



WEAK INTERACTIONS IN MOLECULAR MACHINERY

EDITED BY: Irene Díaz-Moreno and Rivka Isaacson
PUBLISHED IN: Frontiers in Molecular Biosciences



frontiers

Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714
ISBN 978-2-88945-787-8
DOI 10.3389/978-2-88945-787-8

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: researchtopics@frontiersin.org

WEAK INTERACTIONS IN MOLECULAR MACHINERY

Topic Editors:

Irene Díaz-Moreno, University of Seville, Spain

Rivka Isaacson, King's College London, United Kingdom

Diverse cellular processes depend on weak interactions between biological components. In this volume, we bring together a wealth of recent information on this topic, combining original research articles with up-to-date reviews, organized under four separate themes. In our first section on nucleic acid regulation, we include a study of the PII-NAGK-PipX-NtcA regulatory axis of cyanobacteria and discover much about the role of RNA binding protein regulation and cross-talk in the control of AU-rich mRNA. Peptide-mediated weak interactions are our second theme in which we review weak molecular interactions in clathrin-mediated endocytosis, investigate the selectivity of the G7-18NATE inhibitor peptide for the Grb7-SH2 domain target and present new structure and interactions of the TPR domain of Sgt2 with yeast chaperones and Ybr137wp. Our third part focuses on carbohydrates and includes a thorough review of how to use NMR to study transient carbohydrate–protein binding and a structural and functional study of lysostaphin–substrate interaction. In our final section, we look at functional sensors driving weak interactions by presenting the molecular basis of the dual regulation of bacterial iron sulfur cluster biogenesis by CyaY and IscX alongside a review of intramolecular fuzzy interactions involving intrinsically disordered domains. Taken together, our eBook chapters offer some recent insight into this area of scientific understanding which is still expanding exponentially.

Citation: Díaz-Moreno, I., Isaacson, R., eds. (2019). Weak Interactions in Molecular Machinery. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-787-8

Table of Contents

05 Editorial: Weak Interactions in Molecular Machinery

Rivka L. Isaacson and Irene Díaz-Moreno

SECTION 1

DNA AND RNA REGULATION

07 The P_{II} -NAGK-PipX-NtcA Regulatory Axis of Cyanobacteria: A Tale of Changing Partners, Allosteric Effectors and Non-covalent Interactions

Alicia Forcada-Nadal, José Luis Llácer, Asunción Contreras, Clara Marco-Marín and Vicente Rubio

25 RNA Binding Protein Regulation and Cross-Talk in the Control of AU-rich mRNA Fate

Sofía M. García-Mauriño, Francisco Rivero-Rodríguez, Alejandro Velázquez-Cruz, Marian Hernández-Vellisca, Antonio Díaz-Quintana, Miguel A. De la Rosa and Irene Díaz-Moreno

SECTION 2

PEPTIDE-MEDIATED WEAK INTERACTIONS

34 Insight Into the Selectivity of the G7-18NATE Inhibitor Peptide for the Grb7-SH2 Domain Target

Gabrielle M. Watson, William A. H. Lucas, Menachem J. Gunzburg and Jacqueline A. Wilce

42 Structure and Interactions of the TPR Domain of Sgt2 With Yeast Chaperones and Ybr137wp

Ewelina M. Kryzstofinska, Nicola J. Evans, Arjun Thapaliya, James W. Murray, Rhodri M. L. Morgan, Santiago Martinez-Lumbreras and Rivka L. Isaacson

53 Weak Molecular Interactions in Clathrin-Mediated Endocytosis

Sarah M. Smith, Michael Baker, Mary Halebian and Corinne J. Smith

SECTION 3

TARGETING CARBOHYDRATES AND GLYCANS

64 The Use of NMR to Study Transient Carbohydrate–Protein Interactions

Pedro M. Nieto

71 Structural and Functional Insights Into Lysostaphin–Substrate Interaction

Helena Tossavainen, Vytas Raulinaitis, Linda Kauppinen, Ulla Pentikäinen, Hannu Maaheimo and Perttu Permi

SECTION 4

FUNCTIONAL SENSORS DRIVING WEAK INTERACTIONS

85 *The Molecular Bases of the Dual Regulation of Bacterial Iron Sulfur Cluster Biogenesis by CyaY and IscX*

Salvatore Adinolfi, Rita Puglisi, Jason C. Crack, Clara Iannuzzi, Fabrizio Dal Piaz, Petr V. Konarev, Dmitri I. Svergun, Stephen Martin, Nick E. Le Brun and Annalisa Pastore

97 *Intramolecular Fuzzy Interactions Involving Intrinsically Disordered Domains*

Miguel Arbesú, Guillermo Iruela, Héctor Fuentes, João M. C. Teixeira and Miquel Pons



Editorial: Weak Interactions in Molecular Machinery

Rivka L. Isaacson¹ and Irene Díaz-Moreno^{2*}

¹ Department of Chemistry, King's College London, London, United Kingdom, ² cicCartuja, Institute for Chemical Research (IIQ), University of Seville - CSIC, Seville, Spain

Keywords: transient interactions, molecular machineries, structural biology, biophysics, weak contacts

Editorial on the Research Topic

Weak Interactions in Molecular Machinery

Individuals are often sustained by intense relationships with others, although few would deny the importance of those more transient interactions with service providers, colleagues, and casual acquaintances in facilitating our continued existence. These latter associations are necessarily and conveniently weaker. Nobody has time for a deep conversation with everyone they encounter during a day, and society, like so many complex systems, runs on a hierarchy of interaction strengths. Similarly, inside each crowded cell of our bodies, functions arise when communication is established through biomolecules that physically and specifically contact each other in a concerted manner to transmit messages efficiently.

“Molecular sociology” within cells involves binding events between macromolecules, such as protein–protein, protein–nucleic acid, protein–carbohydrate, and protein–membrane interactions, that are acutely orchestrated to sustain the life–death balance. These interactions between biomolecules occur on a wide range of timescales. Stable complexes, with lifetimes ranging from minutes to days, involve high affinity and high specificity binding. Amongst many others, these include irreversible enzyme inhibition and the assembly of proteins supporting the cell's ultrastructure.

Weak complexes are characterized by a fine balance between specificity of binding and fast turnover rate, with the majority displaying equilibrium dissociation constants within the micromolar or even millimolar range. Perhaps counter-intuitively, these weaker molecular recognition mechanisms are not uncommon and play key roles in many biological processes, including electron transfer chains in respiration and photosynthesis and the cell signaling cascades involving kinases and phosphatases. Due to technical limitations, weak complexes remain poorly understood despite their critical role in many biological events inside the cell.

Over the last few years, specific tools have been developed to analyse more transient intermolecular interactions, including NMR paramagnetic relaxation enhancement and kinetic approaches. These methods have recently been complemented by high-throughput techniques to identify novel biomolecules that weakly bind to each other. Such advances make the analysis of the transient biointeractome (the so-called *trans-biointeractome*) more affordable. Overall, *trans-biointeractome* analysis remains highly dependent on the specific technique employed. In addition, the dynamics of the system can enhance the strength of interactions that are weak. Similarly, “one person may not be able to pull a truck from the mud but many would be able to.” This may also be the reason that interactions detected under controlled conditions *in vivo* become extremely difficult to study in isolation. Protein compartmentalization, elevated protein levels at defined moments of the cell cycle, multivalent binding, can increase (temporarily) the affinities of transient interactions, suggesting that the so-called *trans-biointeractome* is inherently plastic and that it evolves during cell lifespan.

OPEN ACCESS

Edited by:

Anastasia S. Politou,
University of Ioannina, Greece

Reviewed by:

Konstantinos Tripsianes,
Masaryk University, Czechia

*Correspondence:

Irene Díaz-Moreno
idiazmoreno@us.es

Specialty section:

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

Received: 26 November 2018

Accepted: 22 December 2018

Published: 23 January 2019

Citation:

Isaacson RL and Díaz-Moreno I (2019)
Editorial: Weak Interactions in
Molecular Machinery.
Front. Mol. Biosci. 5:117.
doi: 10.3389/fmolb.2018.00117

Therefore, the study of weakly interacting systems can be reliably tackled in depth only by an integrative and holistic approach combining *in vitro* strategies with other methods to detect such interactions within the cell. We are delighted by the breadth of techniques and systems covered in the articles of our research topic.

Unsurprisingly, given its well-documented strengths in measuring weak binding, NMR spectroscopy features heavily as a method used in this Special Issue in *Frontiers in Molecular Biosciences*.

Nieto provides a comprehensive mini review on using NMR to investigate binding between carbohydrates and proteins, which is particularly useful given the challenges of isotopically labeling glycans. In their original research article, the group headed by Krysztofinska et al. use NMR alongside X-ray crystallography and isothermal titration calorimetry (ITC) to analyse the structure and binding of different heat shock chaperones, via a carboxylate clamp mechanism, to a co-chaperone involved in targeting proteins to membranes. Multiple proteins competing for the same binding site, as analyzed using NMR, is a theme that continues in the research article by Adinolfi et al. who also employ small-angle X-ray scattering (SAXS), biolayer interferometry (BLI), and native mass-spectrometry (MS) to dissect the dual regulation of bacterial Iron Sulfur cluster biogenesis by two proteins, CyaY and IscX. Tossavainen et al. also use NMR and SAXS, alongside Molecular Dynamics simulations to explore the variable affinity interactions between *Staphylococcus aureus* cell-wall digester, lysostaphin, and its substrates.

Surface plasmon resonance (SPR) is another widely used technique to measure interactions, and along with microarray technology, is employed by Watson et al. to establish the efficacy of a peptide inhibitor in blocking an important weak interaction between an SH2 domain and its phosphotyrosine target. This is a step on the way to rational drug design for cancer.

The importance and unique attributes of intrinsically disordered proteins (IDPs) are becoming increasingly recognized, acknowledged and valued. Arbesú et al. feed into this theme by presenting a fascinating mini review on

the intramolecular “fuzzy” interactions that occur within intrinsically disordered domains of proteins. This mini-review perfectly links with the Frontiers topic on *Function and Flexibility: Friend or Foe?*

Perhaps the most self-contained story in our collection, that truly conveys the variety of weak interactions in a well-characterized pathway, is the review from the group of Forcada-Nadal et al. on the PII-NAGK-PipX-NtcA regulatory axis of cyanobacteria, which beautifully compiles the information that led to an understanding of this network. Clathrin-mediated endocytosis is an elegant process as delineated in a review from Smith et al. who examine the vital roles of weak interactions in this choreographed cellular system. On the other hand, it becomes more and more evident that there is a need for a transition from a static to a dynamic point of view in order to take into account the biological environment during RNA binding and RNA metabolism, as reviewed by the team of García-Mauriño et al. Actually, the understanding of the mRNAs processing in highly dynamic and often transient macromolecular complexes also remains challenging.

The subjects covered in our research topic span a wide range of biological questions and methodology. We believe this is merely the tip of the iceberg and that the future will bring vast insight into the importance of weak interactions in molecular machinery.

AUTHOR CONTRIBUTIONS

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

Conflict of Interest Statement: 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 © 2019 Isaacson and Díaz-Moreno. 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 P_{II} -NAGK-PipX-NtcA Regulatory Axis of Cyanobacteria: A Tale of Changing Partners, Allosteric Effectors and Non-covalent Interactions

Alicia Forcada-Nadal^{1,2}, José Luis Llácer^{1,3}, Asunción Contreras², Clara Marco-Marín^{1,3} and Vicente Rubio^{1,3*}

¹ Instituto de Biomedicina de Valencia del Consejo Superior de Investigaciones Científicas, Valencia, Spain, ² Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, Alicante, Spain, ³ Group 739, Centro de Investigación Biomédica en Red de Enfermedades Raras – Instituto de Salud Carlos III, Valencia, Spain

OPEN ACCESS

Edited by:

Irene Díaz-Moreno,
Universidad de Sevilla, Spain

Reviewed by:

Karl Forchhammer,
Universität Tübingen, Germany
Juan A. Hermoso,
Consejo Superior de Investigaciones
Científicas (CSIC), Spain

*Correspondence:

Vicente Rubio
rubio@ibv.csic.es

Specialty section:

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

Received: 26 August 2018

Accepted: 18 October 2018

Published: 13 November 2018

Citation:

Forcada-Nadal A, Llácer JL,
Contreras A, Marco-Marín C and
Rubio V (2018) The
 P_{II} -NAGK-PipX-NtcA Regulatory Axis
of Cyanobacteria: A Tale of Changing
Partners, Allosteric Effectors and
Non-covalent Interactions
Front. Mol. Biosci. 5:91.
doi: 10.3389/fmolb.2018.00091

P_{II} , a homotrimeric very ancient and highly widespread (bacteria, archaea, plants) key sensor-transducer protein, conveys signals of abundance or poorness of carbon, energy and usable nitrogen, converting these signals into changes in the activities of channels, enzymes, or of gene expression. P_{II} sensing is mediated by the P_{II} allosteric effectors ATP, ADP (and, in some organisms, AMP), 2-oxoglutarate (2OG; it reflects carbon abundance and nitrogen scarcity) and, in many plants, L-glutamine. Cyanobacteria have been crucial for clarification of the structural bases of P_{II} function and regulation. They are the subject of this review because the information gathered on them provides an overall structure-based view of a P_{II} regulatory network. Studies on these organisms yielded a first structure of a P_{II} complex with an enzyme, (N-acetyl-L-glutamate kinase, NAGK), deciphering how P_{II} can cause enzyme activation, and how it promotes nitrogen stockpiling as arginine in cyanobacteria and plants. They have also revealed the first clear-cut mechanism by which P_{II} can control gene expression. A small adaptor protein, PipX, is sequestered by P_{II} when nitrogen is abundant and is released when is scarce, swapping partner by binding to the 2OG-activated transcriptional regulator NtcA, co-activating it. The structures of P_{II} -NAGK, P_{II} -PipX, PipX alone, of NtcA in inactive and 2OG-activated forms and as NtcA-2OG-PipX complex, explain structurally P_{II} regulatory functions and reveal the changing shapes and interactions of the T-loops of P_{II} depending on the partner and on the allosteric effectors bound to P_{II} . Cyanobacterial studies have also revealed that in the P_{II} -PipX complex PipX binds an additional transcriptional factor, PlmA, thus possibly expanding PipX roles beyond NtcA-dependency. Further exploration of these roles has revealed a functional interaction of PipX with PipY, a pyridoxal-phosphate (PLP) protein involved in PLP homeostasis whose mutations in the human ortholog cause epilepsy. Knowledge of cellular levels of the different components

of this P_{II}-PipX regulatory network and of K_D values for some of the complexes provides the basic background for gross modeling of the system at high and low nitrogen abundance. The cyanobacterial network can guide searches for analogous components in other organisms, particularly of PipX functional analogs.

Keywords: protein structure, nitrogen regulation, gene expression regulation, signaling, P_{II} complexes, PipX complexes, NtcA structure and complexes, PlmA

Protein P_{II} was discovered in the late sixties of last century (Stadtman, 2001), when *Escherichia coli* glutamine synthetase (GS) was found to exist in feed-back inhibition susceptible or refractory forms depending on the adenylation state of one tyrosine per GS subunit. P_I and P_{II} were the first and second peaks from a gel filtration column (Shapiro, 1969). P_I is a bifunctional enzyme (ATase) that adenylylates or deadenylylates GS (Jiang et al., 2007). P_{II} controls the activity of the ATase. We now know that P_{II} proteins are highly conserved and very widespread sensors used to transduce energy/carbon/nitrogen abundance signals in all domains of life (Kinch and Grishin, 2002; Sant'Anna et al., 2009). They are found in archaea, bacteria (Gram+ and Gram-), unicellular algae and plants. Many organisms have two or more genes for P_{II} proteins (reviewed in Forchhammer and Lüddecke, 2016), as *E. coli*, that has two paralogous genes encoding P_{II} proteins with distinct functions, one (GlnB) involved in the control of GS, and the other one (GlnK) being involved in the regulation of ammonia entry into the cell. By binding to target proteins, including channels, enzymes, or molecules involved in gene regulation and by altering the function of these target molecules, P_{II} proteins can regulate ammonia entry, nitrogen metabolism and gene expression (Forchhammer, 2008; Llácer et al., 2008). Cyanobacteria, and particularly among them *Synechococcus elongatus* PCC 7942 (hereafter *S. elongatus*), have been and continue to be very useful organisms for studies of P_{II} actions, fuelling structural understanding of P_{II} regulation. Studies on these organisms exemplify very clearly how enzyme activity and gene regulation can be controlled by P_{II} via formation of several complexes (summarized in Figure 1) mediated by weak intermolecular interactions that are crucially regulated by allosteric effectors of the proteins involved in these complexes. This is the focus of the present review.

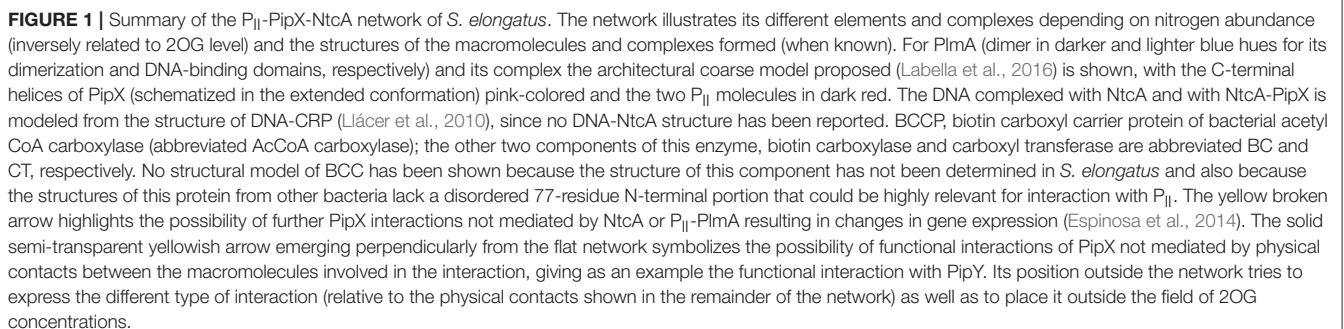
Abbreviations: P_{II}, a homotrimeric signaling protein; GlnK, GlnK3 and GlnB, different paralogous forms of P_{II} proteins; GlnD, the bifunctional enzyme that uridylylates and deuridylylates GlnB in *E. coli*; GS, glutamine synthetase; AmtB and Amt, homologous trimeric bacterial transporters of ammonia; PamA, putative channel of unknown function that is encoded by *sll0985* of *Synechocystis* sp. PCC 6803; NtcA and CRP, homologous homodimeric transcription factors of cyanobacteria and of *E. coli*, respectively; the imperfectly palindromic target DNA sequences to which they bind specifically are called NtcA box and CRP box, respectively; PlmA, putative homodimeric transcription factor of the GntR family that is found in cyanobacteria; P_I or ATase, bifunctional enzyme that adenylylates and deadenylylates glutamine synthetase in *E. coli* and other enterobacteria; NAGK, N-acetyl-L-glutamate kinase; NAG, N-acetyl-L-glutamate; PipX, a small monomeric protein of cyanobacteria that can interact with P_{II} and with NtcA; FRET, fluorescence resonance energy transfer (also known as Förster resonance energy transfer), a phenomenon in which a fluorophore emits light of its characteristic frequency when a nearby different absorbing group is excited by

THE P_{II} SIGNALING PROTEIN

S. elongatus P_{II} (Figure 2), as other P_{II} proteins, is a homotrimer of a polypeptide chain of 112 amino acids that exhibits the ferredoxin fold ($\beta\alpha\beta$)₂ followed by a beta hairpin (Xu et al., 2003). The trimer (Figure 2A) has a hemispheric body nucleated by three antiparallel oblique (relative to the three-fold axis) β -sheets, each one formed by the 4-stranded sheet (topology $\downarrow\beta 2\uparrow\beta 3\downarrow\beta 1\uparrow\beta 4$) of a subunit (see for example subunit B in the central panel of Figure 2A) extended on its $\beta 4$ end by the C-terminal hairpin ($\beta 5$ - $\beta 6$) of an adjacent subunit (subunit A) and on the $\beta 2$ end by the $\beta 2$ - $\beta 3$ hairpin stem (the root of the T-loop, see below) of the other subunit of the trimer (subunit C). The three sheets become continuous on the flat face of the hemispheric body via their $\beta 2$ - $\beta 3$ hairpins (Figure 2A). The subunit sheets encircle like a 3-sided pyramid the three-fold axis, filling the inner space between them with their side-chains. They are covered externally by 6 helices (two per subunit) that run parallel to the β strands, contributing to the rounded shape of the hemispheric trimer (Figure 2A, panel to the right) and to the outer part of its equatorial flat face. In the convex face, three crevices are formed at subunits junctions between adjacent β -sheets, over the $\beta 2$ - $\beta 3$ hairpins (Figure 2B). These crevices host the sites for the allosteric effectors ATP/ADP [and in some species AMP (Palanca et al., 2014)] and 2-oxoglutarate (2OG) that endow P_{II} with its sensing roles (Kamberov et al., 1995; Zeth et al., 2014) (Figure 2B), the nucleotides reflecting the energy status (Fokina et al., 2011) and 2OG reflecting the abundance of carbon and, inversely, the nitrogen richness (see for example Muro-Pastor et al., 2001).

Very salient structural features of P_{II} are the long flexible T-loops (Figures 2A,C) formed by the 18 residues that tip the $\beta 2$ - $\beta 3$ hairpin of each subunit (Xu et al., 2003). These loops are key elements (although not the exclusive ones, see Rajendran et al., 2011 and Schumacher et al., 2015) for P_{II} interaction with its targets (Conroy et al., 2007; Gruswitz et al., 2007; Llácer et al., 2007, 2010; Mizuno et al., 2007; Zhao et al., 2010b; Chellamuthu et al., 2014). By binding at the boundary between the T-loop and the P_{II} body, at the crevice formed between adjacent subunits, the adenine nucleotides and MgATP/2OG promote the adoption by the T-loop of different conformations (Figure 2C) (Fokina et al.,

light; the FRET signal decreases with the 6th power of the distance between the absorbing and emitting groups; 2OG, 2-oxoglutarate, also called α -ketoglutarate; AcCoA carboxylase, acetyl coenzyme A carboxylase; BCCP, biotin carboxyl carrier protein, the protein subunit that hosts the covalently bound biotin in many bacterial AcCoA carboxylases; PLP, pyridoxal phosphate; PipY, product of the gene that in *S. elongatus* is the next downstream of pipX, forming a bicistronic operon with it; it is a PLP-containing protein.



occurs at least in proteobacteria and actinobacteria (Merrick, 2015), but it might be more widespread, since it has also been reported in an archaeon (Pedro-Roig et al., 2013). Structural studies with *E. coli* P_{II} (Palanca and Rubio, 2017) have excluded the stabilization of the T-loop into a fixed conformation by Tyr51 uridylylation, suggesting that the Tyr51-bound UMP physically interacts with the ATase. Although Tyr51 is conserved in cyanobacteria, it is not uridylylated. The T-loop serine 49 (**Figure 2C**, 1st, 2nd and 4th panels from the left) is phosphorylated in *S. elongatus* under conditions of nitrogen starvation by an unknown mechanism (Forchhammer and Tandeau de Marsac, 1994), whereas the phosphatase that dephosphorylates phosphoSer49 has been identified and proven to be 2OG-sensitive (Irmeler et al., 1997).

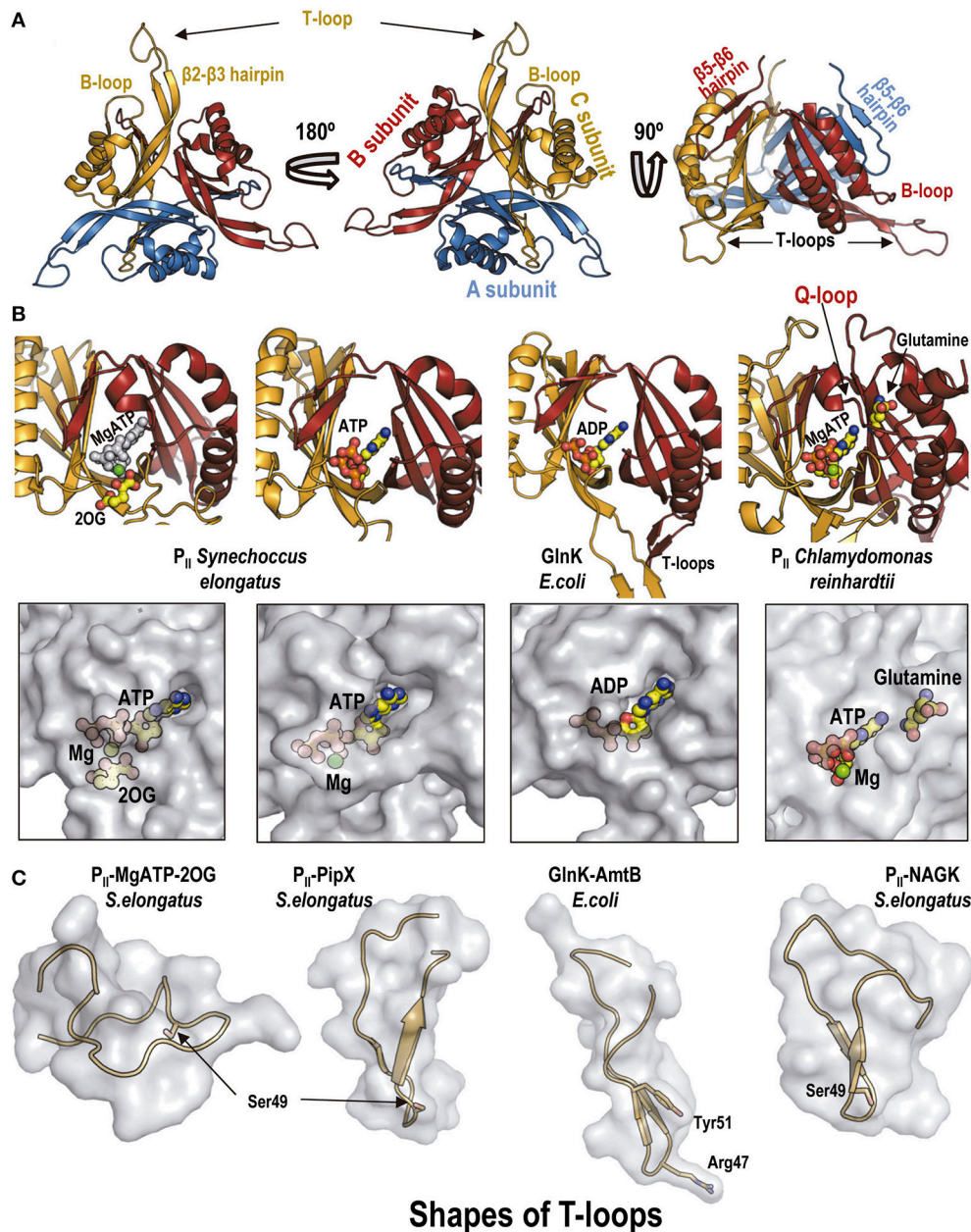


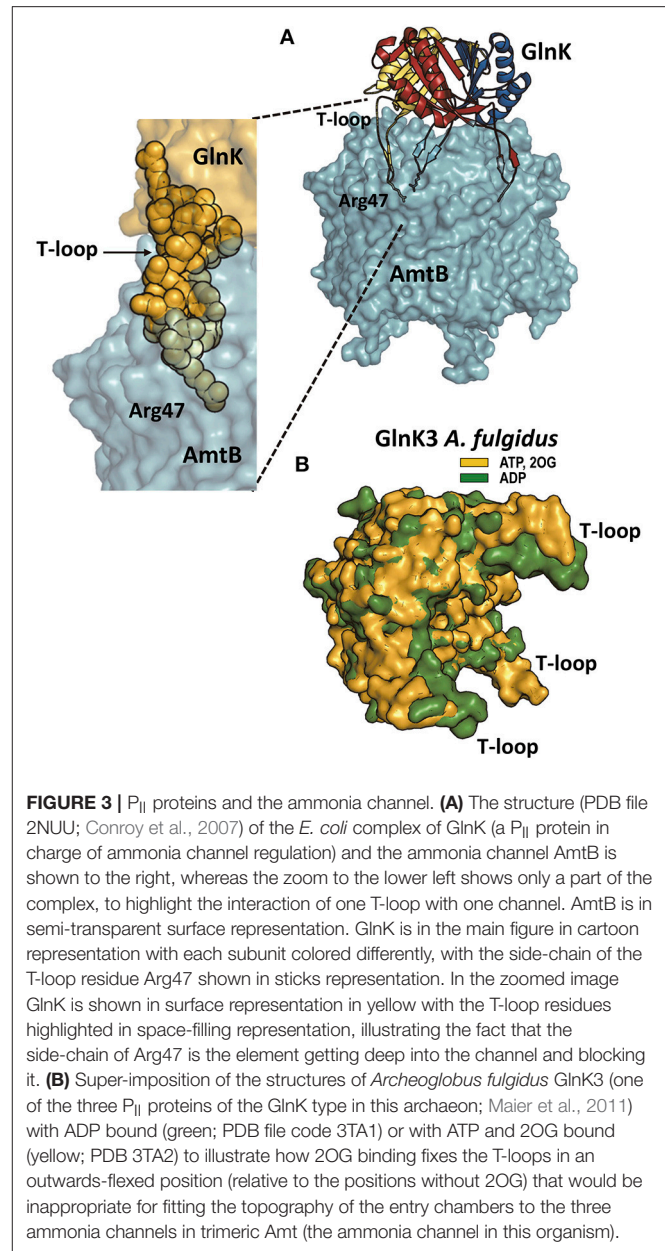
FIGURE 2 | The P_{II} protein. **(A)** Overall view in cartoon representation of *S. elongatus* P_{II} along its three-fold axis from the P_{II} flat face (left), from its convex side (middle), or with the three-fold axis vertical and the flat surface down (right). The structure corresponds to Protein Database (PDB) file 1QY7 (Xu et al., 2003). Each subunit in the trimer is colored differently. Some relevant traits are highlighted. **(B)** The P_{II} allosteric sites shown in cartoon representation (top) and in semi-transparent zoomed surface representation (bottom) in approximately the same pose as in the cartoon representation. For clarity, in the cartoon representations only the two subunits forming each site are shown, in the same colors as in **(A)**. Ligands are shown in sticks and balls representation, with atoms of C, O, N, P, and Mg in yellow, red, blue, orange, and green, respectively, except in the leftmost cartoon figure in which all the atoms of ATP are pale gray to highlight the bound 2OG (colored). Note in the corresponding panel of the bottom row that MgATP and 2OG are nearly fully buried in the P_{II} molecule. The organisms from which the P_{II} derive are indicated in the figure. The two panels to the left belong to isolated *S. elongatus* P_{II} (PDB file 2XZW; Fokina et al., 2010a); the third panels illustrate *E. coli* GlnK taken from its complex with AmtB (PDB 2NUU, Conroy et al., 2007); the rightmost panels show *Chlamydomonas reinhardtii* P_{II}, taken from its complex with *Arabidopsis thaliana* NAGK (PDB 4USJ; Chellamuthu et al., 2014). **(C)** Illustration of different shapes of the T-loops found in distinct complexes with allosteric effectors or with partner proteins. The T-loop is shown in cartoon representation, within a semi-transparent surface representation as if this loop were isolated from the remainder of P_{II} and from the protein partner in the complex. In the third panel, the side chain of Arg47 of *E. coli* GlnK is represented in sticks, given its importance for inhibiting the AmtB channel. Taken, from left to right, from: *S. elongatus* P_{II} with MgATP and 2OG bound (PDB file 2XZW; Fokina et al., 2010a); *S. elongatus* P_{II}-PipX (PDB 2XG8; Ll  cer et al., 2010); *E. coli* GlnK-AmtB complex (PDB 2NUU, Conroy et al., 2007); and P_{II}-NAGK complex (PDB 2V5H; Ll  cer et al., 2007).

FRET studies with engineered fluorescent *S. elongatus* P_{II} used as an ADP and ATP-sensitive probe (Lüddecke and Forchhammer, 2015) have challenged the claim (Radchenko et al., 2013) that P_{II} proteins have a very slow ATPase activity that would regulate P_{II} similarly as the signaling GTPases with bound GTP and GDP. Although this ATPase was reported as a 2OG-triggered switch that appeared an intrinsic trait of P_{II} proteins (Radchenko et al., 2013), the FRET experiments with *S. elongatus* P_{II} (Lüddecke and Forchhammer, 2015) appear to indicate that an endogenous ATPase is not a relevant mechanism for the transition of P_{II} into the ADP state.

P_{II} COMPLEXES WITH CHANNELS

The first structurally solved P_{II} complex was the one of *E. coli* GlnK with the AmtB ammonia channel (Conroy et al., 2007; Gruswitz et al., 2007) (Figure 3A) formed under nitrogen richness conditions. This structure showed that AmtB was inhibited by GlnK because the extended T-loop fits the channel entry, with the insertion into the channel of a totally extended arginine emerging from the T-loop and blocking the channel space (Figure 2C, 3rd panel from the left, and Figure 3A, zoom). The ADP-bound and MgATP/2OG-bound structures of an *Archeoglobus fulgidus* GlnK protein (Maier et al., 2011) indicated that 2OG may prevent GlnK binding to the ammonia channel because of induced flexing outwards (relative to the 3-fold molecular axis) of the T-loops, preventing their topographical correspondence with the three holes of the trimer of ammonia channels (Figure 3B). Interestingly, the T-loops of MgATP/2OG-bound *S. elongatus* P_{II} (Figure 2C, leftmost panel) and *A. fulgidus* GlnK3 (Figure 3B) exhibited different flexed conformations (relative to the ADP-bound extended forms), and thus 2OG-binding by itself does not determine a single T-loop conformation, at least with different P_{II} proteins.

Yeast two hybrid approaches (Osanai et al., 2005) detected the interaction between P_{II} and the putative channel PamA (encoded by *sll0985*) of *Synechocystis* sp. PCC 6803 (from now on *Synechocystis*), but molecular detail on this protein is non-existent, and, therefore, it is uncertain whether such interaction might resemble the GlnK-AmtB interaction. PamA is not conserved in many cyanobacteria, and the most closely related putative protein of *S. elongatus*, the product of the *Synpcc7942_0610* gene, failed to give interaction signal with *S. elongatus* P_{II} in yeast two hybrid assays (Castells, M.A., PhD Dissertation, Universidad de Alicante, 2010), despite the fact that the sequence identity with PamA concentrated in the C-terminal region, where P_{II} binds in *Synechocystis* (Osanai et al., 2005). *In vitro* studies with the recombinantly produced *Synechocystis* PamA and P_{II} showed that their interaction was lost in the presence of ATP and 2OG. Thus, similarly to the GlnK-AmtB and GlnK3-Amt complexes, the P_{II}-PamA complex is formed under conditions of nitrogen abundance. However, T-loop phosphorylation did not dissociate this complex (Osanai et al., 2005). The function of PamA is not known, but its deletion from *Synechocystis* changed the expression of



some NtcA-dependent genes (Osanai et al., 2005) by unclarified mechanisms.

P_{II} COMPLEXES WITH ENZYMES IN CYANOBACTERIA (AND BEYOND)

The complexes of P_{II} with the N-acetyl-L-glutamate kinase (NAGK) enzymes from *S. elongatus* and *Arabidopsis thaliana* presented a very different architecture with respect to the structure of the GlnK-AmtB complex of *E. coli* (Llácer et al., 2007; Mizuno et al., 2007) (Figures 4A,B). The P_{II}-NAGK complex is an activating complex in which the T-loops of P_{II} are flexed (Figure 2C, rightmost panel) and integrated into a

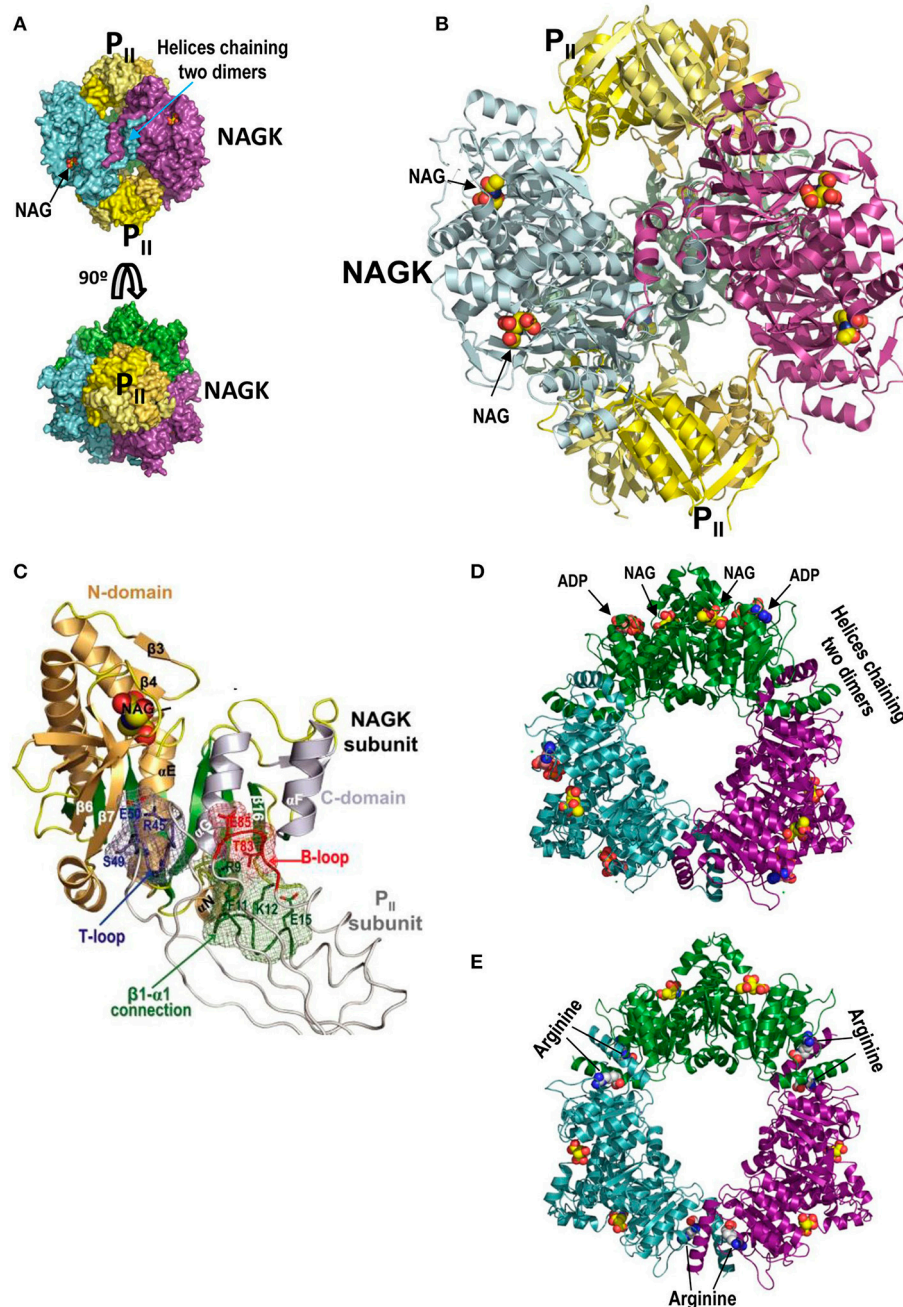


FIGURE 4 | P_{II}-NAGK complex and active and arginine-inhibited NAGK. **(A)** The P_{II}-NAGK complex of *S. elongatus* (PDB 2V5H; Liácer et al., 2010). Surface representations of the complex formed by two P_{II} trimers (yellow) capping on both ends the doughnut-like NAGK hexamer (trimer of dimers; each dimer in a different color). The three-fold axis is vertical (top) or perpendicular to the page (bottom). Figure of J.L. Liácer and V. Rubio taken from Chin (2008). Reprinted with permission from AAAS **(B)**. Cartoon representation of the *S. elongatus* P_{II}-NAGK complex after removing the back NAGK dimer for clarity. The three-fold symmetry axis is vertical. Reprinted from Current Opinion in Structural Biology, 18, Liácer et al., Arginine and nitrogen storage, 673–681, 2008, with permission from Elsevier. **(C)** P_{II} subunit-NAGK subunit contacts. P_{II}, NAGK, and NAG are shown as strings, ribbons, and spheres, respectively. The contacting parts of the T-loop, B-loop, and β1-α1 connection, including some interacting side chains (in sticks), are blue, red, and green, respectively. The surfaces provided by these elements form meshworks of the same colors. The NAGK central β-sheet is green, and other β-strands and the α-helices are brownish and grayish for N- and C-domains, respectively. Some NAGK elements and P_{II} residues are labeled. This figure and its legend reproduce with some modifications a figure and its legend of Liácer et al. (2007). The crystal structure of the complex of P_{II} and acetylglutamate kinase reveals how P_{II} controls the storage of nitrogen as arginine. Copyright (2007) National Academy of Sciences. **(D,E)**, active and inactive conformations, respectively, of hexameric arginine-inhibitable NAGK. The active form is from a crystal of the enzyme from *Pseudomonas aeruginosa* (PDB 2BUF) while the inactive form is from the *Thermotoga maritima* enzyme (PDB 2BTY) (Ramón-Maiques et al., 2006). Note that the inactive form is widened relative to the active form, and that it has arginine sitting on both sides of each interdimeric junction. In the active form the nucleotide (in this case the product ADP rather than the substrate ATP) and NAG sit one in each domain of individual subunits. The NAGK observed in the P_{II}-NAGK complex is in the active form, being stabilized in this form by its contacts with P_{II}.

hybrid (both proteins involved) β -sheet with NAGK, forming also a hybrid ion-pair network (**Figure 4C**; Ll  cer et al., 2007). Apparently this flexing from an extended conformation could occur in two steps (Fokina et al., 2010b). The initial step would be mediated by a smaller loop of P_{II} called the B-loop (**Figures 2A, 4C**). P_{II} binding of 2OG also favors the flexing of the T-loop (**Figure 2C**, leftmost panel) (Fokina et al., 2010a; Truan et al., 2010) although the resulting conformation appears inappropriate for interacting with NAGK. In addition, 2OG can also promote the disassembly of the P_{II}-NAGK complex because certain P_{II} residues like Arg9 that are involved in the binding of 2OG are also involved in the interaction with NAGK (and also with PipX, another target of P_{II}, see below). Therefore, 2OG, an indicator of low ammonia levels (Muro-Pastor et al., 2001), abolishes P_{II}-NAGK complex formation (Maheswaran et al., 2004) (**Figure 1**). In *S. elongatus* 2OG can also promote the disassembly of the P_{II}-NAGK complex by favoring the phosphorylation of Ser49 (Forchhammer and Tandeau de Marsac, 1994; Irmeler et al., 1997), since the bound phosphate sterically prevents formation of the P_{II}-NAGK hybrid β -sheet (Ll  cer et al., 2007).

Plants and cyanobacteria stockpile ammonia as arginine, the protein amino acid with the largest nitrogen content (four atoms per arginine molecule). Arginine-rich proteins are very abundant in plant seeds (VanEtten et al., 1963). Cyanobacteria make non-ribosomally an arginine-rich amino acid polymer called cyanophycin (Oppermann-Sanio and Steinb  chel, 2002; Watzer and Forchhammer, 2018). The arginine stockpiling as arginine-rich macromolecules minimizes the osmotic effect while permitting rapid nitrogen mobilization for protein-building processes such as seed germination and cell multiplication. The selection of NAGK as the regulatory target stems from the fact that in many bacteria (including cyanobacteria) and in plants NAGK controls arginine synthesis via feed-back inhibition by L-arginine (Hoare and Hoare, 1966; Cunin et al., 1986; Lohmeier-Vogel et al., 2005; Beez et al., 2009). This inhibition must be overcome if large amounts of ammonia have to be stored as arginine (Ll  cer et al., 2008). Indeed, the P_{II}-NAGK complex exhibits decreased inhibition by arginine (Maheswaran et al., 2004; Ll  cer et al., 2008).

In arginine-sensitive NAGK (**Figures 4D,E**) the N-terminal α -helix of each subunit interacts with the same helix of an adjacent dimer, chaining three NAGK homodimers into a doughnut-shaped hexameric ring with three-fold symmetry and a central large hole (Ram  n-Maiques et al., 2006). The NAGK reaction (phosphorylation of the γ -COOH of N-acetyl-L-glutamate, NAG, by ATP) occurs within each NAGK subunit. NAG and ATP sit over the C-edge of the central 8-stranded largely parallel β sheet of the N-terminal and C-terminal domains, respectively (**Figure 4D**; Ram  n-Maiques et al., 2002). Catalysis requires the mutual approach of both domains of each subunit to allow the contact of the ATP terminal phosphate with the attacking NAG γ -COOH (Ram  n-Maiques et al., 2002; Gil-Ortiz et al., 2003). Arginine, by binding in each subunit next to the N-terminal α -helix (**Figure 4E**), expands the hexameric ring hampering the contact of the reacting groups and preventing catalysis (Ram  n-Maiques et al., 2006).

In the P_{II}-NAGK complex two P_{II} trimers sit on the three-fold axis of the complex, one on each side of the NAGK ring, making contacts with the inner circumference of this ring (**Figures 4A,B**). Each P_{II} subunit interacts via its T and B loops with each NAGK subunit (**Figure 4C**) gluing the two domains of this last subunit (**Figure 4A**). By restricting NAGK ring expansion (**Figure 4C**) even when arginine is bound, P_{II} renders NAGK highly active (Ll  cer et al., 2007; Mizuno et al., 2007). P_{II} does not compete physically with arginine for its sites on NAGK, simply these sites are widened in the P_{II}-NAGK complex (Ll  cer et al., 2007), resulting in decreased apparent affinity of NAGK for arginine (as reflected in the dependency of the NAGK activity on the arginine concentration). In addition, the hybrid P_{II}-NAGK ion pair network (**Figure 4C**) enhances the apparent affinity for NAG (assessed as the K_m or $S_{0.5}$ value of NAGK for NAG) of cyanobacterial NAGK (Maheswaran et al., 2004; Ll  cer et al., 2007). Overall, the NAGK bound to P_{II} exhibits decreased apparent affinity for arginine and increased activity, rendering NAGK much more active in the presence of arginine than when not bound to P_{II} (Ll  cer et al., 2008), something that is crucial for nitrogen storage as arginine.

NAGK appears to be a P_{II} target only in organisms performing oxygenic photosynthesis (cyanobacteria, algae, and plants, Burillo et al., 2004). P_{II} proteins from plants have lost the ability to bind ADP, while still binding ATP and 2OG (Lapina et al., 2018). In addition, except in *Brassicae*, the C-terminal part of plant P_{II} is extended to form two helical segments and a connecting loop (Q-loop; **Figure 2B**, rightmost panels), creating a novel glutamine site, resulting in glutamine-sensitivity of the P_{II}-NAGK interaction (Chellamuthu et al., 2014). This is not the case with cyanobacterial P_{II}, which binds both ADP and ATP and is glutamine-insensitive (Chellamuthu et al., 2014).

P_{II} has also been shown to interact in plants (Feria-Bourrellier et al., 2010) and bacteria, including cyanobacteria (Rodrigues et al., 2014; Gerhardt et al., 2015; Hauf et al., 2016), with the biotin carboxyl carrier protein (BCCP) of the enzyme acetyl coenzyme A carboxylase (AcCoA carboxylase) (**Figure 1**), although this complex has not been characterized structurally. BCCP is the component that hosts the covalently bound biotin that shuttles between the biotin carboxylase component and the transferase component of AcCoA carboxylase (Rubio, 1986). P_{II}-BCCP complex formation tunes down AcCoA utilization and thus subsequent fatty acid metabolism (Feria-Bourrellier et al., 2010; Gerhardt et al., 2015; Hauf et al., 2016), promoting uses of AcCoA for different purposes than the synthesis of fatty acids, and therefore linking P_{II} to AcCoA and fatty acid metabolism. For interaction, P_{II} has to be in the ATP-bound and 2OG-free form (**Figure 1**) (Gerhardt et al., 2015; Hauf et al., 2016), which are conditions at which P_{II} also binds to NAGK (Ll  cer et al., 2008). Therefore, there could be *in vivo* simultaneous activation of NAGK and inhibition of AcCoA carboxylase by P_{II}. Mutational evidence suggests the involvement of the T-loop in this interaction with AcCoA carboxylase (Hauf et al., 2016), in principle excluding P_{II}-NAGK-BCCP ternary complex formation and raising the possibility of competition between NAGK and BCCP for P_{II}.

The classical example of interaction of P_{II} with an enzymatic target was with the ATase of *E. coli* (see introductory section), with which uridylylated or deuridylylated GlnB (GlnB is one of the two P_{II} proteins of *E. coli*) can interact (Stadtman, 2001). We will not deal with this enzyme here because the P_{II}/ATase/GS cascade of enterobacteria does not appear to have general occurrence, for example in cyanobacteria, and also because we have only partial information on the structure of the ATase (Xu et al., 2004, 2010) and no direct information on the structure of the GlnB-ATase complex, although a model for such complex has been proposed (Palanca and Rubio, 2017).

THE PipX ADAPTOR PROTEIN AND ITS COMPLEX WITH P_{II}

A yeast two hybrid search for proteins interacting with P_{II} in *S. elongatus* identified (Burillo et al., 2004), in addition to NAGK, a small novel protein (89 amino acids) that was named PipX (P_{II}-interacting protein X). This protein was identified later in a search (Espinosa et al., 2006) for proteins interacting with NtcA, the global nitrogen regulator of cyanobacteria (Vega-Palas et al., 1992). PipX binding to P_{II} occurs under conditions of ammonia abundance (Figure 1), the same conditions prevailing for P_{II}-NAGK complex formation (Espinosa et al., 2006). NAGK-PipX competition for P_{II} was revealed in NAGK assays that showed that PipX decreased P_{II}-activation and increased arginine inhibition of NAGK (Llácer et al., 2010), excluding NAGK-P_{II}-PipX ternary complex formation. 2OG binding to P_{II} disassembles the P_{II}-PipX complex (Espinosa et al., 2006; Llácer et al., 2010), leaving PipX free to interact with NtcA (Figure 1).

The crystal structures of P_{II}-PipX complexes of *S. elongatus* (Llácer et al., 2010) and of *Anabaena* sp. PCC7120 (Zhao et al., 2010b) provided the first structural information on PipX (Figure 5A), revealing that it is formed by a compact body folded as a Tudor-like domain (a horseshoe-curved β -sheet sandwich) (Lu and Wang, 2013), followed by two C-terminal helices. In the P_{II}-PipX complex (Figure 5B) the three PipX molecules are enclosed in a cage formed between the flat face of the hemispheric P_{II} trimer and its three fully extended T-loops (see Figure 2C, 2nd panel from the left) emerging perpendicularly to the P_{II} flat face at its edge. The shape and orientation of these T-loops is very different relative to the P_{II} bound to NAGK, Figure 5C). In turn, the caged Tudor-like domains form a homotrimer over the P_{II} flat surface (Figure 5B, bottom), with the PipX self-interaction detected in yeast three-hybrid assays using P_{II} as bridging protein (Llácer et al., 2010). Tudor-like domains characteristically interact with RNA polymerase (Steiner et al., 2002; Deaconescu et al., 2006; Shaw et al., 2008), suggesting that PipX could have some role in gene expression that would be blunted by sequestration of these domains in the P_{II} cage.

In the structure of the *Anabaena* P_{II}-PipX complex, the two C-terminal helices of each PipX molecule lie one along the other in antiparallel orientation ("flexed"), being exposed between two adjacent T-loops in transversal orientation relative to these loops (Zhao et al., 2010b). Recent structural NMR data on isolated PipX showed that when PipX is alone (that

is, not bound to a partner) the C-terminal helices are "flexed" (Figure 5A) (Forcada-Nadal et al., 2017). As shown below, the C-terminal helices of PipX in the NtcA-PipX complex are also flexed (Llácer et al., 2010). However, in the *S. elongatus* P_{II}-PipX complex only one PipX molecule presents the "flexed" conformation, whereas in the other two PipX molecules the C-terminal helix is "extended," not contacting the previous helix and emerging centrifugally outwards from the complex, between two T-loops (Figure 5B) (Llácer et al., 2010). P_{II} binding might facilitate the extension of the PipX C-terminal helix, endowing the P_{II}-PipX complex with a novel surface and novel potentialities for interaction with other components. These novel potentialities were substantiated recently by the identification, in yeast three-hybrid searches (Labella et al., 2016), of interactions of PipX in the P_{II}-PipX complex with the homodimeric transcription factor PlmA (see proposal for the architecture of this complex in Figure 1 bottom left; Labella et al., 2016). Interactions were not observed in yeast two-hybrid assays between PipX or P_{II} and PlmA. Residues involved in three-hybrid interactions, mapped by site-directed mutagenesis, are largely localized in the C-terminal helix of PipX. PlmA belongs to the GntR super-family of transcriptional regulators, but is unique to cyanobacteria (Lee et al., 2003; Hoskisson and Rigali, 2009; Labella et al., 2016). Little is known about PlmA functions other than it is involved in plasmid maintenance in *Anabaena* sp. strain PCC7120 (Lee et al., 2003), in photosystem stoichiometry in *Synechocystis* sp. PCC6803 (Fujimori et al., 2005), in regulation of the highly conserved cyanobacterial sRNA YFR2 in marine picocyanobacteria (Lambrecht et al., 2018), and that it is reduced by thioredoxin, without altering its dimeric nature in *Synechocystis* sp. PCC6803 (Kujirai et al., 2018). The P_{II}-PipX-PlmA ternary complex suggests that PipX can influence gene expression regulation via PlmA, although the PlmA regulon remains to be defined.

THE GENE EXPRESSION REGULATOR NtcA

When ammonia becomes scarce the increasing 2OG levels should determine the disassembly of the P_{II}-NAGK, P_{II}-BCCP, and P_{II}-PipX complexes (Figure 1). These same conditions promote the binding of PipX (see below) to the transcriptional regulator NtcA (Figure 1), an exclusive cyanobacterial factor of universal presence in this phylogenetic group (Vega-Palas et al., 1992; Herrero et al., 2001; Körner et al., 2003). The determination of the structures of NtcA from *S. elongatus* (Figures 6A,B) (Llácer et al., 2010) and from *Anabaena* sp. PCC7120 (Zhao et al., 2010a) confirmed the sequence-based inference (Vega-Palas et al., 1992) that NtcA is a homodimeric transcriptional regulator of the family of CRP (the cAMP-regulated transcriptional regulator of *E. coli*) (McKay and Steitz, 1981; Weber and Steitz, 1987). Similarly to CRP, NtcA has a C-terminal DNA binding domain of the helix-turn-helix type. In CRP, the DNA binding helices of its two C-terminal domains are inserted in two adjacent turns of the major groove of DNA that host the imperfectly palindromic target DNA sequence (called here the CRP box)

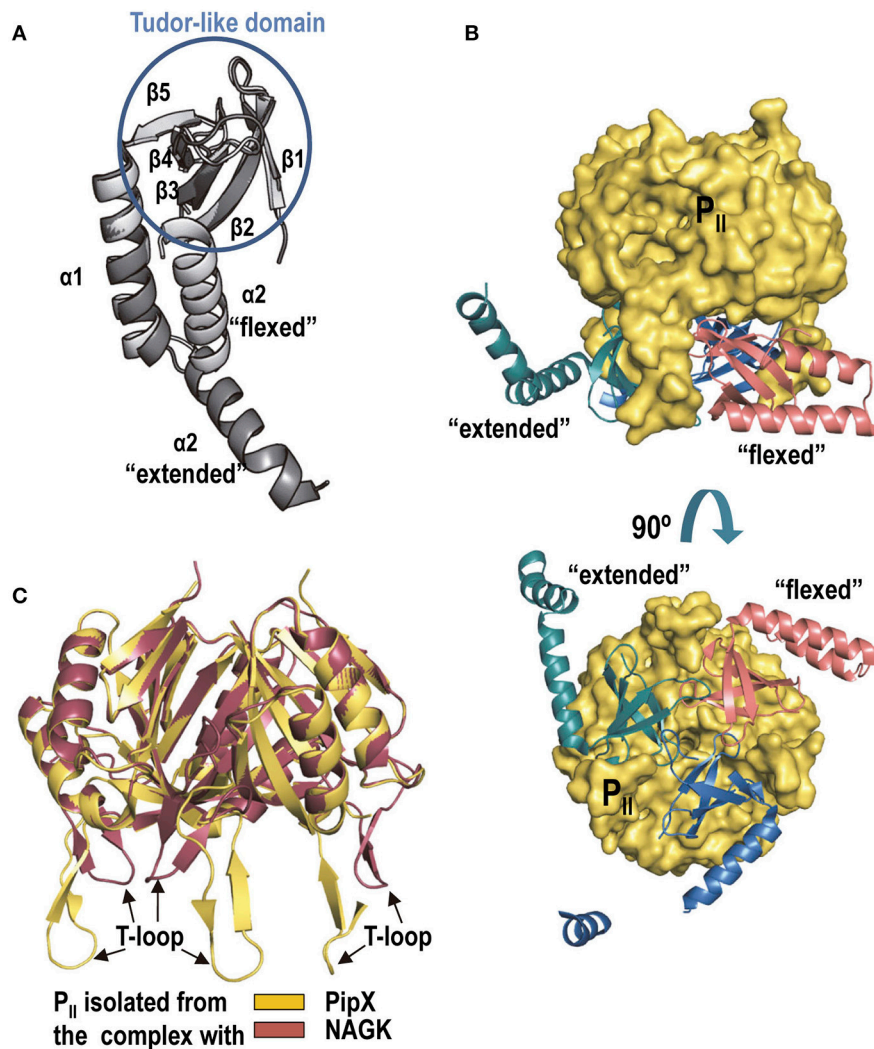


FIGURE 5 | PipX and the P_{II}-PipX complex. **(A)** The structures of PipX in the flexed conformation and in one of the two extended conformations observed in the *S. elongatus* P_{II}-PipX complex are shown (extracted from PDB 2XG8; Ll  cer et al., 2010). Note that the major difference between the two conformations is the large movement of the C-terminal helix around its flexible linker (Forcada-Nadal et al., 2017). The same flexed conformation was observed in the complex with NtcA (see below) and agrees with the data of structural NMR studies on isolated PipX (Forcada-Nadal et al., 2017). The elements of the Tudor-like domain are encircled in a blue circumference. **(B)** The P_{II}-PipX complex of *S. elongatus* viewed with the three-fold axis of P_{II} vertical (top) or in a view along this axis, looking at the flat face of P_{II} (bottom). **(C)** Superimposition of *S. elongatus* P_{II} in the complex with PipX and in that with NAGK. The changes in the T-loops are very patent.

(McKay and Steitz, 1981; Weber and Steitz, 1987). The consensus DNA sequence to which NtcA binds (consensus NtcA box) is quite similar to the consensus CRP box (Berg and von Hippel, 1988; Luque et al., 1994; Jiang et al., 2000; Herrero et al., 2001; Omagari et al., 2008), and thus NtcA and CRP are expected to bind in similar ways to their target DNA sequences.

In vitro studies revealed that 2OG is an NtcA activator (Tanigawa et al., 2002; V  zquez-Berm  dez et al., 2002), increasing NtcA affinity for its target sequences. As in the case of cAMP for CRP, 2OG binds to the NtcA regulatory domain. This domain is responsible for the NtcA dimeric nature (Figure 6A) (Ll  cer et al., 2010; Zhao et al., 2010a). The regulatory domain of NtcA is highly similar to the corresponding domain of CRP (Ll  cer

et al., 2010). The main differences reflect the changes in the characteristics of the site for the allosteric effector that enable the accommodation of 2OG instead of cAMP. Each 2OG molecule interacts in NtcA with the two (one per subunit) long interfacial helices that form the molecular backbone, crossing the molecule in its longer dimension, linking in each subunit both domains (Figures 6A,C) (Ll  cer et al., 2010; Zhao et al., 2010a). 2OG interactions with both interfacial helices favor a twist of one helix relative to the other, dragging the DNA binding domains and helices to apparently appropriate positions and interhelical distance for binding in two adjacent turns of the major groove of DNA where the NtcA box should be found (Figures 6A,B, and inset therein), although the experimental structure of

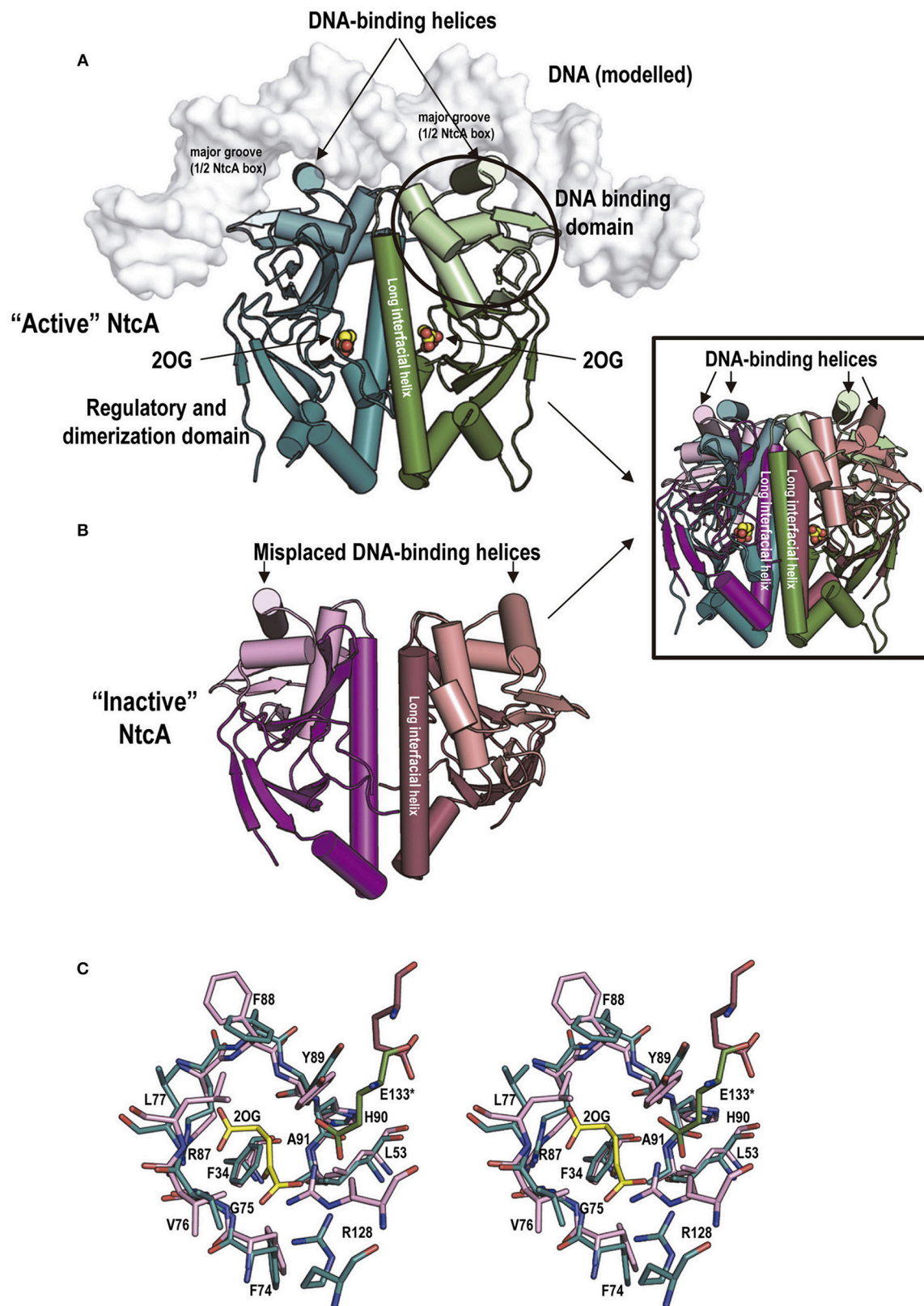


FIGURE 6 | Continued

FIGURE 6 | NtcA structure, 2OG binding to it and associated conformational changes. **(A,B)**, structures of “active” **(A)** and “inactive” **(B)** *S. elongatus* NtcA (PDB files 2XHK and 2XGX, respectively) (Llácer et al., 2010). The two subunits of each dimer are in different colors, with the DNA-binding domains in a lighter hue than the regulatory domain of the same subunit. In the cartoon representation used, helices are shown as cylinders to illustrate best the changes in position of the DNA binding helices and of the long interfacial helices (labeled) upon activation. Bound 2OG is shown in “active” NtcA (in spheres representation, with C and O atoms colored yellow and red, respectively). The DNA, in surface representation in white, has been modeled from the CRP-DNA structure (for details see Llácer et al., 2010). The inset superimposes the “active” and “inactive” forms colored as in the main figure to illustrate the magnitude of the changes. **(C)** Stereo view of sticks representation of the 2OG site residues of the “active” (green) and “inactive” (raspberry) forms of NtcA. The 2OG bound to the “active” form is distinguished by its yellow C atoms. Note that only two residues, both 2OG-interacting and highly polar, experience large changes in their positions between the inactive and the active forms: Arg128 from the long interfacial helix of the subunit that provides the bulk of the residues of the site, and Glu133 from the interfacial helix of the other subunit. They are believed to trigger the changes in the relations between the two interfacial helices that result in NtcA “activation”.

DNA-bound NtcA should be determined to corroborate this proposals.

Although NtcA and CRP boxes are quite similar, plasmon resonance experiments (Forcada-Nadal et al., 2014) revealed that CRP exhibits complete selectivity and specificity for the CRP box, with absolute dependency on the presence of cAMP. In contrast, NtcA had less strict selectivity, since it still could bind to its promoters in the absence of 2OG, although with reduced affinity, and it could also bind to the the CRP promoter tested. Nevertheless, it is unlikely that NtcA could bind *in vivo* to the CRP boxes of cyanobacteria in those species where CRP is also present, given the much higher affinities for the CRP sites of cyanobacterial CRP and the relative concentrations in the cell of both transcriptional regulators (Forcada-Nadal et al., 2014).

While the structures of 2OG-bound NtcA of *S. elongatus* (Llácer et al., 2010) and of *Anabaena* (Zhao et al., 2010a) are virtually identical, the reported structures of “inactive” NtcA of *Anabaena* without 2OG (Zhao et al., 2010a) and of *S. elongatus* (Llácer et al., 2010) differed quite importantly in the positioning of the DNA binding domains (**Figure 1**, “Inactive forms” under “NtcA”), although in both cases the DNA binding helices were misplaced for properly accommodating the NtcA box of DNA, raising the question of whether these structural differences are species-specific or whether “inactive” NtcA can be in a multiplicity of conformations.

PipX AS AN NtcA CO-ACTIVATOR

Soon after PipX was found to interact with NtcA (Espinosa et al., 2006), it was also shown to activate *in vivo* transcription of NtcA-dependent promoters under conditions of low nitrogen availability (Espinosa et al., 2006, 2007). Direct binding studies with the isolated molecules proved that PipX binding to NtcA requires 2OG (Espinosa et al., 2006). Nevertheless, as PipX was not totally essential for transcription of NtcA-dependent promoters (Espinosa et al., 2007; Camargo et al., 2014), it was concluded (1) that PipX was a coactivator of 2OG-activated NtcA-mediated transcription (Espinosa et al., 2006; Llácer et al., 2010); and (2) that the degree of activation by PipX depended on the specific NtcA-dependent promoter (Espinosa et al., 2007; Forcada-Nadal et al., 2014). Detailed plasmon resonance studies (Forcada-Nadal et al., 2014) using sensorchip-bound DNA confirmed for three *Synechocystis* promoters that PipX binding to promoter-bound NtcA has an absolute requirement for 2OG, since no PipX binding was observed when NtcA was bound to the DNA in absence of 2OG. In these studies PipX

increased about one order of magnitude the apparent affinity of NtcA for 2OG. In other *in vitro* experiments with four NtcA-dependent promoters of *Anabaena* sp. PCC 7120, PipX was also found to positively affect NtcA binding to its DNA sites (Camargo et al., 2014). The induction by PipX of increased NtcA affinity for 2OG and for its promoters could account for the PipX-triggered enhancement of NtcA-dependent transcription.

The crystal structure of the NtcA-PipX complex of *S. elongatus* (**Figure 7**) (Llácer et al., 2010) corresponded to one “active” NtcA dimer with one molecule of each 2OG and PipX bound to each subunit. PipX is inserted via its Tudor-like domain (**Figure 7A**), filling a crater-like cavity formed over each NtcA subunit (**Figure 7B**) largely over one regulatory domain, being limited between the DNA binding domain and the long interfacial helix of the same subunit, and the regulatory domain of the other subunit. PipX extensively interacts with the entire crater, with nearly 1200 Å² of NtcA surface covered by each PipX molecule, of which 65% belongs to one subunit (40%, 15% and 10% belonging to the DNA-binding domain, the interfacial helix and the regulatory domain, respectively) and 35% belongs to the regulatory domain (including the interfacial helix) of the other subunit, gluing together the elements of half of the NtcA dimer in its active conformation, stabilizing this conformation (Llácer et al., 2010). This conformation is the one that binds 2OG and that should have high affinity for the DNA, thus explaining the requirement of 2OG for PipX binding and the increased affinities of NtcA for 2OG and DNA when PipX was bound to NtcA (Forcada-Nadal et al., 2014). Since a similar crater-like cavity exists in other transcription factors of the CRP family including CRP (see for example McKay and Steitz, 1981 or Weber and Steitz, 1987) it would be conceivable that PipX-mimicking proteins could exist for these other transcriptional regulators of the CRP family, although PipX cannot do such a role since it does not bind to CRP (Forcada-Nadal et al., 2014). Furthermore, a large set of highly specific contacts (Llácer et al., 2010) ensure the specificity of the binding of PipX to NtcA.

The elements of the Tudor-like domain that interact with NtcA are largely the same that interact with the flat surface of the hemispheric body of P_{II} (many of them mediated by the upper layer of the Tudor-like β-sandwich, particularly strands β1 and β2), predicting total incompatibility for the simultaneous involvement of PipX in the NtcA and P_{II} complexes (Llácer et al., 2010). While the Tudor-like domain monopolizes the contacts of PipX with NtcA, the C-terminal helices of PipX do not participate in these contacts and remain flexed, as in isolated PipX (Forcada-Nadal et al., 2017), protruding away from the

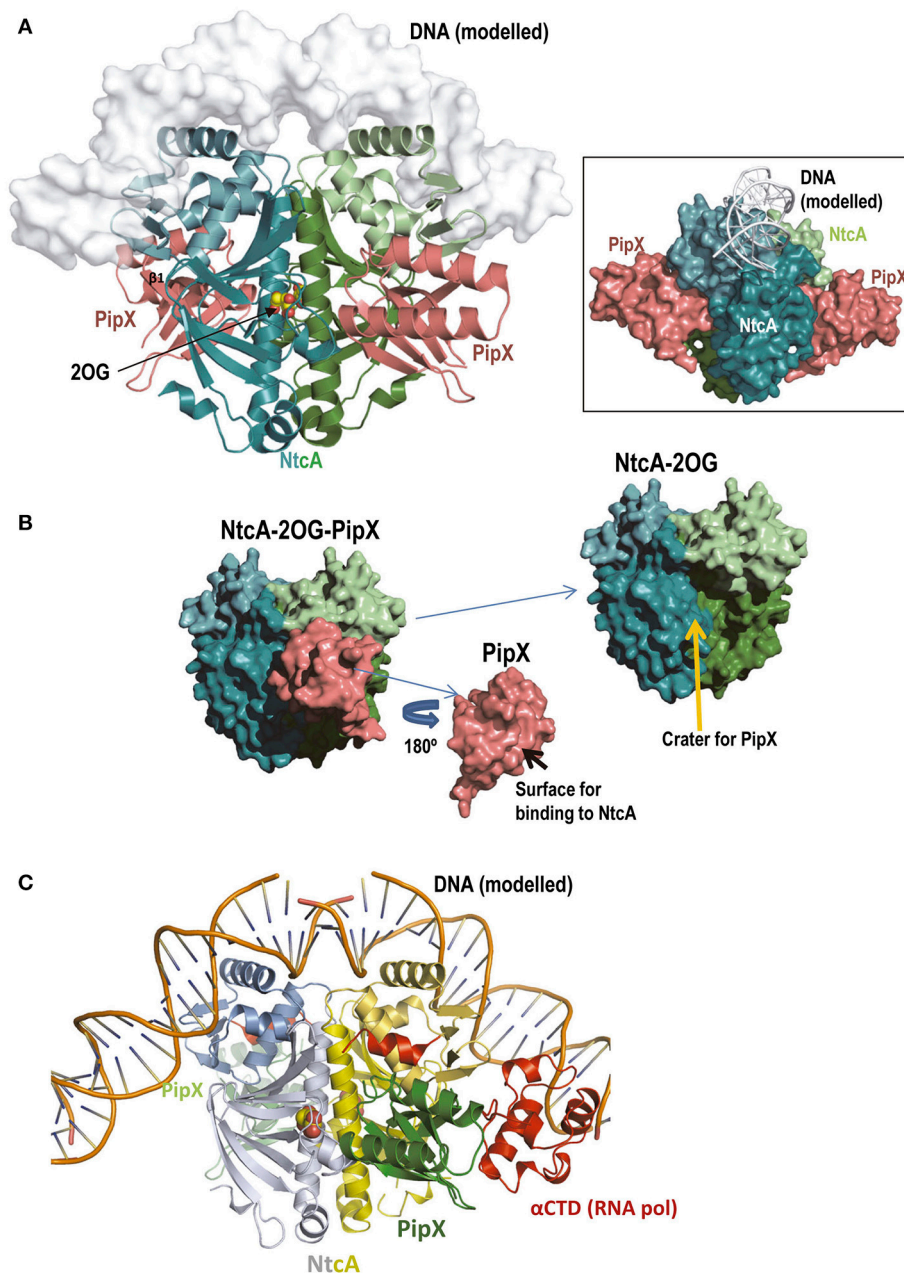


FIGURE 7 | The NtcA-PipX complex. **(A)** Structure of the complex of *S. elongatus* NtcA and PipX, with DNA modeled in semi-transparent surface representation (Llácer et al., 2010). The projection of NtcA differs somewhat from that in **Figure 6A** to allow visualization of both PipX molecules, illustrating the fact that the Tudor-like domain is the part of PipX that binds to NtcA. Note that the two PipX molecules (the asymmetric unit contained an entire complex with two PipX molecules, PDB file 2XKO) are in the “flexed” conformation, that the flexed helices protrude away from the NtcA molecule and that they do not contact the modeled bound DNA (main figure and inset). The *inset* represents the same complex in a different orientation (viewed approximately along the DNA) and with all elements in surface representation except the DNA (shown in cartoon) to better visualize the protrusion in large part due to the flexed helices of both PipX molecules. **(B)** Deconstruction of the NtcA-PipX complex to show in surface representation the crater at the NtcA surface where PipX binds, and the surface of PipX used in this binding. **(C)** A model based on CRP (Llácer et al., 2010) for the complex of the NtcA-PipX complex with DNA and with the C-terminal domain of the α-subunit of RNA polymerase (αCTD), to show that the C-terminal helices of PipX could reach this part of the polymerase. In this figure the C-terminal helix of NtcA is colored red because it has no counterpart in CRP and is involved in the interactions with PipX.

complex (**Figure 7A**). The coactivation functions of PipX for NtcA-mediated transcription could also involve these helices. However, *in vitro* experiments (Llácer et al., 2010; Camargo

et al., 2014) and modeling (based on CRP) of DNA binding by the NtcA-PipX complex (**Figure 7A** and *inset* therein) (Llácer et al., 2010) did not support the idea of PipX binding to DNA.

Alternatively, these helices could interact with RNA polymerase, particularly given the location, in the homologous CRP-DNA complex, of the binding site for the α -subunit of the C-terminal domain (α CTD) of RNA polymerase (Figure 7C; and discussed in Ll  cer et al., 2010). Further structures of P_{II}-PipX bound to DNA alone or with at least some elements of the polymerase are needed to clarify these issues.

THE PipX REGULATORY NETWORK IN QUANTITATIVE PERSPECTIVE

The gene encoding P_{II} could not be deleted in *S. elongatus* unless the *pipX* gene was previously inactivated (Espinosa et al., 2009). Further studies led to the conclusion that decreasing the P_{II}/PipX ratio results in lethality in *S. elongatus*, indicating that PipX sequestration into P_{II}-PipX complexes is crucial for survival and implicating both proteins in the regulation of essential processes (Espinosa et al., 2010, 2018; Laichoubi et al., 2012). The ability of P_{II} to prevent the toxicity of PipX suggests that P_{II} acts as a PipX sink even under conditions in which the affinity for NtcA would be highest, supporting the idea that not all PipX effects are related to its role as NtcA co-activator. Mutational studies (Espinosa et al., 2009, 2010) and massive transcriptomic studies of *S. elongatus* mutants centered on PipX (Espinosa et al., 2014) also support the multifunctionality of PipX, stressing the need for additional studies, including the determination of PlmA functions and the search for further potential PipX-interacting proteins (Figure 1, broken yellow arrow).

Massive proteomic studies (Guerreiro et al., 2014) have estimated the number of chains of each protein of the P_{II}-PipX network (Figure 1) in *S. elongatus* cells. The values obtained (Table 1) are corroborated by those obtained in focused western blot studies for some of these macromolecules (Table 1) (Labella et al., 2016, 2017). These quantitative data give an opportunity to evaluate the possible frequency of the different complexes and macromolecules of the P_{II}-PipX-NtcA network (Figure 1) in one or another form (schematized in Figure 8). Of all the proteins mentioned here until now, P_{II} is by far the most abundant in terms of polypeptide chains (Table 1). In comparison, the sum of all the chains of other known P_{II}-binding proteins represents no more than 20% of the P_{II} chains. Among these molecules is PipX, which only represents ~10% of all the P_{II} chains. This indicates that P_{II} has the potential to sequester all the PipX that is present in the cell (Figure 8). In turn, PlmA could be fully trapped in the P_{II}-PipX-PlmA complex if this complex has the 1:1:1 stoichiometry proposed for it (Figure 1) (Labella et al., 2016), since the number of PlmA chains only represent about 10% of the number of PipX chains. Thus, about 10 and 1% of the P_{II} trimer could be as the respective P_{II}-PipX and P_{II}-PipX-PlmA complexes under nitrogen abundance conditions. In contrast, with nitrogen starvation all of the NtcA could be bound to PipX (Figure 8), given the ~five-fold excess of PipX chains over NtcA chains (Table 1). Thus, assuming that under conditions at which 2OG and ATP reach high levels P_{II} is totally unable to bind PipX, ~80% of the PipX molecules could be free to interact with additional protein partners.

TABLE 1 | Levels of macromolecular players of the P_{II}-NAGK-PipX-NtcA system in *S. elongatus*.

| Protein | Protein chains ^a | | |
|---------------------|-----------------------------|--|-------|
| | N ^o copies/cell | As percentage relative to P _{II} ^b | |
| P _{II} | 60,000 | 100 | (100) |
| PipX | 4,560 | ~8 | (7) |
| NAGK | 2,150 | ~4 | |
| BCCP | 2,410 | 4 | |
| NtcA | 1,000 | ~2 | |
| PlmA | 400 | ~1 | (1) |
| PipY ^c | 1,275 | 2 | (1.6) |
| PamA ^{c,d} | 36 | < 0.1 | |

When given as percentages, the data are relative to the number of P_{II} chains (given the value of 100).

^aData from massive proteomic study (Guerreiro et al., 2014). Percentages within parentheses are data based on immunoquantification in Western blots (Labella et al., 2016, 2017).

^bRounded to the closest integer or to the closest first decimal figure.

^cGiven for reference, since there is no evidence of physical interaction with any of the other proteins.

^dThe physical interaction of this putative channel with P_{II} was found in *Synechocystis* sp. PCC6803, but the findings were not replicated in *S. elongatus* with the homologous product of gene *Synpcc7942_0610*.

These inferences are consistent with the K_D values for the P_{II}-NAGK (Ll  cer et al., 2007) and P_{II}-PipX complexes in the absence of 2OG (Ll  cer et al., 2010) and for the PipX-NtcA complex at high 2OG and ATP (Forcada-Nadal et al., 2014) (~0.08, 7 and 0.09 μ M, respectively). For the estimated cellular levels of the different components (Table 1), assuming a cell volume of 10⁻¹² ml, virtually all the NAGK and ~95% of the PipX could be P_{II}-bound in the absence of 2OG, and ~98% of the NtcA could be PipX-complexed in the presence of 2OG. However, the impacts of varying concentrations of 2OG on the disassembly or assembly of the complexes most likely differ for the various complexes. For example, a two-order of magnitude increase in the K_D value for the P_{II}-NAGK complex due to 2OG binding might have much less impact (a 7% decrease in the amount of NAGK bound to P_{II} would be estimated from the mere total protein levels and K_D value) than a two-order of magnitude increase in the K_D for the P_{II}-PipX complex (an 80% decrease in the amount of PipX bound estimated similarly). These estimations are very crude, since they do not take in consideration that in *S. elongatus* P_{II} phosphorylation prevents NAGK binding (Heinrich et al., 2004), and that this phosphorylation is greatly influenced by the abundance of ammonia (Forchhammer and Tandeau de Marsac, 1994). Furthermore, we have not considered in these estimates the influence of the ATP concentrations, recently shown to decrease *in vivo* in *S. elongatus* upon nitrogen starvation (Doello et al., 2018). Therefore, the situation is much more complex than would be expected from the mere consideration of the abundances of the different proteins and of the K_D values for the non-phosphorylated form of P_{II}. Nevertheless, it appears desirable to estimate the influence of different 2OG levels on K_D values as an important element to take into consideration in future attempts to model the concentrations of the different

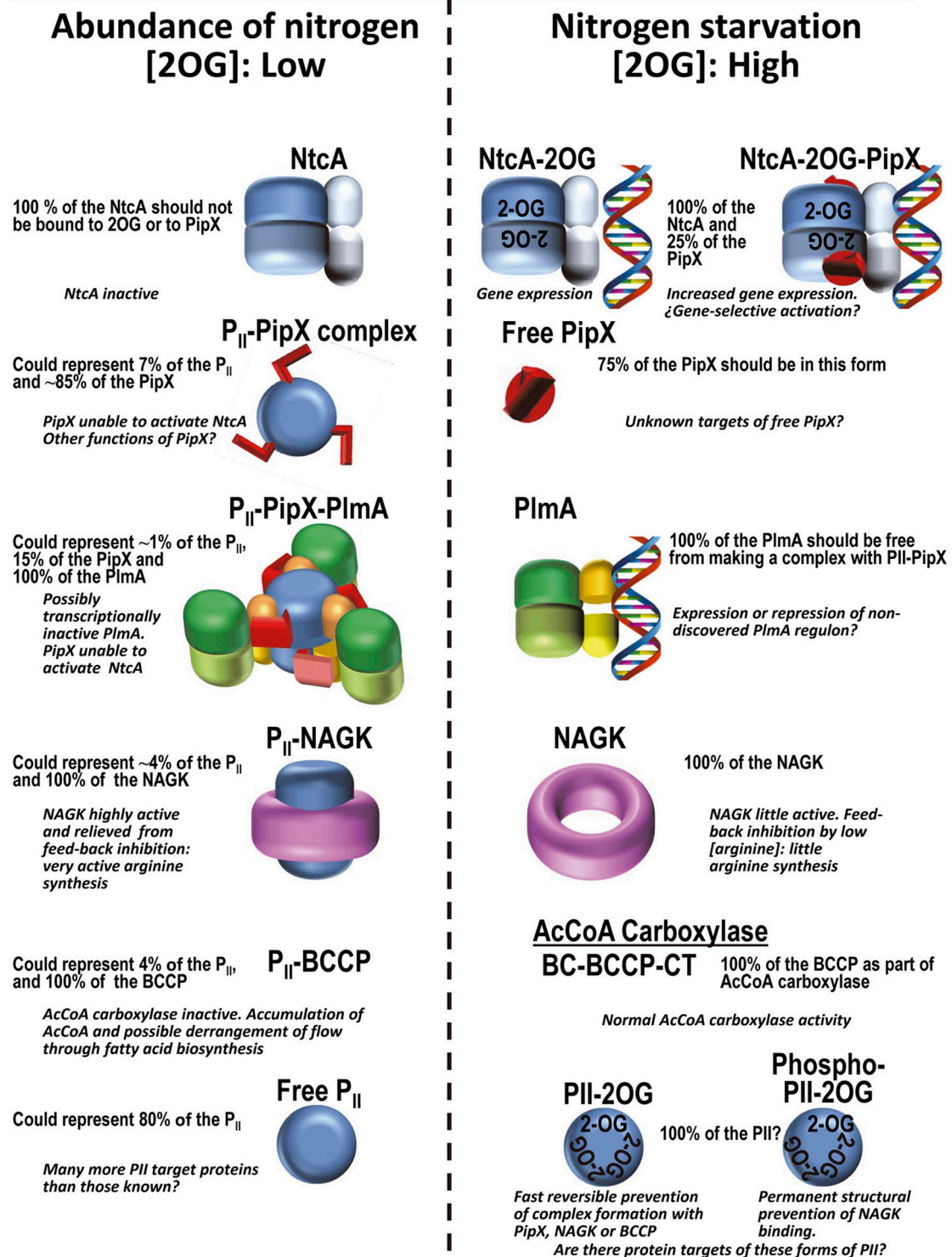


FIGURE 8 | Protein complexes of the P_{II} regulatory system in *S. elongatus* according to availability of ammonium in the cell, and their corresponding functional consequences. The frequencies of the different chains in the various forms are based on the levels of the proteins in the cell found in proteomic studies (Table 1). The P_{II} trimer has been colored blue, PipX and its C-terminal helices are red, PlmA dimers have their DNA binding domains yellow or orange and their dimerization domains in two hues of green, NAGK is shown as a purple crown, and the regulatory and DNA binding domains of NtcA are given dark and light shades of blue, respectively.

complexes of P_{II} and PipX under intermediate conditions of nitrogen richness.

A NOVEL NETWORK MEMBER FUNCTIONALLY RELATED TO PipX

Recently, a novel protein has been identified as belonging to the PipX regulatory network (**Figure 1**, yellow 3D arrow projecting upwards from the plane of the network). In this case no direct protein-protein interaction with PipX has been shown (Labella et al., 2017). This protein (PipY), is the product of the downstream gene in the bicistronic *pipXY* operon. The regulatory influence of PipX on PipY was originally detected in functional, gene expression and mutational studies (Labella et al., 2017). More recently it has been concluded (Cantos et al., 2018) that PipX enhances *pipY* expression in *cis*, preventing operon polarity, a function that might implicate additional interactions of PipX with the transcription and translation machineries, by analogy with the action of NusG paralogues, which are proteins bearing, as PipX, Tudor-like domains. It has been proposed (Cantos et al., 2018) that the *cis*-acting function of PipX might be a sophisticated strategy for keeping the appropriate PipX-PipY stoichiometry.

PipY is an intriguing pyridoxal phosphate-containing protein that is folded as a modified TIM-barrel (**Figure 1**). The PipY structure (Tremiño et al., 2017) gives full structural backing to unsuccessful experimental attempts to show enzymatic activity of PipY and its orthologs (Ito et al., 2013). Because of these negative findings, and given the pleiotropic effects of the inactivation of the PipY orthologs in microorganisms and humans, it has been concluded that these proteins have as yet unclarified roles in PLP homeostasis (Ito et al., 2013; Darin et al., 2016; Prunetti et al., 2016; Labella et al., 2017; Plecko et al., 2017). Interestingly, these proteins are widespread across phyla (Prunetti et al., 2016) and the deficiency of the human ortholog causes vitamin B₆-dependent epilepsy (Darin et al., 2016; Plecko et al., 2017; Tremiño et al., 2018), providing an excellent example that investigations of cyanobacterial regulatory systems like the one summarized here can have far-reaching consequences spanning up to the realm of human and animal pathology. If any lesson can be inferred from all of the above, is that the investigation on P_{II} and particularly on PipX proteins require further efforts.

FINAL REMARKS

The rich P_{II} regulatory network summarized in **Figure 1** of this review even for a unicellular microorganism with a single type of P_{II} protein, attests the importance of P_{II} and of its regulatory processes. This importance, possibly underrecognized until now, is highlighted, for example, by the very wide distribution of P_{II} proteins among microorganisms and plants. Furthermore, in the many organisms with several genes for P_{II} proteins the levels of complexity expected from P_{II} regulatory networks may be much greater than the one presented here. Each one of these paralogous

P_{II} proteins may command a regulatory network, and it would be unlikely that these networks would not be interconnected into a large meshwork that will require the instruments of systems biology to be fully understood.

PipX also deserves deeper attention than received until now. Massive transcriptomics studies (Espinosa et al., 2014) have ascribed to this protein a paramount regulatory role in *S. elongatus*. For full understanding of this role further searches for PipX-interacting or functional partners like PipY appear desirable, with detailed investigation of the molecular mechanisms of the physical or the functional interactions. PlmA merits particular attention to try to characterize the roles of the P_{II}-PipX-PlmA complex. PipY and its orthologs deserve similar attention, to try to define molecularly their PLP homeostatic functions, a need that is made more urgent by the role in pathology of the human ortholog of PipY. In addition to all of this, the structural evidence reviewed here makes conceivable that adaptor proteins capable of stabilizing active conformations of other transcriptional regulators of the CRP family could exist outside cyanobacteria, mimicking the PipX role. In summary, there are many important questions to be addressed arising from the field reviewed here, some within cyanobacteria, but others concerning whether the mechanisms and complexes exemplified here could have a parallel in other bacterial or even plant species. Clearly more investigations on P_{II} and its partners in other phylogenetic groups using the approaches and experimental instruments used to uncover the cyanobacterial P_{II} regulatory network would appear highly desirable.

AUTHOR CONTRIBUTIONS

AF-N, JLL, AC, CM-M, and VR reviewed the literature and their own previous work and contributed to the discussions for writing this review. The main writer was VR, but all the authors contributed to the writing of the manuscript. AF-N and CM-M prepared the figures, with key inputs from VR.

FUNDING

Supported by grants BFU2014-58229-P and BFU2017-84264-P from the Spanish Government.

ACKNOWLEDGMENTS

We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI). We are grateful to ESRF (Grenoble, France), Diamond (Oxfordshire, UK) and Alba (Barcelona, Spain) synchrotrons for access and for staff support to collect the data used for the determination by our group of most of the structures mentioned in this paper, which have been previously published as referred. CM-M holds a contract of CIBERER.

REFERENCES

- Beez, S., Fokina, O., Herrmann, C., and Forchhammer, K. (2009). N-acetyl-L-glutamate kinase (NAGK) from oxygenic phototrophs: P_{II} signal transduction across domains of life reveals novel insights in NAGK control. *J. Mol. Biol.* 389, 748–758. doi: 10.1016/j.jmb.2009.04.053
- Berg, O. G., and von Hippel, P. H. (1988). Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. *J. Mol. Biol.* 200, 709–723.
- Burillo, S., Luque, I., and Fuentes, I. (2004). Interactions between the nitrogen signal transduction protein P_{II} and N-acetyl glutamate kinase in organisms that perform oxygenic photosynthesis. *J. Bacteriol.* 186, 3346–3354. doi: 10.1128/JB.186.11.3346-3354.2004
- Camargo, S., Valladares, A., Forchhammer, K., and Herrero, A. (2014). Effects of PipX on NtcA-dependent promoters and characterization of the cox3 promoter region in the heterocyst-forming cyanobacterium *Anabaena* sp. PCC 7120. *FEBS Lett.* 588, 1787–1794. doi: 10.1016/j.febslet.2014.03.039
- Cantos, C., Labella, J. I., Espinosa, E., and Contreras, A. (2018). The nitrogen regulator PipX acts *in cis* to prevent operon polarity. *Environ. Microbiol. Rep.* doi: 10.1111/1758-2229.12688
- Chellamuthu, V. R., Ermilova, E., Lapina, T., Lüddecke, J., Minaeva, E., Herrmann, C., et al. (2014). A widespread glutamine-sensing mechanism in the plant kingdom. *Cell* 159, 1188–1199. doi: 10.1016/j.cell.2014.10.015
- Chin, G. J. (2008). Editors Choice: Three times two is six. *Science* 318:888. doi: 10.1126/science.318.5852.888a
- Conroy, M. J., Durand, A., Lupo, D., Li, X. D., Bullough, P. A., Winkler, F. K., et al. (2007). The crystal structure of the *Escherichia coli* AmtB-GlnK complex reveals how GlnK regulates the ammonia channel. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1213–1218. doi: 10.1073/pnas.0610348104
- Cunin, R., Glansdorff, N., Pierard, A., and Stalon, V. (1986). Biosynthesis and metabolism of arginine in bacteria. *Microbiol. Rev.* 50, 314–352.
- Darin, N., Reid, E., Prunetti, L., Samuelsson, L., Husain, R. A., Wilson, M., et al. (2016). Mutations in PROSC disrupt cellular pyridoxal phosphate homeostasis and cause vitamin- B6-dependent epilepsy. *Am. J. Hum. Genet.* 99, 1325–1337. doi: 10.1016/j.ajhg.2016.10.011
- Deaconescu, A. M., Chambers, A. L., Smith, A. J., Nickels, B. E., Hochschild, A., Savary, N. J., et al. (2006). Structural basis for bacterial transcription-coupled DNA repair. *Cell* 124, 507–520. doi: 10.1016/j.cell.2005.11.045
- Doello, S., Klotz, A., Makowka, A., Gutekunst, K., and Forchhammer, K. (2018). A specific glycogen mobilization strategy enables rapid awakening of dormant cyanobacteria from chlorosis. *Plant Physiol.* 177, 594–603. doi: 10.1104/pp.18.00297
- Espinosa, J., Castells, M. A., Laichoubi, K. B., and Contreras, A. (2009). Mutations at PipX suppress lethality of P_{II}-deficient mutants of *Synechococcus elongatus* PCC 7942. *J. Bacteriol.* 191, 4863–4869. doi: 10.1128/JB.00557-09
- Espinosa, J., Castells, M. A., Laichoubi, K. B., Forchhammer, K., and Contreras, A. (2010). Effects of spontaneous mutations in PipX functions and regulatory complexes on the cyanobacterium *Synechococcus elongatus* strain PCC 7942. *Microbiology* 156, 1517–1526. doi: 10.1099/mic.0.037309-0
- Espinosa, J., Forchhammer, K., Burillo, S., and Contreras, A. (2006). Interaction network in cyanobacterial nitrogen regulation: PipX, a protein that interacts in a 2-oxoglutarate dependent manner with P_{II} and NtcA. *Mol. Microbiol.* 61, 457–469. doi: 10.1111/j.1365-2958.2006.05231.x
- Espinosa, J., Forchhammer, K., and Contreras, A. (2007). Role of the *Synechococcus* PCC 7942 nitrogen regulator protein PipX in NtcA controlled processes. *Microbiology* 153, 711–718. doi: 10.1099/mic.0.2006/003574-0
- Espinosa, J., Labella, J. I., Cantos, R., and Contreras, A. (2018). Energy drives the dynamic localization of cyanobacterial nitrogen regulators during diurnal cycles. *Environ. Microbiol.* 20, 1240–1252. doi: 10.1111/1462-2920.14071
- Espinosa, J., Rodríguez-Mateos, F., Salinas, P., Lanza, V. F., Dixon, R., de la Cruz, F., et al. (2014). PipX, the coactivator of NtcA, is a global regulator in cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2423–2430. doi: 10.1073/pnas.1404097111
- Feria-Bourrellier, A. B., Valot, B., Guillo, A., Ambard-Bretteville, F., Vidal, J., and Hodges, M. (2010). Chloroplast acetyl-CoA carboxylase activity is 2-oxoglutarate-regulated by interaction of P_{II} with the biotin carboxyl carrier subunit. *Proc. Natl. Acad. Sci. U.S.A.* 107, 502–507. doi: 10.1073/pnas.0910097107
- Fokina, O., Chellamuthu, V. R., Forchhammer, K., and Zeth, K. (2010a). Mechanism of 2-oxoglutarate signaling by the *Synechococcus elongatus* P_{II} signal transduction protein. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19760–19765. doi: 10.1159/000320215
- Fokina, O., Chellamuthu, V. R., Zeth, K., and Forchhammer, K. (2010b). A novel signal transduction protein P_{II} variant from *Synechococcus elongatus* PCC 7942 indicates a two-step process for NAGK-P_{II} complex formation. *J. Mol. Biol.* 399, 410–421. doi: 10.1016/j.jmb.2010.04.018
- Fokina, O., Herrmann, C., and Forchhammer, K. (2011). Signal-transduction protein P_{II} from *Synechococcus elongatus* PCC 7942 senses low adenylate energy charge *in vitro*. *Biochem. J.* 440, 147–156. doi: 10.1042/BJ20110536
- Forcada-Nadal, A., Forchhammer, K., and Rubio, V. (2014). SPR analysis of promoter binding of *Synechocystis* PCC6803 transcription factors NtcA and CRP suggests cross-talk and sheds light on regulation by effector molecules. *FEBS Lett.* 588, 2270–2276. doi: 10.1016/j.febslet.2014.05.010
- Forcada-Nadal, A., Palomino-Schätzlein, M., Neira, J. L., Pineda-Lucena, A., and Rubio, V. (2017). The PipX protein, when not bound to its targets, has its signaling C-terminal helix in a flexed conformation. *Biochemistry* 56, 3211–3224. doi: 10.1021/acs.biochem
- Forchhammer, K. (2008). P_{II} signal transducers: novel functional and structural insights. *Trends Microbiol.* 16, 65–72. doi: 10.1016/j.tim.2007.11.004
- Forchhammer, K., and Lüddecke, J. (2016). Sensory properties of the P_{II} signalling protein family. *FEBS J.* 283, 425–437. doi: 10.1111/febs.13584
- Forchhammer, K., and Tandeau de Marsac, N. (1994). The P_{II} protein in the cyanobacterium *Synechococcus* sp. strain PCC 7942 is modified by serine phosphorylation and signals the cellular N-status. *J. Bacteriol.* 176, 84–91.
- Fujimori, T., Higuchi, M., Sato, H., Aiba, H., Muramatsu, M., Hihara, Y., et al. (2005). The mutant of *slr1961*, which encodes a putative transcriptional regulator, has a defect in regulation of photosystem stoichiometry in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol.* 139, 408–416. doi: 10.1104/pp.105.064782
- Gerhardt, E. C., Rodrigues, T. E., Müller-Santos, M., Pedrosa, F. O., Souza, E. M., Forchhammer, K., et al. (2015). The bacterial signal transduction protein GlnB regulates the committed step in fatty acid biosynthesis by acting as a dissociable regulatory subunit of acetyl-CoA carboxylase. *Mol. Microbiol.* 95, 1025–1035. doi: 10.1111/mmi.12912
- Gil-Ortiz, F., Ramón-Maiques, S., Fita, I., and Rubio, V. (2003). The course of phosphorus in the reaction of N-acetyl-L-glutamate kinase, determined from the structures of crystalline complexes, including a complex with an AlF₄⁻ transition state mimic. *J. Mol. Biol.* 331, 231–244. doi: 10.1016/S0022-2836(03)00716-2
- Gruswitz, F., O'Connell, J., and Stroud, R. M. (2007). Inhibitory complex of the transmembrane ammonia channel, AmtB, and the cytosolic regulatory protein, GlnK, at 1.96 Å. *Proc. Natl. Acad. Sci. U.S.A.* 104, 42–47. doi: 10.1073/pnas.0609796104
- Guerreiro, A. C. L., Benevento, M., Lehmann, R., van Breukelen, B., Post, H., Giansanti, P., et al. (2014). Daily rhythms in the cyanobacterium *Synechococcus elongatus* probed by high-resolution mass spectrometry-based proteomics reveals a small defined set of cyclic proteins. *Mol. Cell. Proteomics* 13, 2042–2055. doi: 10.1074/mcp.M113.035840
- Hauf, W., Schmid, K., Gerhardt, E. C., Huerger, L. F., and Forchhammer, K. (2016). Interaction of the nitrogen regulatory protein GlnB (P_{II}) with biotin carboxyl carrier protein (BCCP) controls Acetyl-CoA levels in the cyanobacterium *Synechocystis* sp. PCC 6803. *Front. Microbiol.* 7:1700. doi: 10.3389/fmicb.2016.01700
- Heinrich, A., Maheswaran, M., Ruppert, U., and Forchhammer, K. (2004). The *Synechococcus elongatus* P_{II} signal transduction protein controls arginine synthesis by complex formation with N-acetyl-L-glutamate kinase. *Mol. Microbiol.* 52, 1303–1314. doi: 10.1111/j.1365-2958.2004.04058.x
- Herrero, A., Muro-Pastor, A. M., and Flores, E. (2001). Nitrogen control in cyanobacteria. *J. Bacteriol.* 183, 411–425. doi: 10.1128/JB.183.2.411-425.2001
- Hoare, D. S., and Hoare, S. L. (1966). Feedback regulation of arginine biosynthesis in blue-green algae and photosynthetic bacteria. *J. Bacteriol.* 92, 375–379.

- Hoskisson, P. A., and Rigali, S. (2009). Chapter 1: variation in form and function the helix-turn-helix regulators of the GntR superfamily. *Adv. Appl. Microbiol.* 69, 1–22. doi: 10.1016/S0065-2164(09)69001-8
- Irmeler, A., Sanner, S., Dierks, H., and Forchhammer, K. (1997). Dephosphorylation of the phosphoprotein P_{II} in *Synechococcus* PCC 7942: identification of an ATP and 2-oxoglutarate-regulated phosphatase activity. *Mol. Microbiol.* 26, 81–90.
- Ito, T., Iimori, J., Takayama, S., Moriyama, A., Yamauchi, A., Hemmi, H., et al. (2013). Conserved pyridoxal protein that regulates Ile and Val metabolism. *J. Bacteriol.* 195, 5439–5449. doi: 10.1128/JB.00593-13
- Jiang, F., Wisén, S., Widersten, M., Bergman, B., and Mannervik, B. (2000). Examination of the transcription factor NtcA-binding motif by *in vitro* selection of DNA sequences from a random library. *J. Mol. Biol.* 301, 783–793. doi: 10.1006/jmbi.2000.4000
- Jiang, P., Mayo, A. E., and Ninfa, A. J. (2007). *Escherichia coli* glutamine synthetase adenyltransferase (ATase, EC 2.7.7.49): kinetic characterization of regulation by P_{II}, P_{II}-UMP, glutamine, and alpha-ketoglutarate. *Biochemistry* 46, 4133–4146. doi: 10.1021/bi0620510
- Kamberov, E. S., Atkinson, M. R., and Ninfa, A. J. (1995). The *Escherichia coli* P_{II} signal transduction protein is activated upon binding 2-ketoglutarate and ATP. *J. Biol. Chem.* 270, 17797–17807.
- Kinch, L. N., and Grishin, N. V. (2002). Expanding the nitrogen regulatory protein superfamily: homology detection at below random sequence identity. *Proteins* 48, 75–84. doi: 10.1002/prot.10110
- Körner, H., Sofia, H. J., and Zumft, W. G. (2003). Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol. Rev.* 27, 559–592. doi: 10.1016/S0168-6445(03)00066-4
- Kujirai, J., Nanba, S., Kadowaki, T., Oka, Y., Nishiyama, Y., Hayashi, Y., et al. (2018). Interaction of the GntR-family transcription factor Sll1961 with thioredoxin in the cyanobacterium *Synechocystis* sp. PCC 6803. *Sci. Rep.* 8:6666. doi: 10.1038/s41598-018-25077-5
- Labella, J. I., Cantos, R., Espinosa, J., Forcada-Nadal, A., Rubio, V., and Contreras, A. (2017). PipY, a member of the conserved COG0325 family of PLP-binding proteins, expands the cyanobacterial nitrogen regulatory network. *Front. Microbiol.* 8:1244. doi: 10.3389/fmicb.2017.01244
- Labella, J. I., Obrebska, A., Espinosa, J., Salinas, P., Forcada-Nadal, A., Tremiño, L., et al. (2016). Expanding the cyanobacterial nitrogen regulatory network: the GntR-like regulator PlmA interacts with the P_{II}-PipX complex. *Front. Microbiol.* 7:1677. doi: 10.3389/fmicb.2016.01677
- Laichoubi, K. B., Espinosa, J., Castells, M. A., and Contreras, A. (2012). Mutational analysis of the cyanobacterial nitrogen regulator PipX. *PLoS ONE* 7:e35845. doi: 10.1371/journal.pone.0035845
- Lambrecht, S. J., Wahlig, J. M. L., and Steglich, C. (2018). The GntR family transcriptional regulator PMM1637 regulates the highly conserved cyanobacterial sRNA Yfr2 in marine picocyanobacteria. *DNA Res.* 25, 489–497. doi: 10.1093/dnares/dsy019
- Lapina, T., Selim, K. A., Forchhammer, K., and Ermilova, E. (2018). The P_{II} signaling protein from red algae represents an evolutionary link between cyanobacterial and Chloroplastida P_{II} proteins. *Sci. Rep.* 8:790. doi: 10.1038/s41598-017-19046-7
- Lee, M. H., Scherer, M., Rigali, S., James, W., and Golden, J. W. (2003). PlmA, a new member of the GntR family, has plasmid maintenance functions in *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* 185, 4315–4325. doi: 10.1128/JB.185.15.4315-4325.2003
- Llácer, J. L., Contreras, A., Forchhammer, K., Marco-Marin, C., Gil-Ortiz, F., Maldonado, R., et al. (2007). The crystal structure of the complex of P_{II} and acetylglutamate kinase reveals how P_{II} controls the storage of nitrogen as arginine. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17644–17649. doi: 10.1073/pnas.0705987104
- Llácer, J. L., Espinosa, J., Castells, M. A., Contreras, A., Forchhammer, K., and Rubio, V. (2010). Structural basis for the regulation of NtcA-dependent transcription by proteins PipX and P_{II}. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15397–15402. doi: 10.1073/pnas.1007015107
- Llácer, J. L., Fita, I., and Rubio, V. (2008). Arginine and nitrogen storage. *Curr. Opin. Struct. Biol.* 18, 673–681. doi: 10.1016/j.sbi.2008.11.002
- Lohmeier-Vogel, E. M., Loukanina, N., Ferrar, T. S., Moorhead, G. B., and Thorpe, T. A. (2005). N-acetyl glutamate kinase from *Daucus carota* suspension cultures: embryonic expression profile, purification and characterization. *Plant. Physiol. Biochem.* 43, 854–861. doi: 10.1016/j.plaphy.2005.07.009
- Lu, R., and Wang, G. G. (2013). Tudor: a versatile family of histone methylation 'readers'. *Trends Biochem. Sci.* 38, 546–555. doi: 10.1016/j.tibs.2013.08.002
- Lüddecke, J., and Forchhammer, K. (2015). Energy sensing versus 2-oxoglutarate dependent ATPase switch in the control of *Synechococcus* P_{II} interaction with its targets NAGK and PipX. *PLoS ONE* 10:e0137114. doi: 10.1371/journal.pone.0137114
- Luque, I., Flores, E., and Herrero, A. (1994). Molecular mechanism for the operation of nitrogen control in cyanobacteria. *EMBO J.* 13, 2862–2869.
- Maheswaran, M., Urbanke, C., and Forchhammer, K. (2004). Complex formation and catalytic activation by the P_{II} signaling protein of N-acetyl-L-glutamate kinase from *Synechococcus elongatus* strain PCC 7942. *J. Biol. Chem.* 279, 55202–55210. doi: 10.1074/jbc.M410971200
- Maier, S., Schleberger, P., Lü, W., Wacker, T., Pflüger, T., Litz, C., et al. (2011). Mechanism of disruption of the Amt-GlnK complex by P_{II}-mediated sensing of 2-oxoglutarate. *PLoS ONE* 6:e26327. doi: 10.1371/journal.pone.0026327
- McKay, D. B., and Steitz, T. A. (1981). Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature* 290, 744–749.
- Merrick, M. (2015). Post-translational modification of P_{II} signal transduction proteins. *Front. Microbiol.* 5:763. doi: 10.3389/fmicb.2014.00763
- Mizuno, Y., Moorhead, G. B. G., and Ng, K. K. S. (2007). Structural basis for the regulation of N-acetylglutamate kinase by P_{II} in *Arabidopsis thaliana*. *J. Biol. Chem.* 282, 35733–35740. doi: 10.1074/jbc.M707127200
- Muro-Pastor, M. I., Reyes, J. C., and Florencio, F. J. (2001). Cyanobacteria perceive nitrogen status by sensing intracellular 2-oxoglutarate levels. *J. Biol. Chem.* 276, 38320–38328. doi: 10.1074/jbc.M105297200
- Omagari, K., Yoshimura, H., Suzuki, T., Takano, M., Ohmori, M., and Sarai, A. (2008). DeltaG-based prediction and experimental confirmation of SYCRP1-binding sites on the *Synechocystis* genome. *FEBS J.* 275, 4786–4795. doi: 10.1111/j.1742-4658.2008.06618.x
- Oppermann-Sanio, F., and Steinbüchel, A. (2002). Occurrence, functions and biosynthesis of polyamides in microorganisms and biotechnological production. *Naturwissenschaften* 89, 11–22. doi: 10.1007/s00114-001-0280-0
- Osana, T., Sato, S., Tabata, S., and Tanaka, K. (2005). Identification of PamA as a P_{II}-binding membrane protein important in nitrogen-related and sugar-catabolic gene expression in *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* 280, 34684–34690. doi: 10.1074/jbc.M507489200
- Palanca, C., Pedro-Roig, L., Llácer, J. L., Camacho, M., Bonete, M. J., and Rubio, V. (2014). The structure of a P_{II} signaling protein from a halophilic archaeon reveals novel traits and high-salt adaptations. *FEBS J.* 281, 3299–3314. doi: 10.1111/febs.12881
- Palanca, C., and Rubio, V. (2017). Effects of T-loop modification on the P_{II}-signalling protein: structure of uridylylated *Escherichia coli* GlnB bound to ATP. *Environ. Microbiol. Rep.* 9, 290–299. doi: 10.1111/1758-2229.12533
- Pedro-Roig, L., Camacho, M., and Bonete, M. J. (2013). *Haloferax mediterranei* GlnK proteins are post-translationally modified by uridylylation. *Proteomics* 13, 1371–1374. doi: 10.1002/pmic.201200465
- Plecko, B., Zweier, M., Begemann, A., Mathis, D., Schmitt, B., Striano, P., et al. (2017). Confirmation of mutations in PROSC as a novel cause of vitamin B6-dependent epilepsy. *J. Med. Genet.* 54, 809–814. doi: 10.1136/jmedgenet-2017-104521
- Prunetti, L., El Yacoubi, B., Schiavon, C. R., Kirkpatrick, E., Huang, L., Bailly, M., et al. (2016). Evidence that COG0325 proteins are involved in PLP homeostasis. *Microbiology* 162, 694–706. doi: 10.1099/mic.0.000255
- Radchenko, M. V., Thornton, J., and Merrick, M. (2013). P_{II} signal transduction proteins are ATPases whose activity is regulated by 2-oxoglutarate. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12948–12953. doi: 10.1073/pnas.1304386110
- Rajendran, C., Gerhardt, E. C. M., Bjelic, S., Gasperina, A., Scarduelli, M., Pedrosa, F. O., et al. (2011). Crystal structure of the GlnZ-DraG complex reveals a different form of P_{II}-target interaction. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18972–18976. doi: 10.1073/pnas.1108038108
- Ramón-Maiques, S., Fernandez-Murga, M. L., Gil-Ortiz, F., Vagin, A., Fita, I., and Rubio, V. (2006). Structural bases of feed-back control of arginine biosynthesis,

- revealed by the structures of two hexameric N-acetylglutamate kinases, from *Thermotoga maritima* and *Pseudomonas aeruginosa*. *J. Mol. Biol.* 356, 695–713. doi: 10.1016/j.jmb.2005.11.079
- Ramón-Maiques, S., Marina, A., Gil-Ortiz, F., Fita, I., and Rubio, V. (2002). Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the amino acid kinase enzyme family, during catalysis. *Structure* 10, 329–342. doi: 10.1016/S0969-2126(02)00721-9
- Rodrigues, T. E., Gerhardt, E. C., Oliveira, M. A., Chubatsu, L. S., Pedrosa, F. O., Souza, E. M., et al. (2014). Search for novel targets of the P_{II} signal transduction protein in bacteria identifies the BCCP component of acetyl-CoA carboxylase as a P_{II} binding partner. *Mol. Microbiol.* 91,751–761. doi: 10.1111/mmi.12493
- Rubio, V. (1986). Enzymatic HCO₃⁻ fixation: a common mechanism for all enzymes involved? *Biosci. Rep.* 6, 335–347.
- Sant'Anna, F. H., Trentini, D. B., de Souto Weber, S., Cecagno, R., da Silva, S. C., and Schrank, I. S. (2009). The P_{II} superfamily revised: a novel group and evolutionary insights. *J. Mol. Evol.* 68, 322–336. doi: 10.1007/s00239-009-9209-6
- Schumacher, M. A., Chinnam, N. B., Cuthbert, B., Tonthat, N. K., and Whitfill, T. (2015). Structures of regulatory machinery reveal novel molecular mechanisms controlling *B. subtilis* nitrogen homeostasis. *Genes Dev.* 29, 451–464. doi: 10.1101/gad.254714.114
- Shapiro, B. M. (1969). The glutamine synthetase deadenylylating enzyme system from *Escherichia coli*. Resolution into two components, specific nucleotide stimulation and cofactor requirements. *Biochemistry* 8, 659–670.
- Shaw, G., Gan, J., Zhou, Y. N., Zhi, H., Subburaman, P., Zhang, R., et al. (2008). Structure of RapA, a Swi2/Snf2 protein that recycles RNA polymerase during transcription. *Structure* 16,1417–1427. doi: 10.1016/j.str.2008.06.012
- Stadtman, E. R. (2001). The story of glutamine synthetase regulation. *J. Biol. Chem.* 276, 44357–44364. doi: 10.1074/jbc.R100055200
- Steiner, T., Kaiser, J. T., Marinkovic, S., Huber, R., and Wahl, M. C. (2002). Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *EMBO J.* 21, 4641–4653. doi: 10.1093/emboj/cdf455
- Tanigawa, R., Shirokane, M., Maeda, S., Omata, T., Tanaka, K., and Takahashi, H. (2002). Transcriptional activation of NtcA-dependent promoters of *Synechococcus* sp. PCC 7942 by 2-oxoglutarate *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4251–4255. doi: 10.1073/pnas.072587199
- Tremiño, L., Forcada-Nadal, A., Contreras, A., and Rubio, V. (2017). Studies on cyanobacterial protein PipY shed light on structure, potential functions, and vitamin B₆-dependent epilepsy. *FEBS Lett.* 591, 3431–3442. doi: 10.1002/1873-3468.12841
- Tremiño, L., Forcada-Nadal, A., and Rubio, V. (2018). Insight into vitamin B₆-dependent epilepsy due to PLPBP (previously PROSC) missense mutations. *Hum. Mutat.* 39, 1002–1013. doi: 10.1002/humu.23540
- Truan, D., Huergo, L. F., Chubatsu, L. S., Merrick, M., Li, X. D., and Winkler, F. K. (2010). A new P_{II} protein structure identifies the 2-oxoglutarate binding site. *J. Mol. Biol.* 400, 531–539. doi: 10.1016/j.jmb.2010.05.036
- VanEtten, C. H., Wolff, I. A., Jones, Q., and Miller, R. W. (1963). Amino acid composition of seeds from 200 angiospermous plant species. *J. Agric. Food Chem.* 11, 399–410. doi: 10.1021/jf60129a016
- Vázquez-Bermúdez, M. F., Herrero, A., and Flores, E. (2002). 2-Oxoglutarate increases the binding affinity of the NtcA (nitrogen control) transcription factor for the *Synechococcus glnA* promoter. *FEBS Lett.* 512,71–74. doi: 10.1016/S0014-5793(02)02219-6
- Vega-Palas, M. A., Flores, E., and Herrero, A. (1992). NtcA, a global nitrogen regulator from the cyanobacterium *Synechococcus* that belongs to the Crp family of bacterial regulators. *Mol. Microbiol.* 6, 1853–1859.
- Watzer, B., and Forchhammer, K. (2018). Cyanophycin synthesis optimizes nitrogen utilization in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. *Appl. Environ. Microbiol.* 84:e01298-18. doi: 10.1128/AEM.01298-18
- Weber, I. T., and Steitz, T. A. (1987). Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution. *J. Mol. Biol.* 198, 311–326.
- Xu, Y., Carr, P. D., Clancy, P., Garcia-Dominguez, M., Forchhammer, K., Florencio, F., et al. (2003). The structures of the P_{II} proteins from the cyanobacteria *Synechococcus* sp. PCC 7942 and *Synechocystis* sp. PCC 6803. *Acta Crystallogr. D Biol. Crystallogr.* 59, 2183–2190. doi: 10.1107/S0907444903019589
- Xu, Y., Carr, P. D., Vasudevan, S. G., and Ollis, D. L. (2010). Structure of the adenylation domain of *E. coli* glutamine synthetase adenylyl transferase: evidence for gene duplication and evolution of a new active site. *J. Mol. Biol.* 396, 773–784. doi: 10.1016/j.jmb.2009.12.011
- Xu, Y., Zhang, R., Joachimiak, A., Carr, P. D., Huber, T., Vasudevan, S. G., et al. (2004). Structure of the N-terminal domain of *Escherichia coli* glutamine synthetase adenylyltransferase. *Structure* 12, 861–869. doi: 10.1016/j.str.2004.02.029
- Zeth, K., Fokina, O., and Forchhammer, K. (2014). Structural basis and target-specific modulation of ADP sensing by the *Synechococcus elongatus* P_{II} signaling protein. *J. Biol. Chem.* 289, 8960–8972. doi: 10.1074/jbc.M113.536557
- Zhao, M. X., Jiang, Y. L., He, Y. X., Chen, Y. F., and Teng, Y. B., Chen, Y., et al. (2010a). Structural basis for the allosteric control of the global transcription factor NtcA by the nitrogen starvation signal 2-oxoglutarate. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12487–12492. doi: 10.1073/pnas.1001556107
- Zhao, M. X., Jiang, Y. L., Xu, B. Y., Chen, Y., Zhang, C. C., and Zhou, C. Z. (2010b). Crystal structure of the cyanobacterial signal transduction protein P_{II} in complex with PipX. *J. Mol. Biol.* 402, 552–559. doi: 10.1016/j.jmb.2010.08.006

Conflict of Interest Statement: 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.

The reviewer JAH declared a shared affiliation, with no collaboration, with several of the authors, AF-N, JL, CM-M, and VR to the handling editor at time of review.

Copyright © 2018 Forcada-Nadal, Llacer, Contreras, Marco-Marín and Rubio. 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.



RNA Binding Protein Regulation and Cross-Talk in the Control of AU-rich mRNA Fate

Sofía M. García-Mauriño, Francisco Rivero-Rodríguez, Alejandro Velázquez-Cruz, Marian Hernández-Vellisca, Antonio Díaz-Quintana, Miguel A. De la Rosa and Irene Díaz-Moreno*

Instituto de Investigaciones Químicas, Centro de Investigaciones Científicas Isla de la Cartuja, Universidad de Sevilla, Consejo Superior de Investigaciones Científicas, Seville, Spain

OPEN ACCESS

Edited by:

Maria Rosaria Conte,
King's College London,
United Kingdom

Reviewed by:

Graeme L. Conn,
Emory University School of Medicine,
United States

Santiago Martínez-Lumbreras,
King's College London,
United Kingdom

Scott A. Tenenbaum,
University at Albany, SUNY,
United States

Teresa Carlomagno,
Leibniz University of Hanover,
Germany

*Correspondence:

Irene Díaz-Moreno
idiazmoreno@us.es

Specialty section:

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

Received: 26 July 2017

Accepted: 04 October 2017

Published: 23 October 2017

Citation:

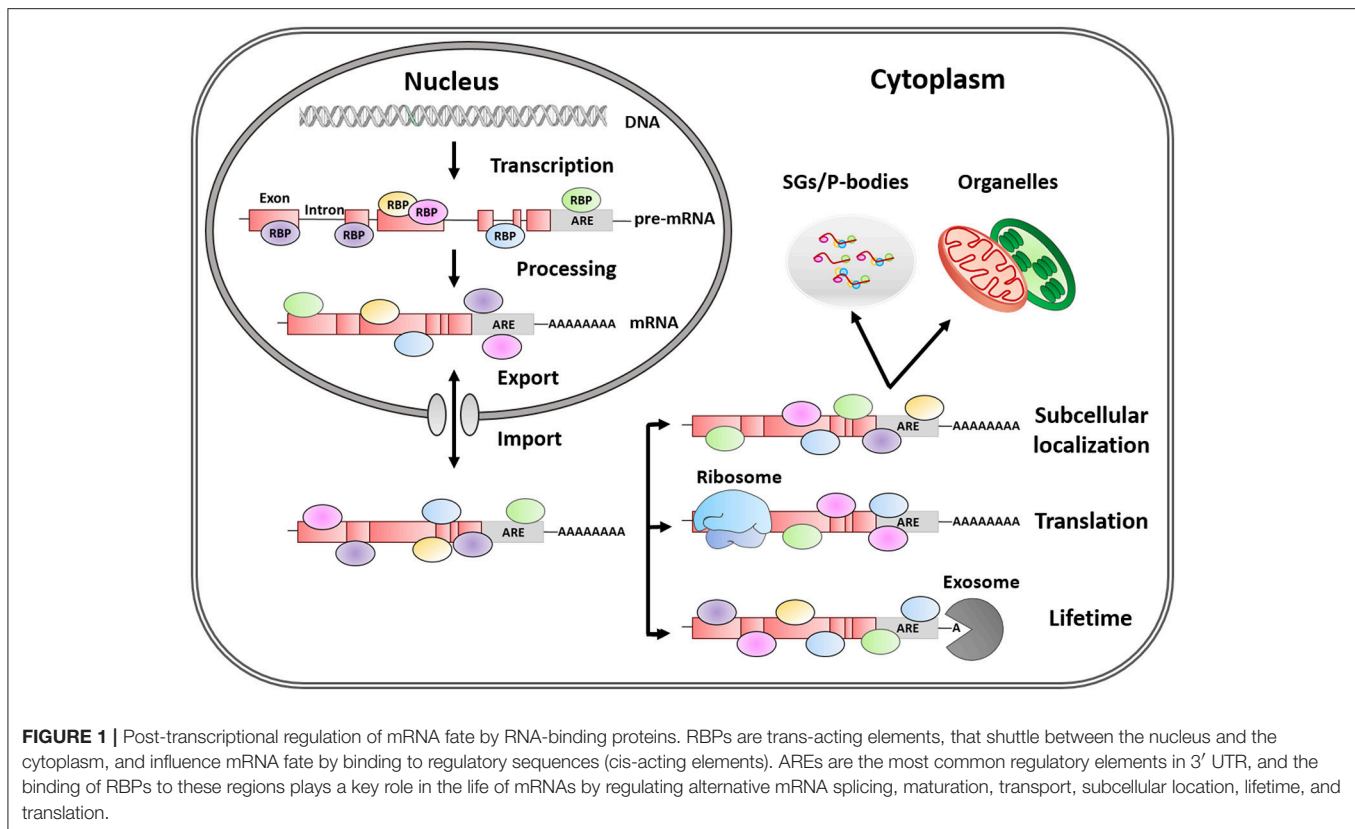
García-Mauriño SM,
Rivero-Rodríguez F, Velázquez-Cruz A,
Hernández-Vellisca M,
Díaz-Quintana A, De la Rosa MA and
Díaz-Moreno I (2017) RNA Binding
Protein Regulation and Cross-Talk in
the Control of AU-rich mRNA Fate.
Front. Mol. Biosci. 4:71.
doi: 10.3389/fmolb.2017.00071

mRNA metabolism is tightly orchestrated by highly-regulated RNA Binding Proteins (RBPs) that determine mRNA fate, thereby influencing multiple cellular functions across biological contexts. Here, we review the interplay between six well-known RBPs (TTP, AUF-1, KSRP, HuR, TIA-1, and TIAR) that recognize AU-rich elements (AREs) at the 3' untranslated regions of mRNAs, namely ARE-RBPs. Examples of the links between their cross-regulations and modulation of their targets are analyzed during mRNA processing, turnover, localization, and translational control. Furthermore, ARE recognition can be self-regulated by several factors that lead to the prevalence of one RBP over another. Consequently, we examine the factors that modulate the dynamics of those protein-RNA transient interactions to better understand the final consequences of the regulation mediated by ARE-RBPs. For instance, factors controlling the RBP isoforms, their conformational state or their post-translational modifications (PTMs) can strongly determine the fate of the protein-RNA complexes. Moreover, mRNA specific sequence and secondary structure or subtle environmental changes are also key determinants to take into account. To sum up, the whole understanding of such a fine tuned regulation is a challenge for future research and requires the integration of all the available structural and functional data by *in vivo*, *in vitro* and *in silico* approaches.

Keywords: mRNA fate, post-transcriptional regulation, RNA binding proteins, stability, translation

POST-TRANSCRIPTIONAL REGULATION OF GENE EXPRESSION BY ARE-RBPs

In eukaryotes, gene expression levels and protein abundance are often correlated but are subjected to a strict regulation. The control of mRNA metabolism allows cells to rapidly adapt to changing environmental conditions. Regulatory processes occurring after mRNA transcription—namely post-transcriptional control—strongly influence mRNA fate and, consequently, final protein levels (Vogel and Marcotte, 2012). Once mRNA transcription occurs in the nucleus, RNA Binding Proteins (RBPs) recognize the primary transcript or pre-mRNA to regulate its alternative splicing, polyadenylation, and capping (**Figure 1**). The generated mature mRNA is then transported to the cytoplasm by various other RBPs. Once in the cytoplasm, RBPs govern the stability, distribution to different cellular compartments and the translation of target mRNAs into their corresponding protein products (Matoulikova et al., 2012).



Within RBPs, ARE-RBPs function as trans-acting factors recognizing cis-acting elements in the 3'-Untranslated Regions (UTR) of eukaryotic mRNA enriched in adenylate and uridylate (AU-rich elements or AREs). AREs are present in 5–8% of human genes with diverse functions such as cell growth and differentiation, signal transduction, apoptosis, nutrient transport, and metabolism. This list is dominated by genes involved in transient processes, which therefore require strict expression control (Barreau et al., 2005). For instance, the length and specific pattern of AREs may contribute to mRNA lifetime (Khabar, 2005). However, the final mRNA fate will be determined by the variable and dynamic ARE-RBPs/mRNA interactions or by RBP competition for the same transcript. Besides, ARE-RBPs bind to AREs via a variety of domains including the so-called RNA-Recognition Motif (RRM), the CCCH tandem zinc finger and the K-Homology domain (KH) (Stoecklin and Anderson,

2006; Clery et al., 2008; Valverde et al., 2008; Daubner et al., 2013). A single protein can contain several of these motifs leading to simultaneous interactions with either multiple targets or multiple sites within a particular target (Shen and Malter, 2015). Additionally, most ARE-RBPs shuttle between nucleus and cytoplasm; and their functions are linked to their specific subcellular distribution (Gama-Carvalho and Carmo-Fonseca, 2001).

In this mini-review we focus on the post-transcriptional regulation exerted by six of the best studied ARE-RBPs whose cross-talk has biological relevance and has been widely reported in the literature. Moreover, we examine the multiple intracellular signals and factors controlling the interactions between these proteins. AU-binding Factor 1 (AUF1), also known as Heterogeneous Nuclear Ribonucleo-Protein D (hnRNP D), is included for being the first identified ARE-RBP (Brewer, 1991). AUF1 is generally considered to promote the decay of target mRNAs, although the stabilization of some other transcripts has been also reported (Xu et al., 2001; Stoecklin and Anderson, 2006). Since AUF1 discovery, 20 additional ARE-RBPs have been identified. That list includes those that primarily promote mRNA degradation, such as Tristetraprolin (TTP) and KH domain-splicing regulatory protein (KSRP) (Gherzi et al., 2004; Sanduja et al., 2011); those stabilizing mRNA, such as Human antigen R (HuR) (Brennan and Steitz, 2001); and translational control proteins, such as T-cell intracellular antigen 1 (TIA-1) and TIA-1-related protein (TIAR) (Kawai et al., 2006; Mazan-Mamczarz et al., 2006).

Abbreviations: AREs, AU-Rich Elements; ARE-RBPs, RNA Binding Proteins that recognize AU-Rich Elements; AUF-1, AU-binding Factor 1; c-fos, Finkel-Biskis-Jinkins murine osteosarcoma viral oncogene homolog; c-myc, Avian myelocytomatosis virus oncogene cellular homolog; COX-2, Cyclooxygenase-2; DDR, DNA Damage Response; hnRNP D, Heterogeneous Nuclear Ribonucleo-Protein D; HuR, Human antigen R; KH, hnRNP K-Homology; KSRP, KH type Splicing Regulatory Protein; PAR, Poly (ADP-Ribose); P-bodies, Processing bodies; PTMs, Post-Translational Modifications; PRD, Prion-Related Domain; RBPs, RNA Binding Proteins; RNP, Ribonucleoprotein Particles; RRM, RNA Recognition Motif; SGs, Stress Granules; TIA-1, T-cell Intracellular Antigen 1; TIAR, TIA-1 Related protein; TNF α , Tumor Necrosis Factor α ; TTP, Tristetraprolin; UTR, Untranslated Region; VEGF, Vascular Endothelial Growth Factor.

INTERPLAY BETWEEN ARE-RBPs IN THE POST-TRANSCRIPTIONAL REGULATION OF mRNAs

It is well-known that the substrates of post-transcriptional control are RNA ribonucleoprotein particles or RNPs containing mRNA molecules covered with RBPs, rather than naked mRNA (Szostak and Gebauer, 2013) (**Figure 1**). However, our understanding of how ARE-RBPs interact with each other at different regulatory levels is rather limited. Noticeably, some RBPs regulate the mRNA that encodes their own gene products, as well as those of other RBP counterparts, establishing self-regulatory loops controlling mRNA metabolism (Pullmann et al., 2007).

A good example of cross-talk between RBPs is the one involving HuR, KSRP and TTP proteins. These three proteins compete with each other for binding to common recognition sequences in the AREs that they regulate. Hence, TTP and KSRP negatively control the stability of several mRNAs—such as *c-fos*, *TNF α* and *COX-2*—whereas HuR generally acts in an opposite way, stabilizing them (Chen et al., 2001, 2002; Dean et al., 2001; Sawaoka et al., 2003; Katsanou et al., 2005; Winzen et al., 2007) with some exceptions (Katsanou et al., 2005; Kim et al., 2009) (**Table 1**). Moreover, TTP acts as a negative regulator of its own mRNA (Tchen et al., 2004; Lin et al., 2007) as well as HuR mRNA, its direct antagonist in mRNA regulation (Al-Ahmadi et al., 2009). On the other hand, HuR acts as a positive translational regulator of both KSRP and HuR mRNAs (Pullmann et al., 2007; Yi et al., 2010); while both proteins regulate the stability of their own mRNAs (Winzen et al., 2007; Al-Ahmadi et al., 2009) (**Table 1**, dashed square). Consequently, the redundant feedback involving KSRP, TTP, and HuR may provide a bi-stable signal transduction circuit in which either all or none of their target mRNAs are stabilized and/or translated. More intriguingly is the role of AUF1 in this regulatory loop as it presents four isoforms generated by alternative splicing of a single mRNA transcript (Wagner et al., 1998) with different RNA-binding affinities and specificities for its target mRNAs—such as *c-fos*, *c-myc*, *TNF α* , *VEGF*, and *COX-2* (Brewer, 1991; Loflin et al., 1999; Lasa et al., 2000; Xu et al., 2001; Fellows et al., 2012).

In addition to recognizing AU-rich sequences at the 3' UTR of target mRNAs, some ARE-RBPs are able to activate splice 5' sites followed by U-rich sequences. This is the case of TIA-1 and TIAR, that upregulate the translation of their own coding mRNAs (Le Guiner et al., 2001). Conversely, although consistent with their functional redundancy, their translation levels are negatively cross-regulated by each other (Le Guiner et al., 2001; Izquierdo and Valcárcel, 2007b; Pullmann et al., 2007) (**Table 1**, dotted square). Interestingly, TIA-1 and TIAR share common functions, acting as negative translational regulators of diverse mRNAs—such as *c-myc*, *TNF α* , *VEGF*, and *COX-2*—and are able to compensate for each other (Gueydan et al., 1999; Piecyk et al., 2000; Zhang et al., 2002; Cok et al., 2003; Dixon et al., 2003; Lu et al., 2009; Hamdollah Zadeh et al., 2015). In addition, it has been shown that HuR positively controls TIA-1 expression by enhancing its mRNA stability (Pullmann et al., 2007). By contrast,

TIA-1 knockdown causes a marked increase in HuR levels, indicating that TIA-1 may contribute to lowering HuR levels in the cell (Kawai et al., 2006) (**Table 1**, black square). This is of great importance because both HuR and TIA-1 bind to cytochrome *c* (*Cc*) mRNA, respectively promoting or inhibiting its translation without affecting its mRNA stability. The struggle between HuR (antiapoptotic factor) and TIA-1 (proapoptotic factor) for the control of *Cc* mRNA translation underlies possible mechanisms to regulate both cellular respiration and programmed cell death. A direct binding between HuR and TIAR mRNA has also been reported (Pullmann et al., 2007) but, unexpectedly, TIAR does not seem to complex with *Cc* mRNA, despite the extensively shared homology between TIAR and TIA-1 (Kawai et al., 2006).

FACTORS THAT MODULATE ARE-RBP/mRNA INTERACTIONS

Several examples of cross-talk between ARE-RBPs highlight that there must be an intricate network of regulatory events that lead to the prevalence of one RBP over the others when recognizing the same mRNA target. Thus, the regulatory activity of RBPs on gene expression is dynamic and adapts to cell conditions continuously. In this section, we briefly describe those factors for which there is evidence of their influence on the interaction between RBPs and their mRNA targets (Supplemental Figure 1).

RBP Isoforms

Alternative splicing is a highly regulated process that allows the synthesis of multiple different transcripts from the same gene, and is therefore an important source of protein diversity and complexity. The slight differences in amino acid sequence between isoforms can be determining for their function (Gallego-Páez et al., 2017). For example, TIA-1 and TIAR present two isoforms—*a* and *b*—in humans. Isoform *a* in TIA-1 and TIAR possesses 11 or 17 extra amino acids, respectively, that are critical for distinct functional properties. For instance, only TIAR isoform *a*—but not TIAR *b* and none of TIA-1 isoforms—has a translational silencing activity on the proteolytic enzyme Human Matrix Metalloproteinase-13 (HMMP13) in HEK293 cells. High levels of expression of HMMP13 have been documented in certain subset of cancers. Therefore, its downregulation by TIAR *a* may act as a tumor suppression mechanism (Yu et al., 2003).

As previously mentioned, AUF1 isoforms come from the alternative splicing of the same pre-mRNA. They differ as a function of the presence or absence of two independent domains encoded by exons 2 and 7. While p37^{AUF1} lacks both domains, p42^{AUF1} and p45^{AUF1} include a 49-amino acid domain encoded by exon 7 and p40^{AUF1} and p45^{AUF1} both contain a 19-amino acid domain encoded by exon 2. Inclusion of the exon 2-encoded sequence reduces the affinity of the first and second binding events of AUF1 dimers toward their mRNA substrates, but incorporation of the exon 7-encoded sequence increases the affinity of the second binding event. The isoform-specific differences provide unique biochemical characteristics that explain the diversity of AUF1 functions and complex regulation (Zucconi et al., 2010).

TABLE 1 | Matrix representation of the interaction of selected RBPs (vertical axis) with the mRNA of those RBPs (Upper table) and several ARE-containing mRNA targets (Lower table, horizontal axis).

| mRNA | | | | | | | | | | | | | |
|-------|---------------------------|--|-----|---|------|-----------------------|-----|--|-------|--|------|----------------------------|--|
| AUF1 | | | TTP | | KSRP | | HuR | | TIA-1 | | TIAR | | |
| AUF1 | Gratacós and Brewer, 2010 | | | n.d. | | Pullmann et al., 2007 | | Pullmann et al., 2007 | | Pullmann et al., 2007 | | Pullmann et al., 2007 | |
| TTP | n.d. | | | Tchen et al., 2004; Lin et al., 2007 | | n.d. | | Al-Ahmadi et al., 2009 | | n.d. | | n.d. | |
| KSRP | n.d. | | | n.d. | | Winzen et al., 2007 | | n.d. | | n.d. | | n.d. | |
| HuR | Pullmann et al., 2007 | | | n.d. | | Pullmann et al., 2007 | | Al-Ahmadi et al., 2009; Yi et al., 2010 | | Pullmann et al., 2007; Kawai et al., 2006 | | Pullmann et al., 2007 | |
| TIA-1 | n.d. | | | n.d. | | n.d. | | Kawai et al., 2006 | | Le Guiner et al., 2001 (S) | | Pullmann et al., 2007 | |
| TIAR | Pullmann et al., 2007 | | | n.d. | | Pullmann et al., 2007 | | Pullmann et al., 2007 | | Pullmann et al., 2007; Izquierdo and Valcárcel, 2007b (S) | | Le Guiner et al., 2001 (S) | |

| ARE-containing mRNA | | | | | | | | | | | | |
|---------------------|---|---------------------|---------------------------------------|------------------------|--------------------|---|---|---|------------------------------|---|-----------------------|--|
| c-fos | | | c-myc | | Cc | | TNFα | | VEGF | | COX-2 | |
| AUF1 | Xu et al., 2001; Chen et al., 2002 | Loflin et al., 1999 | Xu et al., 2001; Liao et al., 2007 | Brewer, 1991 | n.d. | | Xu et al., 2001 | | Fellows et al., 2012 | | Lasa et al., 2000* | |
| TTP | Chen et al., 2001 | | n.d. | | n.d. | | Lai et al., 1999; Chen et al., 2001; Lai and Blackshear, 2001 | | Lee et al., 2010 | | Sawaoka et al., 2003 | |
| KSRP | Chen et al., 2001; Gherzi et al., 2004 | | Yamakoshi et al., 2010 | Trabucchi et al., 2009 | n.d. | | Chen et al., 2001; Gherzi et al., 2004; Winzen et al., 2007 | | n.d. | | Winzen et al., 2007 | |
| HuR | Fan and Steitz, 1998; Peng et al., 1998; Chen et al., 2002 | | Kim et al., 2009 | | Kawai et al., 2006 | Dean et al., 2001; Katsanou et al., 2005 | Katsanou et al., 2005 | Levy et al., 1998; Goldberg-Cohen et al., 2002 | | Sengupta et al., 2003; Katsanou et al., 2005 | Katsanou et al., 2005 | |
| TIA-1 | n.d. | | Wang et al., 2010 | | Kawai et al., 2006 | Piecyk et al., 2000 | | Hamdollah Zadeh et al., 2015 | Hamdollah Zadeh et al., 2015 | Dixon et al., 2003 | | |
| TIAR | n.d. | | Liao et al., 2007 | | n.d. | | Gueydan et al., 1999; Zhang et al., 2002* | | Lu et al., 2009* | | Cok et al., 2003 | |

Blue colors show positive regulation, whereas negative regulations are colored in red. Gray color indicates interactions that have been described but the effects were not examined. n.d., non-described; S, Splicing; *, Postulated regulations. The cross-talk between TTP, KSRP and HuR is highlighted by a dashed square; between HuR and TIA-1 by a black square; and between TIA-1 and TIAR by a dotted square.

RBP Post-Translational Modifications

Post-Translational Modifications (PTMs), such as phosphorylation, isomerization, methylation, NEDDylation, acetylation, and ubiquitination of RBPs have a major influence on their function and/or their affinity toward their targets, with the consequent impact on mRNA stability, turnover and translation efficiency (Lee, 2012). For instance, the phosphorylation of p40^{AUF1} in residues Ser83 and Ser87 influences the sequential binding of dimers to TNF α mRNA (Wilson et al., 2003). Single phosphorylation of Ser83 inhibits by 40% the initial dimer binding to mRNA substrate, whereas Ser87-single phosphorylation induces a 2-fold increase in the affinity of

the second binding event. In addition, when simultaneous phosphorylation of both residues occurs, the negative effect on the binding affinity of Ser83 prevails over the positive effect of Ser87 (Wilson et al., 2003). Several phosphorylation sites have also been identified in TTP (Cao et al., 2006, 2014). Phosphorylated TTP binds with a lower affinity than the dephosphorylated TTP to target AREs (Carballo et al., 2001; Hitti et al., 2006). Phosphorylation of RBPs can also modify their activity without altering the affinity for mRNA targets. Such is the case of TIA-1 and TIAR, whose splicing control over the *Fas* gene sequence determines the expression of the pro-apoptotic membrane-bound form in detriment of the anti-apoptotic

soluble one (Izquierdo and Valcárcel, 2007a). Moreover, HuR methylation has been proposed to increase the nuclear export of HuR, which could be important for mRNA localization (Li et al., 2002). NEDDylation of HuR increases its stability and lifetime, which, in turn, can affect the total levels of HuR target mRNAs due to its main stabilizing action (Embade et al., 2012; Fernández-Ramos and Martínez-Chantar, 2015).

RBP Conformational Changes

ARE-RBPs can undergo conformational changes upon binding to their targets (Ellis and Jones, 2008). These variations can be detected in the contact surface with mRNAs as well as in distant areas, meaning that ARE-RBPs can adapt both the local and global structure. An example of conformational changes that influence ARE recognition has been reported for KSRP. An inter-domain re-arrangement, that orients the two central KH domains and their RNA-binding surfaces creating a two-domain unit, is crucial for its role in ARE-mediated mRNA decay (Supplemental Figure 2) (Díaz-Moreno et al., 2010). Additionally, some of the PTMs mentioned above can also influence the conformation of RBPs. Hence, the phosphorylation of Ser193 within the N-terminal KH motif (KH1) of KSRP leads to the unfolding of this structurally atypical and unstable domain, creating a binding site for 14-3-3 ζ , driving the nuclear localization of KSRP and controlling its mRNA-degradation activity (Díaz-Moreno et al., 2009).

Another important regulation factor is the RBP oligomerization state upon mRNA recognition. HuR RRM1 domain and RRM1-2 di-domain (the main platform of cytoplasmic mRNA binding in HuR) form homodimers in solution (Benoit et al., 2010). This phenomenon is dependent on Cys13, which is able to form disulfide bonds. Such homodimerization may modulate HuR function upon oxidative stress. Moreover, the HuR RRM3 domain has been found to be involved in protein oligomerization and RNA recognition, both functions regulated by the same RRM but using different surfaces at opposite sides of the domain. The conserved Trp261 residue is key for dimerization, as the substitution by glutamic acid alters its dimerization dynamics and stabilizes the monomeric state (Scheiba et al., 2014; Díaz-Quintana et al., 2015).

Cellular Conditions and Stress Response

Eukaryotic cells have evolved sophisticated strategies to overcome stress. One of them is the assembly of Stress Granules (SGs), which allows mRNA translation silencing and protection from degradation. Among RBPs with critical roles in neurodegenerative diseases, TIA-1 proteins are essential in SG formation (Mazan-Mamczarz et al., 2006; Vanderweyde et al., 2012). Hence, under hypoxic conditions, TIA-1 and TIAR block the expression of hypoxia-inducible factor (HIF)-1 α through binding to its ARE-containing mRNA (Gottschald et al., 2010). Inhibition of this transcription factor is enhanced when both RBPs are organized into SGs. In addition, HuR also aggregates into SGs to halt the translation of specific housekeeping mRNAs under stress conditions (Bergalet et al., 2011). The deregulation of SGs results in cytoplasmic accumulation and subsequent

pathologies such as Parkinson and Alzheimer (Vanderweyde et al., 2012).

Variations in pH values can also modulate the binding of TIA-1 to nucleic acids, acting as a pH-dependent molecular switch. The pK_a values of the histidine imidazole groups of TIA-1 RRM2 and RRM3 are substantially higher in complexes with short RNA and DNA oligonucleotides than in the isolated domains. Interestingly, those pK_a values are also controlled by slight environmental pH changes (Cruz-Gallardo et al., 2013, 2015). This fact provides valuable information to understand the pH effect on ARE-RBPs when shuttling among cellular compartments with different pHs (nucleus, cytoplasm, SGs, etc.).

During oxidative stress, AUF1 binding to mRNAs containing 8-oxo-7,8-dihydro-guanine could play a role in the selective elimination of oxidized mRNA by presumably driving their degradation (Ishii et al., 2015). Finally, HuR localization can also be altered upon different stress signals such as UV, actinomycin D or hydrogen peroxide, leading to the cytoplasmic accumulation of the protein. However, after a heat shock treatment, the decrease in HuR protein levels enhances cell survival. This phenomenon is linked to the ubiquitination of Lys182, promoting protein degradation, which finally interferes with the binding of HuR to its target mRNAs (Abdelmohsen et al., 2009).

mRNA Specific Sequence and Conformation

RBPs do not interact with the same affinity with every ARE-containing mRNA; instead, preferences exist for certain sequences. For instance, TIA-1 RRM domains display different binding constants during nucleic acid recognition. Indeed, the central domains (RRM2 and RRM3) constitute the mRNA binding platform of the protein. RRM2 drives the interaction with RNA, and shows the highest affinities for pyrimidine rich sequences. In turn, RRM3 enhances the overall TIA-1 binding affinity for RNA, preferentially interacting with C-rich motifs (Cruz-Gallardo et al., 2014; Wang et al., 2014; Waris et al., 2017). Moreover, HuR and TIAR interact with U- and AU-rich mRNAs *in vitro*, with greater affinity (≈ 10 -fold) for the former ones. This higher affinity for U-rich mRNAs results from a higher association rate constant, mainly derived from the presence of a greater number of effective binding positions (Kim et al., 2011). However, *in vivo* analysis showed that HuR stabilized AU-rich mRNAs to a greater extent than U-rich mRNAs (Brennan and Steitz, 2001). Additionally, the KH domains of KSRP behave as independent binding modules with different affinities for AU-rich mRNAs, explaining the broad range of targets recognized by the protein. While the fourth KH domain (KH4) is primarily responsible for mRNA binding and decay through an essential structural element in its β_4 , KH3 is also necessary to drive the recognition of AU- and G-rich sequences. On the other hand, all KH domains show a clear negative selection for C-rich sequences (García-Mayoral et al., 2007, 2008). Interestingly, many RNA targets of HuR, which acts antagonistically to KSRP, often contain isolated Gs but very rarely Cs (López De Silanes et al., 2004).

Conformational changes in the ARE-mRNA structure have also the potential to regulate the binding affinity of RBPs. These changes may precede the binding of RBPs, as occurs with TNF α mRNA as a consequence of the stabilization of its folding mediated by divalent cations such as Mg²⁺ (Wilson et al., 2001a,b). In addition, the AU-rich motif of TNF α mRNA can also adopt a hairpin-like structure that inhibits specifically p37^{AUF1} binding, but hardly affects its interaction with HuR (Fialcowitz et al., 2005). On the other hand, the association of RBPs can cause local changes in the structure of their cognate mRNAs, which may affect the recruitment of new *trans*-acting factors or establish preferences for one RBP over another. Consequently, these changes would directly impact on the turnover rates of such ARE-containing mRNAs (Wilson et al., 2001b; Zucconi et al., 2010).

DNA Recognition and Role of RBPs in DNA Damage Response

Some ARE-RBPs also have the ability to bind to DNA. Importantly, in the case of TIA-1 and TIAR, it occurs with a markedly higher affinity than both RBPs show for their mRNA targets (Suswam et al., 2005; Waris et al., 2017). In fact, it has been hypothesized that the formation of the RBP-mRNA complexes would require the direct displacement of the RBP from its DNA-binding site by the polymerase. This dual binding capacity of TIA-1 and TIAR could be potentially providing a link between transcription and splicing (Suswam et al., 2005; Mcalinden et al., 2007; Waris et al., 2017).

Interestingly, several RBPs are involved in DNA Damage Response (DDR), being recruited to DNA breaks in a Poly (ADP-Ribose) (PAR)-dependent manner and/or forming liquid-like compartments by phase separation (Kai, 2016). The formation of these phases requires the presence of an unstructured Prion-Related Domain (PRD) like the one that is present in TIA-1 and TIAR proteins (Gills et al., 2004). Importantly, abnormal phase separation by mutated PRD-containing proteins leads to pathological protein aggregation and is associated with neurodegenerative and aging-associated diseases (Kai, 2016).

CONCLUSIONS AND FUTURE PERSPECTIVES

Before being translated into proteins, mRNAs are subjected to a sequential and strict control by RBPs exerted by the

recognition of AREs in their 3'-UTRs. Regulation of mRNA homeostasis through ARE-RBPs allows the fine tuning of responses by controlling mRNA translation, degradation, or storage in diverse eukaryotic cell compartments (Glisovic et al., 2008; Ganguly et al., 2016). As reviewed above, many examples of ARE-RBP interactions have been reported in the literature, but it is still not well-understood how RBP domains collaborate or compete with each other for the modulation of its targets. The proper inspection of such a convoluted interplay between RBPs requires the combination of different methods in order to compensate the specific strengths and weaknesses of each technique. On the other hand, it becomes more and more evident the need of a transition from a static to a dynamic point of view to take into account the biological environment during RNA binding. Consequently, the integration of the information obtained by *in vivo* approaches with the structural data would be of great interest. Moreover, the understanding of the ARE-mRNAs processing in highly dynamic and often transient macromolecular complexes also remains challenging (Rissland, 2017). Finally, the key role of intrinsically disordered connecting linkers between RNA binding domains has acquired significant relevance in the latest reports (Basu and Bahadur, 2016). Altogether, the examples of mRNA-protein interactions by ARE-RBPs herein reviewed highlight the need for integrative studies to fully understand such a fine tuned regulation.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

Financial support was provided by the Andalusian Government (P11-CVI-7216, BIO198); the Spanish Ministry of Economy, Industry and Competitiveness (BFU2015-71017-P); the Spanish Ministry of Education, Culture and Sports (FPU013/04373, FPU016/01513) and the Ramón Areces Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2017.00071/full#supplementary-material>

REFERENCES

- Abdelmohsen, K., Srikantan, S., Yang, X., Lal, A., Kim, H. H., Kuwano, Y., et al. (2009). Ubiquitin-mediated proteolysis of HuR by heat shock. *EMBO J.* 28, 1271–1282. doi: 10.1038/emboj.2009.67
- Al-Ahmadi, W., Al-Ghamdi, M., Al-Haj, L., Al-Saif, M., and Khabar, K. S. A. (2009). Alternative polyadenylation variants of the RNA binding protein, HuR: abundance, role of AU-rich elements and auto-regulation. *Nucleic Acids Res.* 37, 3612–3624. doi: 10.1093/nar/gkp223
- Barreau, C., Paillard, L., and Osborne, H. B. (2005). AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res.* 33, 7138–7150. doi: 10.1093/nar/gki1012
- Basu, S., and Bahadur, R. P. (2016). A structural perspective of RNA recognition by intrinsically disordered proteins. *Cell. Mol. Life Sci.* 73, 4075–4084. doi: 10.1007/s00018-016-2283-1
- Benoit, R. M., Meisner, N.-C., Kallen, J., Graff, P., Hemmig, R., Cèbe, R., et al. (2010). The X-ray crystal structure of the first RNA recognition Motif and site-directed mutagenesis suggest a possible HuR Redox sensing mechanism. *J. Mol. Biol.* 397, 1231–1244. doi: 10.1016/j.jmb.2010.02.043

- Bergalet, J., Fawal, M., López, C., Desjobert, C., Lamant, L., Delsol, G., et al. (2011). HuR-mediated control of C/EBP β mRNA stability and translation in ALK-positive anaplastic large cell lymphomas. *Mol. Cancer Res.* 9, 485–496. doi: 10.1158/1541-7786.MCR-10-0351
- Brennan, C. M., and Steitz, J. A. (2001). HuR and mRNA stability. *Cell. Mol. Life Sci.* 58, 266–277. doi: 10.1007/PL00000854
- Brewer, G. (1991). An A + U-rich element RNA-binding factor regulates c-myc mRNA stability *in vitro*. *Mol. Cell. Biol.* 11, 2460–2466. doi: 10.1128/MCB.11.5.2460
- Cao, H., Deterding, L. J., Venable, J. D., Kennington, E. A., Yates, J. R., Tomer, K. B., et al. (2006). Identification of the anti-inflammatory protein tristetraprolin as a hyperphosphorylated protein by mass spectrometry and site-directed mutagenesis. *Biochem. J.* 394, 285–297. doi: 10.1042/BJ20051316
- Cao, H., Deterding, L. J., and Blackshear, P. J. (2014). Identification of a major phosphopeptide in human Tristetraprolin by phosphopeptide mapping and mass spectrometry. *PLoS ONE* 9:e100977. doi: 10.1371/journal.pone.0100977
- Carballo, E., Cao, H., Lai, W. S., Kennington, E. A., Campbell, D., and Blackshear, P. J. (2001). Decreased sensitivity of Tristetraprolin-deficient cells to p38 inhibitors suggests the involvement of Tristetraprolin in the p38 signaling pathway. *J. Biol. Chem.* 276, 42580–42587. doi: 10.1074/jbc.M104953200
- Chen, C. Y., Gherzi, R., Ong, S. E., Chan, E. L., Raijmakers, R., Pruijn, G. J., et al. (2001). AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell* 107, 451–464. doi: 10.1016/S0092-8674(01)00578-5
- Chen, C. Y., Xu, N., and Shyu, A. B. (2002). Highly selective actions of HuR in antagonizing AU-rich element-mediated mRNA destabilization. *Mol. Cell. Biol.* 22, 7268–7278. doi: 10.1128/MCB.22.20.7268-7278.2002
- Clery, A., Blatter, M., and Allain, F. H. (2008). RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* 18, 290–298. doi: 10.1016/j.sbi.2008.04.002
- Cok, S. J., Acton, S. J., and Morrison, A. R. (2003). The proximal region of the 3'-untranslated region of cyclooxygenase-2 is recognized by a multimeric protein complex containing HuR, TIA-1, TIAR, and the heterogeneous nuclear ribonucleoprotein U. *J. Biol. Chem.* 278, 36157–36162. doi: 10.1074/jbc.M302547200
- Cruz-Gallardo, I., Aroca, A., Gunzburg, M. J., Sivakumaran, A., Yoon, J. H., Angulo, J., et al. (2014). The binding of TIA-1 to RNA C-rich sequences is driven by its C-terminal RRM domain. *RNA Biol.* 11, 766–776. doi: 10.4161/rna.28801
- Cruz-Gallardo, I., Aroca, A., Persson, C., Karlsson, B. G., and Díaz-Moreno, I. (2013). RNA binding of T-cell intracellular antigen-1 (TIA-1) C-terminal RNA recognition motif is modified by pH conditions. *J. Biol. Chem.* 288, 25986–25994. doi: 10.1074/jbc.M113.489070
- Cruz-Gallardo, I., Del Conte, R., Velázquez-Campoy, A., García-Maurino, S. M., and Díaz-Moreno, I. (2015). A non-invasive NMR method based on histidine imidazoles to analyze the pH-modulation of protein-nucleic acid interfaces. *Chem. Eur. J.* 21, 7588–7595. doi: 10.1002/chem.201405538
- Daubner, G. M., Clery, A., and Allain, F. H. (2013). RRM-RNA recognition: NMR or crystallography and new findings. *Curr. Opin. Struct. Biol.* 23, 100–108. doi: 10.1016/j.sbi.2012.11.006
- Dean, J. L. E., Wait, R., Mahtani, K. R., Sully, G., Clark, A. R., and Saklatvala, J. (2001). The 3' untranslated region of tumor necrosis factor alpha mRNA is a target of the mRNA-stabilizing factor HuR. *Mol. Cell. Biol.* 21, 721–730. doi: 10.1128/MCB.21.3.721-730.2001
- Díaz-Moreno, I., Hollingworth, D., Frenkiel, T. A., Kelly, G., Martin, S., Howell, S., et al. (2009). Phosphorylation-mediated unfolding of a KH domain regulates KSRP localization via 14-3-3 binding. *Nat. Struct. Mol. Biol.* 16, 238–246. doi: 10.1038/nsmb.1558
- Díaz-Moreno, I., Hollingworth, D., Kelly, G., Martin, S., García-Mayoral, M., Briata, P., et al. (2010). Orientation of the central domains of KSRP and its implications for the interaction with the RNA targets. *Nucleic Acids Res.* 38, 5193–5205. doi: 10.1093/nar/gkq216
- Díaz-Quintana, A., García-Maurino, S. M., and Díaz-Moreno, I. (2015). Dimerization model of the C-terminal RNA recognition motif of HuR. *FEBS Lett.* 589, 1059–1066. doi: 10.1016/j.febslet.2015.03.013
- Dixon, D. A., Balch, G. C., Kedersha, N., Anderson, P., Zimmerman, G. A., Beauchamp, R. D., et al. (2003). Regulation of Cyclooxygenase-2 expression by the translational silencer TIA-1. *J. Exp. Med.* 198, 475–481. doi: 10.1084/jem.20030616
- Ellis, J. J., and Jones, S. (2008). Evaluating conformational changes in protein structures binding RNA. *Proteins* 70, 1518–1526. doi: 10.1002/prot.21647
- Embade, N., Fernández-Ramos, D., Varela-Rey, M., Beraza, N., Sini, M., Gutierrez De Juan, V., et al. (2012). Murine double minute 2 regulates Hu antigen R stability in human liver and colon cancer through NEDDylation. *Hepatology* 55, 1237–1248. doi: 10.1002/hep.24795
- Fan, X. C., and Steitz, J. A. (1998). Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the *in vivo* stability of ARE-containing mRNAs. *EMBO J.* 17, 3448–3460. doi: 10.1093/emboj/17.12.3448
- Fellows, A., Griffin, M. E., Petrella, B. L., Zhong, L., Parvin-Nejad, F. P., Fava, R., et al. (2012). AUF1/hnRNP D represses expression of VEGF in macrophages. *Mol. Biol. Cell.* 23, 1414–1422. doi: 10.1091/mbc.E11-06-0545
- Fernández-Ramos, D., and Martínez-Chantar, M. L. (2015). NEDDylation in liver cancer: the regulation of the RNA binding protein Hu antigen R. *Pancreatol.* 15, S49–S54. doi: 10.1016/j.pan.2015.03.006
- Fialcowitz, E. J., Brewer, B. Y., Keenan, B. P., and Wilson, G. M. (2005). A hairpin-like structure within an AU-rich mRNA-destabilizing element regulates trans-factor binding selectivity and mRNA decay kinetics. *J. Biol. Chem.* 280, 22406–22417. doi: 10.1074/jbc.M500618200
- Gallego-Páez, L. M., Bordone, M. C., Leote, A. C., Saraiva-Agostinho, N., Ascensao-Ferreira, M., and Barbosa-Morais, N. L. (2017). Alternative splicing: the pledge, the turn, and the prestige: the key role of alternative splicing in human biological systems. *Hum. Genet.* 136, 1015–1042. doi: 10.1007/s00439-017-1790-y
- Gama-Carvalho, M., and Carmo-Fonseca, M. (2001). The rules and roles of nucleocytoplasmic shuttling proteins. *FEBS Lett.* 498, 157–163. doi: 10.1016/S0014-5793(01)02487-5
- Ganguly, K., Giddaluru, J., August, A., and Khan, N. (2016). Post-transcriptional regulation of immunological responses through ribocustering. *Front. Immunol.* 7:161. doi: 10.3389/fimmu.2016.00161
- García-Mayoral, M. F., Díaz-Moreno, I., Hollingworth, D., and Ramos, A. (2008). The sequence selectivity of KSRP explains its flexibility in the recognition of the RNA targets. *Nucleic Acids Res.* 36, 5290–5296. doi: 10.1093/nar/gkn509
- García-Mayoral, M. F., Hollingworth, D., Masino, L., Díaz-Moreno, I., Kelly, G., Gherzi, R., et al. (2007). The structure of the C-terminal KH domains of KSRP reveals a noncanonical motif important for mRNA degradation. *Structure* 15, 485–498. doi: 10.1016/j.str.2007.03.006
- Gherzi, R., Lee, K. Y., Briata, P., Wegmuller, D., Moroni, C., Karin, M., et al. (2004). A KH domain RNA binding protein, KSRP, promotes ARE-directed mRNA turnover by recruiting the degradation machinery. *Mol. Cell.* 14, 571–583. doi: 10.1016/j.molcel.2004.05.002
- Gilks, N., Kedersha, N., Ayodele, M., Shen, L., Stoeklin, G., Dember, L. M., et al. (2004). Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Mol. Biol. Cell.* 15, 5383–5398. doi: 10.1091/mbc.E04-08-0715
- Glisovic, T., Bachorik, J. L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* 582, 1977–1986. doi: 10.1016/j.febslet.2008.03.004
- Goldberg-Cohen, I., Furneaux, H., and Levy, A. P. (2002). A 40-bp RNA element that mediates stabilization of vascular endothelial growth factor mRNA by HuR. *J. Biol. Chem.* 277, 13635–13640. doi: 10.1074/jbc.M108703200
- Gottschald, O. R., Malec, V., Krasteva, G., Hasan, D., Kamlah, F., Herold, S., et al. (2010). TIAR and TIA-1 mRNA-binding proteins co-aggregate under conditions of rapid oxygen decline and extreme hypoxia and suppress the HIF-1 α pathway. *J. Mol. Cell. Biol.* 2, 345–356. doi: 10.1093/jmcb/mjq032
- Gratacós, F. M., and Brewer, G. (2010). The role of AUF1 in regulated mRNA decay. *RNA* 1, 457–473. doi: 10.1002/wrna.26
- Gueydan, C., Droogmans, L., Chalou, P., Huez, G., Caput, D., and Kruys, V. (1999). Identification of TIAR as a protein binding to the translational regulatory AU-rich element of tumor necrosis factor alpha mRNA. *J. Biol. Chem.* 274, 2322–2326. doi: 10.1074/jbc.274.4.2322
- Hamdollah Zadeh, M. A., Amin, E. M., Hoareau-Aveilla, C., Domingo, E., Symonds, K. E., Ye, X., et al. (2015). Alternative splicing of TIA-1 in human colon cancer regulates VEGF isoform expression, angiogenesis, tumour growth and bevacizumab resistance. *Mol. Oncol.* 9, 167–178. doi: 10.1016/j.molonc.2014.07.017
- Hitti, E., Iakovleva, T., Brook, M., Deppenmeier, S., Gruber, A. D., Radzioch, D., et al. (2006). Mitogen-activated protein kinase-activated protein kinase 2 regulates tumor necrosis factor mRNA stability and

- translation mainly by altering Tristetraprolin expression, stability, and binding to adenine/uridine-rich element. *Mol. Cell. Biol.* 26, 2399–2407. doi: 10.1128/MCB.26.6.2399-2407.2006
- Ishii, T., Hayakawa, H., Sekiguchi, T., Adachi, N., and Sekiguchi, M. (2015). Role of AUF1 in elimination of oxidatively damaged messenger RNA in human cells. *Free Radic. Biol. Med.* 79, 109–116. doi: 10.1016/j.freeradbiomed.2014.11.018
- Izquierdo, J. M., and Valcárcel, J. (2007a). Fas-activated serine/threonine kinase (FAST K) synergizes with TIA-1/TIAR proteins to regulate Fas alternative splicing. *J. Biol. Chem.* 282, 1539–1543. doi: 10.1074/jbc.C600198200
- Izquierdo, J. M., and Valcárcel, J. (2007b). Two isoforms of the T-cell intracellular antigen 1 (TIA-1) splicing factor display distinct splicing regulation activities. Control of TIA-1 isoform ratio by TIA-1-related protein. *J. Biol. Chem.* 282, 19410–19417. doi: 10.1074/jbc.M700688200
- Kai, M. (2016). Roles of RNA-binding proteins in DNA damage response. *Int. J. Mol. Sci.* 17:310. doi: 10.3390/ijms17030310
- Katsanou, V., Papadaki, O., Milatos, S., Blackshear, P. J., Anderson, P., Kollias, G., et al. (2005). HuR as a negative posttranscriptional modulator in inflammation. *Mol. Cell.* 19, 777–789. doi: 10.1016/j.molcel.2005.08.007
- Kawai, T., Lal, A., Yang, X., Galban, S., Mazan-Mamczarz, K., and Gorospe, M. (2006). Translational control of cytochrome *c* by RNA-binding proteins TIA-1 and HuR. *Mol. Cell. Biol.* 26, 3295–3307. doi: 10.1128/MCB.26.8.3295-3307.2006
- Khabar, K. S. (2005). The AU-rich transcriptome: more than interferons and cytokines, and its role in disease. *J. Interferon Cytokine Res.* 25, 1–10. doi: 10.1089/jir.2005.25.1
- Kim, H. H., Kuwano, Y., Srikantan, S., Lee, E. K., Martindale, J. L., and Gorospe, M. (2009). HuR recruits let-7/RISC to repress c-Myc expression. *Genes Dev.* 23, 1743–1748. doi: 10.1101/gad.1812509
- Kim, H. S., Wilce, M. C. J., Yoga, Y. M. K., Pardini, N. R., Gunzburg, M. J., Cowieson, N. P., et al. (2011). Different modes of interaction by TIAR and HuR with target RNA and DNA. *Nucleic Acids Res.* 39, 1117–1130. doi: 10.1093/nar/gkq837
- Lai, W. S., and Blackshear, P. J. (2001). Interactions of CCCH zinc finger proteins with mRNA: Tristetraprolin-mediated AU-rich element-dependent mRNA degradation can occur in the absence of a poly(A) tail. *J. Biol. Chem.* 276, 23144–23154. doi: 10.1074/jbc.M100680200
- Lai, W. S., Carballo, E., Strum, J. R., Kennington, E. A., Phillips, R. S., and Blackshear, P. J. (1999). Evidence that Tristetraprolin binds to AU-Rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mRNA. *Mol. Cell Biol.* 19, 4311–4323. doi: 10.1128/MCB.19.6.4311
- Lasa, M., Mahtani, K. R., Finch, A., Brewer, G., Saklatvala, J., and Clark, A. R. (2000). Regulation of cyclooxygenase 2 mRNA stability by the mitogen-activated protein kinase p38 signaling cascade. *Mol. Cell. Biol.* 20, 4265–4274. doi: 10.1128/MCB.20.12.4265-4274.2000
- Lee, E. K. (2012). Post-translational modifications of RNA-binding proteins and their roles in RNA granules. *Curr. Protein Pept. Sci.* 13, 331–336. doi: 10.2174/138920312801619411
- Lee, H. H., Son, Y. J., Lee, W. H., Park, Y. W., Chae, S. W., Cho, W. J., et al. (2010). Tristetraprolin regulates expression of VEGF and tumorigenesis in human colon cancer. *Int. J. Cancer.* 126, 1817–1827. doi: 10.1002/ijc.24847
- Le Guiner, C., Lejeune, F., Galiana, D., Kister, L., Breathnach, R., Stévenin, J., et al. (2001). TIA-1 and TIAR activate splicing of alternative exons with weak 5' splice sites followed by a U-rich stretch on their own pre-mRNAs. *J. Biol. Chem.* 276, 40638–40646. doi: 10.1074/jbc.M105642200
- Levy, N. S., Chung, S., Furneaux, H., and Levy, A. P. (1998). Hypoxic stabilization of vascular endothelial growth factor mRNA by the RNA-binding protein HuR. *J. Biol. Chem.* 273, 6417–6423. doi: 10.1074/jbc.273.11.6417
- Li, H., Park, S., Kilburn, B., Jelinek, M. A., Henschen-Edman, A., Aswad, D. W., et al. (2002). Lipopolysaccharide-induced methylation of HuR, an mRNA-stabilizing protein, by CARM1. *Coactivator-associated arginine methyltransferase. J. Biol. Chem.* 277, 44623–44630. doi: 10.1074/jbc.M206187200
- Liao, B., Hu, Y., and Brewer, G. (2007). Competitive binding of AUF1 and TIAR to MYC mRNA controls its translation. *Nat. Struct. Mol. Biol.* 14, 511–518. doi: 10.1038/nsmb1249
- Lin, N. Y., Lin, C. T., Chen, Y. L., and Chang, C. J. (2007). Regulation of tristetraprolin during differentiation of 3T3-L1 preadipocytes. *FEBS J.* 274, 867–878. doi: 10.1111/j.1742-4658.2007.05632.x
- Loflin, P., Chen, C.-Y. A., and Shyu, A.-B. (1999). Unraveling a cytoplasmic role for hnRNP D in the *in vivo* mRNA destabilization directed by the AU-rich element. *Genes Dev.* 13, 1884–1897. doi: 10.1101/gad.13.14.1884
- López De Silanes, I., Zhan, M., Lal, A., Yang, X., and Gorospe, M. (2004). Identification of a target RNA motif for RNA-binding protein HuR. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2987–2992. doi: 10.1073/pnas.0306453101
- Lu, L., Wang, S., Zheng, L., Li, X., Suswam, E. A., Zhang, X., et al. (2009). Amyotrophic lateral sclerosis-linked mutant SOD1 sequesters Hu antigen R (HuR) and TIA-1-related protein (TIAR): implications for impaired post-transcriptional regulation of vascular endothelial growth factor. *J. Biol. Chem.* 284, 33989–33998. doi: 10.1074/jbc.M109.067918
- Matoukova, E., Michalova, E., Vojtesek, B., and Hrstka, R. (2012). The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biol.* 9, 563–576. doi: 10.4161/rna.20231
- Mazan-Mamczarz, K., Lal, A., Martindale, J. L., Kawai, T., and Gorospe, M. (2006). Translational repression by RNA-binding protein TIAR. *Mol. Cell. Biol.* 26, 2716–2727. doi: 10.1128/MCB.26.7.2716-2727.2006
- Mcalinden, A., Liang, L., Mukudai, Y., Imamura, T., and Sandell, L. J. (2007). Nuclear protein TIA-1 regulates COL2A1 alternative splicing and interacts with precursor mRNA and genomic DNA. *J. Biol. Chem.* 282, 24444–24454. doi: 10.1074/jbc.M702717200
- Peng, S. S., Chen, C. Y., Xu, N., and Shyu, A. B. (1998). RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. *EMBO J.* 17, 3461–3470. doi: 10.1093/emboj/17.12.3461
- Pieczyk, M., Wax, S., Beck, A. R. P., Kedersha, N., Gupta, M., Maritim, B., et al. (2000). TIA-1 is a translational silencer that selectively regulates the expression of TNF- α . *EMBO J.* 19, 4154–4163. doi: 10.1093/emboj/19.15.4154
- Pullmann, R., Kim, H. H., Abdelmohsen, K., Lal, A., Martindale, J. L., Yang, X., et al. (2007). Analysis of turnover and translation regulatory RNA-binding protein expression through binding to cognate mRNAs. *Mol. Cell. Biol.* 27, 6265–6278. doi: 10.1128/MCB.00500-07
- Rissland, O. S. (2017). The organization and regulation of mRNA-protein complexes. *Wiley Interdiscip. Rev. RNA* 2017, 8. doi: 10.1002/wrna.1369
- Sanduja, S., Blanco, F. F., and Dixon, D. A. (2011). The roles of TTP and BRF proteins in regulated mRNA decay. *Wiley Interdiscip. Rev. RNA* 2, 42–57. doi: 10.1002/wrna.28
- Sawaoka, H., Dixon, D. A., Oates, J. A., and Bouteau, O. (2003). Tristetraprolin binds to the 3'-Untranslated Region of cyclooxygenase-2 mRNA: a polyadenylation variant in a cancer cell line lacks the binding site. *J. Biol. Chem.* 278, 13928–13935. doi: 10.1074/jbc.M300016200
- Scheiba, R. M., De Opakua, A. I., Diaz-Quintana, A., Cruz-Gallardo, I., Martínez-Cruz, L. A., Martínez-Chantar, M. L., et al. (2014). The C-terminal RNA binding motif of HuR is a multi-functional domain leading to HuR oligomerization and binding to U-rich RNA targets. *RNA Biol.* 11, 1250–1261. doi: 10.1080/15476286.2014.996069
- Sengupta, S., Jang, B. C., Wu, M. T., Paik, J. H., Furneaux, H., and Hla, T. (2003). The RNA-binding protein HuR regulates the expression of cyclooxygenase-2. *J. Biol. Chem.* 278, 25227–25233. doi: 10.1074/jbc.M301813200
- Shen, Z.-J., and Malter, J. S. (2015). Regulation of AU-rich element RNA binding proteins by phosphorylation and the prolyl isomerase Pin1. *Biomolecules* 5, 412–434. doi: 10.3390/biom5020412
- Stoecklin, G., and Anderson, P. (2006). Posttranscriptional mechanisms regulating the inflammatory response. *Adv. Immunol.* 89, 1–37. doi: 10.1016/S0065-2776(05)89001-7
- Suswam, E. A., Li, Y. Y., Mahtani, H., and King, P. H. (2005). Novel DNA-binding properties of the RNA-binding protein TIAR. *Nucleic Acids Res.* 33, 4507–4518. doi: 10.1093/nar/gki763
- Szostak, E., and Gebauer, F. (2013). Translational control by 3'-UTR-binding proteins. *Brief. Funct. Genomics* 12, 58–65. doi: 10.1093/bfpg/els056
- Tchen, C. R., Brook, M., Saklatvala, J., and Clark, A. R. (2004). The stability of Tristetraprolin mRNA is regulated by mitogen-activated protein kinase p38 and by Tristetraprolin itself. *J. Biol. Chem.* 279, 32393–32400. doi: 10.1074/jbc.M402059200
- Trabucchi, M., Briata, P., García-Mayoral, M., Haase, A. D., Filipowicz, W., Ramos, A., et al. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* 459, 1010–1014. doi: 10.1038/nature08025

- Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains. *FEBS J.* 275, 2712–2726. doi: 10.1111/j.1742-4658.2008.06411.x
- Vanderweyde, T., Yu, H., Varnum, M., Liu-Yesucevitz, L., Citro, A., Ikezu, T., et al. (2012). Contrasting pathology of the stress granule proteins TIA-1 and G3BP in tauopathies. *J. Neurosci.* 32, 8270–8283. doi: 10.1523/JNEUROSCI.1592-12.2012
- Vogel, C., and Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* 13, 227–232. doi: 10.1038/nrg3185
- Wagner, B. J., Demaria, C. T., Sun, Y., Wilson, G. M., and Brewer, G. (1998). Structure and genomic organization of the Human AUF1 Gene: alternative pre-mRNA splicing generates four protein isoforms. *Genomics* 48, 195–202. doi: 10.1006/geno.1997.5142
- Wang, I., Hennig, J., Jagtap, P. K., Sonntag, M., Valcárcel, J., and Sattler, M. (2014). Structure, dynamics and RNA binding of the multi-domain splicing factor TIA-1. *Nucleic Acids Res.* 42, 5949–5966. doi: 10.1093/nar/gku193
- Wang, Z., Kayikci, M., Briesse, M., Zarnack, K., Luscombe, N. M., Rot, G., et al. (2010). iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol.* 8:e1000530. doi: 10.1371/journal.pbio.1000530
- Waris, S., García-Mauriño, S. M., Sivakumaran, A., Beckham, S. A., Loughlin, F. E., Gorospe, M., et al. (2017). TIA-1 RRM23 binding and recognition of target oligonucleotides. *Nucleic Acids Res.* 45, 4944–4957. doi: 10.1093/nar/gkx102
- Wilson, G. M., Lu, J., Sutphen, K., Suarez, Y., Sinha, S., Brewer, B., et al. (2003). Phosphorylation of p40AUF1 regulates binding to A + U-rich mRNA-destabilizing elements and protein-induced changes in ribonucleoprotein structure. *J. Biol. Chem.* 278, 33039–33048. doi: 10.1074/jbc.M305775200
- Wilson, G. M., Sutphen, K., Chuang, K., and Brewer, G. (2001a). Folding of A+U-rich RNA elements modulates AUF1 binding. *Potential roles in regulation of mRNA turnover. J. Biol. Chem.* 276, 8695–8704. doi: 10.1074/jbc.M009848200
- Wilson, G. M., Sutphen, K., Moutafis, M., Sinha, S., and Brewer, G. (2001b). Structural remodeling of an A + U-rich RNA element by cation or AUF1 binding. *J. Biol. Chem.* 276, 38400–38409. doi: 10.1074/jbc.M106509200
- Winzen, R., Thakur, B. K., Dittrich-Breiholz, O., Shah, M., Redich, N., Dhamija, S., et al. (2007). Functional analysis of KSRP interaction with the AU-Rich element of Interleukin-8 and identification of inflammatory mRNA targets. *Mol. Cell. Biol.* 27, 8388–8400. doi: 10.1128/MCB.01493-07
- Xu, N., Chen, C.-Y. A., and Shyu, A.-B. (2001). Versatile role for hnRNP D isoforms in the differential regulation of cytoplasmic mRNA turnover. *Mol. Cell. Biol.* 21, 6960–6971. doi: 10.1128/MCB.21.20.6960-6971.2001
- Yamakoshi, H., Kanoh, N., Kudo, C., Sato, A., Ueda, K., Muroi, M., et al. (2010). KSRP/FUBP2 is a binding protein of GO-Y086, a cytotoxic curcumin analogue. *ACS Med. Chem. Lett.* 1, 273–276. doi: 10.1021/ml1000454
- Yi, J., Chang, N., Liu, X., Guo, G., Xue, L., Tong, T., et al. (2010). Reduced nuclear export of HuR mRNA by HuR is linked to the loss of HuR in replicative senescence. *Nucleic Acids Res.* 38, 1547–1558. doi: 10.1093/nar/gkp1114
- Yu, Q., Cok, S. J., Zeng, C., and Morrison, A. R. (2003). Translational repression of human matrix metalloproteinases-13 by an alternatively spliced form of T-cell-restricted intracellular antigen-related protein (TIAR). *J. Biol. Chem.* 278, 1579–1584. doi: 10.1074/jbc.M203526200
- Zhang, T., Kruys, V., Huez, G., and Gueydan, C. (2002). AU-rich element-mediated translational control: complexity and multiple activities of transactivating factors. *Biochem. Soc. Trans.* 30, 952–958. doi: 10.1042/bst0300952
- Zucconi, B. E., Ballin, J. D., Brewer, B. Y., Ross, C. R., Huang, J., Toth, E. A., et al. (2010). Alternatively expressed domains of AU-rich element RNA-binding protein 1 (AUF1) regulate RNA-binding affinity, RNA-induced protein oligomerization, and the local conformation of bound RNA ligands. *J. Biol. Chem.* 285, 39127–39139. doi: 10.1074/jbc.M110.180182

Conflict of Interest Statement: 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.

The reviewer SM-L and handling Editor declared their shared affiliation.

Copyright © 2017 García-Mauriño, Rivero-Rodríguez, Velázquez-Cruz, Hernández-Vellisca, Díaz-Quintana, De la Rosa and Díaz-Moreno. 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) or licensor 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.



Insight into the Selectivity of the G7-18NATE Inhibitor Peptide for the Grb7-SH2 Domain Target

Gabrielle M. Watson, William A. H. Lucas, Menachem J. Gunzburg and Jacqueline A. Wilce*

Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC, Australia

OPEN ACCESS

Edited by:

Rivka Isaacson,
King's College London,
United Kingdom

Reviewed by:

Krystle J. McLaughlin,
Lehigh University, United States
Emeric Miclet,
Université Pierre et Marie Curie,
France

*Correspondence:

Jacqueline A. Wilce
jackie.wilce@monash.edu

Specialty section:

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

Received: 25 July 2017

Accepted: 13 September 2017

Published: 26 September 2017

Citation:

Watson GM, Lucas WAH,
Gunzburg MJ and Wilce JA (2017)
Insight into the Selectivity of the
G7-18NATE Inhibitor Peptide for the
Grb7-SH2 Domain Target.
Front. Mol. Biosci. 4:64.
doi: 10.3389/fmolb.2017.00064

Growth factor receptor bound protein 7 (Grb7) is an adaptor protein with established roles in the progression of both breast and pancreatic cancers. Through its C-terminal SH2 domain, Grb7 binds to phosphorylated tyrosine kinases to promote proliferative and migratory signaling. Here, we investigated the molecular basis for the specificity of a Grb7 SH2-domain targeted peptide inhibitor. We identified that arginine 462 in the BC loop is unique to Grb7 compared to Grb2, another SH2 domain bearing protein that shares the same consensus binding motif as Grb7. Using surface plasmon resonance we demonstrated that Grb7-SH2 binding to G7-18NATE is reduced 3.3-fold when the arginine is mutated to the corresponding Grb2 amino acid. The reverse mutation in Grb2-SH2 (serine to arginine), however, was insufficient to restore binding of G7-18NATE to Grb2-SH2. Further, using a microarray, we confirmed that G7-18NATE is specific for Grb7 over a panel of 79 SH2 domains, and identified that leucine at the β D6 position may also be a requirement for Grb7-SH2 binding. This study provides insight into the specificity defining features of Grb7 for the inhibitor molecule G7-18NATE, that will assist in the development of improved Grb7 targeted inhibitors.

Keywords: Grb7, SH2 domain, inhibitor specificity, BC loop, peptide inhibitor

INTRODUCTION

For over 20 years, Src homology 2 (SH2) domains have been explored as targets for the development of potential therapeutics (Kraskouskaya et al., 2013; Morlacchi et al., 2014). SH2 domains mediate the formation of protein complexes by recognizing phosphorylated tyrosines (pY) on target tyrosine kinases and subsequently mediating their downstream effects. They are utilized by proteins with a range of functions, including enzymes such as the PLC γ and the Src kinase (Rebecchi and Pentyala, 2000; Roskoski, 2004), transcriptional regulators including STAT 1-6 (Darnell, 1997) and adaptor proteins including growth factor receptor bound protein 7 (Grb7) (Han et al., 2001). The reversible association of the SH2 domain/pY complex allows for the efficient regulation of signaling and is therefore frequently found in proteins mediating intracellular signaling pathways—including those regulating proliferation, migration and growth (Schlessinger and Lemmon, 2003). It is these same signaling networks that are susceptible to dysregulation, leading to the amplification and transmission of signals that drive cells into a cancerous state. Thus, targeting SH2 domains can serve as an effective point for the development of anti-cancer agents that effectively block the downstream effects of signaling.

Grb7 has been identified as a therapeutic target due to its role in the progression of a number of cancers including breast, pancreatic, ovarian and oesophageal cancers (Stein et al., 1994; Tanaka et al., 1997, 2006; Wang et al., 2010). In HER2+ breast cancer, Grb7 is co-overexpressed with HER2, leading to enhanced tumorigenesis and cell proliferation (Bai and Luoh, 2007; Chu et al., 2010; Pradip et al., 2013). While HER2 has a well-known role in breast cancer progression, overexpression of Grb7 has now also been identified as a significant predictor for reduced cancer-free periods and a worse prognosis for breast cancer patients (Ramsey et al., 2011). In pancreatic, oesophageal and triple negative breast cancers, Grb7 drives migratory and invasive events, with Grb7 knockdown inhibiting these processes when tested *in vitro* (Tanaka et al., 1997, 2000, 2006; Gircz et al., 2012). Furthermore, a significant relationship has been identified between Grb7 expression and tumor metastasis in pancreatic and esophageal cancers (Tanaka et al., 1997, 2006).

Grb7 has a multi-domain architecture consisting of an N-terminal proline rich domain, a Ras-associating (RA) domain, pleckstrin homology (PH) domain, a between the PH and SH2 (BPS) domain, and lastly a C-terminal SH2 domain (Shen and Guan, 2004). It is via the SH2 domain that Grb7 interacts with phosphorylated tyrosine kinases including growth factor receptors such as HER2, HER3 and EGFR, as well as cytoplasmic kinases including the focal adhesion kinase (FAK) (Stein et al., 1994; Daly et al., 1996; Han and Guan, 1999). Through these interactions, Grb7 mediates signaling networks controlling proliferation, migration and growth, making the Grb7-SH2 domain an attractive candidate for the development of targeted inhibitors (Han and Guan, 1999; Pero et al., 2003; Pradip et al., 2013).

SH2 domains are extremely prevalent in the proteome, with over 110 proteins bearing the domain. Thus, ensuring target selectivity is a critical aspect of the development of molecules targeting SH2 domains. Despite the high number, SH2 domains display exquisite selectivity for their substrates (Pawson, 2004). SH2 domains contain a well-characterized positively charged cleft for the pY to bind, but it is typically the residues C-terminal to the target pY that dictate the binding specificity for a substrate (Songyang et al., 1993). It has been determined that individual SH2 domains recognize characteristic binding motifs. In the case of Grb7 this motif has been identified as pYXN, where any residue is accommodated at the +1 position, and an asparagine is preferred at the +2 position (Pero et al., 2003). The SH2 domains of Grb2 and Grb7 have low amino acid identity (29%), however, the Grb2-SH2 also recognizes this binding motif, and both adaptor proteins bind to HER2 at pY1139 (Janes et al., 1997). Despite this binding similarity, Grb7 and Grb2 can have remarkable selectivity and an inhibitor has been developed that specifically binds and inhibits Grb7-SH2 (Pero et al., 2002).

The cyclic, non-phosphorylated peptide (named G7-18NATE (cyclo-(CH₂CO-WFEGYDNTFPC)-amide) was developed via phage display and was found to specifically decrease binding

between Grb7 and tyrosine phosphorylated HER family members in breast cancer cell extracts, but have no effect on the interaction between Grb2 and HER3 (Pero et al., 2002). When G7-18NATE was attached to the Penetratin cell permeability sequence, the peptide inhibited growth, migration and proliferation in breast cancer cell lines, and displayed synergistic effects on cell proliferation with the currently available chemotherapeutics Doxorubicin and Trastuzumab (Pero et al., 2007; Pradip et al., 2013). G7-18NATE-Penetratin was also shown to specifically inhibit the interaction between Grb7 and FAK, but not interfere with the Grb2/FAK or Grb2/EGFR interactions (Tanaka et al., 2006). This specificity for the Grb7-SH2 has also been demonstrated *in vitro* with surface plasmon resonance (SPR) experiments confirming that G7-18NATE specifically binds to Grb7-SH2 preferentially over the Grb2-SH2 domain (Gunzburg et al., 2012). G7-18NATE also displayed minimal binding to the SH2 domains of Grb10 and Grb14, proteins that share the same domain structure as Grb7 (with the three Grbs collectively termed the Grb7 family). Derivatives of G7-18NATE have also been developed that bind to the Grb7-SH2 domain with higher affinity than G7-18NATE ($K_D = 0.83 \mu\text{M}$) but with the equivalent specificity (Watson et al., 2015; Gunzburg et al., 2016).

While it has been demonstrated repeatedly that G7-18NATE and its derivatives are specific for Grb7, there has been little investigation into the molecular basis of Grb7 ligand selectivity, except for analysis of the βD6 amino acid (Leu481 in Grb7). It is important to note that the nomenclature to be used to describe SH2 domain secondary structure is as derived by Eck et al. as illustrated in **Figure 1A** (Eck et al., 1993). While Grb10 and Grb14 possess a polar glutamine at the βD6 position, Grb7 has a hydrophobic leucine residue. When Leu481 in Grb7-SH2 was mutated to glutamine, binding to the HER2 receptor was completely abolished (Janes et al., 1997). Conversely, the opposite mutation in Grb14 (glutamine to leucine) imparted binding of Grb14 to the HER2 receptor. Together this established the importance of a leucine at the βD6 position for Grb7-SH2 domain interactions with its target motif. In the case of the Grb2-SH2 domain, however, the βD6 amino acid is a lysine and this would be predicted to interfere with ligand binding. Yet some binding to Grb2-SH2 domain by G7-18NATE still takes place, showing that it is not only the βD6 amino acid that dictates G7-18NATE binding (Gunzburg et al., 2012).

The structure of the Grb7-SH2 domain in complex with G7-18NATE revealed the amino acid residues forming the binding site for the peptide ligand (Ambaye et al., 2011b). It showed the way in which G7-18NATE Y5 and F2 (note that single letter code will be used in reference to the peptide ligand amino acids for clarity) straddle the Grb7-SH2 domain Leu481 residue, helping to visualize the role of this amino acid in the specificity of the interaction. Leu481, and other Grb7-SH2 domain amino acids that form hydrogen bond interactions or contribute to the buried surface area with G7-18NATE are indicated in **Figure 1C**. For the most part, these residues are conserved between Grb7-SH2 domain and the closely related Grb10, Grb14, and Grb2 SH2 domains and, thus, do not help to explain the basis for Grb7 ligand specificity. Only the βE4 methionine and EF4 glutamine are unique to the Grb7-SH2 domain and may act alongside Leu481 to contribute to ligand specificity.

Abbreviations: BPS, between PH and SH2 domain; FAK, focal adhesion kinase; Grb, growth factor receptor bound protein; PH, pleckstrin homology; pY, phosphorylated tyrosine; RA, ras-associating; SPR, surface plasmon resonance; SH2, src homology 2; WT, wild-type.

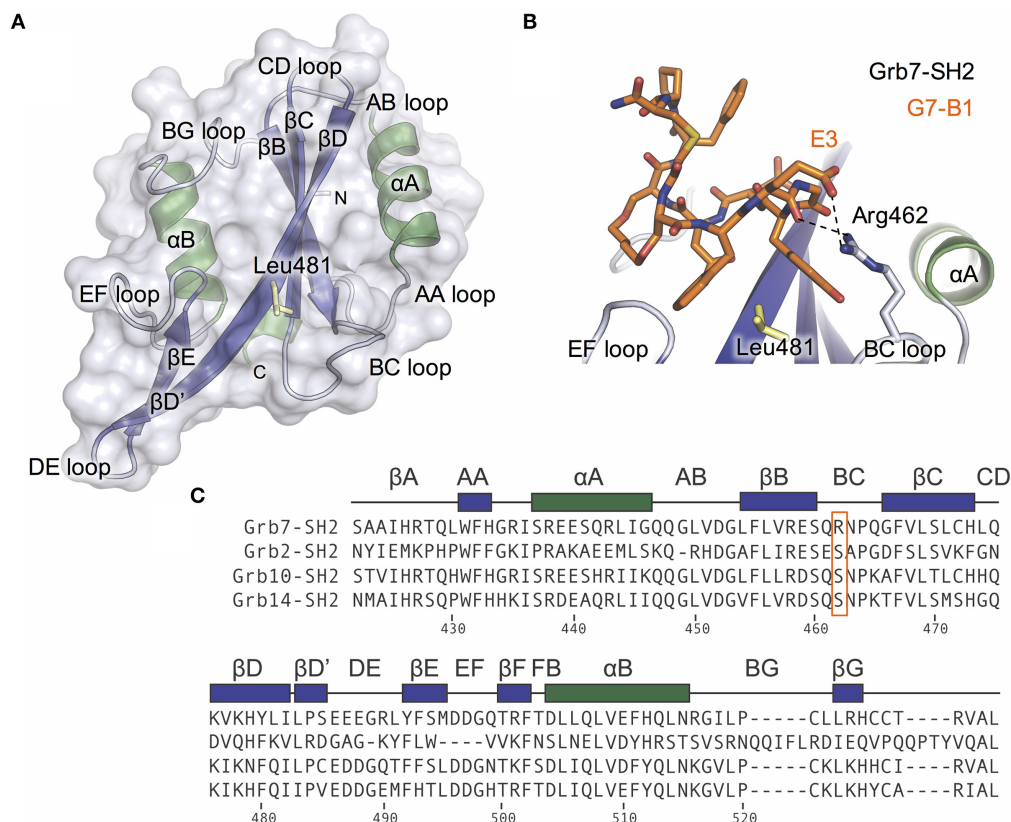


FIGURE 1 | The Grb7 specific amino acid Arg462 may contribute to the Grb7-SH2/G7-18NATE interaction. **(A)** Overview of the Grb7-SH2 domain with SH2 domain nomenclature displayed. Surface representation and loops displayed in gray, α -helices in green and β -sheet in purple. The previously described specificity defining Leu481 is displayed as yellow sticks; **(B)** Close-up of the interaction of the G7-B1 bicyclic peptide (orange sticks) with the Grb7-SH2 domain colored as in **(A)**. Hydrogen bonds between the Grb7-SH2 Arg462 and the peptide E3 are displayed as dashed lines (PDB ID: 5EEQ); **(C)** Protein sequence alignment of the SH2 domains of Grb2, Grb7, Grb10 and Grb14, generated using Clustal Omega (Sievers et al., 2011). Secondary structural elements for the Grb7-SH2 domain are displayed above the alignment at the corresponding position, and sequence numbering for the Grb7-SH2 displayed below the alignment. Grb7 amino acids that form hydrogen bond interactions with the G7-18NATE peptide are identified with filled circles and additional amino acids that contribute to the buried surface area are identified with empty circles. The unique Grb7-SH2 arginine at position 462 (BC2) is highlighted by an orange box.

In this study, we identified an additional amino acid in the Grb7-SH2 BC loop that is involved in binding to G7-18NATE and potentially also contributes to Grb7-SH2 domain binding specificity. Based on another Grb7-SH2/peptide complex X-ray crystal structure, we postulated that the Grb7 BC2 arginine contributes to the Grb7-SH2/peptide interaction. In Grb2, Grb10, and Grb14 the BC2 residue is a serine. Using mutational analysis we demonstrate that this Grb7 residue does indeed contribute to the G7-18NATE interaction, and thus may dictate some of the Grb7 selectivity. We identified that mutating the BC2 residue in the Grb2-SH2 from serine to arginine is not sufficient to confer binding to G7-18NATE indicating that other Grb7 amino acids also contribute to binding strength and specificity. Lastly, using a pY micro-array we confirmed that G7-18NATE is specific for Grb7 compared to a panel of 79 SH2 domains, providing extra verification of the specificity of this peptide. Together this study provides insight into the specificity of G7-18NATE for Grb7 which is critical for the development of effective therapeutics targeted to the Grb7-SH2 domain signaling in cancer.

MATERIALS AND METHODS

Protein and Peptide Preparation

Grb7-SH2 and Grb2-SH2 were prepared as GST fusion proteins and expressed and purified as previously described (Watson et al., 2015). Briefly, Grb7-SH2 (residues 415-532) and Grb2-SH2 (residues 58-160) were incorporated into the pGex2T plasmid and expressed in BL21(DE3)pLysS competent cells following IPTG induction (0.4 mM) at 25°C. Subsequently, glutathione affinity chromatography and size exclusion chromatography were used to purify the soluble GST tagged proteins. To generate the Grb7-SH2 R462S and Grb2-SH2 S90R mutants, site-directed mutagenesis was utilized (Liu and Naismith, 2008). For this, the wild-type SH2 domain constructs were used as templates for polymerase chain reaction with the primers listed in Supplementary Material. The parent plasmid was subsequently removed by treatment with *DpnI* and the presence of the mutation verified by DNA sequencing. The mutant proteins were expressed and purified as per the wild-type proteins. GST alone was purified similarly to the GST-SH2 domain proteins with the

exception that size exclusion chromatography was not necessary for purification.

G7-18NATE (cyclo-(CH₂CO-WFEGYDNTFPC)-amide) was synthesized using standard Fmoc-chemistry and purchased from Purar Chemicals (Australia). The synthesis of G7-18NATE-PB ((cyclo-(CH₂CO-WFEGYDNTFPC-RRMKWKKK(Biotin))-amide)) has been previously described (Ambaye et al., 2011a). The purity of both peptides was >95% as determined using LC-MS.

The final solution concentration of all proteins and peptides used in this study were determined spectroscopically at 280 nm using extinction coefficients predicted by the ProtParam server (Gasteiger et al., 2005).

Binding Studies Using Surface Plasmon Resonance

SPR experiments were performed on a Biacore T100 using CM5 series S sensor chips. The GST tagged proteins were immobilized onto the sensor chip surface by amine coupling an anti-GST antibody to the surface of the chip. For this, firstly the chip was activated by 0.4 M 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride and 0.1 M N-hydroxysuccinimide. The polyclonal antibody (Abcam) was passed over the chip surface at 30 µg.mL⁻¹ before the flow cells were blocked with 1 M ethanolamine. The running buffer used for the immobilization contained 50–150 mM NaPO₄, 150 mM NaCl and 1 mM DTT (pH 7.4), with levels reaching between 3289 RU and 6324 RU. The GST fusion proteins were passed over the active chip surface at 0.7–0.9 µM to achieve final immobilization levels of 838–1283 RU. GST alone was immobilized on the reference flow cell (462–616 RU) as a negative control to allow for double referencing. Lyophilized G7-18NATE was resuspended in running buffer containing 150 mM NaPO₄, 150 mM NaCl and 1 mM DTT (pH 7.4) and injected over the chip surface for 60 s at 30 µL.min⁻¹ in triplicates. Due to machine error some injections were removed before analysis due to presence of spikes that rendered the sensorgrams unable to be interpreted. The interpretable data were analyzed using Scrubber2.0 (BioLogic Software, Campbell, ACT, Australia) and GraphPad Prism.

SH2 Domain pY Microarray

The microarray was conducted by the Protein Array and Analysis Core at the MD Anderson Cancer Center (University of Texas, USA) using standard protocols. Briefly, lyophilized G7-18NATE-PB (100 µg) was resuspended in PBST and incubated with the nitrocellulose slide pre-spotted with GST-fusion SH2 domains (full list provided as Supplementary Material). Each SH2 domain was spotted in duplicate.

RESULTS

The Unique Grb7-SH2 BC2 Amino Acid Engages E3 in the G7-B1 Peptide

We recently characterized the interaction between the Grb7-SH2 and a G7-18NATE derivative peptide, G7-B1 (B for bicyclic) (Gunzburg et al., 2016). In the course of our investigation, we solved the X-ray crystal structure of the Grb7-SH2/G7-B1 complex to 1.6 Å (PDB ID: 5EEQ). This revealed an interesting

feature potentially underlying binding specificity of the Grb7 targeting peptides. We identified that an arginine at the BC2 position (Arg462) in the Grb7 BC loop reaches up and hydrogen bonds with the G7-B1 E3 backbone carbonyl and forms a salt bridge with the E3 sidechain carboxylate (**Figure 1B**; Gunzburg et al., 2016).

A sequence alignment of the SH2 domains of Grb2, Grb7, Grb10, and Grb14 revealed that an arginine at this BC2 position is unique to Grb7 with Grb10, Grb14, and Grb2 possessing a serine at the equivalent position (**Figure 1C**). Although the Arg462/E3 interaction has not been clearly observed in other Grb7-SH2/peptide crystal structures (PDB ID:3PQZ Ambaye et al., 2011b; PDB ID 4X6S Watson et al., 2015; Gunzburg et al., 2016), we postulated that this interaction might occur transiently and help to define the molecular basis of binding specificity of G7-18NATE for Grb7.

Arg462 Contributes to the Binding Affinity between Grb7-SH2/G7-18NATE

To test this hypothesis, we mutated the Grb7-SH2 R462 (BC2) amino acid to a serine (the corresponding residue in the Grb2, Grb10, and Grb14 SH2 domains). Using SPR, we directly compared the binding of the cyclic peptide G7-18NATE to the Grb7-SH2 wild-type (WT) and the mutated Grb7-SH2 (R462S). The Grb7-SH2 proteins were prepared as GST fusions to enable efficient immobilization on the sensor chip surface via anti-GST antibodies. The G7-18NATE peptide was flowed over the sensor chip surface and the measured response at equilibrium used to construct binding curves that were fitted by a single-site binding model.

Upon injection of G7-18NATE, equilibrium was rapidly reached for both WT Grb7-SH2 (**Figure 2A**) and the mutated R462S Grb7-SH2 (**Figure 2B**), followed by rapid dissociation after the 60 s injection was complete. Excellent fits by a 1:1 interaction model were obtained with a *R*² of 0.99 for both interactions. The binding curves indicated that the R462 of the BC loop did indeed contribute to the binding affinity of the G7-18NATE/Grb7-SH2 interaction. While G7-18NATE bound to WT Grb7-SH2 with a *K*_D of 2.43 µM, the R462S mutant had a 3.3-fold reduction in binding affinity for G7-18NATE, with a *K*_D of 7.97 µM (**Figure 2C**; **Table 1**). This suggested that the BC2 arginine does indeed engage with the peptide ligand, and therefore may impart some of the specificity of Grb7 for its substrates.

The Corresponding Grb2 Mutation Is Insufficient to Confer Binding to G7-18NATE

We next postulated that the opposite mutation at the BC2 position in the Grb2-SH2 (serine to arginine) could confer binding of the Grb2-SH2 to G7-18NATE. This would indicate the degree to which Arg462 in Grb7 is responsible for binding specificity. We therefore generated the Grb2-SH2 S90R (BC2) mutant protein using site-directed mutagenesis and tested its binding to G7-18NATE using the same SPR experimental procedure described above. As shown in **Figure 3A**, there was

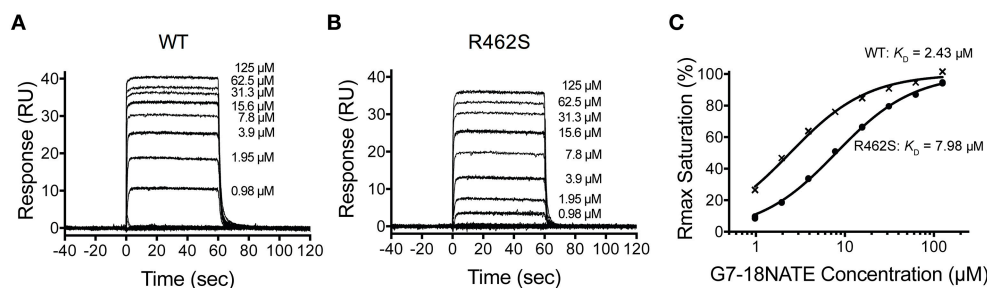


FIGURE 2 | The Grb7 R462S mutant has decreased binding for G7-18NATE. Sensorgrams of the interaction between G7-18NATE with either wild-type (WT) Grb7-SH2 (**A**) or mutated (R462S) Grb7-SH2 (**B**) with the concentrations of peptide tested displayed on the right of the sensorgram; (**C**) The corresponding binding curves were calculated from the responses at equilibrium and fitted by a 1:1 interaction model. All data points from the repeated experiments are displayed in the binding profiles.

TABLE 1 | SPR investigations into the specificity of G7-18NATE for the Grb7-SH2.

| Protein | Equilibrium Dissociation Constant (μM) [#] | Rmax (RU) |
|----------------|--|-----------------|
| Grb7-SH2 WT | 2.43 \pm 0.01 | 39.7 \pm 0.05 |
| Grb7-SH2 R462S | 7.98 \pm 0.03 | 37.9 \pm 0.04 |
| Grb2-SH2 WT | > 125 | 14.2 \pm 0.9 |
| Grb2-SH2 S90R | > 125 | 41.0 \pm 3 |

[#]Equilibrium dissociation constants were calculated from fits by a single-site binding model. Errors displayed are the standard error of the fits.

only a low binding response to the Grb2-SH2 WT following injection of G7-18NATE over the chip surface, reflecting low binding affinity as previously established for Grb2-SH2 compared to Grb7-SH2 (Gunzburg et al., 2012). The binding to G7-18NATE by Grb2-SH2 S90R, however, also gave rise to low responses, though slightly higher than seen for Grb2-SH2 WT (**Figure 3B**). Excellent fits were obtained by a single-site binding model ($R^2 = 0.99$), however the calculated equilibrium dissociation constant was higher than the highest peptide concentration measured (125 μM ; **Figure 3C**). Therefore, we could not reliably determine differences in the K_D between WT Grb2-SH2 and the mutated Grb2-SH2. This demonstrated, nevertheless, that the incorporation of the arginine at the BC2 position was not sufficient to significantly increase the binding of Grb2-SH2 to G7-18NATE.

G7-18NATE Is Specific for the Grb7-SH2 Compared to a Panel of SH2 Domains

In the course of our analysis of G7-18NATE and its derivatives, specificity was assessed by comparing the binding of the peptides to the SH2 domains of Grb2, Grb7, and Grb10, and in some instances Grb14-SH2 (Watson et al., 2015; Gunzburg et al., 2016). While Grb2 shares the same consensus pYXN binding motif, Grb10 and Grb14 share significant sequence identity (approximately 70%) with Grb7-SH2 (Daly et al., 1996). We therefore considered that these were the most closely related SH2 domains for assessment of specificity for Grb7-SH2 using SPR experiments.

In order to determine whether there were other SH2 domains that G7-18NATE could bind to that we had not yet encountered,

we utilized the pY reader protein microarray at the Protein Array and Analysis Core (MD Anderson Cancer Center). Here, a biotinylated G7-18NATE derivative (G7-18NATE-PB) was tested for binding against 79 SH2 domains (a complete list of tested SH2 domains is provided as Supplementary Material). The G7-18NATE derivative also contained a shortened Penetratin sequence (displayed as a schematic in **Figure 4A**), and the presence of this sequence has been previously demonstrated not to interfere with G7-18NATE binding to the Grb7-SH2 (Ambaye et al., 2011a; Watson et al., 2017). As shown in **Figure 4B**, it was clear that G7-18NATE-PB preferentially binds to the Grb7-SH2 over the suite of SH2 domains. Although this array was not quantitative, the level of fluorescence was found to be significantly higher for the positions containing the Grb7-SH2, indicative of peptide binding. This was not due to differences in the amount of protein blotted on the array, as similar fluorescence levels were detected when probed with an anti GST antibody (**Figure 4B**).

There was no observation of fluorescence at the positions corresponding to Grb2, Grb10, or Grb14, consistent with our previously published SPR experiments (Gunzburg et al., 2012). Fluorescence was observed, although at much lower intensity, at the positions corresponding to a previously untested SH2 domain—that of SH2D2A. A sequence alignment of the SH2 domains of Grb7 and SH2D2A revealed that SH2D2A also contains a serine at the BC2 position and, strikingly, a leucine at the βD6 position (**Figure 4C**). Although this is not the only SH2 domain in the array that possesses a leucine at the βD6 position, it is suggestive that leucine at this position is also determining feature for binding the G7-18NATE peptide and its derivatives. Thus, as well as further confirming the specificity of the ligand G7-18NATE for the Grb7-SH2 domain, this experiment has also provided further insights into the defining features of Grb7 that affect ligand binding selectivity.

DISCUSSION

SH2 domains are frequently found in signaling pathways regulating critical processes that are perturbed in aggressive diseases like cancer. Due to this, and their well-characterized

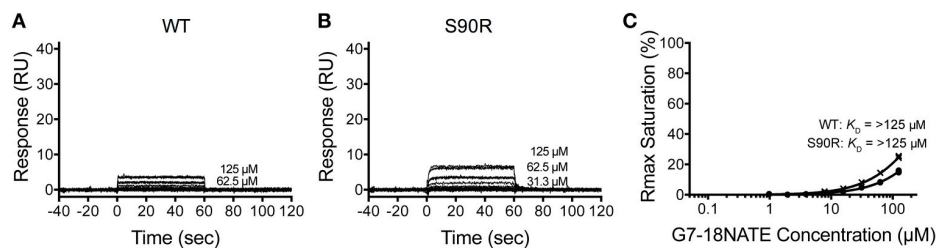


FIGURE 3 | The Grb2 S90R mutant has minimal binding to G7-18NATE. Sensorgrams of the interaction between G7-18NATE with either wild-type (WT) Grb2-SH2 (A) or mutated (S90R) Grb2-SH2 (B) with the concentrations of peptide tested displayed on the right of the sensorgram (lower concentrations omitted for clarity); (C) The corresponding binding curves were calculated from the response at equilibrium and fitted by a 1:1 interaction model. All data points from the repeated experiments are displayed in the binding profiles.

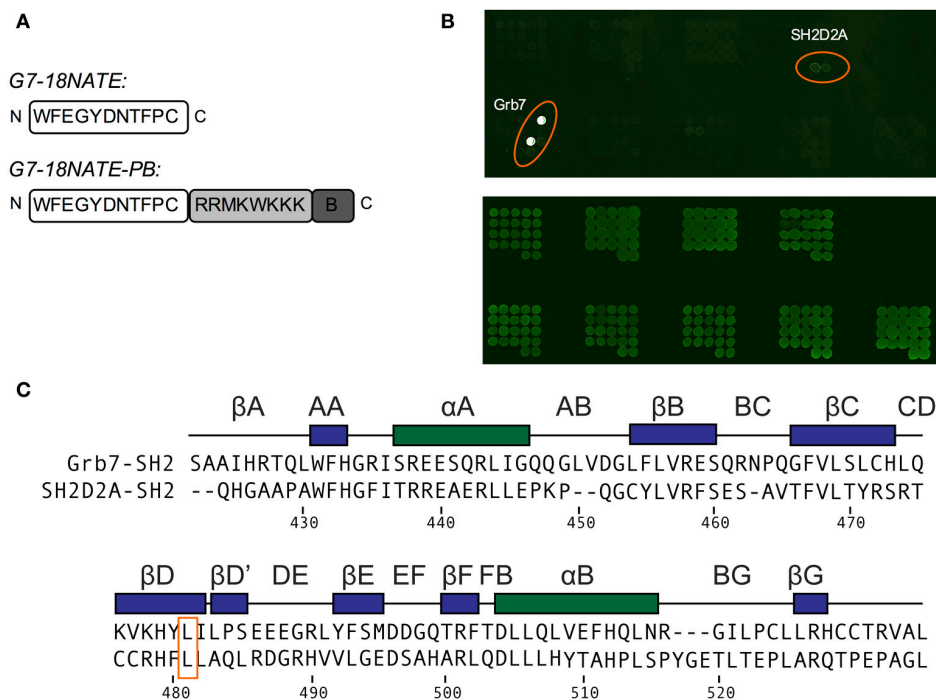


FIGURE 4 | G7-18NATE-PB is specific for Grb7 over a panel of SH2 domains. (A) Schematic of G7-18NATE and G7-18NATE-PB with the single letter amino acid code and the amino (N) and carboxy (C) termini displayed. (B) Represents the biotin moiety. (C) SH2 domain protein microarray. Top Panel: following treatment with 100 μg of G7-18NATE-PB; Bottom Panel: Anti-GST antibody loading control to show equivalent amounts of SH2 domain and the positions of the microarray spots. A list of SH2 domains in the microarray is provided as Supplementary Information. (C) Protein sequence alignment of the SH2 domains of Grb7 and Sh2D2A, generated using Clustal Omega (Sievers et al., 2011). Secondary structural elements for the Grb7-SH2 domain are displayed above the alignment at the corresponding position, and sequence numbering for the Grb7-SH2 displayed below the alignment. The shared leucine at the βD6 position is highlighted by an orange box.

binding mode, a number of SH2 domains have been targeted for the development of novel therapeutics (Kraskouskaya et al., 2013; Morlacchi et al., 2014). However, with over 110 SH2 domains in the proteome, ensuring selectivity for the target SH2 domain is a significant challenge in the development of specific and potent inhibitors. Grb7 is a signaling adaptor protein that has been specifically targeted through the development of cyclic peptide inhibitors that are able to inhibit proliferation and migration in breast cancer cell lines and reduce tumor metastasis in a mouse model of pancreatic cancer (Tanaka et al., 2006; Pradip et al., 2013). The cyclic peptide, G7-18NATE, has been demonstrated

to be specific for the Grb7-SH2 domain over the closely related SH2 domains of Grb2, Grb10, and Grb14, and this specificity has, to date, been attributed to the Grb7 specific leucine at the βD6 position (Janes et al., 1997; Gunzburg et al., 2012). Here, we have investigated whether additional amino acids also contribute to the specificity of Grb7 for its ligands.

Based on the high-resolution X-ray crystal structure of the Grb7-SH2/G7-B1 complex, we identified an arginine in the Grb7-SH2 BC loop that mediates additional hydrogen bond interactions with the inhibitor peptide. We identified that this arginine is unique to Grb7 compared to the SH2 domains of

Grb2, Grb10 and Grb14 (where there is a serine at the comparable position) and considered that this interaction could contribute to the specificity for Grb7 of all the G7-18NATE derived peptide ligands. Using mutagenesis we henceforth determined that Arg462 contributes to the binding affinity between Grb7-SH2 and G7-18NATE. The converse mutation in the Grb2-SH2 domain did not confer binding to G7-18NATE, suggesting other amino acids also define Grb7 specificity. We also probed a panel of 79 SH2 domains with a biotinylated version of G7-18NATE and identified that the peptide is highly specific for the Grb7-SH2 domain. The only other SH2 domain that showed detectable binding by G7-18NATE was that of SH2D2A. Interestingly, while SH2D2A does not possess an arginine at the BC2 position it has a leucine at the β D6 position.

The initial observation that Arg462 may contribute to the binding interaction came from the Grb7-SH2/G7-B1 X-ray crystal structure (PDB ID: 5EEQ). In this structure, the Arg462 side-chain extends and engages the E3 backbone carbonyl in G7-18NATE, as well as the E3 sidechain carboxylic acid. The former of these interactions is likely to be more influential to binding. In previously described crystal structures of cyclic peptides bound to the Grb7-SH2 domain, the E3 sidechain is positioned in various rotamer conformations whereas Arg462 is frequently observed in the proximity of the E3 backbone carbonyl (Gunzburg et al., 2016). Furthermore, peptide ligands without glutamic acid at the E3 position have also been shown to bind the Grb7-SH2 domain (Ambaye et al., 2011b). This was initially identified using a constrained phage display library, enriched for a CX₁FX₂GYDNX₃X₄X₅ motif, whereby tryptophan was selected in approximately 50% of clones at the corresponding E3 position. Thus, G7-18NATE derivatives that are modified at the E3 position are predicted to maintain the hydrogen bond interaction with the Grb7 Arg462 (BC2). It should be noted that the formation of this single hydrogen bond, that may only occur transiently (since it is not observed in all crystal structures), is consistent with the relatively small increase in affinity that is conferred. A 3.3-fold change in K_D corresponds to approximately only 1 kcal/mol change in ΔG , which is less than expected for a well formed hydrogen bond.

The G7-18NATE peptide and its derivatives are constrained in a β -turn conformation with extensive intra-molecular hydrogen bonds. A hydrogen bond between the F2 backbone carbonyl and the Y5 backbone amine positions the E3 backbone carbonyl for engagement with the Arg462 (BC2) sidechain. This suggests that the Arg462/substrate interaction may only occur in ligands that are naturally restricted in a β -turn conformation. To date, no X-ray crystal structures have been solved of Grb7 in complex with an *in vivo* binding partner (such as FAK or HER2) to determine whether or not Arg462 (BC2) contributes to the interaction with these molecules, and defines Grb7 selectivity for natural substrates.

We have also described how the introduction of the arginine mutation to the Grb2-SH2 at the BC2 position was insufficient to confer binding by the G7-18NATE peptide (although it may have made a contribution to binding to Grb10 or Grb14 that are more closely related to Grb7). This suggests that additional

amino acids underlie the specificity of G7-18NATE for Grb7-SH2 over Grb2-SH2. This is supported by our microarray results where the only SH2 domain other than Grb7 to show detectable binding to G7-18NATE was that of SH2D2A, an SH2 domain that contains a leucine at the β D6 position. This finding is consistent with previous reports that report the contribution of leucine at the β D6 position to the specificity of Grb7. When Leu481 is mutated to glutamine (the corresponding amino acid in Grb14), binding to the HER2 receptor is lost, and conversely in Grb14, the glutamine to leucine mutation enables HER2 binding (Janes et al., 1997). The X-ray crystal structure of the Grb7-SH2/G7-18NATE complex revealed how a leucine at this position enhances the interaction (Ambaye et al., 2011b). Leucine at the β D6 position is not, however, sufficient for G7-18NATE binding to SH2 domains, as other SH2 domains in the microarray with leucine present at this position showed no ability to be bound by G7-18NATE. Clearly other amino acid residue differences contribute to binding specificity such as the β E4 methionine or EF4 glutamine that are unique to the Grb7-SH2 domain.

Thus, in this study we have demonstrated that Grb7 Arg462 contributes to G7-18NATE binding, but that specificity is mediated by multiple factors including the leucine at the β D6 position. We have also established that the G7-18NATE peptide is exquisitely specific for Grb7-SH2 over other SH2 domains. Together this provides insight for the development of even more potent ligands that target the Grb7-SH2 domain, and demonstrates that SH2 domains, despite their common structural features, are readily able to differentiate between ligands and thus have high potential as targets for therapeutics development.

AUTHOR CONTRIBUTIONS

GW was responsible for the detailed experimental design, protein overexpression and SPR experiments, figure and manuscript preparation. WL undertook the cloning of SH2 domain mutants and was also involved in protein expression, purification and SPR experiments. MG provided SPR expertise and training. JW designed the conceptual framework for the experiments, oversaw the interpretation of the results and finalized the manuscript with the other authors.

ACKNOWLEDGMENTS

We wish to thank the Monash Micromon DNA sequencing facility, and Danielle Cini and the MD Anderson Cancer Center for assistance with the pY microarray. This work was supported by a grant from the National Health and Medical Research Council (APP3236668) awarded to JW.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2017.00064/full#supplementary-material>

REFERENCES

- Ambaye, N. D., Lim, R. C., Clayton, D. J., Gunzburg, M. J., Price, J. T., Pero, S. C., et al. (2011a). Uptake of a cell permeable G7-18NATE construct into cells and binding with the Grb-7-SH2 domain. *Biopolymers* 96, 181–188. doi: 10.1002/bip.21403
- Ambaye, N. D., Pero, S. C., Gunzburg, M. J., Yap, M., Clayton, D. J., Del Borgo, M. P., et al. (2011b). Structural basis of binding by cyclic nonphosphorylated peptide antagonists of Grb7 implicated in breast cancer progression. *J. Mol. Biol.* 412, 397–411. doi: 10.1016/j.jmb.2011.07.030
- Bai, T., and Luoh, S. W. (2007). GRB-7 facilitates HER-2/Neu-mediated signal transduction and tumor formation. *Carcinogenesis* 29, 473–479. doi: 10.1093/carcin/bgm221
- Chu, P. Y., Li, T. K., Ding, S. T., Lai, I. R., and Shen, T. L. (2010). EGF-induced Grb7 recruits and promotes Ras activity essential for the tumorigenicity of Sk-Br3 breast cancer cells. *J. Biol. Chem.* 285, 29279–29285. doi: 10.1074/jbc.C110.114124
- Daly, R. J., Sanderson, G. M., Janes, P. W., and Sutherland, R. L. (1996). Cloning and characterization of GRB14, a novel member of the GRB7 gene family. *J. Biol. Chem.* 271, 12502–12510. doi: 10.1074/jbc.271.21.12502
- Darnell, J. E. Jr. (1997). STATs and gene regulation. *Science* 277, 1630–1635. doi: 10.1126/science.277.5332.1630
- Eck, M. J., Shoelson, S. E., and Harrison, S. C. (1993). Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. *Nature* 362, 87–91. doi: 10.1038/362087a0
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). *Protein Identification and Analysis Tools on the ExPASy Server*. New York, NY: Humana Press.
- Gircz, O., Calvo, V., Pero, S. C., Krag, D. N., Sparano, J. A., and Kenny, P. A. (2012). GRB7 is required for triple-negative breast cancer cell invasion and survival. *Breast Cancer Res. Treat.* 133, 607–615. doi: 10.1007/s10549-011-1822-6
- Gunzburg, M. J., Ambaye, N. D., Del Borgo, M. P., Pero, S. C., Krag, D. N., Wilce, M. C., et al. (2012). Interaction of the non-phosphorylated peptide G7-18NATE with Grb7-SH2 domain requires phosphate for enhanced affinity and specificity. *J. Mol. Recognit.* 25, 57–67. doi: 10.1002/jmr.2148
- Gunzburg, M. J., Kulkarni, K., Watson, G. M., Ambaye, N. D., Del Borgo, M. P., Brandt, R., et al. (2016). Unexpected involvement of staple leads to redesign of selective bicyclic peptide inhibitor of Grb7. *Sci. Rep.* 6:27060. doi: 10.1038/srep27060
- Han, D. C., and Guan, J. L. (1999). Association of focal adhesion kinase with Grb7 and its role in cell migration. *J. Biol. Chem.* 274, 24425–24430. doi: 10.1074/jbc.274.34.24425
- Han, D. C., Shen, T. L., and Guan, J. L. (2001). The Grb7 family proteins: structure, interactions with other signaling molecules and potential cellular functions. *Oncogene* 20, 6315–6321. doi: 10.1038/sj.onc.1204775
- Janes, P. W., Lackmann, M., Church, W. B., Sanderson, G. M., Sutherland, R. L., and Daly, R. J. (1997). Structural determinants of the interaction between the erbB2 receptor and the Src homology 2 domain of Grb7. *J. Biol. Chem.* 272, 8490–8497. doi: 10.1074/jbc.272.13.8490
- Kraskouskaya, D., Duodu, E., Arpin, C. C., and Gunning, P. T. (2013). Progress towards the development of SH2 domain inhibitors. *Chem. Soc. Rev.* 42, 3337–3370. doi: 10.1039/c3cs35449k
- Liu, H., and Naismith, J. H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* 8:91. doi: 10.1186/1472-6750-8-91
- Morlacchi, P., Robertson, F. M., Klostergaard, J., and McMurray, J. S. (2014). Targeting SH2 domains in breast cancer. *Future Med. Chem.* 6, 1909–1926. doi: 10.4155/fmc.14.120
- Pawson, T. (2004). Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* 116, 191–203. doi: 10.1016/S0092-8674(03)01077-8
- Pero, S. C., Daly, R. J., and Krag, D. N. (2003). Grb7-based molecular therapeutics in cancer. *Expert Rev. Mol. Med.* 5, 1–11. doi: 10.1017/S1462399403006227
- Pero, S. C., Oligno, L., Daly, R. J., Soden, A. L., Liu, C., Roller, P. P., et al. (2002). Identification of novel non-phosphorylated ligands, which bind selectively to the SH2 domain of Grb7. *J. Biol. Chem.* 277, 11918–11926. doi: 10.1074/jbc.M111816200
- Pero, S. C., Shukla, G. S., Cookson, M. M., Flemer, S. Jr., and Krag, D. N. (2007). Combination treatment with Grb7 peptide and Doxorubicin or Trastuzumab (Herceptin) results in cooperative cell growth inhibition in breast cancer cells. *Br. J. Cancer* 96, 1520–1525. doi: 10.1038/sj.bjc.6603732
- Pradip, D., Bouzyk, M., Dey, N., and Leyland-Jones, B. (2013). Dissecting GRB7-mediated signals for proliferation and migration in HER2 overexpressing breast tumor cells: GTP-ase rules. *Am. J. Cancer Res.* 3, 173–195.
- Ramsey, B., Bai, T., Hanlon Newell, A., Troxell, M., Park, B., Olson, S., et al. (2011). GRB7 protein over-expression and clinical outcome in breast cancer. *Breast Cancer Res. Treat.* 127, 659–669. doi: 10.1007/s10549-010-1010-0
- Rebecchi, M. J., and Pentylä, S. N. (2000). Structure, function, and control of phosphoinositide-specific phospholipase C. *Physiol. Rev.* 80, 1291–1335.
- Roskoski, R. Jr. (2004). Src protein-tyrosine kinase structure and regulation. *Biochem. Biophys. Res. Commun.* 324, 1155–1164. doi: 10.1016/j.bbrc.2004.09.171
- Schlessinger, J., and Lemmon, M. A. (2003). SH2 and PTB domains in tyrosine kinase signaling. *Sci. STKE* 2003:RE12. doi: 10.1126/stke.2003.191.re12
- Shen, T. L., and Guan, J. L. (2004). Grb7 in intracellular signaling and its role in cell regulation. *Front. Bio. Sci.* 9, 192–200. doi: 10.2741/1229
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75
- Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., et al. (1993). SH2 domains recognize specific phosphopeptide sequences. *Cell* 72, 767–778. doi: 10.1016/0092-8674(93)90404-E
- Stein, D., Wu, J., Fuqua, S. A., Roonprapunt, C., Jainik, V., D'eustachio, P., et al. (1994). The SH2 domain protein GRB-7 is co-amplified, overexpressed and in a tight complex with HER2 in breast cancer. *EMBO J.* 13, 1331–1340.
- Tanaka, S., Mori, M., Akiyoshi, T., Tanaka, Y., Mafune, K., Wands, J. R., et al. (1997). Coexpression of Grb7 with epidermal growth factor receptor or Her2/erbB2 in human advanced esophageal carcinoma. *Cancer Res.* 57, 28–31.
- Tanaka, S., Pero, S. C., Taguchi, K., Shimada, M., Mori, M., Krag, D. N., et al. (2006). Specific peptide ligand for Grb7 signal transduction protein and pancreatic cancer metastasis. *J. Natl. Cancer Inst.* 98, 491–498. doi: 10.1093/jnci/djj105
- Tanaka, S., Sugimachi, K., Kawaguchi, H., Saeki, H., Ohno, S., and Wands, J. R. (2000). Grb7 signal transduction protein mediates metastatic progression of esophageal carcinoma. *J. Cell Physiol.* 183, 411–415. doi: 10.1002/(SICI)1097-4652(200006)183:3<411::AID-JCP14>3.0.CO;2-Z
- Wang, Y., Chan, D. W., Liu, V. W., Chiu, P., and Ngan, H. Y. (2010). Differential functions of growth factor receptor-bound protein 7 (GRB7) and its variant GRB7v in ovarian carcinogenesis. *Clin. Cancer Res.* 16, 2529–2539. doi: 10.1158/1078-0432.CCR-10-0018
- Watson, G. M., Gunzburg, M. J., Ambaye, N. D., Lucas, W. A., Traore, D. A., Kulkarni, K., et al. (2015). Cyclic peptides incorporating phosphotyrosine mimetics as potent and specific inhibitors of the Grb7 breast cancer target. *J. Med. Chem.* 58, 7707–7718. doi: 10.1021/acs.jmedchem.5b00609
- Watson, G. M., Kulkarni, K., Brandt, R., Del Borgo, M. P., Aguilar, M.-I., and Wilce, J. A. (2017). Shortened penetratin cell-penetrating peptide is insufficient for cytosolic delivery of a Grb7 targeting peptide. *ACS Omega* 2, 670–677. doi: 10.1021/acsomega.6b00561

Conflict of Interest Statement: 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 © 2017 Watson, Lucas, Gunzburg and Wilce. 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) or licensor 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.



Structure and Interactions of the TPR Domain of Sgt2 with Yeast Chaperones and Ybr137wp

Ewelina M. Kryzstofinska¹, Nicola J. Evans¹, Arjun Thapaliya¹, James W. Murray², Rhodri M. L. Morgan², Santiago Martinez-Lumbreras¹ and Rivka L. Isaacson^{1*}

¹ Department of Chemistry, King's College London, London, United Kingdom, ² Department of Life Sciences, Imperial College London, South Kensington, United Kingdom

OPEN ACCESS

Edited by:

Piero Andrea Temussi,
University of Naples Federico II, Italy

Reviewed by:

Filippo Prisci,
University of Essex, United Kingdom

Doriano Lamba,
Consiglio Nazionale Delle Ricerche
(CNR), Italy

*Correspondence:

Rivka L. Isaacson
rivka.isaacson@kcl.ac.uk

Specialty section:

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

Received: 28 July 2017

Accepted: 21 September 2017

Published: 11 October 2017

Citation:

Kryzstofinska EM, Evans NJ, Thapaliya A, Murray JW, Morgan RML, Martinez-Lumbreras S and Isaacson RL (2017) Structure and Interactions of the TPR Domain of Sgt2 with Yeast Chaperones and Ybr137wp. *Front. Mol. Biosci.* 4:68. doi: 10.3389/fmolb.2017.00068

Small glutamine-rich tetratricopeptide repeat-containing protein 2 (Sgt2) is a multi-module co-chaperone involved in several protein quality control pathways. The TPR domain of Sgt2 and several other proteins, including SGTA, Hop, and CHIP, is a highly conserved motif known to form transient complexes with molecular chaperones such as Hsp70 and Hsp90. In this work, we present the first high resolution crystal structures of Sgt2_TPR alone and in complex with a C-terminal peptide PTVEEVD from heat shock protein, Ssa1. Using nuclear magnetic resonance spectroscopy and isothermal titration calorimetry, we demonstrate that Sgt2_TPR interacts with peptides corresponding to the C-termini of Ssa1, Hsc82, and Ybr137wp with similar binding modes and affinities.

Keywords: Sgt2, TPR, carboxylate clamp, NMR, CSP, x-ray crystallography, ITC

INTRODUCTION

Transient interactions between proteins confer functional versatility upon a range of cellular processes, including protein modification, transport, folding, and cell signaling pathways (Perkins et al., 2010). The co-chaperone, Small glutamine-rich tetratricopeptide repeat (TPR) protein alpha (SGTA), is involved in the decision to target various misfolded and mislocalized proteins into their appropriate pathways, upstream of either insertion to the endoplasmic reticulum (ER) or degradation (Hessa et al., 2011; Leznicki and High, 2012; Wunderley et al., 2014; Casson et al., 2016; Shao et al., 2017). SGTA interacts with many proteins and forms transient complexes with chaperones, membrane targeting proteins, and members of the ubiquitin/proteasome system (UPS; Rodrigo-Brenni et al., 2014; Leznicki et al., 2015; Kryzstofinska et al., 2016; Thapaliya et al., 2016).

The yeast ortholog of SGTA, Sgt2, is best understood in the context of post-translational insertion of tail-anchored (TA) proteins into the ER membrane. The majority of membrane proteins undergo targeting to the endoplasmic reticulum in a co-translational process mediated by the signal recognition particle (SRP) as the nascent peptide chain emerges from the ribosomal tunnel. However, TAs are a special case of membrane proteins with obscured targeting signals at the extreme C-terminus. Therefore, their membrane delivery occurs post-translationally via the Guided Entry of Tail-anchored proteins pathway (GET; Schuldiner et al., 2008; Rabu et al., 2009; Borgese and Fasana, 2011; Hegde and Keenan, 2011). Sgt2 captures TA substrates after they are released from the ribosome and passes them on to the Get3 ATPase, the central targeting complex, in a process mediated by the Get4/Get5 heterodimeric scaffolding complex (Chartron et al., 2010; Wang et al., 2010; Simon et al., 2013; Mateja et al., 2015). This is followed by subsequent TA-protein release at the ER membrane, assisted by the Get1/Get2 heterodimeric membrane receptor complex (Wang et al., 2010; Mariappan et al., 2011; Vilardi et al., 2014). A new pathway in yeast has recently

been discovered which is suggested to be the back-up mechanism in the event of GET system failure (Aviram et al., 2016). This involves three proteins, named Snd1, Snd2, and Snd3 (for SRP-independent targeting), which have possible roles in targeting substrates to the ER translocation machinery Sec61 (Aviram et al., 2016).

Importantly for its role, Sgt2 associates with several heat-shock proteins such as Hsp104, Hsc82 (yeast ortholog of Hsp90), and Ssa1/Ssa2 (yeast orthologs of Hsp70), which can bind directly to its central TPR domain (Liou and Wang, 2005). The Hsp70 and Hsp90 chaperones are important parts of the cellular machinery for protein folding, maturation and structural stability (Richter and Buchner, 2001; Pratt and Toft, 2003). They often associate with co-chaperones containing multiple copies of TPR domains (Frydman and Hohfeld, 1997; Pratt, 1997) which help them to facilitate correct folding of client proteins (Wang et al., 2010; Morgan et al., 2015). It has also been proposed that SGTA regulates the ATPase activity and folding rates of Hsp70 (Angeletti et al., 2002).

Recently, Sgt2 was reported to interact with Ybr137wp, a protein of uncharacterized function that is specific to yeast (Yeh et al., 2014). The Sgt2 TPR domain binds to the C-terminal end of Ybr137. Ybr137wp is thought to be a decamer both in its crystal form and in solution (Yeh et al., 2014). The function of Ybr137 is linked to the GET pathway where it is able to rescue the TA protein delivery-defect caused by a GET system that is impaired under starvation conditions (Yeh et al., 2014). However, the exact role for Ybr137wp in the TA targeting mechanism is not understood.

Sgt2 contains an N-terminal dimerization domain (Liou and Wang, 2005; Simon et al., 2013; Tung et al., 2013), followed by the conserved, central TPR domain and a glutamine rich region toward the C-terminus. The N-terminal domain of Sgt2 can directly bind the ubiquitin-like (UBL) domain of Get5 and facilitate the handover of TA substrates downstream onto GET pathway components for membrane delivery (Chartron et al., 2012; Simon et al., 2013; Darby et al., 2014). The C-terminal domain is predicted to be flexible based on SAXS experiments and, is structurally uncharacterized (Chartron et al., 2012). These domains of both Sgt2 and SGTA bind hydrophobic substrates including the TMDs of TA-proteins (Dutta and Tan, 2008; Wang F. et al., 2011; Leznicki et al., 2013). TPR domains typically consist of three or more tandem repeats of a loosely conserved 34 residue motif (Lamb et al., 1995; Smith, 2004). Each tandem motif is formed of two anti-parallel α -helices. TPR domains are well-known for mediating protein-protein interactions (Das et al., 1998). The structure of the human SGTA TPR domain was determined previously by X-ray crystallography (Dutta and Tan, 2008) and, like the yeast ortholog, has been reported to interact directly with Hsp70/Hsp90 chaperones, the proteasomal subunit Rpn13 and a variety of disease-related proteins (Buchanan et al., 2007; Dutta and Tan, 2008; Roberts et al., 2015; Thapaliya et al., 2016). Moreover, the TPR domain structure (including some additional linker residues at the C-terminal) of the Sgt2 homolog from *Aspergillus fumigatus* was solved by crystallography (Chartron et al., 2011).

Several TPR domains of various proteins are known to interact with Hsp70/Hsp90 peptides. The high-resolution structures of such complexes were reported for HOP TPR1 with an Hsp70-derived peptide, HOP TPR2A with an Hsp90-derived peptide (Scheufler et al., 2000) and CHIP TPR with an Hsp70 C-terminal peptide (Zhang et al., 2005; Wang Q. et al., 2011). The common mode of interaction involves the formation of a carboxylate clamp where both the side-chain and main-chain terminal carboxylate groups of the C-terminal aspartic acids of these peptides form salt bridges with conserved arginine residues within the groove of the cochaperone TPR domains. Currently there is no structure of the Sgt2 TPR domain from *Saccharomyces cerevisiae*.

In this study, we report the first high resolution X-ray structure of the Sgt2_TPR at 1.55 Å and also in complex with the PTVEEVD peptide corresponding to the C-terminus of Ssa1, at 2.0 Å. In addition, we have characterized the interaction between Sgt2_TPR and C-terminal protein fragments of Ssa1/Ssa2 and Hsp82 (Hsp70 and Hsp90 in mammals respectively) and Ybr137wp using isothermal titration calorimetry (ITC) and Nuclear magnetic resonance (NMR).

RESULTS

Overall Sgt2_TPR Structure

The structure of Sgt2_TPR was determined by molecular replacement and refined to 1.55 Å resolution (Figure 1A, PDB: 5LYP; data collection and refinement parameters in Table 1). The coordinates of SGTA_TPR (PDB: 2VYI) were used as a search model since there is high structural homology with Sgt2_TPR (57% sequence identity; Figure 1B).

All residues were built into the electron density map except for the C-terminal Val229 and the solvent-exposed sidechains of Glu93 and Asp94. The final model also contains 90 water molecules and a single BO_4 ion from the crystallization condition. The TPR domain of Sgt2 consists of three TPR repeats, comprising six almost identical α -helices and a C-terminal “capping” helix (α_1 = A96-N115; α_2 = Y118-V131; α_3 = A136-L149; α_4 = Y152-I165; α_5 = F170-183Q; α_6 = P186-E200; α_7 = E206-L225) connected by short loops and arranged in a antiparallel fold homologous to that of SGTA_TPR. A structural overlay with the equivalent human domain is shown in Figure 1A (RMSD of 1.13 Å over 135 C α).

Complex Structure of Sgt2_TPR with the C-Terminal Peptide of Ssa1

Initially it was not possible to form a crystal complex of Sgt2_TPR (93–229) with the C-terminal of Ssa1 due to the flexible N-terminal and C-terminal ends of symmetry-related molecules in the crystal occluding the binding interface. However, producing a shorter Sgt2_TPR (96–225) construct by removing three residues (EDD) from the N-terminus and four (EKTV) from the C-terminus resulted in successful crystallization of the Sgt2_TPR/PTVEEVD complex (Figures 2A–C, PDB: 5LYN, statistics in Table 2).

Two copies of the TPR domain were present in the asymmetric unit due to non-crystallographic symmetry (Figure S1A). All

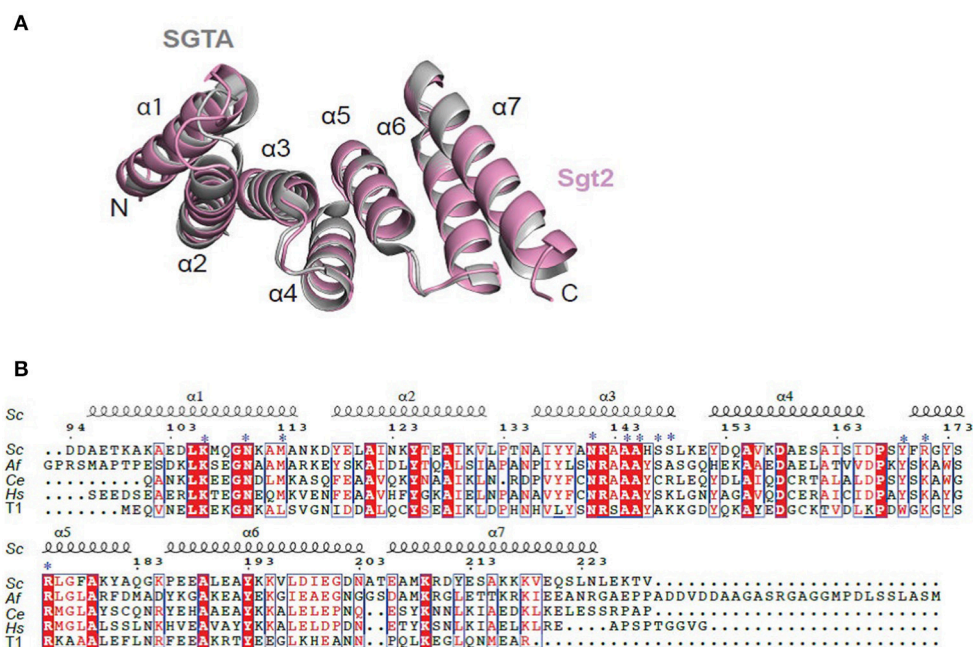


FIGURE 1 | Crystal structure of Sgt2_TPR. **(A)** Superimposition of TPR domains from Sgt2 (magenta, PDB: 5LYN) with human SGTA (gray, PDB: 2VYI). Structures were superposed using secondary-structure matching in ccp4mg (McNicholas et al., 2011). The structures align with RMSD of 1.13 Å over 135 C α . **(B)** Structure-based sequence alignment of SGT TPR domains from *S. cerevisiae* (Sc), *A. fumigatus* (Af), *C. elegans* (Ce), *Homo sapiens* (Hs), and human HOP TPR1 (T1). The residues involved in interaction and forming the two-carboxylate clamp are indicated with asterisks; boxes show conserved residues while red highlights sequence identity, structural motifs are labeled on the top. The residue numbering is from *S. cerevisiae*. Figure generated using ESPript 3.0 server.

Sgt2_TPR residues could be built into electron density maps in chain A and chain B. The two peptides (chains C and D) were partially occupied and were built with care. They were modeled into electron density (Figure 2A) and then verified by producing a simulated annealing omit map (Figure S2). The overall structures of the two chains, A and B, are very similar in their backbones (RMSD of 0.77 Å over 135 C α), with deviations observed for the R171 sidechains possibly due to their flexibility and significant differences in the modeled C and D peptides (Figures S1A,B). The electron density was ambiguous for peptide C at chain B, especially for Pro1 with some electron density appearing in the Fo-Fc map which could be suggesting the presence of another atom. However, neither a zinc ion nor water molecule could be fitted. Nevertheless, all PTVVEVD (1–7) residues were successfully modeled. The occupancy for the peptide chains (C and D) was refined to an Rfactor of 0.158 and an Rfree of 0.202, and both converged to occupancy 0.93. The final model also contained 148 water molecules, nine Zinc ions and a single BO $_4$ ion, which was present in the M9 medium we used for protein expression. Zinc ions were added at peaks of the phased anomalous difference map (DANO).

The interaction between the Sgt2_TPR and the PTVVEVD peptide from Ssa1 is mostly driven by the formation of a two-carboxylate clamp. Most of the electrostatic interactions between the TPR domain and the peptide occur in the C-terminal EEVD region and anchor the peptide in place. Peptide chains C and D, whilst overlapping, show slight conformational differences with an RMSD of 1.24 (Figure 2B). PDBePISA

highlights this difference showing a binding surface area of 474.7 Å 2 between chain B and C, and 514.3 Å 2 between chain A and D, and a difference in solvation energy of binding of -0.8 ΔiG kcal/mol and -3.7 ΔiG kcal/mol respectively. In the interface between chain A and D direct backbone contacts involve hydrogen bond formation between the carboxamide sidechains of Asn141 and Asn110 in Sgt2_TPR, and the sidechain and backbone of the terminal Asp7 of the Ssa1 peptide. Moreover, the sidechain amine of Lys 106 binds to the same carbonyl sidechain of Asp7. The guanidinium group of Arg171 forms a salt-bridge with the carbonyl main chain of Asp7 and forms an additional internal contact with Tyr169. The Arg175 sidechain interacts with the carbonyl main chain of Glu4 and Glu5 of the peptide (Figure 2C). In addition, the N-terminal of the peptide is involved in hydrophobic and van der Waals interactions. Phe178 and Tyr181 contribute to creating hydrophobic pockets and interact with the aliphatic part of the Pro1, Thr2, and Val3 residues. Met113 makes a hydrophobic contact with Val6 of the peptide (Figure S3). The “two-carboxylate clamp” binding mode is characteristic for TPR domains interacting with the conserved C-terminal IEEVD and MEEVD motifs of Hsp70 and Hsp90 chaperones, respectively (Scheufler et al., 2000; Zhang et al., 2005). Comparing the Sgt2_TPR/PTVEVD binding interface with a previously published complex of HOP TPR1/GPTIEVD (Hsp70-derivative; Scheufler et al., 2000) shows that the PTVVEVD peptide occupies the same position at the Sgt2 TPR groove as GPTIEVD. The peptides overlap apart from a difference in the conformation of the main chain

TABLE 1 | Data collection and refinement statistics of Sgt2_TPR.

| | | | |
|--------------------------------|---|------------------------------|----------------------|
| Resolution range | 32.31–1.55 (1.605–1.55) | CC (work) | 0.939 (0.928) |
| Space group | P 21 21 21 | CC (free) | 0.921 (0.863) |
| Unit cell | a: 36.86 Å, b: 50.76 Å c: 67.12 Å α : 90°, β : 90°, γ : 90° | Number of non-hydrogen atoms | 1,157 |
| Total reflections | 232,024 (23272) | Macromolecules | 1,062 |
| Unique reflections | 18,908 (1864) | Ligands | 5 |
| Multiplicity | 12.3 (12.5) | Protein residues | 137 |
| Completeness (%) | 1.00 (1.00) | RMS (bonds) | 0.005 |
| Mean I/sigma(I) | 12.89 (8.72) | RMS (angles) | 0.62 |
| Wilson B-factor | 8.98 Å ² | Ramachandran favored (%) | 100 |
| R-merge | 0.2183 (1.061) | Ramachandran allowed (%) | 0 |
| R-meas | 0.2277 (1.104) | Ramachandran outliers (%) | 0 |
| CC1/2 | 0.995 (0.975) | Rotamer outliers (%) | 0.95 |
| CC* | 0.999 (0.994) | Clashscore | 0.95 |
| Reflections used in refinement | 18,908 (1,864) | Average B-factor | 10.96 Å ² |
| Reflections used for R-free | 937 (89) | Macromolecules | 10.50 |
| R-work | 0.1936 (0.1711) | Ligands | 19.64 |
| R-free | 0.2186 (0.2231) | Solvent | 15.95 |

Statistics for the highest-resolution shell are shown in parentheses.

CC* is a derived quantity that links data and model and estimates the correlation of an observed data set with the underlying true signal.

and sidechain of the terminal Asp7 of GPTIEVD (**Figure 2B**). The alternative Asp7 conformation however overlaps with the Asp7 conformation in our chain C peptide in the complex structure. Comparing the two copies of the TPR-bound peptide indicates that, in the interface between chains B and C, the main difference is in the backbone of the N-terminal of peptide C. This is manifested through small variations in the orientation of Pro1, Thr2 and Val3 sidechains, changing hydrophobic interactions (particularly between Val3 and Phe178 of the TPR), and new electrostatic interactions between Thr2 and Asp211. These differences are not driven by the binding interface of Sgt2_TPR, as the sidechains of A and B largely adopt the same conformation, except for Arg171 due to its inherently flexible sidechain (Najmanovich et al., 2000). Notably, the Arg171 backbone NH was missing in the HSQC spectrum, which is a common feature of flexible sidechains in intermediate exchange on NMR timescales. The change in orientation of chain B Arg171, creates a salt bridge with the sidechain of Glu4. In addition, slight conformational changes in chain C Asp7 facilitate formation of hydrogen bond between the hydroxyl group of Tyr169 and the Asp7 peptide sidechain and the interaction between the sidechains of Asn141 and Asn110 of Sgt2_TPR and the main chain of Asp7 (**Figure S1B**).

Interactions of the TPR Domain of Sgt2 with Yeast Chaperones and Ybr137wp

To understand the molecular details of the interactions between Sgt2_TPR and the C-terminal fragments of Ssa1 (PTVEEVD),

Hsp82 (MEEVD), and Ybr137wp (SLEEDLNLD) we also performed NMR and ITC experiments. We acquired a complete set of NMR triple resonance experiments using the longer construct of Sgt2_TPR (93–229) to facilitate the full backbone assignment (BMRB Accession Number: 27044). All residues were assigned except for Arg171. Reciprocal chemical shift perturbation (CSP) experiments were carried out by titrating the unlabeled peptides PTVEEVD, MEEVD and SLEEDLNLD into ¹⁵N-labeled Sgt2_TPR (93–229) up to a six-fold molar excess of PTVEEVD and MEEVD and a five-fold excess of SLEEDLNLD. The NMR backbone assignment of Sgt2_TPR allowed us to identify the residues involved in the interactions in all three titrations (**Figure S4**, **S5A**, **S6A**). The CSP analysis exhibited a similar pattern for all three peptide titrations (**Figure S7**) and showed binding in a fast exchange regimen, with a number of peaks shifting non-linearly suggesting the formation of an intermediate during the titration. We analyzed the Sgt2_TPR CSPs (**Figures 3A–C**) by applying a titration cut-off at 3 molar equivalents of the peptide for all peaks and then dividing them into two groups. The first group consisted of peak shifts between 0 and 3 molar equivalents and was named “first event,” and the second group, called “second event,” comprised peak shifts that occurred between 3 and 6 (or 5 in the case of SLEEDLNLD) molar equivalents (marked as red and black arrows respectively in **Figure 3B**, **Figure S5B**, **S6B**). The CSP of Sgt2_TPR/PTVEEVD binding for the “first event,” shown on our x-ray structure is shown in **Figure 3A** on the left, and the “second event” on the right. The most perturbed residues for the “first event” correspond to the residues at the binding interface in the x-ray complex structure and/or the neighboring residues.

We also characterized the binding between Sgt2_TPR (93–229) and PTVEEVD, MEEVD, and SLEEDLNLD peptides by ITC (**Figure 3C**). The ITC results indicate a similar binding affinity for all three complexes with dissociation constants (K_d) of $9.04 \pm 0.05 \mu\text{M}$ for Sgt2_TPR/PTVEEVD, $2.95 \pm 0.30 \mu\text{M}$ for Sgt2_TPR/MEEVD and $1.53 \pm 0.05 \mu\text{M}$ for Sgt2_TPR/SLEEDLNLD. The favorable enthalpy and entropy values obtained from ITC suggest that all complex formations were driven by the establishment of both hydrogen bonds and hydrophobic interactions (**Figure 3C**).

DISCUSSION

In this study we provide high-resolution X-ray structures of the free Sgt2_TPR domain and its complex with the last seven amino acids of Ssa1 (Hsp70). We also assign the backbone of Sgt2_TPR using NMR spectroscopy and characterize the interaction between Sgt2_TPR and the extreme C-terminal fragments of Hsc82 (Hsp90) and Ybr137wp. Our structural data clearly show that Ssa1 binds to the TPR domain of Sgt2 via a carboxylate clamp mechanism and we can predict a similar mode of binding for Hsc82 and Ybr137wp from our consistent ITC and NMR data in all three systems. Analysis of the three protein-peptide complexes using the PDBePISA interactive tool indicates similarities. In all three complexes Glu4 and Glu5 are predicted to be involved in formation of hydrogen bonds and the terminal,

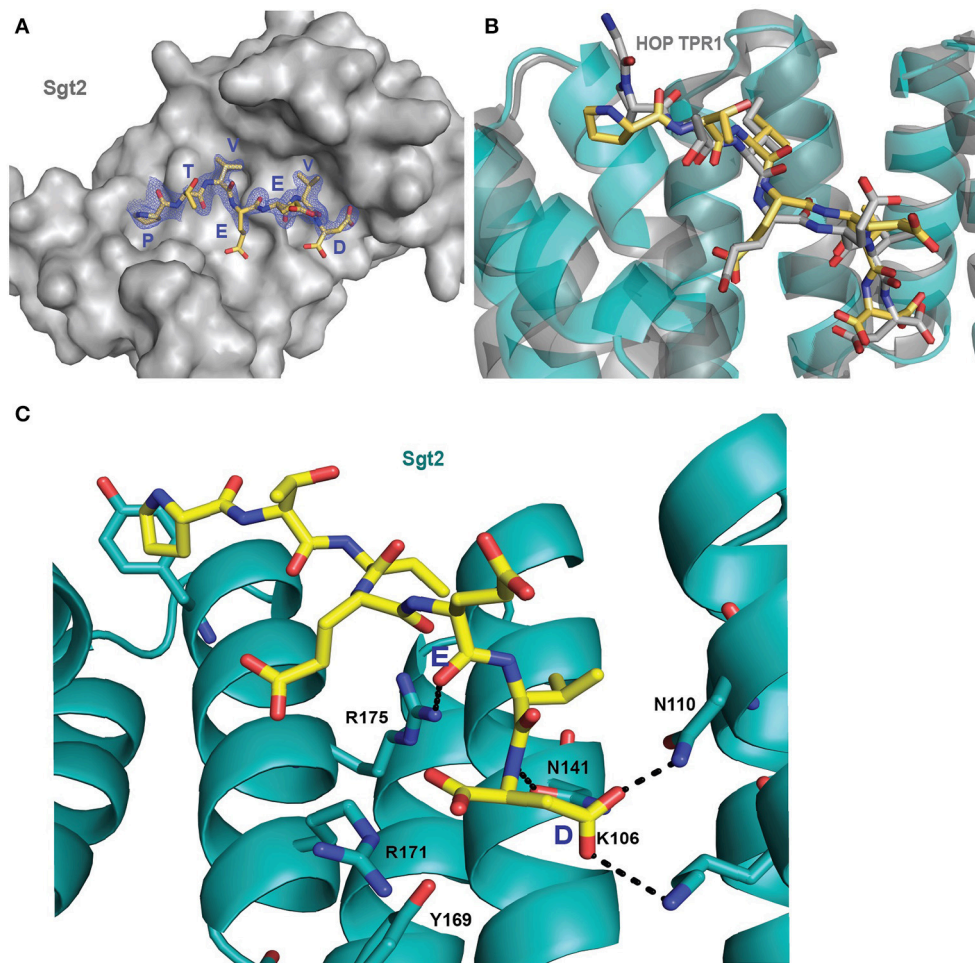


FIGURE 2 | Crystal structure of Sgt2_TPR/PTVEEVD complex. **(A)** The surface representation of Sgt2 hydrophobic groove with bound Ssa1 derived PTVEEVD peptide (PDB: 5LYP). The 2Fo-Fc map for the peptide was contoured at 1.0 σ . **(B)** Superimposition of Sgt2_TPR/PTVEEVD complex (peptide in yellow) onto Hsc70 peptide-bound HOP TPR1A (peptide in gray, PDB: 1ELW) highlighting similarities in peptide conformation at the binding interface. The peptides align with RMSD 0.52 Å. **(C)** Network of interactions formed at the complex interface (chain A and D). Residues shown as sticks are involved in the formation of hydrogen bonds or electrostatic interactions (shown as black dashed lines). Residues K106, N110, N141, R171, R175, and Y169 are involved in the formation of two-carboxylate clamp and M113 is involved in hydrophobic interactions.

Asp7, can form a salt bridge with Arg171, Arg175, and Lys106. In addition, in the case of the SLEEDLNLD peptide there is potential for Glu3 to also be involved in a hydrogen bond formation with Ser148. There is a binding surface area of 514.3 Å² between chains A and D, and a difference in solvation energy of binding of -3.7 Δ iG kcal/mol with PTVEEVD. In modeled examples of MEEVD and SLEEDLNLD there are binding surface areas of 420.8 and 579.5 Å², and differences in solvation energy of binding of -2.7 Δ iG kcal/mol and -2.2 Δ iG kcal/mol, respectively. There are many examples in the literature of carboxylate clamp mechanisms, most of them connecting the C-termini of heat shock proteins with different co-chaperones (Carrigan et al., 2004; Prasad et al., 2010; Panigrahi et al., 2014), but there are also non-chaperone examples which include the recognition of the proteasomal protein Rpn13 by SGTA (Thapaliya et al., 2016) and the interaction of Sgt2 with Ybr137wp presented here.

All carboxylate clamp interactions studied so far, including our recent and past investigations, describe dissociation constants in the low micromolar range by ITC or SPR (Scheufler et al., 2000; Brinker et al., 2002; Worrall et al., 2008; Thapaliya et al., 2016). In addition our NMR data suggest the presence of an intriguing dual-event binding mode during titrations and a widespread perturbation along the whole TPR domain. A detailed analysis of the titration revealed that in the first event only residues in the binding interface appear perturbed, while in the second event the perturbation is not localized to a specific interface. A similar scenario had previously been observed in the Rpn13 interaction with SGTA TPR, where signals all over the TPR were affected upon titration (Thapaliya et al., 2016). This behavior appears conserved for carboxylate clamp recognition whether synthetic peptides or recombinant proteins were used for the titration experiments. It likely

TABLE 2 | Data collection and refinement statistics of Sgt2_TPR complex.

| | | | |
|--------------------------------|---|------------------------------|----------------------|
| Resolution range | 33.59–2.0 (2.071–2.0) | CC (work) | 0.845 (0.345) |
| Space group | P 1 21 1 | CC (free) | 0.725 (0.476) |
| Unit cell | a: 45.49 Å, b: 61.09 Å c: 55.25 Å, α : 90°, β : 108.81°, γ : 90° | Number of non-hydrogen atoms | 2,340 |
| Total reflections | 117,547 (10,486) | Macromolecules | 2,183 |
| Unique reflections | 19189 (1,904) | Ligands | 9 |
| Multiplicity | 6.1 (5.6) | Protein residues | 280 |
| Completeness (%) | 0.99 (0.97) | RMS (bonds) | 0.006 |
| Mean I/sigma(I) | 6.49 (2.84) | RMS (angles) | 0.68 |
| Wilson B-factor | 25.49 Å ² | Ramachandran favored (%) | 99.63 |
| R-merge | 0.354 (0.8515) | Ramachandran allowed (%) | 0.37 |
| R-meas | 0.3851 (0.9392) | Ramachandran outliers (%) | 0 |
| CC1/2 | 0.92 (0.531) | Rotamer outliers (%) | 0 |
| CC* | 0.979 (0.833) | Clashscore | 2.77 |
| Reflections used in refinement | 19,279 (1903) | Average B-factor | 30.91 Å ² |
| Reflections used for R-free | 928 (83) | Macromolecules | 30.60 |
| R-work | 0.1576 (0.1903) | Ligands | 50.77 |
| R-free | 0.2025 (0.2587) | Solvent | 34.38 |

Statistics for the highest-resolution shell are shown in parentheses.

CC* is a derived quantity that links data and model and estimates the correlation of an observed data set with the underlying true signal.

relates to the fact that the TPR domain helices suffer a subtle contraction to enclose the peptide in the TPR groove. The crystallographic structures we have obtained clearly show a slightly more compact conformation of the TPR where helices 1 and 7 are closer to each other in the complex structure than in the unbound TPR (See Figure S8 for a structural alignment). This observation was previously reported for a longer TPR motif (Zeytuni et al., 2011) and proposed for the TPR domains of co-chaperones (Panigrahi et al., 2014), where it was suggested that changes in the curvature of the cradle structure by concerted movement of the helices may be necessary for ligand binding.

The orientation of Hsp peptides in TPR structures varies between proteins. The conserved clamp mode of interactions is consistent, but there are some differences observed for the N-terminal parts of the peptides, which is not surprising given that the carboxylate clamp is the fixed point of attachment. The PTVEEVD peptide adopts an extended conformation within the Sgt2_TPR groove similar to that observed in the structure of the HOP TPR1/GPTIEEVD complex (Scheufler et al., 2000). In contrast, structures of GPTIEEVD/Chip TPR [PDB: 3Q49 (Wang L. et al., 2011) and PDB: 4KBQ Zhang et al., 2005] show the peptide in a curled conformation lining the groove. The structures also vary in orientations of the peptide Pro1. In comparison, we also observe differences between the Pro1, Thr2, and Val3 sidechains in the two chains of our Sgt2_TPR/PTVEEVD complex structure allowing for some flexibility in the association between TPR and Hsp peptides at

the same interface. Sgt2_TPR serves as a binding interface for transient interactions with a variety of chaperones and other proteins. However, the extended conformation of the peptides and their position within the TPR groove allows for a widespread contact surface with TPR domains thus supporting the specific recognition of short amino acid stretches with sufficient affinity to bind (Figure S6). The preceding residues to EEVD are also important for the binding affinity and it has been reported that trimming the peptide sequence to EEVD only, reduced the affinity by at least 10 times (Scheufler et al., 2000). Furthermore, Sgt2 is a homodimer and it can target a broad range of substrates by binding more than one protein simultaneously and bringing them into closer proximity promoting interactions.

Little is known about the role of Ybr137wp in the GET pathway except that it binds to Sgt2 at the same binding interface as heat shock chaperones and that it can influence TA membrane insertion by mediating interactions between Sgt2 and chaperones. Previous ITC binding experiments reported that one full-length Ybr137w decamer is capable of binding to five Sgt2_TPR dimers with a dissociation constant (K_d) of $1.38 \pm 0.09 \mu\text{M}$ (Yeh et al., 2014). This is almost identical to the ITC results we obtained for the association of Sgt2_TPR with the extreme C-terminal nine-residue Ybr137wp-derived peptide (K_d of 1.53 ± 0.05), suggesting that the SLEEDLNLD fragment is sufficient for the interaction. Moreover, it has been shown that removing ESLEEDLNLD from the C-terminal of Ybr137wp abolished the interaction, confirming that this flexible C-terminal region is also necessary for the interaction (Yeh et al., 2014).

Further work is required to define the distinct role of Ybr137wp in ER delivery of tail-anchored membrane proteins and examine whether there is any link between this protein and the recently discovered SND targeting pathway in yeast. This alternative to the GET and SRP mechanisms, is proposed to act as a back-up targeting system (Aviram et al., 2016). It involves three proteins, localized at the ER or in the cytoplasm, Snd1 (encoded by YDR186C), Snd2 (encoded by ENV10, also known as YLR065C) and Snd3 (encoded by PHO88, also known as YBR106W), working together in a joint targeting pathway (Aviram et al., 2016). The function of Ybr137wp is also linked with altering the defect in TA protein delivery and cell viability derived from impairment of the GET system under starvation conditions.

Future investigations will improve our understanding of Ybr137wp function which will shed light on the importance of the carboxylate clamp interaction with Sgt2 that we delineate here.

METHODS

Plasmid Preparation

Gene fragments encoding the Sgt2_TPR (residues 93–229 and 96–225 for the shorter construct) from *S. cerevisiae* were PCR amplified from synthetic cDNA (Life Technologies) and cloned into the *Bam*HI/*Xho*I restriction sites of a home-modified pET28 vector which encodes an N-terminal thioredoxin A fusion protein followed by a hexahistidine tag and tobacco etch virus (TEV) protease cleavage site.

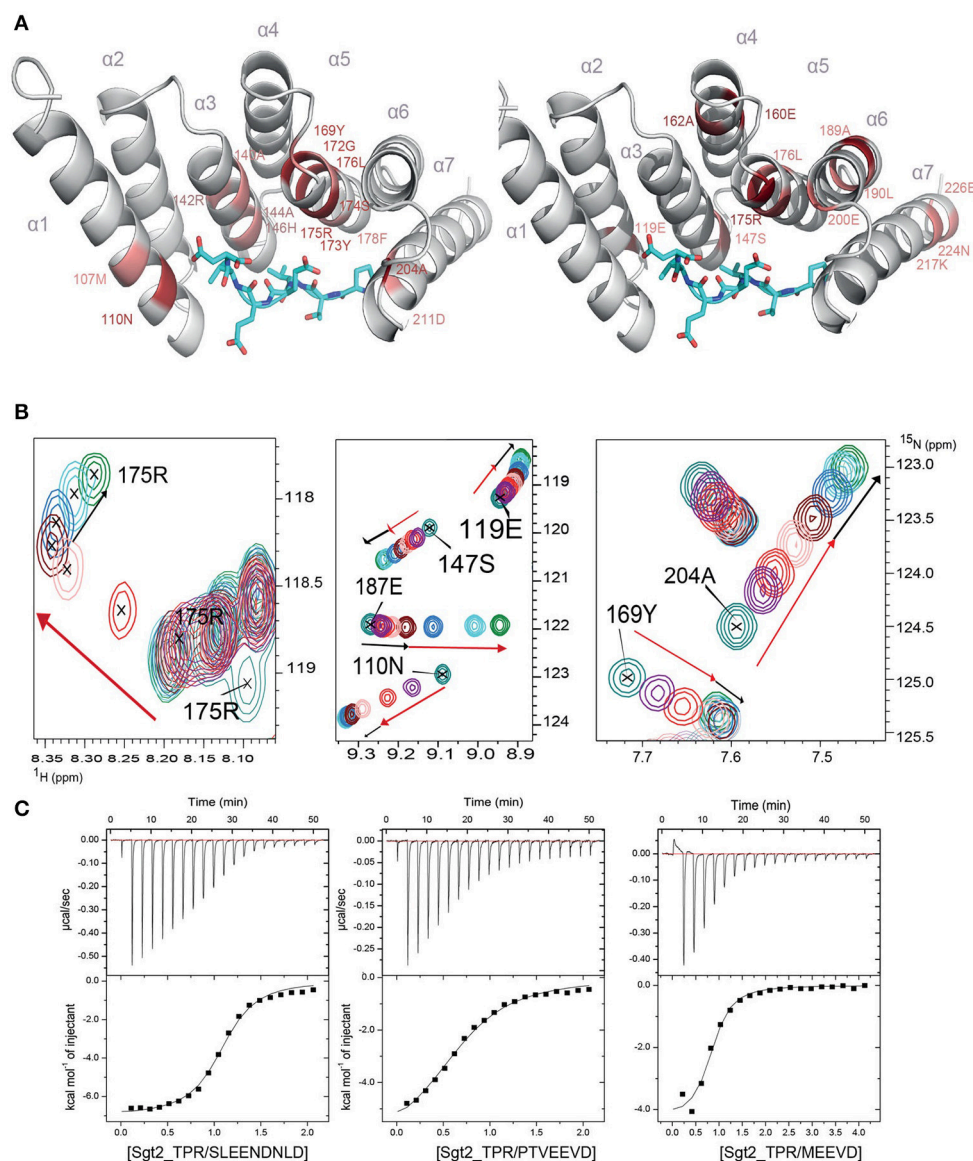


FIGURE 3 | Sgt2_TPR binds yeast chaperones and Ybr137wp. **(A)** Sgt2_TPR domain colored according to reciprocal chemical shift perturbation (CSP) upon additions of unlabeled PTVEEDV peptide. The model on the left shows the titration points between 0 and 1:3 protein/peptide molar ratio and on the right between 1:3 and 1:6 molar ratio. The most perturbed residues were selected at >0.4 ppm (red) and >0.3 ppm (light red) for the model on the left and >0.2 (red) and >0.15 (light red) for the model on the right. **(B)** A small regions of the ^1H - ^{15}N HSQC spectrum of ^{15}N -labeled Sgt2_TPR titrated with unlabeled PTVEEDV at ratios: 1:0 (teal), 1:0.5 (purple), 1:1 (red), 1:2 (light pink), 1:3 (maroon), 1:4 (blue), 1:5 (cyan), and 1:6 (green). Black and red arrows indicate the first and second events in CSPs upon titration respectively. **(C)** ITC data showing binding of Sgt2_TPR to SLEEDLNLD, PTVEEDV, and MEEVD. Binding parameters, determined by ITC were $K_d = 1.53 \pm 0.05 \mu\text{M}$, $\Delta H = -7.07 \pm 0.99 \text{ kcal/mol}$; $\Delta S = 3.19 \text{ kcal/mol-deg}$ for SLEEDLNLD; $K_d = 9.04 \pm 0.05 \mu\text{M}$, $\Delta H = -6.72 \pm 0.07 \text{ kcal/mol}$, $\Delta S = 140 \text{ cal/mol-deg}$ for PTVEEDV and $K_d = 2.95 \pm 0.31 \mu\text{M}$, $\Delta H = -4.51 \pm 0.03 \text{ kcal/mol}$; $\Delta S = 10.9 \text{ cal/mol-deg}$, for MEEVD.

Protein Production

All plasmids carrying Sgt2_TPR were transformed into *E. coli* BL21 (DE3) strain. Typically, protein expression was induced by adding 0.3–0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG) to cultures at $\text{OD}_{600} \approx 0.8$, followed by overnight incubation at 18°C . For ^{15}N -labeled proteins, growth was carried out in M9 media supplemented with labeled ammonium chloride ($>98\%$ ^{15}N , Sigma-Aldrich) and/or glucose ($>99\%$

U- ^{13}C , Sigma-Aldrich). Harvested cells were resuspended in lysis buffer [20 mM potassium phosphate, pH 8.0, 300 mM NaCl, 10 mM Imidazole, 250 μM tris(2-carboxyethyl)phosphine (TCEP)], supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF), and lysed by sonication or using a cell disruptor (Constant Systems Ltd). Cell debris and insoluble material were removed by centrifugation and overexpressed protein recovered from soluble fractions was purified using

nickel affinity chromatography (HisTrapTM HP 5 ml, GE Healthcare). Recombinant proteins were eluted with buffer containing 300 mM imidazole, then dialyzed against cleavage buffer (20 mM potassium phosphate, pH 8.0 and 300 mM NaCl) and digested with homemade TEV protease ($\approx 100 \mu\text{g/ml}$) at 4°C overnight. After TEV cleavage, a second nickel affinity chromatography step was performed to remove fusion protein, histidine tags, undigested protein, and TEV protease; the desired protein was then recovered in the flow through and loaded into a HiLoad 16/60 Superdex 75 column (GE Healthcare), previously equilibrated in buffer containing 10 mM potassium phosphate pH 6.0, 100 mM NaCl and 250 μM TCEP or 20 mM Tris-HCl pH 7.5. Proteins were concentrated using Vivaspinn concentrators with 5K cut-off (Sartorius Stedim) and sample purity and homogeneity was checked by SDS-PAGE, mass spectrometry and NMR. The lyophilized peptides: PTVVEVD (corresponding to Ssa1 C-terminal; residues 634–640), MEEVD (corresponding to C-terminal of Hsp82; residues 705–709) and SLEEDLNLD (corresponding to C-terminal of Ybr137wp; residues 171–179) were purchased from Alpha BioScience (Birmingham, UK) and resuspended in water or an appropriate buffer before use. All peptides were purified and verified by HPLC and mass spectrometry.

NMR Titrations

Sgt2_TPR (residues 93–229) and peptides used for NMR were prepared in 10 mM potassium phosphate pH 6.0, 100 mM NaCl and 250 μM TCEP buffer. Typically, ^1H - ^{15}N HSQC experiments were recorded for each titration point at 25°C and CSP calculated for each amide signal using the following formula, where $\Delta\delta_{1\text{H}}$ and $\Delta\delta_{15\text{N}}$ are the chemical shift differences for the same amide in its free and bound spectra ($\delta_{\text{free}} - \delta_{\text{bound}}$) and for proton and nitrogen values respectively:

$$\Delta\delta^{\text{av}} = \sqrt{((\Delta\delta_{1\text{H}})^2 + (\Delta\delta_{15\text{N}}/5)^2)} \cdot 0.5$$

CSP results were mapped onto the structures using the PyMOL software.

NMR Experiments

Protein samples at concentrations between 500 and 3,000 μM were prepared in 10% D₂O (Sigma Aldrich), 10 mM potassium phosphate pH 6.0, 100 mM NaCl and 250 μM TCEP buffer. All NMR experiments were acquired in 5 mm NMR tubes at 25°C on Bruker Avance spectrometers operating at 500 and 800 MHz equipped with cryoprobes, controlled by the TopSpin 3.1 software package. Backbone assignments were carried out using 3D experiments [HNCO, HNCA, HN(CA)CO, CBCA(CO)NH, and CBCANH] for Sgt2_TPR. All NMR spectra were processed with NMRPipe (Delaglio et al., 1995) and analyzed with CcpNMR Analysis (Vranken et al., 2005).

ITC

ITC measurements were performed at 25°C using an ITC-200 MicroCal microcalorimeter (GE Healthcare) following standard

procedures (Darby et al., 2014). Proteins were prepared in 10 mM potassium phosphate pH 6.0, 100 mM NaCl, 250 μM TCEP. In each titration, 20 injections of 2 μL of peptide solution at a concentration of 500 μM , were added to Sgt2_TPR (residues 93–229) at 50 μM in the reaction cell. Integrated heat data obtained for the titrations, corrected for heats of dilution, were fitted using a nonlinear least-squares minimization algorithm to a theoretical titration curve, using the MicroCal-Origin 7.0 software package. ΔH (reaction enthalpy change in Kcal/mol), K_b (equilibrium binding constant per mole), and n (molar ratio between the proteins in the complex) were the fitting parameters. The reaction entropy, ΔS , was calculated using the $\Delta G = -RT \cdot \ln K_b$ ($R = 8.314 \text{ J/(mol}\cdot\text{K)}$, $T = 298 \text{ K}$) and $\Delta G = \Delta H - T \Delta S$. Dissociation constants (K_d) are shown in the figure legends for each interaction.

Crystallization

Sgt2_TPR (residues 93–229) was concentrated to 35 mg/ml in 20 mM Tris-HCl pH 7.5 buffer and crystals were obtained after 4 days by the vapor-diffusion method at 293K using MRC plates in 0.1 M SPG, pH 6.0, 25% w/v PEG 1500 (PACT premier from Molecular Dimensions) at 20°C (drop volume = 400 nl). In the case of Sgt2_TPR (96–225)/PTVVEVD complex, protein/peptide complex was eluted from a HiLoad 16/60 Superdex 75 column and concentrated to 20 mg/ml in 20 mM Tris-HCl pH 7.5 followed by a further peptide addition (up to protein: peptide molar ratio of 1:3) prior to crystallization. The complex crystallized after 7 days by the vapor-diffusion method at 293 K in 0.2 M zinc Acetate, pH 7.2, 30% w/v PEG 3350. All crystals were harvested in reservoir solution with 20% glycerol before flash cooling in liquid nitrogen.

Data Collection and Processing

A complete dataset was collected from a single crystal on Diamond Beamline I04 for the free Sgt2_TPR dataset and I03 for the Sgt2_TPR/PTVVEVD complex using a Pilatus 6M-F detector and a single wavelength 0.920 Å. Data were processed using Xia2 (Winter et al., 2013) with scaling and merging using Aimless (McNicholas et al., 2011).

Structure Solution and Refinement

The crystal structure of Sgt2_TPR (93–229) was determined by molecular replacement using Phaser (McCoy et al., 2007) with the human SGTA_TPR crystal structure (PDB: 2VYI) used as a search model (57% sequence identity). This structure was then used as the search model to solve the Sgt2_TPR (96–225)/PTVVEVD complex. Both structures were refined using REFMAC5 (Murshudov et al., 2011) and PHENIX (Adams et al., 2010) with manual model building using Coot (Emsley and Cowtan, 2004). Free R-value of 4.9% was used as a cross-validation method for Sgt2_TPR and 4.8% for Sgt2_TPR (96–225)/PTVVEVD. Water molecules, Zn atoms and BO₄ atoms were fitted manually using Coot. The free Sgt2_TPR structure was solved in space group P 21 21 21 with cell parameters: 36.86 Å (a) 50.76 Å (b) 67.12 Å (c) 90.00° (α) 90.00° (β) 90.00° (γ). The refined structure shows very good stereochemistry

(statistics from the MolProbity (Chen et al., 2010, 2015) report are shown in **Table 1**). The complex structure was solved in space group P21 with cell dimensions: 45.49 Å (a) 61.09 Å (b) 55.25 Å (c) 90.00° (α) 108.81° (β) 90.00° (γ) and statistics from the MolProbity (Chen et al., 2010, 2015) report are shown in **Table 2**).

AUTHOR CONTRIBUTIONS

EK, NE, AT, and RI conceived the ideas and designed experiments. EK, NE, and AT performed experiments. EK, NE, AT, JM, RM, SM, and RI analyzed data. All authors contributed toward writing the manuscript.

FUNDING

RI was supported by MRC New Investigator Research Grant: G0900936. RI is funded by BBSRC grants: BB/L006952/1 and BB/N006267/1. AT is funded by BBSRC grant: BB/J014567/1. NMR experiments were performed at the Centre for Biomolecular Spectroscopy, King's College London, established with a Capital Award from the Wellcome Trust. This work was supported by the Francis Crick Institute through provision of access to the MRC Biomedical NMR Centre. The Francis Crick Institute receives its core funding from Cancer Research UK

REFERENCES

- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sec. D Biol. Crystallogr.* 66(Pt 2), 213–221. doi: 10.1107/S0907444909052925
- Angeletti, P. C., Walker, D., and Panganiban, A. T. (2002). Small glutamine-rich protein/viral protein U-binding protein is a novel cochaperone that affects heat shock protein 70 activity. *Cell Stress Chaperones* 7, 258–268. doi: 10.1379/1466-1268(2002)007<0258:SGRPVP>2.0.CO;2
- Aviram, N., Ast, T., Costa, E. A., Arakel, E. C., Chuartzman, S. G., Jan, C. H., et al. (2016). The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. *Nature* 540, 134–138. doi: 10.1038/nature20169
- Borgese, N., and Fasana, E. (2011). Targeting pathways of C-tail-anchored proteins. *Biochim. Biophys. Acta* 1808, 937–946. doi: 10.1016/j.bbame.2010.07.010
- Brinker, A., Scheufler, C., Von Der Mulbe, F., Fleckenstein, B., Herrmann, C., Jung, G., et al. (2002). Ligand discrimination by TPR domains. Relevance and selectivity of EEVD-recognition in Hsp70 x Hop x Hsp90 complexes. *J. Biol. Chem.* 277, 19265–19275. doi: 10.1074/jbc.M109002200
- Buchanan, G., Ricciardelli, C., Harris, J. M., Prescott, J., Yu, Z. C., Jia, L., et al. (2007). Control of androgen receptor signaling in prostate cancer by the cochaperone small glutamine rich tetratricopeptide repeat containing protein α. *Cancer Res.* 67, 10087–10096. doi: 10.1158/0008-5472.CAN-07-1646
- Carrigan, P. E., Nelson, G. M., Roberts, P. J., Stoffer, J., Riggs, D. L., and Smith, D. F. (2004). Multiple domains of the co-chaperone Hop are important for Hsp70 binding. *J. Biol. Chem.* 279, 16185–16193. doi: 10.1074/jbc.M314130200
- Casson, J., McKenna, M., and High, S. (2016). On the road to nowhere: cross-talk between post-translational protein targeting and cytosolic quality control. *Biochem. Soc. Trans.* 44, 796–801. doi: 10.1042/BST20160045
- Charton, J. W., Gonzalez, G. M., and Clemons, W. M. Jr. (2011). A structural model of the Sgt2 protein and its interactions with chaperones and the Get4/Get5 complex. *J. Biol. Chem.* 286, 34325–34334. doi: 10.1074/jbc.M111.277798
- Charton, J. W., Suloway, C. J., Zaslaver, M., and Clemons, W. M. Jr. (2010). Structural characterization of the Get4/Get5 complex and its interaction with Get3. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12127–12132. doi: 10.1073/pnas.1006036107
- Charton, J. W., VanderVelde, D. G., and Clemons, W. M. Jr. (2012). Structures of the Sgt2/SGTA dimerization domain with the Get5/UBL4A UBL domain reveal an interaction that forms a conserved dynamic interface. *Cell Rep.* 2, 1620–1632. doi: 10.1016/j.celrep.2012.10.010
- Chen, V. B., Arendall, W. B., 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. Sec. D Biol. Crystallogr.* 66(Pt 1), 12–21. doi: 10.1107/S0907444909042073
- Chen, V. B., Wedell, J. R., Wenger, R. K., Ulrich, E. L., and Markley, J. L. (2015). MolProbity for the masses-of data. *J. Biomol. NMR* 63, 77–83. doi: 10.1007/s10858-015-9969-9
- Darby, J. F., Kryzstofinska, E. M., Simpson, P. J., Simon, A. C., Leznicki, P., Srisankarajah, N., et al. (2014). Solution structure of the SGTA dimerisation domain and investigation of its interactions with the ubiquitin-like domains of BAG6 and UBL4A. *PLoS ONE* 9:e113281. doi: 10.1371/journal.pone.0113281
- Das, A. K., Cohen, P. W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J.* 17, 1192–1199. doi: 10.1093/emboj/17.5.1192
- 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
- Dutta, S., and Tan, Y. J. (2008). Structural and functional characterization of human SGT and its interaction with Vpu of the human immunodeficiency virus type 1. *Biochemistry* 47, 10123–10131. doi: 10.1021/bi800758a
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sec. D Biol. Crystallogr.* 60(12 Pt 1), 2126–2132. doi: 10.1107/S0907444904019158
- Frydman, J., and Hohfeld, J. (1997). Chaperones get in touch: the Hip-Hop connection. *Trends Biochem. Sci.* 22, 87–92. doi: 10.1016/S0968-0004(97)01005-0
- (FC001029), the UK Medical Research Council (FC001029), and the Wellcome Trust (FC001029).

ACKNOWLEDGMENTS

The authors would like to thank Diamond Light Source for beamtime (proposals mx12579 and mx13597), and the staff of beamlines I03 and I04 for assistance with data collection. The dataset for the free Sgt2_TPR was collected at I04 as part of proposal number mx12579 for the Imperial College London CSB BAG and datasets for Sgt2_TPR/PTVEEVD complex were collected at I03 as part of proposal number mx13597 for King's College University BAG. We thank Dr. R. A. Atkinson (KCL) and Dr. G. Kelly (The Crick Institute) for his assistance with NMR experiments, and Dr. R. Yan (KCL) for insightful scientific discussions. We are grateful to Dr. J. M. Pérez-Cañadillas (Rocasolano Chemical Physical Institute, Madrid, Spain) for providing the home-modified pET28 vector and plasmid encoding TEV protease.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2017.00068/full#supplementary-material>

- Hegde, R. S., and Keenan, R. J. (2011). Tail-anchored membrane protein insertion into the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* 12, 787–798. doi: 10.1038/nrm3226
- Hessa, T., Sharma, A., Mariappan, M., Eshleman, H. D., Gutierrez, E., and Hegde, R. S. (2011). Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature* 475, 394–397. doi: 10.1038/nature10181
- Krysztofinska, E. M., Martinez-Lumbreras, S., Thapaliya, A., Evans, N. J., High, S., and Isaacson, R. L. (2016). Structural and functional insights into the E3 ligase, RNF126. *Sci. Rep.* 6:26433. doi: 10.1038/srep26433
- Lamb, J. R., Tugendreich, S., and Hieter, P. (1995). Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends Biochem. Sci.* 20, 257–259.
- Leznicki, P., and High, S. (2012). SGTA antagonizes BAG6-mediated protein triage. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19214–19219. doi: 10.1073/pnas.1209997109
- Leznicki, P., Korac-Prljic, J., Kliza, K., Husnjak, K., Nyathi, Y., Dikic, I., et al. (2015). Binding of SGTA to Rpn13 selectively modulates protein quality control. *J. Cell Sci.* 128, 3187–3196. doi: 10.1242/jcs.165209
- Leznicki, P., Roebuck, Q. P., Wunderley, L., Clancy, A., Krysztofinska, E. M., Isaacson, R. L., et al. (2013). The association of BAG6 with SGTA and tail-anchored proteins. *PLoS ONE* 8:e59590. doi: 10.1371/journal.pone.0059590
- Liou, S. T., and Wang, C. (2005). Small glutamine-rich tetratricopeptide repeat-containing protein is composed of three structural units with distinct functions. *Arch. Biochem. Biophys.* 435, 253–263. doi: 10.1016/j.abb.2004.12.020
- Mariappan, M., Mateja, A., Dobosz, M., Bove, E., Hegde, R. S., and Keenan, R. J. (2011). The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature* 477, 61–66. doi: 10.1038/nature10362
- Mateja, A., Paduch, M., Chang, H. Y., Szydłowska, A., Kossiakoff, A. A., Hegde, R. S., et al. (2015). Protein targeting. Structure of the Get3 targeting factor in complex with its membrane protein cargo. *Science* 347, 1152–1155. doi: 10.1126/science.1261671
- 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. Crystallogr.* 40(Pt 4), 658–674. doi: 10.1107/S0021889807021206
- McNicholas, S., Potterton, E., Wilson, K. S., and Noble, M. E. (2011). Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. Sec. D Biol. Crystallogr.* 67(Pt 4), 386–394. doi: 10.1107/S0907444911007281
- Morgan, R. M., Pal, M., Roe, S. M., Pearl, L. H., and Prodromou, C. (2015). Tah1 helix-swap dimerization prevents mixed Hsp90 co-chaperone complexes. *Acta Crystallogr. Sec. D Biol. Crystallogr.* 71(Pt 5), 1197–1206. doi: 10.1107/S1399004715004551
- Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., et al. (2011). REFMACS for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sec. D Biol. Crystallogr.* 67(Pt 4), 355–367. doi: 10.1107/S0907444911001314
- Najmanovich, R., Kuttner, J., Sobolev, V., and Edelman, M. (2000). Side-chain flexibility in proteins upon ligand binding. *Proteins* 39, 261–268. doi: 10.1002/(SICI)1097-0134(20000515)39:3<261::AID-PROT90>3.0.CO;2-4
- Panigrahi, R., Whelan, J., and Vrielink, A. (2014). Exploring ligand recognition, selectivity and dynamics of TPR domains of chloroplast Toc64 and mitochondria Om64 from *Arabidopsis thaliana*. *J. Mol. Recognit.* 27, 402–414. doi: 10.1002/jmr.2360
- Perkins, J. R., Diboun, I., Dessailly, B. H., Lees, J. G., and Orengo, C. (2010). Transient protein-protein interactions: structural, functional, and network properties. *Structure* 18, 1233–1243. doi: 10.1016/j.str.2010.08.007
- Prasad, B. D., Goel, S., and Krishna, P. (2010). *In silico* identification of carboxylate clamp type tetratricopeptide repeat proteins in Arabidopsis and rice as putative co-chaperones of Hsp90/Hsp70. *PLoS ONE* 5:e12761. doi: 10.1371/journal.pone.0012761
- Pratt, W. B. (1997). The role of the hsp90-based chaperone system in signal transduction by nuclear receptors and receptors signaling via MAP kinase. *Annu. Rev. Pharmacol. Toxicol.* 37, 297–326.
- Pratt, W. B., and Toft, D. O. (2003). Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Exp. Biol. Med.* 228, 111–133. doi: 10.1177/153537020322800201
- Rabu, C., Schmid, V., Schwappach, B., and High, S. (2009). Biogenesis of tail-anchored proteins: the beginning for the end? *J. Cell Sci.* 122(Pt 20), 3605–3612. doi: 10.1242/jcs.041210
- Richter, K., and Buchner, J. (2001). Hsp90: chaperoning signal transduction. *J. Cell. Physiol.* 188, 281–290. doi: 10.1002/jcp.1131
- Roberts, J. D., Thapaliya, A., Martinez-Lumbreras, S., Krysztofinska, E. M., and Isaacson, R. L. (2015). Structural and functional insights into small, glutamine-rich, tetratricopeptide repeat protein alpha. *Front. Mol. Biosci.* 2:71. doi: 10.3389/fmolb.2015.00071
- Rodrigo-Brenni, M. C., Gutierrez, E., and Hegde, R. S. (2014). Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. *Mol. Cell* 55, 227–237. doi: 10.1016/j.molcel.2014.05.025
- Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., et al. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell* 101, 199–210. doi: 10.1016/S0092-8674(00)80830-2
- Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H. D., et al. (2008). The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell* 134, 634–645. doi: 10.1016/j.cell.2008.06.025
- Shao, S., Rodrigo-Brenni, M. C., Kivlen, M. H., and Hegde, R. S. (2017). Mechanistic basis for a molecular triage reaction. *Science* 355, 298–302. doi: 10.1126/science.aah6130
- Simon, A. C., Simpson, P. J., Goldstone, R. M., Krysztofinska, E. M., Murray, J. W., High, S., et al. (2013). Structure of the Sgt2/Get5 complex provides insights into GET-mediated targeting of tail-anchored membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1327–1332. doi: 10.1073/pnas.1207518110
- Smith, D. F. (2004). Tetratricopeptide repeat cochaperones in steroid receptor complexes. *Cell Stress Chaperones* 9, 109–121. doi: 10.1379/CSC-31.1
- Thapaliya, A., Nyathi, Y., Martinez-Lumbreras, S., Krysztofinska, E. M., Evans, N. J., Terry, I. L., et al. (2016). SGTA interacts with the proteasomal ubiquitin receptor Rpn13 via a carboxylate clamp mechanism. *Sci. Rep.* 6:36622. doi: 10.1038/srep36622
- Tung, J. Y., Li, Y. C., Lin, T. W., and Hsiao, C. D. (2013). Structure of the Sgt2 dimerization domain complexed with the Get5 UBL domain involved in the targeting of tail-anchored membrane proteins to the endoplasmic reticulum. *Acta Crystallogr. Sec. D Biol. Crystallogr.* 69(Pt 10), 2081–2090. doi: 10.1107/S0907444913019379
- Vilardi, F., Stephan, M., Clancy, A., Janshoff, A., and Schwappach, B. (2014). WRB and CAML are necessary and sufficient to mediate tail-anchored protein targeting to the ER membrane. *PLoS ONE* 9:e85033. doi: 10.1371/journal.pone.0085033
- 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
- Wang, F., Brown, E. C., Mak, G., Zhuang, J., and Denic, V. (2010). A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. *Mol. Cell* 40, 159–171. doi: 10.1016/j.molcel.2010.08.038
- Wang, F., Whynter, A., Tung, M., and Denic, V. (2011). The mechanism of tail-anchored protein insertion into the ER membrane. *Mol. Cell* 43, 738–750. doi: 10.1016/j.molcel.2011.07.020
- Wang, L., Liu, Y. T., Hao, R., Chen, L., Chang, Z., Wang, H. R., et al. (2011). Molecular mechanism of the negative regulation of Smad1/5 protein by carboxyl terminus of Hsc70-interacting protein (CHIP). *J. Biol. Chem.* 286, 15883–15894. doi: 10.1074/jbc.M110.201814
- Wang, Q., Liu, Y., Soetandyo, N., Baek, K., Hegde, R., and Ye, Y. (2011). A ubiquitin ligase-associated chaperone holdase maintains polypeptides in soluble states for proteasome degradation. *Mol. Cell* 42, 758–770. doi: 10.1016/j.molcel.2011.05.010
- Winter, G., Lobley, C. M., and Prince, S. M. (2013). Decision making in xia2. *Acta Crystallogr. Sec. D Biol. Crystallogr.* 69(Pt 7), 1260–1273. doi: 10.1107/S0907444913015308
- Worrall, L. J., Wear, M. A., Page, A. P., and Walkinshaw, M. D. (2008). Cloning, purification and characterization of the *Caenorhabditis elegans* small glutamine-rich tetratricopeptide repeat-containing protein. *Biochim. Biophys. Acta* 1784, 496–503. doi: 10.1016/j.bbapap.2007.12.003
- Wunderley, L., Leznicki, P., Payapilly, A., and High, S. (2014). SGTA regulates the cytosolic quality control of hydrophobic substrates. *J. Cell Sci.* 127(Pt 21), 4728–4739. doi: 10.1242/jcs

- Yeh, Y. H., Lin, T. W., Li, Y. C., Tung, J. Y., Lin, C. Y., and Hsiao, C. D. (2014). Structural and functional characterization of ybr137wp implicates its involvement in the targeting of tail-anchored proteins to membranes. *Mol. Cell. Biol.* 34, 4500–4512. doi: 10.1128/MCB.00697-14
- Zeytuni, N., Ozyamak, E., Ben-Harush, K., Davidov, G., Levin, M., Gat, Y., et al. (2011). Self-recognition mechanism of MamA, a magnetosome-associated TPR-containing protein, promotes complex assembly. *Proc. Natl. Acad. Sci. U.S.A.* 108, E480–E487. doi: 10.1073/pnas.1103367108
- Zhang, M., Windheim, M., Roe, S. M., Pegg, M., Cohen, P., Prodromou, C., et al. (2005). Chaperoned ubiquitylation—crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP-Ubc13-Uev1a complex. *Mol. Cell* 20, 525–538. doi: 10.1016/j.molcel.2005.09.023

Conflict of Interest Statement: 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 © 2017 Krysztofinska, Evans, Thapaliya, Murray, Morgan, Martinez-Lumbreras and Isaacson. 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) or licensor 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.



Weak Molecular Interactions in Clathrin-Mediated Endocytosis

Sarah M. Smith, Michael Baker, Mary Halebian and Corinne J. Smith*

School of Life Sciences, University of Warwick, Coventry, United Kingdom

Clathrin-mediated endocytosis is a process by which specific molecules are internalized from the cell periphery for delivery to early endosomes. The key stages in this step-wise process, from the starting point of cargo recognition, to the later stage of assembly of the clathrin coat, are dependent on weak interactions between a large network of proteins. This review discusses the structural and functional data that have improved our knowledge and understanding of the main weak molecular interactions implicated in clathrin-mediated endocytosis, with a particular focus on the two key proteins: AP2 and clathrin.

Keywords: clathrin, endocytosis, adaptor-protein, structural biology, molecular interactions

INTRODUCTION

Biological processes are built on a complex interplay between proteins in the crowded, heterogeneous environment that exists within cells; and the functional protein interactions that are vital to these processes are often weak and transient. Insight into how such interactions are exploited in biological systems can help us understand how individual proteins contribute to functional networks and pathways.

One such pathway is clathrin-mediated endocytosis; a fundamental cellular process that serves to internalize cargo, that is—proteins or nutrients that need to be brought into the cell interior, and is implicated in numerous cellular functions including: nutrient uptake, membrane protein recycling, cell polarity, synaptic vesicle recycling and cell signaling. Defects in clathrin-mediated endocytosis have been linked to numerous pathological conditions such as Alzheimer's Disease, HIV/AIDS and hypercholesterolemia (Goldstein et al., 1985; McMahon and Boucrot, 2011; Zhang et al., 2011).

The main stages of clathrin-mediated endocytosis can be subdivided into 6 main steps: initiation, growth, stabilization, vesicle budding, scission and uncoating (summarized in **Figure 1A**). As the name suggests, this type of endocytosis is characterized by its reliance on a protein called clathrin which interacts with a large network of adaptor proteins during the formation of a clathrin-coated vesicle and selection of cargo for internalization. Since clathrin cannot directly interact with the lipids or proteins of the plasma membrane (Maldonado-Báez and Wendland, 2006), adaptor proteins assist in the assembly of clathrin-coated vesicles by providing a link between clathrin and the membrane-bound cargo. The main adaptor protein that clathrin engages with at the plasma membrane is adaptor protein 2 (AP2). As well as binding to clathrin, AP2 also interacts with a significant number of binding partners which include receptors destined for internalization as well as other adaptor proteins that facilitate endocytosis (summarized in **Figure 1B**). AP2 is a member of a family of five heterotetrameric complexes. These complexes contain 4 types of subunit: two large (~100 kDa), one medium (~50 kDa), and one small (~17 kDa).

Key stages in clathrin-mediated endocytosis, such as receptor recruitment and assembly of the clathrin coat, appear to rely on weak interactions that are based on recognition of short peptide sequences. This review discusses how these weak molecular interactions are exploited by the crucial endocytic components: AP2 and clathrin.

OPEN ACCESS

Edited by:

Rivka Isaacson,
King's College London,
United Kingdom

Reviewed by:

Eileen M. Lafer,
University of Texas Health Science
Center San Antonio, United States
Aaron Neumann,
University of New Mexico,
United States

*Correspondence:

Corinne J. Smith
corinne.smith@warwick.ac.uk

Specialty section:

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

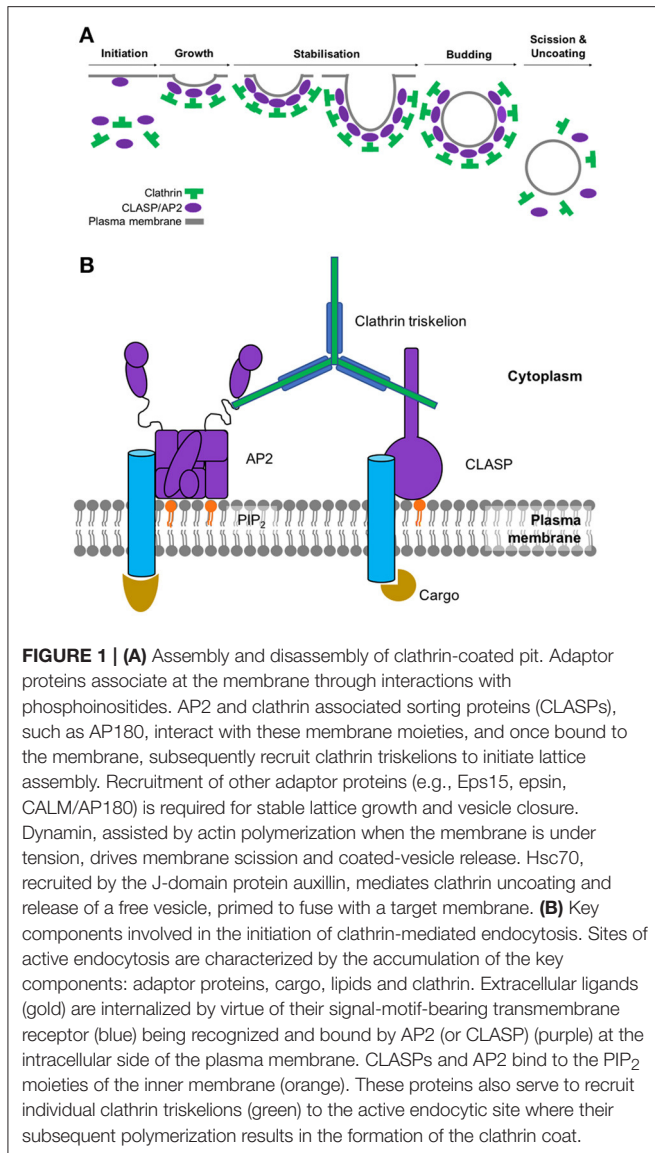
Received: 21 August 2017

Accepted: 11 October 2017

Published: 14 November 2017

Citation:

Smith SM, Baker M, Halebian M and
Smith CJ (2017) Weak Molecular
Interactions in Clathrin-Mediated
Endocytosis. *Front. Mol. Biosci.* 4:72.
doi: 10.3389/fmolb.2017.00072



AP2 INTERACTIONS—THE MOLECULAR BASIS OF RECEPTOR RECRUITMENT AND CLATHRIN COAT FORMATION

AP2 assists receptor internalization through several routes. It interacts directly with two types of internalization motifs (LL and (Y-X-X-Φ) (Φ = hydrophobic residue) found within the cytoplasmic domains of integral membrane protein receptors via its σ (LL) and μ2 (Y-X-X-Φ) subunits (Ohno et al., 1995; Owen and Evans, 1998; Owen et al., 2001; Collins et al., 2002; Kelly et al., 2008; Jackson et al., 2010). It is also associated with receptors indirectly by binding to other adaptors which are themselves (directly) associated with particular receptors, e.g., LDL receptor with autosomal recessive hypercholesterolemia (ARH) and G-protein coupled receptors (GPCRs) with arrestin.

The first receptor internalization motif to be identified was YXXΦ (Ohno et al., 1995). Surface plasmon resonance (SPR)

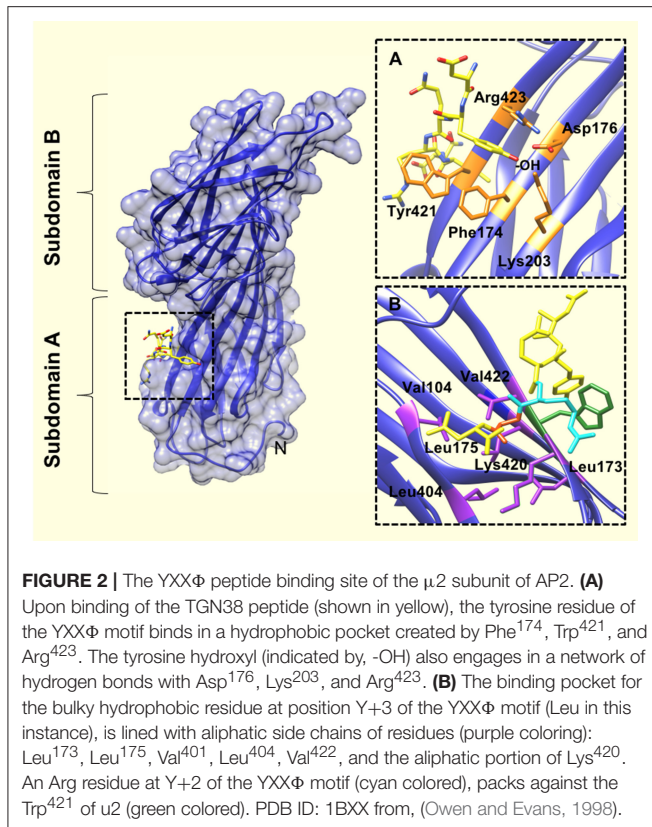
experiments revealed that the YXXΦ motif binds to the μ2 subunit of AP2 with affinities between 10 and 70 μM (Boll et al., 1996; Rapoport et al., 1997). Owen and Evans (1998) gave a structural explanation for the affinity between the aforementioned tyrosine-based motifs and AP2. A 2.7 Å crystal structure of the signal binding domain of μ2 (residues 158–435) complexed with internalization signal peptides from EGFR (Sorkin et al., 1996) and TGN38 (Bos et al., 1993; Humphrey et al., 1993) revealed that hydrophobic pockets accommodate both the tyrosine and leucine residue of the sequence motif. Upon the target peptide binding, these pockets are positioned such that 3 additional H-bonds are made between the backbone of the peptide and the AP2, resulting in β-strand formation. A similar mechanism of increased binding affinity upon correct recognition of key side chains has also been shown in other cases (Lowe et al., 1997).

The tyrosine residue of the YXXΦ sequence motif forms significant interactions with the binding pocket. For example—there are hydrophobic interactions between the tyrosine ring and Trp⁴²¹ and Phe¹⁷⁴. In addition, the tyrosine hydroxyl engages in a network of hydrogen bonds with Asp¹⁷⁶, Lys²⁰³, and Arg⁴²³. The bulky, hydrophobic residue (Φ) at position Y+3 of the internalization motif is also a major determinant of μ2 binding (Ohno et al., 1995; Boll et al., 1996), and binds in a cavity lined with aliphatic residues (Figure 2). Leu, Phe, Met or Ile residues at the Y+3 position could be accommodated in such cavity owing to the size and flexibility of side chains in the pocket.

Collins et al. (2002) obtained the structure of the AP2 core complexed with polyphosphatidylinositol headgroup mimic, inositolhexakisphosphate (IP6); which revealed two potential polyphosphatidylinositol binding sites: one on α and one on μ2. Interestingly, the YXXΦ binding motif (which localizes to the C-terminus of μ2) was occluded by part of the β2 trunk (Figure 2). This conformation of AP2 suggested to the authors a mechanism by which AP2 operates via an open or closed conformation in order to interact with motifs presented at the cell membrane.

This then raised the question of what the “open” conformation of AP2 might look like. Data showed that the distance between the end of a protein’s transmembrane helix and a YXXΦ motif requires only seven amino acids in order to confer efficient internalization (Rohrer et al., 1996). In the “closed” AP2 core structure revealed by Collins et al, the YXXΦ binding site is ~65 Å from the membrane surface; therefore, AP2 must undergo a significant conformational change not only to expose the YXXΦ binding motif, but to also ensure that it is in close enough proximity to the transmembrane cargo.

In 2008, Kelly et al. (2008) revealed the “open” conformation of AP2 upon crystallization of its core region bound to a peptide from CD4 (T-cell cell-surface antigen protein). Analysis of the crystal structure showed that the peptide bound to the core region in an extended conformation, with the LL moiety shown to bind 2 adjacent hydrophobic pockets on the σ2 subunit (Figure 3). SPR experiments showed that the CD4 LL-motif bound to WT AP2 with a K_d of 0.85 μM—considerably higher than the affinity previously shown for YXXΦ bound to the AP2 core. Comparison of this ligand-bound, “open” crystal structure with the previously



published “closed” AP2 structure (Collins et al., 2002), showed that for the LL motif to bind, the N-terminus of $\beta 2$ must be displaced from the surface of $\sigma 2$, in order to expose the hydrophobic binding pocket (Figure 3).

Whilst the “unblocking” of the LL motif binding site was explained by minor conformational changes in AP2 core structure, (Kelly et al., 2008), the YXXΦ motif-binding site remained blocked. As mentioned above, the AP2 core must undergo substantial conformational changes to permit binding of membrane-embedded YXXΦ-containing cargo. To gain molecular insight into the large conformational change of AP2, Jackson et al. (2010) solved the crystal structure of a form of AP2 whereby both LL- and YXXΦ-motif binding sites are occupied. Driven partly by the phosphorylation of Thr¹⁵⁶ on $\mu 2$ (Ricotta et al., 2002), and the electrostatic attraction of the highly positive electrostatic surface of the C-terminal region of this domain (C- $\mu 2$) to the negatively charged lipid head groups of the membrane, C- $\mu 2$ moves to the orthogonal face of the complex, resulting in the LL-motif, YXXΦ-motif and phosphatidyl inositol-4,5-bisphosphate (PtdIns4,5P₂) –binding sites becoming coplanar on the surface of AP2 and therefore suitably positioned for contacting various motifs and/or signals at the plasma membrane. The adoption of an “open” AP2 conformation would therefore cause the $\beta 2$ subunit to move out of the way and no longer occupy the motif binding sites.

Revelation of the “open” and “closed” conformations of the AP2 core structure was a significant milestone, providing mechanistic insight into how this adaptor protein is able to

interact with internalization motifs found on the cytoplasmic tails of receptors. We have so far discussed how these interactions occur at the membrane, but in order for internalization to occur, the coated vesicle itself must form. This process of coat formation is driven by interactions between AP2 and its network of binding partners which bind to the appendage (or “ear”) domains of $\alpha 2$ -adaptin and $\beta 2$ -adaptin. Here, weak interactions play a role as it was found that a number of adaptor proteins binding to the $\alpha 2$ -appendage of AP2 did so via short linear motifs with weak binding affinities. An early crystal structure of this appendage revealed that the domain interface contains tightly packed and mostly hydrophobic residues. Hydrophobic surface potential analysis revealed a single candidate protein-binding sites that was centered around residue, W840 (Owen et al., 1999).

Three linear motifs were found to bind the $\alpha 2$ -appendage domain. These were DP[FW] (Owen et al., 1999; Brett et al., 2002), FXDXF (Collins et al., 2002), and WXX[FW]X[DE] (Ritter et al., 2003; Jha et al., 2004; Walther et al., 2004). Peptides containing these linear motifs were shown to bind the $\alpha 2$ -appendage with relatively low affinities: 120 μ M, 30–50 μ M and 10 μ M, respectively (Owen et al., 1999; Edeling et al., 2006). Furthermore, structural studies showed that peptides corresponding to these motifs bound to the $\alpha 2$ -appendage in an extended conformation (Brett et al., 2002; Mishra et al., 2004; Praefcke et al., 2004; Ritter et al., 2004; Figure 4). Both the DPF/DPW and FXDXF motif bind to the $\alpha 2$ -appendage through an overlapping site in the platform subdomain (Brett et al., 2002), whereas the WXX[FW]X[DE] motif was shown to interact with the sandwich subdomain (Praefcke et al., 2004; Ritter et al., 2004; Figure 4). This additional, distinct peptide binding site on the sandwich subdomain of the $\alpha 2$ -appendage could permit multiple different motifs to bind the appendage, or, could allow multiple motifs of the same type to simultaneously bind, which would increase the avidity of the interaction (Walther et al., 2004).

The $\beta 2$ -appendage domain of AP2 was shown to possess a very similar bilobal structure to the α -appendage (Owen et al., 1999; Traub et al., 1999), with an N-terminal sandwich subdomain that is rigidly attached to a C-terminal platform subdomain (Owen et al., 2000). Also, as with the α -appendage, there was a single patch of highly hydrophobic surface potential on the $\beta 2$ -appendage platform subdomain that indicated a potential ligand interaction site. Charged residues adjacent to the hydrophobic pocket (such as R834, K842, E849, R879, E902, R904, and K917), could provide specificity for ligand-motif binding where the strength of the interaction was predominantly derived from hydrophobic interaction(s). For example, an abundance of positively charged arginine residues could confer electrostatic complementarity to a ligand rich in negatively charged amino acid side chains.

The $\beta 2$ -appendage domain binds a group of proteins that also bind the $\alpha 2$ -appendage domain: AP180, epsin and eps15. Sequence analysis has shown that there is virtually no sequence homology between these proteins/ligands, except that they each contain multiple DPF/W sequences so the authors proposed that the DØF/W motifs are likely to mediate binding to the $\beta 2$ -appendage (Owen et al., 2000). Interestingly, these proteins bind AP2 appendage domains with differing affinities:

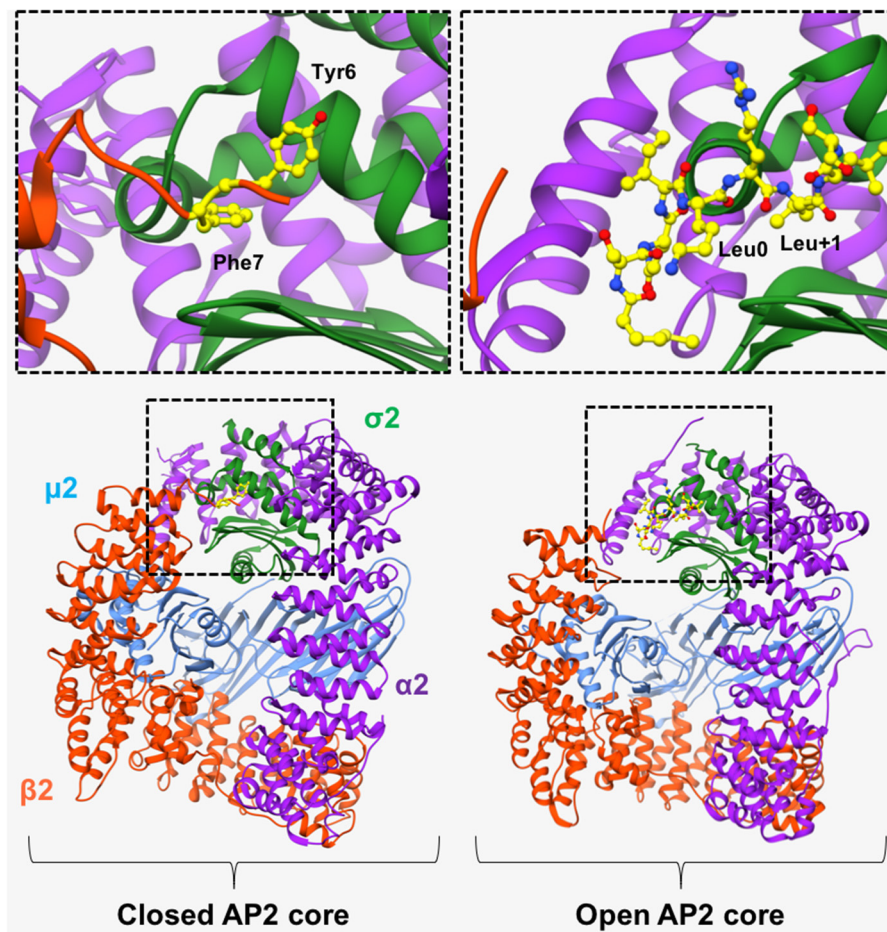


FIGURE 3 | AP2 in “open” and “closed” conformation. (Left) In the basal, “closed” conformation of AP2, two N-terminal aromatic residues (Tyr⁶ and Phe⁷, shown in yellow in top left panel), obstruct the [DE]XXX[LIM] binding site on σ 2-subunit. Right) Binding of a CD4 peptide (shown in yellow in top right panel), causes the β 2-subunit to move outwards, resulting in its N-terminus being expelled, and consequently exposing the hydrophobic binding pocket and allowing the LL-containing peptide to bind. However, the μ 2 subunit remains closely associated with the β 2 subunit and is therefore unable to bind YXX Φ motifs. α 2 subunit—purple. μ 2 subunit—blue. β 2 subunit—orange. σ 2 subunit green. PDB IDs: 2VGL from Collins et al. (2002) and, 2JKR from Kelly et al. (2008).

the α 2-appendage binds appreciable amounts of amphiphysin and has epsin as its high affinity ligand (Owen et al., 1999; Traub et al., 1999); whereas the highest affinity ligand for β 2 appendage domain is eps15, with amphiphysin showing no significant signs of binding (Owen et al., 2000). This suggests that the context of the DPF/W motifs is also a factor in the interaction of these proteins with the AP2 appendage domains.

Owen et al. (2000) also showed that the β 2 appendage together with its hinge region, is able to bind clathrin and also displace AP180, epsin and eps15 that is already bound to the domain. The binding region identified on the β 2 appendage is larger and more open than the α 2-appendage binding domain, which may explain its ability to preferentially bind clathrin. In light of these data, the author's proposed a model for CCV formation: in the cytosol, distant from regions of active endocytosis (Gaidarov et al., 1999; Roos and Kelly, 1999), AP2 appendage domains bind to D Φ F/W motif-containing accessory proteins such as AP180, epsin or eps15. In this region, the clathrin concentration

is low (Goud et al., 1985; Wilde and Brodsky, 1996; Gaidarov et al., 1999) and would therefore be unable to compete with the aforementioned accessory proteins for AP2 binding. Conversely, at sites of active clathrin-mediated endocytosis, high clathrin concentrations would enable clathrin to compete effectively with DPF/W motif-containing accessory proteins for binding to the β 2-appendage of AP2. Once bound, clathrin would be able to polymerize and consequently form a lattice, and recruit more clathrin.

A subset of the AP2 appendage binding accessory proteins are also able to bind cargo. These proteins, termed clathrin-associated sorting proteins (CLASPs), increase the catalog of endocytic cargo that can be recruited by AP2 beyond transmembrane proteins bearing cytoplasmic internalization motifs. Two key examples of the use of CLASPs are low density lipoproteins (LDL) and GPCRs, which are internalized by the CLASP proteins, ARH or Dab2 (Traub, 2005; Maurer and Cooper, 2006), and β -arrestin (Lefkowitz and Shenoy, 2005), respectively.

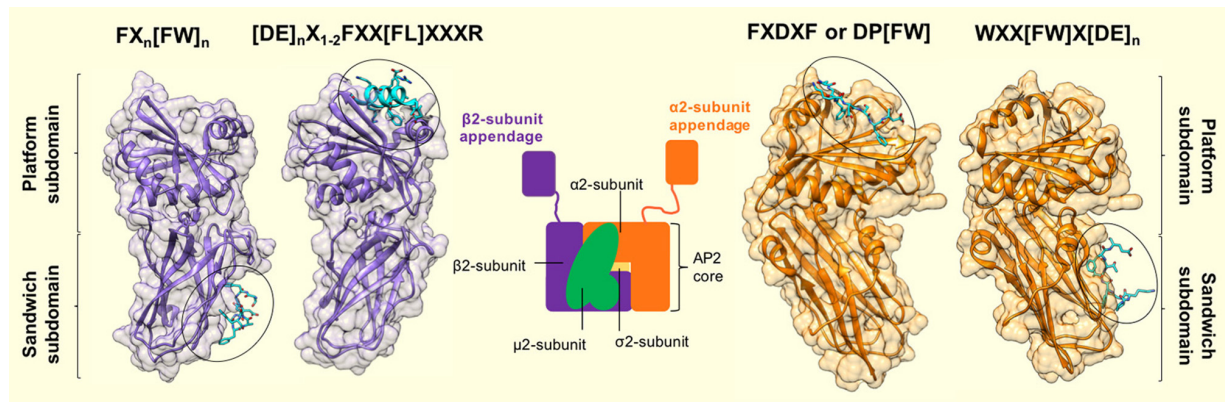


FIGURE 4 | Motif binding to the appendage domains of AP2. The α 2- and β 2-subunit appendages of AP2 (orange and purple, respectively) share a similar bilobal structure each consisting of an N-terminal sandwich subdomain attached to a C-terminal platform subdomain. Each subdomain, of each appendage contains a distinct interaction surface for protein partner binding, which results in a single AP2 molecule possessing 4 separate contact sites. Specificity for each site is conferred by short, interaction motifs which are characterized by aromatic side chains. The sandwich subdomain of the α 2-appendage binds to WXX[FW]X[DE]_n-containing ligands (where X is any amino acid). [PDB code: 1W80, (Praefcke et al., 2004)]. The platform subdomain of the same appendage binds either FXDXF or DP[FW] motifs [PDB code: 1KY7, (Brett et al., 2002)]. The same subdomain of the β 2-subunit binds to [DE]_nX₁₋₂FXX[FL]XXXXR sequences that are presented in a α -helical conformation. Such motifs are present in β -arrestin, ARH and epsins (Edeling et al., 2006; Schmid et al., 2006) [PDB code: 2G30, (Edeling et al., 2006)]. Finally – the sandwich subdomain binds to a Phe-rich motif that is present in proteins, Eps15 and AP180 [PDB code: 2IV9, (Schmid et al., 2006)]. Peptide motif nomenclature as in Traub (2009).

The binding of ARH to AP2 is highly selective for the β 2-appendage (He et al., 2002; Laporte et al., 2002; Mishra et al., 2002). The 1.6 Å crystal structure of a β 2-appendage in complex with an ARH-derived peptide (²⁵²DDGLDEAFSRLAQSR) (Edeling et al., 2006), revealed a completely different mode of interaction in comparison to all other known appendage ligands. The ARH peptide adopted an α -helical conformation which bound a deep groove on the top of the β 2-appendage (Figure 4). Analysis of the helix-binding region showed that Leu262 of the ARH peptide (termed [FL] pocket) was accommodated in a hydrophobic pocket of the β 2 platform subdomain. Phe259 of the ARH peptide fitted into an adjacent, complementary hydrophobic pocket on the β 2-appendage which the authors denote the “[F] pocket,” and the side chain of the residue Arg266 extends along a small channel on the surface of the β 2 subdomain ([R] pocket) which forms hydrogen bonds with acidic residues Glu902 and Glu849. For these F, [FL] and [R] pocket interactions to occur, the α -helical motif must fit into its binding groove, therefore providing the specificity for binding. In agreement with this, a 2.5 Å crystal structure of the β 2-appendage co-complexed with a Eps15 peptide and β -arrestin confirmed that the core motif for interaction with this AP2 appendage is: DxxFxxFxxxR, and exhibits an alpha helical conformation (Schmid et al., 2006).

Furthermore, β -arrestin, which has already been shown to bind the β 2 appendage (Laporte et al., 2000; Kim and Benovic, 2002; Milano et al., 2002), displays significant sequence similarity with the β 2-appendage binding motif of ARH at its C-terminus. Subsequent sequence analysis and mutagenesis of this C-terminal peptide region demonstrated the importance of the FXX[FL]XXXXR motif in binding the platform subdomain of the β 2-appendage (Edeling et al., 2006). Isothermal titration calorimetry (ITC) measurements confirmed that this region (³⁸³DDDIVFEDFARQLKG) of β -arrestin binds the β 2-appendage with a K_d of 2.6 μ M; a value very similar to the

K_d of ARH peptide of 2.4 μ M (Mishra et al., 2005). Such affinity values are comparatively higher than those for a YXX Φ motif binding to the μ 2 subunit of AP2 (Boll et al., 1996; Rapoport et al., 1997).

What's more, protein database searching revealed that mammalian epsins 1 and 2 also possess FXX[FL]XXXXR motifs located in their unstructured region, which ITC experiments confirmed to bind the β 2-appendage (Edeling et al., 2006). Further analysis showed that epsin, ARH and β -arrestin also contain acidic residues N-terminal to the proximal phenylalanine and thus the β 2-appendage binding motif is more accurately described as: [DE]_nX₁₋₂FXX[FL]XXXXR.

Taken together, these data reveal fundamental differences in the mode of interaction between the β 2 platform domain and CLASPs compared to other appendage-ligand interactions. Instead of numerous, avidity-based interaction motif repeats, β -arrestin, ARH and epsin contain only a single [DE]_nX₁₋₂FXX[FL]XXXXR motif, which adopts an α -helical conformation to bind the β 2 appendage. Therefore, not only are there charge and hydrophobic components to the interaction between ligand and AP2, but extra specificity is conferred by the requirement that the CLASP motif folds into a helix with the interacting residues on one face of the helix.

In addition to high affinity interactions between AP2 appendages and accessory proteins bearing a single appendage binding motif, α 2- and β 2-appendages also engage in high avidity interactions. SPR experiments between immobilized α 2- or β 2-appendages and the motif domain of Eps15, showed very tight interactions such that an off-rate could not be measured (Schmid et al., 2006). It was assumed that these tight interactions were due to the presence of multiple appendage interaction sites in a single protein domain of Eps15. Therefore, if appendages are linked/bound to the same surface there is a high avidity for ligand interaction that is much stronger

than the sum of individual affinities (Praefcke et al., 2004). So in the context of clathrin-mediated endocytosis, such an environment would be akin to “assembly-zones,” where AP2 is clustered at the membrane, presenting multiple appendages that are available for accessory protein binding. Proteins with multiple appendage interaction sites will not only aid adaptor clustering, but the presentation of many juxta-positioned motifs leads to an increased affinity for the adaptor appendage. Therefore, individual weak affinity interactions between AP2 and its ligands can make significant contributions to protein-protein interactions, providing there are multiple copies.

Schmid et al. (2006) proposed that clathrin coated pit (CCP) formation proceeds as a result of high avidity interactions of accessory proteins being replaced by the weak interactions of the clathrin coat with adaptors; meaning that initially low affinity (and therefore readily reversible) interactions between cargo and adaptors, between adaptors and accessory proteins, and between accessory proteins and clathrin, are used to build the network.

CLATHRIN-ADAPTOR INTERACTIONS MEDIATED BY SHORT PEPTIDE MOTIFS

A pivotal step upon the recruitment of clathrin to sites of endocytosis is the interaction between individual clathrin triskelia and an array of accessory proteins that assist in the

formation of a clathrin-coated pit. Clathrin has more than 20 binding partners and interacts with most of these via a 7-bladed beta-propeller domain at its N-terminus (Figures 5A,B). Interactions between the clathrin N-terminal domain (TD) and peptides corresponding to multiple binding motifs are in the micromolar range. Thus, weak interactions also feature in the role of clathrin as well as AP2 in endocytosis.

The mode of binding of adaptor proteins for clathrin was investigated by Dell’Angelica et al. (1998) using a combination of GST pull-down assays and mutagenesis. Through this they identified a segment of residues (SLLDLDDFN817–825) in the β 3-appendage of AP-3 that contributed to binding to the clathrin TD. Subsequent sequence analysis also identified similar amino acid residues in other adaptor proteins that have been shown to mediate interaction(s) with clathrin; namely—amphiphysin II (Ramjaun and McPherson, 1998), segments of arrestin3 (Krupnick et al., 1997) and in the clathrin-binding region of β 1 and β 2 (Shih et al., 1995). Alignment of these sequences enabled the definition of a motif for clathrin-binding, comprised of acidic and bulky hydrophobic residues, L(L, I)(D, E, N)(L, F)(D, E), termed the “clathrin box” motif (Figures 5B,C). Although they vary between proteins, clathrin-box motifs are highly-conserved and are found in many proteins known to interact with the TD, for example—AP1, epsin, AP180 and amphiphysin (Shih et al., 1995; Drake et al., 2000; Kirchhausen, 2000).

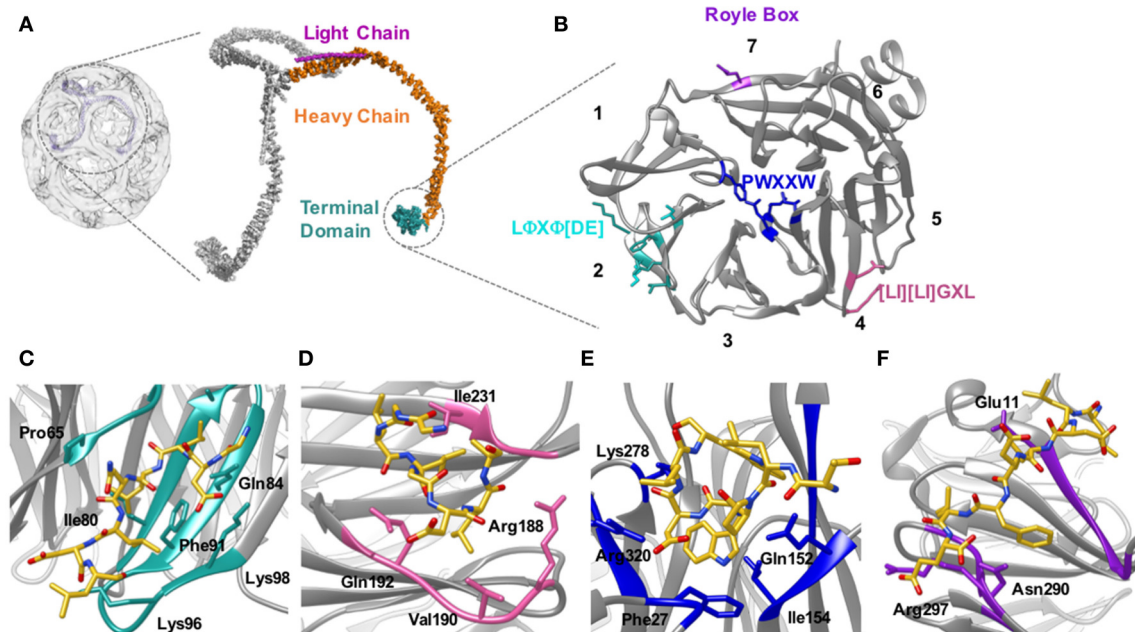


FIGURE 5 | Location of the clathrin heavy chain N-terminal domain (TD) and the location of the adaptor binding sites. **(A)** Clathrin forms a polymerized lattice structure around the growing vesicle in concert with adaptor proteins. The functional monomer, the triskelion, is formed of a trimer of heavy chains (~190 kDa, orange) with a smaller light chain (~25 kDa Pink) located along the top edge near the dimerization domain. The TD (cyan) is the primary binding location for adaptor proteins, and is located on the inside of the cage closest to the plasma membrane. **(B)** The TD is a 7-bladed β propeller that has 4 known sites for binding adaptor proteins. Site 1 between blades 1 and 2 is known as the clathrin box site (L Φ X Φ [DE]), (turquoise); Site 2 situated in the center of the propeller is known as the W-box (PWXXW), (blue); Site 3 is the Arrestin-box ([L][L]GXL), (Pink); and Site 4 is the Royle Box which as yet has no defined interaction sequence, (purple). The numbers indicate the blade number. Peptides or protein bound with the 4 sites are indicated in the following four panels: **(C)** Clathrin box site of β 2 adaptin (CGDLLNLDLG) bound to site 1; **(D)** β arrestin1L peptide (ALDGLLG) bound to site 3; **(E)** amphiphysin peptide (TLPWDLWTT) bound to site 2; **(F)** an amphiphysin peptide bound to site 4. Structures are derived from PDB codes: 3IYV **(A)** (Fotin et al., 2004), 5M5R **(B,C)** (Muenzner et al., 2017), 3GD1 **(D)** (Kang et al., 2009), 1UTC **(E)** (Miele et al., 2004), 5M5T **(F)** (Muenzner et al., 2017).

A number of X-ray structures of the clathrin TD provided both its tertiary structure and the binding location of peptides from several different adaptor proteins, revealing 3 independent binding sites for clathrin binding partners on this relatively small (40 kDa) (ter Haar et al., 2000) domain.

Almost two decades ago, ter Haar et al. (1998) showed that the clathrin TD comprised a 7-bladed beta-propeller structure linked to a series of short alpha helices which formed the start of the clathrin “leg” region. ter Haar et al. (2000) then determined structures of complexes of clathrin TD with peptides derived from adaptors β -arrestin 2, and the β -subunit of AP3. The residues contacting the TD in both structures were consistent with the five-residue clathrin box motif identified previously: L Φ X Φ [DE] (where x denotes any amino acid, Φ denotes a bulky hydrophobic residue and [DE] is a glutamate or aspartate). Both structures revealed very similar peptide interactions with each peptide binding in an extended conformation in a groove between blades 1 and 2 of the propeller structure (Figures 5B,C). The sharing of the same binding site by both peptides was surprising given previous evidence suggesting that β -arrestin 2 and AP2 could bind at different sites on the TD (Goodman et al., 1997).

The situation became more complex when a 2.3 Å crystal structure of clathrin TD bound to a peptide of amphiphysin 1 revealed a second TD-binding motif, PWXXW (termed “the W-box”), to bind at a site remote from the “clathrin-box” binding site (Miele et al., 2004; Figure 5B). The presence of a second motif-binding site had previously been suggested by biochemical data which indicated that the binding sequence, PWDLW, could bind to the TD without competing with the canonical L Φ X Φ [DE] clathrin-box motif (Ramjaun and McPherson, 1998; Slepnev et al., 2000; Drake and Traub, 2001).

Unlike the clathrin-box motif (which adopts an extended conformation when bound to the TD), the bound W-box was compact and helical, buried in a solvent-exposed cavity of complementary shape on the membrane-proximal “top” surface of the TD; a location spatially distinct from where the clathrin-box peptides bind (Miele et al., 2004; Figure 5E). Affinity measurements showed that the W-box peptide binds the TD with a similar affinity (Kd of 28 μ M) to that of the clathrin-box peptide (Kd of 22 μ M).

Finally, a third, spatially distinct adaptor binding site was identified by Kang et al. (2009), who showed that an extended surface loop of the arrestin 2 long isoform occupied a site between blades 4 and 5 of the TD, which binds peptides with motif [LI][LI]GxL—termed the “arrestin-box” (Figures 5B,D).

The above crystal structures revealed the location of adaptor protein binding sites to the TD; however, the role of these interactions in coat assembly has been difficult to define. The structures obtained so far are of peptides corresponding to clathrin-binding motifs co-crystallized with the clathrin TD. It would be interesting to know how TD binding to the full-length clathrin-binding domains of these proteins compares but this will be hard to achieve by crystallography (ter Haar et al., 2000; Miele et al., 2004) owing to the unstructured nature of these clathrin binding regions. AP180, implicated in clathrin assembly, is one such example. It has a 33 kDa N-terminal ANTH domain, which is involved in membrane binding; and a largely unstructured 58

kDa C-terminal region that is responsible for clathrin-binding and assembly. More specifically—the C-terminal region binds to the TD of the clathrin heavy chain (Morgan et al., 2000), and self-homology analyses of this region showed that it contains 12 repeats, each \sim 23 aa in length, and containing a single DLL/DLF sequence per repeat. The large number of clathrin binding motifs along the length of the AP180 sequence suggests that organization of AP180 binding to clathrin must go beyond a straight forward 1:1 interaction between a single DLL motif and the clathrin TD.

MORE COMPLEX INTERACTIONS—MULTIPLE DLL MOTIFS

The potential for complex binding interactions between clathrin binding motifs and clathrin TD led Zhuo et al. (2010) to further investigate the binding of DLL and DLF motifs in AP180 to clathrin TD. Zhuo et al. (2010) demonstrated that the DLL and DLF sequences within the clathrin binding site are critical for clathrin binding, and bind clathrin TD relatively weakly, with Kd values in the \sim 2 \times 10^{−4} M range. The weak binding of these sites to the clathrin TD and the observation that chemical exchange kinetics are in the intermediate to fast-exchange regimen (Schlosshauer and Baker, 2004) indicate that both association and dissociation rates for these interactions are rapid, with dissociation rates in the range of 2 \times 10³ and 4 \times 10³ s^{−1}, and association rates in the range of 1 \times 10⁷ to 2 \times 10⁷ M^{−1} s^{−1}.

In light of these data, the authors were able to expand on previous models for how AP180 mediates clathrin coat assembly. AP180 binds to the membrane via its ANTH domain, resulting in its unstructured, flexible C-terminal region being exposed to the cytoplasm and available for binding any clathrin molecules encountered. Although the clathrin binding site of AP180 binds clathrin weakly with rapid dissociation rates, the likelihood of clathrin diffusing away is minimized given that each AP180 molecule contains up to 12 clathrin binding sites; therefore, if a clathrin molecule unsuccessfully binds one site, it is possible that it can interact with the many other clathrin binding sites. What's more, the rapid dissociation rates mean that each triskelion is able to move and reorient itself, enabling interactions with other triskelia to be established. Those clathrin-clathrin interactions that occur will determine both the geometry and stability of the clathrin lattice. In this way, weak binding by multiple clathrin triskelia to binding sites dispersed throughout the AP180 sequence allows efficient recruitment of clathrin to endocytic sites and dynamic assembly of the clathrin lattice. This example of weak, multi-valent binding in combination with intrinsic disorder of a protein binding partner is able to create a highly dynamic mode of protein-protein interaction.

COMPLEXITY OF BINDING TO MULTIPLE CLATHRIN TERMINAL DOMAIN SITES

The crystal structures of ter Haar et al. (2000), Miele et al. (2004), and Kang et al. (2009) suggested that clathrin-box motif, W-box motif and arrestin splice loop 5 peptides bind uniquely to

individual sites on the TD, respectively, giving a total of 3 binding sites. However, data published since then has suggested that such a situation is likely to be an oversimplification. For example, yeast epsin (Ent2p) was shown to bind a TD where clathrin-box and W-box binding sites were mutated. Additionally, deletion of the Ent2p C-terminal clathrin-box sequence eliminated Ent2p binding to the TD (Collette et al., 2009). Together, these data indicate that clathrin-box sequences are able to bind the TD at site(s) distinct from site 1 or 2. In fact, a fourth adaptor binding site was identified by Willox and Royle (2012). This study found that mutating all 3 binding sites did not block clathrin/AP2 mediated endocytosis in human cell lines whereas deleting the TD inhibited endocytosis. Using an *in silico* approach they were able to identify a conserved patch located $\sim 120^\circ$ relative to the other binding sites. This site encompasses the end of strand d of blade 7 and the helical segment in the loop connecting blades 7 and 1 (Figures 5B,F). Mutating E11 to K on this 4th site, in combination with mutations to the other 3 sites, resulted in the same phenotype as shown by deletion of the TD. The identity of the motif conferring binding to this site remains undefined.

A deeper understanding of adaptor binding to the four, distinct adaptor-binding sites on the clathrin TD must account for observations that three out of four of the aforementioned binding sites can be mutationally eliminated without causing loss of CME (Willox and Royle, 2012).

In an effort to gain insight into the ambiguities regarding adaptor/accessory protein binding to clathrin TD, Zhuo et al. (2015) adopted a solution-based NMR approach to study the interaction of clathrin TD with clathrin-box peptides derived from AP2 adaptor protein and the accessory protein, AP180. Results showed that these peptides simultaneously bound the clathrin-box site, the W-box site and the β -arrestin splice loop site of a single TD with a similar, low affinity (Kd values in the range of 800–900 μ M). The high promiscuity and stoichiometry of binding of peptide to the TD could be a reflection of the functional redundancy of these sites, and could also be important for the dynamic reorganization of the clathrin TD during endocytosis.

In agreement with biochemical data that showed clathrin only precipitated in GST-binding assays upon immobilization of a high density of clathrin-box peptides to a GST-resin (Drake et al., 2000; Drake and Traub, 2001), the weak molecular interactions between clathrin-box peptides and the clathrin TD (Zhuo et al., 2015) means that multiple interactions are required for a stable association of adaptor/accessory protein with clathrin. Furthermore, these data also suggest that each TD can bind up to 3 such peptides, which not only increases the potential avidity of peptide-TD interaction, but also offers an explanation as to why individual binding sites in the TD can be mutationally eradicated without significantly compromising CME (Lemmon and Traub, 2012). Also, it has been proposed that weak molecular interactions between TD and peptides would facilitate the dynamic reorganization of clathrin during lattice assembly (Zhuo et al., 2010). The temporal regulation of this event and the fact that it involves transfer of clathrin between different adaptor and accessory proteins during the process of internalizing cargo (Drake et al., 2000; McMahon and

Boucrot, 2011), enables the development of our understanding of clathrin-mediated endocytosis. The authors (Zhuo et al., 2015) propose that the differing affinities and number of clathrin binding sequences in an adaptor/accessory protein could be an important factor in aiding clathrin transfer: tighter binding, or more clathrin binding sequences could displace a protein that has weaker or fewer clathrin binding elements.

Muenzner et al. (2017) endeavored to investigate the suggested potential degeneracy of clathrin binding, (Willox and Royle, 2012; Zhuo et al., 2015) by resolving high resolution structures of clathrin TD complexed with cellular and viral peptide motifs. In contrast to previous crystallographic structures (ter Haar et al., 2000; Miele et al., 2004), where the co-complexed peptide was shown only to bind a single site of the clathrin TD, the structures resolved by Muenzner and colleagues demonstrated that 2 distinct sequence motifs (arrestin-box and the clathrin-box), can bind the arrestin box binding site of clathrin TD.

Furthermore, the authors also note that the sequences capable of binding the Royle box are somewhat variable (amphiphysin I clathrin-binding motif peptide: ETLDDLDFLE and hepatitis D virus large antigen peptides: SDILFPADS and SPRLPLLES), preventing the unambiguous identification of a consensus binding sequence. Thus, they suggest that the model of “1 consensus motif binds a single peptide-binding site on the clathrin TD” may require revision since binding could rely on the peptide’s structural environment upon contacting the TD (Muenzner et al., 2017).

The fact that a clathrin TD is capable of simultaneously binding multiple adaptors emphasises the dynamic nature of clathrin-adaptor interactions. The authors (Muenzner et al., 2017) go on to discuss that differences in the affinity of protein-protein interactions come as a result of differing rates of dissociation (Pollard, 2010). Weak molecular interactions, such as those between clathrin and its adaptor proteins (Shih et al., 1995; Zhuo et al., 2015), are on the order of approximately 1 per second (Pollard, 2010). Given that the timeframe of complete clathrin-coated pit formation and disassembly is ~ 90 s (Loerke et al., 2009), we would anticipate that adaptors undergo rapid cycles of binding and dissociation from clathrin, which would enable the recruitment of many different adaptor proteins to a given clathrin TD. Also, the promiscuity of clathrin motif binding would permit a single adaptor protein that contains multiple clathrin interaction motifs (e.g., AP180 Zhuo et al., 2010), to simultaneously bind multiple sites on clathrin TD, consequently increasing the affinity of the interaction.

CONCLUSION

This review of the structural and functional experiments that investigated the binding between cargo, adaptor and accessory proteins, as well as clathrin, has demonstrated the different ways weak molecular interactions are exploited in clathrin-mediated endocytosis.

AP2 is a key regulatory factor in clathrin-mediated endocytosis, and its activation commences upon recruitment and subsequent low-affinity interactions with PtdIns4,5P2 of

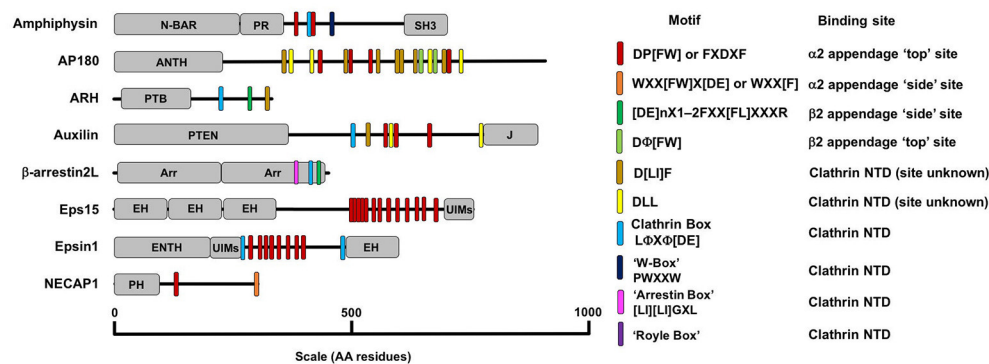


FIGURE 6 | A diagram detailing the AP2 and clathrin binding motifs present in a number of adaptor proteins with diverse structure and function. Motifs are listed in the key on the right along with their binding location on AP2 or clathrin. The other domains detailed in the figure are as follows: ANTH, AP180 N-Terminal Homology Domain; Arr, Arrestin Domain; CC, coiled-coil domain; EH, Epsin Hand; ENTH, Epsin N-Terminal Homology Domain; J, J-domain; SH3, SRC Homology domain 3; PH, Plextrin Homology domain; PR, Proline Rich domain; PTB, Phosphotyrosine Binding domain; N-BAR, N-terminal Bin/Amphiphysin/Rvs domain; UIM, Ubiquitin Interacting Motif.

the plasma membrane (Höning et al., 2005). Phosphorylation of Thr156 of the μ 2 domain causes AP2 to adopt an “open” conformation allowing it to interact with the large network of other accessory proteins and clathrin.

The evolution of AP2 to adopt “open” and “closed” conformations allows this protein to spatially and temporally control different stages of cargo internalization, from the point of cargo recognition and sorting, to downstream clathrin-coated pit formation.

Together, the AP2 core and its appendage domains bind their respective ligands via motifs (summarized in **Figure 6**), which engage in weak molecular interactions. This network of low-affinity protein interactions provides not only high avidity and specificity, but also reversibility of protein interactions that allows for rapid exchange of binding partners, accounting for the dynamic nature of CME *in vivo* (Avinoam et al., 2015).

Whilst low-affinity interactions are common throughout clathrin-mediated endocytosis, AP2 is also able to engage in high affinity interactions with some accessory proteins. These interactions are conferred by the requirement that a ligand motif binds in an α -helical conformation, as opposed to an extended conformation that is more commonly used. These two different modes of interaction between AP2 appendages, along with the multiple variants in binding motifs, explain how AP2 is able to act as a central hub for interactions between adaptors and clathrin. However, a greater understanding of how these motifs interact with and compete with each other for AP2 would greatly enhance our understanding of how adaptors and cargo are spatially and temporally regulated *in vivo*.

The clathrin TD has been identified as the major interaction site on clathrin for adaptor proteins. Initial crystallographic structures identified 3 potential binding sites for specific adaptor motifs (**Figure 5B**), suggesting that adaptors would bind to discrete locations on the TD (ter Haar et al., 2000; Miele et al.,

2004; Kang et al., 2009). However, recent studies both *in vivo* and *in vitro* have identified a 4th binding site (**Figures 5B,F**) and present evidence that the binding of these motifs is much more degenerate than previously expected, with peptides of a given sequence found to bind in more than one location (Willox and Royle, 2012; Zhuo et al., 2015; Muenzner et al., 2017). As with AP2-adaptor interactions, a greater understanding of the relative affinities of these motifs or associated proteins would give us better insight into how adaptor recruitment is regulated in CME.

To conclude, clathrin-mediated endocytosis is a versatile pathway, not just in terms of the diversity of cargos that can be internalized, or in the large number of accessory and adaptor proteins used, but it also in the pivotal role of weak molecular interactions orchestrating and controlling the internalization of specific cargo and its delivery to early endosomes.

AUTHOR CONTRIBUTIONS

All authors (SS, MB, MH, and CS) fulfill the “author criteria” of: Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; and Drafting the work or revising it critically for important intellectual content; and Final approval of the version to be published; and Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

ACKNOWLEDGMENTS

SS and CS are funded by BBSRC Research grant BB/N008391/1. CS was also supported by a Royal Society Leverhulme Trust Senior Research Fellowship LT150036. MH is funded by the MRC grant number MR/J003964/1. MB thanks the BBSRC Midlands Integrative Biosciences Training Partnership (MIBTP) for support.

REFERENCES

- Avinoam, O., Schorb, M., Beese, C. J., Briggs, J. A. G., and Kaksonen, M. (2015). Endocytic sites mature by continuous bending and remodeling of the clathrin coat. *Science* 348:1369. doi: 10.1126/science.aaa9555
- Boll, W., Ohno, H., Songyang, Z., Rapoport, I., Cantley, L. C., Bonifacino, J. S., et al. (1996). Sequence requirements for the recognition of tyrosine-based endocytic signals by clathrin AP-2 complexes. *EMBO J.* 15, 5789–5795.
- Bos, K., Wraight, C., and Stanley, K. K. (1993). TGN38 is maintained in the trans-Golgi network by a tyrosine-containing motif in the cytoplasmic domain. *EMBO J.* 12, 2219–2228.
- Brett, T. J., Traub, L. M., and Fremont, D. H. (2002). accessory protein recruitment motifs in clathrin-mediated endocytosis. *Structure* 10, 797–809. doi: 10.1016/S0969-2126(02)00784-0
- Collette, J. R., Chi, R. J., Boettner, D. R., Fernandez-Golbano, I. M., Plemel, R., Merz, A. J., et al. (2009). Clathrin functions in the absence of the terminal domain binding site for adaptor-associated clathrin-box motifs. *Mol. Biol. Cell* 20, 3401–3413. doi: 10.1091/mbc.E08-10-1082
- Collins, B. M., McCoy, A. J., Kent, H. M., Evans, P. R., and Owen, D. J. (2002). Molecular architecture and functional model of the endocytic AP2 complex. *Cell* 109, 523–535. doi: 10.1016/S0092-8674(02)00735-3
- Dell'Angelica, E. C., Klumperman, J., Stoorvogel, W., and Bonifacino, J. S. (1998). Association of the AP-3 adaptor complex with clathrin. *Science* 280:431. doi: 10.1126/science.280.5362.431
- Drake, M. T., Downs, M. A., and Traub, L. M. (2000). Epsin binds to clathrin by associating directly with the clathrin-terminal domain: evidence for cooperative binding through two discrete sites. *J. Biol. Chem.* 275, 6479–6489. doi: 10.1074/jbc.275.9.6479
- Drake, M. T., and Traub, L. M. (2001). Interaction of two structurally distinct sequence types with the clathrin terminal domain β -propeller. *J. Biol. Chem.* 276, 28700–28709. doi: 10.1074/jbc.M104226200
- Edeling, M. A., Mishra, S. K., Keyel, P. A., Steinhauser, A. L., Collins, B. M., Roth, R., et al. (2006). Molecular switches involving the AP-2 β 2 appendage regulate endocytic cargo selection and clathrin coat assembly. *Dev. Cell* 10, 329–342. doi: 10.1016/j.devcel.2006.01.016
- Fotin, A., Cheng, Y., Sliz, P., Grigorieff, N., Harrison, S. C., Kirchhausen, T., et al. (2004). Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* 432, 573–579. doi: 10.1038/nature03079
- Gaidarov, I., Santini, F., Warren, R. A., and Keen, J. H. (1999). Spatial control of coated-pit dynamics in living cells. *Nat. Cell Biol.* 1, 1–7.
- Goldstein, J. L., Brown, M. S., Anderson, R. G. W., Russell, D. W., and Schneider, W. J. (1985). Receptor-mediated endocytosis: concepts emerging from the LDL receptor system. *Annu. Rev. Cell Biol.* 1, 1–39. doi: 10.1146/annurev.cb.01.110185.000245
- Goodman, O. B., Krupnick, J. G., Gurevich, V. V., Benovic, J. L., and Keen, J. H. (1997). Arrestin/Clathrin Interaction: localisation of the arresting binding locus to the clathrin terminal domain. *J. Biol. Chem.* 272, 15017–15022. doi: 10.1074/jbc.272.23.15017
- Goud, B., Huet, C., and Louvard, D. (1985). Assembled and unassembled pools of clathrin: a quantitative study using an enzyme immunoassay. *J. Cell Biol.* 100, 521–527. doi: 10.1083/jcb.100.2.521
- He, G., Gupta, S., Yi, M., Michael, P., Hobbs, H. H., and Cohen, J. C. (2002). ARH is a modular adaptor protein that interacts with the LDL receptor, clathrin, and AP-2. *J. Biol. Chem.* 277, 44044–44049. doi: 10.1074/jbc.M208539200
- Höning, S., Ricotta, D., Krauss, M., Späte, K., Spolaore, B., Motley, A., et al. (2005). Phosphatidylinositol(4,5)-bisphosphate regulates sorting signal recognition by the clathrin-associated adaptor complex AP2. *Mol. Cell* 18, 519–531. doi: 10.1016/j.molcel.2005.04.019
- Humphrey, J. S., Peters, P. J., Yuan, L. C., and Bonifacino, J. S. (1993). Localization of TGN38 to the trans-Golgi network: involvement of a cytoplasmic tyrosine-containing sequence. *Jo. Cell Biol.* 120, 1123–1135. doi: 10.1083/jcb.120.5.1123
- Jackson, L. P., Kelly, B. T., McCoy, A. J., Gaffry, T., James, L. C., Collins, B. M., et al. (2010). A Large-scale conformational change couples membrane recruitment to cargo binding in the AP2 clathrin adaptor complex. *Cell* 141, 1220–1229. doi: 10.1016/j.cell.2010.05.006
- Jha, A., Agostinelli, N. R., Mishra, S. K., Keyel, P. A., Hawryluk, M. J., and Traub, L. M. (2004). A novel AP-2 adaptor interaction motif initially identified in the long-splice isoform of synaptojanin 1, SJ170. *J. Biol. Chem.* 279, 2281–2290. doi: 10.1074/jbc.M305644200
- Kang, D. S., Kern, R. C., Puthenveedu, M. A., von Zastrow, M., Williams, J. C., and Benovic, J. L. (2009). Structure of an arrestin2-clathrin complex reveals a novel clathrin binding domain that modulates receptor trafficking. *J. Biol. Chem.* 284, 29860–29872. doi: 10.1074/jbc.M109.023366
- Kelly, B. T., McCoy, A. J., Späte, K., Miller, S. E., Evans, P. R., Höning, S., et al. (2008). A structural explanation for the binding of endocytic dileucine motifs by the AP2 complex. *Nature* 456, 976–979. doi: 10.1038/nature07422
- Kim, Y.-M., and Benovic, J. L. (2002). Differential roles of arrestin-2 interaction with clathrin and adaptor protein 2 in G protein-coupled receptor trafficking. *J. Biol. Chem.* 277, 30760–30768. doi: 10.1074/jbc.M204528200
- Kirchhausen, T. (2000). Clathrin. *Annu. Rev. Biochem.* 69, 699–727. doi: 10.1146/annurev.biochem.69.1.699
- Krupnick, J. G., Goodman, O. B., Keen, J. H., and Benovic, J. L. (1997). Arrestin/Clathrin Interaction: localisation of the clathrin binding domain of non visual arresting to the carboxyl terminus. *J. Biol. Chem.* 272, 15011–15016. doi: 10.1074/jbc.272.23.15011
- Laporte, S. A., Miller, W. E., Kim, K.-M., and Caron, M. G. (2002). β -arrestin/AP-2 interaction in G protein-coupled receptor internalisation: identification of a β -arrestin binding site in β 2-adaptin. *J. Biol. Chem.* 277, 9247–9254. doi: 10.1074/jbc.M108490200
- Laporte, S. A., Oakley, R. H., Holt, J. A., Barak, L. S., and Caron, M. G. (2000). The interaction of β -arrestin with the AP-2 adaptor is required for the clustering of β 2-adrenergic receptor into clathrin-coated pits. *J. Biol. Chem.* 275, 23120–23126. doi: 10.1074/jbc.M002581200
- Lefkowitz, R. J., and Shenoy, S. K. (2005). Transduction of receptor signals by β -arrestins. *Science* 308:512. doi: 10.1126/science.1109237
- Lemmon, S. K., and Traub, L. M. (2012). Getting in touch with the clathrin terminal domain. *Traffic* 13, 511–519. doi: 10.1111/j.1600-0854.2011.01321.x
- Loerke, D., Mettlen, M., Yasar, D., Jaqaman, K., Jaqaman, H., Danuser, G., et al. (2009). Cargo and dynamin regulate clathrin-coated pit maturation. *PLoS Biol.* 7:e1000057. doi: 10.1371/journal.pbio.1000057
- Lowe, E. D., Noble, M. E., Skamnik, V. T., Oikonomakos, N. G., Owen, D. J., and Johnson, L. N. (1997). The crystal structure of a phosphorylase kinase peptide substrate complex: kinase substrate recognition. *EMBO J.* 16, 6646–6658. doi: 10.1093/emboj/16.22.6646
- Maldonado-Báez, L., and Wendland, B. (2006). Endocytic adaptors: recruiters, coordinators and regulators. *Trends Cell Biol.* 16, 505–513. doi: 10.1016/j.tcb.2006.08.001
- Maurer, M. E., and Cooper, J. A. (2006). The adaptor protein Dab2 sorts LDL receptors into coated pits independently of AP-2 and ARH. *J. Cell Sci.* 119, 4235. doi: 10.1242/jcs.03217
- McMahon, H. T., and Boucrot, E. (2011). Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.* 12, 517–533. doi: 10.1038/nrm3151
- Miele, A. E., Watson, P. J., Evans, P. R., Traub, L. M., and Owen, D. J. (2004). Two distinct interaction motifs in amphiphysin bind two independent sites on the clathrin terminal domain [beta]-propeller. *Nat. Struct. Mol. Biol.* 11, 242–248. doi: 10.1038/nsmb736
- Milano, S. K., Pace, H. C., Kim, Y.-M., Brenner, C., and Benovic, J. L. (2002). Scaffolding functions of arrestin-2 revealed by crystal structure and mutagenesis. *Biochemistry* 41, 3321–3328. doi: 10.1021/bi015905j
- Mishra, S. K., Hawryluk, M. J., Brett, T. J., Keyel, P. A., Dupin, A. L., Jha, A., et al. (2004). Dual engagement regulation of protein interactions with the AP-2 adaptor α appendage. *J. Biol. Chem.* 279, 46191–46203. doi: 10.1074/jbc.M408095200
- Mishra, S. K., Keyel, P. A., Edeling, M. A., Dupin, A. L., Owen, D. J., and Traub, L. M. (2005). Functional dissection of an AP-2 β 2 appendage-binding sequence within the autosomal recessive hypercholesterolemia protein. *J. Biol. Chem.* 280, 19270–19280. doi: 10.1074/jbc.M501029200
- Mishra, S. K., Watkins, S. C., and Traub, L. M. (2002). The autosomal recessive hypercholesterolemia, (ARH) protein interfaces directly with the clathrin-coat machinery. *Proc. Natl. Acad. Sci.* 99, 16099–16104. doi: 10.1073/pnas.252630799
- Morgan, J. R., Prasad, K., Hao, W., Augustine, G. J., and Lafer, E. M. (2000). A conserved clathrin assembly motif essential for synaptic vesicle endocytosis. *J. Neurosci.* 20:8667.

- Muenzner, J., Traub, L. M., Kelly, B. T., and Graham, S. C. (2017). Cellular and viral peptides bind multiple sites on the N-terminal domain of clathrin. *Traffic* 18, 44–57. doi: 10.1111/tra.12457
- Ohno, H., Stewart, J., Fournier, M. C., Bosshart, H., Rhee, I., Miyatake, S., et al. (1995). Interaction of tyrosine-based sorting signals with clathrin-associated proteins. *Science* 269, 1872–1875. doi: 10.1126/science.7569928
- Owen, D. J., and Evans, P. R. (1998). A structural explanation for the recognition of tyrosine-based endocytotic signals. *Science* 282:1327. doi: 10.1126/science.282.5392.1327
- Owen, D. J., Setiadi, H., Evans, P. R., McEver, R. P., and Green, S. A. (2001). A third specificity-determining site in μ 2 adaptin for sequences upstream of Yxx Φ sorting motifs. *Traffic* 2, 105–110. doi: 10.1034/j.1600-0854.2001.020205.x
- Owen, D. J., Vallis, Y., Noble, M. E. M., Hunter, J. B., Dafforn, T. R., Evans, P. R., et al. (1999). A Structural explanation for the binding of multiple ligands by the α -adaptin appendage domain. *Cell* 97, 805–815. doi: 10.1016/S0092-8674(00)80791-6
- Owen, D. J., Vallis, Y., Pearse, B. M. F., McMahon, H. T., and Evans, P. R. (2000). The structure and function of the β 2-adaptin appendage domain. *EMBO J.* 19, 4216–4227. doi: 10.1093/emboj/19.16.4216
- Pollard, T. D. (2010). A guide to simple and informative binding assays. *Mol. Biol. Cell*, 21, 4061–4067. doi: 10.1091/mbc.E10-08-0683
- Praefcke, G. J. K., Ford, M. G. J., Schmid, E. M., Olesen, L. E., Gallop, J. L., Peak-Chew, S.-Y., et al. (2004). Evolving nature of the AP2 α -appendage hub during clathrin-coated vesicle endocytosis. *EMBO J.* 23, 4371–4383. doi: 10.1038/sj.emboj.7600445
- Ramjaun, A. R., and McPherson, P. S. (1998). Multiple amphiphysin II splice variants display differential clathrin binding: identification of two distinct clathrin-binding sites. *J. Neurochem.* 70, 2369–2376. doi: 10.1046/j.1471-4159.1998.70062369.x
- Rapoport, I., Miyazaki, M., Boll, W., Duckworth, B., Cantley, L. C., Shoelson, S., et al. (1997). Regulatory interactions in the recognition of endocytic sorting signals by AP-2 complexes. *EMBO J.* 16, 2240–2250. doi: 10.1093/emboj/16.9.2240
- Riccotta, D., Conner, S. D., Schmid, S. L., von Figura, K., and Höning, S. (2002). Phosphorylation of the AP2 μ subunit by AAK1 mediates high affinity binding to membrane protein sorting signals. *J. Cell Biol.* 156, 791–795. doi: 10.1083/jcb.200111068
- Ritter, B., Denisov, A. Y., Philie, J., Deprez, C., Tung, E. C., Gehring, K., et al. (2004). Two WXXF-based motifs in NECAPs define the specificity of accessory protein binding to AP-1 and AP-2. *EMBO J.* 23, 3701–3710. doi: 10.1038/sj.emboj.7600378
- Ritter, B., Philie, J., Girard, M., Tung, E. C., Blondeau, F., and McPherson, P. S. (2003). Identification of a family of endocytic proteins that define a new alpha-adaptin ear-binding motif. *EMBO Rep.* 4, 1089–1095. doi: 10.1038/sj.embor.7400004
- Rohrer, J., Schweizer, A., Russell, D., and Kornfeld, S. (1996). The targeting of Lamp1 to lysosomes is dependent on the spacing of its cytoplasmic tail tyrosine sorting motif relative to the membrane. *J. Cell Biol.* 132, 565–576. doi: 10.1083/jcb.132.4.565
- Roos, J., and Kelly, R. B. (1999). The endocytic machinery in nerve terminals surrounds sites of exocytosis. *Curr. Biol.* 9, 1411–1414. doi: 10.1016/S0960-9822(00)80087-1
- Schlosshauer, M., and Baker, D. (2004). Realistic protein–protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. *Protein Sci.* 13, 1660–1669. doi: 10.1110/ps.03517304
- Schmid, E. M., Ford, M. G. J., Burtay, A., Praefcke, G. J. K., Peak-Chew, S.-Y., Mills, I. G., et al. (2006). Role of the AP2 β -appendage hub in recruiting partners for clathrin-coated vesicle assembly. *PLoS Biol.* 4:e262. doi: 10.1371/journal.pbio.0040262
- Shih, W., Gallusser, A., and Kirchhausen, T. (1995). A clathrin-binding site in the hinge of the 2 chain of mammalian AP-2 complexes. *J. Biol. Chem.* 270, 31083–31090. doi: 10.1074/jbc.270.52.31083
- Slepnev, V. I., Ochoa, G.-C., Butler, M. H., and De Camilli, P. (2000). Tandem arrangement of the clathrin and AP-2 binding domains in amphiphysin 1 and disruption of clathrin coat function by amphiphysin fragments comprising these sites. *J. Biol. Chem.* 275, 17583–17589. doi: 10.1074/jbc.M910430199
- Sorkin, A., Mazzotti, M., Sorkina, T., Scotto, L., and Beguinot, L. (1996). Epidermal growth factor receptor interaction with clathrin adaptors is mediated by the Tyr974-containing internalization motif. *J. Biol. Chem.* 271, 13377–13384. doi: 10.1074/jbc.271.23.13377
- ter Haar, E., Harrison, S. C., and Kirchhausen, T. (2000). Peptide-in-groove interactions link target proteins to the β -propeller of clathrin. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1096–1100. doi: 10.1073/pnas.97.3.1096
- ter Haar, E., Musacchio, A., Harrison, S. C., and Kirchhausen, T. (1998). Atomic structure of clathrin: a β propeller terminal domain joins an α zigzag linker. *Cell* 95, 563–573. doi: 10.1016/S0092-8674(00)81623-2
- Traub, L. M. (2005). Common principles in clathrin-mediated sorting at the Golgi and the plasma membrane. *Biochim. Biophys. Acta* 1744, 415–437. doi: 10.1016/j.bbamcr.2005.04.005
- Traub, L. M. (2009). Tickets to ride: selecting cargo for clathrin-regulated internalization. *Nat. Rev. Mol. Cell Biol.* 10, 583–596. doi: 10.1038/nrm2751
- Traub, L. M., Downs, M. A., Westrich, J. L., and Fremont, D. H. (1999). Crystal structure of the α appendage of AP-2 reveals a recruitment platform for clathrin-coat assembly. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8907–8912. doi: 10.1073/pnas.96.16.8907
- Walther, K., Diril, M. K., Jung, N., and Haucke, V. (2004). Functional dissection of the interactions of stonin 2 with the adaptor complex AP-2 and synaptotagmin. *Proc. Natl. Acad. Sci. U.S.A.* 101, 964–969. doi: 10.1073/pnas.0307862100
- Wilde, A., and Brodsky, F. M. (1996). *In vivo* phosphorylation of adaptors regulates their interaction with clathrin. *J. Cell Biol.* 135, 635–645. doi: 10.1083/jcb.135.3.635
- Willox, A. K., and Royle, S. J. (2012). Functional analysis of interaction sites on the N-terminal domain of clathrin heavy chain. *Traffic* 13, 70–81. doi: 10.1111/j.1600-0854.2011.01289.x
- Zhang, F., Zang, T., Wilson, S. J., Johnson, M. C., and Bieniasz, P. D. (2011). Clathrin facilitates the morphogenesis of retrovirus particles. *PLoS Pathog.* 7:e1002119. doi: 10.1371/journal.ppat.1002119
- Zhuo, Y., Cano, K. E., Wang, L., Ilangovan, U., Hinck, A. P., Sousa, R., et al. (2015). Nuclear magnetic resonance structural mapping reveals promiscuous interactions between clathrin-box motif sequences and the N-terminal domain of the clathrin heavy chain. *Biochemistry* 54, 2571–2580. doi: 10.1021/acs.biochem.5b00065
- Zhuo, Y., Ilangovan, U., Schirf, V., Demeler, B., Sousa, R., Hinck, A. P., et al. (2010). Dynamic interactions between clathrin and locally structured elements in a disordered protein mediate clathrin lattice assembly. *J. Mol. Biol.* 404, 274–290. doi: 10.1016/j.jmb.2010.09.044

Conflict of Interest Statement: 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 © 2017 Smith, Baker, Halebian and Smith. 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) or licensor 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 Use of NMR to Study Transient Carbohydrate—Protein Interactions

Pedro M. Nieto*

Glycosystems Laboratory, Instituto de Investigaciones Científicas, cicCartuja, CSIC/USE, Seville, Spain

Carbohydrates are biologically ubiquitous and are essential to the existence of all known living organisms. Although they are better known for their role as energy sources (glucose/glycogen or starch) or structural elements (chitin or cellulose), carbohydrates also participate in the recognition events of molecular recognition processes. Such interactions with other biomolecules (nucleic acids, proteins, and lipids) are fundamental to life and disease. This review focuses on the application of NMR methods to understand at the atomic level the mechanisms by which sugar molecules can be recognized by proteins to form complexes, creating new entities with different properties to those of the individual component molecules. These processes have recently gained attention as new techniques have been developed, while at the same time old techniques have been reinvented and adapted to address newer emerging problems.

Keywords: NMR, protein–carbohydrate interaction, STD-NMR, transfer-NOESY, transient interactions

OPEN ACCESS

Edited by:

Rivka Isaacson,
King's College London,
United Kingdom

Reviewed by:

Gyula Batta,
University of Debrecen, Hungary
Peter J. Simpson,
Imperial College London,
United Kingdom

*Correspondence:

Pedro M. Nieto
pedro.nieto@iiq.csic.es

Specialty section:

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

Received: 08 January 2018

Accepted: 26 March 2018

Published: 11 April 2018

Citation:

Nieto PM (2018) The Use of NMR to
Study Transient
Carbohydrate—Protein Interactions.
Front. Mol. Biosci. 5:33.
doi: 10.3389/fmolb.2018.00033

INTRODUCTION

Glycobiology, which can be defined as the study of the structure, chemistry, biosynthesis, and biological functions of glycans and their derivatives, is fundamental in many critical biological processes (Varki, 1993, 2017). A challenge of structural glycobiology is to reconcile the large variety of 3D shapes that carbohydrates can assume with the high degree of selectivity found between closely related glycans (DeMarco and Woods, 2008). In addition, the binding constants between individual carbohydrates and proteins are generally low. Under these conditions, most of the interactions are in the fast exchange region of the NMR chemical shift scale and a single averaged set of signals is detected. This behavior can be exploited for structural elucidation. The rationale behind transient or transfer experiments is that if the equilibrium is fast enough and the property is dependent on the correlation time, it is possible to observe properties characteristic of the bound state in the averaged ligand signals, due to the larger correlation time of the bound form (Hyde et al., 1980; Feeney, 2000; Meyer and Peters, 2003).

In order to understand the evolution of NMR methods for the study of glycans, the scarcity of simple methods for obtaining isotopically-labeled glycans needs to be kept in mind. When such methods are available, the use of isotope-filtered/edited NMR experiments should be considered. When there is no such alternative, however, transfer techniques need to be employed. These can be applied to systems in fast equilibrium in the relaxation time scale and to properties weighted by the correlation time. Under these circumstances, the average of the NMR property obtained is biased toward the minor population of the bound carbohydrate due to its larger correlation time, even in the presence of an excess of free ligand. This is the basis of the transfer NOE technique and its analogs (Hyde et al., 1980; Ni, 1994). The advent of methodological improvements, monodimensional analogs with better signal to noise ratio, has effectively relaunched the transfer

techniques (Stott et al., 1997). Other parameters that can be used to study molecular complexes in a transient state are $T_{1\rho}$ (selective longitudinal relaxation time), or T_2 (transversal relaxation time) and ligand-detected ^1H relaxation dispersion because they depend on the correlation time, which is a function of the size of the molecule. Theoretical descriptions of T_1 and T_2 and of their application to the analysis of binding constants have been reviewed by Stockman and Dalvit (2002) and Peng et al. (Lepre et al., 2004).

A further experiment that benefits from the difference in the correlation times is the STD (saturation transfer difference) experiment (Mayer and Meyer, 1999; Meyer and Peters, 2003). This experiment uses the faster transfer of saturation from the receptor, caused by its long correlation time, that is further transferred to the ligand within the complex and that finally is observed in signals corresponding to the free-state. Since its formulation, the STD technique has experienced a rapid growth, both in its range of applications and in the number of labs using it. Finally, WaterLOGSY uses protonated water molecules to distinguish those that are in fast exchange from those that are buried in the interface with the receptor and therefore reflects the interaction surfaces between receptor and ligand.

NMR CHARACTERISTICS OF FREE CARBOHYDRATES

In general, the principal sources of glycans are *via* chemical synthesis or natural product isolation. In each case, they are almost completely restricted to non-isotopically-labeled compounds. As, from an NMR viewpoint, there are few experimental restrictions for unlabeled-carbohydrates, the analysis conditions must nevertheless be carefully documented and quantified.

Coupling Constants

Three-bond H – H coupling constants depend on the dihedral angles, and consequently they are key to determining the ring conformation. In general, most hexoses are monokonformational and are generally in a $^4\text{C}_1$ or $^1\text{C}_4$ chair conformation. Due to the cyclic nature of the sugars, many redundant interprotonic three-bond coupling constants are available and it is relatively simple to define the conformation from the Cremer–Pople polar coordinates (Cremer and Pople, 1975). Residual dipolar coupling constants provide also data about the relative orientation of two vectors in partially oriented media (Martin-Pastor and Bush, 2001). In carbohydrates, few examples of conformational equilibrium are known; however, one such example is the iduronate residue, which can exist in at least three conformations, $^1\text{C}_4$, $^4\text{C}_1$, and $^2\text{S}_0$, easily distinguishable by their coupling constants. Generally, internal iduronate residues are found in fast equilibrium between $^1\text{C}_4$ and $^2\text{S}_0$ (Figure 1A,B) (Mulloy et al., 1993). This equilibrium is fundamental in the molecular recognition events in which heparin is involved (Canales et al., 2005) and also has been studied by residual dipolar coupling (Jin et al., 2009).

NOE, Distances: The ISPA (Isolated Spin Pair Approach)

NOE (and other related parameters such as ROE, transverse NOE, and off-resonance ROE)-based distance constraints are often used to study the glycosidic angle behavior and can thus be used for the determination of the 3D shapes of saccharides. The scarcity of experimental restrictions in the analysis of carbohydrates implies a more accurate quantification of the NOE. Fortunately, for small molecules, distances are easily quantifiable from the NOE by means of an analysis of the growing NOE curve at several mixing times (Macura et al., 1986). Then, assuming equal motional behaviors for an undetermined distance and for a known reference, the initial NOE growth rate (cross-relaxation rate) relationship is inversely proportional to the sixth power of the interprotonic distance (Neuhaus and Williamson, 2000).

Applications of this methodology have increased since the arrival of modern selective 1D NOESY methods based on the double pulsed field gradients spin echo technique (Stott et al., 1997). This scheme has advantages over the bidimensional analogs; it is faster and has better resolution. Moreover, as the protons are relaxing by $T_{1\rho}$ instead of T_1 , the growing curves have longer linearity, and thus better-fitting and more precise results can be obtained (Hu and Krishnamurthy, 2006; Munoz-Garcia et al., 2013). The technique has also been applied to off-resonance ROESY, so as to obtain reference-independent distances. Using this approach, it has been possible to obtain the interglycosidic distances for a strongly anisotropic heparin hexasaccharide in which the interprotonic correlation time depends upon the orientation of the vector relative to the axis of the molecule (Munoz-Garcia et al., 2013).

NMR CHARACTERISTICS OF BOUND SMALL GLYCANS

The ability of NMR to observe and analyze specific signals of individual atoms, focusing analysis on a particular aspect of a complex without the need to solve its entire structure, is the most important consideration.

A bound ligand has NMR properties that are governed by its geometry in the bound state and these can therefore be different to those of the ligand in its free state. In a situation of fast equilibrium in the NMR timescale, the magnitude of the parameter that is measured corresponds to the weighted average of the NMR properties of the free and the bound states. For the special case of correlation time-dependent properties, the weight of the properties derived from the complexed ligand is so large that it is predominant in the averaged values, even at low molar ratio (Ni, 1994; Neuhaus and Williamson, 2000).

Transfer NOE (TR-NOE)

Most conformational analyses of sugars have been based on NOE data. For the determination of the structures of ligands bound to receptor proteins, transfer NOE (TR-NOE)-based experiments are an excellent tool (Ni, 1994). In suitable conditions, a fast equilibrium in the NMR timescale is established between the free ligand (a small to medium-sized fast-tumbling molecule

with a short correlation time) and the ligand bound to the high-molecular-weight receptor (a large and slow-tumbling molecule with a long correlation time) and therefore behaving as the receptor, with a large and fast-growing negative NOE.

This averaged situation is reflected in the observation of a large-molecule negative NOE in the signals of the ligand (see **Figures 1C,D**). In general, modifications of the magnetic field strength or changes in the sample conditions such as temperature

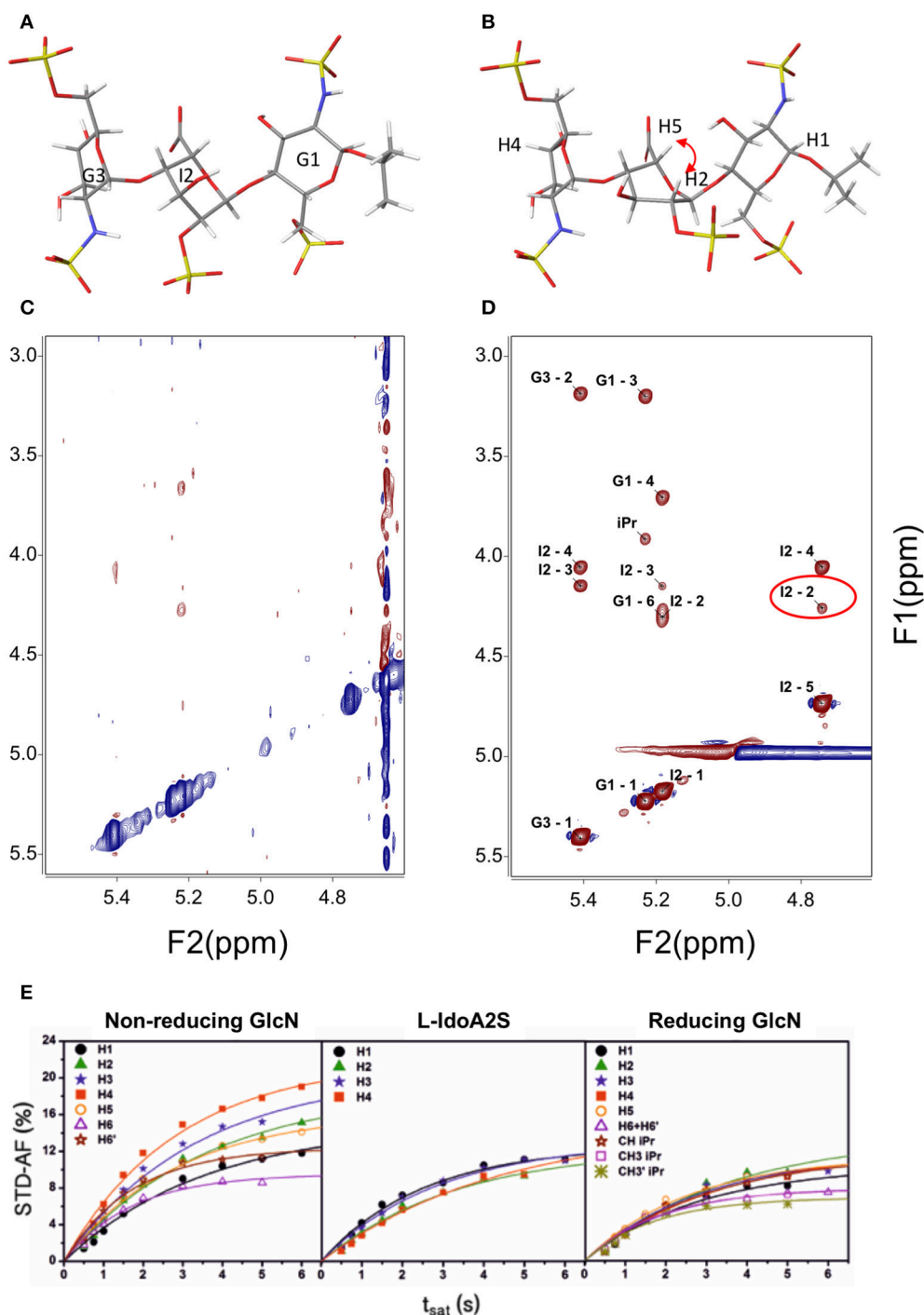


FIGURE 1 | Structure of a heparin trisaccharide with the central iduronate in: (A) 1C_4 and (B) 2S_0 conformation; (C) NOESY 500 MHz and 600 ms t_{mix} ; (D) transfer NOESY (800 MHz, 200 ms t_{mix})—the labels correspond to the assignment of cross-peaks—notice the peak I2 H2—H5, a NOESY peak exclusive to the 2S_0 conformation; and (E) STD—affinity factor growing curves for the extraction of the initial rates for quantitative analysis (Muñoz-García et al., 2015).

or viscosity can be used for fine adjustment if the free ligand is not clearly in the fast-tumbling regime.

In such a case, it is possible to obtain the bound ligand NOE by the analysis of the ligand signals (Ni, 1994; Neuhaus and Williamson, 2000). As a result, the qualitative analysis considering the changes between the free ligand conformation and that of the bound ligand is direct. On the other hand, the quantitative interpretation of the spectrum suffers from the potential effects of spin diffusion. In this situation, complete relaxation matrix calculations (CORCEMA or Mardigras (Borgias and James, 1990; Jayalakshmi and Krishna, 2002) should be used.

In principle, any experiment that measures interprotonic relaxation rates can be used for the analysis of the structure of the bound ligand. Thus, ROESY has been used to improve the experiment, taking advantage of its lower sensitivity to spin diffusion (Poveda et al., 1997). As the ROESY experiment suffers from some spurious effects, transverse-ROESY (Hwang and Shaka, 1992), is better suited as it is less prone to spin-lock artifacts. In addition, the quiet-NOESY experiment has been used. In this experiment, in the middle of the mixing time a biselective signal inversion is applied and a selective band inversion pulse is included, which avoids diffusion through the use of inverted spins (Zwahlen et al., 1994).

Special mention should be made of the monodimensional analogs (Stott et al., 1997). By selecting a single signal, not only is a better resolution achieved due to the change from 2D to 1D but a new dependence on $T_{1\text{sel}}$ is created and the NOE exhibits a longer linearity, a feature that is particularly useful when precise distances are required (Hu and Krishnamurthy, 2006).

STD-NMR

The STD-NMR (Mayer and Meyer, 1999) experiment consists of the difference between two experiments. It is undertaken with low-power irradiation during the relaxation delay on a sample in equilibrium that comprises a large excess of ligand(s) (from 10:1 up to 1,000:1) relative to the receptor (a large molecule), which is present at low concentration (nM to μM). In one experiment, recorded as the reference, dummy irradiation far from the signals is performed, while in the other experiment some signals from the receptor are selectively irradiated. If binding occurs, magnetization from the receptor is transferred to the ligand through close contacts with the receptor in the complex, and the effect will emerge in the difference experiment (Mayer and Meyer, 1999; Meyer and Peters, 2003).

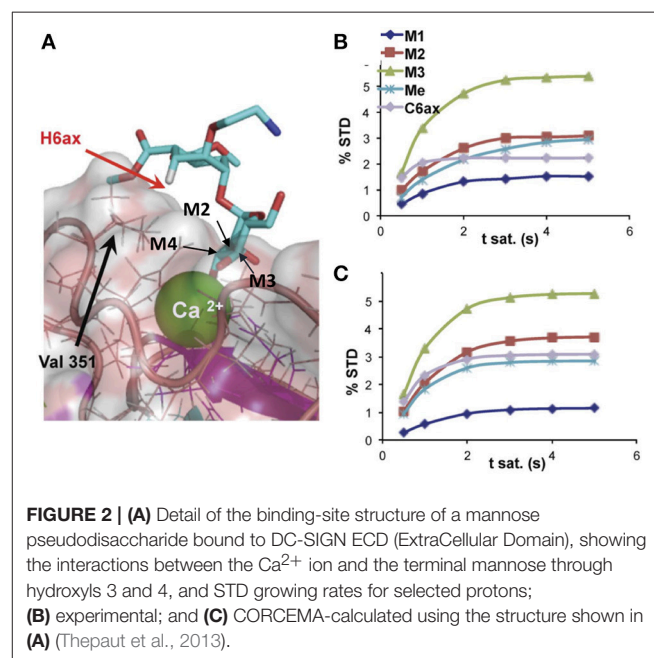
Since its formulation, several applications have been explored for STD-NMR. The first application proposed for the technique was ligand screening; this consists of the deconvolution of a library of potential binding molecules by using several compounds at once in each experiment (Henrichsen et al., 1999). The molecules that bind better will show STD signals. Epitope mapping was the next application proposed. The experiment can be used to determine the relative importance of diverse regions of the ligand in the interaction with the receptor (Mayer and Meyer, 2001). The technique can also be used for affinity constant evaluation. The first application in which calculation of

a binding constant was described used the Cheng and Prusoff equation, since the STD values are biased by the relaxation properties of the ligand proton being considered (Meyer and Peters, 2003). Latterly, the initial growth rate of the STD affinity factor has been used, in order to avoid this dependence upon the relaxation of the protons, see **Figure 1E** (Angulo et al., 2010). One of the potential complications in the STD is peak overlap that can make difficult the interpretation of the results. One of the solutions relies on adding another transfer step, then the peaks will be spread in two dimensions: STD-TOCSY (Mayer and Meyer, 1999) and STD-HMQC (Vogtherr and Peters, 2000). Similarly, spectral editing can also be performed using other nuclei ^{15}N or ^{19}F (Kövér et al., 2007; Diercks et al., 2009).

Using methods based on the relaxation matrix, it is possible to calculate the theoretical STD of a given complex. Thus, a quantitative analysis of the ligand within the complex can be carried out using a 3D structure. Based on a 3D model of the complex, using some NMR parameters and considering the saturation time and the irradiation frequency among others, CORCEMA-ST can estimate the different values of saturation transferred from the protein to a particular proton of the ligand (Jayalakshmi and Krishna, 2004, 2005). An example is given in **Figure 2**. This methodology has also been extended to the case of multiple binding modes within the same binding site (Angulo et al., 2008) and to the iterative refinement of a complex structure (Jayalakshmi and Krishna, 2005).

Selective T_1 ($T_{1\text{sel}}$) and T_2

$T_{1\text{sel}}$ and T_2 also depend on correlation times and in principle they can be used to study association processes for fast equilibrium between ligand and receptor, observing the



ligand signals at low concentration of the receptor. These values can be used for the analysis of binding constants by performing a titration (García-Jiménez et al., 2017). A series of $T_{1\rho}/T_2$ values are obtained at different ratios of ligand to receptor, and by fitting these to the variation in the ligand/receptor ratio, K_D or IC_{50} can be extracted. Recently the application of 1H relaxation dispersion measurements has been described for weak binding systems using both T_2 and $T_{1\rho}$ (Moschen et al., 2015; Trigo-Mouriño et al., 2017).

WaterLOGSY

The WaterLOGSY experiment (Dalvit et al., 2000) takes into account explicitly the water molecules surrounding the macromolecular complex. It is easy to implement and consists essentially of a selective excitation of the water molecules in a sample prepared in H_2O , followed by a NOESY mixing period and a water suppression scheme for detection (Dalvit et al., 2001).

This experiment traces the mechanisms that follow the magnetization from excited water to ligand protons. It can be a direct transfer of magnetization from bulk water to ligand, through exchangeable protons from the receptor or via water molecules buried into the receptor interface. These molecules have tumbling times closer to the receptor than to the bulk water or ligand. Due to the slow correlation times of the bound ligand or the buried water molecules the peaks through the two later mechanisms will have the opposite sign as the direct transfer of magnetization of the free compounds (Stockman and Dalvit, 2002).

The WaterLOGSY competition experiment is used for ligand library deconvolution, in which several relay pathways are used constructively, transferring bulk water magnetization to the ligand in a selective manner (Dalvit et al., 2002). In this experiment, the resonances of non-binding compounds appear with opposite sign, corresponding to NOE in the narrowing limit, and they tend to be weaker than those of the interacting ligands.

NEW DEVELOPMENTS

A recent application of the WaterLOGSY is the LOGSY titration (Geist et al., 2017). From the slopes of the WaterLOGSY vs the protein concentration curves, LOGSY-titration factors can be obtained, that can be used for epitope mapping alternative to STD-AF₀. In this case, a great advantage is that isolated water molecules are always in the narrowing limit of the NOE curve.

Complementary information can be also obtained using paramagnetic tags attached to carbohydrates. At least two methods can be exploited: PCS, where changes in chemical shifts are produced and PRE, where additional relaxation is caused. In both cases the effects are proportional to the distance to the paramagnetic center. The paramagnetic center can be either a Lanthanide (Gao et al., 2016; Canales et al., 2017) or nitroxide carrying spin labels like TEMPO.

INTEGRATING THE RESULTS. FROM NMR DATA TO 3D STRUCTURES

The process of deriving a 3D structure from NMR data is not straightforward and it requires the application of force field modeling, with experimental constraints taken from the experimental data. For carbohydrates, accurate distances have to be used and complemented by high-quality theoretical data; this is because an accurate description of the flexibility is required. This often implies the use of explicit solvent MD (molecular dynamics) calculations for the free ligand.

Several force field programs have been developed for carbohydrates, the most widely used being Amber, making use of the parameters for carbohydrates developed by Woods (GLYCAM) (Kirschner et al., 2008); CHARMM (Brooks et al., 2009), optimized by Mackerel; and GROMACS (Hess et al., 2008). These force fields are compatible with other force fields developed for proteins and peptides and can be used in mixed systems calculations.

Using the experimental NMR data from a complex, we could obtain a 3D structure of the ligand within the complex using transfer NOESY evaluation, or at least qualitative data relative to the potential conformational changes upon binding (Ni, 1994). Finally, CORCEMA-ST software is capable of calculating the STD values for individual protons (Krishna and Jayalakshmi, 2008), and of verifying the bound ligand conformation. STD₀-AF have been used for the definition of the structure, position, and conformation of the ligand within the complex (Enríquez-Navas et al., 2011). This approach can also be used to tackle more complex situations, such as multiple binding modes in one or more than one site (Angulo et al., 2008). Finally, it has been used to refine NMR structures, using STD-NMR intensity-restrained optimization (Jayalakshmi and Krishna, 2005).

FUTURE PERSPECTIVES

NMR provides tools for the analysis and study of low-resolution 3D structures of protein-carbohydrate complexes. A particular advantage is that there are no upper limits to the size of the receptor and the analyses can be performed even with complete cells. In order to improve the reliability of the results, however, efforts should be made to achieve an integrated semi-quantitative method combining all the transfer NMR techniques that can deliver spatial resolution, namely transfer NOESY, STD-NMR, and WaterLOGSY. Recently, promising new methods based on PCS (Pseudo Contact Shifts) have emerged. In the case of the lanthanide paramagnetic tags using PCS they can be applied to protein-carbohydrate complexes with the advantage of the larger dispersion of the signals induced by the lanthanide linked to the carbohydrate (Canales et al., 2017) or the protein (Gao et al., 2016).

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

FUNDING

Research described in this paper has been performed with financial support from MINECO grants (CTQ2012-32605, CTQ2015-70134-P), Junta de Andalucía (FQM-1303), co-financed by European Regional Development Funds (ERDF). The author acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative

REFERENCES

- Angulo, J., Diaz, I., Reina, J. J., Tabarani, G., Fieschi, F., Rojo, J., et al. (2008). Saturation Transfer Difference (STD) NMR spectroscopy characterization of dual binding mode of a mannose disaccharide to DC-SIGN. *ChemBioChem* 9, 2225–2227. doi: 10.1002/cbic.200800361
- Angulo, J., Enriquez-Navas, P. M., and Nieto, P. M. (2010). Ligand-Receptor Binding Affinities from Saturation Transfer Difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. *Chemistry* 16, 7803–7812. doi: 10.1002/chem.200903528
- Borgias, B. A., and James, T. L. (1990). Mardigras - A procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure. *J. Magn. Reson.* 87, 475–487.
- Brooks, B. R., Brooks, C. L. III, Mackerell, A. D. Jr., Nilsson, L., Petrella, R. J., Roux, B., et al. (2009). CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30, 1545–1614. doi: 10.1002/jcc.21287
- Canales, A., Angulo, J., Ojeda, R., Bruix, M., Fayos, R., Lozano, R., et al. (2005). Conformational flexibility of a synthetic glycosylaminoglycan bound to a fibroblast growth factor. FGF-1 recognizes both the (1)C(4) and (2)S(O) conformations of a bioactive heparin-like hexasaccharide. *J. Am. Chem. Soc.* 127, 5778–5779. doi: 10.1021/ja043363y
- Canales, A., Boos, I., Perkams, L., Karst, L., Luber, T., Karagiannis, T., et al. (2017). Breaking the limits in analyzing carbohydrate recognition by NMR spectroscopy: resolving branch-selective interaction of a tetra-antennary N-glycan with lectins. *Angew. Chem. Int.* 56, 14987–14991. doi: 10.1002/anie.201709130
- Cremer, D., and Pople, J. A. (1975). General definition of ring puckering coordinates. *J. Am. Chem. Soc.* 97, 1354–1358. doi: 10.1021/ja00839a011
- Dalvit, C., Flocco, M., Knapp, S., Mostardini, M., Perego, R., Stockman, B. J., et al. (2002). High-throughput NMR-based screening with competition binding experiments. *J. Am. Chem. Soc.* 124, 7702–7709. doi: 10.1021/ja020174b
- Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M., and Stockman, B. (2001). WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *J. Biomol. NMR* 21, 349–359. doi: 10.1023/A:1013302231549
- Dalvit, C., Pevarello, P., Tato, M., Veronesi, M., Vulpetti, A., and Sundstrom, M. (2000). Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J. Biomol. NMR* 18, 65–68. doi: 10.1023/A:1008354229396
- DeMarco, M. L., and Woods, R. J. (2008). Structural Glycobiology: a game of snakes and ladders. *Glycobiology* 18, 426–440. doi: 10.1093/glycob/cwn026
- Diercks, T., Ribeiro, J. P., Cañada, F. J., André, S., Jiménez-Barbero, J., and Gabius, H.-J. (2009). Fluorinated carbohydrates as lectin ligands: versatile Sensors in 19F-detected saturation transfer difference NMR spectroscopy. *Chemistry* 15, 5666–5668. doi: 10.1002/chem.200900168
- Enriquez-Navas, P. M., Marradi, M., Padro, D., Angulo, J., and Penadés, S. (2011). A solution NMR study of the interactions of oligomannosides and the Anti-HIV-1 2G12 antibody reveals distinct binding modes for Branched Ligands*. *Chemistry* 17, 1547–1560. doi: 10.1002/chem.201002519
- Feeney, J. (2000). NMR studies of ligand binding to dihydrofolate reductase. *Angew. Chem.* 39, 290–312. doi: 10.1002/(SICI)1521-3773(20000117)39:2<290::AID-ANIE290>3.0.CO;2-1
- Gao, Q., Chen, C. Y., Zong, C., Wang, S., Ramiah, A., Prabhakar, P., et al. (2016). Structural Aspects of Heparan Sulfate Binding to Robo1-Ig1-2. *ACS Chem. Biol.* 11, 3106–3113. doi: 10.1021/acscchembio.6b00692
- through its Unit of Information Resources for Research (URICI).
- ## ACKNOWLEDGMENTS
- The author is grateful to his colleagues (J. L. de Paz, J. Rojo, S. Gil-Caballero, M. J. Garcia, and M. Torres) for their careful reading of the manuscript and their participation in the studies cited.
- García-Jiménez, M. J., Gil-Caballero, S., Canales, Á., Jiménez-Barbero, J., de Paz, J. L., and Nieto, P. M. (2017). Interactions between a heparin trisaccharide library and FGF-1 analyzed by NMR methods. *Int. J. Mol. Sci.* 18:1293. doi: 10.3390/ijms18061293
- Geist, L., Mayer, M., Cockcroft, X.-L., Wolkerstorfer, B., Kessler, D., Engelhardt, H., et al. (2017). Direct NMR probing of hydration shells of protein ligand interfaces and its application to drug design. *J. Med. Chem.* 60, 8708–8715. doi: 10.1021/acs.jmedchem.7b00845
- Henrichsen, D., Ernst, B., Magnani, J. L., Wang, W. T., Meyer, B., and Peters, T. (1999). Bioaffinity NMR spectroscopy: identification of an E-selectin antagonist in a substance mixture by transfer NOE. *Angew. Chem. Int.* 38, 98–102.
- Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theor. Comput.* 4, 435–447. doi: 10.1021/ct700301q
- Hu, H., and Krishnamurthy, K. (2006). Revisiting the initial rate approximation in kinetic NOE measurements. *J. Magn. Reson.* 182, 173–177. doi: 10.1016/j.jmr.2006.06.009
- Hwang, T. L., and Shaka, A. J. (1992). Cross relaxation without TOCSY - transverse rotating frame overhauser spectroscopy. *J. Am. Chem. Soc.* 114, 3157–3159. doi: 10.1021/ja00034a083
- Hyde, E. I., Birdsall, B., Roberts, G. C. K., Feeney, J., and Burgen, A. S. V. (1980). Proton nuclear magnetic-resonance saturation transfer studies of coenzyme binding to lactobacillus-casei dihydrofolate-reductase. *Biochemistry* 19, 3738–3746. doi: 10.1021/bi00557a015
- Jayalakshmi, V., and Krishna, N. R. (2002). Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *J. Magn. Reson.* 155, 106–118. doi: 10.1006/jmr.2001.2499
- Jayalakshmi, V., and Krishna, N. R. (2004). CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities. *J. Magn. Reson.* 168, 36–45. doi: 10.1016/j.jmr.2004.01.017
- Jayalakshmi, V., and Krishna, N. R. (2005). Determination of the conformation of trimethoprim in the binding pocket of bovine dihydrofolate reductase from a STD-NMR intensity-restrained CORCEMA-ST optimization. *J. Am. Chem. Soc.* 127, 14080–14084. doi: 10.1021/ja054192f
- Jin, L., Hricovini, M., Deakin, J. A., Lyon, M., and Uhrin, D. (2009). Residual dipolar coupling investigation of a heparin tetrasaccharide confirms the limited effect of flexibility of the iduronic acid on the molecular shape of heparin. *Glycobiology* 19, 1185–1196. doi: 10.1093/glycob/cwp105
- Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L., et al. (2008). GLYCAM06: a generalizable Biomolecular force field. *Carbohydrates. J. Comput. Chem.* 29, 622–655. doi: 10.1002/jcc.20820
- Kövr, K. E., Groves, P., Jiménez-Barbero, J., and Batta, G. (2007). Molecular recognition and screening using a 15N group selective STD NMR method. *J. Am. Chem. Soc.* 129, 11579–11582. doi: 10.1021/ja0732911
- Krishna, N. R., and Jayalakshmi, V. (2008). Quantitative analysis of STD-NMR spectra of reversibly forming ligand-receptor complexes. *Top. Curr. Chem.* 15–54. doi: 10.1007/128_2007_144
- Lepre, C. A., Moore, J. M., and Peng, J. W. (2004). Theory and applications of NMR-based screening in pharmaceutical research. *Chem. Rev.* 104, 3641–3675. doi: 10.1021/cr030409h
- Macura, S., Farmer, B. T., and Brown, L. R. (1986). An Improved method for the determination of cross-relaxation rates from NOE data. *J. Magn. Reson.* 70, 493–499. doi: 10.1016/0022-2364(86)90143-5

- Martin-Pastor, M., and Bush, C. A. (2001). Refined structure of a flexible heptasaccharide using ^1H - ^{13}C and ^1H - ^1H NMR residual dipolar couplings in concert with NOE and long range scalar coupling constants. *J. Biomol. NMR* 19, 125–139. doi: 10.1023/A:1008327926009
- Mayer, M., and Meyer, B. (1999). Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chem. Int.* 38, 1784–1788.
- Mayer, M., and Meyer, B. (2001). Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* 123, 6108–6117. doi: 10.1021/ja0100120
- Meyer, B., and Peters, T. (2003). NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew. Chem. Int.* 42, 864–890. doi: 10.1002/anie.200390233
- Moschen, T., Wunderlich, C. H., Spitzer, R., Levic, J., Micura, R., Tollinger, M., et al. (2015). Ligand-detected relaxation dispersion NMR spectroscopy: dynamics of preQ1–RNA Binding. *Angew. Chem. Int.* 54, 560–563. doi: 10.1002/ange.201409779
- Mulloy, B., Forster, M. J., Jones, C., and Davies, D. B. (1993). Nmr and molecular-modeling studies of the solution conformation of heparin. *Biochem. J.* 293, 849–858. doi: 10.1042/bj2930849
- Muñoz-García, J. C., Chabrol, E., Vivès, R. R., Thomas, A., de Paz, J. L., Rojo, J., et al. (2015). Langerin–heparin interaction: two binding sites for small and large ligands as revealed by a combination of NMR spectroscopy and cross-linking mapping experiments. *J. Am. Chem. Soc.* 137, 4100–4110. doi: 10.1021/ja511529x
- Munoz-Garcia, J. C., Solera, C., Carrero, P., de Paz, J. L., Angulo, J., and Nieto, P. M. (2013). 3D structure of a heparin mimetic analogue of a FGF-1 activator. A NMR and molecular modelling study. *Organ. Biomol. Chem.* 11, 8269–8275. doi: 10.1039/c3ob41789a
- Neuhaus, D., and Williamson, M. P. (2000). *The Nuclear Overhauser Effect in Structural and Conformational Analysis*. New York, NY: Wiley-VCH.
- Ni, F. (1994). Recent developments in transferred NOE methods. *Prog. Nuclear Magn. Reson. Spectrosc.* 26(Pt 6), 517–606. doi: 10.1016/0079-6565(94)90000-0
- Poveda, A., Asensio, J. L., MartinPastor, M., and JimenezBarbero, J. (1997). Solution conformation and dynamics of a tetrasaccharide related to the Lewis(x) antigen deduced by NMR relaxation measurements. *J. Biomol. NMR* 10, 29–43. doi: 10.1023/A:1018395627017
- Stockman, B. J., and Dalvit, C. (2002). NMR screening techniques in drug discovery and drug design. *Prog. Nuclear Magn. Reson. Spectrosc.* 41, 187–231. doi: 10.1016/S0079-6565(02)00049-3
- Stott, K., Keeler, J., Van, Q. N., and Shaka, A. J. (1997). One-dimensional NOE experiments using pulsed field gradients. *J. Magn. Reson.* 125, 302–324. doi: 10.1006/jmre.1997.1110
- Thepaut, M., Guzzi, C., Sutkeviciute, I., Sattin, S., Ribeiro-Viana, R., Varga, N., et al. (2013). Structure of a glycomimetic ligand in the carbohydrate recognition domain of C-type Lectin DC-SIGN. Structural requirements for selectivity and ligand design. *J. Am. Chem. Soc.* 135, 2518–2529. doi: 10.1021/ja3053305
- Trigo-Mouriño, P., Griesinger, C., and Lee, D. (2017). Label-free NMR-based dissociation kinetics determination. *J. Biomol. NMR* 69, 229–235. doi: 10.1007/s10858-017-0150-5
- Varki, A. (1993). Biological roles of oligosaccharides - all of the theories are correct. *Glycobiology* 3, 97–130. doi: 10.1093/glycob/3.2.97
- Varki, A. (2017). Biological roles of glycans. *Glycobiology* 27, 3–49. doi: 10.1093/glycob/cww086
- Vogtherr, M., and Peters, T. (2000). Application of NMR based binding assays to identify key hydroxy groups for intermolecular recognition. *J. Am. Chem. Soc.* 122, 6093–6099. doi: 10.1021/ja0001916
- Zwahlen, C., Vincent, S. J. F., Dibari, L., Levitt, M. H., and Bodenhausen, G. (1994). Quenching spin-diffusion in selective measurements of transient overhauser effects in nuclear magnetic resonance. Applications to Oligonucleotides. *J. Am. Chem. Soc.* 116, 362–368. doi: 10.1021/ja00080a041

Conflict of Interest Statement: The author declares 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 © 2018 Nieto. 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 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 and Functional Insights Into Lysostaphin–Substrate Interaction

Helena Tossavainen¹, Vytas Raulinaitis², Linda Kauppinen³, Ulla Pentikäinen^{3,4,5}, Hannu Maaheimo⁶ and Perttu Permi^{1,3*}

¹ Department of Chemistry, Nanoscience Center, University of Jyväskylä, Jyväskylä, Finland, ² Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, Helsinki, Finland, ³ Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland, ⁴ Institute of Biomedicine, University of Turku, Turku, Finland, ⁵ Turku Centre for Biotechnology, Turku, Finland, ⁶ VTT Technical Research Centre of Finland Ltd., Espoo, Finland

OPEN ACCESS

Edited by:

Maria Rosaria Conte,
King's College London,
United Kingdom

Reviewed by:

Filippo Prischi,
University of Essex, United Kingdom
Vladimir I. Polshakov,
Lomonosov Moscow State University,
Russia

*Correspondence:

Perttu Permi
permtu.permi@jyu.fi

Specialty section:

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

Received: 19 April 2018

Accepted: 12 June 2018

Published: 03 July 2018

Citation:

Tossavainen H, Raulinaitis V,
Kauppinen L, Pentikäinen U,
Maaheimo H and Permi P (2018)
Structural and Functional Insights Into
Lysostaphin–Substrate Interaction.
Front. Mol. Biosci. 5:60.
doi: 10.3389/fmolb.2018.00060

Lysostaphin from *Staphylococcus simulans* and its family enzymes rapidly acquire prominence as the next generation agents in treatment of *S. aureus* infections. The specificity of lysostaphin is promoted by its C-terminal cell wall targeting domain selectivity toward pentaglycine bridges in *S. aureus* cell wall. Scission of these cross-links is carried out by its N-terminal catalytic domain, a zinc-dependent endopeptidase. Understanding the determinants affecting the efficiency of catalysis and strength and specificity of interactions lies at the heart of all lysostaphin family enzyme applications. To this end, we have used NMR, SAXS and molecular dynamics simulations to characterize lysostaphin structure and dynamics, to address the inter-domain interaction, the enzyme-substrate interaction as well as the catalytic properties of pentaglycine cleavage in solution. Our NMR structure confirms the recent crystal structure, yet, together with the molecular dynamics simulations, emphasizes the dynamic nature of the loops embracing the catalytic site. We found no evidence for inter-domain interaction, but, interestingly, the SAXS data delineate two preferred conformation subpopulations. Catalytic H329 and H360 were observed to bind a second zinc ion, which reduces lysostaphin pentaglycine cleaving activity. Binding of pentaglycine or its lysine derivatives to the targeting domain was found to be of very low affinity. The pentaglycine interaction site was located to the N-terminal groove of the domain. Notably, the targeting domain binds the peptidoglycan stem peptide Ala-D-γ-Glu-Lys-D-Ala-D-Ala with a much higher, micromolar affinity. Binding site mapping reveals two interaction sites of different affinities on the surface of the domain for this peptide.

Keywords: lysostaphin, NMR structure, pentaglycine, peptidoglycan, protein dynamics, SH3b domain, *Staphylococcus aureus*, substrate binding

INTRODUCTION

The gram-positive pathogen *Staphylococcus aureus* causes numerous diseases, ranging from minor skin abscesses to meningitis and even toxic shock syndrome (Lowy, 1998). *S. aureus* bacteremia is estimated to be fatal in up to 30 % of cases (van Hal et al., 2012) and annually incur billions in overall medical costs just in the US (Lee et al., 2013) and hundreds of millions of additional hospitalization

costs in the EU (Köck et al., 2010). The problem is exacerbated by the pathogen being carried by one third of human population (Chambers and Deleo, 2009) and its vigor to develop resistance against drugs mobilized for its treatment. Hence, hospital-acquired *S. aureus* cases are particularly severe.

For seven decades treatment of *S. aureus* infections primarily relied on antibiotics interfering with the synthesis of its cell wall, specifically, the penicillin methicillin and the glycopeptide vancomycin. Inhibitors of protein synthesis have been recently added to the arsenal of therapeutic means (Brickner et al., 2008). However, the historic propensity of bacteria to rapidly develop resistance to small molecules impels the conception of new treatment paradigms (Taubes, 2008). Today *S. aureus* appears as one of the 12 most health-threatening pathogens in the list of resistant bacteria compiled by WHO to promote development of new antibiotics (WHO Mediacentre, 2017).

Pentaglycine cross-bridges of peptidoglycan (PG) are a distinct feature of the *S. aureus* cell wall and provide innate selectivity as targets (Vollmer et al., 2008). Some members of the *Staphylococcus* genus have taken advantage of this feature and secrete bactericidal enzymes which cleave this cross-bridge to lyse competing *S. aureus* (Schindler and Schuhardt, 1964; Browder et al., 1965; Sugai et al., 1997), with their own intrinsic resistance engineered by serine replacements of the glycine residues (Thumm and Götz, 1997; Tschierske et al., 1997). Scission of pentaglycine is executed by endopeptidases of the lysostaphin family, which also includes *S. aureus* inherent autolytic enzymes that maintain the integrity of the cell wall during cell life cycle (Ramadurai et al., 1999; Ercoli et al., 2015; Raulinaitis et al., 2017a).

A modular organization is common to many peptidoglycan hydrolases (Szweda et al., 2012). Mature lysostaphin consists of an N-terminal catalytic (CAT) domain, and a C-terminal cell wall targeting (CWT) domain (Baba and Schneewind, 1996). Its X-ray crystal structure has been reported recently (Sabala et al., 2014). The CAT domain forms a barrel fold typical for the MEROPS M23 family zinc-dependent endopeptidases, which bear conserved catalytic as well as Zn²⁺-coordinating amino acid residues (Rawlings et al., 2008). A fourteen-residue linker connects the CAT domain to the CWT domain. The latter assumes the fold corresponding to SH3b domains (Lu et al., 2006). The SH3b domains are the bacterial counterpart of the ubiquitous SH3 domains found in eukaryotes (Ponting et al., 1999; Whisstock and Lesk, 1999). These are small interaction modules present in proteins of signaling pathways, which interact with proline-rich sequences in binding partners (Saksela and Permi, 2012). The canonical PXXP binding site on the surface of SH3 domains is partially blocked in the SH3b domains, and the pentaglycine interaction site of SH3b domains has been located to a groove structure formed by a 20-residue N-terminal extension not present in the smaller SH3 domains (Lu et al., 2006; Hirakawa et al., 2009; Gu et al., 2014). Despite the linker, evidence for CAT–CWT domain interaction *in vitro* has been presented, although the same study underscored the benefits of flexibility between the domains *in vivo* (Lu et al., 2013).

Surprisingly little is known about the details of the interaction between lysostaphin and PG. Biochemical characterization has

revealed that the CWT domain is indispensable for lysostaphin specificity, and that an intact pentaglycine cross-bridge is crucial for the CWT–PG interaction (Gründling and Schneewind, 2006). The interaction can be inhibited by PG fragments containing multiple cross-linked murein subunits, but not with smaller murein mono- or dimers. Lipoproteins, cell wall-anchored proteins and wall teichoic acids appear not to have a role in the interaction (Gründling and Schneewind, 2006). Similarly, PG binding of *Staphylococcus capitis* ALE-1 CWT domain, which is highly similar in sequence and structure to the lysostaphin CWT domain, is dependent on the length and amino acid composition of the cross-bridge, showing strong preference for Gly₅ bridges (Lu et al., 2006). The complex structure of lysostaphin CWT domain with pentaglycine peptide is a landmark in the characterization of the SH3b–PG interaction (PDB ID 5LEO). In this structure pentaglycine resides in the N-terminal groove, consistent with previous mutational, computational and NMR studies.

Medical applications of recombinant lysostaphin are advancing and some have entered clinical trials (Nelson et al., 2012). Lysostaphin CWT domain fusion proteins have also been considered as a promising source of alternative chemotherapeutics (Schmelcher et al., 2012; Osipovitch and Griswold, 2015; Jagielska et al., 2016). In addition, lysostaphin has been engineered to suppress its immunogenic potential (Blazanovic et al., 2015). It is clear that the key factor in the success of all lysostaphin family enzyme applications is understanding the determinants affecting efficiency of catalysis and strength and specificity of interactions. To this end, we have used NMR, SAXS and molecular dynamics (MD) simulations to characterize lysostaphin structure and dynamics, to address the inter-domain interaction, the enzyme-substrate interaction as well as the catalytic properties of pentaglycine cleavage in solution.

MATERIALS AND METHODS

Production and Purification of Lysostaphin and Its CWT Domain

Gene for mature lysostaphin (residues 251–493) was synthesized de novo at GenScript (NJ, USA) and cloned into pGEX-2T vector (GE Healthcare Life Sciences) at the BamHI and EcoRI cloning sites. Resulting plasmid was transformed into *E. coli* BL21 (DE3) strain. DNA encoding the CWT domain (residues 402–493) was amplified (primers 5′-CGCGGATCCTGGAAAACCAATAAATATGGCACGC-3′ and 3′-CGCGAATTCTTATTTGA TGGTGC CCCACAGG-5′) using this plasmid as template and was cloned into pGEX-2T vector separately at the BamHI and EcoRI cloning sites as well.

Protein expression and purification was carried out as described in detail previously (Raulinaitis et al., 2017a). Overnight cell cultures were inoculated into a fresh medium. LB medium was used for non-labeled protein production and when labeled protein was required for NMR data collection, cells were grown in minimal resource M9 medium (40 mM Na₂HPO₄, 22 mM KH₂PO₄, 12 mM NaCl, 1 mM MgSO₄, 100 μM CaCl₂,

10 μ M FeSO₄, 0.001 % w/w biotin, and 0.1 mg/ml ampicillin, pH 7.0). Upon cells reaching OD₆₀₀, protein expression was induced with isopropyl β -D-1-thiogalactopyranoside (IPTG), 0.5 mM final concentration, and further incubated for 3 h. Harvested cells were disrupted by sonication, and soluble fraction was loaded on Glutathione Sepharose 4 Fast Flow resin (GE Healthcare). Lysostaphin was released from resin-bound glutathione tag by thrombin and protein was further purified by gel filtration on HiLoad 16/60 Superdex 75 column (GE Healthcare Life Sciences). Removal of innately bound metal ions was carried out with EDTA, which was thereafter dialyzed away in PBS buffer and protein was concentrated with Amicon[®] 5000 MWCO centrifugal filters.

NMR Spectroscopy

Spectra for chemical shift assignment and structure determination were acquired at 308 K on a Varian INOVA 800 MHz spectrometer equipped with a cryogenically cooled triple resonance probehead equipped with an actively shielded z-gradient. Resonance assignment was carried out with the following set of experiments: ¹H, ¹⁵N HSQC, ¹H, ¹³C HSQC, ¹H, ¹³C CT-HSQC, HNCACB, CBCA(CO)NH, (H)CC(CO)NH, H(CCO)NH, HBHA(CO)NH, (HB)CB(CGCD)HD, (HB)CB(CGCDCE)HE, ¹H, ¹⁵N NOESY-HSQC, ¹H, ¹³C NOESY-HSQC in 7% D₂O/93% H₂O, and ¹H, ¹³C HSQC-NOESY in 100% D₂O. The latter five spectra were employed for the assignment of the aromatic residues as well as the source of distance restraints for structure determination. Additional structural restraints were derived from the chemical shifts in the form of TALOS-N (Shen and Bax, 2013) dihedral angle restraints and applied for residues in secondary structures as well as H₂O to D₂O exchange experiment in the form of hydrogen bond restraints for amides experiencing slow exchange. Histidine tautomeric state could be derived from the peak pattern in a ¹H, ¹⁵N HMBC spectrum or the C δ 2-C ϵ 1 chemical shift difference for four out of nine histidines (Barraud et al., 2012). Spectra were processed with NMRPipe (Delaglio et al., 1995) and analyzed with Sparky (T. D. Goddard and D. G. Kneller, University of California, San Francisco). The automated NOE peak assignment-structure determination routine of CYANA 2.1 (López-Méndez and Güntert, 2006) was used to generate an ensemble of 20 lowest target energy structures which were further refined with AMBER 16 (Case et al., 2005) in explicit solvent. Fifteen lowest AMBER energy structures were chosen to represent the solution state structure of lysostaphin.

The pentaglycine binding site on the CWT domain was determined by comparing the ¹H, ¹⁵N, and ¹H, ¹³C HSQC spectra of full-length lysostaphin and CWT domain acquired in the presence and absence of a large amount, ~170 molar excess, of KGGGGG and GGGGGK hexapeptides (Sigma-Aldrich). Binding characteristics of a PG peptide mimic were determined by titration of the CWT domain with A-D-EKGGGGGA-D-EK-D-A and comparing ¹H, ¹⁵N HSQC spectra acquired at approximate peptide to CWT domain molar ratios of 0, 0.5, 1, 2, 5, 10, and 24. Similarly, the pentapeptide A-D- γ -EK-D-A-D-A (Bachem) and muramyl dipeptide *N*-acetylmuramyl-L-alanyl-D-isoglutamine hydrate (Carbosynth Ltd) were titrated

to a CWT domain sample, and ¹H, ¹⁵N HSQC spectra were collected at approximate peptide to CWT domain molar ratios of 0, 0.5, 1, 2, 4, 8, 16, 32, 45, and 60. CSPs, calculated as $\Delta\delta = (\Delta\delta H^2 + (0.154 \times \Delta\delta N)^2)^{1/2}$, were plotted as a function of the molar ratio of titrated ligand to CWT domain. Dissociation constants for individual residues were obtained by nonlinear least squares fitting, as implemented in the program xcrvfit (www.bionmr.ualberta.ca/bds/software/xcrvfit). Site-specific dissociation constants were calculated as averages of dissociation constants of individual residues within the site. In all binding site studies protein to zinc concentration ratio was 1:1.

CAT domain-SH3 domain interaction was probed by replacing zinc with paramagnetic manganese (1:1 protein to cation ratio) and observing peak bleaching in a ¹H, ¹⁵N HSQC spectrum.

¹⁵N T₁, T₂ and heteronuclear NOE spectra were acquired for one- and three-zinc bound lysostaphin in the presence and absence of substrate pentaglycine on a Bruker AVANCE III HD 800 MHz spectrometer, equipped with a TCI ¹H/¹³C/¹⁵N cryoprobe. For T₁ relaxation delays of 20, 60, 100, 200, 400, 600, 900, 1,400, 2,000, and 2,600 ms were used. For T₂ the delays were multiples of loop length 16.96 ms, with the loop counter set to 1, 2, 3, 4, 5, 6, 7, 8, 10, and 12. The recycle delay was set to 3.1 and 2.6 s for the T₁ and T₂ experiments, respectively. All spectra were acquired as pseudo 3D spectra and Fourier transformed with NMRPipe. Peak intensities were fitted to decaying exponential function as implemented in Sparky. Heteronuclear NOE values were calculated as the intensity ratios of peaks from a pair of spectra measured with and without ¹H presaturation during the recycle delay. The relaxation delay was set to 5 s. The CWT domain relaxation spectra were recorded on a Bruker AVANCE III 500 MHz spectrometer. For T₁, relaxation delays of 20, 60, 100, 200, 400, 600, 800, 1,200, 1,600, and 2,000 ms. The recycle delay was set to 2.5. For T₂, the loop counter was set to 1, 2, 4, 8, 10, 12, 14, and 16 and the recycle delay was 1.5 s. Heteronuclear NOE spectra were acquired as for full-length lysostaphin.

Pentaglycine cleavage by lysostaphin was monitored by ¹H NMR as described previously (Raulinaitis et al., 2017a). Reactions took place in phosphate buffer saline (PBS), pH 7.2, with >90% D₂O by volume as a solvent and final volumes were 600 μ L. Initial pentaglycine (Sigma-Aldrich) and lysostaphin concentrations were 1 mM and 2.25 μ M, respectively. Metal ions (ZnCl₂, CoCl₂, CuCl₂, and MnCl₂) were added as needed. Residual activity of lysostaphin “apo” form was taken into account in activity calculations. Measurements were carried out by using a Bruker Avance III 600 MHz spectrometer, equipped with a QCI ¹H/¹³C/¹⁵N/³¹P cryoprobe and a SampleJet automated sample changer. The preheater of the sample changer was used for sample incubation when product formation was measured as a time series with 3-h intervals for 48 h. Single time-point samples were incubated in an incubator for 21 h and reactions were quenched by heating at 90°C for 15 min before proton spectra acquisitions at 37°C. The data were processed with Bruker TopSpin 3.5 software.

MD Simulations

MD simulations in explicit solvent were performed with AMBER 16 (Case et al., 2005) using the ff14SB force field. Lysostaphin CYANA ensemble structures 1, 5, 7, and 14 were selected as targets of simulations. The topology and coordinate files were generated with the LEaP program in AMBER 16. The zinc ion was modeled using the cationic dummy atom approach (Pang, 1999; Pang et al., 2000). The protein molecule was placed in a cubic box with a minimum solute-box distance of 10 Å, and solvated with TIP3P water molecules. Nine chloride ions were added to maintain electrical neutrality. After minimization, heating and equilibration of the system, the production 30-ns MD simulations were performed with periodic boundary conditions at 300 K. The temperature was maintained by using the Langevin thermostat, whereas the pressure was kept at 1 bar using the Berendsen barostat (Berendsen et al., 1984). The time step was set to 2 fs. Long-range electrostatic interactions were treated using the Particle Mesh Ewald method (Darden et al., 1993) with a cut-off of 10 Å. Bond lengths involving hydrogen atoms were constrained by SHAKE (Ryckaert et al., 1997). Analyses of the trajectories were carried out with CPPTRAJ (Roe and Cheatham, 2013) and VMD (Humphrey et al., 1996).

SAXS

BM29 beamline, ESRF, Grenoble, France was used for collecting the SAXS data on a PILATUS 1M image plate using sample to detector distance of 2.9 m and a wavelength of 1.0 Å (momentum transfer range $0.01 < q < 5 \text{ nm}^{-1}$). The measurements were carried out at 288 K in PBS buffer. Three different protein concentrations (1.0, 3.0, and 5.0 mg/ml) were used in the data acquisition. PRIMUS (Konarev et al., 2003) in ATSAS software package (Franke et al., 2017) was utilized in the data processing. To assess the conformational variability of lysostaphin, EOM (Tria et al., 2015) was used. In EOM calculations, lysostaphin CAT domain (residues 251–384) and CWT domain (residues 410–493) of the NMR structure reported here, were used as rigid bodies, and residues 385–409 as random coil. 10 000 randomized conformations were generated for lysostaphin based on amino acid sequence and 3D-structure of CAT and CWT domains of lysostaphin using both random coil and native options in EOM. The scattering profiles of these randomly generated conformations were compared, and representative 20 structures whose scattering curve fits to the experimental scattering curve were selected by EOM algorithm. EOM run was repeated 100 times to provide statistical data about lysostaphin radius of gyration (R_g) and maximum dimensions (D_{max}) distribution.

RESULTS AND DISCUSSION

Lysostaphin Catalytic Groove Embracing Loops are Highly Mobile

The solution NMR structure of lysostaphin was determined based on distance restraints from NOESY spectra, ϕ/ψ dihedral restraints derived from chemical shifts using TALOS-N (Shen and Bax, 2013) and hydrogen bond restraints obtained from a H/D exchange experiment (Figure 1A and Table 1). In line with the X-ray crystal structure (Sabala et al., 2014),

lysostaphin is structurally arranged into two domains. The domains are connected by a fourteen-residue linker, which is devoid of persistent structure. The domains have random mutual orientations in the ensemble of structures. The CAT domain shares the structural features of a MEROPS M23 metallopeptidase domain with a β -sheet core and four loops forming the bottom and the rims of the catalytic groove, respectively (Rawlings et al., 2008). In one end of the groove reside the catalytic residues surrounding a zinc cation. The CWT domain has the SH3b domain, barrel-like all β -sheet fold.

While the core of the CAT domain is well-defined, and nicely superimposable with the crystal structure (RMSDs of 0.7 and 1.3 Å for backbone and heavy atoms, respectively), the catalytic groove embracing loops and the catalytic histidines are conformationally dispersed, which results in a poorer overall structural match in these regions. The conformational spread in the NMR ensemble is the result of solvent exposure and flexibility of residues in these regions (Figure 1B and Supplementary Figure 1). Fast amide proton exchange with solvent is likely to be present in the loops, because these do not display or have severely broadened peaks in a ^1H , ^{15}N HSQC spectrum. Additionally, the catalytic histidines probably exhibit solvent and/or conformational exchange, because their signals were not observed in a histidine-region ^1H , ^{15}N HMBC spectrum, showing cross-peaks between N δ 1, N ϵ 2, H δ 2, and H ϵ 1. Although carbon-bound H δ 2 or H ϵ 1 do not exchange with solvent, they can be influenced by solvent or conformational exchange elsewhere in the aromatic ring. This manifests as additional line-broadening of histidine ring protons and disappearance of catalytic histidine correlations from the HMBC. In addition to exchange, micro-to milli-second time scale motions are present in the catalytic groove (Figure 2 and Table 2), as identified by high R_1R_2 products derived from ^{15}N relaxation measurements (Kneller et al., 2002).

Catalytic groove conformational dynamics is evident also from MD simulations. The root mean square fluctuation, RMSF, emphasizes the flexible nature of three of the groove embracing loops (Figure 1B and Supplementary Figure 1). Consistent with the NMR ensemble, the outer half of the N-terminal loop, residues G270–G277, is highly dynamic, whereas the inner half is restrained by transient backbone hydrogen bonds within the loop, Y267–G281, G268–H279 as well as the H279N ϵ 2–Zn $^{2+}$ interaction. The loop encompassing residues N368–Q376 is the least flexible. It is transiently stabilized by backbone S371–T374 and side chain S369–T374 hydrogen bonds, the former twisting the loop into a three-residue 3_{10} helix.

The MD simulation reveals several possible mechanisms contributing to the disappearance of the catalytic histidine HMBC signals (Supplementary Figure 1). Firstly, while the two histidine rings remain approximately coplanar during the simulation, they slide relative to each other. Secondly, the aromatic ring of F285, next to H360, is able to rotate. Thirdly, the loop closest to the catalytic histidines, residues T353–Y355 thereof in particular, adopt a multitude of conformations at a fluctuating distance from the catalytic histidines. Conformational exchange between all these different structural arrangements is likely to broaden the HMBC signal beyond detection. Moreover,

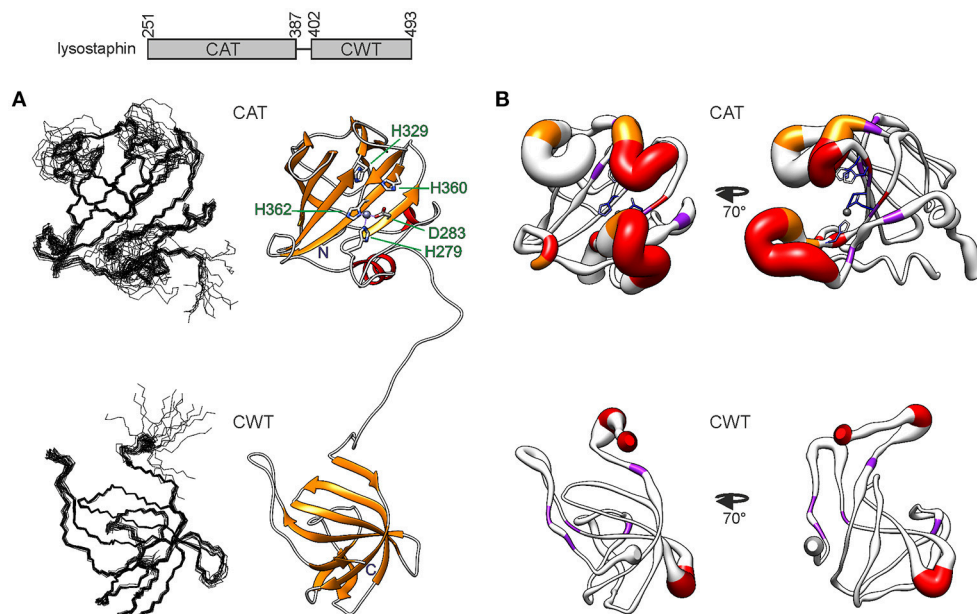


FIGURE 1 | Structure of *S. simulans* lysostaphin. On top, the domain structure is depicted. The N-terminal CAT domain is composed of 137 residues and the C-terminal CWT domain of 92 residues. The domains are linked by a flexible 14-residue linker. **(A)** Solution NMR structure of lysostaphin. On the left, the two domains are represented as ensembles of 15 superimposed structures. On the right, a representative member of the ensemble is shown in a ribbon diagram. The catalytic site residues in the CAT domain are labeled. Zinc is shown as a gray sphere. **(B)** CAT (top) and CWT domain dynamics. The heavy atom RMSF obtained from the MD simulation is depicted as the radius of the ribbon. Residues which lack or display strongly attenuated amide peak in the ^1H , ^{15}N HSQC spectrum are colored in red and orange, respectively. Residues which experience μs -ms dynamics are colored in purple. Catalytic site residues in the CAT domain are colored according to their heavy atom RMSF, blue depicting low RMSF values and lighter colors higher RMSF values. The structure on the left has the same orientation as the structures in **(A)**, whereas the structure on the right has been rotated 70 degrees around the vertical axis.

perhaps most importantly, H329 and H360 are transiently accessible to solvent, leading to signal broadening through exchange between tautomeric, ionic and H/D states. Notably, the outward orientation of H360 aromatic ring in some of the NMR ensemble structures appears to be energetically unfavorable as during the MD simulations the ring turns toward the catalytic groove.

Of the M23 family endopeptidases, for which the structure has been determined, lysostaphin CAT domain has the highest amino acid sequence identity with the sequences of LytM and LytU, 49 and 44%, respectively. LytM and LytU are *S. aureus* autolytic endopeptidases taking part in PG remodeling by cleaving the pentaglycine cross-bridges. The three structures are remarkably similar. Backbone atom RMSDs over the β -sheet core are 0.7 and 0.9 Å for LytM and LytU, respectively, when overlaid with the lysostaphin CAT domain. Besides structural correspondence they share a similar dynamic behavior with a stable core and mobile catalytic site and surrounding loops (Raulinaitis et al., 2017a,b).

The CWT domain is compact, stabilized by extensive hydrogen bonding within loops as well as between loops and the rest of the molecule, and high similarity is found between the solution and crystal structures. Backbone and heavy atom RMSDs are 0.5 and 1.1 Å, respectively, for all residues in the closest matching pair. Some dynamic features are, however, observed. Micro- to millisecond time scale motions are present in five residues, K412, T428, R476, G486, and L488 and four

residues, G401, K406, N466, S467 are prone to fast exchange with solvent, all in loops (**Figure 1B**).

Lysostaphin Domains Do Not Interact *in Vitro*

No inter-domain NOE peaks were observed in the NOESY spectra, strongly indicating that the two domains do not possess a stable interaction interface. The lack of chemical shift differences between the isolated CWT domain and the CWT domain in full-length lysostaphin (Supplementary Figure 2) provides additional solid evidence for non-interacting domains. Nevertheless, we further inspected the possibility of domain interaction by replacing the CAT-bound zinc cation with manganese and analyzing peak bleaching in a ^1H , ^{15}N HSQC spectrum. Unpaired electrons in isotropic paramagnetic probes such as manganese induce transverse paramagnetic relaxation enhancement (PRE) leading to line-broadening, affecting peak intensity, up to 35 Å from the paramagnetic center depending on the paramagnetic probe (Clore and Iwahara, 2009). In Mn^{2+} -bound lysostaphin peaks disappeared or broadened only in the CAT domain (Supplementary Figure 3), at a maximum average distance of 23.6 Å from the paramagnetic center, leaving peaks from the far-most backside of this domain intact. Conceivably, as the linker allows such an arrangement, the domains could interact through this backside CAT domain region without the CWT domain peaks experiencing line-broadening.

TABLE 1 | NMR restraints and structural statistics for the ensemble of 15 lysostaphin conformers of least restraint violations tabulated for the whole enzyme and domains separately.

| | Lysostaphin | CAT | CWT |
|---|-----------------|-------------|-------------|
| COMPLETENESS OF RESONANCE ASSIGNMENTS (%)^a | | | |
| Backbone | 93.3 | 92.1 | 98.3 |
| Side chain, aliphatic | 93.7 | 93.3 | 97.4 |
| aromatic | 95.7 | 91.7 | 99.5 |
| EXPERIMENTAL RESTRAINTS | | | |
| Distance restraints | | | |
| Total | 4312 | 2208 | 2070 |
| Short range ($i-j \leq 1$) | 1950 | 1017 | 903 |
| Medium range ($1 < i-j < 5$) | 448 | 245 | 199 |
| Long range ($i-j \geq 5$) | 1914 | 946 | 968 |
| H-bond restraints | 44 | 30 | 14 |
| Dihedral restraints (TALOS-N) | 277 | 153 | 124 |
| No. of restraints per restrained residue | 22.9 | 19.8 | 24.8 |
| No. of long-range restraints per restrained residue | 9.7 | 8.6 | 11.0 |
| RESIDUAL RESTRAINTS VIOLATIONS | | | |
| Average no. of distance violations per structure | | | |
| 0.1–0.2 Å | 10.9 | | |
| 0.2–0.5 Å | 1.1 (max 0.3 Å) | | |
| >0.5 Å | 0 | | |
| Average no. of dihedral angle violations per structure | | | |
| 1–5° | 0 | | |
| >5° | 0 | | |
| MODEL QUALITY^b | | | |
| RMSD backbone atoms (Å) | | 0.6 | 0.3 |
| RMSD heavy atoms (Å) | | 1.0 | 0.7 |
| RMSD bond lengths (Å) | 0.014 | | |
| RMSD bond angles (°) | 2.0 | | |
| MOLPROBITY RAMACHANDRAN STATISTICS (%)^b | | | |
| Most favored regions | 97.0 | 97.0 | 96.9 |
| Allowed regions | 3.0 | 2.9 | 3.1 |
| Disallowed regions | 0.0 | 0.1 | 0.0 |
| GLOBAL QUALITY SCORES (RAW/Z SCORE)^b | | | |
| Verify3D | 0.44/–0.32 | 0.43/–0.48 | 0.46/0.00 |
| ProsaII | 0.33/–1.32 | 0.20/–1.86 | 0.45/–0.83 |
| PROCHECK(ϕ - ψ) | –0.43/–1.38 | –0.40/–1.26 | –0.48/–1.57 |
| PROCHECK (all) | –0.40/–2.37 | –0.36/–2.13 | –0.43/–2.54 |
| Molprobtity clash score | 1.19/1.32 | 0.45/1.45 | 2.30/1.13 |
| MODEL CONTENTS | | | |
| No. of ordered residues ^b | 202 | 113 | 89 |
| Total no. of residues | 245 | 139 | 92 |
| BMRB accession number | 34121 | | |
| PDB ID code | 5NMY | | |

^aBackbone includes C α , C β , N, and H atoms, except the N-terminal amide. For side chains, excluded are the highly exchangeable groups (Lys, amino, Arg, guanido, Ser/Thr/Tyr hydroxyl, His $\delta 1/\epsilon 2$), as well as all unprotonated carbons and nitrogens. ^bOrdered residues: 253–270, 279–308, 313–350, 361–387, 402–465, 468–492. Computed using PSVS (Bhattacharya et al., 2007).

The PRE experiment remaining inconclusive, we turned to the ¹⁵N relaxation data and overall rotational correlation times (τ_c), which can also be used to untangle the confines of inter-domain mobility. The τ_c s derived from ¹⁵N R₂, R₁ rates are notably different for the CAT and CWT domains, 7.8 and 6.0 ns at 35°C,

respectively (Table 2). Also, the τ_c of the CWT domain in full-length lysostaphin is significantly higher than that of the isolated CWT domain, 4.1 ns. These indicate that there is a clear motional hindrance by the neighboring domain in full-length lysostaphin. There still is, however, substantial flexibility, considering that a

unified complex encompassing both domains has a predicted τ_c of 13.4 ns.

To further investigate the structure and flexibility of lysostaphin in solution, SAXS data were collected. An ensemble optimization (EOM) method was employed to analyse the SAXS data. In EOM calculations, lysostaphin CAT (residues 251–384) and CWT (410–493) domains of the NMR structure reported here were used as rigid bodies, and residues 385–409 as random coil. EOM runs yielded reproducible ensembles

neatly fitting the experimental data with χ^2 -values around 1.4 (Figure 3C). The results clearly show that lysostaphin adopts two different conformations in solution; a more favored compact subpopulation with average R_g and D_{max} of 27 and 82 Å, respectively, and a less favored extended subpopulation, with average R_g and D_{max} of 37 Å and 110 Å (Figures 3A,B,D). The D_{max} of the predominant conformation is consistent with an average domain separation of 22–23 Å. This is considerably shorter than the length encompassed by a fully extended

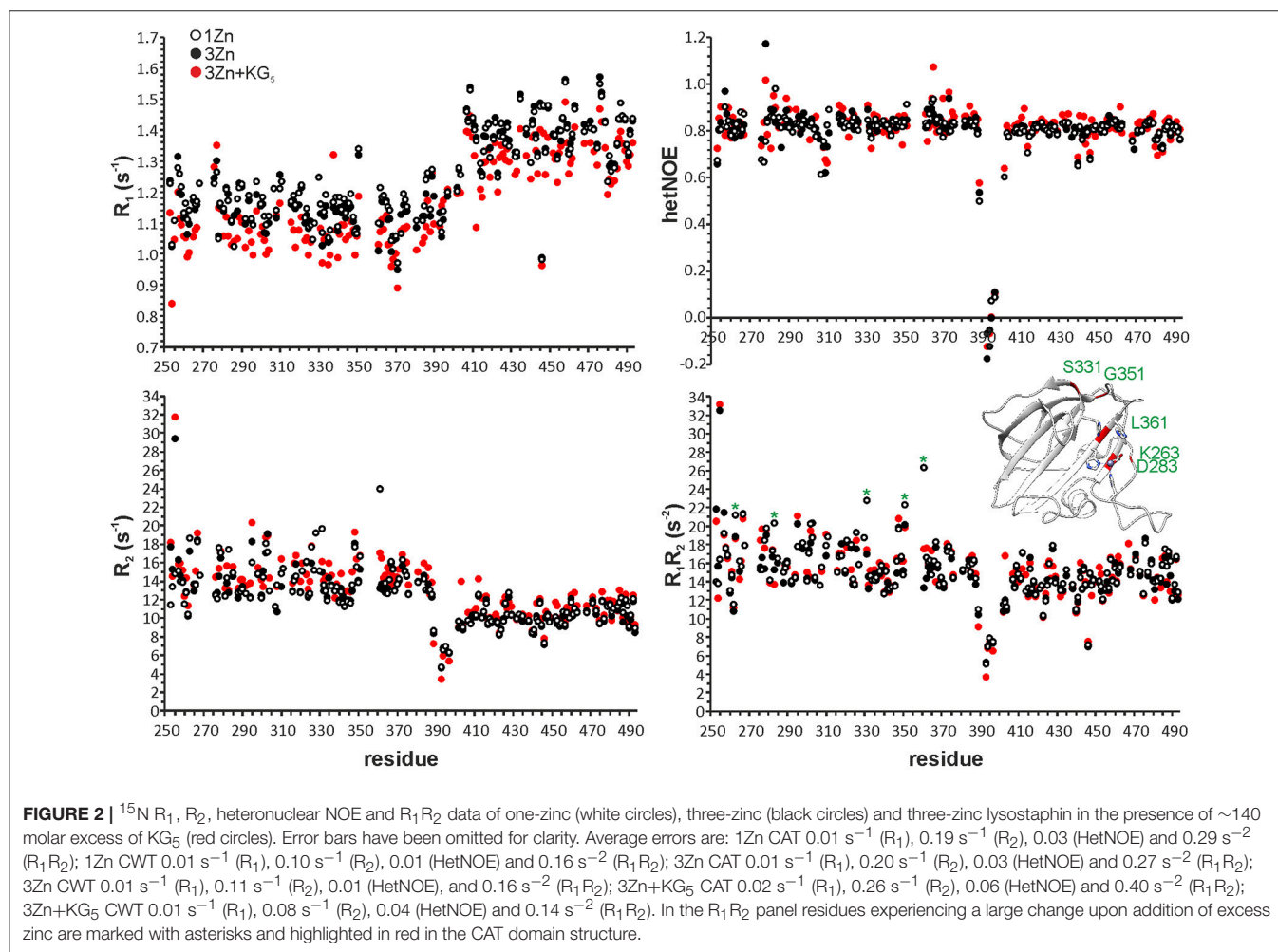
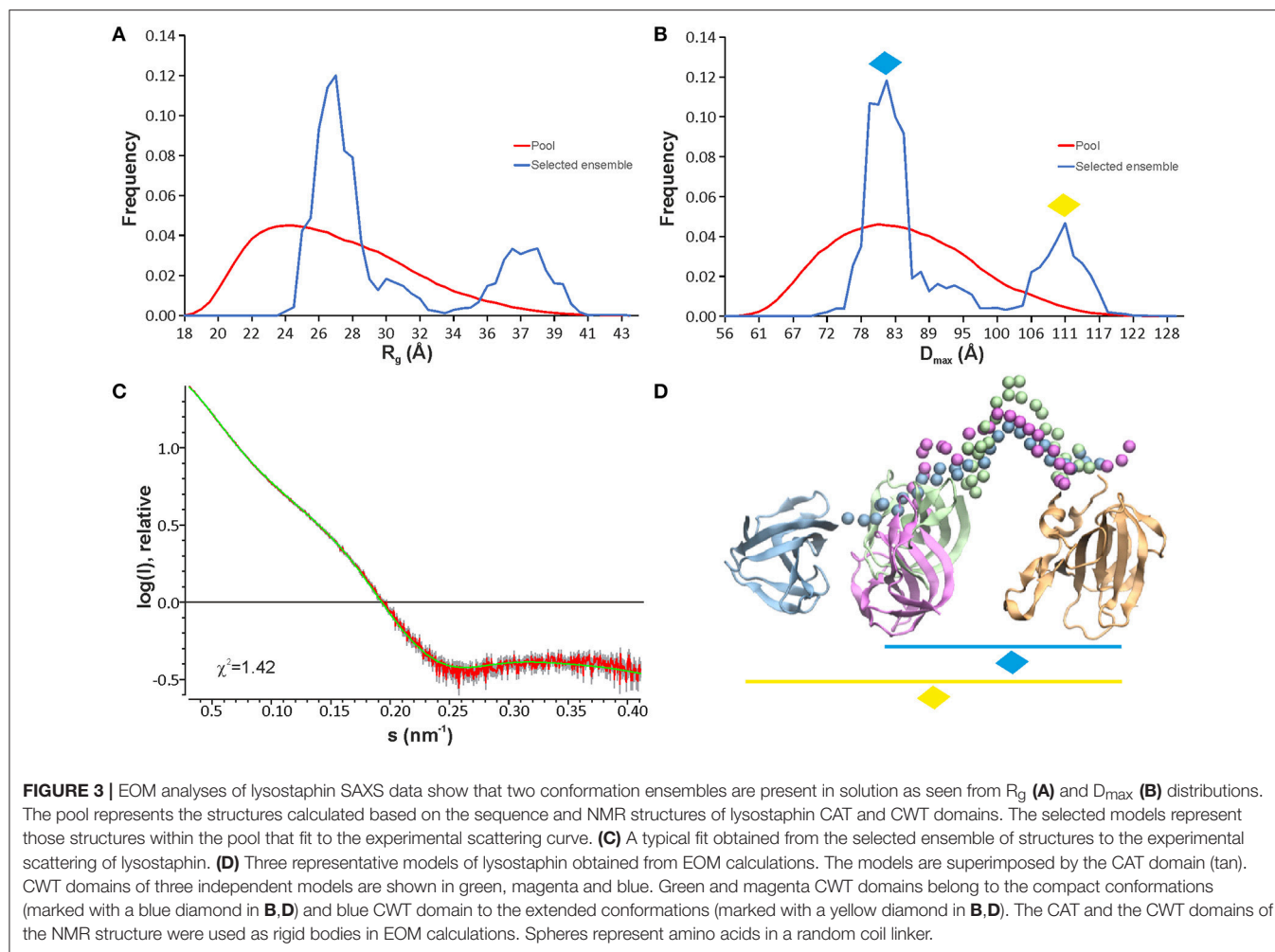


TABLE 2 | Average R_1 , R_2 and hetNOE values and standard deviations derived from ^{15}N relaxation experiments at 800 MHz, 35°C .

| Cofactor/Ligand | Domain | $R_1 \text{ (s}^{-1}\text{)}$ | $R_2 \text{ (s}^{-1}\text{)}$ | hetNOE | $\tau_c \text{ (ns)}$ |
|---------------------|--------|-------------------------------|-------------------------------|-----------------|-----------------------|
| 1Zn | CAT | 1.15 ± 0.07 | 13.60 ± 1.49 | 0.83 ± 0.05 | 7.9 ± 0.5 |
| | CWT | 1.37 ± 0.09 | 10.10 ± 1.08 | 0.79 ± 0.06 | 6.0 ± 0.4 |
| 3Zn | CAT | 1.13 ± 0.07 | 13.61 ± 1.51 | 0.83 ± 0.06 | 7.9 ± 0.5 |
| | CWT | 1.38 ± 0.08 | 9.86 ± 1.01 | 0.80 ± 0.04 | 5.9 ± 0.3 |
| 3Zn + KG_5 | CAT | 1.06 ± 0.07 | 14.78 ± 1.69 | 0.84 ± 0.07 | 8.6 ± 0.6 |
| | CWT | 1.31 ± 0.08 | 10.59 ± 1.23 | 0.82 ± 0.07 | 6.3 ± 0.4 |

Residues with significant contribution from internal motions were not included in the calculation of averages (Tjandra, 1996).



conformation of the linker containing two prolines, ~ 42 Å. This raises the possibility of the presence of secondary structure in the linker or interactions of the linker with one or other domain. Neither is, however, supported by secondary chemical shifts or NOE peaks. Moreover, significantly lower R_1R_2 products and heteronuclear NOEs found for linker residues K389-W402 with respect to the values in the CAT and CWT domains indicate that the linker is highly flexible. However, two sets of peaks were assigned for linker residues V394-P398, indicating the presence of two preferred conformations within this stretch, likely due to P398 cis/trans isomerism. This potentially offers a functional benefit to lysostaphin as it allows sampling of a larger conformational space and readjustment of domain orientations upon scission of pentaglycine bridges near its anchoring site (Aitio et al., 2012).

A dynamic inter-domain interaction in lysostaphin has been presented previously (Lu et al., 2013). Although the long flexible linker indeed allows the spatial proximity of the domains, there are no indications of such interaction in our data. The linker easily allows cleavage of adjacent and nearby pentaglycine bridges in PG near the bridge that serves as an anchor. Considering the architecture and dimensions of the PG lattice (Kim et al., 2015),

formation of a CAT and CWT “envelope” around the very same target pentapeptide, which is also fixated between perpendicular peptidoglycan stem peptides that it connects, is very improbable due to steric restrictions. The current data remain in support of the flexible arrangements between partly independent domains.

Binding of a Second Zinc by the Catalytic Histidines Reduces Lysostaphin Catalytic Activity

Proton NMR allows for direct and real-time reaction monitoring devoid of intricacies of cellular suspension (Raulinaitis et al., 2017a). Scission of pentaglycine by lysostaphin was monitored for 48 h by ^1H NMR and under conditions of this study its catalytic rate appeared to be limited only by saturation with the substrate (Supplementary Figure 4). The substrate turnover rate by the enzyme in the first 3 h of the reaction is estimated to be in the order of 0.006 s^{-1} .

We used the method to assess lysostaphin catalytic activity dependence on different metal cofactors (**Figure 4**). Samples were first incubated for 21 h and then quenched by heating, followed by acquisition of a ^1H spectrum. Results indicate that

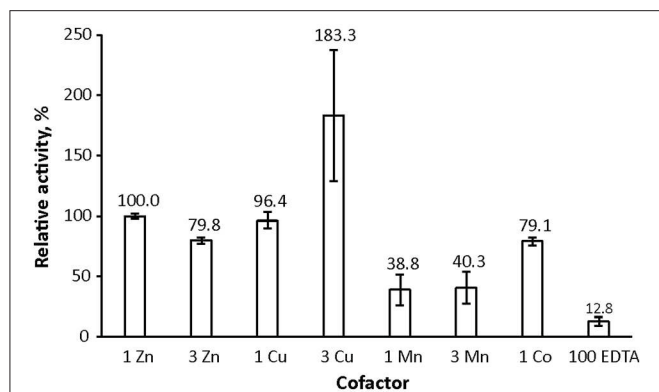


FIGURE 4 | Lysostaphin activity dependence on zinc, cobalt, copper and manganese cofactors. Lysostaphin was incubated with pentaglycine (2.25 μ M and 1 mM concentrations, respectively) in PBS buffer, pH 7.2, for 21 h at 37°C. Total cofactor ratios to protein were 1:1 and 3:1, EDTA excess was 70-fold. Purified enzyme, despite extensive treatment with EDTA during purification to remove any cofactors bound to the enzyme, nevertheless displayed residual activity. Activity was adjusted to one zinc-bound lysostaphin as 100% reference point. Bars represent standard error of the mean (S.E.M.) of at least three independent measurements.

metal cations Zn^{2+} , Co^{2+} , and Cu^{2+} are suitable cofactors of lysostaphin, whereas Mn^{2+} is a weak cofactor. The activity increases when Cu^{2+} is present in excess. This behavior might be explained by Cu^{2+} being a better cofactor with a lower affinity, the enzyme reaching full metal cation saturation at higher cation concentrations. M23 family endopeptidases are recognized as zinc-dependent enzymes, yet other metal ions have been found to partially restore their activity (Firczuk et al., 2005; Wang et al., 2011) and recently Co^{2+} ions have been shown to yield a hyperactive LytU (Raulinaitis et al., 2017a). A partial inhibition of lysostaphin activity is observed in the excess of zinc. Even under the inhibition by an extra zinc ion, the enzyme retains its fold, as demonstrated by its ^1H , ^{15}N HSQC spectrum displaying features of a structured protein, very similar to those of the one-zinc form (Supplementary Figure 5). Notably, chemical shift perturbations, CSPs, are observed for peaks of residues in the catalytic groove as well as the surrounding loops. Changes in the dynamics are also observed (Figure 2). The majority of the largest changes occur for residues having the largest exchange contribution in the one-zinc form, located in the vicinity of the catalytic site. At these sites, the R_1R_2 products decrease, which indicates reduction of exchange in the μs -ms timescale, and structure stabilization. Loop amides remain unprotected from fast solvent exchange, though, as no additional peaks appear into the ^1H , ^{15}N HSQC spectrum. Conclusively, two additional cross-peak sets appear into a ^1H , ^{15}N HMBC spectrum with ^{15}N chemical shifts corresponding to those of a zinc-coordinating nitrogen (Banci et al., 1998). Hence, binding of a second zinc, coordinated by the catalytic histidines, reduces catalytic activity.

The inhibitory effect of second zinc binding was also demonstrated and explained by the arrest of catalytic histidines in the autolytic LytU, where the process may

have a regulatory role (Raulinaitis et al., 2017a). Whereas it is conceivable for an intracellular autolysin, such regulation by inhibition would be irrational for an extracellular bactericide like lysostaphin and is likely physiologically irrelevant.

Pentaglycine Interacts Transiently With the N-Terminal Groove in the CWT Domain

NMR suits particularly well to study weak or transient molecular interactions. We performed CWT domain binding site mapping with hexapeptides G_5K and KG_5 (Figure 5A and Supplementary Figure 6). The presence of lysine significantly enhanced the solubility of the peptide, which was essential in order to reach the large excess of ligand needed to achieve noticeable CSPs. The two pentaglycine derivatives produced similar amide peak CSPs when present in ~ 170 molar excess. Ligand interaction with CWT was independent of the presence of the CAT domain, as deduced by the similar CSPs observed for the CWT domain in full-length lysostaphin in the presence of ~ 140 molar excess of KG_5 (data not shown).

Largest peak shifts, $\Delta\delta \geq 0.10$ ppm, in the presence of G_5K were observed for N405, Y411, G430-F432, V452, M453, and Y472. Residues in strands β_1 and β_2 and the loop in between (K403-K412) form the bottom of the pentaglycine binding groove, and residues G430-F432 from the loop between strands β_3 and β_4 and Y472 in strand β_7 the ceiling of the groove. Binding site mapping demonstrates CSPs for a delimited set of residues, without secondary, rearrangement-induced perturbations. Our data are nicely in accordance with the crystal structure of the lysostaphin CWT-pentaglycine complex (PDB ID 5LEO), which reveals that only very small structural rearrangements, namely few sidechain reorientations, occur upon binding of pentaglycine. The dissociation constant of the interaction is estimated to be ultra-weak, > 10 mM, in line with those observed for other pentapeptide-SH3b-type CWT domain interactions (Gu et al., 2014; Benešik et al., 2017).

The ^{15}N relaxation data show interesting differences between free and pentaglycine-bound three-zinc lysostaphin. Both domains in the complex have larger τ_{cs} than those in the free three-zinc form, 8.6 vs. 7.9 ns for the CAT and 6.3 vs. 5.9 ns for the CWT domains (Table 2). The difference is significant and cannot be attributed to the change in molecular weight only, considering that an increase of ~ 430 Da in molecular weight is predicted to increase τ_{c} by 0.2 ns only. Exchange between the complex and free forms could be the source of the larger τ_{cs} , and indeed many local changes in the R_1R_2 products are observed. However, there is an equal number of positive and negative changes with values larger than $\langle R_1R_2 \rangle \pm 1$ SD, and the average R_1R_2 products remain the same for both domains. We gather that the additional rise can be interpreted in terms of transient inter-domain interaction in the presence of the ligand, or as a change in the degree of anisotropy of molecular tumbling induced by peptide or inter-domain interaction. Future studies in alignment media will help to derive information about relative domain orientation and inter-domain motions. The linker region shows

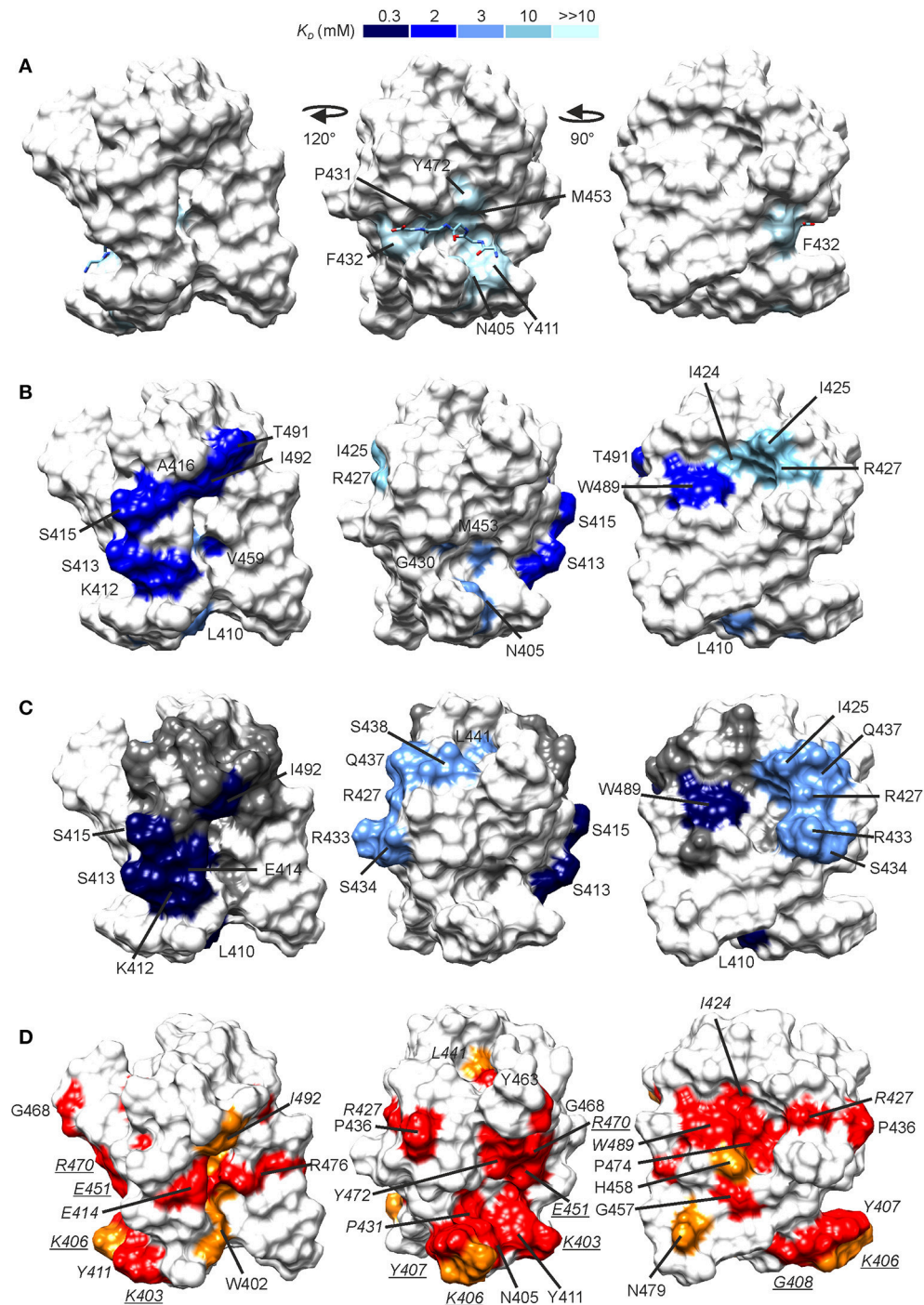


FIGURE 5 | Peptide binding sites as revealed by NMR titrations. CSPs induced by G₅K **(A)**, A-D-EK-GGGGG-A-D-EK-D-A **(B)**, and A-D-γ-EK-D-A-D-A **(C)** mapped onto the CWT domain structure. The surfaces are colored by the associated dissociation constant, the scale of which is given on the top of the figure. In **(A)**, the location of the bound pentaglycine is given from a structural superimposition of the free NMR structure and the X-ray CWT-substrate complex with PDB ID 5LEO. In **(C)**, between the high- and low affinity sites lies a group of residues which demonstrated a curved peak movement (see Supplementary Figure 4). These are shown in gray. **(D)** Conserved residues in *S. aureus*-targeting CWT domains mapped onto the lysostaphin CWT domain structure. Conservation was determined from a sequence alignment with ClustalW (Larkin et al., 2007) of lysostaphin CWT-like *S. aureus*-targeting sequences in the UniProt database (The UniProt Consortium, 2017). In red are shown fully conserved residues and in orange residues with strongly similar properties. Conserved residues for which large CSPs were observed are labeled in red, whereas conserved residues for which smaller CSPs were observed (0.06–0.09 ppm, not colored in A–C for clarity) are labeled in underlined italics. Sequence alignment is given in Supplementary Figure 5.

similar dynamics in the absence and presence of the peptide, revealing that this stretch remains flexible, loosely connecting the domains.

The CWT Domain Shows Higher Affinity Toward the Stem Peptide

We reasoned that *in vivo* the CWT–pentaglycine interaction is likely to be bolstered by auxiliary contacts from the glycan backbone or the stem peptides in the PG lattice (for a description of the chemical structure of PG see Vollmer et al., 2008). We found that the CWT domain had no significant affinity for the muramyl dipeptide *N*-acetylmuramyl-L-A-D- γ -Q hydrate as deduced by the very small CSPs observed at ~ 80 molar excess of the peptide (data not shown).

Titration of the CWT domain with a synthetic peptide A-D-EKGGGGGA-D-EK-D-A, which would resemble the stem-cross-bridge-stem sequence in native PG, resulted, instead, in significant CSPs at considerably lower concentration ratios, ~ 24 molar excess. In addition to CSPs of residues in the pentaglycine groove, CSPs were observed for two groups of peaks with distinct binding affinities. If viewing the CWT domain the pentaglycine groove laying horizontally toward the viewer, the N-terminus of the bound pentaglycine on the right (PDB ID 5LEO), then K412-S413, S415-A416, V459, T491, and I492 form an elongated patch on the right side of the domain (**Figure 5B**). K412 is located between the pentaglycine and the elongated sites. These residues have an average K_D of ~ 2.3 mM. Residues I424-I425 and R427 form a small patch on the left side of the CWT domain with an average K_D of ~ 10.5 mM. W489 is located between the elongated and the small sites, but with a K_D of ~ 2.3 mM belongs to the elongated site. The peptide is not large enough to reach simultaneously the three surfaces. Considering that residues in the pentaglycine site have a much higher affinity toward this peptide than to pentaglycine or KG₅/G₅K (average K_D of ~ 3.2 mM) we suspect that one peptide molecule binds via its tail to the elongated site, and boosts the affinity of the pentaglycine site for the pentaglycine moiety of the peptide.

Titration with a shorter peptide A-D- γ -EK-D-A-D-A resulted in a complex peak behavior with three distinct patterns of peak position movement (**Figure 6**). Cross-peaks of residues L410, K412-S415, V452, W489, and I492 moved linearly upon peptide addition and reached saturation at ~ 32 molar excess of peptide. The dissociation constant was determined to be ~ 0.3 mM, notably lower than that for pentaglycine or the longer peptide. These residues mostly overlap with the elongated site of the previous titration (**Figure 5C**). Cross-peaks of the second group of residues also moved linearly, but began prominent movement and reached saturation at later steps of the titration, resulting in a dissociation constant of ~ 3 mM. These residues form a patch on the surface, which covers the small site of the previous titration (I425, R427, and L473) and has extensions of polar residues (R433, S434, Q437, and S438). The third group of peaks exhibited non-linear peak movement. For some of these residues the CSP are as large as for those in the two other groups, which rules out unspecific multiple binding as the cause of nonlinearity. Instead, as these residues are located

between the first two surface patches, and include residues in the hydrophobic core, a conformational rearrangement brought about binding to the high-affinity site might be responsible for the curved titrations. Core residues V461 and V475 are probably involved in this rearrangement because their methyl peaks substantially shift in the ^1H spectra along the titration (Supplementary Figure 7).

The importance of R296 and W358 in ALE-1 SH3b binding to *S. aureus* PG has been demonstrated by selective mutations (Lu et al., 2006). Binding of ALE-1 R296A and W358A mutants was reduced 3- and 2-fold, respectively, as compared to wild type binding. These residues correspond to R427 and W489 of lysostaphin CWT, belonging to the lower- and higher affinity sites, respectively, although structurally very close on the surface of CWT. R427 and W489 are strictly conserved in *S. aureus*-targeting SH3b domains (**Figure 5D**, Supplementary Figure 8). Based on the titration results and residue conservation, other residues which are likely to have an impact on CWT PG binding other than those in the pentaglycine binding groove are E414 and I492 of the high-affinity site.

We generated model structures of the CWT domain in complex with pentaglycine-stem peptide fragments (Supplementary Figure 9) with the assumption that the pentaglycine moiety occupies the N-terminal binding groove as in the structure 5LEO. The models indicate that the stem peptides on either side of pentaglycine seem to be too short to reach the CSPs observed in the presence of the A-D- γ -EK-D-A-D-A peptide. Also, CSPs in the region of the linkage between pentaglycine and the stem peptide are missing from the CWT domain surface. We gather that the observed CSPs might arise from a structural rearrangement occurring upon binding (see before), or be the result of stem peptide contacts present within the PG architecture, that is from stem peptides not directly linked to the pentaglycine recognized by the N-terminal groove. According to the parallel-stem architecture proposed by Kim et al. (2015) for *S. aureus* PG, contacts with stem peptides from further up or down in the densely packed glycan chains are probable.

These results clearly support the idea of formation of affinity-heaving accessory contacts with the PG network. Altogether it is advisable to consider using larger PG fragments in SH3b interaction studies in order to overcome the very low affinity of these domains toward pentaglycine (Gu et al., 2014; Benešik et al., 2017).

CWT–PG association studies have demonstrated that CWT binding to *S. aureus* cells cannot be blocked with excess of pentaglycine (Gründling and Schneewind, 2006), which is in accord with the extremely low affinity found for the pentaglycine–CWT interaction in the present study. The authors also demonstrated that inhibition of binding was achieved only with PG fragments containing multiple cross-linked murein subunits, not with smaller murein mono- or dimers. This suggests that the CWT domain affinity for *S. aureus* PG has to be higher than that observed here for the five-residue peptide, ~ 300 μM . It also suggests that the PG lattice structure is needed for maximal affinity, only cross-linked fragments with multiple murein moieties being capable of reproducing this structure. In

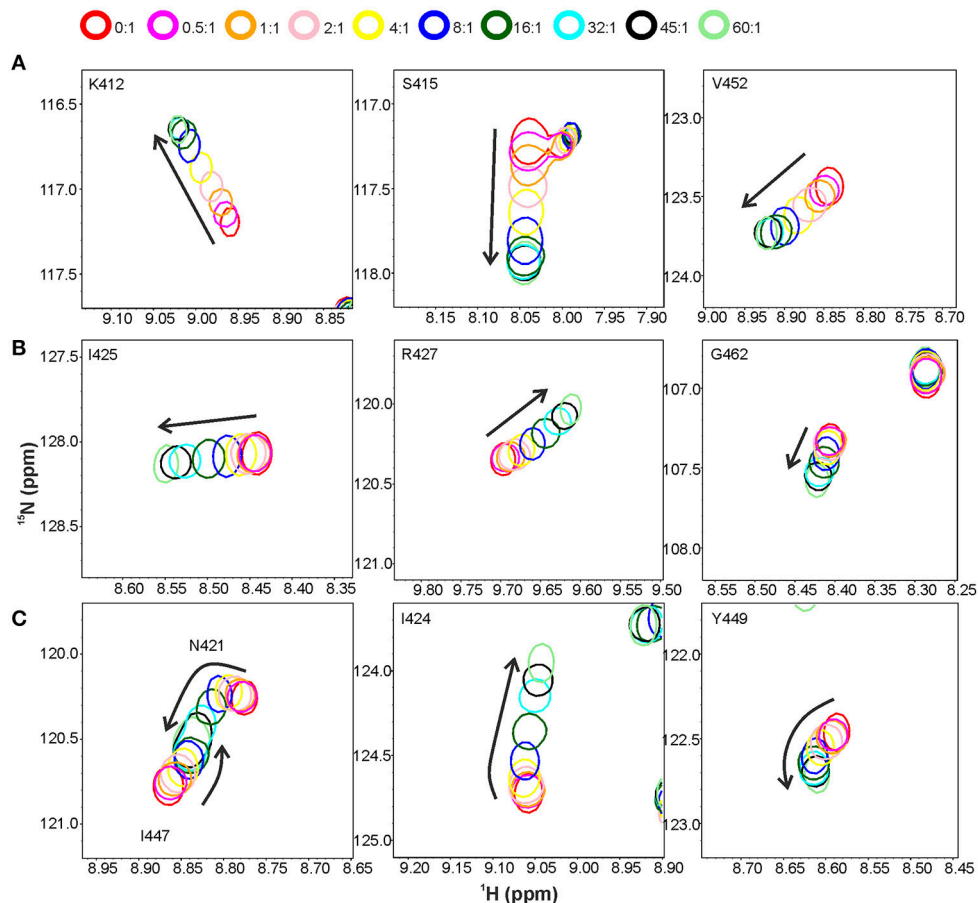


FIGURE 6 | Titration of the CWT domain with the peptide A-D-γ-EK-D-A-D-A. Shown are overlaid sections of ^1H , ^{15}N HSQC spectra recorded at peptide to protein concentration ratios of 0:1 (red contours), 0.5:1 (magenta), 1:1 (orange), 2:1 (pink), 4:1 (yellow), 8:1 (blue), 16:1 (dark green), 32:1 (cyan), 45:1 (black) and 60:1 (light green). Three types of peak movement were observed: **(A)** linear, with saturation at $\sim 32:1$ concentration ratio, **(B)** linear, with saturation at $\sim 60:1$ concentration ratio, and **(C)** non-linear peak movement.

other words, binding to the pre-arranged PG lattice is likely to be entropically more favorable than binding, through induced folding, to a disordered peptide nonetheless containing the same interaction sites.

CONCLUSIONS

This study reveals details on structure and dynamics within and between the lysostaphin domains, on metal ion preference of the CAT domain, as well as characterizes substrate binding sites and affinities of the CWT domain. Remarkably, the affinity of the CWT domain was found to be much higher toward the stem peptide than toward pentaglycine. Also, the CWT domain affinity for pentaglycine increased when the latter was annexed to a stem peptide-like moiety. These findings underscore the importance of auxiliary contacts in CWT PG recognition. The stem peptide analogs interacted with two surfaces on both sides of the N-terminal pentaglycine binding groove of the CWT domain. Residues E414, R427, W489, and I492 are predicted to be crucial to the PG interaction. Taking into

account the PG architecture, rather than just the pentaglycine bridge, is imperative to appreciate the factors contributing to the catalytic activity of the CAT domain, to recognize the structural components influencing the interaction between lysostaphin and PG and to engineer novel lysostaphin- or CWT-based antimicrobials.

DATA AVAILABILITY STATEMENT

The assigned chemical shifts and coordinates of lysostaphin have been deposited in the BioMagResBank (<http://www.bmrb.wisc.edu/>) and the PDB (<https://www.rcsb.org/>) with accession codes 34121 and 5NMY, respectively.

AUTHOR CONTRIBUTIONS

PP conceived and designed the research. VR and LK produced and purified the proteins. VR and HM performed enzyme activity studies. UP performed SAXS measurements and SAXS

data analysis. HT carried out MD simulations. HT, LK, and PP conducted CWT domain NMR studies. HT and PP performed full-length lysostaphin NMR studies. All authors wrote the paper.

FUNDING

The study was supported by the grants from Academy of Finland (number 288235 to PP and 28348 to UP) and Sigrid Juselius foundation to PP and the Helsinki Graduate Program in Biotechnology and Molecular Biology and Centre for International Mobility (CIMO) to VR.

REFERENCES

- Aitio, O., Hellman, M., Skehan, B., Kesti, T., Leong, J. M., Saksela, K., et al. (2012). Enterohaemorrhagic *Escherichia coli* exploits a tryptophan switch to hijack host f-actin assembly. *Structure* 20, 1692–1703. doi: 10.1016/j.str.2012.07.015
- Baba, T., and Schneewind, O. (1996). Target cell specificity of a bacteriocin molecule: a C-terminal signal directs lysostaphin to the cell wall of *Staphylococcus aureus*. *EMBO J.* 15:4789–4797.
- Banci, L., Benedetto, M., Bertini, I., Del Conte, R., Piccioli, M., and Viezzoli, M. S. (1998). Solution structure of reduced monomeric Q133M2 copper, zinc superoxide dismutase (SOD). Why is SOD a dimeric enzyme? *Biochemistry* 37, 11780–11791. doi: 10.1021/bi9803473
- Barraud, P., Schubert, M., and Allain, F. H. (2012). A strong ¹³C chemical shift signature provides the coordination mode of histidines in zinc-binding proteins. *J. Biomol. NMR* 53, 93–101. doi: 10.1007/s10858-012-9625-6
- Beneš, M., Nováček, J., Janda, L., Dopitová, R., Pernisová, M., Melková, K., et al. (2017). Role of SH3b binding domain in a natural deletion mutant of Kayvirus endolysin LysF1 with a broad range of lytic activity. *Virus Genes* 54, 130–139. doi: 10.1007/s11262-017-1507-2
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi: 10.1063/1.448118
- Bhattacharya, A., Tejero, R., and Montelione, G. T. (2007). Evaluating protein structures determined by structural genomics consortia. *Proteins* 66, 778–795. doi: 10.1002/prot.21165
- Blazanovic, K., Zhao, H., Choi, Y., Li, W., Salvat, R. S., Osipovitch, D. C., et al. (2015). Structure-based redesign of lysostaphin yields potent antistaphylococcal enzymes that evade immune cell surveillance. *Mol. Ther. Methods. Clin. Dev.* 2:15021. doi: 10.1038/mtm.2015.21
- Brickner, S. J., Barbachyn, M. R., Hutchinson, D. K., and Manninen, P. R. (2008). Linezolid (ZYVOX), the first member of a completely new class of antibacterial agents for treatment of serious gram-positive infections. *J. Med. Chem.* 51, 1981–1990. doi: 10.1021/jm800038g
- Browder, H. P., Zygmunt, W. A., Young, J. R., and Tavormina, P. A. (1965). Lysostaphin: enzymatic mode of action. *Biochem. Biophys. Res. Commun.* 19, 383–389. doi: 10.1016/0006-291X(65)90473-0
- Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., et al. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688. doi: 10.1002/jcc.20290
- Chambers, H. F., and Deleo, F. R. (2009). Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nat. Rev. Microbiol.* 7, 629–641. doi: 10.1038/nrmicro2200
- Clore, G. M., and Iwahara, J. (2009). Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.* 109, 4108–4139. doi: 10.1021/cr900033p
- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397
- 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
- Ercoli, G., Tani, C., Pezzicoli, A., Vacca, I., Martinelli, M., Pecetta, S., et al. (2015). LytM proteins play a crucial role in cell separation, outer membrane composition, and pathogenesis in nontypeable *Haemophilus influenzae*. *MBio* 6:e02575. doi: 10.1128/mBio.02575-14
- Firczuk, M., Mucha, A., and Bochtler, M. (2005). Crystal structures of active LytM. *J. Mol. Biol.* 354, 578–590. doi: 10.1016/j.jmb.2005.09.082
- Franke, D., Petoukhov, M. V., Konarev, P. V., Panjikovich, A., Tuukkanen, A., Mertens, H. D. T., et al. (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* 50, 1212–1225. doi: 10.1107/S1600576717007786
- Gründling, A., and Schneewind, O. (2006). Cross-linked peptidoglycan mediates lysostaphin binding to the cell wall envelope of *Staphylococcus aureus*. *J. Bacteriol.* 188, 2463–2472. doi: 10.1128/JB.188.7.2463-2472.2006
- Gu, J., Feng, Y., Feng, X., Sun, C., Lei, L., Ding, W., et al. (2014). Structural and biochemical characterization reveals LysGH15 as an unprecedented “EF-hand-like” calcium-binding phage lysin. *PLoS Pathog.* 10:e1004109. doi: 10.1371/journal.ppat.1004109
- Hirakawa, H., Akita, H., Fujiwara, T., Sugai, M., and Kuhara, S. (2009). Structural insight into the binding mode between the targeting domain of ALE-1 (92AA) and pentaglycine of peptidoglycan. *Protein Eng. Des. Sel.* 22, 385–391. doi: 10.1093/protein/gzp014
- 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
- Jagielska, E., Chojnacka, O., and Sabala, I. (2016). LytM fusion with SH3b-like domain expands its activity to physiological conditions. *Microb. Drug Resist.* 22, 461–469. doi: 10.1089/mdr.2016.0053
- Kim, S. J., Chang, J., and Singh, M. (2015). Peptidoglycan architecture of Gram-positive bacteria by solid-state NMR. *Biochim. Biophys. Acta* 1848, 350–362. doi: 10.1016/j.bbammem.2014.05.031
- Kneller, J. M., Lu, M., and Bracken, C. (2002). An effective method for the discrimination of motional anisotropy and chemical exchange. *J. Am. Chem. Soc.* 124, 1852–1853. doi: 10.1021/ja017461k
- Köck, R., Becker, K., Cookson, B., van Gemert-Pijnen, J. E., Harbarth, S., Kluytmans, J., et al. (2010). Methicillin-resistant *Staphylococcus aureus* (MRSA): burden of disease and control challenges in Europe. *Euro Surveill.* 15:19688. doi: 10.2807/ese.15.41.19688-en
- Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* 36, 1277–1282. doi: 10.1107/S0021889803012779
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Lee, B. Y., Singh, A., David, M. Z., Bartsch, S. M., Slayton, R. B., Huang, S. S., et al. (2013). The economic burden of community-associated methicillin-resistant *Staphylococcus aureus* (CA-MRSA). *Clin. Microbiol. Infect.* 19, 528–536. doi: 10.1111/j.1469-0691.2012.03914.x
- López-Méndez, B., and Güntert, P. (2006). Automated protein structure determination from NMR spectra. *J. Am. Chem. Soc.* 128, 13112–13122. doi: 10.1021/ja061136l
- Lowy, F. D. (1998). *Staphylococcus aureus* infections. *N. Engl. J. Med.* 339, 520–532. doi: 10.1056/NEJM199808203390806

ACKNOWLEDGMENTS

We acknowledge CSC-IT Center for Science Ltd. for the allocation of computational resources. ESRF is thanked for providing the beamline BM29 access.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2018.00060/full#supplementary-material>

- Lu, H., Gu, M. G., Huang, Q., Huang, J. J., Lu, W. Y., Lu, H., et al. (2013). Hydrogen/deuterium exchange mass spectrometry and site-directed disulfide cross-linking suggest an important dynamic interface between the two lysostaphin domains. *Antimicrob. Agents Chemother.* 57, 1872–1881. doi: 10.1128/AAC.02348-12
- Lu, J. Z., Fujiwara, T., Komatsuzawa, H., Sugai, M., and Sakon, J. (2006). Cell wall-targeting domain of glycylglycine endopeptidase distinguishes among peptidoglycan cross-bridges. *J. Biol. Chem.* 281, 549–558. doi: 10.1074/jbc.M509691200
- Nelson, D. C., Schmelcher, M., Rodriguez-Rubio, L., Klumpp, J., Pritchard, D. G., Dong, S., et al. (2012). Endolysins as antimicrobials. *Adv. Virus Res.* 83, 299–365. doi: 10.1016/B978-0-12-394438-2.00007-4
- Osipovitch, D. C., and Griswold, K. E. (2015). Fusion with a cell wall binding domain renders autolysin LytM a potent anti-Staphylococcus aureus agent. *FEMS Microbiol. Lett.* 362, 1–7. doi: 10.1093/femsle/fnu035
- Pang, Y. P. (1999). Novel zinc protein molecular dynamics simulations: steps toward antiangiogenesis for cancer treatment. *J. Mol. Model.* 5, 196–202. doi: 10.1007/s008940050119
- Pang, Y. P., Xu, K., Yazal, J. E., and Prendergas, F. G. (2000). Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Sci.* 9, 1857–1865.
- Ponting, C. P., Aravind, L., Schultz, J., Bork, P., and Koonin, E. V. (1999). Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J. Mol. Biol.* 289, 729–745. doi: 10.1006/jmbi.1999.2827
- Ramadurai, L., Lockwood, K. J., Nadakavukaren, M. J., and Jayaswal, R. K. (1999). Characterization of a chromosomally encoded glycylglycine endopeptidase of *Staphylococcus aureus*. *Microbiology (Reading, Engl.)* 145 (Pt 4), 801–808. doi: 10.1099/13500872-145-4-801
- Raulinaitis, V., Tossavainen, H., Aitio, O., Juuti, J. T., Hiramatsu, K., Kontinen, V., et al. (2017a). Identification and structural characterization of LytU, a unique peptidoglycan endopeptidase from the lysostaphin family. *Sci. Rep.* 7:6020. doi: 10.1038/s41598-017-06135-w
- Raulinaitis, V., Tossavainen, H., Aitio, O., Seppala, R., and Permi, P. (2017b). ¹H, ¹³C and ¹⁵N resonance assignments of the new lysostaphin family endopeptidase catalytic domain from *Staphylococcus aureus*. *Biomol. NMR Assign.* 11, 69–73. doi: 10.1007/s12104-016-9722-7
- Rawlings, N. D., Morton, F. R., Kok, C. Y., Kong, J., and Barrett, A. J. (2008). MEROPS: the peptidase database. *Nucleic Acids Res.* 36, D320–D325. doi: 10.1093/nar/gkm954
- Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095. doi: 10.1021/ct400341p
- Ryckaert, J.-P., Ciccotti, 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. doi: 10.1016/0021-9991(77)90098-5
- Sabala, I., Jagielska, E., Bardelang, P. T., Czapinska, H., Dahms, S. O., Sharpe, J. A., et al. (2014). Crystal structure of the antimicrobial peptidase lysostaphin from *Staphylococcus simulans*. *FEBS J.* 281, 4112–4122. doi: 10.1111/febs.12929
- Saksela, K., and Permi, P. (2012). SH3 domain ligand binding: what's the consensus and where's the specificity? *FEBS Lett.* 586, 2609–2614. doi: 10.1016/j.febslet.2012.04.042
- Schindler, C. A., and Schuhradt, V. T. (1964). Lysostaphin: a new bacteriolytic enzyme for the Staphylococcus. *Proc. Natl. Acad. Sci. U.S.A.* 51, 414–421. doi: 10.1073/pnas.51.3.414
- Schmelcher, M., Powell, A. M., Becker, S. C., Camp, M. J., and Donovan, D. M. (2012). Chimeric phage lysins act synergistically with lysostaphin to kill mastitis-causing *Staphylococcus aureus* in murine mammary glands. *Appl. Environ. Microbiol.* 78, 2297–2305. doi: 10.1128/AEM.07050-11
- 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
- Sugai, M., Fujiwara, T., Akiyama, T., Ohara, M., Komatsuzawa, H., Inoue, S., et al. (1997). Purification and molecular characterization of glycylglycine endopeptidase produced by *Staphylococcus capitis* EPK1. *J. Bacteriol.* 179, 1193–1202. doi: 10.1128/jb.179.4.1193-1202.1997
- Szweda, P., Schiellmann, M., Kotlowski, R., Gorczyca, G., Zalewska, M., and Milewski, S. (2012). Peptidoglycan hydrolases-potential weapons against *Staphylococcus aureus*. *Appl. Microbiol. Biotechnol.* 96, 1157–1174. doi:10.1007/s00253-012-4484-3
- Taubes, G. (2008). The bacteria fight back. *Science* 321, 356–361. doi: 10.1126/science.321.5887.356
- The Uniprot Consortium (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45, D158–D169. doi: 10.1093/nar/gkw1099
- Thumm, G., and Götz, F. (1997). Studies on polysostaphin processing and characterization of the lysostaphin immunity factor (Lif) of *Staphylococcus simulans* biovar staphylophilus. *Mol. Microbiol.* 23, 1251–1265. doi: 10.1046/j.1365-2958.1997.2911657.x
- Tjandra, N. (1996). Anisotropic rotational diffusion of perdeuterated HIV protease from ¹⁵N NMR relaxation measurements at two magnetic fields. *J. Biomol. NMR* 8, 273–284. doi: 10.1007/BF00410326
- Tria, G., Mertens, H. D., Kachala, M., and Svergun, D. I. (2015). Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr* 2, 207–217. doi: 10.1107/S205225251500202X
- Tschierske, M., Ehler, K., Strandén, A. M., and Berger-Bächi, B. (1997). Lif, the lysostaphin immunity factor, complements FemB in staphylococcal peptidoglycan interpeptide bridge formation. *FEMS Microbiol. Lett.* 153, 261–264. doi: 10.1111/j.1574-6968.1997.tb12583.x
- van Hal, S. J., Jensen, S. O., Vaska, V. L., Espedido, B. A., Paterson, D. L., and Gosbell, I. B. (2012). Predictors of mortality in *Staphylococcus aureus* bacteremia. *Clin. Microbiol. Rev.* 25, 362–386. doi: 10.1128/CMR.05022-11
- Vollmer, W., Blanot, D., and de Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* 32, 149–167. doi: 10.1111/j.1574-6976.2007.00094.x
- Wang, X., Yang, X., Yang, C., Wu, Z., Xu, H., and Shen, Y. (2011). Crystal structure of outer membrane protein NMB0315 from *Neisseria meningitidis*. *PLoS ONE* 6:e26845. doi: 10.1371/journal.pone.0026845
- Whisstock, J. C., and Lesk, A. M. (1999). SH3 domains in prokaryotes. *Trends Biochem. Sci.* 24, 132–133. doi: 10.1016/S0968-0004(99)01366-3
- WHO Mediacentre (2017). WHO Publishes List of Bacteria For Which New Antibiotics are Urgently Needed. Available online at: <http://www.who.int/mediacentre/news/releases/2017/bacteria-antibiotics-needed/en/>

Conflict of Interest Statement: HM was employed by company VTT Technical Research Centre of Finland Ltd.

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.

Copyright © 2018 Tossavainen, Raulinaitis, Kauppinen, Pentikäinen, Maaheimo and Permi. 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 Molecular Bases of the Dual Regulation of Bacterial Iron Sulfur Cluster Biogenesis by CyaY and IscX

Salvatore Adinolfi^{1†}, Rita Puglisi^{1†}, Jason C. Crack², Clara Iannuzzi^{1,3}, Fabrizio Dal Piaz⁴, Petr V. Konarev^{5,6}, Dmitri I. Svergun⁷, Stephen Martin⁸, Nick E. Le Brun² and Annalisa Pastore^{1,9*}

¹ The Wohl Institute, King's College London, London, United Kingdom, ² School of Chemistry, University of East Anglia, Norwich Research Park, Norwich, United Kingdom, ³ DBBGP, Università degli Studi della Campania "L. Vanvitelli," Naples, Italy, ⁴ Dipartimento di Medicina, Chirurgia e Odontoiatria "Scuola Medica Salernitana"/DIPMED, Università di Salerno, Fisciano, Italy, ⁵ A.V. Shubnikov Institute of Crystallography of Federal Scientific Research Centre "Crystallography and Photonics" of Russian Academy of Sciences, Moscow, Russia, ⁶ National Research Centre "Kurchatov Institute," Moscow, Russia, ⁷ European Molecular Biology Laboratory, Hamburg, Germany, ⁸ The Francis Crick Institute, London, United Kingdom, ⁹ Department of Molecular Medicine, University of Pavia, Pavia, Italy

OPEN ACCESS

Edited by:

Irene Diaz-Moreno,
Universidad de Sevilla, Spain

Reviewed by:

Antonio J. Diaz Quintana,
Universidad de Sevilla, Spain
Michael Sattler,
Technische Universität München,
Germany

*Correspondence:

Annalisa Pastore
annalisa.pastore@crick.ac.uk

†Present Address:

Salvatore Adinolfi,
Dipartimento di Scienza e Tecnologia
del Farmaco, Università di Torino, Italy

†These authors have contributed
equally to this work.

Specialty section:

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

Received: 09 November 2017

Accepted: 22 December 2017

Published: 02 February 2018

Citation:

Adinolfi S, Puglisi R, Crack JC,
Iannuzzi C, Dal Piaz F, Konarev PV,
Svergun DI, Martin S, Le Brun NE and
Pastore A (2018) The Molecular Bases
of the Dual Regulation of Bacterial Iron
Sulfur Cluster Biogenesis by CyaY and
IscX. *Front. Mol. Biosci.* 4:97.
doi: 10.3389/fmolb.2017.00097

IscX (or YfhJ) is a protein of unknown function which takes part in the iron-sulfur cluster assembly machinery, a highly specialized and essential metabolic pathway. IscX binds to iron with low affinity and interacts with IscS, the desulfurase central to cluster assembly. Previous studies have suggested a competition between IscX and CyaY, the bacterial ortholog of frataxin, for the same binding surface of IscS. This competition could suggest a link between the two proteins with a functional significance. Using a hybrid approach based on nuclear magnetic resonance, small angle scattering and biochemical methods, we show here that IscX is a modulator of the inhibitory properties of CyaY: by competing for the same site on IscS, the presence of IscX rescues the rates of enzymatic cluster formation which are inhibited by CyaY. The effect is stronger at low iron concentrations, whereas it becomes negligible at high iron concentrations. These results strongly suggest the mechanism of the dual regulation of iron sulfur cluster assembly under the control of iron as the effector.

Keywords: desulfurase, enzyme activity, frataxin, isc operon, iron chaperone, iron sulfur cluster

INTRODUCTION

Iron and sulfur are elements essential for life thanks to their unique redox properties. Yet, they are highly toxic. An efficient way to store them in cells in a nontoxic form is through formation of iron sulfur (Fe-S) clusters, labile prosthetic groups involved in several essential metabolic pathways (for a review Roche et al., 2013; Py and Barras, 2014). Assembly of Fe-S clusters is carried out by highly conserved machines which, in prokaryotes, are encoded by the *suf*, *nif*, and *isc* operons. *Isc* is the most general machine with highly conserved orthologs in eukaryotes.

In *E. coli*, the *isc* operon contains eight genes (i.e., *iscR*, *iscS*, *iscU*, *iscA*, *fdx*, *hscA*, *hscB*, and *iscX*; Zheng et al., 1998). Among the corresponding gene products, the most important players are the cysteine desulfurase IscS (EC 2.8.1.7), which converts cysteine to alanine and IscS-bound persulfide (Flint, 1996), and IscU, a transient scaffold protein which forms a complex with IscS (Agar et al., 2000; Urbina et al., 2001). The last component of the machine according to the order of genes in the operon is IscX (also known as YfhJ), a small acidic protein about which very little is known (Tokumoto and Takahashi, 2001; Tokumoto et al., 2002). IscX, the Cinderella of the *isc* operon, is

not essential, in contrast to the other *isc* proteins (Tokumoto and Takahashi, 2001). It is exclusively present in prokaryotes and in eukaryotes of the Apicomplexa, where it is highly conserved (Pastore et al., 2006). Based on its phylogenetic occurrence, IscX seems to depend on the presence of IscS whereas the reverse is not the case (Pastore et al., 2006).

The structure of *E. coli* IscX consists of a classical helix-turn-helix fold often found in transcription regulators (Shimomura et al., 2005; Pastore et al., 2006). *In vitro* studies have shown that IscX is able to bind IscS, thus suggesting a role for IscX as a molecular adaptor (Pastore et al., 2006; Shi et al., 2010; Kim et al., 2014). IscX also binds to iron, although with an affinity at least one order of magnitude weaker than that of CyaY, through a negatively charged surface, the same by which it recognizes IscS (Pastore et al., 2006). An exposed negatively charged iron binding surface which overlaps with the surface of interaction with IscS is also a feature present in another protein, CyaY (the ortholog in bacteria of the eukaryotic frataxin). This protein has attracted much attention because in humans it is associated with Friedreich's ataxia (Pastore and Puccio, 2013). In contrast to IscX, frataxins are proteins highly conserved from bacteria to high eukaryotes (Gibson et al., 1996) and are essential in eukaryotes (Li et al., 1999). In prokaryotes, CyaY is external to the *isc* operon but has extensively been implicated in Fe-S cluster assembly (Huynen et al., 2001). We have in the past shown that CyaY is an IscS regulator, which dictates the enzymatic assembly of Fe-S clusters (Adinolfi et al., 2009; Prischi et al., 2010a). Intriguingly, IscX and CyaY compete for the same site on IscS (Prischi et al., 2010a; Shi et al., 2010; Kim et al., 2014). A genetic interaction between CyaY and IscX was also demonstrated by a recent study which has validated a role of IscX as a new bona fide Fe-S cluster biogenesis factor (Roche et al., 2015). In some species, CyaY and IscX seem to replace each other.

These data raise the compelling question of whether there could be a functional link between CyaY and IscX, which could both elucidate the function of IscX and explain why these two proteins compete for the same binding site. Given the high conservation between prokaryotic and eukaryotic frataxins and desulfurases, answering this question could also inspire new studies on the regulation of eukaryotic frataxin. Using a complementary approach which makes use of enzymology, cross-linking, and structural methods, we provide conclusive evidence indicating that IscX is a modulator of CyaY that switches off the inhibitory properties of CyaY as a function of the iron concentration. At low iron concentrations, IscS is under IscX control which has no inhibitory capacity. At high iron concentrations the system becomes controlled by CyaY. Based on our results, we suggest a general scheme that supports a role of CyaY as an iron sensor and provides the first testable indications concerning the function of IscX.

MATERIALS AND METHODS

Protein Production

All proteins used are from *E. coli*. Their sequences were subcloned in a pET-24d vector modified as fusion proteins with His-tagged glutathione-S-transferase (GST). The constructs

were expressed in BL21(DE3). Bacteria expressing IscU were grown in Luria Broth enriched medium containing 8.3 mg ZnSO₄ (Prischi et al., 2010b) to stabilize its fold. The proteins were purified as previously described (Musco et al., 2000; Pastore et al., 2006) by affinity chromatography using Ni-NTA agarose gel and cleaved from the His-GST tag by Tobacco virus protease (TEV) protease. The mixture was reloaded on Ni-NTA gel to separate the His-tagged GST. This was further purified by gel-filtration chromatography on a Sephadex G-75 column. All the purification steps were performed in the presence of 20 mM β -mercaptoethanol. The protein purities were checked by SDS-PAGE and by mass-spectrometry of the final product. Protein concentrations were determined by absorbance at 280 nm using $\epsilon_{280\text{ nm}}$ of 41,370 or 19,480 or 28,990 M⁻¹ cm⁻¹ for IscS or IscX or CyaY, respectively.

Biolayer Interferometry

All experiments were performed in 20 mM HEPES (pH 7.5), 150 mM NaCl, 2 mM TCEP, 0.5 mg/mL BSA on an Octet Red instrument (ForteBio, Inc., Menlo Park, CA) operating at 25°C. Streptavidin coated biosensors with immobilized biotinylated IscS were exposed to different concentrations of IscX (0–50 μ M). Alternatively, binding affinity between IscS and IscX was evaluated by competition with CyaY. For this experiment, streptavidin coated biosensors with immobilized biotinylated CyaY were exposed to 10 μ M IscS at different concentrations of IscX (0–60 μ M). Apparent K_d values were estimated by fitting the response intensity as a function of the concentration. Since BLI does not give information about reaction stoichiometry, the data were fitted to a simple 1:1 binding model using non-linear least squares methods (Levenberg Marquardt algorithm).

Enzymatic Assays

All enzymatic experiments to form Fe-S clusters on IscU were performed in an anaerobic chamber (Belle technology) under nitrogen atmosphere. The kinetics were followed at 456 nm as a function of time by absorbance spectroscopy using a Cary 50Bio Varian spectrophotometer. The initial rates were measured by incubating 1 μ M IscS, 50 μ M IscU, 250 μ M Cys, 2 mM DTT, and 25 μ M Fe²⁺ in 50 mM Tris-HCl (pH 7.5) and 150 mM NaCl. When added CyaY was 5 μ M. The reactions were initiated after half an hour incubation by addition of 1 μ M IscS and 250 μ M Cys. Each measurement was repeated at least five times on different batches of proteins. The data were always reconfirmed by CD performed under the same conditions.

Cross Linking Assays

A mixture of 8 μ M IscS was mixed with increasing molar ratios of IscX (0–80 μ M) in PBS buffer. Bis[sulfosuccinimidyl]suberate was added to the protein to a final concentration of 2.5 mM. The reaction mixture was incubated at room temperature for 30 min and quenched by adding Tris-HCl (pH 8) to a final concentration of 20 mM. The quenching reaction was incubated at room temperature for 15 min. Identification of cross-linked peptides was performed using a classic mass-based peptide mapping approach (see Supplementary Material). A further confirmation of the identity of the cross-linked peptides was achieved using

the StravoX software, performing the “shuffle sequences but keep protease site” decoy analysis.

Mass Spectrometry under Non-denaturing Conditions

Solutions of IscS ($3\ \mu\text{M}$) were mixed with increasing concentrations of IscX (0–16 IscX/IscS molar ratios). Samples were incubated at room temperature for 5 min before being loaded in a 500 μl gas-tight syringe (Hamilton) and infused directly in a Bruker micrOTOF-QIII mass spectrometer (Bruker Daltonics) operating in the positive ion mode. The ESI-TOF was calibrated online using ESI-L Low Concentration Tuning Mix (Agilent Technologies) and subsequently re-calibrated offline in the 4,000–8,000 m/z region. MS data were acquired over the m/z range 4,000–8,000 continuously for 10 min. For LC-MS, an aliquot of IscS or IscX was diluted with an aqueous mixture of 2% (v/v) acetonitrile, 0.1% (v/v) formic acid, and loaded onto a ProsSwift RP-1S column ($4.6 \times 50\ \text{mm}$, Thermo Scientific) attached to an Ultimate 3000 uHPLC system (Dionex, Leeds, UK). Processing and analysis of MS experimental data were carried out using Compass DataAnalysis v4.1 (Bruker Daltonik). Neutral mass spectra were generated using the ESI Compass v1.3 Maximum Entropy deconvolution algorithm. Titration data were fitted using the program DynaFit (BioKin, CA, USA). For further details see Supplementary Material.

SAXS Measurements

SAXS data were collected on the EMBL P12 beamline on the storage ring PETRA III (DESY, Hamburg, Germany; Blanchet et al., 2015). Solutions of IscX-IscS complexes were measured at 25°C in a concentration range 0.5–10.0 mg/mL at 1:1, 1:2, 1:20, and 1:40 molar ratios. The data were recorded using a 2 M PILATUS detector (DECTRIS, Switzerland) at a sample-detector distance of 4.0 m and a wavelength of $\lambda = 0.124\ \text{nm}$, covering the range of momentum transfer $0.04 < s < 5.0\ \text{nm}^{-1}$ ($s = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle). No measurable radiation damage was detected by comparison of 20 successive time frames with 50 ms exposures. The data were averaged after normalization to the intensity of the transmitted beam and the scattering of the buffer was subtracted using PRIMUS (Konarev et al., 2003). For further details on data treatment see Supplementary Material.

RESULTS

IscX Has Different Effects on the Kinetics of Cluster Formation as a Function of Concentration

We started by exploring the effect of IscX on the rates of enzymatic assembly of the Fe-S cluster on IscU using an assay in which the cluster forms, under strict anaerobic conditions, through IscS-mediated conversion of cysteine to alanine and persulfide and is reconstructed on IscU (Adinolfi et al., 2009). We performed the experiment at increasing concentrations of IscX in the range 1–50 μM using 1 μM IscS, 50 μM IscU, 250 μM Cys, 2 mM DTT, and $25\ \mu\text{M}\ \text{Fe}^{2+}$. We did not observe significant

variations of the kinetics (Figures 1A,B) up to ca. 10 μM , a range in which the initial rates in the presence or absence of IscX were superposable, revealing that IscX does not have inhibitory effects under these conditions. Higher concentrations of IscX (from 20 to 50 μM , i.e., higher IscX/IscS ratios) led instead to inhibition of the reaction. Similar observations were

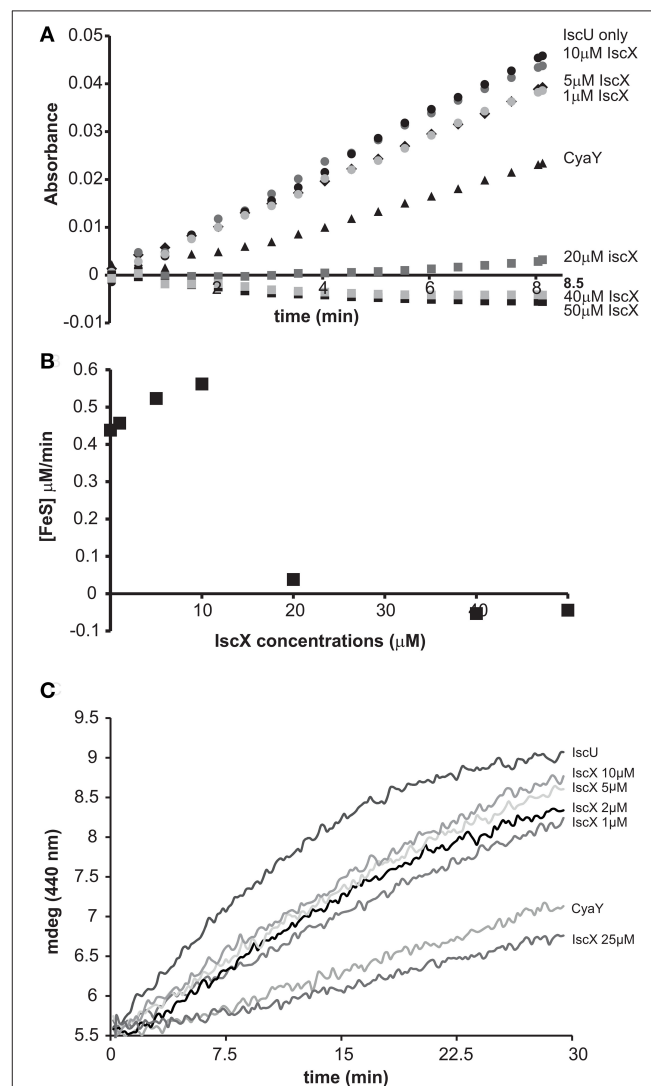


FIGURE 1 | Effect of increasing concentrations of IscX on the enzymatic kinetics of Fe-S cluster formation on IscU. **(A)** Kinetics of cluster formation at increasing concentrations of IscX as measured by absorbance at room temperature. **(B)** Plot of the initial rates at increasing concentrations of IscX. Negative values are due to imperfect subtraction from the baseline and should be considered as zero (full inhibition). Only the initial part of the kinetics for each IscX concentration point was considered. The slope of these linear regions (for each IscX concentration) was used as the initial velocity. A molar extinction coefficient of $10.5\ \text{mM}^{-1}\ \text{cm}^{-1}$ was assumed to convert the absorbance in concentration of Fe-S cluster produced. **(C)** Kinetics followed by CD. The assays were carried out using 1 μM IscS, 50 μM IscU, 250 μM Cys, 2 mM DTT, and $25\ \mu\text{M}\ \text{Fe}^{2+}$ and increasing concentrations of IscX as indicated. For comparison, the experiment was also repeated in the presence of 5 μM CyaY. The error bars are not shown for sake of clarity but were typically within 5% or less.

made using circular dichroism (CD) (**Figure 1C**). The results at higher IscX/IscS ratios are in agreement with a previous study (Kim et al., 2014) in which the kinetics were performed at high concentrations of both Fe^{2+} (125 μM) and IscX (25 μM), but the results at lower IscX/IscS ratios are new and surprising. Why does IscS have a different behavior at low and high ratios? There are different hypotheses which could explain these observations. First, the effect we observed at high IscX/IscS ratios could be explained by considering that IscX binds to iron and at high iron:IscX molar ratios (i.e., 80:1) aggregates (Pastore et al., 2006; Prischi et al., 2010a). This could lead to a loss of available iron which could be sequestered by IscX. However, it was noted that aggregation occurs only when the assay is carried out at very low ionic strength. When salt is present, as was the case here, aggregation is not observed. Second, at low ratios the IscX occupancy on IscS could be too low to detect an effect because the complex has low affinity. Third, there could be a secondary binding site for IscX on IscS which is occupied only when the primary site is fully occupied. Only the 2:1 IscX-IscS complex would behave as an inhibitor.

The Presence of IscU Does Not Influence the Affinity of IscX for IscS

To test the complex occupancy, we reconsidered the dissociation constants. We had previously estimated by fluorescence labeling and calorimetry dissociation constants (K_d s) of 12 and 20 μM for the binary IscX-IscS and CyaY-IscS complexes respectively (Pastore et al., 2006; Prischi et al., 2010a). The IscX-IscS complex should thus be >50% populated under the conditions of our cluster formation assay. However, the presence of IscU could in principle modify these affinities as it is observed for binding to IscS when both IscU and CyaY are present. Also, although CyaY and IscX compete for the same site of IscS, the IscS dimer could have, at low IscX-IscS ratios, simultaneous occupancy of both CyaY and IscX. We used Biolayer Interferometry (BLI), a technique which can measure weak molecular interactions between several partners, to test this possibility. When we immobilized IscS on the surface and titrated with IscX only, we obtained a K_d of $8 \pm 3 \mu\text{M}$ for the IscX-IscS binary complex, in excellent agreement with the previous measurements (data not shown). When we immobilized CyaY on the surface, saturated with IscS and titrated with increasing quantities of IscX we obtained a K_d of $6 \pm 2 \mu\text{M}$ for the IscX-IscS complex (**Figures 2A,B**). Thus, the presence of CyaY on IscS does not modify the affinity of IscX for the desulfurase, and so major allosteric effects or binding cooperativity between the two proteins are unlikely. When we tested binding of IscX to the IscU-IscS complex (immobilizing IscS saturated with IscU), we obtained a K_d of $8 \pm 2 \mu\text{M}$, a value comparable to the one obtained in the absence of IscU (**Figures 2C,D**). These data indicate that the presence of IscU bound to IscS does not affect the affinity for IscX and that the affinity of IscX for IscS is higher than that of CyaY. Thus, the effect observed at different IscX-IscS ratios cannot be ascribed to insufficient occupancy.

IscX Has Two Different Binding Sites on IscS

BLI proved inconclusive toward the presence of a secondary binding site. This could either be because we did not explore sufficiently high molar ratios, or because BLI depends on there being a significant change in the distance between the sensor's internal reference layer and the solvent interface. For formation of some complexes this can be rather small and thus undetectable. We thus used electrospray ionization (ESI) mass spectrometry (MS) under non-denaturing conditions, a technique which, if optimized, provides direct information on all complexes present in a solution. The m/z spectrum (4,100–6,300 m/z) of IscS displayed well-resolved charge states (**Figure S1A**). The deconvoluted spectrum of IscS revealed a major peak at 91,035 Da consistent with the presence of dimeric IscS (predicted mass 91,037 Da) (**Figure S1B** and **Table S1**). Occasionally, shoulder peaks at +32 Da intervals on the high mass side of the peak were observed; these are likely due to the presence of one or more sulfane sulfur atoms, as previously observed for other proteins (Crack et al., 2017). An unknown adduct at +304 Da relative to the main IscS peak was also observed (**Figure S1B**). Addition of IscX to IscS to an 8:1 IscX/IscS molar ratio resulted in significant changes in the m/z spectrum. A complex pattern of new charge states, superimposed over those of IscS, were observed, consistent with the presence of IscX-IscS complexes (**Figure S2**). The deconvoluted spectrum (spanning 90 to 125 kDa) revealed the presence of four new species, corresponding to dimeric IscS in complex with up to 4 IscX molecules (at intervals of ~ 7.9 kDa) (**Figure 3A**). The observed mass of IscX was 7935 Da, +76 Da higher than the predicted mass of 7859 Da. LC-MS showed that a minor amount (<1%) of IscX was detected at the expected mass of 7859 Da, suggesting a covalent modification of the majority of the IscX protein. This is most likely due to adduct formation of IscX with β -mercaptoethanol during purification to give a mixed disulfide, which would have a predicted mass of 7935 Da ($7859 + 78 - 2 = 7935$ Da). Collision induced dissociation (CID) data supported this conclusion (**Figure S3**).

When increasing concentrations of IscX were added to IscS, we observed the gradual formation of dimeric IscS complexes containing 1 to 4 IscX monomers (**Figures 3B–E**). The complex of IscS with a single IscX protein, $(\text{IscX})(\text{IscS})_2$, formed readily at low levels of IscX ($[\text{IscX}]/[\text{IscS}] \approx 0.3$) and maximized at $[\text{IscX}]/[\text{IscS}] \approx 4$ (**Figure 3B**). The $(\text{IscX})_2(\text{IscS})_2$ complex was detectable at a $[\text{IscX}]/[\text{IscS}]$ ratio ≈ 1 and reached maximum abundance by $[\text{IscX}]/[\text{IscS}] \approx 8$ (**Figure 3C**). $(\text{IscX})_3(\text{IscS})_2$ and $(\text{IscX})_4(\text{IscS})_2$ complexes were evident at $[\text{IscX}]/[\text{IscS}]$ ratios ≈ 4 and ≈ 8 , respectively (**Figures 3D–F**). The data were analyzed according to a sequential binding model. The resulting fit of the data revealed that binding of the first two IscX molecules to form $(\text{IscX})(\text{IscS})_2$ and $(\text{IscX})_2(\text{IscS})_2$ occurs with a similar affinity, $K_d = 13.7 \pm 0.4 \mu\text{M}$, consistent with the binding constant obtained by fluorescence and BLI. Binding of the third and fourth IscX molecules [to form $(\text{IscX})_3(\text{IscS})_2$ and $(\text{IscX})_4(\text{IscS})_2$] was also found to occur with comparable affinity, with a $K_d = 170.0 \pm 10 \mu\text{M}$, that is significantly lower than the primary site. That the binding behavior is well described by two pairs of dissociation

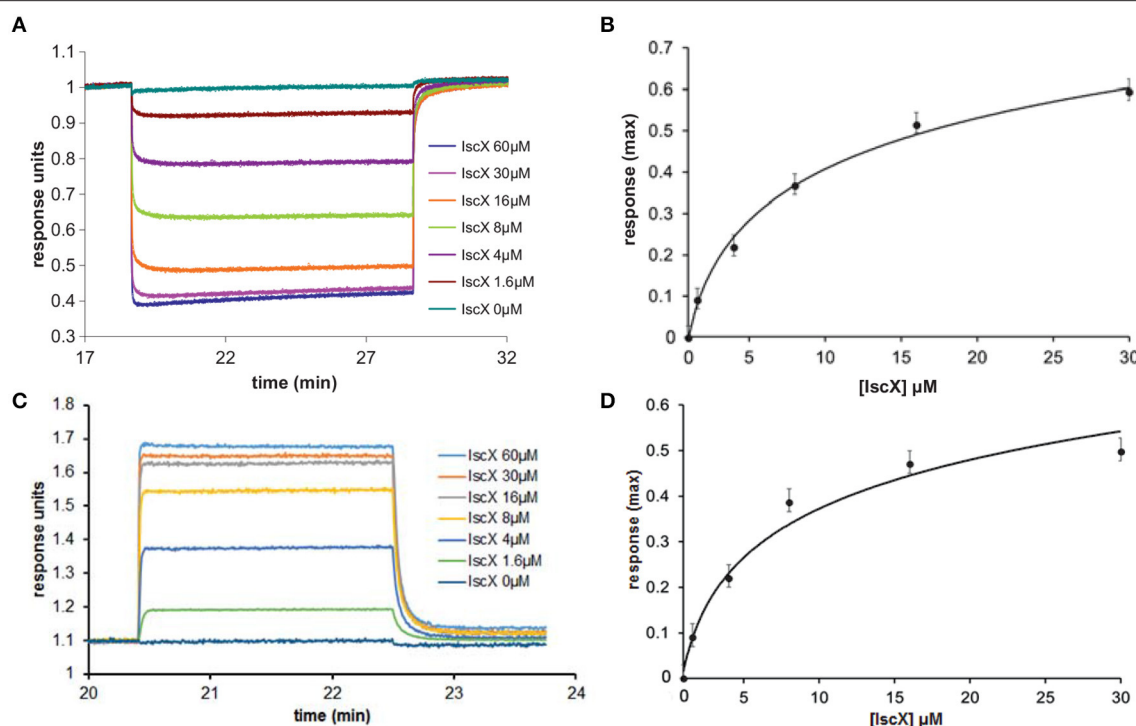


FIGURE 2 | Quantification of the binding affinities of the IscX-IscS-IscU and IscX-IscS complexes to assess the influence of a third component. **(A)** Plot of the interferometry response obtained by immobilizing CyaY saturated by IscS (10 μ M) and titrating with increasing concentrations of IscX (0–60 μ M). **(B)** Plot of the response reduction as a function of IscX concentration. **(C)** Plot of the interferometry response obtained by immobilizing IscS saturated by IscU (10 μ M) and titrating with increasing concentrations of IscX (0–60 μ M). **(D)** Plot of the response as a function of IscX concentrations.

constants indicates that the IscS dimer contains two equivalent pairs of IscX binding sites, with one primary (high affinity) and one secondary (low affinity) site per IscS protein.

This evidence conclusively supports the presence of two binding sites for IscX on IscS.

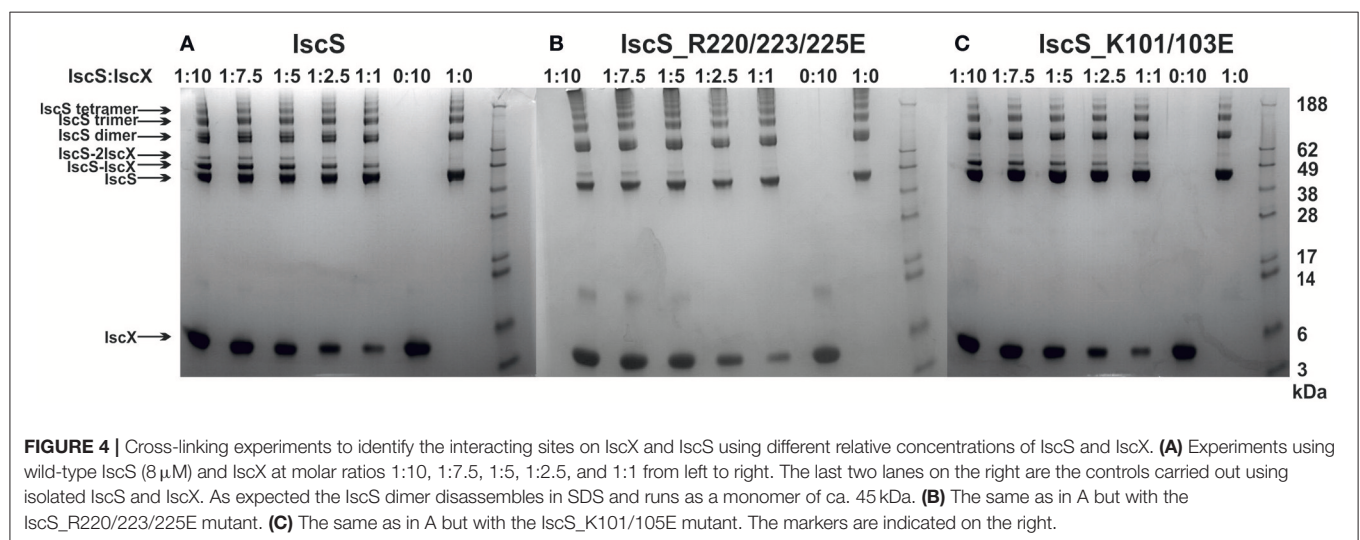
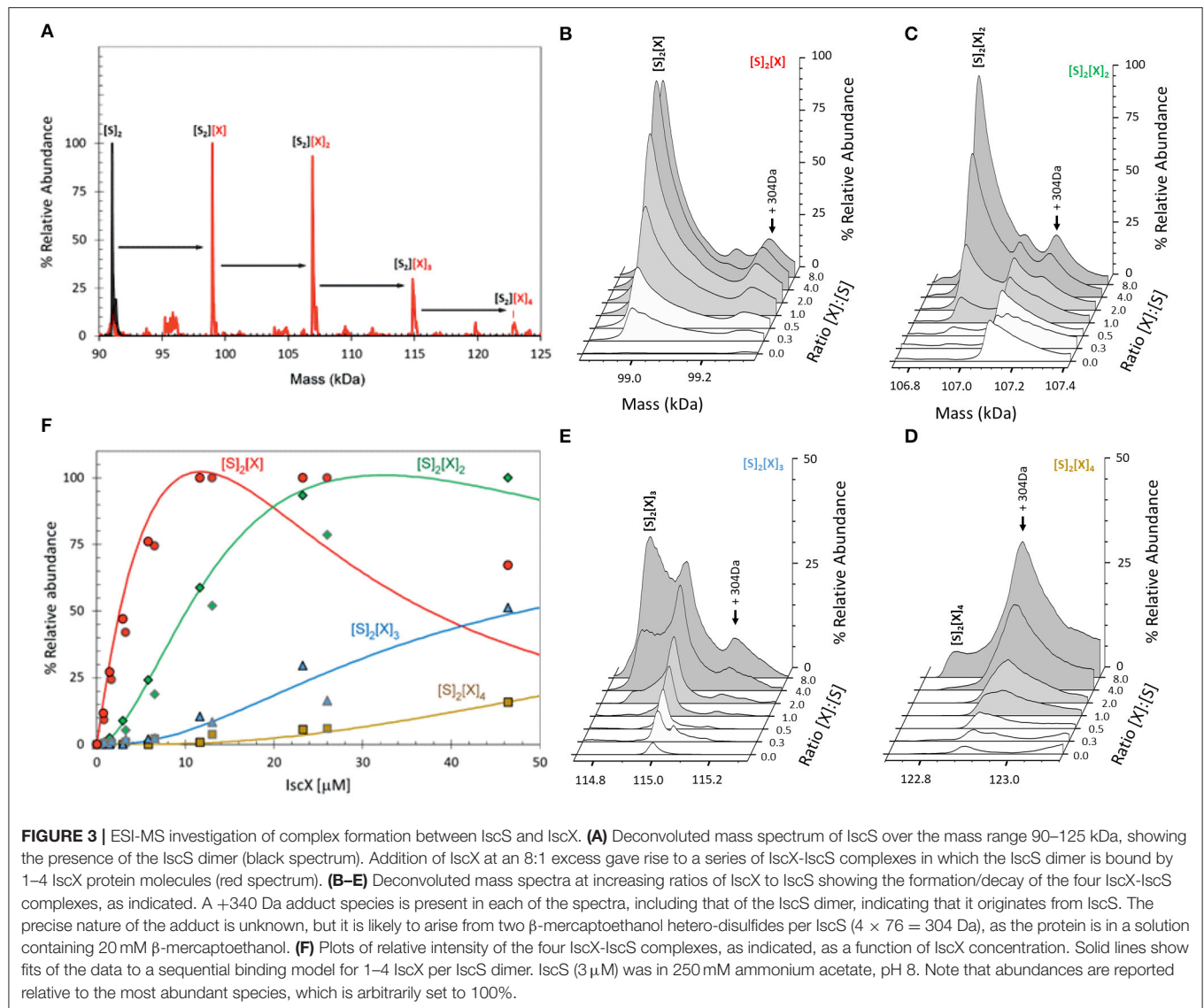
Mapping the Two Binding Sites by Cross-Linking

We next used cross-linking experiments to map the IscX binding sites on IscS, attempting to transform the non-covalent interaction in a stable covalent bond. In this assay, a cross-linking agent was added to link covalently proteins in close spatial proximity (Watson et al., 2012). We used bis[sulfosuccinimidyl]suberate (BS3), a cross-linking agent that reacts with primary amino groups up to ~ 11.5 Å apart, to mixtures of IscS and IscX at different molar ratios. The reaction produced two covalently attached protein complexes with apparent molecular weights of 52 and 60 kDa, tentatively identified on PAGE as 1:1 and a 2:1 IscX-IscS complexes respectively (Figure 4). Mass spectrometry (MS) confirmed the presence of only these two proteins in both bands. Identification of the cross-link nature was achieved by subjecting to trypsin digestion the two species, as well as the isolated IscS. The obtained peptide mixtures were then analyzed by MALDI/MS (Table S2). Comparison between the spectra acquired for IscS and for the two complexes allowed us to detect a signal at m/z

3121.5 ± 0.8 in the lower band which corresponds to the peptide 85–101 of IscS bound to the 1–9 region of IscX (theoretical mass 3120.6 Da). This value is uniquely compatible with a BS3-mediated cross-link involving Lys4 of IscX and Lys92, Lys93, or Lys95 of IscS. This identification was confirmed by HR LC-MS analyses and MS/MS data (Figures S4, S5). In the upper band, we could identify the same fragment together with another one (m/z 2424.2 \pm 0.5), which involves cross-linking between Lys101 of IscS and Lys4 of IscX (theoretical mass 2423.3 Da). As a control, we used an IscS mutant (IscS_R220/223/225E) which abolishes the interactions of IscS with IscX and CyaY (Prischi et al., 2010a; Kim et al., 2014). We also used IscS_K101/105E which includes one of the lysines implicated in cross-linking with IscX. Formation of the IscX-IscS complex was strongly affected in both cases: binding was almost completely abolished with IscS_R220/223/225E, whereas the secondary binding site was barely detectable with the IscS_K101/105E mutant (Figure 4). As a control, when the same cross-linker was used with mixtures of CyaY and IscS only one band was observed (data not shown). These results map the positions of the two binding sites on IscS.

Structural Characterization of the Secondary Binding Site of IscX on IscS

Since it is difficult to assess the significance of these binding sites without a working structural model, we used small-angle X-ray scattering (SAXS) to grasp the shape of the 1:1 and 2:1



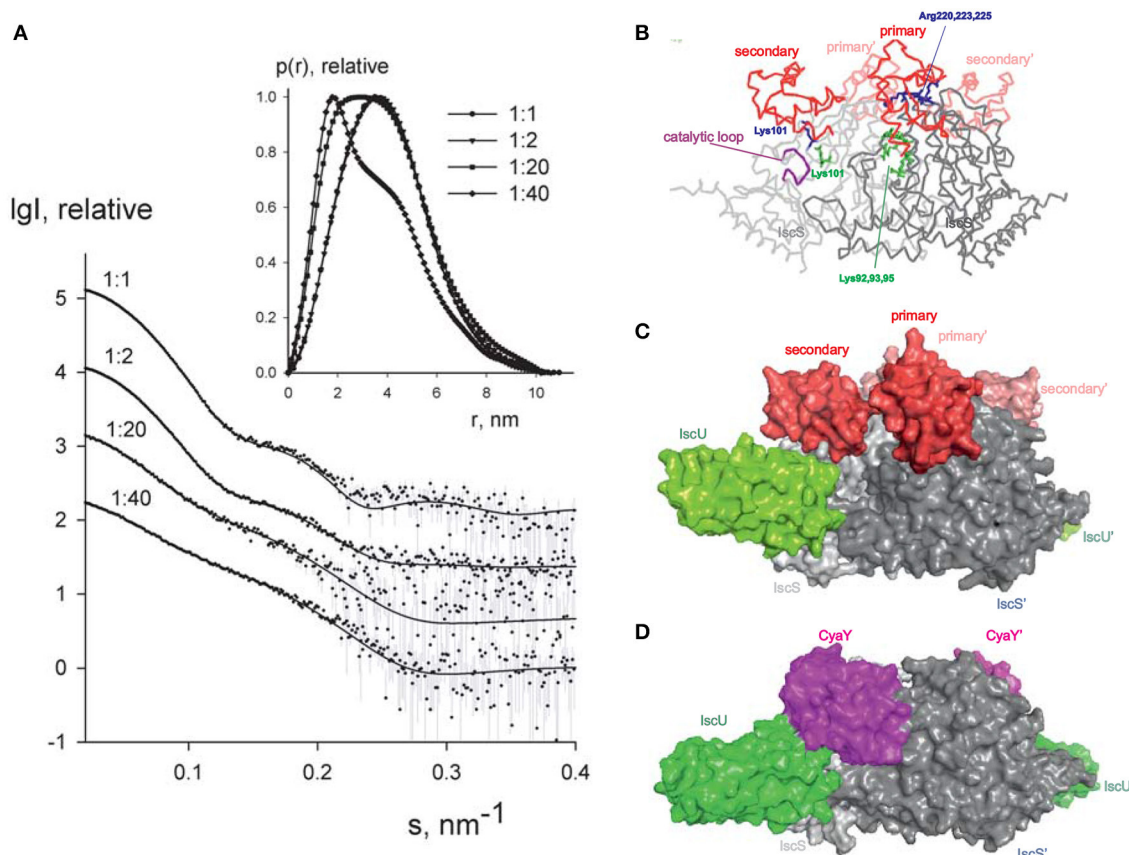


FIGURE 5 | SAXS measurements. **(A)** X-ray scattering patterns for IscX-IscS at molar ratios of 1:1 (at 5 mg/ml solute concentration), 2:1 (at 3 mg/ml), 20:1 and 40:1 (both at 0.5 mg/ml). The experimental data are displayed as dots with error bars. The scattering from a mixture of two rigid body models (1:1 and 2:1 IscX-IscS complexes) obtained by SASREFMX (for data with 1:1 molar ratio) and the fits by mixtures of unbound IscX, IscS dimers, and 1:1/2:1 rigid body complexes (for data at 2:1, 20:1, and 40:1 molar ratios) as obtained by OLIGOMER are shown by solid lines. The plots display the logarithm of the scattering intensity as a function of the momentum transfer. The distance distribution functions are indicated in the insert. **(B)** Backbone representation of the 2:1 IscX-IscS complex. The two IscS protomers are shown in different shades of gray. Residues which are either involved in cross-linking or have an effect on binding are indicated. The catalytic loop of IscS is indicated in purple. **(C)** Same as in **(B)** in a full atom representation. For reference, IscU (in green) is included (3LVL) to clarify the relative positions of the components. **(D)** Comparison with the model previously obtained for the ternary complex CyaY-IscS-IscU (Shi et al., 2010). Light and dark gray represent each of the protomers of the IscS dimer.

complexes. We recorded the scattering intensity patterns $I(s)$ as a function of the momentum transfer s and the computed distance distribution function of IscX-IscS data (at 1:1 molar ratio, 5 mg/ml; **Figure 5A**). The symmetric shape of the distance distribution function $p(r)$ (**Figure 5A, inset**) is typical of a globular protein. The D_{\max} value of the IscX-IscS complex (11.0 ± 0.5 nm) is comparable with that found for the CyaY-IscS complex and significantly lower than that observed for the IscU-IscS complex (12.1 nm). This is compatible with an arrangement in which IscX does not bind along the main axis of the IscS dimer but laterally increasing the IscS shape globularly (Shi et al., 2010). The experimental R_g and excluded volume (V_{Porod}) (3.09 ± 0.04 and 135 ± 10 nm³, respectively) are almost equal to the IscS homodimer (3.09 ± 0.04 and 136 ± 10 nm³, respectively) which suggests a partial dissociation of the complex (Table S3). The molecular mass (92 ± 5 kDa) is also lower than that expected for a 1:2 IscX-IscS complex with full occupancy

(117 kDa). This is in agreement with the presence of unbound species in solution due to the weak binding constants. The R_g and V_{Porod} significantly drop when excesses of IscX were used (3.01 ± 0.03 nm/56 ± 5 nm³ for the 20:1 ratio and 2.75 ± 0.03 nm/31 ± 4 nm³ for the 40:1 ratio). The higher the volume fraction of small species in the mixture (i.e., IscX molecules), the smaller is the average estimated excluded volume of the system. To further prove complex formation upon addition of IscX, all scattering curves recorded at different concentrations and different stoichiometries were analyzed using a model-independent Singular Value Decomposition (SVD) approach (Golub and Reinsch, 1970). SVD clearly points to the presence of four components contributing to the scattering data in agreement with the presence of free IscS, free IscX, and their 1:1 and 2:1 complexes (Figure S6). To construct a low resolution model of the IscX-IscS complex, rigid body modeling was used taking into account the possibility of 1:1 and 2:1 complexes. The interaction

sites of IscS and IscX determined by NMR and cross-linking experiments were used as distance restraints between IscX and the possible binding sites of IscS (Pastore et al., 2006; Kim et al., 2014). Multiple runs of rigid body modeling program SASREFMX starting from random initial configurations with applied P2 symmetry yielded models of the IscS/IscX complexes fitting well the scattering data taken at a 1:1 molar ratio with a discrepancy $\chi^2 = 1.7$. The presence of unbound (free) components was taken into account by using free IscS dimers and IscX monomers as independent components in addition to the SASREFMX models. Their relative volume fractions were further refined with the program OLIGOMER to best fit all the available data (Table S4). The volume fraction of free IscS decreases at a large excess of IscX consistent with an almost full occupancy of IscX. While it is difficult to distinguish between the fitting for 1:1 and 2:1 complexes, the fits assuming both types of complexes are somewhat better than those obtained assuming the presence of 2:1 complexes only. We estimated ca. 60–70% of 2:1 IscX-IscS complexes. The structure of the 1:1 complex is in excellent agreement with a previous model obtained by the same hybrid technique (Kim et al., 2014), IscX sits in the catalytic pocket of IscS contributed by both protomers of the IscS dimer (Figures 5B,C). This is the site where also CyaY binds (Prischi et al., 2010a; Kim et al., 2014; Figure 5D) thus explaining the direct competition between the two molecules. Experimental validation of this binding site has been supported by mutation studies of IscS (Shi et al., 2010). The second site is close but shifted toward the IscU-binding site.

The presented models of 2:1 and 1:1 IscX:IscS complexes should be treated as tentative given the potential ambiguity of SAXS modeling and the limitations of rigid body approach assuming that no conformational rearrangements are observed upon complex formation. Still, the obtained models permitted us to neatly fit the complete SAXS data set recorded at different conditions, agree with the cross-linking data and also offer a logical explanation for the observed effects. We have previously demonstrated by molecular dynamics that binding of CyaY in the primary site restricts the motions of the catalytic loop which is thought to transfer persulfide, bound to Cys237, from the catalytic site to IscU (di Maio et al., 2017). As compared to CyaY, IscX is smaller and would not be sufficient to block the movement of the catalytic loop as long as only the primary site of IscS is occupied. In this position, IscX simply prevents CyaY from binding, having higher affinity, but does not act as an inhibitor. However, binding of another IscX molecule with its N-terminus close to K101, as observed by cross-linking, can strongly increase the steric hindrance and result in inhibition of iron sulfur cluster formation resulting in an effect similar to that observed for CyaY.

Understanding the Competition between IscX and CyaY

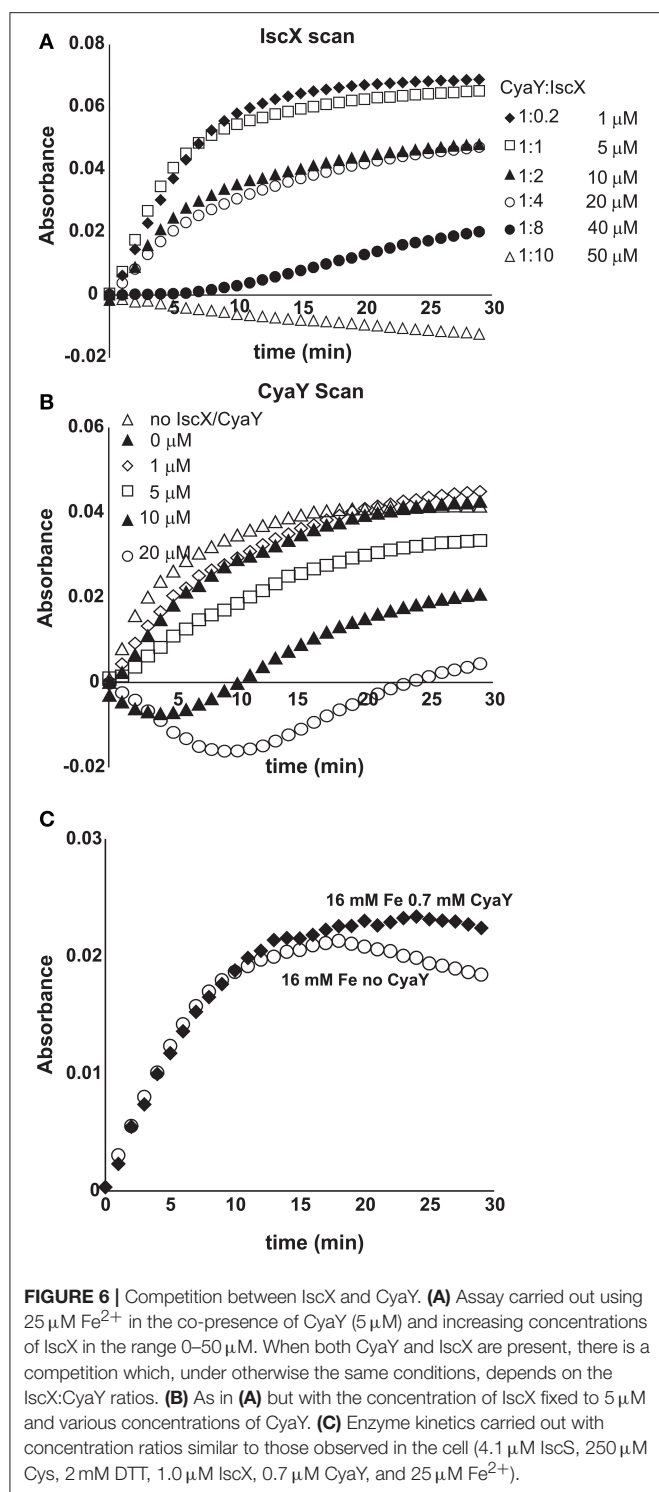
We then explored more closely the conditions under which IscX competes with CyaY (Shi et al., 2010; Kim et al., 2014). We first repeated the enzymatic assays keeping the conditions unchanged (1 μ M IscS, 50 μ M IscU, 250 μ M Cys, 2 mM DTT, and 25 μ M Fe^{2+}), adding CyaY (5 μ M) and progressively increasing the

concentration of IscX in the range 0–50 μ M. Below 10 μ M (i.e., below 2:1 IscX-CyaY and 10:1 IscX-IscS molar ratios), where IscX has no effect on IscS activity as demonstrated above, the experiment results in a progressive rescuing of CyaY inhibition (Figure S7), in agreement with a competition between the two proteins for the same binding site on IscS. The effect of IscX becomes noticeable at substoichiometric concentrations as compared to CyaY, as expected from the respective dissociation constants. However, even under the maximal condition of rescuing activity (10 μ M), IscX is able to reach only ca. 70% of the rate compared to a control experiment performed in the absence of CyaY where there is no inhibition. This reflects the fact that, at relatively high IscX/IscS ratios, the secondary binding site, which also leads to inhibition, starts being populated probably favored by a conformational change or an electrostatic rearrangement. Above a 10:1 IscX/IscS ratio, increase of the IscX concentration results in progressively marked inhibition (Figure 6A). When we titrated instead the system with CyaY (0–20 μ M), keeping IscX fixed at 5 μ M, we started observing a decrease of the intensities at 1:1 IscX/CyaY molar ratios and a marked inhibition for higher CyaY ratios (Figure 6B).

We then put our results within the context of the protein concentrations observed in *E. coli*. Several independent studies have reported these under different growth conditions (Ishihama et al., 2008; Taniguchi et al., 2010; Li et al., 2014; Schmidt et al., 2016). Under non-stress conditions, both CyaY and IscX are usually, but not always, present at substoichiometric ratios as compared to IscS, with IscX in excess over CyaY (Li et al., 2014). Different conditions can however remarkably change the molar ratios. We thus repeated the enzymatic assay using concentrations comparable to the non-stress conditions to verify that IscX controls the reaction under these conditions (4.1 μ M IscS, 250 μ M Cys, 2 mM DTT, 1.0 μ M IscX, 0.7 μ M CyaY, and 25 μ M Fe^{2+}). We used IscU in a large excess (50 μ M) because it is the reporter. We preferred to use this protein rather than another reporter because the IscU binding site on IscS does not interfere with CyaY or IscX (Pastore et al., 2006; Prischi et al., 2010a). At these ratios we do not observe inhibition indicating that the reaction is under the control of IscX (Figure 6C). Taken together, these results fully confirm the competing role of IscX and CyaY.

The Modulation Effect of IscX and CyaY on Cluster Formation on IscU Is Iron Dependent

Finally, we tested the effect of the competition between CyaY and IscX on Fe-S cluster formation on IscU as a function of increasing iron concentrations keeping the other concentrations unchanged (1 μ M IscS, 50 μ M IscU, 250 μ M Cys, 2 mM DTT) because we had already observed that CyaY inhibition is enhanced by iron (Adinolfi et al., 2009; Figures 7A,B). As a control, we compared the variation of the initial rates as a function of iron concentration in the presence of IscU or CyaY individually. Cluster formation in the presence of the acceptor IscU only increased with the concentration of iron. The presence of CyaY (5 μ M) drastically reduced the rate of cluster formation. The difference between the



initial rates in the absence and in the presence of CyaY increased at increasing concentrations of iron. Addition of IscX to the reaction together with CyaY (1 and 5 μM respectively) generated a “rescuing” effect but only at low iron concentrations. Under these conditions, the effect reached a maximum around 15 μM

iron. At higher iron concentrations, the curve obtained in the presence of both CyaY and IscX became similar to that obtained with CyaY alone. This implies that, at lower iron concentrations, IscX has higher affinity for IscS than does CyaY, impairing the ability of CyaY to inhibit cluster formation. The effect is reversed at higher iron concentrations where CyaY must acquire a higher affinity.

To dissect whether iron dependence is controlled by CyaY, IscX or both proteins together, we fixed the concentration of IscX to 1 μM and repeated the experiments varying the concentrations of Fe^{2+} from 10 to 100 μM in the absence of CyaY and in the presence or in the absence of IscX. We observed that the initial rates increased at increasing Fe^{2+} concentrations but were independent of the presence of IscX (**Figure 7C**). Thus, the interaction of IscX to IscS is Fe^{2+} independent in this range of concentrations. This is at variance with CyaY whose interaction with IscS is clearly iron mediated, as previously demonstrated (Adinolfi et al., 2009): the same experiment carried out with CyaY showed a marked difference in the absence and presence of CyaY which increases with iron.

We thus must conclude that IscX acts as a modulator of CyaY inhibition, resulting from the ability of CyaY to sense iron and bind to IscS more tightly at high iron concentrations.

DISCUSSION

Since its discovery (Beinert, 2000), iron-sulfur cluster biogenesis has rapidly become an important topic of investigation both because it is a molecular machine at the very basis of life and because of its crucial role in an increasing number of human diseases (Rouault, 2012). The study of this topic in the bacterial system is also specifically important given the important role that iron-sulfur cluster in bacterial infections (Bridwell-Rabb et al., 2012). Much progress has been made in understanding this metabolic pathway but several questions have so far remained unanswered. Why does IscS bind different proteins using the same surface and how is recognition regulated? What is the role of IscX in the *isc* machine? What is the relationship between IscX and CyaY? We have now tentative answers to these questions.

Previous studies carried out at high IscX concentrations (>10 μM) led to the conclusion that IscX, like CyaY, is an inhibitor of cluster formation (Kim et al., 2014). We could reproduce these results but demonstrated that they strongly depend on the relative and absolute concentrations used *in vitro*. At high IscX-IscS ratios (i.e., >10:1), the inhibitory effect of IscX is the consequence of a secondary binding site on IscS which becomes populated only once the primary site is saturated. At low IscX-IscS ratios, IscX has no effect on cluster formation but efficiently competes with CyaY and modulates its strong inhibitory power also at substoichiometric ratios.

This is an important result which changes our perspective on IscX: this protein is not just “another frataxin-like protein” but the modulator of the inhibitory function of CyaY in bacteria; it is what silences CyaY. Our data are fully consistent with and explain in molecular terms a previous

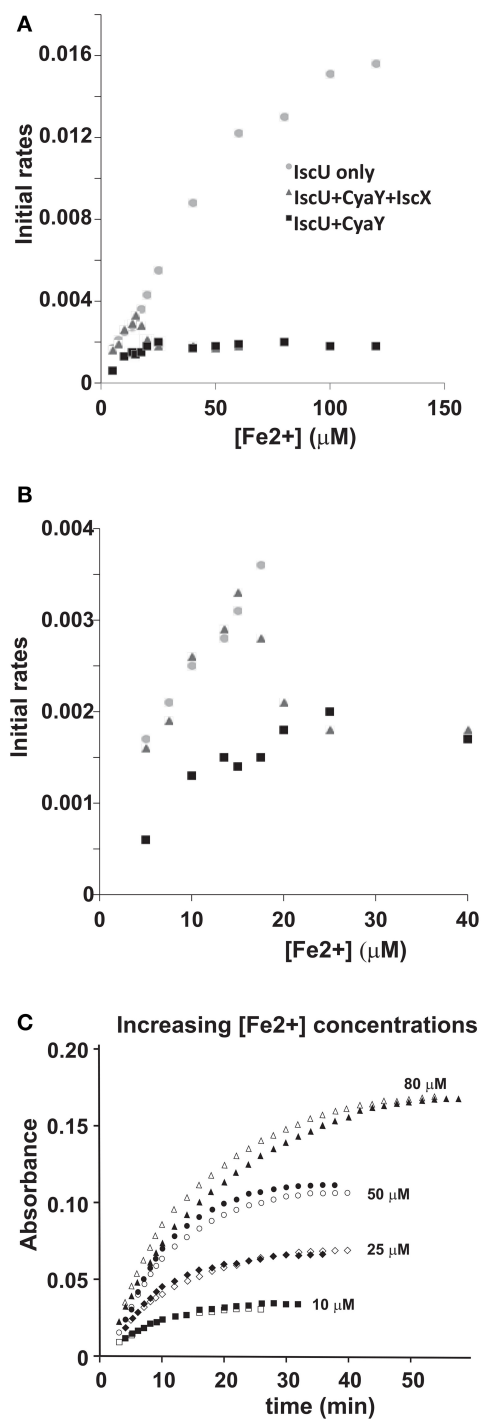


FIGURE 7 | Iron dependence of the relative effect of CyaY and IscX on cluster assembly enzymatic kinetics on IscU. **(A)** Initial rates of kinetics of cluster formation on IscU as a function of iron concentration in the presence of IscU only (gray circle), IscU and CyaY (black rectangles) and IscU, IscX, and CyaY (gray triangles). **(B)** The same as in **(A)** but expanded for clarity. **(C)** Assays carried out in the absence and in the presence of IscX (1 μM) at increasing concentrations of iron (10 μM, 25 μM, 50 μM, 80 μM) to assess whether IscX is affected by iron. No CyaY was added to the assay. All experiment contained 1 μM IscS, 50 μM IscU, 250 μM Cys, 2 mM DTT.

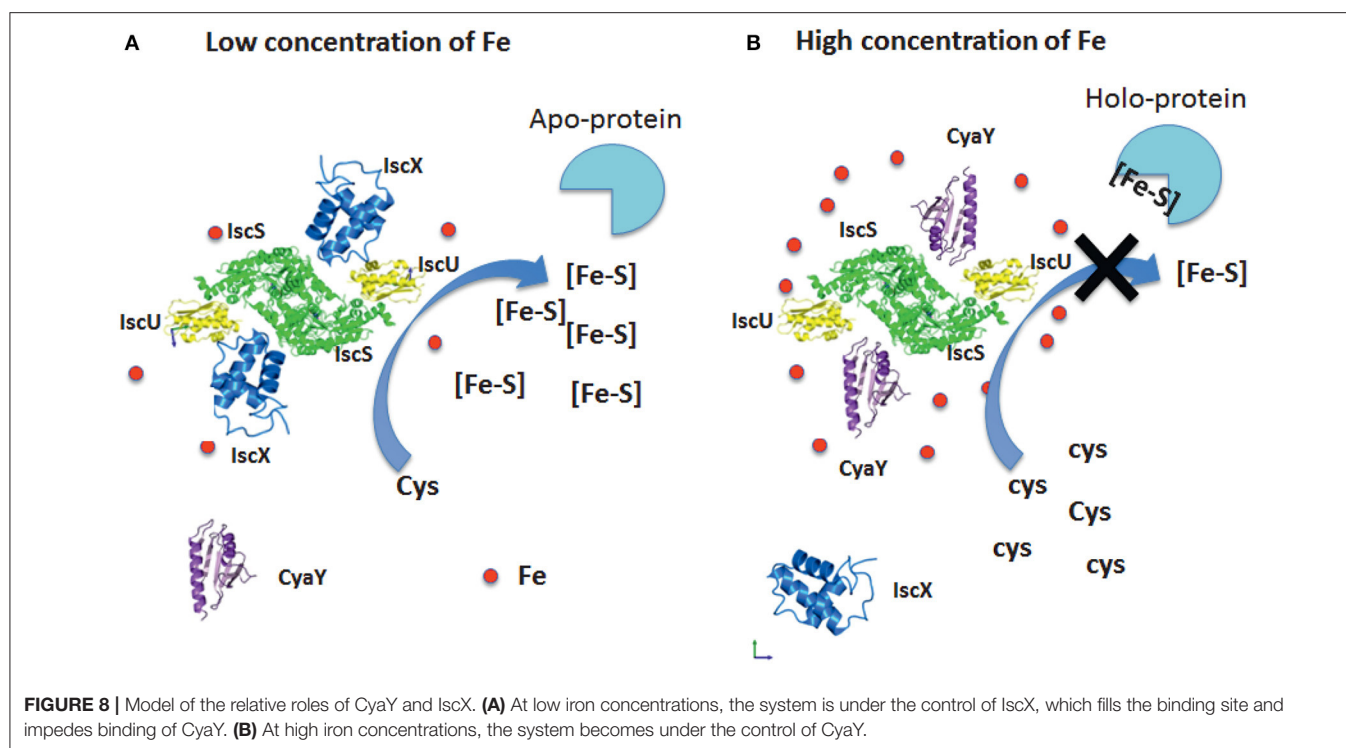
report that conclusively demonstrated a genetic interaction between CyaY and IscX and an additive positive effect on cluster maturation upon deletion of both genes (Roche et al., 2015).

Although the tentative model of 2:1 IscX-IscS complex constructed from SAXS and cross-linking experiments has its limitations, it clearly supports the presence of a secondary binding site and explains the enzymatic assays. The two proteins compete for the same binding site on IscS but IscX is appreciably smaller than CyaY. In the complex with only the primary site occupied, IscX can thus allow free movement of the catalytic loop, which transports the persulfide from the active site to IscU (di Maio et al., 2017). Occupancy of the secondary binding site would interfere instead with the loop movement by steric hindrance, thus producing an inhibitory effect. We also demonstrated that the effect of IscX on cluster formation depends on the iron concentration and that the iron sensor is CyaY and not IscX, whose effect on IscS is iron independent.

We can thus suggest a model based on our results which fully explains the role of both proteins (**Figure 8**). At low iron concentrations, cluster formation would be under the control of IscX and thus continue unperturbed. At higher concentrations, CyaY would have higher affinity for IscS and thus control the cluster formation rate. Additionally, it is likely that, under iron stress conditions, CyaY is up regulated (or IscS down regulated). Independent studies have indeed reported the concentrations of proteins in *E. coli* (Ishihama et al., 2008; Taniguchi et al., 2010; Li et al., 2014; Schmidt et al., 2016) under different growth conditions. Under normal conditions, IscX and CyaY are in substoichiometric ratios as compared to IscS with CyaY being also lower than, or equimolar, to IscX. The concentrations of the components can however vary considerably (up to 6–8 times) (Ishihama et al., 2008; Taniguchi et al., 2010; Li et al., 2014; Schmidt et al., 2016) according to the conditions, suggesting a tight regulation in which the ratios between the concentrations of IscS, IscX, and CyaY are fine-tuned according to the cellular conditions. While we started investigating the possibility of a secondary binding site of IscX on IscS in the attempt to explain a different behavior at high and low molar ratios, it is tempting to speculate that also the secondary binding site could under given conditions have a physiologic significance: even just a partial occupancy could further modulate the regulation of IscS.

Such a system thus represents a sophisticated way to control the enzymatic properties of IscS: a tight regulation is required for maintaining the number of Fe-S clusters formed under tight control at all times, to match the number of apoproteins present. If this balance is lost, there will be too many “highly reactive” Fe-S clusters not being delivered which will be degraded.

From an evolutionary perspective, our results explain the presence of IscX in prokaryotes and not in most eukaryotes: bacterial IscS is a fully active enzyme, whereas the eukaryotic ortholog Nfs1 is inactive in the absence of its activators frataxin and the eukaryotic specific Isd11 and ACP (Foury and Cazzalini, 1997; Calabrese et al., 2005). As demonstrated in a recent paper, this behavior can be explained by a different quaternary assembly of the desulfurase complex with Isd11 mediated by



the acyl carrier protein in this structure (Foury and Cazzalini, 1997; Calabrese et al., 2005). These results indicate a different regulation of Fe-S cluster biogenesis in eukaryotes compared to prokaryotes. We propose the mechanistic basis of the bacterial regulation based on our results. We can also speculate that, as a regulator of the inhibitory function of CyaY in bacteria, IscX could have disappeared in the passage from prokaryotes to eukaryotes where the desulfurase was deactivated making the double repression unnecessary. This hypothesis would fully explain the difference in the function of frataxin between prokaryotes and eukaryotes, and suggests that IscX is the evolutionary “missing link.”

In conclusion, our results place the role of IscX into a new and completely different perspective and provide one of the most elegant examples of a double regulation of an enzyme in which inhibition and counter-inhibition are achieved by exploiting a triple-component system in which two proteins compete for the same binding site on an enzyme depending on the concentration levels of an effector, iron. Similar examples are not rare in nature and have been already described, for instance, in muscles where calcium is able to discriminate between different conformational states of tropomyosin and troponin I to determine different

conformational states of myosin. Through our findings, we can account for the previous literature and suggest a new perspective for the quest to fully elucidate the roles of frataxin and IscX and their functions in the *isc* machine.

AUTHOR CONTRIBUTIONS

SA and RP: carried out the enzymatic work; CI and SM: the BLI measurements; FD, JC, and NL: the mass spectrometry analysis; PK and DS: the SAXS measurements; AP: wrote the paper and coordinated the research.

ACKNOWLEDGMENTS

This work was supported by MRC (grant number U117584256) and BBSRC (grant BB/P006140/1).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2017.00097/full#supplementary-material>

REFERENCES

- Adinolfi, S., Iannuzzi, C., Prisci, F., Pastore, C., Iametti, S., Martin, S., et al. (2009). Bacterial frataxin CyaY is the gatekeeper of iron-sulfur cluster formation catalyzed by IscS. *Nat. Struct. Mol. Biol.* 16, 390–396. doi: 10.1038/nsm.1579
- Agar, J. N., Krebs, C., Frazzon, J., Huynh, B. H., Dean, D. R., and Johnson, M. K. (2000). IscU as a scaffold for iron-sulfur cluster biosynthesis: sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU. *Biochemistry* 39, 7856–7862. doi: 10.1021/bi000931n
- Beinert, H. (2000). Iron-sulfur proteins: ancient structures, still full of surprises. *J. Biol. Inorg. Chem.* 5, 2–15. doi: 10.1007/s007750050002

- Blanchet, C. E., Spilotros, A., Schwemmer, F., Graewert, M. A., Kikhney, A., Svergun, D. I., et al. (2015). Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.* 48(Pt 2), 431–443. doi: 10.1107/S160057671500254X
- Bridwell-Rabb, J., Iannuzzi, C., Pastore, A., and Barondeau, D. P. (2012). Effector role reversal during evolution: the case of frataxin in Fe-S cluster biosynthesis. *Biochemistry* 51, 2506–2514. doi: 10.1021/bi201628j
- Calabrese, V., Lodi, R., Tonon, C., D'Agata, V., Sapienza, M., Scapagnini, G., Butterfield, D. A., et al. (2005). Oxidative stress, mitochondrial dysfunction and cellular stress response in Friedreich's ataxia. *J. Neurol. Sci.* 233, 145–162. doi: 10.1016/j.jns.2005.03.012
- Crack, J. C., Thomson, A. J., and Le Brun, N. E. (2017). Mass spectrometric identification of intermediates in the O₂-driven [4Fe-4S] to [2Fe-2S] cluster conversion in FNR. *Proc. Natl. Acad. Sci. U.S.A.* 114, E3215–E3223. doi: 10.1073/pnas.1620987114
- di Maio, D., Chandramouli, B., Yan, R., Brancato, G., and Pastore, A. (2017). Understanding the role of dynamics in the iron sulfur cluster molecular machine. *Biochim. Biophys. Acta* 1861(1 Pt A), 3154–3163. doi: 10.1016/j.bbagen.2016.07.020
- Flint, D. H. (1996). *Escherichia coli* contains a protein that is homologous in function and N-terminal sequence to the protein encoded by the *nifS* gene of *Azotobacter vinelandii* and that can participate in the synthesis of the Fe-S cluster of dihydroxy-acid dehydratase. *J. Biol. Chem.* 271, 16068–16074.
- Foury, F., and Cazzalini, O. (1997). Deletion of the yeast homologue of the human gene associated with Friedreich's ataxia elicits iron accumulation in mitochondria. *FEBS Lett.* 411, 373–377. doi: 10.1016/S0014-5793(97)00734-5
- Gibson, T. J., Koonin, E. V., Musco, G., Pastore, A., and Bork, P. (1996). Friedreich's ataxia protein: phylogenetic evidence for mitochondrial dysfunction. *Trends Neurosci.* 19, 465–468. doi: 10.1016/S0166-2236(96)20054-2
- Golub, G. H., and Reinsch, C. (1970). Singular value decomposition and least squares solutions. *Numer. Math.* 14, 403–420. doi: 10.1007/BF02163027
- Huynen, M. A., Snel, B., Bork, P., and Gibson, T. J. (2001). The phylogenetic distribution of frataxin indicates a role in iron-sulfur cluster protein assembly. *Hum. Mol. Genet.* 10, 2463–2468. doi: 10.1093/hmg/10.21.2463
- Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F. U., Frishman, D. (2008). Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* 9:102. doi: 10.1186/1471-2164-9-102
- Kim, J. H., Bothe, J. R., Frederick, R. O., Holder, J. C., and Markley, J. L. (2014). Role of IscX in iron-sulfur cluster biogenesis in *Escherichia coli*. *J. Am. Chem. Soc.* 136, 7933–7942. doi: 10.1021/ja501260h
- Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003). PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* 36, 1277–1282. doi: 10.1107/S0021889803012779
- Li, D. S., Ohshima, K., Jiralerspong, S., Bojanowski, M. W., and Pandolfo, M. (1999). Knock-out of the *cyaY* gene in *Escherichia coli* does not affect cellular iron content and sensitivity to oxidants. *FEBS Lett.* 456, 13–16. doi: 10.1016/S0014-5793(99)00896-0
- Li, G. W., Burkhardt, D., Gross, C., and Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624–635. doi: 10.1016/j.cell.2014.02.033
- Musco, G., Stier, G., Kolmerer, B., Adinolfi, S., Martin, S., Frenkiel, T. (2000). Towards a structural understanding of Friedreich's ataxia: the solution structure of frataxin. *Structure* 8, 695–707. doi: 10.1016/S0969-2126(00)00158-1
- Pastore, A., and Puccio, H. (2013). Frataxin: a protein in search for a function. *J. Neurochem.* 126(Suppl. 1), 43–52. doi: 10.1111/jnc.12220
- Pastore, C., Adinolfi, S., Huynen, M. A., Rybin, V., Martin, S., Pastore, A. et al. (2006). YfhJ, a molecular adaptor in iron-sulfur cluster formation or a frataxin-like protein? *Structure* 14, 857–867. doi: 10.1016/j.str.2006.02.010
- Prischi, F., Konarev, P. V., Iannuzzi, C., Pastore, C., Adinolfi, S., Pastore, A., et al. (2010a). Structural bases for the interaction of frataxin with the central components of iron-sulphur cluster assembly. *Nat. Commun.* 1:95. doi: 10.1038/ncomms1097
- Prischi, F., Pastore, C., Carroni, M., Iannuzzi, C., Adinolfi, S., Temussi, P., et al. (2010b). Of the vulnerability of orphan complex proteins: the case study of the *E. coli* IscU and IscS proteins. *Protein Expr. Purif.* 73, 161–166. doi: 10.1016/j.pep.2010.05.003
- Py, B., and Barras, F. (2014). Iron and sulfur in proteins. How does the cell build Fe-S clusters, cofactors essential for life? *Med. Sci.* 30, 1110–1122. doi: 10.1051/medsci/20143012014
- Roche, B., Aussel, L., Ezraty, B., Mandin, P., Py, B., and Barras, F. (2013). Iron/sulfur proteins biogenesis in prokaryotes: formation, regulation and diversity. *Biochim. Biophys. Acta* 1827, 455–469. doi: 10.1016/j.bbmbio.2012.12.010
- Roche, B., Huguenot, A., Barras, F., and Py, B. (2015). The iron-binding CyaY and IscX proteins assist the ISC-catalyzed Fe-S biogenesis in *Escherichia coli*. *Mol. Microbiol.* 95, 605–623. doi: 10.1111/mmi.12888
- Rouault, T. A. (2012). Biogenesis of iron-sulfur clusters in mammalian cells: new insights and relevance to human disease. *Dis. Model. Mech.* 5, 155–164. doi: 10.1242/dmm.009019
- Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrne, E., Volkmer, B., Callipo, L., et al. (2016). The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* 34, 104–110. doi: 10.1038/nbt.3418
- Shi, R., Proteau, A., Villarroja, M., Moukadiri, I., Zhang, L., Trempe, J., et al. (2010). Structural basis for Fe-S cluster assembly and tRNA thiolation mediated by IscS protein-protein interactions. *PLoS Biol.* 8:e1000354. doi: 10.1371/journal.pbio.1000354
- Shimomura, Y., Takahashi, Y., Kakuta, Y., and Fukuyama, K. (2005). Crystal structure of *Escherichia coli* YfhJ protein, a member of the ISC machinery involved in assembly of iron-sulfur clusters. *Proteins* 60, 566–569. doi: 10.1002/prot.20481
- Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Xie, X. S. et al. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329, 533–538. doi: 10.1126/science.1188308
- Tokumoto, U., and Takahashi, Y. (2001). Genetic analysis of the *isc* operon in *Escherichia coli* involved in the biogenesis of cellular iron-sulfur proteins. *J. Biochem.* 130, 63–71. doi: 10.1093/oxfordjournals.jbchem.a002963
- Tokumoto, U., Nomura, S., Minami, Y., Mihara, H., Kato, S., Kurihara, T., et al. (2002). Network of protein-protein interactions among iron-sulfur cluster assembly proteins in *Escherichia coli*. *J. Biochem.* 131, 713–719. doi: 10.1093/oxfordjournals.jbchem.a003156
- Urbina, H. D., Silberg, J. J., Hoff, K. G., and Vickery, L. E. (2001). Transfer of sulfur from IscS to IscU during Fe/S cluster assembly. *J. Biol. Chem.* 276, 44521–44526. doi: 10.1074/jbc.M106907200
- Watson, A. A., Mahajan, P., Mertens, H. D., Deery, M. J., Zhang, W., Laue, E. D., et al. (2012). The PHD and chromo domains regulate the ATPase activity of the human chromatin remodeler CHD4. *J. Mol. Biol.* 422, 3–17. doi: 10.1016/j.jmb.2012.04.031
- Zheng, L., Cash, V., L., Flint, D. H., and Dean, D. R. (1998). Assembly of iron-sulfur clusters. Identification of an *iscSUA-hscBA-fdx* gene cluster from *Azotobacter vinelandii*. *J. Biol. Chem.* 273, 13264–13272. doi: 10.1074/jbc.273.21.13264

Conflict of Interest Statement: 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.

The reviewer AD and handling Editor declared their shared affiliation.

Copyright © 2018 Adinolfi, Puglisi, Crack, Iannuzzi, Dal Piaz, Konarev, Svergun, Martin, Le Brun and Pastore. 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 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.



Intramolecular Fuzzy Interactions Involving Intrinsically Disordered Domains

Miguel Arbesú*, Guillermo Iruela, Héctor Fuentes, João M. C. Teixeira and Miquel Pons*

BioNMR Laboratory, Inorganic and Organic Chemistry Department, University of Barcelona, Barcelona, Spain

Structural disorder is an essential ingredient for function in many proteins and protein complexes. Fuzzy complexes describe the many instances where disorder is maintained as a critical element of protein interactions. In this minireview we discuss how intramolecular fuzzy interactions function in signaling complexes. Focussing on the Src family of kinases, we argue that the intrinsically disordered domains that are unique for each of the family members and display a clear fingerprint of long range interactions in Src, might have critical roles as functional sensor or effectors and mediate allosteric communication via fuzzy interactions.

OPEN ACCESS

Edited by:

Irene Díaz-Moreno,
Universidad de Sevilla, Spain

Reviewed by:

Antonio J. Díaz Quintana,
Universidad de Sevilla, Spain
Monika Fuxreiter,
University of Debrecen, Hungary

*Correspondence:

Miguel Arbesú
miguelarbesu@gmail.com
Miquel Pons
mpons@ub.edu

Specialty section:

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

Received: 02 February 2018

Accepted: 03 April 2018

Published: 30 April 2018

Citation:

Arbesú M, Iruela G, Fuentes H,
Teixeira JMC and Pons M (2018)
Intramolecular Fuzzy Interactions
Involving Intrinsically Disordered
Domains. *Front. Mol. Biosci.* 5:39.
doi: 10.3389/fmolb.2018.00039

Keywords: fuzzy complexes, intrinsically disordered proteins, Src family kinases, allostery, fuzzy domains, paramagnetic relaxation enhancement

A large majority of proteins are built from domains, classically defined as functional autonomous folding units. In a typical divide-and-conquer approach, the structure-function analysis proceeds through the characterization of the individual domains followed by the study of their mutual interactions. This approach makes a clear distinction between the “functional” domains and the linkers separating them.

The same strategy is taken in the analysis of the also very abundant multiprotein complexes, in which the individual proteins are considered the building blocks (equivalent to domains of multidomain proteins). Key components of multiprotein complexes are scaffolding proteins, which would play the role of linkers in multidomain proteins.

The current view of protein-protein interactions is quite dynamic and intrinsically disordered regions (IDR) are increasingly recognized as key players.

In this minireview we shall summarize some important aspects of (intermolecular) protein binding by disordered proteins and extend them to the case of interdomain (i.e., intramolecular) binding using the c-Src family of kinases as an example.

PROTEIN INTERACTIONS BY INTRINSICALLY DISORDERED PROTEINS

Intrinsically disordered proteins (IDP) or proteins with long IDR form a significant portion of the proteome of eukaryotes and are specially prevalent in signaling and regulation complexes (Iakouchcheva et al., 2002).

Protein complexes involving IDRs span a wide range of affinities and lifetimes as well as specificities (Tompá et al., 2015). A recent analysis of K_d value statistics in the curated DIBS database of IDR-folded protein complexes (Schad et al., 2017) confirms a wide range of affinities spanning from the subnanomolar to the millimolar regimes.

In their review on experimental thermodynamic data from binary protein complexes involving IDPs or ordered proteins, Teilum et al. (2015) found that the ΔG° values of their data sets involving IDPs were on average only 2.5 kcal mol⁻¹ less stable than the values from complexes between ordered proteins. In the two sets, isothermal enthalpy-entropy compensation was observed, a general phenomenon in biomolecular recognition processes (Chodera and Mobley, 2013). Thus, favorable binding enthalpy is associated to a loss of entropy due to massive reduction of structural freedom degrees.

The interaction surfaces involving folded proteins or IDRs showed similar amino acid composition and size and the distribution of ΔH° values were statistically equal; so, all the destabilizing contribution had an entropic origin ($-T\Delta S^\circ > 0$).

An entropic cost for binding an IDR is intuitively expected, however, the surprising result is its relatively small value suggesting compensatory mechanisms are an important component of IDR interactions. The importance of entropy compensation is highlighted by the similarity in the distribution of ΔH° values, a very interesting result in itself that emphasizes the underlying short-range similarities between protein-protein interactions involving folded and disordered proteins.

Flock et al. (2014) have reviewed the importance of entropy control to tune IDR function. The importance of entropy in the formation of complexes endows IDR-involving complexes with their unique functional characteristics as molecular rheostats and signal integrators, able to respond in a precise, continuous and dynamic way to varying combinations of inputs with specific outputs.

The formation of a stable complex between two proteins (i.e., with a negative $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$) can be achieved by optimizing the enthalpy gain ($\Delta H^\circ < 0$), increasing the entropic gain ($-T\Delta S^\circ < 0 \Rightarrow \Delta S^\circ > 0$) or minimizing the entropic loss ($-T\Delta S^\circ \approx 0$). A fundamental aspect of the interplay between enthalpy and entropy components is their “locality.”

Enthalpy effects usually reflect local short-range interactions and may be considered additive and proportional to the contact surface. Thus, large enthalpy components are usually associated to large contact surfaces, although these interfaces do not have to be necessarily continuous. Electrostatic contributions to the enthalpy, however, are long range. They often drive the partners together (thus reducing the translational and possibly rotational entropy of the system) and, in the formed complex, enable dynamic interactions that minimize the entropy loss upon complex formation. A recent example of a picomolar interaction between two IDPs without a significant loss of flexibility is driven by electrostatics (Borgia et al., 2018).

INTERACTING ELEMENTS AND MULTIVALENCY

Entropy contains “local” components associated to the degree of structure achieved by the contact regions of the two interacting partners, as well as more global contributions

of which we may distinguish (i) the effect of regions that can remain highly flexible in the complex (thus not contributing an entropic penalty to binding), (ii) the possible preexistence of long range intramolecular contacts restricting the conformational freedom in the free IDR (therefore minimizing the loss of entropy upon complex formation), and (iii) the configurational entropy arising from multiple alternative binding poses (“microstates”) contributing to the bound state.

The first two situations reduce the entropic cost of binding through IDRs and correspond to the strategies of not to pay (i) or pre-pay (ii). The third situation actually contributes an entropic gain.

In the interacting regions, pre-pay strategies may take the form of preformed structural elements retained in the complex (Davey et al., 2012; Pancsa and Fuxreiter, 2012) or bound solvent molecules that are retained in the complex in water-mediated interactions (London et al., 2010).

The dominant role of entropy in protein interactions is not restricted to IDPs. The changes in internal dynamics of the catabolite activator protein (CAP), measured by NMR in the entire protein, explain the dramatic changes in affinity observed in CAP variants that form complexes with identical interfaces (Tzeng and Kalodimos, 2012).

In IDRs the change in entropy upon binding is determined by the interplay between local and global effects. Short linear motifs (LM) play an important role in IDR interfaces. Although the definition of LM is based on bioinformatic studies, they can be interpreted using structural and dynamic concepts. LM are often formed by hydrophobic residues grafted onto a malleable template (Fuxreiter et al., 2007). The expected lower enthalpy of the interaction by short elements, as compared to the large rigid interfaces between ordered proteins, can be partially compensated by the fact that IDPs often adopt extended conformations permitting short motifs to establish a variety of interactions through virtually any element of their backbone or side chains, thus their interacting interfaces have a larger effective area per residue than those of ordered proteins (Gunasekaran et al., 2003). In addition, these short stretches are modular recognition elements that can be combined to form multivalent complexes. Thus, a favorable binding enthalpy, comparable to that found associated to a large, rigid interface, can be achieved by weaker but multiple sparse anchoring elements (Cumberworth et al., 2013). The participating groups may be difficult to identify either experimentally, because interactions are weak, or statistically, because they may appear in a variety of combinations that are not repeated “motifs” (Van Roey et al., 2014).

The number and intrinsic properties of individual interacting regions, as well as the size and dynamic properties of the spacers between them collectively, and therefore non-linearly, determine the binding properties of IDRs.

The effects of multivalent binding have been recognized for a long time, beyond the field of IDPs. The strength by which a multivalent antibiotic binds to its antigen, termed avidity (Crothers and Metzger, 1972), can be explained by the fact that when one of the sites is bound to its cognate site receptor, a second site located close-by binds

cooperatively, basically because of the lower entropic cost of a (pseudo)-intramolecular interaction (Kitov and Bundle, 2003). If the linker connecting the two sites is flexible, the average distance between the sites is the main factor determining the cooperativity. If the flexibility is limited, the linker may also modulate the relative orientation of the components of the second interacting site. A consequence of the model is that avidity may be modulated either by modifying the interacting sites or the flexibility of the linker (Cerofolini et al., 2013).

The avidity model assumes multiple binding sites but the interacting partners for each site are not interchangeable. The scenario in which the multiple interaction sites in one of the molecules can interact with, and therefore compete for, the same site of the second molecule is referred to as **allovalency** (Klein et al., 2003). Allovalency predicts a dependency on the number of interacting sites, not through simultaneous cooperative binding, because all of them compete for a single site in the second molecule, but through local concentration and rebinding. In the defining example, the binding of Sic1 to Cdc4, the number of interacting sites is actively modulated by the random phosphorylation of up to ten serine and threonine residues, showing a sharp increase in the fraction of bound form after six of them are phosphorylated (Mittag et al., 2010).

FUZZY COMPLEXES AND MULTIVALENCY

Fuzzy complexes, introduced by Tompa and Fuxreiter (2008) describe binding situations in which at least one of the elements in the complex remains dynamic. Therefore, the complex cannot be properly described by a defined structure but has the characteristics of an heterogeneous ensemble. Importantly, the interaction heterogeneity of the fuzzy complex is an essential component of the functional outcome of complex formation. The functional character of the retained disorder, thus, differentiates a fuzzy complex from a complex including a random region in non-specific contact with the partner. An expanded repertoire of examples can be found in recent reviews (Fuxreiter, 2012; Fuxreiter and Tompa, 2012; Sharma et al., 2015; Miskei et al., 2017).

Structural disorder in fuzzy complexes represents a continuum, from rather rigid polymorphic complexes displaying static disorder with only a few alternative conformations to highly dynamic random complexes. The proportion between regions directly involved in short range contacts and connector regions decreases in this series. Individual regions contributing to $\Delta H^\circ < 0$ become smaller but may increase in number, thus a favorable enthalpy contribution can be retained. Splitting the interaction interface in many smaller areas, each binding weakly and with high promiscuity, enhances binding degeneracy that contributes an additional entropic term, $\Delta S_{\text{configurational}}^\circ$, which reflects the contribution stemming from the different forms in which the IDR and its partner can associate.

A recent experimental example is the detailed study of the thermodynamics of the fuzzy complex between the C-terminal IDR of antitoxin CcdA, which adopts α helical structure at the

time of binding the toxin dimer CcdB (Hadži et al., 2017). The authors perform a series of mutations that affect contacting and non-contacting residues. Their results show that mutations in residues not directly involved in protein:protein interaction reduce the degree of structuration both in the bound and free forms ($\Delta\Delta S_{\text{conformational}}^\circ \approx 0$), but also promote alternative isoenergetic configurations ($\Delta\Delta S_{\text{configurational}}^\circ > 0$ and $\Delta\Delta H^\circ \approx 0$) thus minimizing the particular ΔG° of the mutant complex.

NMR FINGERPRINT OF LONG-RANGE ORGANIZATION OF IDR

Fuzziness is not associated to promiscuous binding. The selectivity is encoded in the dynamic, non random, organization of distant potentially interacting regions.

Operationally, a very efficient method to map a set of long range interactions is by measuring the paramagnetic relaxation enhancement (PRE) along the sequence induced by one or several paramagnetic tags (Clare and Iwahara, 2009). Since paramagnetic effects are sensitive to transient interactions and efficient over considerable distances, in the case of disordered proteins the key aspect is to differentiate specific from random coil effects. In this respect, the Konrat's group have introduced the concept of paramagnetic relaxation interference (PRI) by comparing the simultaneous effect of two paramagnetic centers (Kurzbach et al., 2016) with the sum of the individual effects. A difference between these values requires that the two sites move in a correlated fashion. An alternative, often simpler, approach is to compare the observed PREs with the predictions of a random coil model. The Δ PRE analysis (Arbesú et al., 2017) clearly identifies the relevant transient contacts in IDRs.

Δ PRE analysis of several paramagnetically tagged forms of the intrinsically disordered N-terminus of c-Src in the presence and in the absence of the folded neighboring SH3 domain show a very similar profile, confirming the presence of a conserved set of non-random long-range interactions and validating the use of the term **domain** for this intrinsically disordered region (Arbesú et al., 2017). Substitution of residues important for IDR pre-organization and interdomain contacts showed that the Δ PRE profiles are generally conserved upon different perturbations—e.g., same profile trends, location of maxima and minima, etc.—but can also reflect the functional loss of interactions—i.e., consistent contact reduction upon substitution. Δ PRE profiling thus provides a structural signature that captures non random ensemble conformational preferences and their associated dynamics. This simple method enables facile comparison for carrying out functional analysis based on mutations.

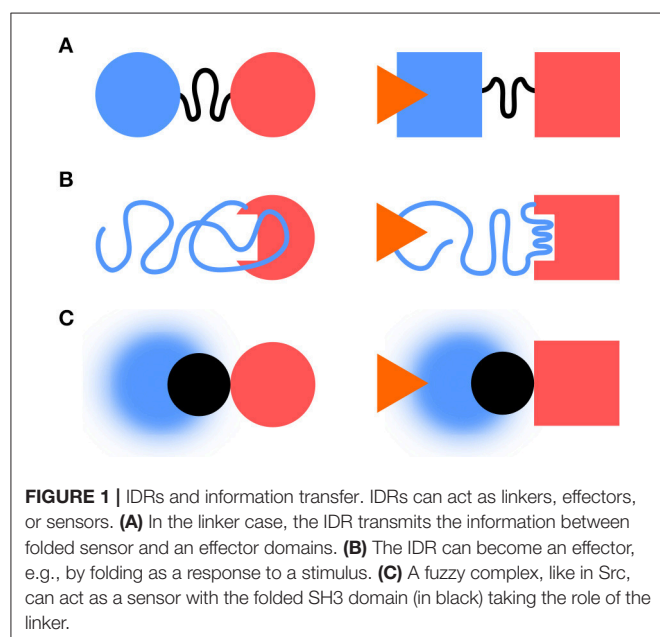
In an analysis of the Pfam database, which identifies domains based on multiple sequence alignments, Tompa et al. (2009) found that a substantial number of the sequence defined domains contained disordered regions and confirmed that disordered domains are inheritable, evolvable, and functional units. Some domains, such as the Unique domain, which is the most discriminating feature of the distinct Src family kinases (SFK),

is not identified as a domain using multisequence alignment methods. This is not surprising since sequence variability is its defining characteristic. We argue that the requirement of “autonomous folding,” which would identify a domain without using sequence conservation data, could be replaced in the field of IDPs by that of a conserved, non-random, set of long range interactions.

The PRE and chemical shift perturbation analysis of wild-type Src as well as a number of mutated or truncated variants, also showed the interaction between the intrinsically disordered domain and specific regions of the SH3 domain. Interestingly, the most affected regions in the folded scaffold could be mapped to the loops that decorate its surface. The interaction between the disordered N-terminal region of c-Src and the SH3 domain has the characteristics associated to fuzzy complexes: (i) the disordered region remains highly dynamic, as seen by NMR, (ii) its overall dimension is affected by the presence of the SH3 domain, as seen by Small Angle X-ray Scattering, (iii) the local perturbations sensed by chemical shifts are affected by modifications in distant, well defined parts of the protein, and (iv) mutations in the disordered region cause strong functional effects in the entire protein.

INTRAMOLECULAR FUZZY COMPLEXES AS SIGNAL SENSORS

Classical descriptions of multidomain signaling proteins distinguish between regulatory/sensor and catalytical/effector domains. IDR can act as linkers, effectors or sensors (**Figure 1**). IDR-mediated signaling enables complex regulatory behavior, including multiple signal integration and rheostat-like graded responses (Tomba, 2014).



IDRs can act as linkers through which information is propagated to distant regions. This can occur without concomitant structuration through remodeling of the protein free energy landscape affecting the conformer populations and causing specific functional outputs (Tsai et al., 1999; Hilser and Thompson, 2007; Ma et al., 2011; Montlagh et al., 2014). Examples include the DNA binding Ets-1 transcription factor (Pufall, 2005), the Sic1 cell cycle protein (Mittag et al., 2008), or the *Drosophila* Ultrabithorax transcription factor (Liu et al., 2008).

IDR conformational ensembles can be modified by “external” signals and modulated by “internal” parameters, such as post-translational modifications. Thus, they can act as sensitive **sensors** with tunable selectivity and sensitivity. In a recent work, the effect of a small drug interacting with a disordered region of p27 was shown to cause a shift in its conformational landscape (Ban et al., 2017) stressing the capacity of IDRs as sensor of their environment (and not trivially, as drug targets).

Borrowing concepts from information theory (Shannon and Weaver, 1949), the capacity to transfer information is determined by the signaling event rate, and the size of the set formed by possible events it permits. Fuzzy complexes provide fast interconversion dynamics and a large set of configurations in the interface. Thus, fuzzy interfaces have the ability to act as high-capacity channels.

FUZZY INTERACTIONS IN SRC FAMILY KINASES

The Src N-terminal regulatory element (SNRE) studied by our group suggests an additional class of IDR allostery, in which the disordered region acts a sensor but the connecting element is a folded SH3 domain.

Mutations in the Unique domain of c-Src induce strong phenotypes in Src-dependent colorectal cancer cells (Arbesú et al., 2017). The Unique domain participates in a number of interactions with proteins such as calmodulin (Pérez et al., 2013) or N-methylaspartate receptor (Gingrich et al., 2004), lipids (Pérez et al., 2013) and is subjected to phosphorylation (Amata et al., 2014) and proteolytic processing (Hossain et al., 2013). In order to integrate these capabilities into a functional sensor-activator pair, the nature of the connector becomes a key issue. The SH3 domain has been shown to act as a scaffold of a fuzzy intramolecular complex (Maffei et al., 2015; Arbesú et al., 2017). These findings suggests that the SH3 domain may have a dual role in c-Src regulation: the traditionally recognized one, as a sensor (docking site) of polyproline peptide motifs, as well as that of a connector, relaying the information sensed by the preceding IDR (**Figure 2**). An enhanced capacity of SH3 motifs to interact with intrinsically disordered regions has been suggested (Beltrao and Serrano, 2005). Recently, the N-terminal IDR of Abl kinase has been shown to modulate its activity through the SH3 domain (Saleh et al., 2017).

Recent NMR data (Tong et al., 2017) as well as SAXS studies (Bernadó et al., 2008) confirm that the interaction

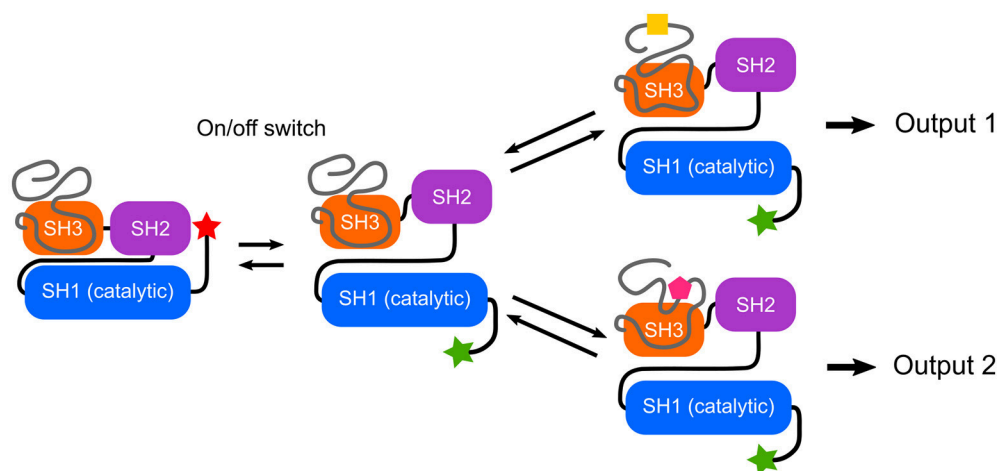


FIGURE 2 | Switches and rheostats in Src Family Kinases. The conserved SH3-SH2-SH1 multidomain cassette implements an on/off switch based on the interaction between the phosphorylated C-terminal tail and the SH2 domain, as well as interactions involving the SH3 domain. The intrinsically disordered N-terminal region forms a fuzzy complex with the SH3 domain and enables a program that could direct different qualitative and quantitative outputs depending on the environment. Thus, the IDRs can function as selective rheostats. Within the Src Family of Kinases, the high homology among of the SH3-SH2-SH1 cassette contrasts with the uniqueness of the disordered N-terminal regions, suggesting a modular architecture in which specific responses are programmed in the IDR.

between the SH3-SH2 regulatory domains and the kinase (SH1) domain is conserved in the active form of cSrc, i.e., in the absence of the autoinhibitory interaction between pTyr527 and the SH2 domain. The NMR results show that this interaction is dynamic and suggests that modulation of the interdomain dynamics may contribute to modulate c-Src activity.

In spite of their large sequence divergence, the IDR regions of the various SFK show coevolution with their respective SH3 domains, suggesting that a fuzzy interaction such as the one found in c-Src may be a functional element in all SFKs (Arbesú et al., 2017). The large sequence variations in the Unique domains contrasts with the very high homology displayed by the SH3-SH2-SH1 cassette, suggesting that the Unique domain has evolved to read the distinct environments required by each SFK.

The Δ PRE method is a simple and robust analytical approach to generate a fingerprint of the long-range interactions within intrinsically disordered domains. The complete processing tools are part of the Farseer software (Teixeira et al., in press).

Simulations based on fuzzy logic recapitulate many features of a kinase network (Aldridge et al., 2009). Proteins like the SFKs can be considered as algorithms reading complex signaling inputs to generate the proper responses. Thus, fuzzy interactions by IDRs may, in fact, be implementing fuzzy logic at the level of individual proteins.

AUTHOR CONTRIBUTIONS

This is a mini-review article based on the Ph.D. thesis of MA, supervised by MP and with contributions from GI, HF, and JT working in fuzzy complexes of Src Family Kinases.

ACKNOWLEDGMENTS

We acknowledge the financial support from the Spanish MINECO (BIO2016-78006R) co-financed with EU structural funds, and the Fundació Marató de TV3 (2013-2830/31) and insightful discussions with M. Fuxreiter, O. Millet, and R. Crehuet during the defence of M.A. thesis.

REFERENCES

- Aldridge, B. B., Saez-Rodriguez, J., Muhlich, J. L., Sorger, P. K., and Lauffenburger, D. A. (2009). Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling. *PLoS Comput. Biol.* 5:e1000340. doi: 10.1371/journal.pcbi.1000340
- Amata, I., Maffei, M., and Pons, M. (2014). Phosphorylation of unique domains of Src family kinases. *Front. Genet.* 5:181. doi: 10.3389/fgene.2014.00181
- Arbesú, M., Maffei, M., Cordeiro, T. N., Teixeira, J. M. C., Pérez, Y., Bernadó, P., et al. (2017). The unique domain forms a fuzzy intramolecular complex in Src family kinases. *Structure* 25, 630.e4–640.e4. doi: 10.1016/j.str.2017.02.011
- Ban, D., Iconaru, L. I., Ramanathan, A., Zuo, J., and Kriwacki, R. W. (2017). A small molecule causes a population shift in the conformational landscape of an intrinsically disordered protein. *J. Am. Chem. Soc.* 139, 13692–13700. doi: 10.1021/jacs.7b01380
- Beltrao, P., and Serrano, L. (2005). Comparative genomics and disorder prediction identify biologically relevant SH3 protein interactions. *PLoS Comput. Biol.* 1:e26. doi: 10.1371/journal.pcbi.0100026
- Bernadó, P., Pérez, Y., Svergun, D. I., and Pons, M. (2008). Structural characterization of the active and inactive states of src kinase in solution by small-angle X-ray scattering. *J. Mol. Biol.* 376, 492–505. doi: 10.1016/j.jmb.2007.11.066
- Borgia, A., Borgia, M. B., Bugge, K., Kissling, V. M., Heidarrsson, P. O., Fernandes, C. B., et al. (2018). Extreme disorder in an ultrahigh-affinity protein complex. *Nature* 555, 61–66. doi: 10.1038/nature25762

- Cerofolini, L., Fields, G. B., Fragai, M., Gerales, C. F., Luchinat, C., Parigi, G., et al. (2013). Examination of matrix metalloproteinase-1 in solution. *J. Biol. Chem.* 288, 30659–30671. doi: 10.1074/jbc.M113.477240
- 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
- Clore, G. M., and Iwahara, J. (2009). Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.* 109, 4108–4139. doi: 10.1021/cr900033p
- Crothers, D. M., and Metzger, H. (1972). The influence of polyvalency on the binding properties of antibodies. *Immunochemistry* 9, 341–357. doi: 10.1016/0019-2791(72)90097-3
- Cumberworth, A., Lamour, G., Babu, M. M., and Gsponer, J. (2013). Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochem. J.* 454, 361–369. doi: 10.1042/BJ20130545
- Davey, N. E., Van Roey, K., Weatheritt, R. J., Toedt, G., Uyar, B., Altenberg, B., et al. (2012). Attributes of short linear motifs. *Mol. Biosyst.* 8, 268–281. doi: 10.1039/c1mb05231d
- Flock, T., Weatheritt, R. J., Latysheva, N. S., and Madan Babu, M. M. (2014). Controlling entropy to tune the functions of intrinsically disordered regions. *Curr. Opin. Struct. Biol.* 26, 62–72. doi: 10.1016/j.sbi.2014.05.007
- Fuxreiter, M. (2012). Fuzziness: linking regulation to protein dynamics. *Mol. Biosyst.* 8, 168–177. doi: 10.1039/C1MB05234A
- Fuxreiter, M., and Tompa, P. (eds.). (2012). *Fuzziness*. New York, NY: Springer US.
- Fuxreiter, M., Tompa, P., and Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* 23, 950–956. doi: 10.1093/bioinformatics/btm035
- Gingrich, J. R., Pelkey, K. A., Fam, S. R., Huang, Y., Petralia, R. S., Wenthold, R. J., et al. (2004). Unique domain anchoring of Src to synaptic NMDA receptors via the mitochondrial protein NADH dehydrogenase subunit 2. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6237–6242. doi: 10.1073/pnas.0401413101
- Gunasekaran, K., Tsai, C.-J., Kumar, S., Zanuy, D., and Nussinov, R. (2003). Extended disordered proteins: targeting function with less scaffold. *Trends Biochem. Sci.* 28, 81–85. doi: 10.1016/S0968-0004(03)00003-3
- Hadži, S., Mernik, A., Podlipnik, C., Loris, R., and Lah, J. (2017). *The Thermodynamic Basis of the Fuzzy Interaction of an Intrinsically Disordered Protein*. *Angew. Chemie Int. Ed.* 1–5.
- Hilser, V. J., and Thompson, E. B. (2007). Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8311–8315. doi: 10.1073/pnas.0700329104
- Hossain, M. I., Roulston, C. L., Kamaruddin, M. A., Chu, P. W., Ng, D. C., Dusting, G. J., et al. (2013). A truncated fragment of Src protein kinase generated by calpain-mediated cleavage is a mediator of neuronal death in excitotoxicity. *J. Biol. Chem.* 288, 9696–9709. doi: 10.1074/jbc.M112.419713
- Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z., and Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* 323, 573–584. doi: 10.1016/S0022-2836(02)00969-5
- Kitov, P. I., and Bundle, D. R. (2003). On the nature of the multivalency effect: a thermodynamic model. *J. Am. Chem. Soc.* 125, 16271–16284. doi: 10.1021/ja038223n
- Klein, P., Pawson, T., and Tyers, M. (2003). Mathematical modeling suggests cooperative interactions between a disordered polyvalent ligand and a single receptor site. *Curr. Biol.* 13, 1669–1678. doi: 10.1016/j.cub.2003.09.027
- Kurzbach, D., Vanas, A., Flamm, A. G., Tarnoczi, N., Kontaxis, G., Maltar-Stremečki, N., et al. (2016). Detection of correlated conformational fluctuations in intrinsically disordered proteins through paramagnetic relaxation interference. *Phys. Chem. Chem. Phys.* 18, 5753–5758. doi: 10.1039/C5CP04858C
- Liu, Y., Matthews, K. S., and Bondos, S. E. (2008). Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the drosophila hox protein ultrabithorax. *J. Biol. Chem.* 283, 20874–20887. doi: 10.1074/jbc.M80375200
- London, N., Movshovitz-Attias, D., and Schueler-Furman, O. (2010). The structural basis of peptide-protein binding strategies. *Structure* 18, 188–199. doi: 10.1016/j.str.2009.11.012
- Ma, B., Tsai, C.-J., Hailoglu, T., and Nussinov, R. (2011). Dynamic allostery: linkers are not merely flexible. *Structure* 19, 907–917. doi: 10.1016/j.str.2011.06.002
- Maffei, M., Arbesú, M., Le Roux, A.-L., Amata, I., Roche, S., and Pons, M. (2015). The SH3 domain acts as a scaffold for the N-terminal intrinsically disordered regions of c-Src. *Structure* 23, 893–902. doi: 10.1016/j.str.2015.03.009
- Miskei, M., Antal, C., and Fuxreiter, M. (2017). FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. *Nucleic Acids Res.* 45, D228–D235. doi: 10.1093/nar/gkw1019
- Mittag, T., Marsh, J., Grishaev, A., Orlicky, S., Lin, H., Sicheri, F., et al. (2010). Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* 18, 494–506. doi: 10.1016/j.str.2010.01.020
- Mittag, T., Orlicky, S., Choy, W.-Y., Tang, X., Lin, H., Sicheri, F., et al. (2008). Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17772–17777. doi: 10.1073/pnas.0809222105
- Montlagh, H. N., Wrabl, J. O., and Hilser, V. J. (2014). The ensemble nature of allostery. *Nature* 508, 331–339. doi: 10.1038/nature13001
- Pancsa, R., and Fuxreiter, M. (2012). Interactions vis intrinsically disordered regions: what kind of motifs? *Life* 64, 513–520. doi: 10.1002/iub.1034
- Pérez, Y., Maffei, M., Igea, A., Amata, I., Gairí, M., Nebreda, A. R., et al. (2013). Lipid binding by the unique and SH3 domains of c-Src suggests a new regulatory mechanism. *Sci. Rep.* 3:1295. doi: 10.1038/srep01295
- Pufall, M. A. (2005). Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* 309, 142–145. doi: 10.1126/science.111915
- Saleh, T., Rossi, P., and Kalodimos, C. G. (2017). Atomic view of the energy landscape in the allosteric regulation of Abl kinase. *Nature Struct. Mol. Biol.* 24, 893–901. doi: 10.1038/nsmb.3470
- Schad, E., Fichó, E., Pancsa, R., Simon, I., Dosztányi, Z., and Mészáros, B. (2017). DIBS: a repository of disordered binding sites mediating interactions with ordered proteins. *Bioinformatics* 34, 535–537. doi: 10.1093/bioinformatics/btx640
- Shannon, C. E., and Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, IL: Univ of Illinois Press.
- Sharma, R., Raduly, Z., Miskei, M., and Fuxreiter, M. (2015). Fuzzy complexes: specific binding without complete folding. *FEBS Lett.* 589, 2533–2542. doi: 10.1016/j.febslet.2015.07.022
- Teilmann, K., Olsen, J. G., and Kragelund, B. B. (2015). Globular and disordered—the non-identical twins in protein-protein interactions. *Front. Mol. Biosci.* 2:40. doi: 10.3389/fmolb.2015.00040
- Teixeira, J. M. C., Skinner, S., Arbesú, M., Breeze, A. L., and Pons, M. (in press). Farseer-NMR: automatic treatment, analysis and plotting of large, multi-variable NMR data. *J. Biomol. NMR*.
- Tompa, P. (2014). Multisteric regulation by structural disorder in modular signaling proteins: an extension of the concept of allostery. *Chem. Rev.* 114, 6715–6732. doi: 10.1021/cr4005082
- Tompa, P., and Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* 33, 2–8. doi: 10.1016/j.tibs.2007.10.003
- Tompa, P., Fuxreiter, M., Oldfield, C. J., Simon, I., Dunker, K., and Uversky, V. N. (2009). Close encounters of the third kind: disordered domains and the interaction of proteins. *Bioessays* 31, 328–335. doi: 10.1002/bies.200800151
- Tompa, P., Schad, E., Santos, A., and Kalmar, L. (2015). Intrinsically disordered proteins: emerging interaction specialists. *Curr. Opin. Struct. Biol.* 35, 49–59. doi: 10.1016/j.sbi.2015.08.009

- Tong, M., Pelton, J. G., Gill, M. L., Zhang, W., Picart, F., and Seeliger, M. A. (2017). Survey of solution dynamics in Src kinase reveals allosteric cross talk between the ligand binding and regulatory sites. *Nat. Commun.* 8, 2160. doi: 10.1038/s41467-017-02240-6
- Tsai, C. J., Kumar, S., Ma, B., and Mussinov, R. (1999). Folding funnels, binding funnels, and protein function. *Protein Sci.* 8, 1181–1190. doi: 10.1110/ps.8.6.1181
- Tzeng, S.-R., and Kalodimos, C. G. (2012). Protein activity regulation by conformational entropy. *Nature* 488, 236–240. doi: 10.1038/nature11271
- Van Roey, K., Uyar, B., Weatheritt, R. J., Dinkel, H., Seiler, M., Budd, A., et al. (2014). Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* 114, 6733–6778. doi: 10.1021/cr400585q

Conflict of Interest Statement: 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.

The reviewer AJDQ and handling Editor declared their shared affiliation.

Copyright © 2018 Arbesú, Iruela, Fuentes, Teixeira and Pons. 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 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: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

[@frontiersin](https://twitter.com/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