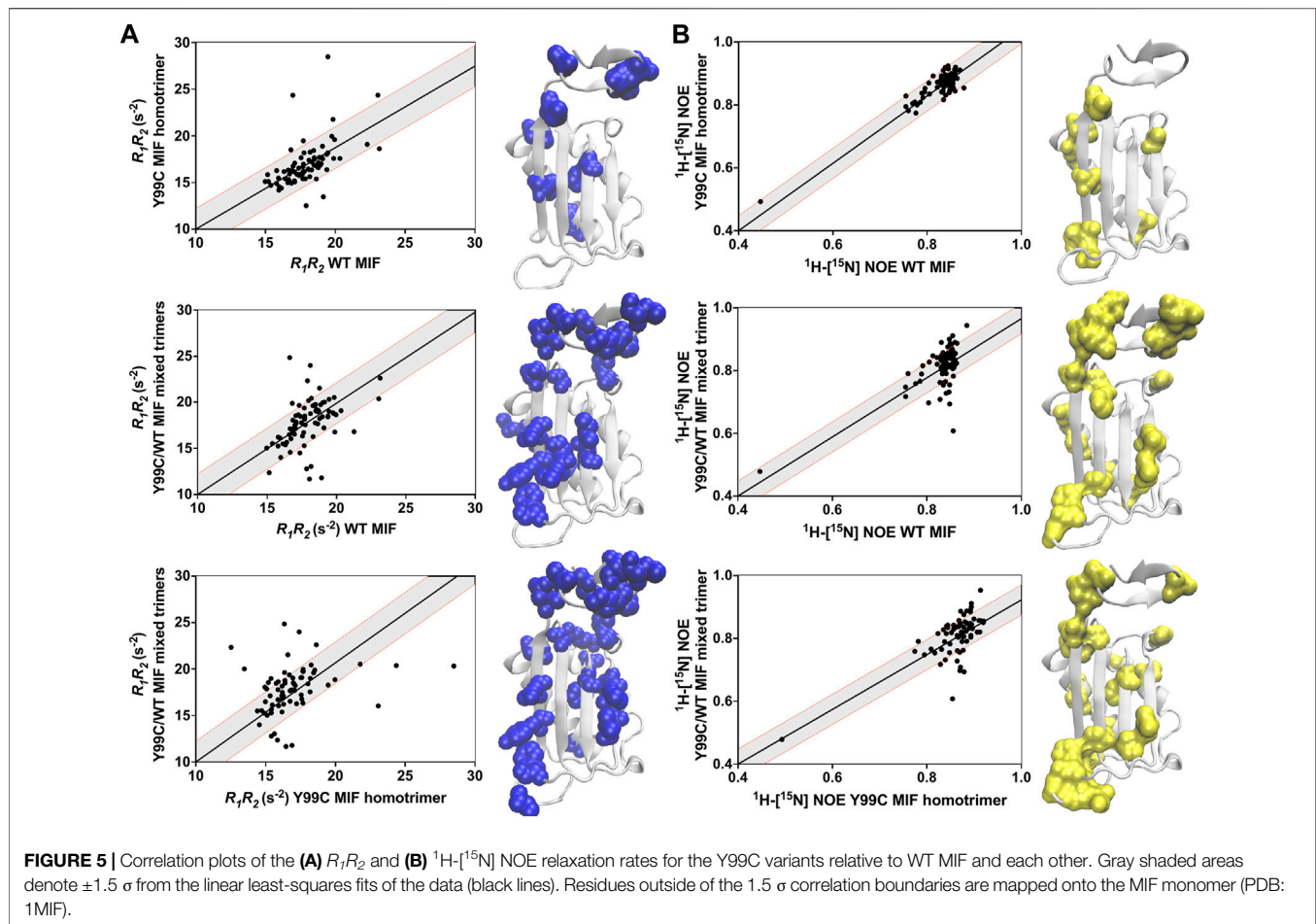
FIGURE 4 | (A) ^1H - ^{15}N TROSY HSQC spectral overlay of WT MIF (gray), the Y99C homotrimer (red), and the Y99C/WT mixed trimers (blue) **(B)** ^1H - ^{15}N combined chemical shift perturbations ($\Delta\delta$) for the Y99C homotrimer (top) and the Y99C/WT mixed trimers (bottom). Pink bars denote residues that are broadened beyond detection, and blue lines represent 1.5σ above the 10% trimmed mean of all shifts **(C)** Significant chemical shift perturbations observed for the Y99C homotrimer (middle, red spheres) and Y99C/WT mixed trimers (right, blue spheres) are mapped onto the MIF structure (PDB: 1MIF) for comparison to the previously characterized MIF allosteric pathways (left) affecting tautomerase activity (yellow) and CD74 activation (teal). Pink spheres denote residues broadened beyond detection. Residues after position 99 are mapped onto the adjacent monomer, as reported for the CD74 activation site.

formation, this is the first evidence that heterozygous MIF samples can be studied with molecular detail in solution.

In order to assess the impact of the Y99C mutation on dynamically-driven signaling in MIF, we quantified longitudinal (R_1) and transverse (R_2) relaxation rates and the heteronuclear ^1H - ^{15}N NOE to report fluctuations in the protein backbone (**Supplementary Figure S4**). We postulated that the dynamic profile of the Y99C/WT mixed trimers would be “in between” that of WT MIF and the Y99C homotrimer, as was observed structurally by chemical shift perturbations of the Y99C variants (**Figure 4B**). Interestingly, we observe a greater overall deviation from WT-like dynamics in the Y99C/WT mixed trimers. In fact, there are limited differences in flexibility between WT MIF and the Y99C homotrimer. Only a few sites of significantly altered R_1 and R_2 relaxation rates for the Y99C homotrimer are observed, including Ile67, proximal to the His62 allosteric site, and Ile96 and Ala104, proximal to the Tyr99 allosteric site and the CD74 binding site (**Supplementary Figure S4,5**). Further, the ^1H - ^{15}N NOE for the Y99C homotrimer reveals a slight global elevation relative to WT MIF, suggesting that the Y99C homotrimer is actually less flexible overall than WT MIF. In contrast, there is a large variance in the dynamic profile of the Y99C/WT mixed

trimers. Regions with altered R_1 and R_2 relaxation rates include residues at the monomer-monomer interface (3, 5, 32, 36, 49), residues proximal to the His62 allosteric site (63, 67), and residues near Tyr99 and the CD74 allosteric pathway (95–98, 100–102, 107, 113). The increased flexibility of the Y99C/WT mixed trimers is most apparent in the ^1H - ^{15}N NOE, where a global depression of NOE values is observed, including significant deviations from WT MIF in functionally relevant regions of the protein similar to those observed by R_1 and R_2 relaxation rates.

We compared the dynamics of the Y99C homotrimer and Y99C/WT mixed trimers to those of WT MIF, and to each other, using correlation plots of the R_1R_2 relaxation rates (suggestive of μs – ms molecular motions), analyzed as a product to minimize anisotropic molecular tumbling artefacts and ^1H - ^{15}N NOEs, reporting on ps–ns dynamics (**Figure 5**). Deviations from linearity in these plots denote substantial differences between the samples, allowing us to pinpoint specific residues affected by disruption of the allosteric site. There is a much lower correlation between the dynamic profiles of WT MIF and the Y99C/WT mixed trimers, indicating that the Y99C/WT mixed trimers exhibit altered flexibility to a much greater extent than the Y99C homotrimers, consistent with our analysis of the raw data. Further, the dynamics of the Y99C variants are distinct



from each other, even more so than either variant relative to WT MIF. Residues outside of the 1.5σ correlation boundaries, mapped onto the MIF monomer in **Figure 5**, highlight regions of altered flexibility suggestive of motion on multiple timescales. Most notably, we observe dynamic changes at the C-terminus and the solvent channel β -sheet, regions that allosterically mediate MIF function (**Figure 4C**), suggesting that the Y99C/WT mixed trimers may be functionally distinct from WT MIF and the Y99C homotrimers. The heightened flexibility observed for the Y99C/WT mixed trimers may be, in part, due to the heterogeneity of the Y99C/WT mixed trimers, as the subtle loss of symmetry of the heterotrimers may disrupt stabilizing interactions, despite maintaining a relatively similar overall MIF structure by NMR.

The Y99C Mutation Modulates MIF Enzymatic Activity and Ability to Activate CD74 *in vivo*

The stimulation of neutrophil recruitment in murine lungs is tied to CD74 receptor activation *in vivo* (Takahashi et al., 2009). Our previous work with a Y99A variant showed significant abrogation of neutrophil recruiting activity, suggesting a coupling of this allosteric residue to the C-terminal CD74 binding site (Pantouris et al., 2018). We measured the percentage of neutrophils and total

protein content as biomarkers for CD74 activity and alveolar-capillary leak/pulmonary edema, respectively, in bronchoalveolar lavage (BAL) fluid obtained from mice sacrificed after 6 h of administration of a vehicle (saline) only or vehicle with Y99C MIF. The neutrophil influx (**Supplementary Figure S5**) was significantly increased in the WT MIF group (compared to controls), but significantly decreased in groups administered the Y99C homotrimeric variant, similar to Y99A (Pantouris et al., 2018). The decrease in lung inflammation is likely due to the same factor as the Y99A variant; the absence of Tyr99 and the inability of the variant to undergo correlated motions with a number of proximal and distal residues originally shown by molecular simulations (Pantouris et al., 2018). NMR studies of Y99C (**Figure 4A,B**, **Figure 5**) show perturbation in residues Ile64, Lys66, Trp108, and Asn109, which are involved in CD74 receptor binding and activation. We also cannot eliminate the two sidechain orientations of Cys99, the local changes it causes at receptor binding residues, and other residues in its proximity.

We previously noted that increasing the radius of the solvent channel opening by substituting Tyr99 with other residues is correlated to reduced catalytic activity (Pantouris et al., 2020). Though we observed a solvent channel opening for Y99C similar to that of Y99A, the catalytic activity of the Y99C homotrimers has reduced (55%) catalytic efficiency relative to WT MIF and

compared to Y99A (77%) (Pantouris et al., 2020) (**Supplementary Figure S6**). Our previous work evaluated a range of amino acid substitutions at position 99, including glycine, alanine, and phenylalanine. While the catalytic efficiencies of Y99G and Y99F were found to be 10 and 143%, respectively, compared to WT MIF, Y99A only exhibited a small decrease in catalytic activity to 77%, suggesting that amino acids with side chain sizes between the extremes of glycine and phenylalanine are less functionally predictable, and there are likely additional factors that influence activity. The catalytic efficiency of the Y99C/WT mixed trimers was predicted to be between that of the Y99C homotrimer and WT MIF, as there are fewer changes (mutations) compared to the Y99C homotrimer. Interestingly, the catalytic efficiency of the Y99C/WT mixed trimers is reduced to ~28% of WT MIF. This result may be attributed to the significant increase in flexibility observed by NMR for the Y99C/WT mixed trimers, particularly at the C-terminus near the monomer-monomer interface and in proximity to the allosteric sites on the solvent channel β -sheet, where signaling, ligand binding, and/or active site chemistry may be altered (**Supplementary Figure S4,5**). We observed limited differences in flexibility of WT MIF and the Y99C homotrimer, which may be a contributing factor in the higher activity of the Y99C homotrimer compared to the heterotrimeric variants. In addition to the altered dynamic profile of the Y99C/WT mixed trimers, the reduced enzymatic efficiency could be due to the importance of symmetry at the allosteric site and its effects in catalytic activity. The use of heterotrimers was thus important to suggest symmetry is also significant at the allosteric site.

DISCUSSION

The unexpected discovery of an allosteric site at the center of the MIF trimer has spawned protein engineering and small molecule studies aimed at modulating two non-overlapping functions of MIF from a single location. This approach highlights a strategy to make the catalytic activity, CD74 agonism and antagonism, and conformational ensemble of MIF more intuitively controllable. Mutagenesis studies employed here attempted to probe dynamics at the gating residue of the solvent channel. Engineering of an intersubunit disulfide via mutagenesis at a different residue (N110C) was used to study MIF aggregation by the compound ebselen via interaction with a natural MIF cysteine (Cys80) and the effect of a locked trimer on the catalytic and CD74 activity of MIF (Ouertatani-Sakouhi et al., 2010; Fan et al., 2013). The intersubunit disulfide induced global conformational changes, a higher-order oligomer, and decreased catalytic and CD74 activation (Fan et al., 2013). The X-ray structure of the Y99C MIF homotrimer did not reveal the presence of a disulfide between any two of three subunits. Interestingly, it did reveal two conformations for each thiol group (per monomer). The Y99C homotrimer abolished CD74 activation *in vivo* but possessed reduced enzymatic activity, supporting the role of Tyr99 as a common allosteric node within two distinct allosteric pathways. We also demonstrate in this study a proof-of-principle that MIF can be engineered to form Y99C/WT MIF heterotrimers with different structural and biophysical properties than WT MIF or

the corresponding Y99C homotrimeric variant. Surprisingly, the catalytic efficiency of the mixed heterotrimeric variants was lower than Y99C. While NMR studies reveal only slight differences in the dynamics of the Y99C homotrimer and WT MIF, the Y99C/WT mixed trimers studied here exhibit heightened flexibility at critical regions of the protein, as well as the absence of symmetry compared to the Y99C homotrimer, suggesting that the slight loss in symmetry of the heterotrimers may play an important role in the altered dynamics and functional effects of the Y99C/WT mixed trimers. Future studies will attempt to isolate the heterotrimers to study their structures and dynamics individually.

The advance in engineering heterotrimeric MIF species will allow us to characterize the biological activities and biophysical properties of this and other heterotrimeric mutants to provide insight into how this protein carries out its multi-functional activities. It is noteworthy that two other examples of heterotrimeric proteins exist for use as an anti-inflammatory, and in the other case, as a causative agent for inflammatory disease. A strategy of structure-based design was used to convert the inflammatory homotrimeric tumor necrosis factor (TNF) protein into a heterotrimeric mutant that inactivates the native TNF via its sequestration from its two receptors (Steed et al., 2003). A disease-causing example involves more than 100 autosomal dominant mutations in the extracellular binding domain of the heterotrimeric tumor necrosis factor receptor 1 (TNFR1) responsible for the TNF Receptor-Associated Periodic Syndrome (TRAPS). The underlying symptoms are recurrent and prolonged episodes of fever and inflammation (Cudrici et al., 2020). The TNFR2-Fc therapeutic, etanercept, has limited efficacy toward TRAPS, demonstrating that the interactions between TNF and its receptors is not the basis of the disease. This supports the reduced binding of TNF to the TNFR1 heterotrimeric T50M variant and absence of almost any binding to the C33Y and C52F mutants, even though these receptor mutants are present on the cell surface (Todd et al., 2004). Studies initially indicated that variant-containing TNFR1 heterotrimers were mostly retained within the ER, resulting in ER stress in the context of inflammatory signaling and secretion of pro-inflammatory cytokines (Simon et al. 2010). Unexpectedly, the inflammation of TRAPS was not driven by the traditional unfolded protein response, but rather was mediated by enhanced generation of mitochondrial ROS (Simon et al., 2010; Bulua et al., 2011). Our studies of Y99 MIF variants serve as another example of a homotrimeric and heterotrimeric protein with unique biological effects. Heterotrimeric MIF species also provide an additional avenue for characterizing the allosteric network proposed to control the distinct functions of MIF, including the role of inter- and intra-subunit signaling.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the online repository BMRB, <https://bmrbl.io/>. The names of the accession number(s) are as follows: BMRB entry assigned accession number: 51223 Title: "1H15N HSQC chemical shifts and Relaxation Parameters for Y99C MIF Homotrimers." BMRB entry assigned accession number: 51224 Title: "1H15N HSQC chemical shifts and Relaxation Parameters for Y99C MIF Mixed Trimers."

ETHICS STATEMENT

The animal study was reviewed and approved by The animal study protocol was approved by the Institutional Animal Care and Use Committee of Cooper University Healthcare, Camden, NJ.

AUTHOR CONTRIBUTIONS

GP and RM performed crystallization, data collection and crystallographic analysis. ES purified homotrimeric Y99C and WT/Y99C mixed trimers, and collected and analyzed NMR and CD data. RM performed steady-state kinetics. NKM collected and analyzed the analytical ultracentrifugation. DS and VB were involved in animal experiments. Everyone

contributed to the manuscript and MO, GL and EL wrote the final draft.

FUNDING

This work was supported by the Intramural Research Programs of the National Institute of Arthritis and Musculoskeletal and Skin Diseases (Z01-AR041198 to MJO).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.783669/full#supplementary-material>

REFERENCES

- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. D Biol. Cryst.* 66 (Pt 2), 213–221. doi:10.1107/S0907444909052925
- Bernhagen, J., Krohn, R., Lue, H., Gregory, J. L., Zernecke, A., Koenen, R. R., et al. (2007). MIF Is a Noncognate Ligand of CXC Chemokine Receptors in Inflammatory and Atherogenic Cell Recruitment. *Nat. Med.* 13 (5), 587–596. doi:10.1038/nm1567
- Bulua, A. C., Simon, A., Maddipati, R., Pelletier, M., Park, H., Kim, K. Y., et al. (2011). Mitochondrial Reactive Oxygen Species Promote Production of Proinflammatory Cytokines and are Elevated in TNFR1-Associated Periodic Syndrome (TRAPS). *J. Exp. Med.* 208 (3), 519–33. doi:10.1084/jem.20102049
- Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. D Biol. Cryst.* 66 (Pt 1), 12–21. doi:10.1107/S0907444909042073
- Cho, Y., Crichtlow, G. V., Vermeire, J. J., Leng, L., Du, X., Hodsdon, M. E., et al. (2010). Allosteric Inhibition of Macrophage Migration Inhibitory Factor Revealed by Ibudilast. *Proc. Natl. Acad. Sci.* 107 (25), 11313–11318. doi:10.1073/pnas.1002716107
- Cho, Y., Vermeire, J. J., Merkel, J. S., Leng, L., Du, X., Bucala, R., et al. (2011). Drug Repositioning and Pharmacophore Identification in the Discovery of Hookworm MIF Inhibitors. *Chem. Biol.* 18 (9), 1089–1101. doi:10.1016/j.chembiol.2011.07.011
- Crichtlow, G. V., Fan, C., Keeler, C., Hodsdon, M., and Lolis, E. J. (2012). Structural Interactions Dictate the Kinetics of Macrophage Migration Inhibitory Factor Inhibition by Different Cancer-Preventive Isothiocyanates. *Biochemistry* 51 (38), 7506–7514. doi:10.1021/bi3005494
- Cudrici, C., Deutch, N., and Aksentijevich, I. (2020). Revisiting TNF Receptor-Associated Periodic Syndrome (TRAPS): Current Perspectives. *Ijms* 21 (9), 3263. doi:10.3390/ijms21093263
- Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* 6 (3), 277–293. doi:10.1007/bf00197809
- Emsley, P., and Cowtan, K. (2004). Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. D Biol. Cryst.* 60 (Pt 12 Pt 1), 2126–2132. doi:10.1107/S0907444904019158
- Fan, C., Rajasekaran, D., Syed, M. A., Leng, L., Loria, J. P., Bhandari, V., et al. (2013). MIF Intersubunit Disulfide Mutant Antagonist Supports Activation of CD74 by Endogenous MIF Trimer at Physiologic Concentrations. *Proc. Natl. Acad. Sci.* 110 (27), 10994–10999. doi:10.1073/pnas.1221817110
- Hudson, J. D., Shoaibi, M. A., Maestro, R., Carnero, A., Hannon, G. J., and Beach, D. H. (1999). A Proinflammatory Cytokine Inhibits P53 Tumor Suppressor Activity. *J. Exp. Med.* 190 (10), 1375–1382. doi:10.1084/jem.190.10.1375
- Lacy, M., Kontos, C., Brandhofer, M., Hille, K., Gröning, S., Sinitski, D., et al. (2018). Identification of an Arg-Leu-Arg Tripeptide that Contributes to the Binding Interface between the Cytokine MIF and the Chemokine Receptor CXCR4. *Sci. Rep.* 8 (1), 5171. doi:10.1038/s41598-018-23554-5
- Lee, W., Tonelli, M., and Markley, J. L. (2015). NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics* 31 (8), 1325–1327. doi:10.1093/bioinformatics/btu830
- Leng, L., Metz, C. N., Fang, Y., Xu, J., Donnelly, S., Baugh, J., et al. (2003). MIF Signal Transduction Initiated by Binding to CD74. *J. Exp. Med.* 197 (11), 1467–1476. doi:10.1084/jem.20030286
- Li, G.-Q., Xie, J., Lei, X.-Y., and Zhang, L. (2009). Macrophage Migration Inhibitory Factor Regulates proliferation of Gastric Cancer Cells via the PI3K/Akt Pathway. *Wjg* 15 (44), 5541–5548. doi:10.3748/wjg.15.5541
- Loria, J. P., Rance, M., and Palmer, III, A. G., 3rd (1999). A TROSY CPMG Sequence for Characterizing Chemical Exchange in Large Proteins. *J. Biomol. NMR* 15 (2), 151–155. doi:10.1023/a:1008355631073
- Lubetsky, J. B., Dios, A., Han, J., Aljabari, B., Ruzsicska, B., Mitchell, R., et al. (2002). The Tautomerase Active Site of Macrophage Migration Inhibitory Factor Is a Potential Target for Discovery of Novel Anti-Inflammatory Agents. *J. Biol. Chem.* 277 (28), 24976–24982. doi:10.1074/jbc.m203220200
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic Software. *J. Appl. Cryst.* 40 (Pt 4), 658–674. doi:10.1107/S0021889807021206
- Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006). HKL-3000: the Integration of Data Reduction and Structure Solution - from Diffraction Images to an Initial Model in Minutes. *Acta Crystallogr. D Biol. Cryst.* 62 (Pt 8), 859–866. doi:10.1107/s0907444906019949
- Mitchell, R. A., Metz, C. N., Peng, T., and Bucala, R. (1999). Sustained Mitogen-Activated Protein Kinase (MAPK) and Cytoplasmic Phospholipase A2 Activation by Macrophage Migration Inhibitory Factor (MIF). *J. Biol. Chem.* 274 (25), 18100–18106. doi:10.1074/jbc.274.25.18100
- Ouertatani-Sakouhi, H., El-Turk, F., Fauvet, B., Cho, M.-K., Pinar Karpinar, D., Le Roy, D., et al. (2010). Identification and Characterization of Novel Classes of Macrophage Migration Inhibitory Factor (MIF) Inhibitors with Distinct Mechanisms of Action. *J. Biol. Chem.* 285 (34), 26581–26598. doi:10.1074/jbc.M110.113951
- Pantouris, G., Syed, M. A., Fan, C., Rajasekaran, D., Cho, T. Y., Rosenberg, E. M., Jr., et al. (2015). An Analysis of MIF Structural Features that Control Functional Activation of CD74. *Chem. Biol.* 22 (9), 1197–1205. doi:10.1016/j.chembiol.2015.08.006
- Pantouris, G., Ho, J., Shah, D., Syed, M. A., Leng, L., Bhandari, V., et al. (2018). Nanosecond Dynamics Regulate the MIF-Induced Activity of CD74. *Angew. Chem. Int. Ed.* 57, 7116–7119. doi:10.1002/anie.201803191
- Pantouris, G., Khurana, L., Ma, A., Skeens, E., Reiss, K., Batista, V. S., et al. (2020). Regulation of MIF Enzymatic Activity by an Allosteric Site at the Central Solvent Channel. *Cel Chem. Biol.* 27 (6), 740–750. doi:10.1016/j.chembiol.2020.05.001

- Philo, J. S., Yang, T. H., and LaBarre, M. (2004). Re-examining the Oligomerization State of Macrophage Migration Inhibitory Factor (MIF) in Solution. *Biophys. Chem.* 108 (1-3), 77–87. doi:10.1016/j.bpc.2003.10.010
- Pravda, L., Sehnal, D., Toušek, D., Navrátilová, V., Bazgier, V., Berka, K., et al. (2018). MOLEonline: a Web-Based Tool for Analyzing Channels, Tunnels and Pores (2018 Update). *Nucleic Acids Res.* 46 (W1), W368–W373. doi:10.1093/nar/gky309
- Rajasekaran, D., Gröning, S., Schmitz, C., Zierow, S., Drucker, N., Bakou, M., et al. (2016). Macrophage Migration Inhibitory Factor-CXCR4 Receptor Interactions. *J. Biol. Chem.* 291, 15881–15895. doi:10.1074/jbc.M116.717751
- Rosengren, E., Bucala, R., Åman, P., Jacobsson, L., Odh, G., Metz, C. N., et al. (1996). The Immunoregulatory Mediator Macrophage Migration Inhibitory Factor (MIF) Catalyzes a Tautomerization Reaction. *Mol. Med.* 2 (1), 143–149. doi:10.1007/bf03402210
- Rosengren, E., Åman, P., Thelin, S., Hansson, C., Ahlfors, S., Björk, P., et al. (1997). The Macrophage Migration Inhibitory Factor MIF Is a Phenylpyruvate Tautomerase. *FEBS Lett.* 417 (1), 85–88. doi:10.1016/s0014-5793(97)01261-1
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 Years of Image Analysis. *Nat. Methods* 9 (7), 671–675. doi:10.1038/nmeth.2089
- Schuck, P. (2000). Size-distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* 78 (3), 1606–1619. doi:10.1016/s0006-3495(00)76713-0
- Sehnal, D., Svobodová Vařeková, R., Berka, K., Pravda, L., Navrátilová, V., Banáš, P., et al. (2013). MOLE 2.0: Advanced Approach for Analysis of Biomacromolecular Channels. *J. Cheminform* 5 (1), 39. doi:10.1186/1758-2946-5-39
- Shah, D., Torres, C., and Bhandari, V. (2019). Adiponectin Deficiency Induces Mitochondrial Dysfunction and Promotes Endothelial Activation and Pulmonary Vascular Injury. *FASEB J.* 33 (12), 13617–13631. doi:10.1096/fj.201901123R
- Simon, A., Park, H., Maddipati, R., Lobito, A. A., Bulua, A. C., Jackson, A. J., et al. (2010). Concerted Action of Wild-type and Mutant TNF Receptors Enhances Inflammation in TNF Receptor 1-Associated Periodic Fever Syndrome. *Proc. Natl. Acad. Sci.* 107 (21), 9801–9806. doi:10.1073/pnas.0914118107
- Steed, P. M., Tansey, M. G., Zalevsky, J., Zhukovsky, E. A., Desjarlais, J. R., Szymkowski, D. E., et al. (2003). Inactivation of TNF Signaling by Rationally Designed Dominant-Negative TNF Variants. *Science* 301 (5641), 1895–1898. doi:10.1126/science.1081297
- Stivers, J. T., Abeygunawardana, C., Mildvan, A. S., Hajipour, G., and Whitman, C. P. (1996a). 4-Oxalocrotonate Tautomerase: pH Dependence of Catalysis and pKa Values of Active Site Residues. *Biochemistry* 35 (3), 814–823. doi:10.1021/bi9510789
- Stivers, J. T., Abeygunawardana, C., Mildvan, A. S., Hajipour, G., Whitman, C. P., and Chen, L. H. (1996b). Catalytic Role of the Amino-Terminal Proline in 4-oxalocrotonate Tautomerase: Affinity Labeling and Heteronuclear NMR Studies. *Biochemistry* 35 (3), 803–813. doi:10.1021/bi951077g
- Subramanya, H. S., Roper, D. I., Dauter, Z., Dodson, E. J., Davies, G. J., Wilson, K. S., et al. (1996). Enzymatic Ketoneization of 2-hydroxybutyrate: Specificity and Mechanism Investigated by the crystal Structures of Two Isomerases. *Biochemistry* 35 (3), 792–802. doi:10.1021/bi951732k
- Sun, H. W., Bernhagen, J., Bucala, R., and Lolis, E. (1996). Crystal Structure at 2.6-Å Resolution of Human Macrophage Migration Inhibitory Factor. *Proc. Natl. Acad. Sci.* 93 (11), 5191–5196. doi:10.1073/pnas.93.11.5191
- Suzuki, M., Sugimoto, H., Nakagawa, A., Tanaka, I., Nishihira, J., and Sakai, M. (1996). Crystal Structure of the Macrophage Migration Inhibitory Factor from Rat Liver. *Nat. Struct. Mol. Biol.* 3 (3), 259–266. doi:10.1038/nsb0396-259
- Swope, M., Sun, H. W., Blake, P. R., and Lolis, E. (1998). Direct Link between Cytokine Activity and a Catalytic Site for Macrophage Migration Inhibitory Factor. *EMBO J.* 17 (13), 3534–3541. doi:10.1093/emboj/17.13.3534
- Takahashi, K., Koga, K., Linge, H. M., Zhang, Y., Lin, X., Metz, C. N., et al. (2009). Macrophage CD74 Contributes to MIF-Induced Pulmonary Inflammation. *Respir. Res.* 10 (1), 33. doi:10.1186/1465-9921-10-33
- Todd, I., Radford, P. M., Draper-Morgan, K.-A., McIntosh, R., Bainbridge, S., Dickinson, P., et al. (2004). Mutant Forms of Tumour Necrosis Factor Receptor I that Occur in TNF-Receptor-Associated Periodic Syndrome Retain Signalling Functions but Show Abnormal Behaviour. *Immunology* 113 (1), 65–79. doi:10.1111/j.1365-2567.2004.01942.x
- Wang, Y., An, R., Umanah, G. K., Park, H., Nambiar, K., Eacker, S. M., et al. (2016). A Nuclease that Mediates Cell Death Induced by DNA Damage and poly(ADP-Ribose) Polymerase-1. *Science* 354 (6308), aad6872. doi:10.1126/science.aad6872
- Wang, Y., Chen, Y., Wang, C., Yang, M., Wang, Y., Bao, L., et al. (2021). MIF Is a 3' Flap Nuclease that Facilitates DNA Replication and Promotes Tumor Growth. *Nat. Commun.* 12 (1), 2954. doi:10.1038/s41467-021-23264-z
- Weber, C., Kraemer, S., Drechsler, M., Lue, H., Koenen, R. R., Kapurniotou, A., et al. (2008). Structural Determinants of MIF Functions in CXCR2-Mediated Inflammatory and Atherogenic Leukocyte Recruitment. *Proc. Natl. Acad. Sci.* 105 (42), 16278–16283. doi:10.1073/pnas.0804017105
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., et al. (2011). Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr. D Biol. Cryst.* 67 (Pt 4), 235–242. doi:10.1107/S0907444910045749

Conflict of Interest: Author NKM was employed by the company KBI Biopharma.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Skeens, Pantouris, Shah, Manjula, Ombrello, Maluf, Bhandari, Lisi and Lolis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: frontiersin.org/about/contact



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership