

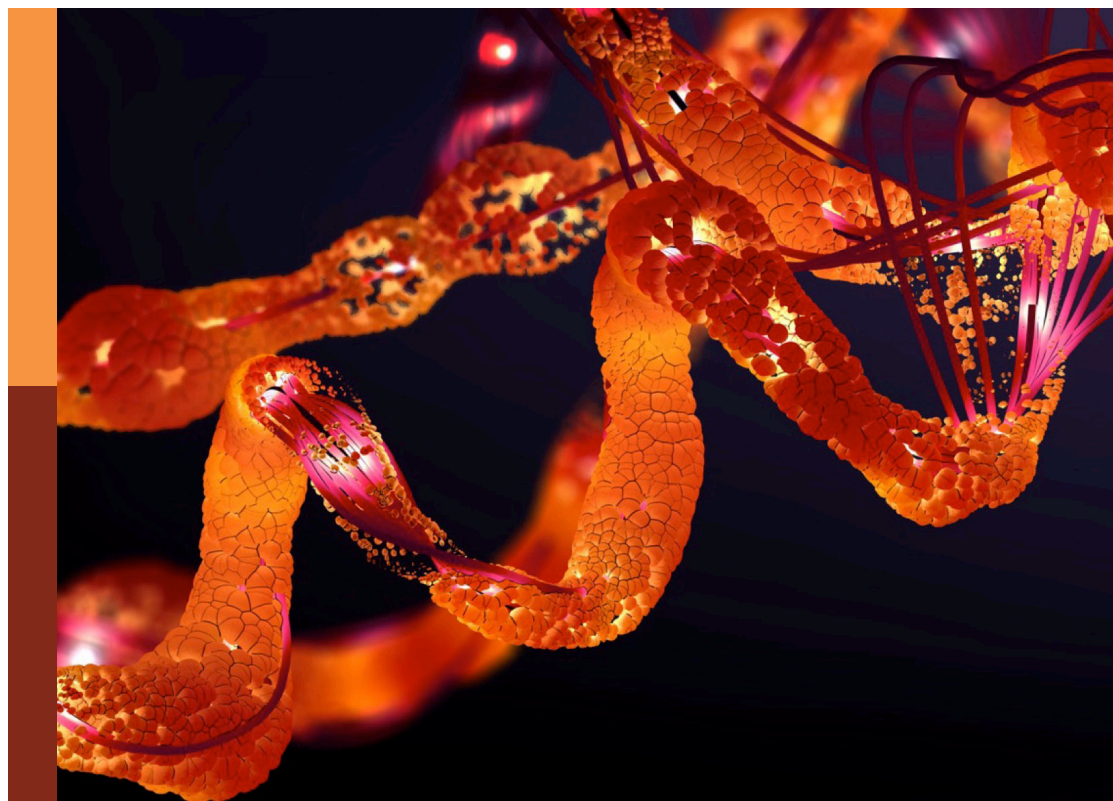
# Integration of structural biology data in lead drug discovery and optimization

**Edited by**

Marco Nardini, Pietro Roversi and Gianluca Molla

**Published in**

Frontiers in Molecular Biosciences



## FRONTIERS EBOOK COPYRIGHT STATEMENT

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

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

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

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

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

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

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

ISSN 1664-8714  
ISBN 978-2-83251-682-9  
DOI 10.3389/978-2-83251-682-9

## About Frontiers

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

## Frontiers journal series

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

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: [frontiersin.org/about/contact](https://frontiersin.org/about/contact)

# Integration of structural biology data in lead drug discovery and optimization

## Topic editors

Marco Nardini — University of Milan, Italy

Pietro Roversi — Institute of Agricultural Biology and Biotechnology, National Research Council (CNR), Italy

Gianluca Molla — University of Insubria, Italy

## Citation

Nardini, M., Roversi, P., Molla, G., eds. (2023). *Integration of structural biology data in lead drug discovery and optimization*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-83251-682-9

## Table of contents

- 05 Editorial: Integration of structural biology data in lead drug discovery and optimization  
Pietro Roversi, Gianluca Molla and Marco Nardini
- 07 Fragment-Based Drug Discovery by NMR. Where Are the Successes and Where can It Be Improved?  
Luca G. Mureddu and Geerten W. Vuister
- 21 Using Structure-guided Fragment-Based Drug Discovery to Target *Pseudomonas aeruginosa* Infections in Cystic Fibrosis  
Sheikh Mohammed Arif, R. Andres Floto and Tom L. Blundell
- 44 Recent Approaches to the Identification of Novel Microtubule-Targeting Agents  
Susanna Eli, Rossella Castagna, Marina Mapelli and Emilio Parisini
- 53 Structural Basis of Human Dimeric  $\alpha$ -Amino- $\beta$ -Carboxymuconate- $\epsilon$ -Semialdehyde Decarboxylase Inhibition With TES-1025  
Michele Cianci, Nicola Giacchè, Lucia Cialabrini, Andrea Carotti, Paride Liscio, Emiliano Rosatelli, Francesca De Franco, Massimiliano Gasparrini, Janet Robertson, Adolfo Amici, Nadia Raffaelli and Roberto Pellicciari
- 62 Experiences From Developing Software for Large X-Ray Crystallography-Driven Protein-Ligand Studies  
Nicholas M. Pearce, Rachael Skyner and Tobias Krojer
- 73 High-Throughput Native Mass Spectrometry Screening in Drug Discovery  
Agni F. M. Gavrilidou, Kleitos Sokratous, Hsin-Yung Yen and Luigi De Colibus
- 88 Targeting the Essential Transcription Factor HP1043 of *Helicobacter pylori*: A Drug Repositioning Study  
Federico Antoniciello, Davide Roncarati, Annamaria Zannoni, Elena Chiti, Vincenzo Scarlato and Federica Chiappori
- 103 Structural Characterization of *Mycobacterium abscessus* Phosphopantetheine Adenylyl Transferase Ligand Interactions: Implications for Fragment-Based Drug Design  
Sherine E. Thomas, William J. McCarthy, Jamal El Bakali, Karen P. Brown, So Yeon Kim, Michal Blaszczyk, Vitor Mendes, Chris Abell, R. Andres Floto, Anthony G. Coyne and Tom L. Blundell
- 115 Innovative Approach for a Classic Target: Fragment Screening on Trypanothione Reductase Reveals New Opportunities for Drug Design  
Annarita Fiorillo, Gianni Colotti, Cécile Exertier, Anastasia Liuzzi, Francesca Seghetti, Alessandra Salerno, Jessica Caciolla and Andrea Ilari



**126 A Fe<sup>2+</sup>-dependent self-inhibited state influences the druggability of human collagen lysyl hydroxylase (LH/PLOD) enzymes**

Luigi Scietti, Elisabetta Moroni, Daiana Mattoteia, Marco Fumagalli, Matteo De Marco, Lisa Negro, Antonella Chiapparino, Stefano A. Serapian, Francesca De Giorgi, Silvia Faravelli, Giorgio Colombo and Federico Forneris

**142 Crystal polymorphism in fragment-based lead discovery of ligands of the catalytic domain of UGGT, the glycoprotein folding quality control checkpoint**

Alessandro T. Caputo, Roberta Ibba, James D. Le Cornu, Benoît Darlot, Mario Hensen, Colette B. Lipp, Gabriele Marcianò, Snežana Vasiljević, Nicole Zitzmann and Pietro Roversi



## OPEN ACCESS

EDITED AND REVIEWED BY  
Annalisa Pastore,  
King's College London, United Kingdom

\*CORRESPONDENCE  
Pietro Roversi,  
✉ [pietro.roversi@ibba.cnr.it](mailto:pietro.roversi@ibba.cnr.it)

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

†These authors have contributed equally to this work

RECEIVED 16 January 2023  
ACCEPTED 23 January 2023  
PUBLISHED 01 February 2023

CITATION  
Roversi P, Molla G and Nardini M (2023),  
Editorial: Integration of structural biology  
data in lead drug discovery  
and optimization.  
*Front. Mol. Biosci.* 10:1145834.  
doi: 10.3389/fmolb.2023.1145834

COPYRIGHT  
© 2023 Roversi, Molla and Nardini. 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.

# Editorial: Integration of structural biology data in lead drug discovery and optimization

Pietro Roversi<sup>1,2\*†</sup>, Gianluca Molla<sup>3†</sup> and Marco Nardini<sup>4†</sup>

<sup>1</sup>IBBA-CNR Unit of Milano, Institute of Agricultural Biology and Biotechnology, Milano, Italy, <sup>2</sup>Department of Molecular and Cell Biology, Leicester Institute of Structural and Chemical Biology, University of Leicester, Leicester, United Kingdom, <sup>3</sup>The Protein Factory 2.0, Dipartimento di Biotechnologie e Scienze della Vita, Università degli Studi dell'Insubria, Varese, Italy, <sup>4</sup>Department of Biosciences, University of Milano, Milano, Italy

## KEYWORDS

lead drug discovery, fragment based drug discovery, integrative biology, virtual high-throughput screening, drug repositioning

## Editorial on the Research Topic

Integration of structural biology data in lead drug discovery and optimization

Proteins, nucleic acids, and their complexes represent the vast majority of drug targets. The rational identification, design and optimization of a drug-lead compound targeting specific macromolecular sites is driven by the knowledge of the atomic interactions between the target itself and the small-molecule ligand under investigation. Of course, additional constraints besides the drug candidate's molecular and chemical properties must be taken into consideration (e.g. bioavailability), but the knowledge of the molecular details of the lead-target interactions also plays a key role in the prediction and rationalisation of those properties. Thus, structural knowledge contributes both to the design and generation of initial lead compounds and to the medicinal chemistry process bridging from lead to final active pharmaceutical drug in the clinic.

Traditionally, macromolecular X-ray crystallography has been the experimental structural technique of choice for structure-based lead drug discovery (SBLDD), with Cryo-EM the latest addition to the structural biology toolbox (Cabral et al., 2022). Native Mass Spectrometry (nMS) now enters the same arena: the technique's high sensitivity, simplicity, speed, wide dynamic range, low protein and ligand requirement, and the possibility of automation make nMS an integral component of the drug discovery pipeline, especially for primary screening (Gavriilidou et al.). When the dynamics of drug-target association matter, the great flexibility and adaptability of NMR provide for qualitative and quantitative insights at each point of the drug development process, aiding hit-identification and detection of weak binders, a paramount advantage in early stages of lead drug discovery (LDD) (Mureddu and Vuister).

Of course, advances in structural bioinformatics (e.g., novel and/or improved data processing workflows or application of machine learning approaches) is just as important to drug discovery as the advances in the experimental techniques themselves. Progress on the experimental front will have to go hand in hand with changes in mindset both on the side of depositors and users of macromolecular models, as exemplified by developments in software and algorithms for X-ray Fragment Based Lead Discovery (FBLD) such as batch data processing methods, X-ray diffraction data analysis and presentation, modelling, refinement, deposition of structures (Pearce et al.) and automation of crystal polymorph assignment (Caputo et al.).

Determining the crystal structure of an enzyme-inhibitor complex may be considered the simplest example of integration of structural biology data in lead drug optimization, but the

impact of such a simple experiment can often change the direction of the research trajectory. The study by [Cianci et al.](#) constitutes an example of an experimentally determined crystal structure correcting initial *in silico* predictions of a ligand-binding pose. Their crystal structure of human  $\alpha$ -amino- $\beta$ -carboxymuconate- $\epsilon$ -semialdehyde decarboxylase in complex with TES-1025 revealed unforeseen protein-ligand interactions, allowing the redefinition of the paradigmatic principles of the medicinal chemistry effort to improve potency and selectivity of the drug.

FBLD often affords the two-birds-with-one-stone simultaneous identification of novel lead scaffolds and previously unknown accessory allosteric drug binding sites on the target. The FBLD effort by [Fiorillo et al.](#), devoted to finding ligands of *T. brucei* trypanothione reductase, identified 12 new ligands, binding five different sites, two of which had not been exploited for inhibition previously. Crystal structures were also a key determinant in the success of the fragment screening campaigns carried out by [Thomas et al.](#) to inhibit *M. abscessus* and *M. tuberculosis* phosphopantetheine adenylyltransferase (PPAT). This study expands the chemical space of mycobacterial PPAT inhibition by discovering novel ligand-binding interactions in already known target sites, as well as previously undescribed ligand-induced cryptic sites.

When tackling the daunting challenge of discovering novel antibiotics, several parallel FBLD efforts targeting different previously neglected bacterial proteins hasten success. [Arif et al.](#) review computational approaches (such as structure modelling and assessment of the suitability of a protein as target for novel antibiotics) that informed the search of novel molecules binding previously untargeted pathogenic *Pseudomonas aeruginosa* proteins.

Structural information about conventional inhibitors can also serve as a starting point for more cutting-edge, unconventional molecules to be designed, as ([Eli et al.](#)) illustrate in their review on the identification of novel tubulin modulators and their delivery strategies. Novel classes of such molecules include PROTACs as well as light-sensitive compounds that can be activated with high spatial-temporal accuracy and therefore represent promising tools for precision-targeted chemotherapy.

Conversely, when a drug candidate affects the activity of its target in ways that cannot be detailed by experimental structural studies, integration of molecular dynamics (MD) and biochemical data can come to the rescue. The study on the effect of  $\text{Fe}^{2+}$  chelating compounds on the activity of human collagen lysyl hydroxylase by [Sciatti et al.](#) reports on a fine balance between  $\text{Fe}^{2+}$ -dependent enzymatic activity and  $\text{Fe}^{2+}$ -induced self-inhibited enzyme conformations. The latter could not be captured by conventional

structure-based approaches, but MD simulations were successful in rationalising the enzymatic activity data and boosting inhibitor development.

Last but not least, SBLDD can be speeded up by combining the power of *in silico* virtual high-throughput screening (VHTS) with the wealth of pharmacological information intrinsically carried by a drug repositioning approach. [Antoniciello et al.](#) successfully applied VHTS to a large number of known drugs already in clinic trials, checking their capability to inhibit the DNA-binding of the *Helicobacter pylori* HP1043 essential transcription factor. Three lead compounds that gave good activity *in vitro* turned out not to belong to the same chemical scaffold and to bind different protein sites, thus broadening the scope of the medicinal chemistry effort.

Overall, the contributions to this Research Topic highlight recent advances in structural biology, computational approaches, and high-throughput methods which have provided researchers with a huge amount of bewildering data and enabled novel approaches to data analysis and assessment. In this perspective, integrative structural biology converts structural information into chemical knowledge, feeding high quality experimental data into the drug development pipeline, thus supporting faster and more efficient delivery of drugs to the clinic.

## Author contributions

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

## Conflict of interest

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

## Publisher's note

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

## Reference

Cabral, A., Cabral, J. E., and McNulty, R. (2022). Cryo-EM for small molecules. *Curr. Protoc.* 2, e632. doi:10.1002/cpz1.632



# Fragment-Based Drug Discovery by NMR. Where Are the Successes and Where can It Be Improved?

Luca G. Mureddu and Geerten W. Vuister\*

Leicester Institute of Structural and Chemical Biology, Department of Molecular and Cell Biology, University of Leicester, Leicester, United Kingdom

## OPEN ACCESS

### Edited by:

Marco Nardini,  
University of Milan, Italy

### Reviewed by:

Giovanna Musco,  
San Raffaele Hospital (IRCCS), Italy  
Cristina Airola,  
University of Milano-Bicocca, Italy

### \*Correspondence:

Geerten W. Vuister  
gv29@leicester.ac.uk

### Specialty section:

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

**Received:** 13 December 2021

**Accepted:** 24 January 2022

**Published:** 18 February 2022

### Citation:

Mureddu LG and Vuister GW (2022)  
Fragment-Based Drug Discovery by  
NMR. Where Are the Successes and  
Where can It Be Improved?  
Front. Mol. Biosci. 9:834453.  
doi: 10.3389/fmolb.2022.834453

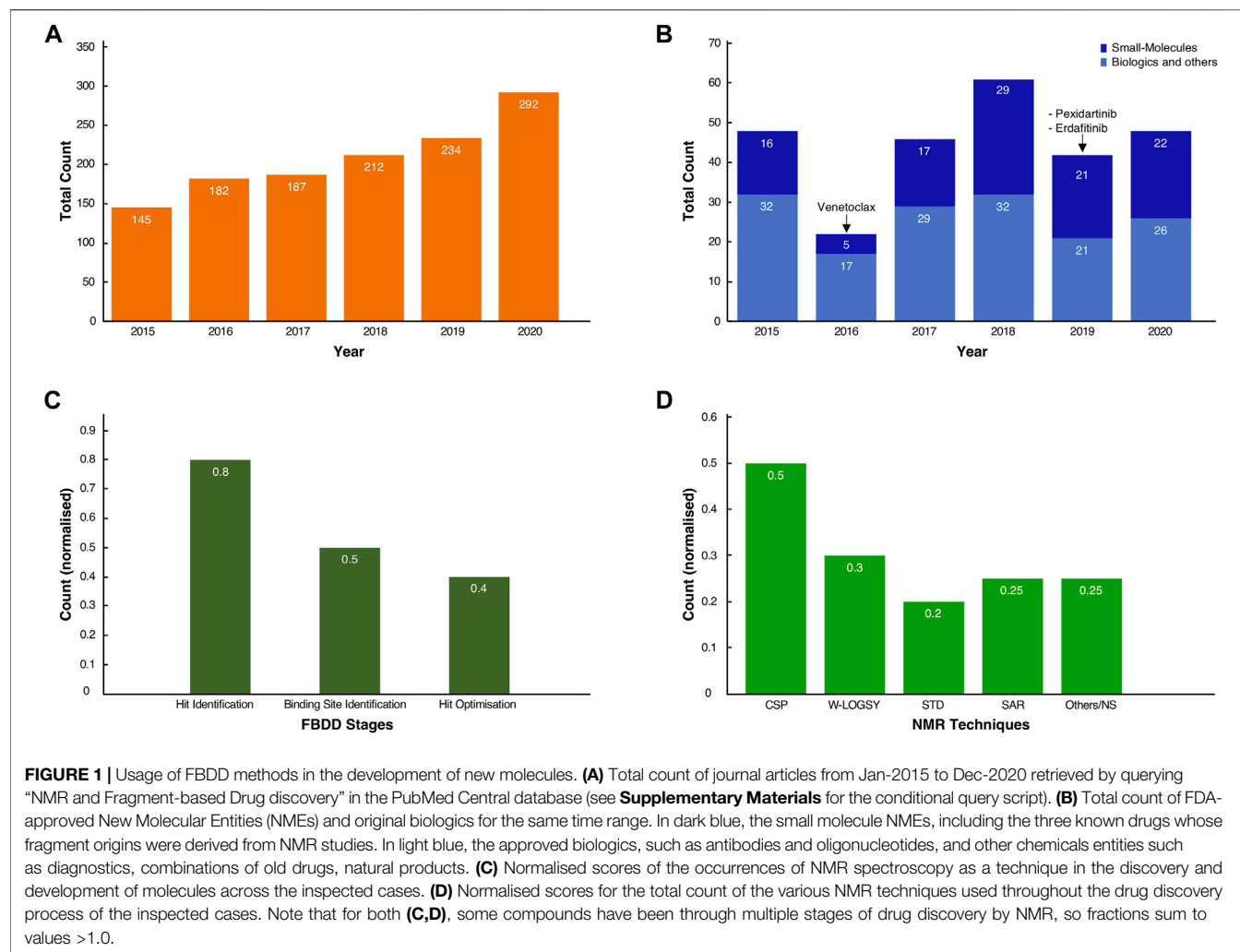
Over the last century, the definitions of pharmaceutical drug and drug discovery have changed considerably. Evolving from an almost exclusively serendipitous approach, drug discovery nowadays involves several distinct, yet sometimes interconnected stages aimed at obtaining molecules able to interact with a defined biomolecular target, and triggering a suitable biological response. At each of the stages, a wide range of techniques are typically employed to obtain the results required to move the project into the next stage. High Throughput Screening (HTS) and Fragment Based Drug Design (FBDD) are the two main approaches used to identify drug-like candidates in the early stages of drug discovery. Nuclear Magnetic Resonance (NMR) spectroscopy has many applications in FBDD and is used extensively in industry as well as in academia. In this manuscript, we discuss the paths of both successful and unsuccessful molecules where NMR had a crucial part in their development. We specifically focus on the techniques used and describe strengths and weaknesses of each stage by examining several case studies. More precisely, we examine the development history from the primary screening to the final lead optimisation of AZD3839 interacting with BACE-1, ABT-199 interacting with BCL<sub>2</sub>/XL and S64315 interacting with MCL-1. Based on these studies, we derive observations and conclusions regarding the FBDD process by NMR and discuss its potential improvements.

**Keywords:** Fragments Based Drug Discovery, NMR-FBDD, Venetoclax, BCL-2, BACE-1, MCL-1

## INTRODUCTION

Fragment Based Drug Discovery, or FBDD for short, is nowadays a well-established and common approach adopted by many pharmaceutical companies and academic groups (Erlanson et al., 2016). The rationale behind FBDD has been extensively reviewed and entire books have been written, including some with references to clinical candidates (Klon, 2015; Rees, 2016). The general concept of FBDD is straightforward; it starts with the

**Abbreviations:** BACE, Beta-Site APP Cleaving Enzyme; BCL, B-Cell Leukaemia/Lymphoma; CPMG, Carr-Purcell-Meiboom-Gill; Cryo-EM, Cryogenic electron microscopy; CSP, Chemical Shift Perturbation; FDA, Food and Drug Administration; HSQC, Heteronuclear Single Quantum Coherence Spectroscopy; MBP, Maltose-Binding Protein; MCL, Myeloid Cell Leukaemia; MD, Molecular Dynamics; MW, Molecular Weight; NMR, Nuclear Magnetic Resonance; NME, New Molecular Entities; NOE, Nuclear Overhauser Effect; NOESY, Nuclear Overhauser Effect Spectroscopy; PAIRS, Pan-Assay Interference Compounds; PSA, Polar Surface Area; SAR, Structure Activity Relationship; SPR, Surface Plasmon Resonance; STD, Saturation-Transfer Difference; W-LOGSY, Water-Ligand Observed via Gradient Spectroscopy.



generation of libraries of small molecules called “fragments.” The size of these libraries varies from a few hundred to thousands of molecules (for industrial cases) (Singh et al., 2018), usually prepared following the so-called “rule of three,” i.e. a molecular weight smaller than 300Da, lipophilicity as expressed by LogP smaller than 3, and a maximum number of hydrogen bond donors and acceptors less than 3 (Congreve et al., 2003). For efficiency reasons the binding against a target of interest is usually evaluated for each of these molecules in so-called “mixtures” of 5–10 compounds in a single experiment (Mercier and Powers, 2005). The strategy of using small molecules instead of large entities allows for a more efficient exploration of the chemical space, defined as the ensemble of all possible molecular conformations presenting drug-like properties, and estimated to be a staggering  $\sim 10^{60}$  molecules (Reymond et al., 2010). The FBDD approach provides for a great chemical variety to probe this chemical space and has many other benefits, such as cost and time reduction in the data analysis.

Using fragments as a starting point in the early stages of drug discovery has been demonstrated to be a viable

approach for producing compounds that are specifically tailored to their targets (Hughes et al., 2011). This approach also increases the novelty of standard drugs and enables the path of chemical optimisation to be monitored, for example by restricting the lipophilicity issue observed in lead molecules obtained by high-throughput screening (HTS) (Author Anonyms, 2007; Erlanson et al., 2016). However, since the fragments are much smaller compared to traditional lead-like molecules, their binding affinity to a target of interest is nearly always low ( $\mu\text{M}$  to  $\text{mM}$ ). Therefore, it requires a technique, such as NMR, that is capable of detecting these weak interactions (Ortega-Roldan et al., 2009; Vinogradova and Qin, 2012).

In this manuscript, we focus exclusively on the relevance and impact of NMR spectroscopy on the generation of new clinical drugs. Through a literature review, we establish the impact of NMR in FBDD. We performed an in-depth assessment of three case studies that establish the impact of various NMR techniques on different stages of drug development. We establish patterns in the drug-development optimization process and formulate proposals for its improvement.

## RESULTS

To appreciate the influence on NMR in the current developments, we queried the PubMed database with the “fragment-based drug discovery” and “NMR” keywords. It revealed a steady and significant increase in the number of publications, growing from 145 in 2015 to 292 in 2020, for a total of 1,200 journal publications over this 6-year period (**Figure 1A**). A further search through the FDA database revealed, over the same time span, that several drugs were unmistakably obtained from fragments (**Figure 1B**) (Deeks, 2016; The ASCO Post, 2019). Excluding biologics (antibodies, oligonucleotides etc.) and other chemical compounds (diagnostics, combination of old drugs etc.), original small molecules make up a significant proportion, typically ~40% or more, of the total FDA approvals in any given year (**Figure 1B**).

From the FDA database analysis, it appeared that only Venetoclax was developed with the help of at least one NMR technique in the early stages of the discovery. Whilst this may seem to be a low number relative to the total number of FDA approvals, this finding is likely explained by the fact that not every breakthrough in the drug development process obtained by NMR will have been shared in the public domain. Furthermore, despite our exhaustive literature search (*vide infra*), the origin of new molecular entities remained often obscure or very difficult to trace.

The true impact of NMR techniques in the development of original lead-molecules appeared clearer from the inspection of the origins of the clinical compounds. We performed an in-depth analysis of all FBDD-derived molecules that are or have been clinical candidates at any time in the past, up to December 2020, and for which relevant information was publicly accessible through journal articles, tables or from Dan Erlanson’s “Practical Fragments” blog (Erlanson, 2018). Our analysis showed that NMR is used in all three stages of the FBDD process (**Figure 1C**). For sixteen out of twenty clinical compounds (80%), we traced that NMR was used for identifying the initial binding fragments, the so called “hit identification” stage. For the subsequent FBDD stages, which include the “binding site identification” and “hit optimisation,” NMR was used less often, i.e. in 50 and 40% of the cases, respectively, (**Figure 1C**), with other techniques, such as X-ray crystallography, being preferred.

NMR-derived compounds were identified mostly by a number of ligand-detected, one-dimensional (1D) NMR techniques, such as Water-Ligand Observed via Gradient Spectroscopy (WaterLOGSY) (Antanasijevic et al., 2014), saturation transfer difference (STD) (Viegas et al., 2011; Meyer et al., 2004), or the so-called  $T_{1\rho}$ -experiment (Hajduk et al., 1997). In contrast, target-based two-dimensional (2D) NMR techniques, such as the chemical shift perturbation (CSP) experiment (Mureddu and Vuister, 2019), were used for the hit and/or binding site validation. Lastly, the so-called SAR (structure activity relationship) by NMR method (Fesik et al., 1997; Shuker et al., 1996), which employs mostly NOE-related techniques and multi-dimensional NMR experiments, was mainly used for hit-growing and linking guidance during the optimisation stage of the FBDD

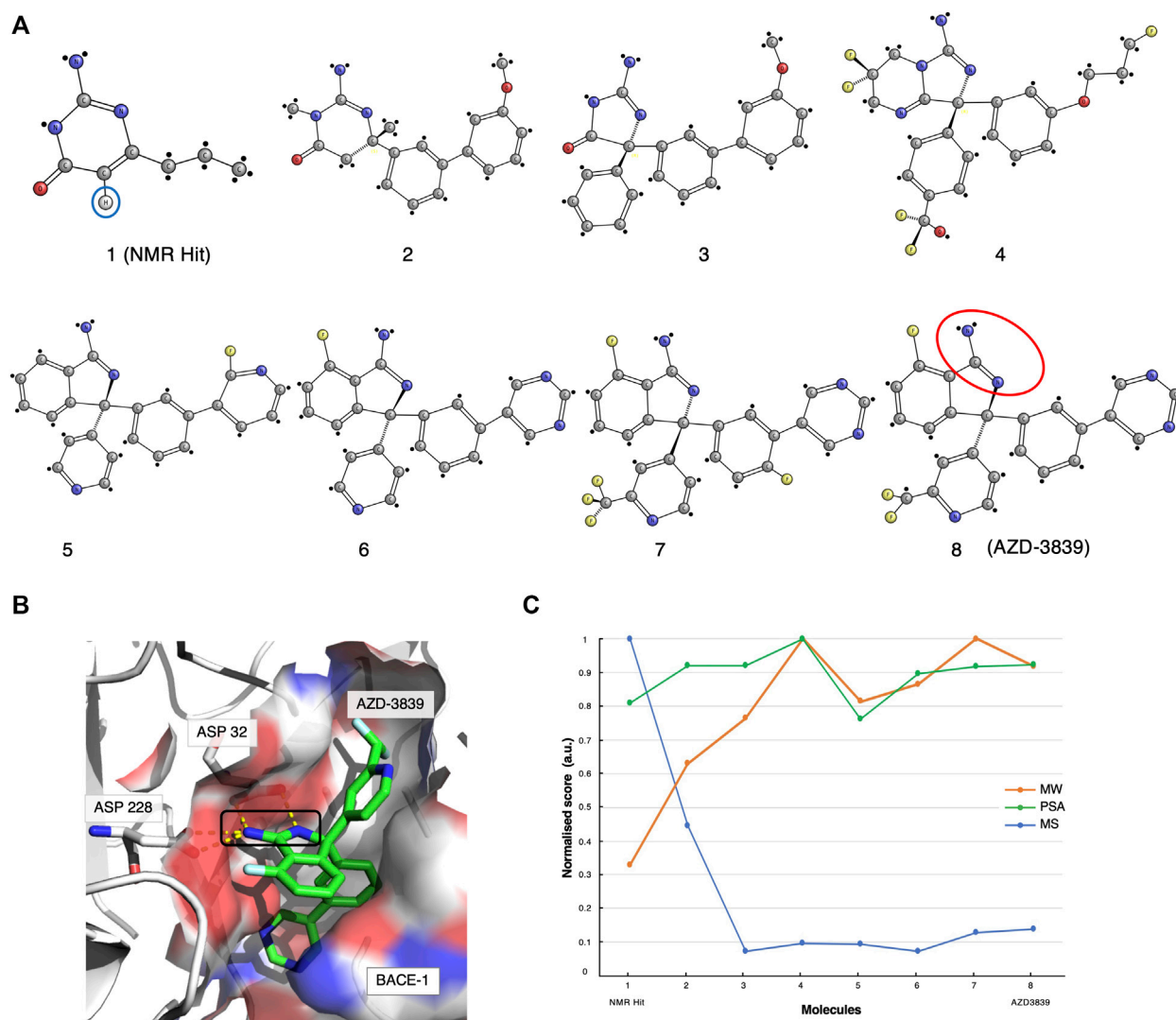
process (**Figure 1D**) (Petros et al., 2006) (see also **Supplementary Table S1**).

The 1D ligand-detected techniques are considered a gold standard in NMR screening, as these do not require expensive protein labelling and therefore can be used against a broad range of molecular targets (Ciulli and Abell, 2007). Furthermore, the various expression systems of the target, e.g. bacteria, insects or human-derived cells, and other common limitations, such as molecular weights, are not of critical importance for these 1D techniques (Campos-Olivas, 2011). Moreover, the inherent versatility of NMR has also allowed the detection of the binding activity of small-molecules to receptors in their native environments and in real-time, a strategy called *in-cell* NMR (Primikyri et al., 2018). Many common NMR techniques have been used in or adapted for *in-cell* NMR (Kang, 2019). However, STD and Tr-NOESY techniques (Mari et al., 2010) have been successfully employed without the limiting step of isotopic labelling (Primikyri et al., 2018).

In addition, 1D ligand-detected techniques can also be utilised in difficult cases where expression and/or purification of the target macromolecule is a limiting factor and only nanomolar concentrations can be obtained. Most importantly, the richness of information acquired in a small amount of time (i.e. minutes per sample) allows the analysis to be performed in a high-throughput fashion (Dalvit et al., 2002). However, 1D ligand-detected experiments are not suitable for detecting the binding sites on the target, and higher dimensionality NMR techniques, including chemical-shift perturbation mapping (Williamson, 2013; Mureddu and Vuister, 2019), are often required. The latter enables the monitoring of target residues that are most likely to be interacting with the fragments, providing validation of binding as well as guidance on the next stage of development (Williamson, 2013).

Fragment optimisation is best achieved where a high-resolution 3D molecular structure of the target is available. While there are several techniques capable of resolving molecular structures, the simplicity and the rapid throughput associated with X-ray crystallography, it makes this the preferred method whenever possible (Erlanson, 2015). However, often targets of interest cannot always be exhaustively assessed by X-ray crystallography. For example, complexes displaying a highly flexible mode of interaction with the target molecule are best inspected by NMR (Teague, 2003; Leone et al., 2010), as crystal packing forces preclude the molecular adaptation required for complex formation. Moreover, the crystal lattice also might not allow the ligand to permeate through to the binding pockets (Zheng et al., 2015). In contrast, the NMR technique can provide unambiguous information on the various orientations of the ligand with respect to the target, referred to as poses. These poses can be combined with computational methods for designing drug-like compounds with improved binding and pharmacological properties. Interestingly, the emergence of Artificial Intelligence (AI)-driven structure elucidation, such as AlphaFold-2 (Jumper et al., 2021) and RoseTTA fold (Baek et al., 2021) provides an additional avenue for the fragment-optimization stage. A description of the most recurrent NMR techniques to elucidate





**FIGURE 2 |** The AZD-3839 case-study. **(A)** The optimisation pathway: from NMR hits to AZD-3839. Compound-1 represents the hit initially identified from the Water-LOGSY NMR study. The blue circle highlights the crucial isocytosine aromatic proton. Compounds were optimised through a series of crystallography-based methods to yield the final compound-8 (AZD-3839) (Geschwindner et al., 2007; Jeppsson et al., 2012), yet preserving the original amidine motif (red circle) already present in compound-1. **(B)** Molecular structure representation of BACE-1 (PDB code: 4B05) and the main interaction between the catalytic groove (Asp32 and Asp228) and the amidine group of AZD-3839, first observed in the NMR-discovered hit (black rectangle). **(C)** Molecular similarity, as expressed by the Tanimoto coefficient (MS, Blue), normalised molecular weight (MW, orange) and polar surface area (PSA, green) scores for the eight compounds on the development path of AZD-3839.

structure information is available in the supplementary materials (**Supplementary Table S1**).

In subsequent sections of this manuscript, we present three case studies that employed a variety of NMR techniques and therefore can be considered templates of NMR-FBDD. The first case study explores the development of a compound denoted as AZD-3839. It originated from fragments identified by a ligand-detected primary screening using the Water-LOGSY technique. The second case study examines the history of the FDA-approved drug ABT-199, commercially called Venetoclax, which was also derived by FBDD. The development of this drug relied on a

variety of target-detected NMR methods. Finally, the third case study analyses a compound known as S64315, a recent FDA-approved drug, which illustrates the role of NMR in FBDD through a combination of ligand-detected and target-based NMR techniques.

### Case Study 1: AZD3839 and BACE-1

$\beta$ -Site Amyloid precursor protein Cleaving Enzyme-1 (BACE-1) was identified over 20 years ago as a key component in Alzheimer disease (AD) pathogenesis (Vassar et al., 1999; Venugopal et al., 2008). BACE-1 is responsible for the initial cleavage of the



amyloid precursor protein into smaller amyloid  $\beta$ -peptides (A $\beta$ ), whose accumulation in brain cells is believed to be one of the underlying causes of AD progression (Vassar et al., 2009). Not surprisingly, BACE-1 is a therapeutic target and a number of academic groups and pharmaceutical companies have placed considerable efforts into the research and development of new inhibitors in the hope of limiting or blocking the formation of A $\beta$  (Ghosh et al., 2012; Goyal et al., 2014; Erlanson and Janke, 2016; Hung and Fu, 2017).

BACE-1 is characterised by an internal groove created by two lobes (S1 and S2), modulated by a loop ("flap") which reveals the aspartyl catalytic site. The flap is highly dynamic, and the presence of an inhibitor can determine whether the macromolecule is in the "open" or "closed" state, thus modulating access to the catalytic pocket. The identification of two crucial aspartic acid residues, i.e. Asp32 and Asp228, has for many years driven the drug development process and optimisation of fragments (Erlanson and Janke, 2016). An exhaustive list of early fragments and their respective primary screening techniques is given by Erlanson and Jahnke in their "Lessons and Outlook" book (Erlanson and Janke, 2016).

A great example of the history of a complete development of a BACE-1 inhibitor is given by the compound AZD-3839, where the initial fragment was identified from 1D NMR studies. According to Geschwindner *et al.* (Geschwindner et al., 2007) at AstraZeneca the choice of NMR for this case provided a compromise between scalability of large fragment libraries with sufficient data-output on the one hand, and ensuring a robust method for detecting very weak binding at low ligand concentration on the other hand. By eliminating non-specific binders, the process also had the advantage of reducing false positives from the analysis.

The original screen using the Water-LOGSY 1D NMR technique was conducted on a 2000-compound library with four fragments per mixture, yielding a relatively low hit rate of 0.5%. Compound-1 (Figures 1, 2A, ) was identified as a binding hit. The hallmark intensity "sign-flip" of the Water-LOGSY signals resulting from this compound was clearly identifiable in the NMR spectra, suggesting binding to BACE-1 (Geschwindner et al., 2007). Crucially, as a control, the authors performed a competition experiment in the presence of a stronger known binder, showing a noticeable intensity reduction (signal becomes more negative) only for the singlet peak from the isocytosine aromatic H<sup>5</sup> proton at around 5.65 ppm (Figure 2A, compound 1, blue circle). This validation assay reduced potential false positives by identifying fragments that displayed weak binding Water-LOGSY responses but did not show any changes upon the addition of the competitor. Compound-1 was eventually selected for further optimisation steps.

Meanwhile, through parallel crystallographic studies performed by Astex Therapeutics (Murray et al., 2007) an optimised compound that preserved the original amidine motif was developed. The amidine motif was confirmed to be responsible for the strong interaction to the catalytic aspartates (Figure 2B). Subsequently, through a series of

substitutions on the crucial scaffold, the molecule was morphed into the isoindole present in the final compound. Furthermore, the introduction of fluoro atoms improved the permeability of the molecule and the brain exposure by "shielding" the reactive amidine. Lastly, additional molecular interaction surface on the molecule, needed for the interaction of the molecule with the adjacent S3 and the flap (Edwards et al., 2007) (Figure 2B), was created by extension with additional aromatic moieties.

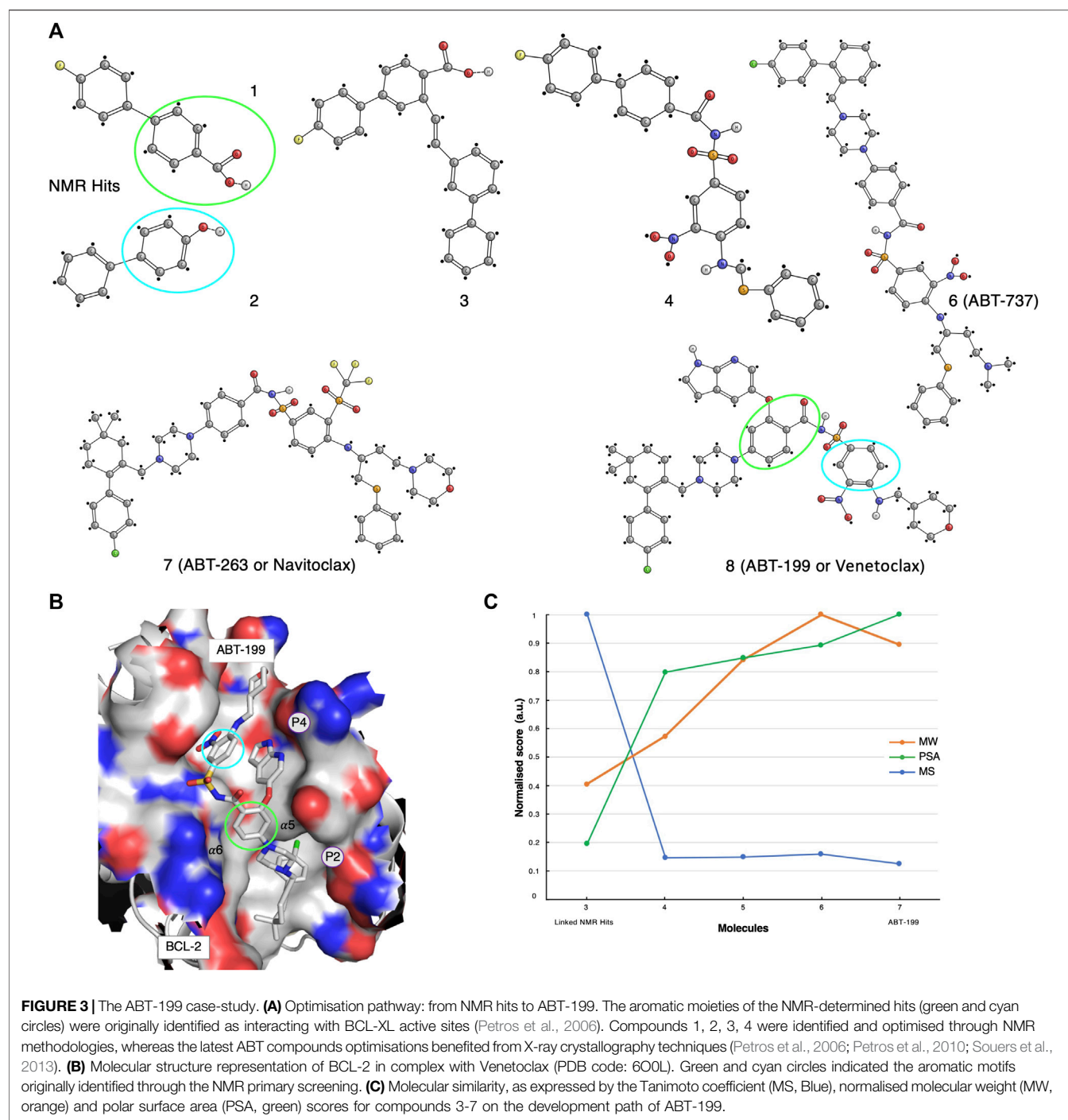
The various steps of this hit optimisation process clearly show how the initial NMR-derived fragment has undergone a series of dramatic changes. The magnitude of these changes can be assessed from the similarities of the molecular characteristics for each component, as given by its molecular weight (MW), polar surface area (PSA) and by the Tanimoto coefficient (Bajusz et al., 2015) which measures molecular similarity (MS, Figure 2C). A prominent drop of the Tanimoto coefficient is observed from the first NMR-detected compound-1 to compound-2 and further for compound-3. However, from compound-4 to the final AZD-3839 molecule, a much smaller variation in the score is observed. A different trend was observed for the molecular weight of successive compounds which showed a constant increment up to compound-4, followed by only minor changes towards the final AZD-3839. Interestingly, the final compound was characterised by a ~10% smaller molecular weight compared to its predecessor yet attained a slightly increased PSA value (Figure 2C).

AZD-3839 appeared to be a very promising drug candidate and underwent clinical phase-1 trials. Unfortunately, it was stopped from patient administration, probably due to its high affinity to the hERG ion channel and resulting related side-effects (Berg et al., 2012). Nevertheless, this case demonstrated that NMR was crucial for determining the first hit from the primary screening containing the essential amidine-fragment. This motif proved to be of a critical importance in the interactions to BACE-1, and as a result it was preserved through the long path of further chemical optimisations that resulted in the final AZD-3839 compound.

## Case Study 2: ABT-199 (Venetoclax) and BCL-2/xL

The second case study concerns the analysis of ABT-199, commercially referred to as Venetoclax, that obtained FDA-approval as BCL-2 inhibitor in 2016. This drug was selected for our studies for two reasons: firstly, the large impact of NMR throughout its development pathway and secondly because the Abbott NMR group, who initiated the studies on ABT-199, have pioneered the so-called "SAR by NMR" method that also underpinned its development (Shuker et al., 1996; Fesik et al., 1997)  $\alpha$ .

ABT-199 is an inhibitor of the anti-apoptotic proteins BCL-2, BCL-xL, and BCL-w (Souers et al., 2013). These proteins play a pivotal role in cell survival; not surprisingly, they are over-expressed in many cancers and they are directly linked to initiation, progression and therapy resistance occurrences



(Choudhary et al., 2015). The BCL members are  $\alpha$  helical proteins; two of these, BCL-2 and BCL-xL, share four domains, BH4 ( $\alpha 1$ ), BH3 ( $\alpha 2$ ), BH1 (partially  $\alpha 4$ ), and BH2 (partially  $\alpha 6$  and  $\alpha 7$ ) plus the transmembrane, TM, motif. The two central hydrophobic helices ( $\alpha 5$  and  $\alpha 6$ ) together with the amphipathic  $\alpha 1-4$  and  $\alpha 7$  together form an elongated hydrophobic groove in the so-called BH1, BH2, BH3 regions (Kelekar and Thompson, 1998) (Figure 3B). The BH3 region, in

particular, is responsible for the interaction with other proapoptotic proteins such as BAK and BAX, rendering it a druggable site of interest (Letai et al., 2002).

The early inhibitor-discovery process was started by screening a large library of small molecules using 2D target-detected approaches, which led to the identification of several molecular candidates (Figure 3A, compounds 1 and 2). The hypothetical binding mechanism was elucidated through

$^{15}\text{N}$ -HSQC chemical shift perturbations (CSP) experiments (Petros et al., 2006). From the CSP results, it was possible to derive that the fluoro-biaryl acid region of compound-1 interacted with the BCL-xL hydrophobic groove. In fact, a series of shifts were observed for the peaks assigned to BCL-xL residues Gly94, Gly138, and Gly196, located in this groove (Petros et al., 2006). However, the study of the complex of BCL-xL with its binding partner BAK suggested the existence of an additional binding interface. Therefore, a second NMR screening was carried-out in the presence of a large excess of compound-1, with the aim of saturating the first site of interaction and while screening for potential hits to the second interface (Petros et al., 2006). Compound-2 (**Figure 3A**) was identified and eventually chosen to be used for chemical linkage to compound-1, in-line with the “SAR by NMR” approach (Petros et al., 2006). Multiple linkers, derived on the basis of various poses of compound-1 in the BH3 binding groove, were explored in order to improve the overall potency of the resulting molecule. Finally, a ~200-fold improvement in binding affinity was established for compound-3 when compared to the original compound-1 (**Figure 3A**).

The first molecular model of the complex of BCL-xL with compound-3 was then derived on the basis of nine intermolecular NOEs (Petros et al., 2006). Although these NOEs were indicative of an interaction with both binding interfaces, it was concluded that compound-3 did not adopt optimal or ideal conformations. Consequently, new linkers and a new set of chemical reactions for connecting the fragments were explored.

Compound-4 was eventually synthesised and structurally evaluated on the basis of protein-ligand NOEs extracted from  $3\text{D}^{13}\text{C}$ -edited and  $^{12}\text{C}$ -filtered NOESY spectra, that allowed the molecule to be docked in the BH3 groove (Petros et al., 2006). Compound-4 was further optimised for those parts that were solvent-exposed. These parts of the molecule were replaced with polar substituents, including a 2-dimethylaminoethyl group in the linker. In addition, the insertion of a new piperazine ring resulted in compound-6, also referred to as ABT-737 (**Figure 3A**). ABT-737 displayed biological activity in the presence of human serum. A crystal structure at 2.2 Å resolution of BCL in complex with ABT-737, solved after the original NMR-determined binding pose of compound-4, validated the latter (Lee et al., 2007). The structure showed that ABT-737 interacted with the two binding interfaces formed by the hydrophobic pockets, P4 and P2, of BCL-2 and BCL-XL; including two hydrogen bonds from the thiophenyl and the 1-chloro-4-(4,4-dimethylcyclohex-1-enyl)benzene moieties to residues Gly138 and Glu96, respectively.

An *in-vivo* analysis suggested that synergetic therapy was required to inhibit the anti-apoptotic activity of the BCL family, while simultaneously promoting the activity of the pro-apoptotic proteins (BAX and BAK) (Oltersdorf et al., 2005). Therefore, the Abbott group also developed the ABT-263 molecule. After an initial positive assessment on multiple cellular lines, where ABT-263 reported stronger inhibitory actions, presumably by targeting both BCL-xL and BCL-2, advanced clinical studies unfortunately revealed major

physiological pitfalls such as thrombocytopenia (Vandenberg and Cory, 2013).

Meanwhile, the project progressed to the final compound ABT-199 (Venetoclax) from ABT-263 through a series of substitutions (Vandenberg and Cory, 2013). In addition, new 3D crystal structures of BCL-2 in complex with various ligands were made publicly available (PDB codes listed in **Supplementary Material and Methods**) (Souers et al., 2013; Lee et al., 2007; Birkinshaw et al., 2019). The ABT-199 molecule (**Figure 3A**, compound-8) incorporated several crucial modifications compared to its ABT-263 (**Figure 3A**, compound-7) predecessor; a pivotal H-bond to Asp103 (corresponding to Glu96 in BCL-xL) was identified, thus providing an increased affinity to both the BCL-2 and BCL-xL P4 pockets (Souers et al., 2013) (**Figure 3B**).

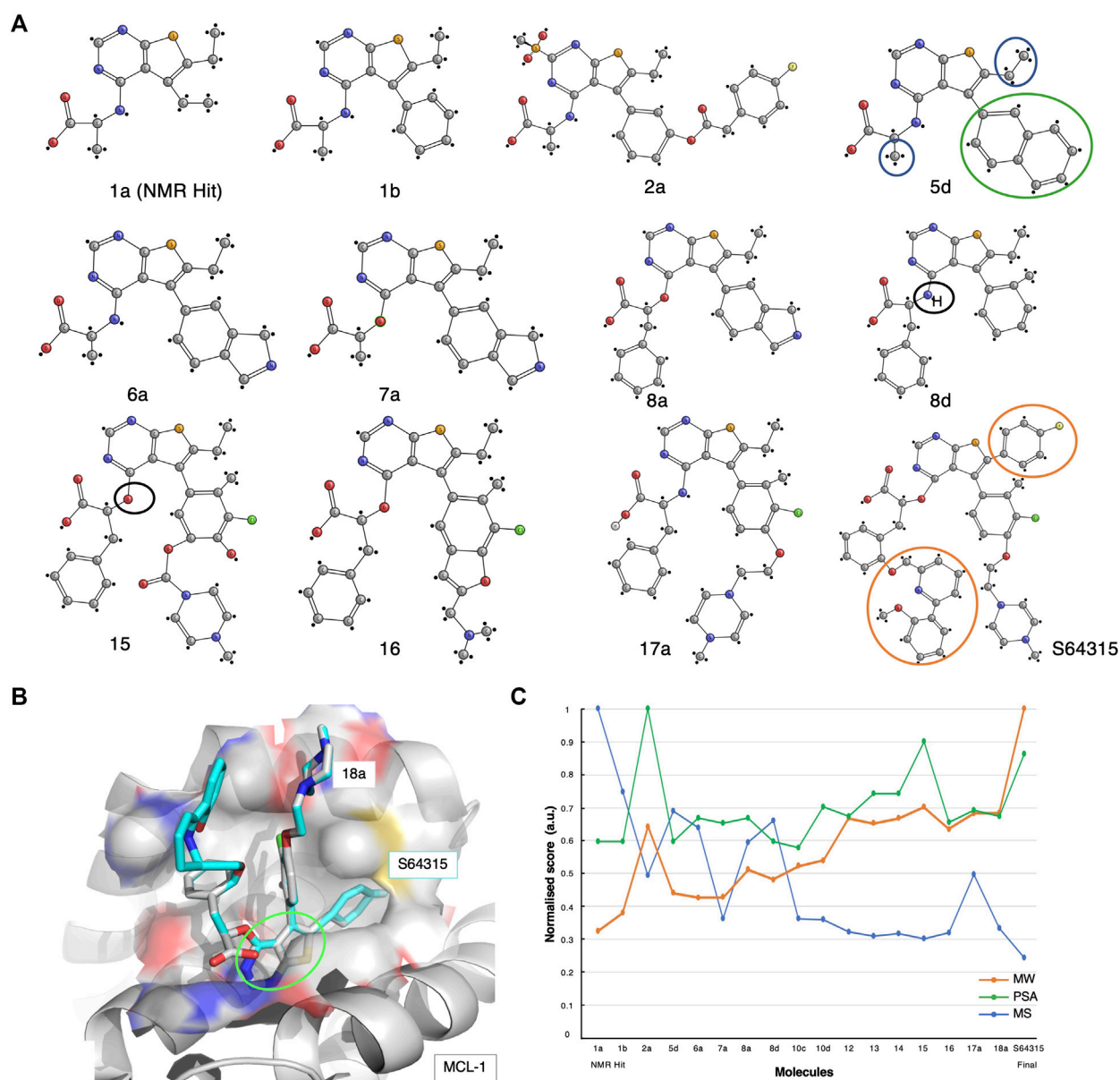
We performed a molecular fingerprint analysis for all available molecules in the development process. However, as the initial NMR-detected fragments underwent a linkage step, the molecular similarities were assessed with respect to compound-3 (**Figure 3A**). Similarly to the AZD-3839 case study, a drastic drop in the Tanimoto coefficient was observed from compound-3 to the following optimised forms (**Figure 3C**). Interestingly, ABT-199 showed a reduced molecular weight and increased polar surface area compared with its predecessor compounds yet keeping an overall structural similarity relative to compound 4.

The ~20-years development history of the ABT-199 compound revealed a multitude of challenges, including the initial failure in obtaining crystals of complexes with the first leads and various other *in-vivo* difficulties, which were not predictable from a structural point of view. However, our analysis shows that the development of Venetoclax most likely would not have been achieved without the crucial data obtained from multiple NMR techniques at the various stages of the ABT-199 drug-development process.

### Case Study 3: S64315/MIK665 and MCL-1

The third case study presents an overview of the most crucial optimisation steps in the development of the molecule S64315, also known as MIK665 (Szlávik et al., 2019) (**Figure 4A**). S64315 is one of the most recent inhibitors currently being tested in clinical trials for targeting the BCL anti-apoptotic family, MCL-1 (Maragno et al., 2019). A series of studies indicated that MCL-1 is over-expressed in many cancer types (multiple myeloma, lymphomas, leukaemia) and therefore it is widely recognised as a druggable target (Albershardt et al., 2011). MCL-1 shares the highly conserved BH3 binding groove with other members of the family such as BCL-xL and BCL-2 (*vide infra*). This groove is essential for interacting and sequestering the pro-apoptotic proteins resulting in an increased cell survival.

The development of specific inhibitors for MCL-1 which target the BH3 groove has proven to be challenging (S. Soderquist and Eastman, 2016). Researchers at Vernalis, together with collaborators, engaged in extensive efforts in their studies of this complex. These studies maximised the potency of an initial hit that was obtained from a ligand-detected NMR

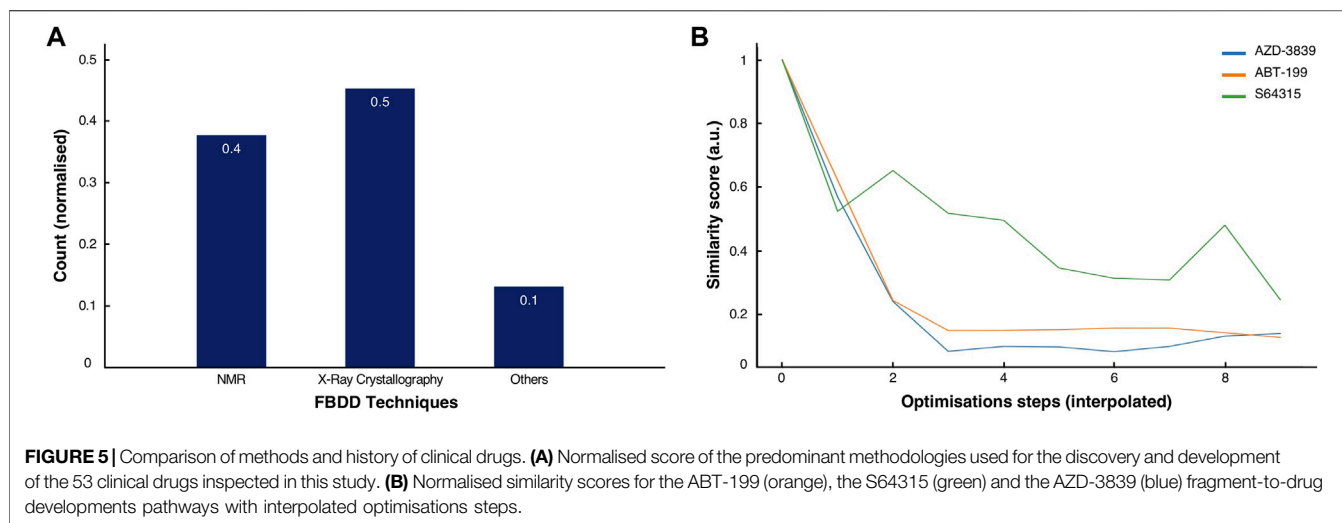


**FIGURE 4 |** The S64315 case-study. **(A)** Optimisation pathway: from NMR hits to S64315. All compound nomenclatures are identical to those used in the original manuscript (Szlávik et al., 2019) for an easier comparability. Compound **1a** represents the initially identified thienopyrimidine core by ligand-detected 1D NMR techniques. The green and blue circles for compound **5d** highlight the chemical groups that gave rise to crucial NOEs that suggested the initial molecule binding poses (Szlávik et al., 2019). **(B)** Molecular structure representation of a model of MCL-1 in complex with compound **18a** (PDB code: 6QYO). The green ellipse highlights the original thienopyrimidine motif first identified by an 1D-NMR screening experiment (Szlávik et al., 2019). **(C)** Molecular similarity, as expressed by the Tanimoto coefficient (MS, Blue), normalised molecular weight (MW, orange) and polar surface area (PSA, green) scores for the twelve compounds on the development path of S64315 AZD-3839.

screening, and resulted in the most promising MCL-1 inhibitor to date, S64315/MIK665 (Szlávik et al., 2019). However, several difficulties had to be overcome during its development, such as the lack of 3D atomic structures resulting from the poor expression and purification of MCL-1 in human-cell lines (Szlávik et al., 2019). Despite this, the protein availability proved adequate for the initial NMR-based screening.

A thousand initial compounds, pooled in groups of eight, were screened using various 1D NMR techniques, such as STD, WaterLOGSY and relaxation experiments (cf. **Supplementary Table S1**), to reveal several potential binding hits. Due to low signal-to-noise ratios of the screening experiments because of limited sample availability, hits were further validated using 2D NMR  $^{15}\text{N}$ -HSQC titrations. In addition, to overcome the lack of a





detailed 3D molecular structure, a new approach for determining ligand poses and guiding the drug development optimisation process was developed. This approach, referred to by the researchers as the NMR-guided model (NGM), employs 3D NMR methodology, i.e., X-filtered NOESYs ( $^{13}\text{C}$ -edited,  $^{13}\text{C}$ ,  $^{15}\text{N}$ -filtered), to identify crucial NOEs between ligands and the target. The information resulting from these NMR studies was combined with high-throughput computational docking studies, allowing for a more accurate classification of binding poses. From the NMR results, multiple compounds with various chemical functionalities were explored, of which a class of compounds comprising a thienopyrimidine group were believed to be the most promising. Particularly, NMR-derived compound-1a was used as the initial fragment towards the development of the S64315 drug (Figure 4A). Following a series of substitutions for the compound's ethyl group, multiple variants were tested on BCL-2, BCL-xL and MCL-1. Some of the newly synthesised molecules showed comparable affinity toward all three targets (Szlávik et al., 2019). Using the  $^{15}\text{N}$ -HSQC technique, it was possible to estimate  $K_d$  values for most of these, which ultimately allowed the selection of compound-5days as the highest affinity binder for MCL-1.

The NOEs derived from the analysis of the compound-5days/MCL-1 complex indicated several potential contacts. In particular, contacts between the naphthyl ring (Figures 4A, 5D, green circle) and the MCL-1 side chains of Ala227, Met231, Val249, Val253, and Thr266 were observed as well as between a methyl group (Figures 4A, 5D, blue circles) and the side chains of Met231, Val249, Val253, Leu267 (Szlávik et al., 2019). To further investigate the BH3 binding region's molecular flexibility, the researchers assessed various possible docking poses using multiple structural ensembles. This approach enabled a better estimation of the possible allowed geometries that were consistent with the experimental NOE information. Ultimately, the preferred molecular orientation

consisted of the carboxylic acid pointing toward the solvent region, and the naphthyl group toward the S2 pocket. Different conformations and variations of the ligand molecule interacting with the hydrophobic groove were also assessed. This was achieved by modifying the core of the original compounds by inserting various aliphatic substituents and testing the different rotational properties of the resulting aryl ethers and anilines (Figure 4A, 8days-15, black circles).

At a much later stage, crystallographic structures of the MCL-1 complex and some variants became available (PDB codes listed in **Supplementary Material and Methods**), allowing for more detailed studies of several fragments and their binding modes.

Multiple optimisation steps were carried out, eventually leading to S64315 (Figure 4A, compound-8a). This final compound presented new crucial ortho-substituents, such as the fluorobenzene and methoxyphenyl-pyridine group (Figure 4A, orange circles), which were responsible for the increased selectivity for MCL-1 compared to its precursor. Here, we have generated a model of MCL-1 in complex with compound-18a (Szlávik et al., 2019) using the MCL-1 crystal structure (PDB code 6QYO) by manually overlaying the S64315 molecule onto compound-18a (Figure 4B). The model shows the thienopyrimidine motif, already observed in the original NMR-derived compound-1a, deeply buried in the hydrophobic groove of the BH3 binding domain.

As for the previous case-studies, we inspected the molecular characteristics using the normalised MW and PSG scores and the Tanimoto score for all available compounds (Figure 4C). In line with the observations for both ABT-199 and AZD-3839, albeit somewhat less prominent, the result of this analysis again shows the characteristic initial decrease in molecular similarity from the first fragment to the following variants, indicating the significant changes during the initial steps of development. Interestingly, several compounds mid-way through the development (i.e., compounds 5 days, 8 days) showed a higher Tanimoto coefficient compared to the

initially optimised fragments, suggesting a more careful optimisation process rather than a revolutionary approach to the first NMR-derived hit. Starting from compound-10 only smaller changes occur together with increased PSA scores. Surprisingly, the final compound appeared to differ the most from its direct precursors. This compound also showed an increased MW and a reduced PSA score compared to its previous three variants.

The search for an MCL-1 inhibitor started several years ago from the identification of a first hit obtained through primary screening by NMR. The process illustrates the huge amount of work required to bring an initial hit to a final lead drug candidate, which included the efforts of multiple academic and industrial laboratories. The failure of crystallisation trials during the early stages of the project, plus the inherent flexibility of the MCL-1 BH3 binding groove, made NMR spectroscopy uniquely capable of driving the project forward. The S64315 compound is currently under evaluation in the clinical phase-1 trials, which provides hope for patients affected by a variety of cancer types.

## DISCUSSION AND CONCLUSION

We explored the development histories of the AZD-3839, ABT-199 and S64315 compounds (Jeppsson et al., 2012; Souers et al., 2013; Szálavik et al., 2019), from the primary screening to the final lead optimisation, focusing on their target interactions and the rationale behind their optimisations. These cases highlighted the underlying role of NMR techniques during all drug discovery phases and their impact throughout each stage.

With multiple compounds in clinical phases, NMR has demonstrated a key role in the process of fragment-based drug discovery (Singh et al., 2018) (**Figure 5A**). In 2016, ABT-199, commonly known as Venetoclax, was the first confirmed FDA-approved drug derived largely by NMR-FBDD (Erlanson, 2012; Souers et al., 2013; Hubbard, 2016). The development of other fragment-derived drugs, e.g., Vemurafenib (approved in 2011) (Flaherty et al., 2011), Erdafitinib (approved in 2019) (Perera et al., 2017; The ASCO Post, 2019), Pexidartinib (approved in 2020) (Benner et al., 2020) and Sotorasib (approved in 2021) (Ostrem et al., 2013), were driven by both NMR, X-ray crystallography and other techniques.

Interestingly, Vemurafenib, and Pexidartinib, both developed by Plexxikon, the phase-III drug Capivasertib (Addie et al., 2013) developed by Astra-Zeneca, and many other clinical candidate molecules, all include the identical 7-azaindole fragment (Irie and Sawa, 2018; Qhobosheane et al., 2020). This unique structural composition presents both crucial hydrogen-bond acceptor and donor groups making this small molecule able potentially to interact with over 90 different kinds of kinase active sites, which has been considered encapsulating the entire human Kinome (Irie and Sawa, 2018).

In spite of this distinctive case of a generally adaptable building block, over the years, different methodologies have been explicitly developed and proven crucial for enhancing the success rate in the drug discovery process (Pellecchia, 2010). The great flexibility and adaptability of NMR provides for qualitative and quantitative

insights at each point of the drug development process (Campos-Olivas, 2011; Pellecchia, 2010). From the case studies, it emerged that NMR is predominately used in the primary screening, also known as “hit-identification” (**Figure 1C**). The detection of weak binders, defined as small-molecules presenting transient interactions to targets or unfavourably high dissociation rates, is paramount in the early stages of FBDD (Wang et al., 2017). The identification of such compounds represents a clear advantage as it shows its natural ability to bind its target and succeed among cocktails of other sub-molar molecules. The crucial chemical nature of these molecules, i.e., their “core”, is often used as a scaffold and preserved during their evolution to final drugs (**Figures 3A, 4A**).

However, the process hit-identification by NMR has also shown a number of drawbacks (Davis and Erlanson, 2013). For example, the usage of only a single ligand-detected 1D technique for identifying binding fragments may prove to give erroneous results (Davis and Erlanson, 2013). Hence, the recommendation is to use at least two NMR techniques, such as STD and Water-LOGSY (Davis and Erlanson, 2013) in parallel.

An inevitable consequence of using weakly binding fragments is the so-called “non-specific” binding event (Mercier et al., 2009). A simple strategy for alleviating this issue was employed in the development of AZD-3839. By recording competition experiments using a potent known, but clinically unsuitable, ligand for the same target non-specific binding molecules were identified and excluded from further development.

Furthermore, from the case studies it appeared that the full analytical power of 1D NMR spectroscopy was often not exploited. Instead, the NMR data appear to be used solely as a binary result, probably also due to a lack of proper computational- and data-analysis tools available at the time the research was conducted (Mureddu et al., 2020). The data obtained from Water-LOGSY and STD experiments offer further quantitative information (Meyer et al., 2004; Cala and Krimm, 2015). The SAR by Water-LOGSY, for example, suggests a scoring factor to identify the most exposed portion of the molecule. Assessing all data that can be derived from 1D and multi-dimensional NMR experiments can eventually provide insights into the ligand binding pose (Raingeval et al., 2019).

Upon validation of fragment hits, the next stage is usually the exploration of potential binding sites on the target. Chemical shift perturbation, or CSP, is so far the most popular NMR technique used for this task (Williamson, 2013; Mureddu and Vuister, 2019). CSP has been widely used as a standard for molecules that progressed into clinical phases. Nevertheless, a CSP analysis potentially might drive researchers in wrong directions, and final conclusions should not be based on this approach alone. Common errors observed in practice include subjective judgments or misinterpretation of shifting peaks, especially in crowded regions of the spectra, leading to overestimating of the CSP effects. Furthermore, in some instances, compounds have been shown to change the pH of the solution, resulting in false positive CSPs (Davis and Erlanson, 2013). By performing appropriate control experiments, such errors might hopefully be avoided.

NMR has often been associated with a requirement for daunting and time-consuming data analysis. There may be a multitude of other undescribed factors, but NMR's lack of modern, more practical, quicker, and unbiased methods, alongside with automated data analysis routines has comparatively slowed the entire NMR-FBDD process, prompting a need for improvements in all these aspects (Mureddu et al., 2020). To this end, we have recently developed a versatile and flexible data-analysis program called AnalysisScreen (Mureddu et al., 2020), part of the CcpNmr Analysis package (Skinner et al., 2016), which presents dedicated provision and capabilities for the analysis of all forms of NMR data used in the NMR-FBDD process, including an integrated CSP analysis (Mureddu and Vuister, 2019).

An additional key step before the expensive and laborious optimisation processes of candidate compounds begins is a minute hit chemical assessment; for example, by employing Pan-Assay Interference Compounds (PAINS) protocols (Baell and Nissink, 2018). Applying these filters to hits or families of hits can help identify erroneous binders. PAINS-flagged molecules can exert photo-reactivity, redox-activity and other undesirable chemical phenomena, which can lead to non-specific or unwanted biological activities. Unfortunately, it is also wise not to rely solely on PAINS filters. A recent analysis showed that many PAINS-flagged molecules had been wrongly evaluated by the applied filters, either as false negatives or false positives (Capuzzi et al., 2017).

Ayotte *et al.* proposed the use of a NMR CPMG series to detect potential aggregation of compounds in mixtures (Ayotte et al., 2019), another potentially complicating effect. In addition, the authors also pointed out that aggregation can be solvent-dependent, and thus a minor adjustment of the sample composition might improve or worsen the outcomes of screening experiments (Ayotte et al., 2019).

The methodological and data-analysis improvements are fundamental to design the molecules from the early stages and onward. From the analysis of the three cases, it emerged that molecular optimisations are mainly guided by multiple manual moiety substitutions, followed by their chemical synthesis and re-evaluation. Although this might generate potential leads, a meticulous use of computational approaches could likely have accelerated this process further. Molecular docking studies aided the final lead generation of the ABT-199 compound; however, combining NMR and molecular docking can still introduce mistakes and whereas docking alone can also not be fully trusted (Chen, 2015). Even the newest scoring functions, implemented using artificial intelligence (AI), likely still present significant limitations (Gabel et al., 2014), as incomplete or erroneous classification of existing experimental data can compromise and bias their validity (Uçar et al., 2020).

Despite these setbacks, we firmly expect that newer chemoinformatics and AI algorithms, together with improved high-performance computing resources for rapid parallel in-silico drug screening using Molecular Dynamic (MD) simulations,

will replace the current optimisation stages of the FBDD protocols. Several algorithms can already complement the experimental data validation (Bienstock, 2011). In addition to experimental methods, such as X-ray crystallography, NMR and electron microscopy, AI-driven structure elucidation by tools such as AlphaFold-2 (Jumper et al., 2021) and RoseTTA fold (Baek et al., 2021) can also provide the 3D molecular information required for the fragment-optimization stage. Ideally, the computational approaches for fragment optimization should simultaneously take the interaction, the experimental data such as derived by NMR, as well as biological aspects into consideration. For example, the robustness of a candidate compound or potential drug resistance could also be assessed early in the development process, thus integrating multiple fields of drug-discovery and pharmacology in a holistic way that scientists alone could not have attained (Dey and Cafilisch, 2008).

By analysing the molecular characteristics of the various compounds, we have identified correlations in the development patterns among the three cases studied. By comparing the Tanimoto coefficients (Bajusz et al., 2015) for the three final compounds (**Figures 2C, 3C, 4C**), we speculate that expanding the molecule (via growing/linking methods) did not improve the binding affinity. This suggests that just covering the conformational molecular interaction space does not necessarily lead to higher affinity drugs. **Figure 5B** displays the normalised Tanimoto scores for the three different fragment-to-drug evolutions. Clearly, the development of AZD-3839 and ABT-199 display a highly similar pattern, which is dissimilar from S64315. An obvious conclusion could be that they simply differ in their optimisation protocols. However, it could equally suggest that the optimization of S64315 is not yet complete and multiple additional changes might occur before “converging” to the ultimate drug.

In conclusion, although only a few drugs approved by FDA have an NMR fragment-based trackable history that is easily accessible from public domain data, an analysis of relevant publications over the last 6 years shows an impressive number of journal articles reporting on discoveries of molecules in the pre-clinical stages in which NMR had a crucial role (**Figure 1A**). One of these molecules, Asciminib (Schoepfer et al., 2018), has only just recently (2021) been granted a FDA Breakthrough Therapy designation (Furnari et al., 2021), and it constitutes yet another marvellous example of the crucial role of NMR experiments in every stage of its development (Schoepfer et al., 2018).

In all, the overall methodological advances in the steps leading from initial hit to candidate-drug provide great hope that, compared to the time currently taken, new potent and selective drugs will soon become much more rapidly available.

## DATA AVAILABILITY STATEMENT

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



accession number(s) can be found below: <https://github.com/VuisterLab/scripts.git>.

## AUTHOR CONTRIBUTIONS

LM researched the literature, wrote the Python-code scripts, created tables and figures. LM and GV analysed the data. LM and GV wrote the manuscript.

## FUNDING

LM acknowledges his stipend provided by MRC-IMPACT PhD programme (Grant MR/NO13913/1) and GV acknowledges

funding of the CCPN project by MRC (Grants MR/L000555/1 and MR/P00038X/1).

## ACKNOWLEDGMENTS

We thank Dr Victoria A. Higman for proofreading the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.834453/full#supplementary-material>

## REFERENCES

- Addie, M., Ballard, P., Buttar, D., Crafter, C., Currie, G., Davies, B. R., et al. (2013). Discovery of 4-Amino-N-[(1s)-1-(4-Chlorophenyl)-3-Hydroxypropyl]-1-(7h-Pyrrolo[2,3-D]pyrimidin-4-Yl)piperidine-4-Carboxamide (AZD5363), an Orally Bioavailable, Potent Inhibitor of Akt Kinases. *J. Med. Chem.* 56, 2059–2073. doi:10.1021/jm301762v
- Albershardt, T. C., Salerni, B. L., Soderquist, R. S., Bates, D. J. P., Pletnev, A. A., Kisselev, A. F., et al. (2011). Multiple BH3 Mimetics Antagonize Antiapoptotic MCL1 Protein by Inducing the Endoplasmic Reticulum Stress Response and Up-Regulating BH3-Only Protein NOXA. *J. Biol. Chem.* 286, 24882–24895. doi:10.1074/jbc.m111.255828
- Antanasijevic, A., Ramirez, B., and Caffrey, M. (2014). Comparison of the Sensitivities of WaterLOGSY and Saturation Transfer Difference NMR Experiments. *J. Biomol. NMR* 60, 37–44. doi:10.1007/s10858-014-9848-9
- Author Anonyms (2007). A Decade of Drug-Likeness. *Nat. Rev. Drug Discov.* 6, Ayotte, Y., Marando, V. M., Vaillancourt, L., Bouchard, P., Heffron, G., Coote, P. W., et al. (2019). Exposing Small-Molecule Nanoentities by a Nuclear Magnetic Resonance Relaxation Assay. *J. Med. Chem.* 62, 7885–7896. doi:10.1021/acs.jmedchem.9b00653
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* 373, 871–876. doi:10.1126/science.abj8754
- Baell, J. B., and Nissink, J. W. M. (2018). Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017-Utility and Limitations. *ACS Chem. Biol.* 13, 36–44. doi:10.1021/acscchembio.7b00903
- Bajusz, D., Rácz, A., and Héberger, K. (2015). Why Is Tanimoto index an Appropriate Choice for Fingerprint-Based Similarity Calculations? *J. Cheminform.* 7, 20. doi:10.1186/s13321-015-0069-3
- Benner, B., Good, L., Quiroga, D., Schultz, T. E., Kassem, M., Carson, W. E., et al. (2020). Pexidartinib, a Novel Small Molecule Csf-1r Inhibitor in Use for Tenosynovial Giant Cell Tumor: A Systematic Review of Pre-clinical and Clinical Development. *Dddt* 14, 1693–1704. doi:10.2147/DDDT.S253232
- Berg, S., Kolmodin, K., Karlström, S., von Berg, S., Söderman, P., Holenz, J., et al. (2012). Design and Synthesis of  $\beta$ -site Amyloid Precursor Protein Cleaving Enzyme (BACE1) Inhibitors with *In Vivo* Brain Reduction of  $\beta$ -amyloid Peptides. *J. Med. Chem.* 55, 9346–9361.
- Bienstock, R. J. (2011). Overview: Fragment-Based Drug Design. *Libr. Des. Search Methods Appl. Fragm. Drug Des.* 1076, 1–26. doi:10.1021/bk-2011-1076.ch001
- Birkinshaw, R. W., Gong, J. N., Luo, C. S., Lio, D., White, C. A., Anderson, M. A., et al. (2019). Structures of BCL-2 in Complex with Venetoclax Reveal the Molecular Basis of Resistance Mutations. *Nat. Commun.* 10, 2385. doi:10.1038/s41467-019-10363-1
- Cala, O., and Krimm, I. (2015). Ligand-orientation Based Fragment Selection in STD NMR Screening. *J. Med. Chem.* 58, 8739–8742. doi:10.1021/acs.jmedchem.5b01114
- Campos-Olivas, R. (2011). NMR Screening and Hit Validation in Fragment Based Drug Discovery. *Ctmc* 11, 43–67. doi:10.2174/156802611793611887
- Capuzzi, S. J., Muratov, E. N., and Tropsha, A. (2017). Phantom PAINS: Problems with the Utility of Alerts for Pan-Assay Interference Compounds. *J. Chem. Inf. Model.* 57, 417–427. doi:10.1021/acs.jcim.6b00465
- Chen, Y.-C. (2015). Beware of Docking! *Trends Pharmacol. Sci.* 36, 78–95. doi:10.1016/j.tips.2014.12.001
- Choudhary, G. S., Al-harbi, S., Mazumder, S., Hill, B. T., Smith, M. R., Bodo, J., et al. (2015). MCL-1 and BCL-xL-dependent Resistance to the BCL-2 Inhibitor ABT-199 Can Be Overcome by Preventing PI3K/AKT/mTOR Activation in Lymphoid Malignancies. *Cell Death Dis* 6, e1593. doi:10.1038/cddis.2014.525
- Ciulli, A., and Abell, C. (2007). Fragment-based Approaches to Enzyme Inhibition. *Curr. Opin. Biotechnol.* 18, 489–496. doi:10.1016/j.copbio.2007.09.003
- Congreve, M., Carr, R., Murray, C., and Jhoti, H. (2003). A 'Rule of Three' for Fragment-Based lead Discovery? *Drug Discov. Today* 8, 876–877. doi:10.1016/s1359-6446(03)02831-9
- 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
- Davis, B. J., and Erlanson, D. A. (2013). Learning from Our Mistakes: The 'unknown Knowns' in Fragment Screening. *Bioorg. Med. Chem. Lett.* 23, 2844–2852. doi:10.1016/j.bmcl.2013.03.028
- Deeks, E. D. (2016). Venetoclax: First Global Approval. *Drugs* 76, 979–987. doi:10.1007/s40265-016-0596-x
- Dey, F., and Cafisch, A. (2008). Fragment-based De Novo Ligand Design by Multiobjective Evolutionary Optimization. *J. Chem. Inf. Model.* 48, 679–690. doi:10.1021/ci700424b
- Edwards, P. D., Albert, J. S., Sylvester, M., Aharony, D., Andisik, D., Callaghan, O., et al. (2007). Application of Fragment-Based Lead Generation to the Discovery of Novel, Cyclic Amidine  $\beta$ -Secretase Inhibitors with Nanomolar Potency, Cellular Activity, and High Ligand Efficiency. *J. Med. Chem.* 50, 5912–5925. doi:10.1021/jm070829p
- Erlanson, D. A. (2012). Introduction to Fragment-Based Drug Discovery. *Top. Curr. Chem.* 317, 1–32. doi:10.1007/128\_2011\_180
- Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W., and Jhoti, H. (2016). Twenty Years on: The Impact of Fragments on Drug Discovery. *Nat. Rev. Drug Discov.* 15, 605–619. doi:10.1038/nrd.2016.109
- Erlanson, D. A. (2018). *Fragments in the Clinic*. 2018 edition. <http://practicalfragments.blogspot.com/2018/10/fragments-in-clinic-2018-edition.html>.
- Erlanson, Daniel A., and Janke, W. (2016). "BACE Inhibitors," in *Fragment-based Drug Discovery, Lessons and Outlook*, 333–338.
- Erlanson, D. A. (2015). Personal Essay: Fragments in the Blogosphere. *RSC Drug Discov. Ser.*, P019–P030. doi:10.1039/9781782620938-fp019
- Fesik, S. W., Shuker, S. B., Hajduk, P. J., and Meadows, R. P. (1997). SAR by NMR: an NMR-Based Approach for Drug Discovery. *Protein Eng.* 10, 73.
- Flaherty, K. T., Yasothan, U., and Kirkpatrick, P. (2011). Vemurafenib. *Nat. Rev. Drug Discov.* 10, 811–812. doi:10.1038/nrd3579

- Furnari, F. R., Treskow, A., and Masow, J. (2021). *Novartis Receives FDA Breakthrough Therapy Designations for Investigational STAMP Inhibitor Asciminib (ABL001) in Chronic Myeloid Leukemia* Novartis Global Media Relations and Communications. <https://www.novartis.com/news/media-releases>.
- Gabel, J., Desaphy, J., and Rognan, D. (2014). Beware of Machine Learning-Based Scoring Functions-On the Danger of Developing Black Boxes. *J. Chem. Inf. Model.* 54, 2807–2815. doi:10.1021/ci500406k
- Geschwindner, S., Olsson, L.-L., Albert, J. S., Deinum, J., Edwards, P. D., de Beer, T., et al. (2007). Discovery of a Novel Warhead against  $\beta$ -Secretase through Fragment-Based Lead Generation. *J. Med. Chem.* 50, 5903–5911. doi:10.1021/jm070825k
- Ghosh, A. K., Brindisi, M., and Tang, J. (2012). Developing  $\beta$ -secretase Inhibitors for Treatment of Alzheimer's Disease. *J. Neurochem.* 120, 71–83. doi:10.1111/j.1471-4159.2011.07476.x
- Goyal, M., Dhanjal, J. K., Goyal, S., Tyagi, C., Hamid, R., and Grover, A. (2014). Development of Dual Inhibitors against Alzheimer's Disease Using Fragment-Based QSAR and Molecular Docking. *Biomed. Res. Int.* 2014, 1–12. doi:10.1155/2014/979606
- Hajduk, P. J., Olejniczak, E. T., and Fesik, S. W. (1997). One-dimensional Relaxation- and Diffusion-Edited NMR Methods for Screening Compounds that Bind to Macromolecules. *J. Am. Chem. Soc.* 119, 12257–12261. doi:10.1021/ja9715962
- Hubbard, R. E. (2016). *Fragment-based Drug Discovery: Lessons and Outlook*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. doi:10.1002/9783527683604
- Hughes, T. V., Baldwin, I., and Churcher, I. (2011). Fragment-based Drug Discovery - from Hit Discovery to FDA Approval: Lessons Learned and Future Challenges. *Int. Drug Discov.* 6.
- Hung, S.-Y., and Fu, W.-M. (2017). Drug Candidates in Clinical Trials for Alzheimer's Disease. *J. Biomed. Sci.* 24, 47. doi:10.1186/s12929-017-0355-7
- Irie, T., and Sawa, M. (2018). 7-azaindole: A Versatile Scaffold for Developing Kinase Inhibitors. *Chem. Pharm. Bull.* 66, 29–36. doi:10.1248/cpb.c17-00380
- Jeppsson, F., Eketjäll, S., Jansson, J., Karlström, S., Gustavsson, S., Olsson, L.-L., et al. (2012). Discovery of AZD3839, a Potent and Selective BACE1 Inhibitor Clinical Candidate for the Treatment of Alzheimer Disease. *J. Biol. Chem.* 287, 41245–41257. doi:10.1074/jbc.m112.409110
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2
- Kang, C. (2019). Applications of In-Cell NMR in Structural Biology and Drug Discovery. *Ijms* 20, 139. doi:10.3390/ijms20010139
- Kelekar, A., and Thompson, C. B. (1998). Bcl-2-family Proteins: The Role of the BH3 Domain in Apoptosis. *Trends Cell Biol.* 8, 324–330. doi:10.1016/s0962-8924(98)01321-x
- Klon, A. E. (2015). *Fragment-Based Methods in Drug Discovery*. Springer.
- Lee, E. F., Czabotar, P. E., Smith, B. J., Deshayes, K., Zobel, K., Colman, P. M., et al. (2007). Crystal Structure of ABT-737 Complexed with Bcl-xL: Implications for Selectivity of Antagonists of the Bcl-2 Family. *Cell Death Differ* 14, 1711–1713. doi:10.1038/sj.cdd.4402178
- Leone, V., Marinelli, F., Carloni, P., and Parrinello, M. (2010). Targeting Biomolecular Flexibility with Metadynamics. *Curr. Opin. Struct. Biol.* 20, 148–154. doi:10.1016/j.sbi.2010.01.011
- Letai, A., Bassik, M. C., Walensky, L. D., Sorcinelli, M. D., Weiler, S., and Korsmeyer, S. J. (2002). Distinct BH3 Domains Either Sensitize or Activate Mitochondrial Apoptosis, Serving as Prototype Cancer Therapeutics. *Cancer Cell* 2, 183–192. doi:10.1016/s1535-6108(02)00127-7
- Maragno, A. L., Mistry, P., Kotschy, A., Szlavik, Z., Murray, J., Davidson, J., et al. (2019). “Abstract 4482: S64315 (MIK665) Is a Potent and Selective Mcl1 Inhibitor with strong Antitumor Activity across a Diverse Range of Hematologic Tumor Models,” in Proceedings: AACR Annual Meeting 2019, March 29–April 3, 2019, Atlanta, GA. doi:10.1158/1538-7445.am2019-4482
- Mari, S., Invernizzi, C., Spitaleri, A., Alberici, L., Ghitti, M., Bordignon, C., et al. (2010). 2D TR-NOESY Experiments Interrogate and Rank Ligand-Receptor Interactions in Living Human Cancer Cells. *Angew. Chem. Int. Ed.* 49, 1071–1074. doi:10.1002/anie.200905941
- Mercier, K. A., and Powers, R. (2005). Determining the Optimal Size of Small Molecule Mixtures for High Throughput NMR Screening. *J. Biomol. NMR* 31, 243–258. doi:10.1007/s10858-005-0948-4
- Mercier, K., Shortridge, M., and Powers, R. (2009). A Multi-step NMR Screen for the Identification and Evaluation of Chemical Leads for Drug Discovery. *Chts* 12, 285–295. doi:10.2174/13862070978581738
- Meyer, B., Klein, J., Mayer, M., Meinecke, R., Möller, H., Neff, A., et al. (2004). Saturation Transfer Difference NMR Spectroscopy for Identifying Ligand Epitopes and Binding Specificities. *Ernst Schering Res. Found. Workshop* 44, 149–167. doi:10.1007/978-3-662-05397-3\_9
- Mureddu, L. G., Ragan, T. J., Brooksbank, E. J., and Vuister, G. W. (2020). CcpNmr AnalysisScreen, a New Software Programme with Dedicated Automated Analysis Tools for Fragment-Based Drug Discovery by NMR. *J. Biomol. NMR* 74, 565–577. doi:10.1007/s10858-020-00321-1
- Mureddu, L., and Vuister, G. W. (2019). Simple High-Resolution NMR Spectroscopy as a Tool in Molecular Biology. *FEBS J.* 286, 2035–2042. doi:10.1111/febs.14771
- Murray, C. W., Callaghan, O., Chessari, G., Cleasby, A., Congreve, M., Frederickson, M., et al. (2007). Application of Fragment Screening by X-ray Crystallography to  $\beta$ -Secretase. *J. Med. Chem.* 50, 1116–1123. doi:10.1021/jm0611962
- Oltersdorf, T., Elmore, S. W., Shoemaker, A. R., Armstrong, R. C., Augeri, D. J., Belli, B. A., et al. (2005). An Inhibitor of Bcl-2 Family Proteins Induces Regression of Solid Tumours. *Nature* 435, 677–681. doi:10.1038/nature03579
- Ortega-Roldan, J. L., Jensen, M. R., Brutscher, B., Azuaga, A. I., Blackledge, M., and van Nuland, N. A. J. (2009). Accurate Characterization of Weak Macromolecular Interactions by Titration of NMR Residual Dipolar Couplings: Application to the CD2AP SH3-C-ubiquitin Complex. *Nucleic Acids Res.* 37, e70. doi:10.1093/nar/gkp211
- Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A., and Shokat, K. M. (2013). K-Ras(G12C) Inhibitors Allosterically Control GTP Affinity and Effector Interactions. *Nature* 503, 548–551. doi:10.1038/nature12796
- Pellecchia, M. (2010). NMR Spectroscopy in Fragment Based Drug Design. *Top. Med. Chem.* 5, 125–139.
- Perera, T. P. S., Jovcheva, E., Mevellec, L., Vialard, J., De Lange, D., Verhulst, T., et al. (2017). Discovery and Pharmacological Characterization of JNJ-42756493 (Erdafitinib), a Functionally Selective Small-Molecule FGFR Family Inhibitor. *Mol. Cancer Ther.* 16, 1010–1020. doi:10.1158/1535-7163.mct-16-0589
- Petros, A. M., Dinges, J., Augeri, D. J., Baumeister, S. A., Betebenner, D. A., Bures, M. G., et al. (2006). Discovery of a Potent Inhibitor of the Antiapoptotic Protein Bcl-xL from NMR and Parallel Synthesis. *J. Med. Chem.* 49, 656–663. doi:10.1021/jm0507532
- Petros, A. M., Huth, J. R., Oost, T., Park, C.-M., Ding, H., Wang, X., et al. (2010). Discovery of a Potent and Selective Bcl-2 Inhibitor Using SAR by NMR. *Bioorg. Med. Chem. Lett.* 20, 6587–6591. doi:10.1016/j.bmcl.2010.09.033
- Primikyri, A., Sayyad, N., Quilici, G., Vrettos, E. I., Lim, K., Chi, S. W., et al. (2018). Probing the Interaction of a Quercetin Bioconjugate with Bcl-2 in Living Human Cancer Cells with In-Cell NMR Spectroscopy. *FEBS Lett.* 592, 3367–3379. doi:10.1002/1873-3468.13250
- Qhobosheane, M. A., Legoabe, L. J., Josselin, B., Bach, S., Ruchaud, S., Petzer, J. P., et al. (2020). Synthesis and Evaluation of 7-azaindole Derivatives Bearing Benzocycloalkane Motifs as Protein Kinase Inhibitors. *Bioorg. Med. Chem.* 28, 115468. doi:10.1016/j.bmc.2020.115468
- Raingeval, C., Cala, O., Brion, B., Le Borgne, M., Hubbard, R. E., and Krimm, I. (2019). 1D NMR WaterLOGSY as an Efficient Method for Fragment-Based lead Discovery. *J. Enzyme Inhib. Med. Chem.* 34, 1218–1225. doi:10.1080/14756366.2019.1636235
- Rees, D. (2016). Fragment-based Drug Discovery: Lessons and Outlook. *ChemMedChem* 11, 1667. doi:10.1002/cmdc.201600256
- Reymond, J.-L., Van Deursen, R., Blum, L. C., and Ruddigkeit, L. (2010). Chemical Space as a Source for New Drugs. *Med. Chem. Commun.* 1, 30–38. doi:10.1039/c0md00020e
- Schoepfer, J., Jahnke, W., Berellini, G., Buonamici, S., Cotesta, S., Cowan-Jacob, S. W., et al. (2018). Discovery of Asciminib (ABL001), an Allosteric Inhibitor of the Tyrosine Kinase Activity of BCR-ABL1. *J. Med. Chem.* 61, 8120–8135. doi:10.1021/acs.jmedchem.8b01040
- Shuker, S. B., Hajduk, P. J., Meadows, R. P., and Fesik, S. W. (1996). Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* 274, 1531–1534. doi:10.1126/science.274.5292.1531

- Singh, M., Tam, B., and Akabayov, B. (2018). NMR-fragment Based Virtual Screening: A Brief Overview. *Molecules* 23, 233–260. doi:10.3390/molecules23020233
- Skinner, S. P., Fogh, R. H., Boucher, W., Ragan, T. J., Mureddu, L. G., and Vuister, G. W. (2016). CcpNmr AnalysisAssign: a Flexible Platform for Integrated NMR Analysis. *J. Biomol. NMR* 66, 111–124. doi:10.1007/s10858-016-0060-y
- Souers, A. J., Levenson, J. D., Boghaert, E. R., Ackler, S. L., Catron, N. D., Chen, J., et al. (2013). ABT-199, a Potent and Selective BCL-2 Inhibitor, Achieves Antitumor Activity while Sparing Platelets. *Nat. Med.* 19, 202–208. doi:10.1038/nm.3048
- S. Soderquist, R., and Eastman, A. (2016). BCL2 Inhibitors as Anticancer Drugs: A Plethora of Misleading BH3 Mimetics. *Mol. Cancer Ther.* 15, 2011–2017. doi:10.1158/1535-7163.mct-16-0031
- Szlávik, Z., Ondi, L., Csekei, M., Pascal, A., Szabo, Z. B., Radics, G., et al. (2019). Structure-guided Discovery of a Selective Mcl-1 Inhibitor with Cellular Activity. *J. Med. Chem.* 62, 6913–6924.
- Teague, S. J. (2003). Implications of Protein Flexibility for Drug Discovery. *Nat. Rev. Drug Discov.* 2, 527–541. doi:10.1038/nrd1129
- The ASCO Post (2019). FDA grants Accelerated Approval to Erdafitinib for Metastatic Urothelial Carcinoma. *Case Med. Res.* doi:10.31525/fda1-ucm635910.htm
- Uçar, M. K., Nour, M., Sindi, H., and Polat, K. (2020). The Effect of Training and Testing Process on Machine Learning in Biomedical Datasets. *Math. Probl. Eng.* 2020, 1–17.
- Vandenberg, C. J., and Cory, S. (2013). ABT-199, a New Bcl-2-specific BH3 Mimetic, Has *In Vivo* Efficacy against Aggressive Myc-Driven Mouse Lymphomas without Provoking Thrombocytopenia. *Blood* 121, 2285–2288. doi:10.1182/blood-2013-01-475855
- Vassar, R., Bennett, B. D., Babu-Khan, S., Kahn, S., Mendiaz, E. A., Denis, P., et al. (1999).  $\beta$ -Secretase Cleavage of Alzheimer's Amyloid Precursor Protein by the Transmembrane Aspartic Protease BACE. *Science* 286, 735–741. doi:10.1126/science.286.5440.735
- Vassar, R., Kovacs, D. M., Yan, R., and Wong, P. C. (2009). The  $\gamma$ -Secretase Enzyme BACE in Health and Alzheimer's Disease: Regulation, Cell Biology, Function, and Therapeutic Potential. *J. Neurosci.* 29, 12787–12794. doi:10.1523/jneurosci.3657-09.2009
- Venugopal, C., Demos, C., Jagannatha Rao, K., Pappolla, M., and Sambamurti, K. (2008). Beta-Secretase: Structure, Function, and Evolution. *Cnsnddt* 7, 278–294. doi:10.2174/187152708784936626
- Viegas, A., Manso, J., Nobrega, F. L., and Cabrita, E. J. (2011). Saturation-transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* 88, 990–994. doi:10.1021/ed101169t
- Vinogradova, O., and Qin, J. (2012). NMR as a Unique Tool in Assessment and Complex Determination of Weak Protein-Protein Interactions. *Top. Curr. Chem.* 326, 35–45. doi:10.1007/128\_2011\_216
- Wang, J., Guo, Z., Fu, Y., Wu, Z., Huang, C., Zheng, C., et al. (2017). Weak-binding Molecules Are Not Drugs?-Toward a Systematic Strategy for Finding Effective Weak-Binding Drugs. *Brief. Bioinform.* 18, 321–332. doi:10.1093/bib/bbw018
- Williamson, M. P. (2013). Using Chemical Shift Perturbation to Characterise Ligand Binding. *Prog. Nucl. Magn. Reson. Spectrosc.* 73, 1–16. doi:10.1016/j.pnmrs.2013.02.001
- Zheng, H., Handing, K. B., Zimmerman, M. D., Shabalin, I. G., Almo, S. C., and Minor, W. (2015). X-ray Crystallography over the Past Decade for Novel Drug Discovery - where Are We Heading Next? *Expert Opin. Drug Discov.* 10, 975–989. doi:10.1517/17460441.2015.1061991

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

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

Copyright © 2022 Mureddu and Vuister. 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.



# Using Structure-guided Fragment-Based Drug Discovery to Target *Pseudomonas aeruginosa* Infections in Cystic Fibrosis

Sheikh Mohammed Arif<sup>1</sup>, R. Andres Floto<sup>2,3</sup> and Tom L. Blundell<sup>1\*</sup>

<sup>1</sup>Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom, <sup>2</sup>Molecular Immunity Unit, Department of Medicine University of Cambridge, MRC-Laboratory of Molecular Biology, Cambridge, United Kingdom, <sup>3</sup>Cambridge Centre for Lung Infection, Royal Papworth Hospital, Cambridge, United Kingdom

## OPEN ACCESS

### Edited by:

Pietro Roversi,  
Italian National Research Council, Italy

### Reviewed by:

Tzvia Springer,  
Concordia University (Wisconsin),  
United States  
Fang Bai,  
Nankai University, China

### \*Correspondence:

Tom L. Blundell  
tlb20@cam.ac.uk

### Specialty section:

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

**Received:** 18 January 2022

**Accepted:** 23 February 2022

**Published:** 30 March 2022

### Citation:

Arif SM, Floto RA and Blundell TL  
(2022) Using Structure-guided  
Fragment-Based Drug Discovery to  
Target *Pseudomonas aeruginosa*  
Infections in Cystic Fibrosis.  
Front. Mol. Biosci. 9:857000.  
doi: 10.3389/fmolb.2022.857000

Cystic fibrosis (CF) is progressive genetic disease that predisposes lungs and other organs to multiple long-lasting microbial infections. *Pseudomonas aeruginosa* is the most prevalent and deadly pathogen among these microbes. Lung function of CF patients worsens following chronic infections with *P. aeruginosa* and is associated with increased mortality and morbidity. Emergence of multidrug-resistant, extensively drug-resistant and pandrug-resistant strains of *P. aeruginosa* due to intrinsic and adaptive antibiotic resistance mechanisms has failed the current anti-pseudomonal antibiotics. Hence new antibacterials are urgently needed to treat *P. aeruginosa* infections. Structure-guided fragment-based drug discovery (FBDD) is a powerful approach in the field of drug development that has succeeded in delivering six FDA approved drugs over the past 20 years targeting a variety of biological molecules. However, FBDD has not been widely used in the development of anti-pseudomonal molecules. In this review, we first give a brief overview of our structure-guided FBDD pipeline and then give a detailed account of FBDD campaigns to combat *P. aeruginosa* infections by developing small molecules having either bactericidal or anti-virulence properties. We conclude with a brief overview of the FBDD efforts in our lab at the University of Cambridge towards targeting *P. aeruginosa* infections.

**Keywords:** FBDD, fragment-based drug discovery, antibiotics, cystic fibrosis, *Pseudomonas*, anti-virulence

## INTRODUCTION

Cystic fibrosis (CF) is a life-shortening autosomal-recessive Mendelian disease affecting approximately 100,000 people worldwide (Cystic Fibrosis Foundation, 2017 Annual Data Report, 2021; United Kingdom; Cystic Fibrosis Registry 2020 Annual Data Report, 2021). In the 1950s, the majority of the CF patients did not live beyond infancy (Davis, 2006; Elborn, 2016). Four decades later, the life expectancy had improved to 31 years (Cystic Fibrosis Foundation. 2017. Cystic Fibrosis Foundation patient registry. 2016 annual data report. Cystic Fibrosis Foundation, Bethesda, MD.). At present, with further improvement in the diagnosis and treatment, the median survival age of CF patients has increased to 50.6 years in United Kingdom (UK Cystic Fibrosis Registry 2020 Annual Data Report, 2021) and 59 years in United states (UK Cystic Fibrosis Registry 2020 Annual Data Report, 2021). The majority of morbidity and mortality in CF is caused by chronic bacterial lung



infections (Elborn, 2016; Molina and Hunt, 2017). CF is caused by bi-allelic deleterious mutations in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) protein, leading to defects in chloride and bicarbonate ion transport across epithelial surfaces, and consequently the production of thick secretions that disrupt mucociliary clearance in the lungs, predisposing to chronic bacterial infections, and progressive inflammatory lung disease.

The most important bacterial infection in CF is *P. aeruginosa*, a Gram-negative bacterium that, in addition to causing opportunistic and hospital-acquired infections, can cause chronic respiratory infections in individuals with underlying inflammatory lung disease, including CF. Approximately 40% of the adult CF population in the United Kingdom is currently chronically infected with *P. aeruginosa* (2019 CF registry), leading to increased mortality and morbidity (Flume et al., 2007; Crull et al., 2018) associated with the emergence of multi-drug resistant organisms and antibiotic failure.

## Current Treatment Regimens for *P. aeruginosa* Infections and Need for Novel Therapeutics

At present, nine categories of anti-pseudomonal antibiotics are used to treat *P. aeruginosa* infections including penicillin- $\beta$ -lactamase combinations (piperacillin-tazobactam and ticarcillin-clavulanate), cephalosporins (Ceftazidime, Cefepime, Cefoperazone and Cefiderocol), a monobactam (aztreonam), fluoroquinolones (Ciprofloxacin, Levofloxacin, Prulifloxacin, Delafloxacin and Finafloxacin), a phosphonic acid derivative (Fosfomycin), carbapenems (Doripenem, Imipenem/cilastatin and Meropenem), novel  $\beta$ -lactams with  $\beta$ -lactamase inhibitors (ceftazidime-avibactam, ceftolozane/tazobactam, Imipenem/cilastatin-relebactam, Meropenem-vaborbactam), aminoglycosides (Tobramycin, Gentamicin, Amikacin and Plazomicin) and polymyxins (Colistin and PolymyxinB) (Ibrahim et al., 2020).

*P. aeruginosa* is equipped with a high level of intrinsic antibiotic resistance owing to restricted outer membrane permeability, efflux systems that pump antibiotics out of the cell and production of antibiotic-inactivating enzymes such as  $\beta$ -lactamases. *P. aeruginosa* can also readily acquire antibiotic resistance through mutations and acquisition of resistance plasmids (Potron et al., 2015; Pang et al., 2019), leading to the emergence of multidrug-resistant (MDR), extensively drug-resistant (XDR) and pandrug-resistant (PDR) strains of *P. aeruginosa* (El Zowalaty et al., 2015; Bassetti et al., 2018). In addition, *P. aeruginosa* possesses adaptive resistance mechanisms against antibiotics, including biofilm-mediated resistance and the formation of multidrug-tolerant persister cells which further limit the effectiveness of current antibiotic treatments (Pang et al., 2019). As a consequence this pathogen is listed in the “critical” category of antibiotic-resistant “priority pathogens” published recently by WHO (WHO, 2017) (Shrivastava et al., 2018). Hence, there is an urgent need for new antibiotics along with the discovery and development of novel potential therapeutic strategies such as quorum sensing inhibition, lectin inhibition,

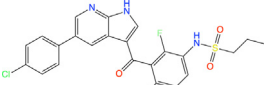
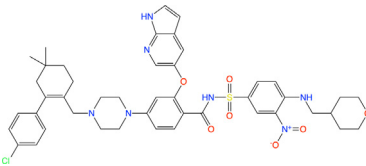
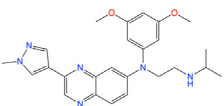
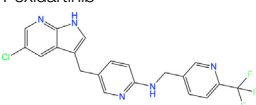
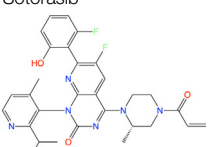
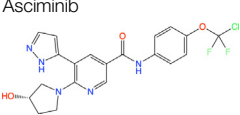
iron chelation, phage therapy, vaccine strategy, nanoparticles, antimicrobial peptides and electrochemical scaffolds, which present new avenues against *P. aeruginosa* infections (Pang et al., 2019).

## Fragment-Based Drug Discovery Approach and Work Flow

Structure-guided fragment-based drug discovery (FBDD) is a powerful approach, now widely used both in academia and industries to produce novel high-quality drug-like molecules (Blundell et al., 2002; Murray and Rees, 2009; Scott et al., 2012; Mashalidis et al., 2013; Erlanson et al., 2016). This approach involves screening a library consisting of small “fragments” (MW < 300 Da) of drug-like molecules against a defined target protein, using various biophysical, biochemical and structural biology methods. Compared with the traditional high-throughput screens of drug-sized molecules, the binding and subsequent growth of low molecular weight fragments to drug-like compounds allows a more extensive exploration of chemical space even when using small libraries and can lead to superior molecules. Although fragments are usually weak binders, they can bind to hotspots that allow well-defined interactions with the target protein. These can be entropically favourable due to displacement of previously organised bound water molecules (Radoux et al., 2016). The three-dimensional binding mode of these fragments is explored by determining the structure of the fragment-bound target using X-ray crystallography, NMR spectroscopy or cryoelectron microscopy, and guided by these structures the fragments are chemically “grown” or “linked” to form a larger molecule with higher affinity and drug-like properties. Over the past 20 years, FBDD has achieved significant success by delivering six FDA approved drugs namely, vemurafenib (Bollag et al., 2012), venetoclax (Souers et al., 2013), erdafitinib (Perera et al., 2017), pexidartinib (Tap et al., 2015), sotoracib (Blair, 2021) and Asciminib (Eşkazan, 2021) (Table 1). In addition, more than 40 molecules discovered by FBDD are currently in active clinical development (Erlanson, 2021). Four of these molecules: capivasertib (AstraZeneca/Astex/CRUK), lanabecestat (Astex/AstraZeneca/Lilly), pelabresib (CP-0610) (Constellation) and verubecestat (Merk) have reached in the phase 3 of the clinical trial (Erlanson, 2021). However, the use of FBDD for the development of anti-pseudomonal molecules has been limited to only a handful of studies, some of which will be discussed later in this review, but focusing on the contribution of our own lab.

Until recently one of the main limitations of fragment-based approaches was the need to have crystals with sufficient resolution to see the fragments. This is now less of a problem as many companies such as Astex and Astra Zeneca, as well as our own academic work, has been radically transformed by the use of cryo-EM. The power of cryo-EM to visualise small molecules is illustrated by recent work in both our academic work and Astex company work, where cryo-EM is now being widely used for fragment-based drug discovery. A recent example from our academic work is on visualising small drug molecules in the 4,000 amino acid DNA-PKcs published in Nature (Liang et al., 2022).

**TABLE 1 |** Drugs on the market developed by FBDD approach and their biological targets.

Drug	Biological target	Description
 Vemurafenib	BRAF V600E mutant	Vemurafenib, the first drug developed using FBDD, marketed as Zelboraf, is an inhibitor of B-raf enzyme that lead to programmed cell death in melanoma cell lines
 Venetoclax	Bcl-2	Venetoclax is a BCL2 homology domain 3 (BH3) mimetic that blocks the anti-apoptotic B-cell lymphoma-2 (Bcl-2) protein leading to programmed cell death of chronic lymphocytic leukemia (CLL) cells
 Erdafitinib	FGFR	Marketed under the brand name Balversa, erdafitinib is a small molecule inhibitor of fibroblast growth factor receptor (FGFR, a subset tyrosine kinase) used for the treatment of bile duct cancer, gastric cancer and esophageal cancer
 Pexidartinib	CSF-1R	Marketed as Turalio, pexidartinib is kinase inhibitor that blocks the activity of colony-stimulating factor-1 receptor (CSF-1R). It is used to treat of adults with asymptomatic tenosynovial giant cell tumor (TGST)
 Sotorasib	KRAS G12C mutant	Sold under the brand name Lumakras and Lumykras, Sotorasib targets the common mutation G12C in KRAS protein associated with various forms of cancers. It is an inhibitor of RAS gtpase family of protein and used to treat non-small-cell lung cancer (NSCLC)
 Asciminib	Bcr-ABL	Asciminib (Scemblix) is a protein kinase inhibitor which specifically targets the ABL myristoyl pocket of the fusion protein Bcr-ABL. It is used to treat Philadelphia chromosome-positive chronic myeloid leukemia (Ph + CML)

We now review the steps involved in a typical target-oriented structure-guided fragment based drug discovery (**Figure 1**).

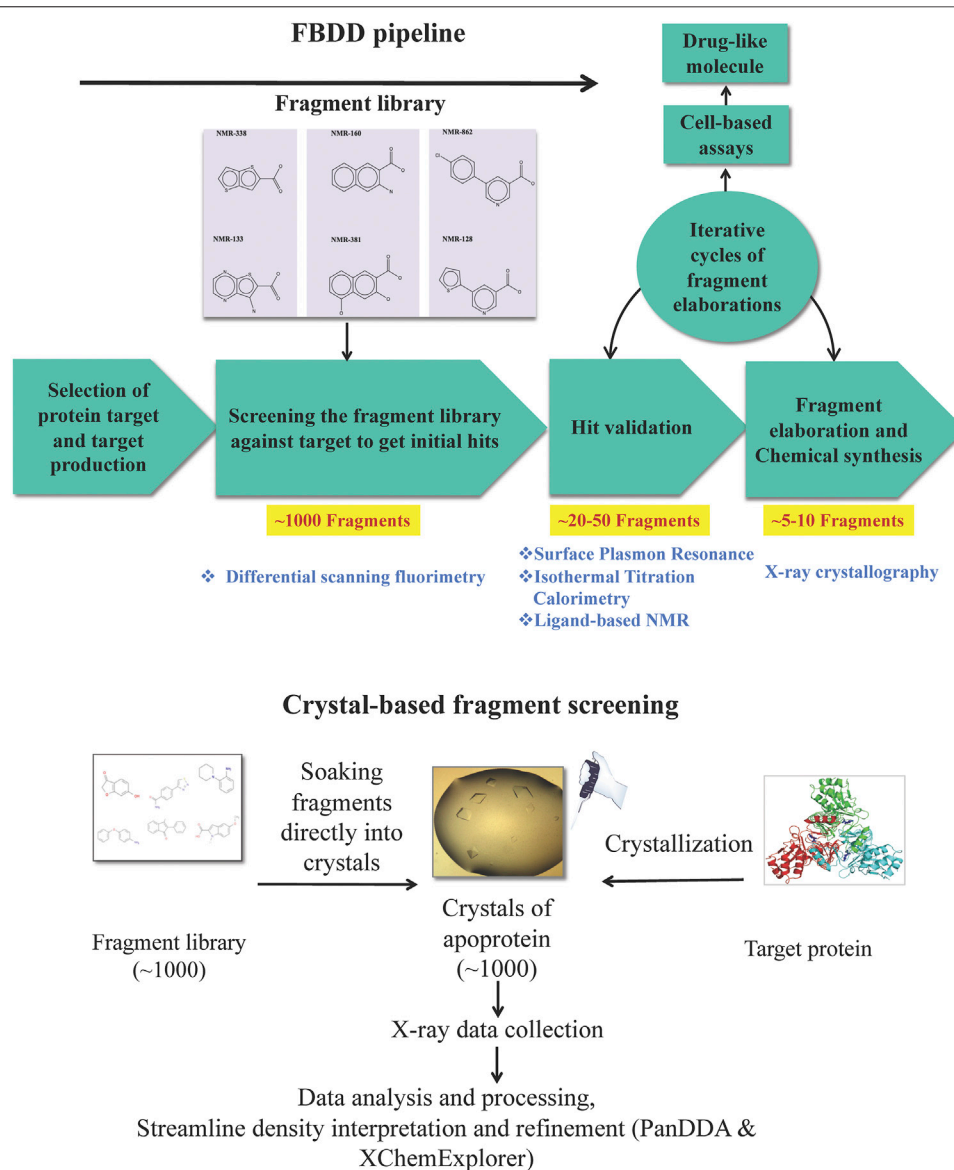
## Target Selection

Target selection is the first step in our drug discovery pipeline. A systematic analysis of literature on gene essentiality is performed in order to select essential drug-targets. These studies include various molecular genetic approaches, such as gene trapping, homologous recombination, transposon insertion mutagenesis (Van Opijnen and Camilli, 2013; Chao et al., 2016), and more recently developed CRISPR/CAS9 gene editing system (Wang et al., 2015; Evers et al., 2016; Morgens et al., 2016; Senturk et al., 2017; Bartha et al., 2018; You et al., 2020), to identify essential genes by first incorporating inactivating mutations in them and then determining the viability of the cells or organism carrying these mutations. The next step is to look for the similarity of these essential genes to those of the human genome. Targets similar to a gene in the human genome or human gut microbiome are not pursued in order to avoid cross-reactivity of the developed antibiotic and maintain the health of the microbiome. The protein products of these genes are considered for selecting a target. This is followed by the ligandability/druggability analysis by detecting and evaluating the ligand binding sites using modelling programs, such as the  $\alpha$ -shape

based approach (Edelsbrunner and Mücke, 1994) in MOE (MOLECULAR OPERATING ENVIRONMENT, 2016 Integrated Computer-Aided Molecular Design Platform), SiteMap (Halgren, 2009), LIGSITE (Huang and Schroeder, 2006), CASTp (Dundas et al., 2006) and more recently by generating fragment HOTSPOT maps (Radoux et al., 2016), taking advantage of the available 3-dimensional structure for potential target proteins. Cryptic binding sites can be detected by molecular dynamics simulations (Kuzmanic et al., 2020) and the cosolvent mapping method (Yang and Wang, 2010). In cases where the 3-dimensional structures are not available, they can be modelled using software such as Modeller (Šali and Blundell, 1993; Eswar et al., 2006) and more recently developed deep-learning algorithms, such as AlphaFold (Senior et al., 2020), AlphaFold2 (Jumper et al., 2021) and RoseTTAFold (Baek et al., 2021), and the resulting structure can be then analysed for ligandability. Other criteria while selecting a target include preference for non-membrane proteins and availability of well-established biochemical/functional assays in order to achieve a speedy drug-discovery process.

## Target Protein Production

Once the target protein is selected, the corresponding gene is cloned into an expression vector using molecular biology



**FIGURE 1 |** Fragment-based drug discovery (FBDD). The various biochemical, biophysical and structural biology techniques used in a FBDD pipeline are illustrated.

techniques such as PCR (Mullis, 1990; Garibyan and Avashia, 2013; Hoseini and Sauer, 2015), restriction digestion (Roberts, 2005) and ligation (Cohen et al., 1973) resulting in an expression plasmid. The target protein is often overexpressed in *Escherichia coli* using this expression plasmid. The overexpressed protein is purified in large quantity using protein chemistry techniques such as affinity chromatography (Rodriguez et al., 2020), ion-exchange chromatography (Walls and Walker, 2017) and size exclusion chromatography (Burgess, 2018; Held and Kilz, 2021).

### Design/Selection of Fragment Library for a Selected Target

Designing/selecting a generic fragment library for a fragment-screening campaign in general involves considerations such as diversity, solubility, molecular weight (MW), cLogP, polar

surface area (PSA), Fsp3 (Lovering et al., 2009; Wei et al., 2020), natural product-likeness and the number of the fragments (Mahmood and Ramachandraiah, 2021). On the other hand, there are certain criteria that are important to consider while assembling/choosing a fragment library with high chemical and structural diversity to be used for screening against a selected target. These involve ligand and protein structure-based approaches of library design/selection (Schuffenhauer et al., 2005). The ligand-based approach involves the knowledge of binding mode of known ligands for the target and the fragments that can undergo similar interactions with the target are selected. Importantly, such interactions are often conserved over the whole target family and hence a fragment identified using this approach can be used to target other members of the family. An example of



such approach include selection of metal chelators in a fragment screening campaign against gelatinase (Wang et al., 2002). Commercial fragment libraries dedicated for matrix metalloprotease (MMPs) such as Chelator Fragment Libraries (CFL 1.1), comprising a range of metal chelating moieties (Agrawal et al., 2008) and protease inhibitor-enriched library (Baell and Holloway, 2010), have been deliberately chosen to target a matrix metalloproteinase, LasB (Garner et al., 2012; Kany et al., 2018a). Another example of target-focussed libraries are halogen-enriched fragment libraries (HEFLibs), comprising chemical probes that identify halogen bonds as the main feature of binding mode, and originally constructed to find chemical moieties that stabilise p53 mutants (Heidrich et al., 2019). Structure-based approaches for design/selection of the fragment library may involve in silico screening as described for DNA gyrase ligands (Boehm et al., 2000). The hit rates in fragment screening with such target-focused fragment libraries are much higher than those with generic fragment libraries.

### Fragment Screening to Identify Hits

Biophysical, biochemical, structural biology and computational methods are used to screen a library of fragments (~1,000) against the target. The most popular biophysical method used for screening is differential scanning fluorimetry (DSF), often known as thermal shift assay, a technique that measures the denaturation temperature (melting temperature,  $T_m$ ) of the protein (Senisterra et al., 2012) and allows the detection of compounds that increase the  $T_m$  of a target protein on binding by promoting protein stability (Niesen et al., 2007). Surface plasmon resonance (SPR), a label-free technique, is another biophysical technique that is used for direct screening of fragment libraries (Navratilova and Hopkins, 2010). Fragment-screening by SPR is advantageous owing to its cost-effectiveness, the possibility of high-throughput mode and the requirement of a very small amount of protein (Neumann et al., 2007; Chavanieu and Pugn  re, 2016). NMR spectroscopy can also be used to perform initial screening, possibly using cocktails of fragments to accelerate the procedure (Leach, 2006).

Microscale thermophoresis (MST), another well-established biophysical technique to quantify any kind of biomolecular interaction (Jerabek-Willemsen et al., 2014; Mueller et al., 2017), is based on thermophoretic mobility, the directed motions of biomolecules and macromolecular complexes in solution in a temperature gradient; these strongly depend on molecular properties such as size, charge, hydration shell or conformation. MST enables the identification of compounds whose binding to a target changes the thermophoretic mobility of the target (Asmari et al., 2018), including even the weak binders, such as fragments (Linke et al., 2016). It has been demonstrated that MST is amenable for implementation into high-throughput screening cascades (Rainard et al., 2018) and has potential to maximize the efficacy of fragment screening campaigns owing to a high degree of automation in the technique generating quantitative data for affinity ranking in a rapid and precise manner (Rainard et al., 2018). Biolayer interferometry (BLI), especially useful for targeting

protein–protein interactions (Wartchow et al., 2011), and nanoelectrospray ionization mass spectrometry (ES-MS) (Maple et al., 2012) are other methods that have recently been used for fragment screening and ranking.

Crystal-based screening of a fragment library has also been developed (Patel et al., 2014), and this is routinely practiced at the XChem facility of Diamond Light Source, United Kingdom as a highly streamlined process, allowing more than 1,000 fragments to be screened in less than a week (Douangamath et al., 2020. <https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening.html>). For this, many crystals of the target protein are produced and the fragments are soaked in these crystals using the acoustic droplet ejection technique (Collins et al., 2017). High-throughput X-ray diffraction data collection (36 h of unattended beamtime) is carried out on the resulting crystals at the dedicated beamline, IO4-1, of the Diamond Light Source. The processing of X-ray data is performed automatically while the data are being collected, using the dedicated data processing softwares, such as xia2 (Winter 2010) among others. The structure solution and map analysis is performed by the XChemExplorer package (Krojer et al., 2017) including modules such as DIMPLE (Difference Map PipeLine) for initial refinement and automated difference map calculation to allow for quick assessment of X-ray data to see if a ligand has bound to the structure; AceDRG (Long et al., 2017) for generation of ligand restraints; REFMAC (Murshudov et al., 2011) for refinement; PanDDA (Pearce et al., 2017) for the streamline density interpretation and hit detection; a number of tools from PHENIX (Adams et al., 2010) for structure validation; and Coot (Emsley et al., 2010) for automated model building. The advantage of crystal-based screening is that it can detect fragments that have very weak binding affinities to the target protein.

To cut the cost of experiments and as a pre-screen, virtual docking of the fragment library on the three-dimensional structure of target protein, known as virtual screening (VS), is often performed in order to obtain initial hits (Bielska et al., 2011; Singh et al., 2018; Yamaotsu and Hirono, 2018; Gimeno et al., 2019). However, it is of note that these hits are hypothetical and warrant experimental validation (Zhu et al., 2013). VS is advantageous in cases where a very large library comprising millions of scaffolds has to be screened, where it can reduce the number of molecules needed to be screened experimentally by other biophysical/biochemical methods. VS campaign can be supplemented by machine and deep learning principles, especially when dealing with large data sets (Melville et al., 2009). Machine learning approaches in ligand-based VS can address complex compound classification problems and help predict new active molecules (Lavecchia, 2015). Deep learning in the field has emerged in recent years and it has been demonstrated to have utility beyond bioactivity prediction to other drug discovery problems such as *de novo* molecular design, synthesis prediction and biological image analysis (Chen et al., 2018).

### Hit Validation

Hits from initial screening are confirmed by ligand-based NMR spectroscopy (Hajduk et al., 1997; Mayer and Meyer, 1999; Dalvit et al., 2001). Real-time fragment-binding affinity and kinetics can be determined using SPR (Yadav et al., 2012; Capelli et al., 2020) and thermodynamics of fragment binding is determined using

isothermal titration calorimetry (ITC) (Srivastava and Yadav, 2019). The 3D structures of protein-fragment complexes are determined using X-ray crystallography to instruct the process of fragment elaboration. Functional biochemical assays are performed to measure the inhibition activity of the fragments.

### Fragment to Lead Optimization and Cell-Based Assays

Guided by the 3D structure of protein-fragment complexes the fragments are elaborated using growing (Hoffer et al., 2018), merging (Nikiforov et al., 2016; Miyake et al., 2019) or linking (Bancet et al., 2020), whichever suits in order to generate compounds with higher affinity and drug-like properties. In-silico molecular modelling is a remarkably beneficial tool as, together with X-ray and cryo-electron microscopy, it provides a means to develop a suitable linker to attach to the low-affinity molecules (Scoffin and Slater, 2015; de Souza Neto et al., 2020). We do this with the help of our chemistry collaborators. The elaborated compounds are then tested by ITC for their affinities to target and functional assays for their inhibition activity against the target. An iterative cycle of fragment growing/merging/linking, followed by biophysical, biochemical and structural analyses while maintaining high ligand efficiency is established. The final compounds, often achieving nanomolar affinities and having potent *in vitro* inhibition, are then used for cell-based assays to check their minimal inhibitory concentrations (MICs) (Kowalska-Krochmal and Dudek-Wicher, 2021). We do this with the help of our cell biologist collaborators.

### Expected Outcomes

Small “fragments” of drug-like molecules may bind to hotspots on a defined protein target. The initial screening of a library of fragments (~1,000) often results in several fragments (~20–50) binding to a target. Hit validation should lead to 5–10 fragments having significant affinity to the target. Fragments can be chemically “grown” or “linked” to deliver compounds with nanomolar binding affinities to the target and altering its activity. Such small-molecule inhibitors developed against essential protein targets can be developed into lead compounds. Further chemical modification of the chemical structure of the lead compound can lead to improvement in potency, selectivity and pharmacokinetic parameters. The pharmacologically active moiety thus obtained may have poor drug likeness which can usually be further modified chemically to result in a more drug-like compound for testing biologically or clinically. Such a compound should be able to compromise the growth and survival of the pathogen inside the host organism.

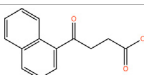
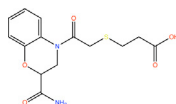
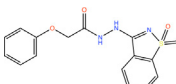
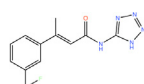
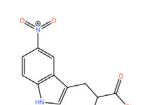
## ANTIBIOTIC STRATEGIES TARGETING *P. AERUGINOSA*

### Targeting Lipopolysaccharide Biosynthesis Pathway

#### Fragment-Based Drug Discovery on LpxA and LpxD

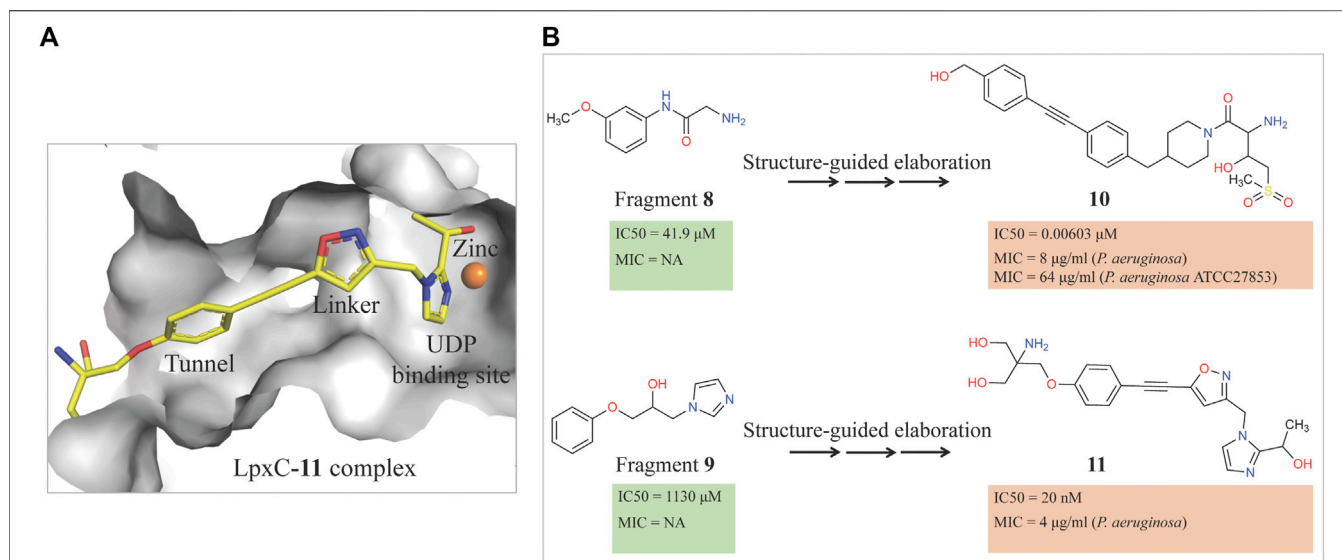
The presence of an outer membrane in Gram-negative bacteria (GNB) protects the bacterial cell not only from the harsh environment but also antibiotics (Koch, 2017).

**TABLE 2 |** Binding affinities to PaLpxA and PaLpxD of the novel small-molecule scaffolds identified following an FBDD campaign. A few of them show dual binding having affinities to both PaLpxA and PaLpxD. NA: binding affinity could not be determined; NB: No binding.

Compound	Structure	LpxA (μM)	LpxD (μM)
1		NA	NA
2		19.5	36.7
3		16.7	NB
4		13.6	NB
5		2.1	NB

Lipopolysaccharide (LPS), the major component of the outer leaflet of this membrane (Raetz and Whitfield, 2002), are virulence factors essential in many clinically important GNBs, such as *P. aeruginosa*, where it plays important roles in the structural integrity of the bacteria and its defence against the host; hence the enzymes of the LPS biosynthesis pathway are attractive drug targets (Cryz et al., 1984; King et al., 2009). Although there are currently no antibiotics targeting LPS biosynthesis, compounds inhibiting this biochemical pathway can lead to the development of new antibiotics with novel mechanism of action (Jackman et al., 2000; Joo, 2015). Lipid A, a glucosamine disaccharide that is connected to multiple fatty acid chains of various lengths, is the minimal component of LPS required for cellular viability in most GNBs (Anderson et al., 1993; Rotella, 1997; Raetz and Whitfield, 2002). Moreover, lipid A is the antigenic determinant of LPS that triggers septic shock. The enzymes of the LPS biosynthesis pathway, including lipid A biosynthesis, are attractive drug targets for therapeutic interventions. An *in silico* study suggests that, in *P. aeruginosa*, a total of thirteen enzymes are involved in this LPS biosynthesis, of which seven enzymes matched with the list of candidate essential genes obtained by transposon mutagenesis study (Perumal et al., 2007). These enzymes have no human homologues and hence they can serve as potential drug targets.

LpxA, LpxC and LpxD are the first three enzymes in lipid A biosynthesis pathway. Of these, LpxC has been extensively targeted by antibiotic discovery leading to the development of many small molecule inhibitors with antibacterial properties (Kalinin and Holl, 2017). In contrast, LpxA and LpxD have remained largely unexplored for the development of small



**FIGURE 2 | (A)** Location of compound **11** [2-(1S-hydroxyethyl)-imidazole] in the *Pa*LpxC-**11** complex structure (pdb id 7ci9) determined by X-ray crystallography (Yamada et al., 2020). The protein binding site is represented as grey surface and the compound as stick. The zinc ion is shown as an orange sphere. **(B)** Structure-guided elaboration of fragments **8** and **9**, leading to compounds **10** and **11**, respectively with enhanced potencies and inhibition properties.

molecule inhibitors and are targeted mostly by antibacterial peptides (Williams et al., 2006; Jenkins and Dotson, 2012; Jenkins et al., 2014; Dangkulwanich et al., 2019). LpxA and LpxD are amenable to dual-targeting inhibitors due to their structural similarity and the advantage of such inhibitors include increased potency and reduced likelihood of resistance against the inhibitors (Silver, 2007). FBDD against LpxA and LpxD to develop dual-targeting inhibitor was performed by Kyle G. Kroeck and co-workers (126). An initial fragment screening by virtual docking of a subset of the ZINC fragment library of John J. Irwin and Brian K. Shoichet (Irwin and Shoichet, 2005), followed by SPR characterization of binding fragments and X-ray crystallography analysis of enzyme-fragment complexes to verify binding and to reveal binding modes for lead optimization have identified novel small-molecule scaffolds that can serve as starting point for future inhibitor discovery (Kroeck et al., 2019). Several of these novel ligands, which have shown dual-binding activity (Table 2), have the potential to result in small molecule inhibitors targeting LpxA and LpxD simultaneously. For instance compound 1 in its complex with LpxA and LpxD adopted similar binding poses in the two acyl-chain binding pockets revealing both the structural similarities and differences that can guide the development of future dual-targeting inhibitors. Interestingly, these structures also highlight additional binding hot spots shared by the two enzymes that can be exploited for further lead optimization.

### Fragment-Based Drug Discovery on LpxC

The bacterial enzyme UDP-3-O-acyl-N-acetylglucosamine deacetylase (LpxC) is an attractive target for the development of novel therapeutic agents (Erwin, 2016; Chen et al., 2019). It is essential to most Gram-negative bacteria

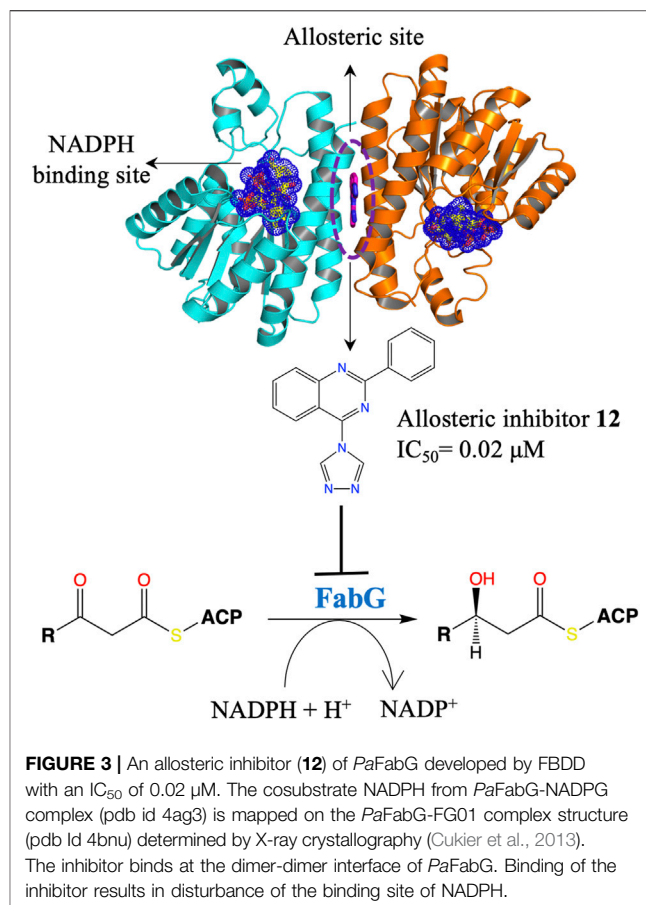
including *P. aeruginosa* (Raetz and Whitfield, 2002, 2002) and catalyzes the removal of an N-acetyl group from UDP-3-O-acyl-N-acetylglucosamine (which constitutes the core of Lipid A), the first committed step in biosynthesis of the LPS, which is an essential component of the bacterial cell wall. Several series of inhibitors have been developed against *P. aeruginosa* LpxC (*Pa*LpxC) employing non-FBDD approach (Erwin, 2016; Kalinin and Holl, 2017; Chen et al., 2019) exploiting the following features of molecules at the binding site: 1) a zinc-chelating motif, 2) a polar group occupying the UDP binding pocket, 3) a linker that is mainly hydrophobic but sometimes makes at least one hydrogen bond interaction, 4) an extended hydrophobic moiety that binds in the tunnel, and 5) often addition of groups out into the solvent at the end of the tunnel to modulate compound properties. The most advanced compounds carry a hydroxamate moiety that coordinates the zinc ion at the core of the enzyme. One of these compounds, ACHN-975 (Cohen et al., 2019; Lpxc et al., 2019), a hydroxamate-based histone deacetylase inhibitor (HDACI) entered clinical trial and was approved by FDA for its use in oncology applications but was discontinued beyond oncology (Lpxc et al., 2019) due to its off-target side effects associated with hydroxamate group (Shen and Kozikowski, 2016).

Therefore, an FBDD campaign, to identify LpxC inhibitors with a non-hydroxamate metal-coordinating group, was started (Yamada et al., 2020). 1,152 compounds from the Vernalis Research fragment library (Baurin et al., 2004; Chen and Hubbard, 2009) were screened against *Pa*LpxC by ligand-observed NMR (STD, water-LOGSY, and CPMG), using cocktails of 6 fragments (Hubbard et al., 2007; Hubbard and Murray, 2011) followed by singleton NMR competition assays to look for the competitive binding of

the fragments with respect to binding of the tool compounds **6** and **7**, (hydroxamate group containing inhibitors reported previously) to LpxC. The 28 fragments showing competitive binding were further assessed in the Fluorescence Polarization (FP) binding assay followed by Fluorescamine-Based Functional Activity Assay. Some of which such as **8** showed clear inhibition with an  $IC_{50}$  of 41.9  $\mu$ M. Two zinc-chelating fragments, a glycine fragment **8** and an imidazole fragment **9**, both of which were competitive with **6**, were selected for further investigation. Both of these were shown to stabilize LpxC in a TSA experiment further confirming the binding. The crystal structure of PaLpxC-fragment **8** complex guided the initial medicinal chemistry and iterative cycle of protein-ligand co-crystallization, confirming binding poses, structure-based modeling predicting the binding poses and guiding the possibility of enhancing potency and chemical modification. This led to the development of a compound **10** with many fold increase in the activity (functional  $IC_{50}$  0.00603  $\mu$ M) as compared to initial fragment **8** (functional  $IC_{50}$  41.9  $\mu$ M). However, this compound showed only weak antibacterial activity even in the presence of phenylalanine-arginine  $\beta$ -naphthylamide (PA $\beta$ N) (an efflux pump inhibitor) (MIC of 8  $\mu$ g/ml against *P. aeruginosa* and 64  $\mu$ g/ml against *P. aeruginosa* ATCC27853), so this series was not pursued any further. Similar medicinal chemistry on imidazole fragment **9** resulted in an advanced lead compound 2-(1 Hydroxyethyl) imidazole **11** which exhibited low nanomolar inhibition of PaLpxC (functional  $IC_{50}$  of 20 nM) and a minimal inhibitory concentration (MIC) of 4  $\mu$ g/ml against *P. aeruginosa* (Figure 2). Further optimization and *in vivo* efficacy measurement are under consideration for this compound and will be described in a future publication. The lead compounds of both the series exhibited significant selectivity towards zinc metalloenzyme, LpxC. It was demonstrated that maintaining the zinc-chelation motif as it was in the fragments, engineering aliphatic linkers of appropriate length extending the fragments in the hydrophobic tunnel of LpxC and appending various hydrophobic groups at the para position of the benzyl successfully increased the potencies of compounds with retained ligand efficiencies.

## Role of Aeropath Project in Fragment-Based Drug Discovery

An EU-funded project AEROPATH, coordinated by University of Dundee, was launched in November 2008 with the aim of identifying, characterising and exploiting novel drug targets from the Gram-negative bacterium, *P. aeruginosa*, by applying a multidisciplinary approach encompassing target validation, structural characterization, assay development and hit identification from small molecule libraries following a highthroughput or fragment-based screening campaigns. Derivation of the structural models of the potential targets was one of the central aims of this project as the structural data allow druggability analysis of the active sites (Krasowski et al., 2011; Radoux et al., 2016) and



support the curation of the structure-activity relationship of the identified ligands. Towards the end of this project, 102 targets were selected based on the available genome with preliminary annotation of *P. aeruginosa* strain PAO1 together with gene essentiality studies and other considerations such as feasibility of enzyme assays, chemogenomics information and an appropriate balance of novel uncharacterized proteins versus established targets for antibacterial drug design (Moynie et al., 2013). *De novo* structures of 39 of these targets were determined using X-ray crystallography and NMR. In addition, the structures of more than 60 complexes involving substrate, cofactor and inhibitors have been determined and published by the consortium. Crystal structures of eight targets including both hypothetical uncharacterized protein and metabolic enzymes from various functional classes were reported before the end of the project (Moynie et al., 2013). This plethora of structural information was envisaged to aid the FBDD campaigns.

## Targeting Fatty Acid Biosynthesis Pathway Fragment-Based Drug Discovery to Target FabG

The fatty acid synthesis type II (FAS II) system of bacteria has been identified as an attractive target for therapeutic interventions and several antibiotics targeting this pathway are



known, such as triclosan, isoniazid and thiolactomycin (Campbell and Cronan, 2001; Zhang et al., 2004; Heath and Rock, 2006; Parsons and Rock, 2011; Pan et al., 2012). FabG is a NADPH-dependent 3-oxoacyl-acyl carrier protein (ACP) reductase that plays a key role in the FAS II system of pathogenic microorganisms and has been identified as an attractive drug target. It catalyses the reduction of 3-oxoacyl-ACP to 3-D-hydroxyacyl-ACP intermediates during the elongation cycle of fatty acid biosynthesis (Rock and Jackowski, 2002; Goodman and McFadden, 2008; Chan and Vogel, 2010). FabG qualifies to be a promising drug target due to its essentiality, high conservation across bacteria and presence of only a single isoform in most of the bacterial species (Zhang et al., 2004). Inhibitors of *P. aeruginosa* FabG (*PaFabG*) identified include largely natural product extracts that pose significant drug development challenges and hence none of them have reached the clinic (Zhang and Rock, 2004; Tasdemir et al., 2006; Wickramasinghe et al., 2006; Sohn et al., 2008; Zhang et al., 2008). As part of the AEROPATH project, Cyprian D. Cukier and coworkers have established the essentiality of *fabG* gene in *P. aeruginosa* using gene knockout procedure and following a FBDD approach they then developed a series of novel *PaFabG* inhibitors with IC<sub>50</sub> values in nanomolar to low micromolar range and ligand efficiencies in the range of 0.37–0.53 (Cukier et al., 2013). Although the compounds show no phenotypic response in the Gram-negative *P. aeruginosa* either due to poor penetration of the compounds through the Gram-negative cell wall or to rapid efflux of the compounds, the diverse chemotypes of these inhibitors presents a number of options for optimization to increase intracellular concentrations. Structural investigation of 16 *PaFabG*-inhibitor complexes by X-ray crystallography in this study reveals that inhibitors bind at a novel allosteric site (cryptic binding site) located at the dimer-dimer interfaces of *PaFabG* (Figure 3) and this binding induces the conformational changes that propagate to the active site and results in the disturbance of the catalytic triads (residues S141, Y154, and K158) with loss in the binding affinity of cosubstrate NADH, thus inhibiting the enzyme. Kinetic analysis of inhibition suggested a noncompetitive mode of inhibition with respect to NADH.

## ANTI-VIRULENCE STRATEGIES AGAINST *P. AERUGINOSA*

Antibiotic development to treat *P. aeruginosa* infections is undergoing a crisis due to the rapid evolution and spread of resistance in bacteria against the current antibiotics. Interfering with bacterial virulence network including virulence factors (proteases, elastase, endotoxins, and polycyanin), lipopolysaccharides, flagella, extracellular polysaccharides, and type II, III, IV and VI secretion system, instead of targeting their viability (growth and survival), to combat *P. aeruginosa* infections offers tantalizing prospects of novel antimicrobials and this approach has gained momentum in recent years to overcome today's crisis in antibacterial development (Papaioannou et al., 2013; Anantharajah et al., 2016; Gao et al., 2017; Boulant et al., 2018; Ranjbar et al., 2019). Moreover, this strategy will reduce the

propensity to induce resistance as it removes the strong selection pressure imposed by bacteriostatic or bactericidal agents.

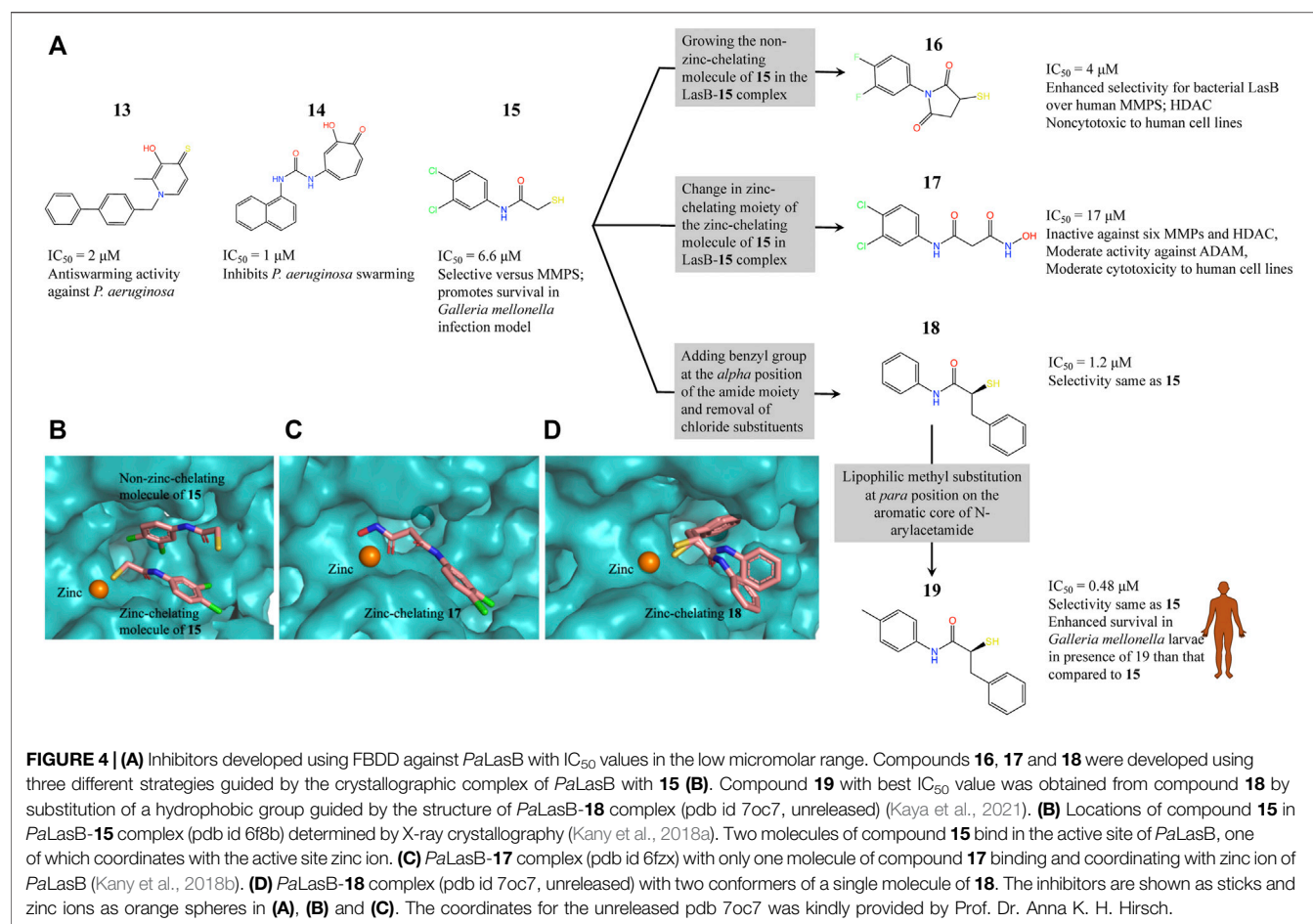
## Targeting Quorum-Sensing Systems to Develop “Second Generation” Antibiotics

Inhibiting quorum-sensing systems (QSS) and the regulators that promote biofilm formation is one such novel strategy to attenuate *P. aeruginosa* virulence (O'Loughlin et al., 2013; Shao et al., 2020; Manos, 2021). Four QS pathways (*pqs*, *iqs*, *las*, *rhl*) have been identified in *P. aeruginosa* (Lee et al., 2013). Of these, *pqs* and *las* mediated QSS have been targeted using FBDD and this will be described in the following section of this review.

## Fragment-Based Drug Discovery on LasB

LasB (or pseudolysin, *Pseudomonas* elastase B) is the most abundant extracellular collagenase protease secreted by *P. aeruginosa* with a hydrolysis activity against a broad spectrum of substrate proteins from the host, causing damage to host tissues, disruption of the host immune response and promoting inflammation associated with *P. aeruginosa* virulence and disease pathology (Wretling and Pavlovskis, 1983; Saint-Criq et al., 2018; Galdino et al., 2019). Since LasB is extracellular the direct inhibitors are not required to cross the difficult-to-penetrate *P. aeruginosa* cell membrane and the fact that the target belongs to a class of validated drug targets (metalloprotease), against which there are already clinically useful drugs, makes LasB an attractive antivirulence target for therapeutic intervention (Galdino et al., 2019). Synthesis of small-molecule inhibitors of *P. aeruginosa* LasB (*PaLasB*) had started more than 40 years ago and since then eighteen inhibitors have been developed (Everett and Davies, 2021). The FBDD approach to develop inhibitors against *PaLasB* started in 2012 by screening a library of 96 metal chelating fragments (CFL-1.1) (a library designed specifically with fragments which can coordinate with the metal ion in the active site of metalloproteins) against *PaLasB* and the initial hits were followed up by medicinal chemistry optimization program (Agrawal et al., 2008; Garner et al., 2012). This led to the development of a thiopyridone **13** (Garner et al., 2012) with an IC<sub>50</sub> of 2.73 μM for *PaLasB* but was promiscuous with regard to other metalloproteases, and the tropolone **14** (Fullagar et al., 2013) exhibiting similar potency (IC<sub>50</sub> of 1 μM for LasB) but had better selectivity over human metalloproteases, such as matrix metalloproteases (MMPs) and carbonic anhydrase II (Figure 4A). They were the first potent non-peptidic small molecule antagonists discovered (Cathcart et al., 2011) and the first targeted compounds exhibiting antiswarming activity (Garner et al., 2012).

With the aim of expanding the chemical space of *PaLasB* inhibitors, a functional screening based on FRET-based *in vitro* assay (Nishino and Powers, 1980) was performed against *PaLasB* using 330 fragments (Maybridge Fragment Library) and a protease inhibitor-enriched library (Baell and Holloway, 2010) comprising of 1,192 low molecular weight compounds (Kany et al., 2018a). This study led to the development of a thiobenzamide (N-(3,4-dichlorophenyl)-2-sulfanylamide) **15**, LasB inhibitor with thiol


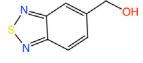



warhead, with an  $IC_{50}$  of 6.6  $\mu M$  (Figure 4A). This was selective with regard to human MMPs. The X-ray structure of PaLasB-**15** complex revealed that binding of **15** to LasB does not necessarily lead to closure of the binding site (Kany et al., 2018a), unlike that described for thermolysin-like proteases (Adekoya et al., 2015), enzymes with high structural similarity to LasB. Moreover, two molecules of **15** were observed in the structure (Figure 4B) and the binding of the second molecule was supposed to be supported by the binding of the first molecule. However, the Hill coefficient of 1 in the *in vitro* assay for **15** suggested that only one binding event was necessary for full inhibition of PaLasB. Nevertheless, this structure with two binding sites for **15** paved the way for the development of novel LasB inhibitors targeting the open conformation of the enzyme. Attempts to merge these two molecules into one *N*-benzylamide derivative did not provide the desired activity. Growing the non-zinc-chelating molecule of **15** in the PaLasB-**15** complex led to the development of *N*-arylsuccinimide **16** with an  $IC_{50}$  of 4  $\mu M$ , which showed, 1) significant selectivity for bacterial LasB over human MMPs and three other off-targets and 2) no signs of cytotoxicity in human cell lines (Konstantinović et al., 2020). In another subsequent study, the zinc-chelating molecule of **15** in the PaLasB-**15** complex was modified by changing the zinc-chelating moiety leading to a hydroxamate **17** (Kany et al., 2018b) with an  $IC_{50}$  of 17  $\mu M$  with moderate cytotoxic effects towards mammalian cell lines. Since only

one molecule of **17** was observed in the crystal structure of PaLasB-**17** complex (Figure 4C), unlike that of **15** in PaLasB-**15** complex, and **17** was less susceptible to oxidation in air, the authors preferred **17** over **15** for further exploration. Moreover, **17** could undergo the characteristic hinge-bending motion resulting in closure of LasB binding site unlike open conformation as observed in PaLasB-**15** complex. Overall, given their modest activities, the authors suggest that the molecules such as **15**, **16** and **17** should be considered as starting points for further chemistry rather than as potential drug development candidates which could pave the way for the rational development of selective protease inhibitors as potential new antibiotics.

Exploiting the alternative binding modes of **15** to guide efficient fragment growing resulted in a series of compounds with better activities (Kaya et al., 2021). Compound **18** thus developed had an  $IC_{50}$  of 1.2  $\mu M$  (Figure 4A) and maintains the selectivity as **15**. PaLasB-**18** complex structure was determined by x-ray crystallography. Here again, the binding of **18** to PaLasB leads to closure of the binding site. This structure provided a deeper understanding of the possible interactions in the surrounding unoccupied space and complemented with docking analysis it paved the way for further optimization. A focused, substrate-inspired structure-based optimization of **18** (substitution of methyl group at the *para* position on the aromatic

**TABLE 3 |** Fragment-like inhibitors of PaPqsD obtained by a FBDD campaign with two fragments showing strong inhibition (85%) of PaPqsD.

Fragment	Structure	Inhibition of PaPqsD
20		>99%
21		58%
22		80%

core of N-arylacetamide) resulted in compound **19** (Figure 4A) with fourteen-fold boost in activity ( $IC_{50}$  0.48  $\mu$ M) compared to **15** (173).

*In vivo* efficacy of these compounds was investigated using an insect model, *Galleria mellonella* and PA14, a virulent strain of *P. aeruginosa*. Since *Galleria mellonella* and mice show similar virulence patterns when infected with mutant PA14 strains (PA14, virulent strain of *P. aeruginosa*) this insect is likely to be a good model system to perform *in vivo* studies (Jander et al., 2000). Administering the anti-LasB compound **15** thus generated by FBDD campaign has shown significant increase in the survival of *Galleria mellonella* larvae infected with PA14. For instance, injection of 2.5 nM of **15** in the PA14 infected larvae increased the survival of the larvae from 43 to 73% after 65 h (Kany et al., 2018a). Compound **19** with better  $IC_{50}$  showed enhanced *in vivo* efficacy compared to **15**, thereby accelerating the translational path (Kaya et al., 2021). Antivirulent agents targeting quorum sensing of *P. aeruginosa* have previously been shown to enhance the survival of *Galleria mellonella* larvae infected with PA14 (Lu et al., 2014; Thomann et al., 2016). These compounds attain this efficacy by inhibiting the swarming and biofilm formation rather than being bacteriocidal. For instance, presence of compound **17** reduces the

formation of biofilm and release of extracellular DNA by *P. aeruginosa*. The advantage of such inhibitors is that they are capable of disrupting several important bacterial resistance mechanisms and hence open novel avenues to combat multidrug resistant strains of *P. aeruginosa*.

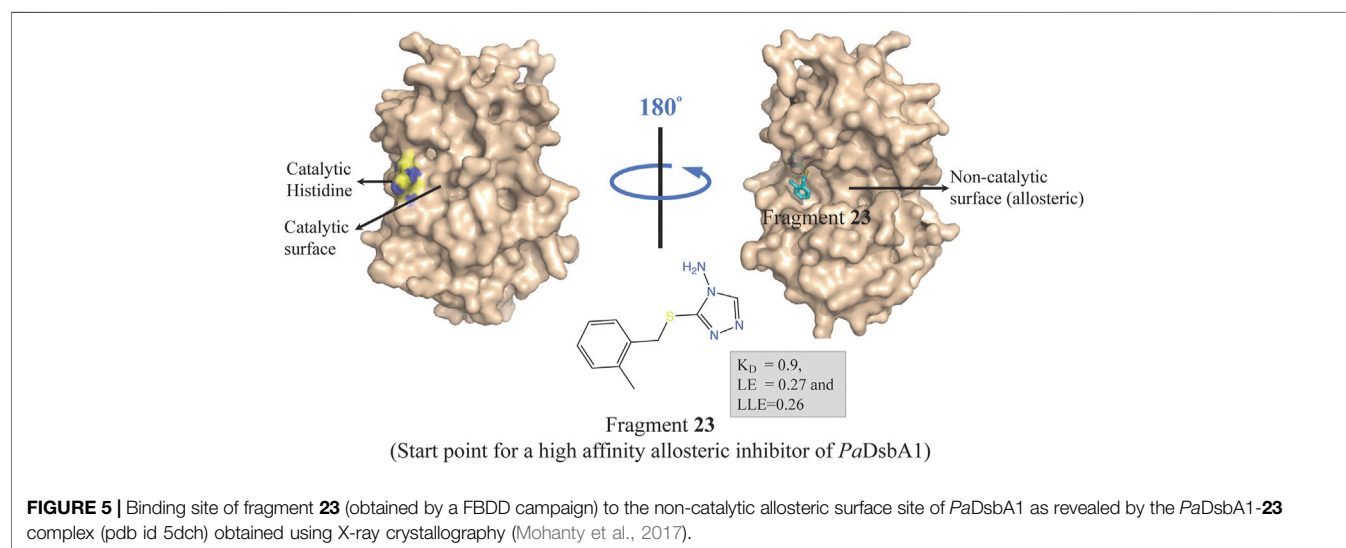
### Fragment-Based Drug Discovery on PqsD

PqsD is a key enzyme in the biosynthesis of signal molecules 2-heptyl-4-hydroxyquinoline (HHQ) and *Pseudomonas* quinolone signal (PQS) that are involved in the regulation of virulence factor (pyocyanine, elastase B, lectin A, rhamnolipids, and hydrogen cyanide) production and biofilm formation in *P. aeruginosa* (Van Delden and Iglewski, 1998; Diggle et al., 2003; Déziel et al., 2005; Yang et al., 2009). It has been shown that a mutant *P. aeruginosa* having a transposon insertion in the *pqsA* gene (deficient in HHQ and PQS production) forms less biofilm than the wild type (Gallagher et al., 2002; Múskén et al., 2010). Using a ligand-based approach the first class of PqsD inhibitors were identified which repressed HHQ and PQS production and biofilm formation in *P. aeruginosa*, validating PqsD as an attractive anti-biofilm target for the development of novel anti-infectives (Storz et al., 2012). Elisabeth Weidel and co-workers performed an SPR-based fragment screening against *P. aeruginosa* PqsD (PaPqsD) using a library of 500 fragments (Maybridge) with high structural diversity covering large chemical space in order to identify new scaffold for drug discovery (Koch, 2017). This screen resulted in identification of three fragments exhibiting a moderate inhibition of PaPqsD (Table 3). Two compounds showed strong inhibition of PaPqsD and hence these may be the starting point for future investigations.

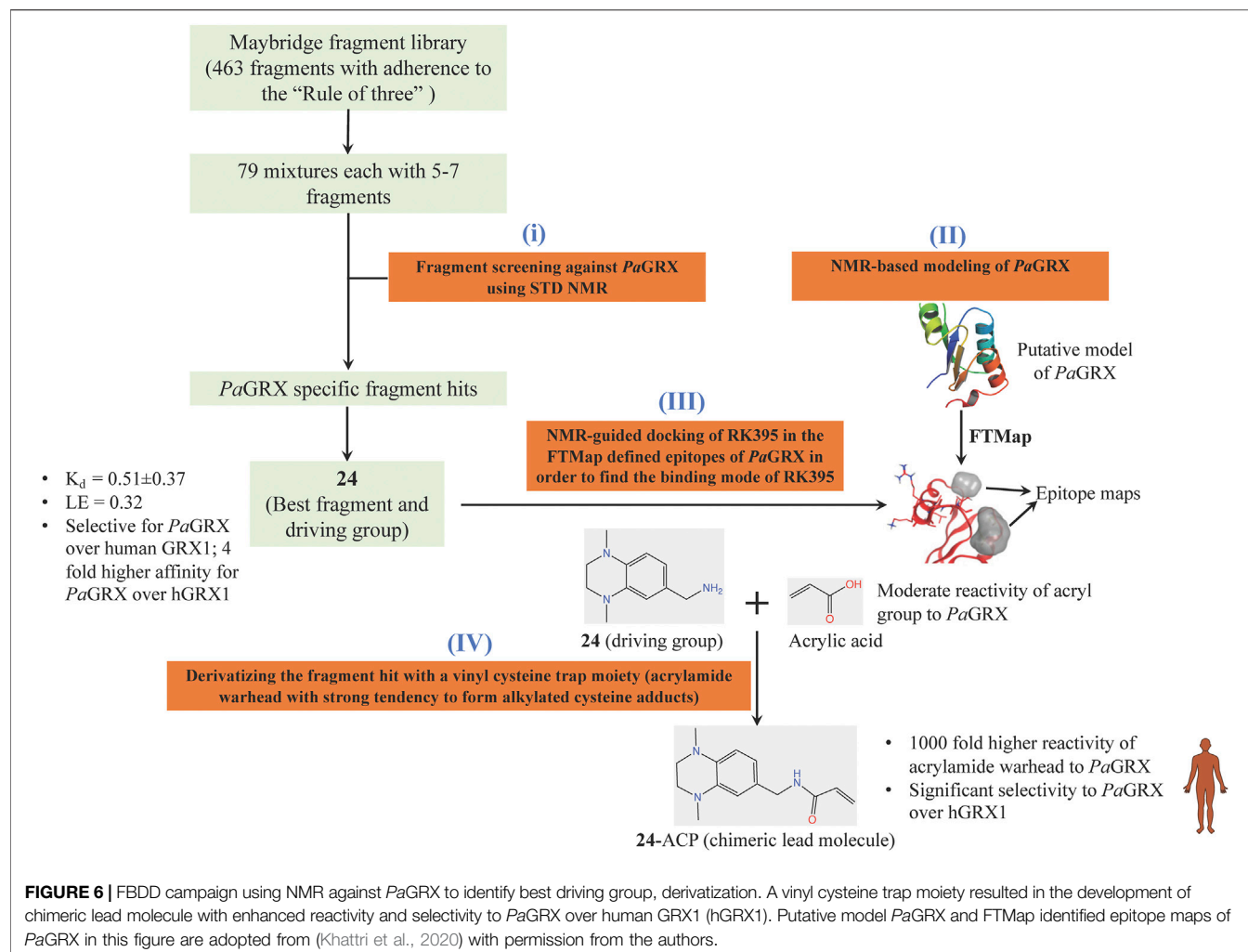
### Targeting Disulfide Mediated Protein Folding

#### Fragment-Based Drug Discovery on DsbA1

Oxidative protein folding is essential for assembly and function of many secreted and membrane proteins, and







DSB (disulfide bond) proteins that catalyze the disulfide bond formation are essential for virulence of many Gram-negative bacteria and hence these proteins are targets for novel antibacterial drugs (Heras et al., 2009). DsbA, an enzyme of disulfide oxidoreductase family, and its membrane-bound partner DsbB together catalyze the oxidative folding of disulfide bond containing proteins, many of which are virulence factors including secreted toxins and cell surface components, such as adhesins and pili. DsbA1 in *P. aeruginosa* (*PaDsbA1*) plays a pivotal role in the oxidative folding of virulence factors qualifying it as an attractive target for the development of new anti-virulence antimicrobials (Braun et al., 2001; Urban et al., 2001; Ha et al., 2003). In a FBDD approach, *PaDsbA1* was screened against a library of 1,137 fragments (Bradley et al., 2013) using ligand-detected STD NMR, which identified small molecules that bind selectively to *PaDsbA1* over *E. coli* DsbA (*EcDsbA*) suggesting the feasibility of species-specific development of narrow-spectrum inhibitor of *PaDsbA1* (Mohanty et al., 2017). Structural characterization of the complex of *PaDsbA1* with fragment 23 (highest affinity fragment with  $K_D$  of 0.9, ligand efficiency 0.27 and lipophilic

ligand efficiency of 0.26) using both X-ray crystallography and HADDOCK model revealed that the fragment is positioned at an interface between the thioredoxin (TRX) and helical domains of *PaDsbA1* on the non-catalytic face (Figure 5) of the enzyme. Unfortunately, fragment 1 had no inhibitory effect on the enzymatic activity of *PaDsbA1* as shown in an *in vitro* model-peptide folding assay. Nevertheless, these findings represent a starting point for the development of high affinity allosteric inhibitors of *PaDsbA1*.

### Fragment-Based Drug Discovery on GRX: Development of Covalent Inhibitors of *P. aeruginosa* by Fragment-Based Drug Discovery

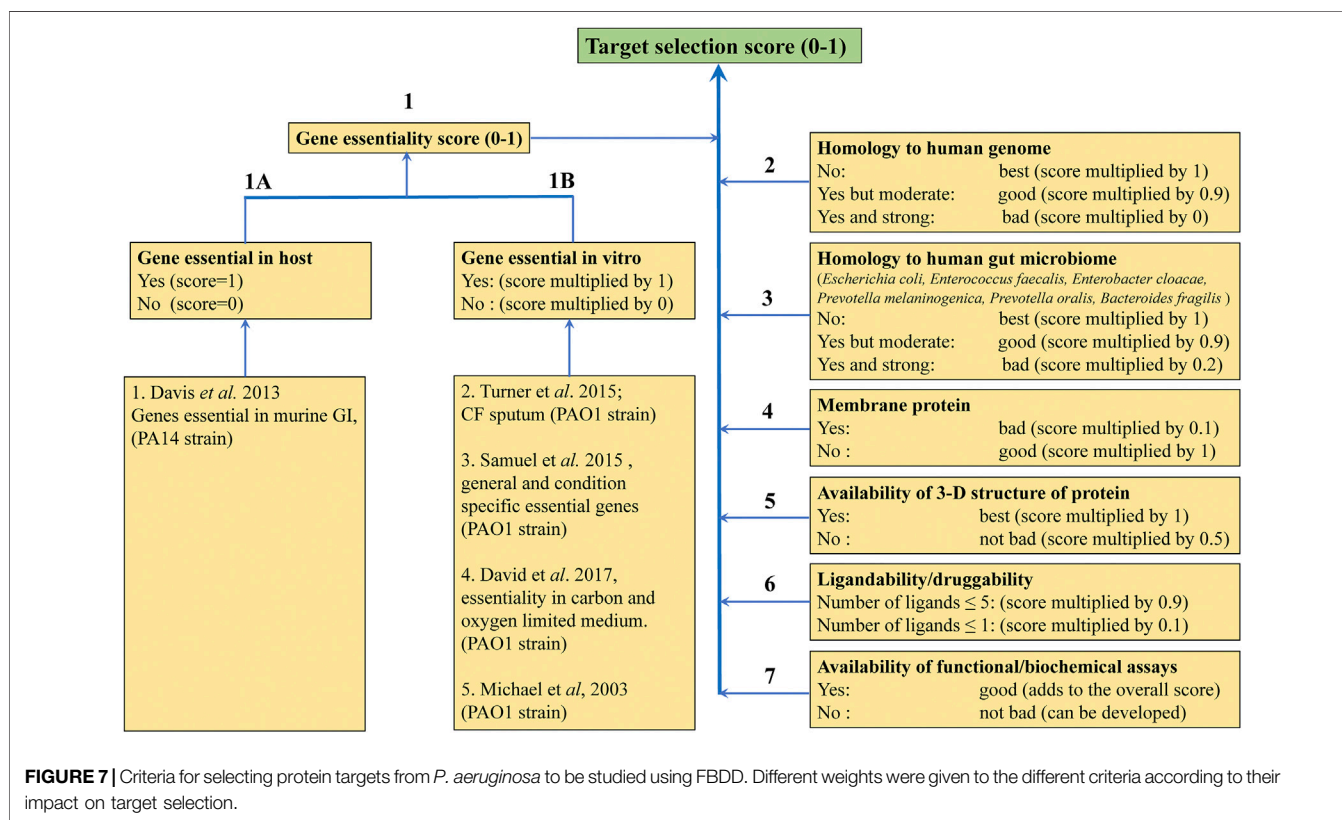
Inspired by the need to develop novel classes of drug molecules targeting unconventional drug targets using a mechanism that can circumvent the efflux-mediated resistance mechanism, covalent inhibitors were sought. The present strategy involves combining a covalently reactive functional group with targeting moieties (of compounds)

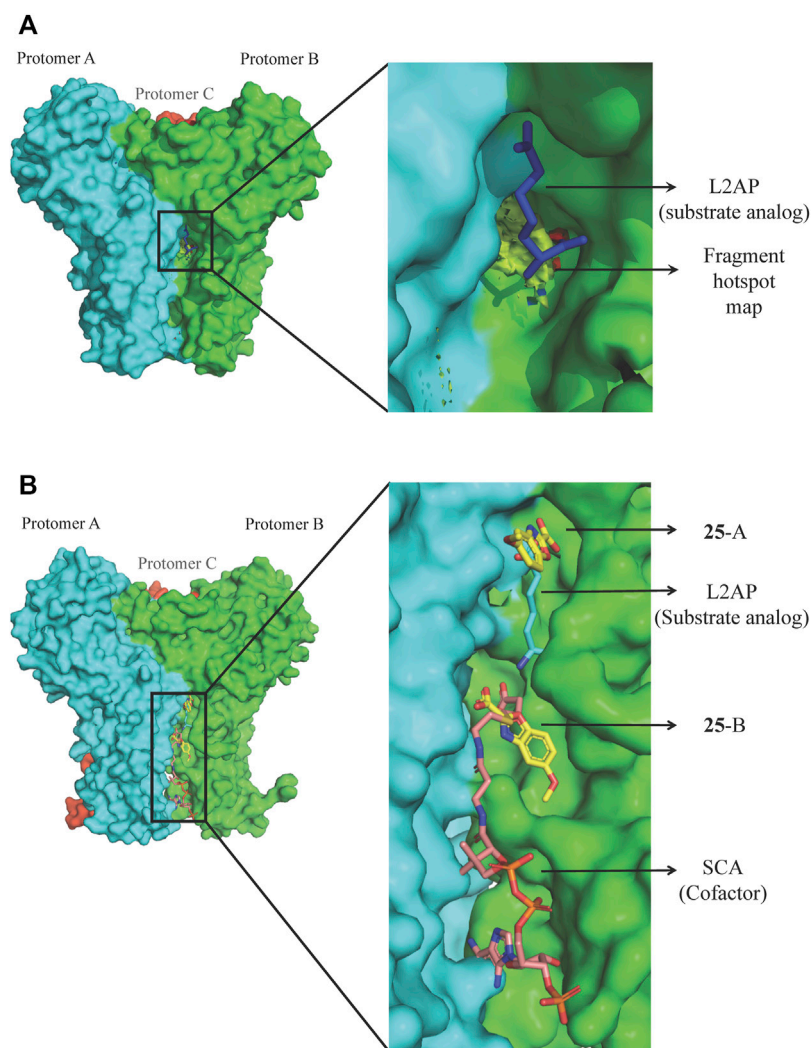
selective for species-specific proteins from important biochemical pathways that might have been missed due to cross-reactivity issues to the host organism. Such atypical protein targets are less likely to contain resistance-conveying mutations. Moreover, if the targeted pathway is fundamentally essential to the bacterial metabolism then infectious colonies are less likely to undergo viable mutations. Such targets from *P. aeruginosa* are the glutaredoxins. Glutaredoxins are component of thiol-disulfide glutaredoxins systems of bacteria that favour reducing conditions for the correct disulfide bonding of the functional protein and therefore they are employed by bacteria to defend against oxidative stress imposed by host (Norambuena et al., 2012). Hence glutaredoxins are considered as potential drug targets. An NMR-led FBDD campaign targeted the *P. aeruginosa* glutaredoxin (*PaGRX*) (Khatti et al., 2020). This work involved: 1) generating *PaGRX*-specific fragment hits by screening 463 fragment molecules using independent STD NMR measurements, 2) NMR-based modeling of *PaGRX*, 3) NMR-guided docking of hits and 4) derivatising the fragment hit with a vinyl cysteine trap moiety (acrylamide warhead with strong tendency to form alkylated cysteine adducts) to generate the chimeric lead (Figure 6). The authors show that mM to  $\mu$ M selectivity can be achieved for a few fragments against orthologous proteins and

the promising fragment can be optimised to enhance the selectivity by choosing an appropriate warhead. For instance, fragment 24 which binds to *PaGRX* with moderate specificity ( $K_d = 0.51 \pm 0.37$  and  $LE = 0.32$ ), when coupled with acrylic acid warhead showed enhanced specificity towards the enzyme. This can be further developed to result in to thiol-transferase inhibitory drug candidate against *P. aeruginosa*.

### Targetting Effector Proteins of Type III Secretion System of *P. aeruginosa*

Type III secretion system (T3SS) plays a pivotal role in the virulence and development of antimicrobial resistance of *P. aeruginosa* by providing a diverse range of virulence factors. Readers of this review are referred to a recent review by Gertrudis Horna and Joaquim Ruiz (Horna and Ruiz, 2021b) for more details on the T3SS machinery of *P. aeruginosa*. T3SS is aimed to inject the effectors in host-cells, subverting cellular machinery and neutralising the host immune responses, thereby enhancing bacterial survival rates in a hostile environment within the macrophages. Hence, T3SS impairment opens up opportunities for developing antimicrobial agents to combat *P. aeruginosa* infections avoiding antimicrobial pressure on this and other microorganisms (Aburto-Rodríguez et al., 2021; Horna and



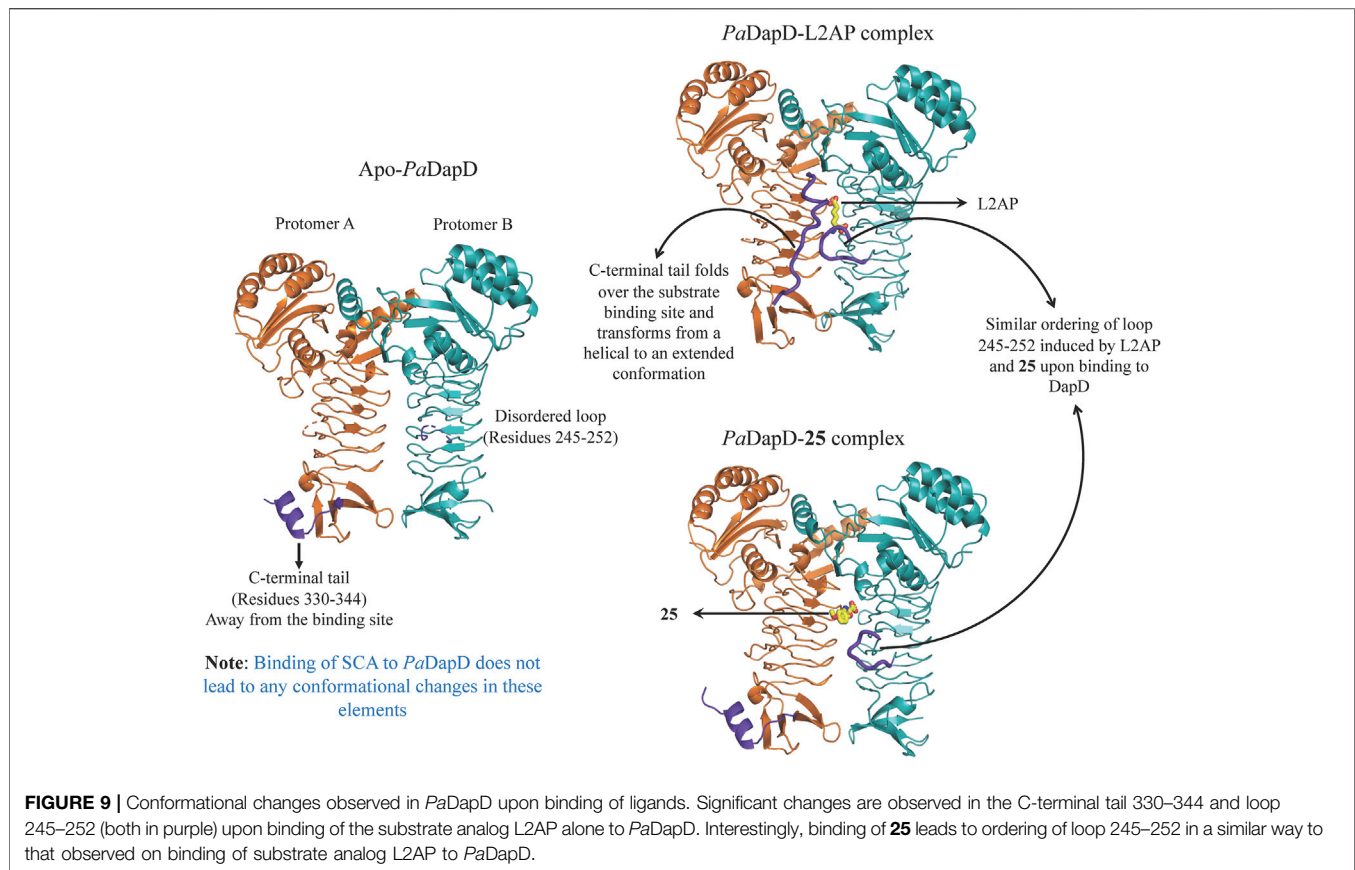


**FIGURE 8 |** FBDD against PaDapD. **(A)** Binding site and fragment hotspot maps of PaDapD. The yellow region in the zoomed view represents the hydrophobic map while hydrogen-bond acceptor and hydrogen-bond donors are represented in red and blue, respectively. **(B)** ligands from various PaDapD-ligand complexes mapped on the structure of PaDapD-25 complex. 25 binds at two different sites in the protomer-protomer interfaces of PaDapD. 25 at site A overlaps with the binding site of the substrate analog L2AP, while that at site B overlaps with the binding site of the cofactor SCA.

Ruiz, 2021a). Until now six effector proteins (ExoS, ExoT, ExoU, ExoY, PemA, PemB) have been reported to be encoded by the T3SS of *P. aeruginosa* (Hueck, 1998; Barbieri and Sun, 2004; Hauser, 2009; Burstein et al., 2015), while four new putative effectors have recently been proposed (Zelikman et al., 2020). Of these, ExoU, ExoS, ExoT and ExoY have been studied most extensively. Drug development other than FBDD campaign have produced small molecule inhibitors of ExoU (Lee et al., 2007; Foulkes et al., 2021), ExoS (Arnoldo et al., 2008) and an inhibitor leading to decrease in secretion of ExoT and ExoY (Sheremet et al., 2018). Most of the components are yet to be investigated for inhibitor development. Thus, there is scope for the development of small molecule inhibitors using FBDD targeting these and other unexplored effectors which could interfere with these effectors and enhance the internalisation of *P. aeruginosa* infections by macrophages.

## FRAGMENT-BASED DRUG DISCOVERY AGAINST *P. AERUGINOSA* AT UNIVERSITY OF CAMBRIDGE

In 2017, United Kingdom Cystic Fibrosis Trust announced a £10 million research partnership with the University of Cambridge to create the first United Kingdom Cystic Fibrosis Innovation Hub with an aim to develop life-changing new treatments for people with CF. The CF innovation hub is hosted by Department of Medicine and led by Professor Andres Floto. It harnesses a multidisciplinary approach with the aim of delivering new treatments for bacterial infections, chronic inflammation, and lung repair for patients with CF. The Arif and Blundell contribution to the United Kingdom Cystic Fibrosis Innovation Hub at the University of Cambridge is to develop new antibiotics against *P. aeruginosa* and *M. abscessus*. We use computational approaches to understand the gene products that



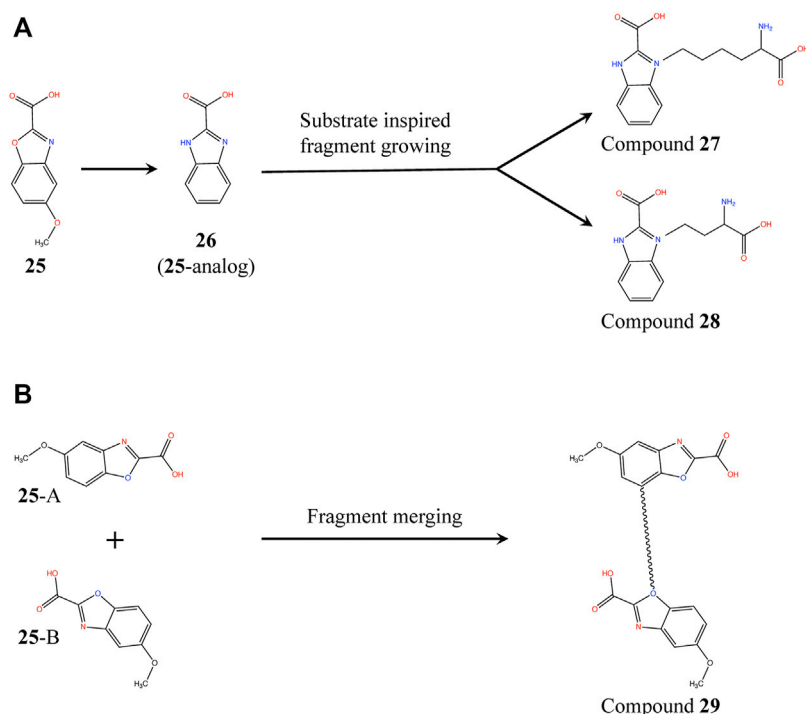
might be appropriate targets by modelling three-dimensional structures (Šali and Blundell, 1993), often of multiprotein assemblies, understanding the impacts of mutations (Pires et al., 2014; Pandurangan et al., 2017; Pandurangan and Blundell, 2020) and assessing the essentiality of the gene. We then use structure-guided fragment-based drug discovery, pioneered in Cambridge in Astex in 1999 (see review Blundell et al., 2002) (Blundell et al., 2002), to develop candidate molecules in our biochemistry and structural biology laboratories and in collaboration with medicinal chemists in the group founded by Professor Chris Abell in Cambridge. Arif and Blundell also work very closely with Prof. Andres Floto, Research Director of the Cambridge Centre for Lung Infection at Papworth Hospital, Cambridge. His team carries out *in vivo* biological assays to test for the effects of elaborated compounds (inhibitors) on the cell system.

Funded as part of the Cystic Fibrosis Hub we began to focus on targets from *P. aeruginosa* and *M. abscessus*. A number of considerations for target selection such as gene essentiality *in vitro* and *in vivo* (Liberati et al., 2006; Skurnik et al., 2013; Lee et al., 2015; Turner et al., 2015; Basta et al., 2017), for example having little or no sequence identity to human or gut microbiome counterparts, the availability of apo and liganded crystal structures that suggest good ligandability, cellular localization of the protein, feasibility of functional/enzymatic assays for the target, were applied to prepare a list of targets from *P. aeruginosa* with ranking based on associated target selection scores (Figure 7).

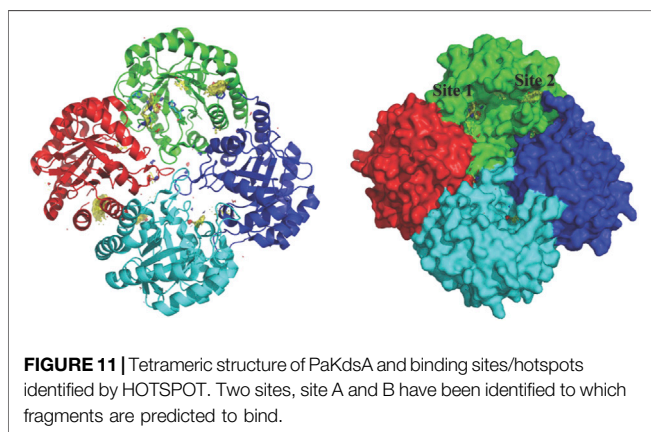
## Targeting DapD: An Essential Protein of the Lysine Biosynthesis Pathway

Bacterial genetic studies have suggested that the lysine biosynthesis pathway is essential and offers several potential antibacterial enzyme targets that could be explored (Cox et al., 2000; Hutton et al., 2007; Gillner et al., 2013). DapD (2,3,4,5-tetrahydropyridine-2,6-dicarboxylate *N*-succinyltransferase), a product of an essential gene involved in the lysine biosynthesis pathway, was selected as an initial target. *P. aeruginosa* DapD (*PaDapD*) was ranked at the very top in the list of potential targets. The lack of human homolog of DapD and maximum identity of 36% to a homolog from human gut microbiome qualifies it for a potential target as this would likely avoid any cross reactivity of the developed inhibitors. DapD catalyses the conversion of cyclic tetrahydrodipicolinate (THDP) into the acyclic *N*-succinyl-*L*-2-amino-6-oxopimelate using succinyl-CoA (SCA) as cofactor. Detailed structural analyses of the available complexes with substrate and cofactors suggest that a long narrow crevice is formed at each interface of the *PaDapD* trimer where the reactants bind (Schnell et al., 2012). Analysis using the HOTSPOT server (50) strongly suggests that this binding site is suitable for the binding of fragments (Figure 8A). We speculated that elaborated small molecules that bind to these sites will potentially block the binding of substrate/cofactor and will eventually inactivate the enzyme.





**FIGURE 10 |** Fragment growing and fragment merging strategies based on the structure of *PaDapD*-**25** and *PaDapD*-L2AP complexes. **(A)** substrate-inspired growing of an **25**-analog (**26**) based on the structures of *PaDapD*-**26** and *PaDapD*-L2AP complexes resulting in compound **27** and **28**. The feasibility of the fragment growing is enhanced when performed on **26**, **(B)** Merging of **25** at sites A and B of *PaDapD* employing suitable linker based on the structure of *PaDapD*-**25** complex to develop compound **29**. The binding and inhibition of these compounds to *PaDapD* needs to be investigated.

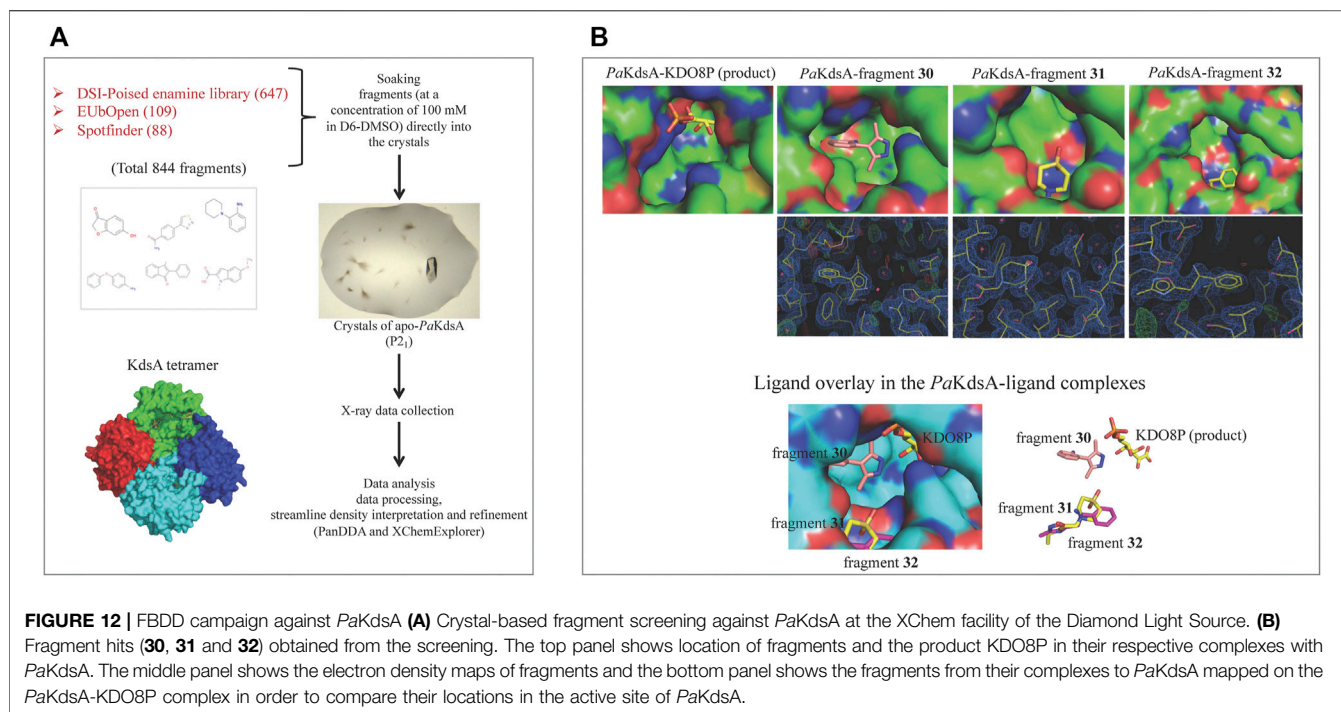


An initial fragment screening employing the differential scanning fluorimetry (DSF) using our in-house library of 960 fragments resulted in 10 promising compounds that could bind to the protein, as all led to a positive shift in the thermal melt temperature of the protein. We then produced crystal structures of apo-*PaDapD* and binary complexes of *PaDapD* with fragment **25**, the cofactor succinyl CoA (SCA); with the substrate analogue L-2-amino pimelic acid (L2AP). SCA binds in the narrow pocket at the interface formed between the N-terminal, middle and C-terminal

part of the neighbouring protomers. L2AP binds at the N-terminal region of the pocket near succinyl  $\beta$ -mercaptoethylamine moiety of SCA (**Figure 8B**). **25** binds at a location (**25-A**) coinciding with the L2AP binding site. At higher concentrations of **25** a second binding site (**25-B**) is observed that overlaps with part of the SCA binding site (**Figure 8B**). The structure of *PaDapD*-L2AP complex differs from the apo-*PaDapD* complex in having the loop (residues 245–252) ordered and interacting with the ligand (**Figure 9**). Another large change is observed in the C-terminal tail region, which, unlike in the apo-*PaDapD* and *PaDapD*-SCA complexes where it was helical, is extended and covers the active site resulting in a cage like structure of the binding pocket (**Figure 9**). A comparison of the binary complexes of L2AP and **25** demonstrates a similarity in the conformational changes observed upon their binding to the active site (**Figure 9**), which might guide the design of inhibitors against the enzyme.

Proximity of **25** and substrate analog L2AP suggests a structure-guided substrate-inspired growing of the fragment (**Figure 10A**) in order to potentiate it for binding to *PaDapD*. Another strategy could be to merge the fragments on the basis of the two binding sites observed in the *PaDapD*-**25** complex structure using an appropriate length of linker (**Figure 10B**). The former could be achieved by first synthesizing an analog (**26**) to which substituents could be chemically added. We had planned to perform the syntheses with the help of our chemist collaborator (Dr J. Mayol-Llinas and Prof. Chris Abell) but these have been interrupted by the untimely death





of Professor Chris Abell. The plan remains that the resulting compounds (compounds 27, 28 and 29) will then be tested for the binding to *PaDapD* using DSF, ITC and SPR.

## Targeting KdsA: An Enzyme in Lipopolysaccharide Biosynthesis Pathway

As LPS has an important role in the structural integrity of the bacterium and its defence against the host, the enzymes of the LPS biosynthesis pathway are attractive drug targets. Four of them KdsA, KdsB, LpxC and WaaG were selected for homology modeling with the aim of developing inhibitors against them. Subsequently, the essentiality of KdsA for the survival and virulence of the *P. aeruginosa* was validated experimentally (Perumal et al., 2011).

KdsA, a 3-deoxy-D-manno-octulosonate 8-phosphate (KDO8P) synthase, catalyses the condensation reaction between D-arabinose 5-phosphate (A5P) and phosphoenolpyruvate (PEP). This enzymatic reaction plays an essential role in the synthesis and assembly process of lipopolysaccharides of most Gram-negative bacteria and is therefore an attractive target for the design of novel antibacterial drugs. The crystal structures of *E. coli* KdsA as binary complexes with the substrate, PEP, and with a mechanism-based inhibitor ( $K_d = 0.4 \mu\text{M}$ ), gave insight about its mechanism of action and inhibition. Interestingly, KdsA belongs to a family of PEP-utilizing enzymes, two of which, UDPGlcNAc enolpyruvoyl transferase (MurZ) and 5-enolpyruvoylshikimate-3-phosphate synthase (EPSPS), are targeted by the antibiotic Fosfomycin and by the herbicide glyphosate, respectively. The crystal structure of *P. aeruginosa*

KdsA (*PaKdsA*) in its binary complex with phosphoenolpyruvate (PEP) was reported in the year 2013 (Nelson et al., 2013). *PaKdsA* is a tetrameric enzyme and the binding site/hotspot of the *P. aeruginosa* KdsA was probed by generating fragment hotspot maps using the *HOTSPOT* program (50), which suggested the presence of promising pocket for targeting new therapeutics (Figure 11).

A fragment library of 480 non-redundant fragments was used for screening against *PaKdsA* employing DSF. 56 fragments showed negative shift in melting temperature while 14 fragments showed positive  $\Delta T_m$  shift. Unfortunately, none of them showed any binding to *PaKdsA* as determined using ITC and X-ray crystallography.

The protein was also used to produce crystals of apo protein for a crystal-based fragment screening, and in complex with the substrates, PEP and A5P, in order to locate the active site which could be used for a virtual screening campaign. The attempt with PEP resulted in *PaKdsA*-PEP complex and in the presence of A5P resulted in *PaKdsA*-A5P complex. When the two substrates were mixed and used for cocrystallization, a product (KDO8P) bound crystal, *PaKdsA*-KDO8P, was obtained. PEP binds deep in the binding pocket and A5P binds above PEP in a more solvent exposed location. KDO8P in its complex with *PaKdsA* spans the location of both PEP and A5P.

A high-throughput crystal-based fragment screening against *PaKdsA* was carried out at the XChem facility of Diamond Light Source (Figure 12A) using a total of 886 crystals, 42 for solvent characterisation and 844 for soaking fragments from three different fragment libraries, namely, DSI poised (Enamine) (647 fragments), EUBOpen (109 fragments) and Spotfinder (88 fragments). 17 hits were obtained after this screening,

most of the fragments were bound at the sites generated by the crystal lattice while three fragments (**30**, **31** and **32**) were bound near the active site of the enzyme (**Figure 12B**). The fragments binding near the active site could be starting points for hit-to-lead compound development against PaKdsA.

## CONCLUSION

Our review of structure-guided fragment-based drug discovery to target *P. aeruginosa* infections demonstrates its great potential in the design of new medicines to combat infections in the lungs of cystic fibrosis patients. However, although *P. aeruginosa* is a common strict aerobic bacterium that can cause disease in plants and animals, including humans, research towards the development of molecules to target it, using fragment-based approaches, remains at an early stage. Nevertheless, as our review demonstrates, the studies are encouraging, although small in number. In the future, further investment will be required in taking leads through the clinic and this is as always a challenge for genetic diseases, which affect a small percentage of the population, even those that are relatively common in wealthy Western nations of North America and Western Europe. The challenge is exacerbated by the fact that this ubiquitous multidrug-resistant pathogen, *P. aeruginosa*, exhibits advanced antibiotic resistance mechanisms. We remain grateful to organisations such as the Cystic Fibrosis Trust that are giving generous support

to research groups such as ours in moving new candidates into the clinic.

## AUTHOR CONTRIBUTIONS

SA contributed to the conception, writing, and generation of graphical figures in this work. RF and TB contributed to the writing of the article and approved the submitted version.

## FUNDING

TLB, RAF and SMA thank the Cystic Fibrosis Trust for funding support (RG70975) and for support through the Cystic Fibrosis Trust Innovation Hub (RG92393).

## ACKNOWLEDGMENTS

TB, RF and SA thank the Cystic Fibrosis Trust for funding support (PHZJ\_395, RG 70975) and for support through the Cystic Fibrosis Trust Innovation Hub (PHZJ\_585, RG92393). SA and TB thank the editors for the invitation to participate in this research topic "Integration of Structural Biology Data in Lead Drug Discovery and Optimization". SA thanks Prof. Anna K. H. Hirsch for kindly providing the coordinates of the unreleased pdb 7oc7.

## REFERENCES

- Aburto-Rodríguez, N. A., Muñoz-Cázares, N., Castro-Torres, V. A., González-Pedrajo, B., Díaz-Guerrero, M., García-Contreras, R., et al. (2021). Antipathogenic Properties of the Combination of a T3ss Inhibitory Halogenated Pyrrolidone with C-30 Furanone. *Molecules* 26, 7635. doi:10.3390/molecules26247635
- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. D Biol. Cryst.* 66, 213–221. doi:10.1107/S0907444909005295
- Adekoya, O. A., Sjøli, S., Wuxiuer, Y., Bilot, I., Marques, S. M., Santos, M. A., et al. (2015). Inhibition of Pseudolysin and Thermolysin by Hydroxamate-Based MMP Inhibitors. *Eur. J. Med. Chem.* 89, 340–348. doi:10.1016/j.ejmech.2014.10.009
- Agrawal, A., Johnson, S. L., Jacobsen, J. A., Miller, M. T., Chen, L.-H., and Pellecchia, M. (2010). Chelator Fragment Libraries for Targeting Metalloproteinases. Available at: <https://doi.org/10.1002/cmdc.200900516>
- Anantharajah, A., Mingeot-Leclercq, M.-P., and Van Bambeke, F. (2016). Targeting the Type Three Secretion System in *Pseudomonas aeruginosa*. *Trends Pharmacol. Sci.* 37, 734–749. doi:10.1016/j.tips.2016.05.011
- Anderson, M. S., Bull, H. G., Galloway, S. M., Kelly, T. M., Mohan, S., Radika, K., et al. (1993/19865). UDP-N-acetylglucosamine Acyltransferase of *Escherichia coli*. The First Step of Endotoxin Biosynthesis Is Thermodynamically Unfavorable. *J. Biol. Chem.* 268, 19858–19865. doi:10.1016/s0021-9258(19)36592-5
- Arnoldo, A., Curak, J., Kittanakom, S., Chevelev, I., Lee, V. T., Sahebol-Amri, M., et al. (2008). Identification of Small Molecule Inhibitors of *Pseudomonas aeruginosa* Exoenzyme S Using a Yeast Phenotypic Screen. *Plos Genet.* 4, e1000005. doi:10.1371/journal.pgen.1000005
- Asmari, M., Ratih, R., Alhazmi, H. A., and El Deeb, S. (2018). Thermophoresis for Characterizing Biomolecular Interaction. *Methods* 146, 107–119. doi:10.1016/j.ymeth.2018.02.003
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* 373, 871–876. doi:10.1126/science.abj8754
- Baell, J. B., and Holloway, G. A. (2010). New Substructure Filters for Removal of pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* 53, 2719–2740. doi:10.1021/jm901137j
- Bancet, A., Raingeval, C., Lomberget, T., Le Borgne, M., Guichou, J.-F., and Krimm, I. (2020). Fragment Linking Strategies for Structure-Based Drug Design. *J. Med. Chem.* 63, 11420–11435. doi:10.1021/acs.jmedchem.0c00242
- Barbieri, J. T., and Sun, J. (2004). *Pseudomonas aeruginosa* ExoS and ExoT. *Rev. Physiol. Biochem. Pharmacol.* 152, 79–92. doi:10.1007/s10254-004-0031-7
- Bartha, I., Di Iulio, J., Venter, J. C., and Telenti, A. (2018). Human Gene Essentiality. *Nat. Rev. Genet.* 19, 51–62. doi:10.1038/nrg.2017.75
- Bassetti, M., Vena, A., Croxatto, A., Righi, E., and Guery, B. (2018). How to Manage *Pseudomonas aeruginosa* Infections. *Dic* 7, 1–18. doi:10.7573/dic.212527
- Basta, D. W., Bergkessel, M., and Newman, D. K. (2017). Identification of Fitness Determinants during Energy-Limited Growth Arrest in *Pseudomonas aeruginosa*. *MBio* 8, 1–17. doi:10.1128/mBio.01170-17
- Baurin, N., Aboul-El, F., Barril, X., Davis, B., Drysdale, M., Dymock, B., et al. (2004). Design and Characterization of Libraries of Molecular Fragments for Use in NMR Screening against Protein Targets. *J. Chem. Inf. Comput. Sci.* 44, 2157–2166. doi:10.1021/ci049806z
- Bielska, E., Lucas, X., Czerwonec, A., M. Kasprzak, J., H. Kaminska, K., and M. Bujnicki, J. (2011). Virtual Screening Strategies in Drug Design - Methods and Applications. *bta* 3, 249–264. doi:10.5114/bta.2011.46542
- Blair, H. A. (2021). Sotorasib: First Approval. *Drugs* 81, 1573–1579. doi:10.1007/s40265-021-01574-2
- Blundell, T. L., Jhoti, H., and Abell, C. (2002). High-throughput Crystallography for lead Discovery in Drug Design. *Nat. Rev. Drug Discov.* 1, 45–54. doi:10.1038/nrd706

- Boehm, H.-J., Boehringer, M., Bur, D., Gmuender, H., Huber, W., Klaus, W., et al. (2000). Novel Inhibitors of DNA Gyrase: 3D Structure Based Biased Needle Screening, Hit Validation by Biophysical Methods, and 3D Guided Optimization. A Promising Alternative to Random Screening. *J. Med. Chem.* 43, 2664–2674. doi:10.1021/jm000017s
- Bollag, G., Tsai, J., Zhang, J., Zhang, C., Ibrahim, P., Nolop, K., et al. (2012). Vemurafenib: the First Drug Approved for BRAF-Mutant Cancer. *Nat. Rev. Drug Discov.* 11, 873–886. doi:10.1038/nrd3847
- Boulant, T., Boudehen, Y.-M., Filloux, A., Plesiat, P., Naas, T., and Dortet, L. (2018). Higher Prevalence of PldA, a *Pseudomonas aeruginosa* Trans-kingdom H2-type VI Secretion System Effector, in Clinical Isolates Responsible for Acute Infections and in Multidrug Resistant Strains. *Front. Microbiol.* 9, 1–7. doi:10.3389/fmicb.2018.02578
- Bradley, C. D., Craig, J. M., Jamie, S. S., and Martin, J. S. (2013). Design and Evaluation of the Performance of an NMR Screening Fragment Library. *Aust. J. Chem.* 66, 1465–1472.
- Braun, P., Ockhuijsen, C., Eppens, E., Koster, M., Bitter, W., and Tommassen, J. (2001). Maturation of *Pseudomonas aeruginosa* Elastase. *J. Biol. Chem.* 276, 26030–26035. doi:10.1074/jbc.M007122200
- Burgess, R. R. (2018). A Brief Practical Review of Size Exclusion Chromatography: Rules of Thumb, Limitations, and Troubleshooting. *Protein Expr. Purif.* 150, 81–85. doi:10.1016/j.pep.2018.05.007
- Burstein, D., Satanower, S., Simovitch, M., Belnik, Y., Zehavi, M., Yerushalmi, G., et al. (2015). Novel Type III Effectors in *Pseudomonas aeruginosa*. *MBio* 6, 3–8. doi:10.1128/mBio.00161-15
- Campbell, J. W., and Cronan, J. E. (2001). Bacterial Fatty Acid Biosynthesis: Targets for Antibacterial Drug Discovery. *Annu. Rev. Microbiol.* 55, 305–332. doi:10.1146/annurev.micro.55.1.305
- Capelli, D., Parravicini, C., Pochetti, G., Montanari, R., Temporini, C., Rabuffetti, M., et al. (2020). Surface Plasmon Resonance as a Tool for Ligand Binding Investigation of Engineered GPR17 Receptor, a G Protein Coupled Receptor Involved in Myelination. *Front. Chem.* 7, 1–14. doi:10.3389/fchem.2019.00910
- Cathcart, G. R. A., Quinn, D., Greer, B., Harriott, P., Lynas, J. F., Gilmore, B. F., et al. (2011). Novel Inhibitors of the *Pseudomonas aeruginosa* Virulence Factor LasB: A Potential Therapeutic Approach for the Attenuation of Virulence Mechanisms in Pseudomonal Infection. *Antimicrob. Agents Chemother.* 55, 2670–2678. doi:10.1128/AAC.00776-10
- Chan, D. I., and Vogel, H. J. (2010). Current Understanding of Fatty Acid Biosynthesis and the Acyl Carrier Protein. *Biochem. J.* 430, 1–19. doi:10.1042/BJ20100462
- Chao, M. C., Abel, S., Davis, B. M., and Waldor, M. K. (2016). The Design and Analysis of Transposon Insertion Sequencing Experiments. *Nat. Rev. Microbiol.* 14, 119–128. doi:10.1038/nrmicro.2015.7
- Chavanieu, A., and Pugnère, M. (2016). Developments in SPR Fragment Screening. *Expert Opin. Drug Discov.* 11, 489–499. doi:10.1517/17460441.2016.1160888
- Chen, A. Y., Adamek, R. N., Dick, B. L., Credille, C. V., Morrison, C. N., and Cohen, S. M. (2019). Targeting Metalloenzymes for Therapeutic Intervention. *Chem. Rev.* 119, 1323–1455. doi:10.1021/acs.chemrev.8b00201
- Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., and Blaschke, T. (2018). The Rise of Deep Learning in Drug Discovery. *Drug Discov. Today* 23, 1241–1250. doi:10.1016/j.drudis.2018.01.039
- Chen, I.-J., and Hubbard, R. E. (2009). Lessons for Fragment Library Design: Analysis of Output from Multiple Screening Campaigns. *J. Comput. Aided. Mol. Des.* 23, 603–620. doi:10.1007/s10822-009-9280-5
- Cohen, F., Aggen, J. B., Andrews, L. D., Assar, Z., Boggs, J., Choi, T., et al. (2019). Optimization of LpxC Inhibitors for Antibacterial Activity and Cardiovascular Safety. *ChemMedChem* 14, 1560–1572. doi:10.1002/cmdc.201900287
- Cohen, S. N., Chang, A. C. Y., Boyer, H. W., and Helling, R. B. (1973). Construction of Biologically Functional Bacterial Plasmids *In Vitro*. *Proc. Natl. Acad. Sci.* 70, 3240–3244. doi:10.1073/pnas.70.11.3240
- Collins, P. M., Ng, J. T., Talon, R., Nekrosiute, K., Krojer, T., Douangamath, A., et al. (2017). Gentle, Fast and Effective crystal Soaking by Acoustic Dispensing. *Acta Cryst. Sect D Struct. Biol.* 73, 246–255. doi:10.1107/S205979831700331X
- Cox, R. J., Sutherland, A., and Vederas, J. C. (2000). Bacterial Diaminopimelate Metabolism as a Target for Antibiotic Design. *Bioorg. Med. Chem.* 8, 843–871. doi:10.1016/S0968-0896(00)00044-4
- Crull, M. R., Somayaji, R., Ramos, K. J., Caldwell, E., Mayer-Hamblett, N., Aitken, M. L., et al. (2018). Changing Rates of Chronic *Pseudomonas aeruginosa* Infections in Cystic Fibrosis: A Population-Based Cohort Study. *Clin. Infect. Dis.* 67, 1089–1095. doi:10.1093/cid/ciy215
- Cryz, S. J., Pitt, T. L., Fürer, E., and Germanier, R. (1984). Role of Lipopolysaccharide in Virulence of *Pseudomonas aeruginosa*. *Infect. Immun.* 44, 508–513. doi:10.1128/iai.44.2.508-513.1984
- Cukier, C. D., Hope, A. G., Elamin, A. A., Moynie, L., Schnell, R., Schach, S., et al. (2013). Discovery of an Allosteric Inhibitor Binding Site in 3-Oxo-Acyl-ACP Reductase from *Pseudomonas aeruginosa*. *ACS Chem. Biol.* 8, 2518–2527. doi:10.1021/cb4005063
- Cystic Fibrosis Foundation 2020 Annual Data Report (2021). *Cystic Fibrosis Foundation 2020 Annual Data Report*.
- Cystic Fibrosis Foundation (2017). *Cystic Fibrosis Foundation Patient Registry. 2016 Annual Data Report*. Bethesda, MD: Cystic Fibrosis Foundation.
- 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
- Dangkulwanich, M., Raetz, C. R. H., and Williams, A. H. (2019). Structure Guided Design of an Antibacterial Peptide that Targets UDP-N-Acetylglucosamine Acyltransferase. *Sci. Rep.* 9, 1–10. doi:10.1038/s41598-019-40418-8
- Davis, P. B. (2006). Cystic Fibrosis since 1938. *Am. J. Respir. Crit. Care Med.* 173, 475–482. doi:10.1164/rccm.200505-840OE
- de Souza Neto, L. R., Moreira-Filho, J. T., Neves, B. J., Maidana, R. L. B. R., Guimarães, A. C. R., Furnham, N., et al. (2020). In Silico Strategies to Support Fragment-To-Lead Optimization in Drug Discovery. *Front. Chem.* 8, 1–18. doi:10.3389/fchem.2020.00093
- Dézil, E., Gopalan, S., Tampakaki, A. P., Lépine, F., Padfield, K. E., Saucier, M., et al. (2005). The Contribution of MvfR to *Pseudomonas aeruginosa* Pathogenesis and Quorum Sensing Circuitry Regulation: Multiple Quorum Sensing-Regulated Genes Are Modulated without Affecting lasRI, rhlRI or the Production of N-Acyl-L-Homoserine Lactones. *Mol. Microbiol.* 55, 998–1014. doi:10.1111/j.1365-2958.2004.04448.x
- Diggle, S. P., Winzer, K., Chhabra, S. R., Worrall, K. E., Cámara, M., and Williams, P. (2003). The *Pseudomonas aeruginosa* Quinolone Signal Molecule Overcomes the Cell Density-Dependency of the Quorum Sensing Hierarchy, Regulates Rhl-dependent Genes at the Onset of Stationary Phase and Can Be Produced in the Absence of LasR. *Mol. Microbiol.* 50, 29–43. doi:10.1046/j.1365-2958.2003.03672.x
- Douangamath, A., Powell, A., Fearon, D., Collins, P. M., Talon, R., Krojer, T., et al. (2020). Achieving Efficient Fragment Screening at XChem Facility at Diamond Light Source Published. doi:10.3791/62414
- Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006). CASTp: Computed Atlas of Surface Topography of Proteins with Structural and Topographical Mapping of Functionally Annotated Residues. *Nucleic Acids Res.* 34, W116–W118. doi:10.1093/nar/gkl282
- Edelsbrunner, H., and Mücke, E. P. (1994). Three-Dimensional Alpha Shapes. *ACM Trans. Graph.* 13, 43–72. doi:10.1145/174462.156635
- El Zowalaty, M. E., Al Thani, A. A., Webster, T. J., El Zowalaty, A. E., Schweizer, H. P., Nasrallah, G. K., et al. (2015). *Pseudomonas aeruginosa*: Arsenal of Resistance Mechanisms, Decades of Changing Resistance Profiles, and Future Antimicrobial Therapies. *Future Microbiol.* 10, 1683–1706. doi:10.2217/fmb.15.48
- Elborn, J. S. (2016). Cystic Fibrosis. *The Lancet* 388, 2519–2531. doi:10.1016/S0140-6736(16)00576-6
- Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and Development of Coot. *Acta Crystallogr. D Biol. Cryst.* 66, 486–501. doi:10.1107/S0907444910007493
- Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W., and Jhoti, H. (2016). Twenty Years on: The Impact of Fragments on Drug Discovery. *Nat. Rev. Drug Discov.* 15, 605–619. doi:10.1038/nrd.2016.109
- Erlanson, D. (2021). Fragments in the Clinic. Available at: <https://practicalfragments.blogspot.com/2021/11/fragments-in-clinic-2021-edition.html>.
- Erwin, A. L. (2016). Antibacterial Drug Discovery Targeting the Lipopolysaccharide Biosynthetic Enzyme LpxC. *Cold Spring Harb. Perspect. Med.* 6, a025304. doi:10.1101/cshperspect.a025304



- Eşkazan, A. E. (2021). Asciminib in Chronic Myeloid Leukemia: many Questions Still Remain to Be Answered. *Blood Cancer J.* 11, 10–11. doi:10.1038/s41408-021-00475-7
- Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., et al. (2006). Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinformatics* 15, 1–30. doi:10.1002/0471250953.bi0506s15
- Everett, M. J., and Davies, D. T. (2021). *Pseudomonas aeruginosa* Elastase (LasB) as a Therapeutic Target. *Drug Discov. Today* 26, 2108–2123. doi:10.1016/j.drudis.2021.02.026
- Evers, B., Jastrzebski, K., Heijmans, J. P. M., Grønrum, W., Beijersbergen, R. L., and Bernards, R. (2016). CRISPR Knockout Screening Outperforms shRNA and CRISPRi in Identifying Essential Genes. *Nat. Biotechnol.* 34, 631–633. doi:10.1038/nbt.3536
- Flume, P. A., O'Sullivan, B. P., Robinson, K. A., Goss, C. H., Mogayzel, P. J., Willey-Courand, D. B., et al. (2007). Cystic Fibrosis Pulmonary Guidelines. *Am. J. Respir. Crit. Care Med.* 176, 957–969. doi:10.1164/rccm.200705-664OC
- Foulkes, D. M., McLean, K., Zheng, Y., Sarsby, J., Haneef, A. S., Fernig, D. G., et al. (2021). A Pipeline to Evaluate Inhibitors of the *Pseudomonas aeruginosa* Exotoxin U. *Biochem. J.* 478, 647–668. doi:10.1042/BCJ20200780
- Fullagar, J. L., Garner, A. L., Struss, A. K., Day, J. A., Martin, D. P., Yu, J., et al. (2013). Antagonism of a Zinc Metalloprotease Using a Unique Metal-Chelating Scaffold: Tropolones as Inhibitors of *P. aeruginosa* Elastase. *Chem. Commun.* 49, 3197–3199. doi:10.1039/c3cc41191e
- Galdino, A. C. M., de Oliveira, M. P., Ramalho, T. C., de Castro, A. A., Branquinho, M. H., and Santos, A. L. S. (2019). Anti-Virulence Strategy against the Multidrug-Resistant Bacterial Pathogen *Pseudomonas aeruginosa*: Pseudolysin (Elastase B) as a Potential Druggable Target. *Cpps* 20, 471–487. doi:10.2174/1389203720666190207100415
- Gallagher, L. A., McKnight, S. L., Kuznetsova, M. S., Pesci, E. C., and Manoil, C. (2002). Functions Required for Extracellular Quinolone Signaling by *Pseudomonas aeruginosa*. *J. Bacteriol.* 184, 6472–6480. doi:10.1128/JB.184.23.6472-6480.2002
- Gao, X., Mu, Z., Qin, B., Sun, Y., and Cui, S. (2017). Structure-based Prototype Peptides Targeting the *Pseudomonas aeruginosa* Type VI Secretion System Effector as a Novel Antibacterial Strategy. *Front. Cel. Infect. Microbiol.* 7. doi:10.3389/fcimb.2017.00411
- Gariyban, L., and Avashia, N. (2013). Polymerase Chain Reaction. *J. Invest. Dermatol.* 133, 1–4. doi:10.1038/jid.2013.1
- Garner, A. L., Struss, A. K., Fullagar, J. L., Agrawal, A., Moreno, A. Y., Cohen, S. M., et al. (2012). 3-Hydroxy-1-alkyl-2-methylpyridine-4(1H)-thiones: Inhibition of the *Pseudomonas aeruginosa* Virulence Factor LasB. *ACS Med. Chem. Lett.* 3, 668–672. doi:10.1021/ml300128f
- Gillner, D. M., Becker, D. P., and Holz, R. C. (2013). Lysine Biosynthesis in Bacteria: A Metalloprotease as a Potential Antimicrobial Target. *J. Biol. Inorg. Chem.* 18, 155–163. doi:10.1007/s00775-012-0965-1
- Gimeno, A., Ojeda-Montes, M., Tomás-Hernández, S., Cereto-Massagué, A., Beltrán-Debón, R., Mulero, M., et al. (2019). The Light and Dark Sides of Virtual Screening: What Is There to Know? *Ijms* 20, 1375. doi:10.3390/ijms20061375
- Goodman, C., and McFadden, G. (2008). Fatty Acid Synthesis in Protozoan Parasites: Unusual Pathways and Novel Drug Targets. *Cpd* 14, 901–916. doi:10.2174/138161208784041088
- Ha, U.-H., Wang, Y., and Jin, S. (2003). DsbA of *Pseudomonas aeruginosa* Is Essential for Multiple Virulence Factors. *Infect. Immun.* 71, 1590–1595. doi:10.1128/IAI.71.3.1590-1595.2003
- Hajduk, P. J., Olejniczak, E. T., and Fesik, S. W. (1997). One-dimensional Relaxation- and Diffusion-Edited NMR Methods for Screening Compounds that Bind to Macromolecules. *J. Am. Chem. Soc.* 119, 12257–12261. doi:10.1021/ja9715962
- Halgren, T. A. (2009). Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* 49, 377–389. doi:10.1021/ci800324m
- Hauser, A. R. (2009). The Type III Secretion System of *Pseudomonas aeruginosa*: Infection by Injection. *Nat. Rev. Microbiol.* 7, 654–665. doi:10.1038/nrmicro2199
- Heath, R. J., and Rock, C. O. (2006). Fatty Acid Biosynthesis as a Target for Novel Antibacterials. *Curr. Opin. Investig. Drugs* 5, 146–153.
- Heidrich, J., Sperl, L. E., and Boeckler, F. M. (2019). Embracing the Diversity of Halogen Bonding Motifs in Fragment-Based Drug Discovery-Construction of a Diversity-Optimized Halogen-Enriched Fragment Library. *Front. Chem.* 7. doi:10.3389/fchem.2019.00009
- Held, D., and Kilz, P. (2021). Size-exclusion Chromatography as a Useful Tool for the Assessment of Polymer Quality and Determination of Macromolecular Properties. *Chem. Teach. Int.* 3, 77–103. doi:10.1515/cti-2020-0024
- Heras, B., Shouldice, S. R., Totsika, M., Scanlon, M. J., Schembri, M. A., and Martin, J. L. (2009). DSB Proteins and Bacterial Pathogenicity. *Nat. Rev. Microbiol.* 7, 215–225. doi:10.1038/nrmicro2087
- Hoffer, L., Voitovich, Y. V., Raux, B., Carrasco, K., Muller, C., Fedorov, A. Y., et al. (2018). Integrated Strategy for Lead Optimization Based on Fragment Growing: The Diversity-Oriented-Target-Focused-Synthesis Approach. *J. Med. Chem.* 61, 5719–5732. doi:10.1021/acs.jmedchem.8b00653
- Horna, G., and Ruiz, J. (2021a). Type 3 Secretion System as an Anti-pseudomonal Target. *Microb. Pathogenesis* 155, 104907. doi:10.1016/j.micpath.2021.104907
- Horna, G., and Ruiz, J. (2021b). Type 3 Secretion System of *Pseudomonas aeruginosa*. *Microbiol. Res.* 246, 126719. doi:10.1016/j.micres.2021.126719
- Hoseini, S., and Sauer, M. G. (2015). Molecular Cloning Using Polymerase Chain Reaction, an Educational Guide for Cellular Engineering. *J. Biol. Eng.* 9, 2. doi:10.1186/1754-1611-9-2
- Huang, B., and Schroeder, M. (2006). LIGSITEcsc: Predicting Ligand Binding Sites Using the Connolly Surface and Degree of Conservation. *BMC Struct. Biol.* 6, 11–19. doi:10.1186/1472-6807-6-19
- Hubbard, R., Davis, B., Chen, L., and Drysdale, M. (2007). The SeeDs Approach: Integrating Fragments into Drug Discovery. *Ctmc* 7, 1568–1581. doi:10.2174/156802607782341109
- Hubbard, R. E., and Murray, J. B. (2011). *Experiences in Fragment-Based lead Discovery*. 1st ed. Elsevier, 509–531. doi:10.1016/B978-0-12-381274-2.00020-0
- Hueck, C. J. (1998). Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants. *Microbiol. Mol. Biol. Rev.* 62, 379–433. doi:10.1128/mmbr.62.2.379-433.1998
- Hutton, C. A., Perugini, M. A., and Gerrard, J. A. (2007). Inhibition of Lysine Biosynthesis: An Evolving Antibiotic Strategy. *Mol. Biosyst.* 3, 458–465. doi:10.1039/b705624a
- Ibrahim, D., Jabbour, J.-F., and Kanj, S. S. (2020). Current Choices of Antibiotic Treatment for *Pseudomonas aeruginosa* Infections. *Curr. Opin. Infect. Dis.* 33, 464–473. doi:10.1097/QCO.0000000000000677
- Irwin, J. J., and Shoichet, B. K. (2005). ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* 45, 177–182. doi:10.1021/ci049714+
- Jackman, J. E., Fierke, C. A., Tumey, L. N., Pirrung, M., Uchiyama, T., Tahir, S. H., et al. (2000). Antibacterial Agents that Target Lipid A Biosynthesis in Gram-Negative Bacteria. *J. Biol. Chem.* 275, 11002–11009. doi:10.1074/jbc.275.15.11002
- Jander, G., Rahme, L. G., and Ausubel, F. M. (2000). Positive Correlation between Virulence of *Pseudomonas aeruginosa* Mutants in Mice and Insects. *J. Bacteriol.* 182, 3843–3845. doi:10.1128/JB.182.13.3843-3845.2000
- Jenkins, R. J., and Dotson, G. D. (2012). Dual Targeting Antibacterial Peptide Inhibitor of Early Lipid A Biosynthesis. *ACS Chem. Biol.* 7, 1170–1177. doi:10.1021/cb300094a
- Jenkins, R. J., Heslip, K. A., Meagher, J. L., Stuckey, J. A., and Dotson, G. D. (2014). Structural Basis for the Recognition of Peptide RJPXD33 by Acyltransferases in Lipid A Biosynthesis. *J. Biol. Chem.* 289, 15527–15535. doi:10.1074/jbc.M114.564278
- Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H. M., Duhr, S., Baaske, P., et al. (2014). MicroScale Thermophoresis: Interaction Analysis and beyond. *J. Mol. Struct.* 1077, 101–113. doi:10.1016/j.molstruc.2014.03.009
- Joo, S. H. (2015). Lipid A as a Drug Target and Therapeutic Molecule. *Biomolecules Ther.* 23, 510–516. doi:10.4062/biomolther.2015.117
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2
- Kalinin, D. V., and Holl, R. (2017). LpxC Inhibitors: a Patent Review (2010–2016). *Expert Opin. Ther. Patents* 27, 1227–1250. doi:10.1080/13543776.2017.1360282
- Kany, A. M., Sikandar, A., Haupenthal, J., Yahiaoui, S., Maurer, C. K., Proschak, E., et al. (2018a). Binding Mode Characterization and Early *In Vivo* Evaluation of Fragment-like Thiols as Inhibitors of the Virulence Factor LasB from *Pseudomonas aeruginosa*. *ACS Infect. Dis.* 4, 988–997. doi:10.1021/acscinfdis.8b00010



- Kany, A. M., Sikandar, A., Yahiaoui, S., Hauptenthal, J., Walter, I., Empting, M., et al. (2018b). Tackling *Pseudomonas aeruginosa* Virulence by a Hydroxamic Acid-Based LasB Inhibitor. *ACS Chem. Biol.* 13, 2449–2455. doi:10.1021/acscchembio.8b00257
- Kaya, C., Walter, I., Yahiaoui, S., Sikandar, A., Alhayek, A., Konstantinović, J., et al. (2021). Substrate-Inspired Fragment Merging and Growing Affords Efficacious LasB Inhibitors. *Angew. Chem. Int. Ed.* 61, 1–6. doi:10.1002/anie.202112295
- Khattari, R. B., Morris, D. L., Bilinovich, S. M., Manandhar, E., Napper, K. R., Sweet, J. W., et al. (2020). Identifying Ortholog Selective Fragment Molecules for Bacterial Glutaredoxins by NMR and Affinity Enhancement by Modification with an Acrylamide Warhead. *Molecules* 25, https://www.mdpi.com/1420-3049/25/1/147. doi:10.3390/molecules25010147
- King, J. D., Kocincová, D., Westman, E. L., and Lam, J. S. (2009). Review: Lipopolysaccharide Biosynthesis in *Pseudomonas Aeruginosa*. *Innate Immun.* 15, 261–312. doi:10.1177/1753425909106436
- Koch, G. (2017). Composing Compound Libraries for Hit Discovery – Rationality-Driven Preselection or Random Choice by Structural Diversity? *Chimia (Aarau)* 71, 643. doi:10.2307/j.ctvncw0d0.18
- Konstantinović, J., Yahiaoui, S., Alhayek, A., Hauptenthal, J., Schöner, E., Andreas, A., et al. (2020). N-Aryl-3-mercaptosuccinimides as Antivirulence Agents Targeting *Pseudomonas aeruginosa* Elastase and *Clostridium* Collagenases. *J. Med. Chem.* 63, 8359–8368. doi:10.1021/acscimedchem.0c00584
- Kowalska-Krochmal, B., and Dudek-Wicher, R. (2021). The Minimum Inhibitory Concentration of Antibiotics: Methods, Interpretation, Clinical Relevance. *Pathogens* 10 (2), 165. https://www.mdpi.com/2076-0817/10/2/165 doi:10.3390/pathogens10020165
- Krasowski, A., Muthas, D., Sarkar, A., Schmitt, S., and Brenk, R. (2011). DrugPred: A Structure-Based Approach to Predict Protein Druggability Developed Using an Extensive Nonredundant Data Set. *J. Chem. Inf. Model.* 51, 2829–2842. doi:10.1021/ci200266d
- Krause, K. M., Haglund, C. M., Hebner, C., Serio, A. W., Lee, G., Nieto, V., et al. (2019). Potent LpxC Inhibitors with *In Vitro* Activity against Multidrug-Resistant *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* 63, e00977–19. doi:10.1128/AAC.00977-19
- Kristam, R., and Gosu, R. (2021). “Selection and Identification of Fragment Library,” in *Methods for Fragments Screening Using Surface Plasmon Resonance*. Editors S. M. Zaheer, and R. Gosu, 53–57. doi:10.1007/978-981-16-1536-8\_7
- Kroec, K. G., Sacco, M. D., Smith, E. W., Zhang, X., Shoun, D., Akhtar, A., et al. (2019). Discovery of Dual-Activity Small-Molecule Ligands of *Pseudomonas aeruginosa* LpxA and LpxD Using SPR and X-ray Crystallography. *Sci. Rep.* 9, 1–12. doi:10.1038/s41598-019-51844-z
- Krojer, T., Talon, R., Pearce, N., Collins, P., Douangamath, A., Brandao-Neto, J., et al. (2017). The XChem Explorer Graphical Workflow Tool for Routine or Large-Scale Protein-Ligand Structure Determination. *Acta Cryst. Sect. D Struct. Biol.* 73, 267–278. doi:10.1107/S2059798316020234
- Kuzmanic, A., Bowman, G. R., Juarez-Jimenez, J., Michel, J., and Gervasio, F. L. (2020). Investigating Cryptic Binding Sites by Molecular Dynamics Simulations. *Acc. Chem. Res.* 53, 654–661. doi:10.1021/acs.accounts.9b00613
- Lavecchia, A. (2015). Machine-learning Approaches in Drug Discovery: Methods and Applications. *Drug Discov. Today Today* 20, 318–331. doi:10.1016/j.drudis.2014.10.012
- Leach, A. R. (2006). RSC Biomolecular Sciences RSC Publishing Presents an Excerpt from : Structure-Based Drug Discovery an Overview. *Mol. Biosyst.* 2, 429–446.
- Lee, J., Wu, J., Deng, Y., Wang, J., Wang, C., Wang, J., et al. (2013). A Cell-Cell Communication Signal Integrates Quorum Sensing and Stress Response. *Nat. Chem. Biol.* 9, 339–343. doi:10.1038/nchembio.1225
- Lee, S. A., Gallagher, L. A., Thongdee, M., Staudinger, B. J., Lippman, S., Singh, P. K., et al. (2015). General and Condition-specific Essential Functions of *Pseudomonas Aeruginosa*. *Proc. Natl. Acad. Sci. USA* 112, 5189–5194. doi:10.1073/pnas.1422186112
- Lee, V. T., Pukatzki, S., Sato, H., Kikawada, E., Kazimirova, A. A., Huang, J., et al. (2007). Pseudolipase A Is a Specific Inhibitor for Phospholipase A 2 Activity of *Pseudomonas aeruginosa* Cytotoxin ExoU. *Infect. Immun.* 75, 1089–1098. doi:10.1128/IAI.01184-06
- Liang, S., Thomas, S. E., Chaplin, A. K., Hardwick, S. W., Chirgadze, D. Y., and Blundell, T. L. (2022). Structural Insights into Inhibitor Regulation of the DNA Repair Protein DNA-PKcs. *Nature* 601, 643–648. doi:10.1038/s41586-021-04274-9
- Liberati, N. T., Urbach, J. M., Miyata, S., Lee, D. G., Drenkard, E., Wu, G., et al. (2006). An Ordered, Nonredundant Library of *Pseudomonas aeruginosa* Strain PA14 Transposon Insertion Mutants. *Proc. Natl. Acad. Sci.* 103, 2833–2838. doi:10.1073/pnas.0511100103
- Linke, P., Amaning, K., Maschberger, M., Vallee, F., Steier, V., Baaske, P., et al. (2016). An Automated Microscale Thermophoresis Screening Approach for Fragment-Based Lead Discovery. *J. Biomol. Screen.* 21, 414–421. doi:10.1177/1087057115618347
- Long, F., Nicholls, R. A., Emsley, P., Gražulis, S., Merkys, A., Vaitkus, A., et al. (2017). AceDRG: A Stereochemical Description Generator for Ligands. *Acta Cryst. Sect. D Struct. Biol.* 73, 112–122. doi:10.1107/S2059798317000067
- Lovering, F., Bikker, J., and Humblet, C. (2009). Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical success. *J. Med. Chem.* 52, 6752–6756. doi:10.1021/jm901241e
- Lu, C., Maurer, C. K., Kirsch, B., Steinbach, A., and Hartmann, R. W. (2014). Overcoming the Unexpected Functional Inversion of a PqsR Antagonist in *Pseudomonas Aeruginosa*: An *In Vivo* Potent Antivirulence Agent Targeting pqs Quorum Sensing. *Angew. Chem. Int. Ed.* 53, 1109–1112. doi:10.1002/anie.201307547
- Manos, J. (2021). Current and Emerging Therapies to Combat Cystic Fibrosis Lung Infections. *Microorganisms* 9, 1874. doi:10.3390/microorganisms9091874
- Maple, H. J., Garlish, R. A., Rigau-Roca, L., Porter, J., Whitcombe, I., Prosser, C. E., et al. (2012). Automated Protein-Ligand Interaction Screening by Mass Spectrometry. *J. Med. Chem.* 55, 837–851. doi:10.1021/jm201347k
- Mashalidis, E. H., Sledz, P., Lang, S., and Abell, C. (2013). A Three-Stage Biophysical Screening cascade for Fragment-Based Drug Discovery. *Nat. Protoc.* 8, 2309–2324. doi:10.1038/nprot.2013.130
- Mayer, M., and Meyer, B. (1999). Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int. Ed.* 38, 1784–1788. doi:10.1002/(sici)1521-3773(19990614)38:12<1784::aid-anie1784>3.0.co;2-q
- Melville, J., Burke, E., and Hirst, J. (2009). Machine Learning in Virtual Screening. *Cchts* 12, 332–343. doi:10.2174/138620709788167980
- Miyake, Y., Itoh, Y., Hatanaka, A., Suzuma, Y., Suzuki, M., Kodama, H., et al. (2019). Identification of Novel Lysine Demethylase 5-selective Inhibitors by Inhibitor-Based Fragment Merging Strategy. *Bioorg. Med. Chem.* 27, 1119–1129. doi:10.1016/j.bmc.2019.02.006
- Mohanty, B., Rimmer, K., McMahon, R. M., Headey, S. J., Vazirani, M., Shouldice, S. R., et al. (2017). Fragment Library Screening Identifies Hits that Bind to the Non-catalytic Surface of *Pseudomonas aeruginosa* DsbA1. *PLoS One* 12, e0173436–20. doi:10.1371/journal.pone.0173436
- Molecular Operating Environment (2016). Integrated Computer-Aided Molecular Design Platform. Available at: https://www.chemcomp.com/Products.htm.
- Molina, S. A., and Hunt, W. R. (2017). Cystic Fibrosis: An Overview of the Past, Present, and the Future. *Elsevier Inc.*, 219–249. doi:10.1016/B978-0-12-803809-3.00012-9
- Morgens, D. W., Deans, R. M., Li, A., and Bassik, M. C. (2016). Systematic Comparison of CRISPR/Cas9 and RNAi Screens for Essential Genes. *Nat. Biotechnol.* 34, 634–636. doi:10.1038/nbt.3567
- Moynie, L., Schnell, R., McMahon, S. A., Sandalova, T., Boulkerou, W. A., Schmidberger, J. W., et al. (2013). The AEROPATH Project targeting *Pseudomonas Aeruginosa*: Crystallographic Studies for Assessment of Potential Targets in Early-Stage Drug Discovery. *Acta Cryst. Sect. F* 69, 25–34. doi:10.1107/S1744309112044739
- Mueller, A. M., Breitsprecher, D., Duhr, S., Baaske, P., Schubert, T., and Längst, G. (2017). MicroScale Thermophoresis: A Rapid and Precise Method to Quantify Protein-Nucleic Acid Interactions in Solution. *Methods Mol. Biol.* 1654, 151–164. doi:10.1007/978-1-4939-7231-9\_10
- Mullis, K. B. (1990). The Unusual Origin of the Polymerase Chain Reaction. *Sci. Am.* 262, 56–65. doi:10.1038/scientificamerican0490-56
- Murray, C. W., and Rees, D. C. (2009). The Rise of Fragment-Based Drug Discovery. *Nat. Chem* 1, 187–192. doi:10.1038/nchem.217
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., et al. (2011). REFMAC5 for the Refinement of Macromolecular crystal Structures. *Acta Crystallogr. D Biol. Cryst.* 67, 355–367. doi:10.1107/S0907444911001314
- Müsken, M., Di Fiore, S., Dötsch, A., Fischer, R., and Häussler, S. (2010). Genetic Determinants of *Pseudomonas aeruginosa* Biofilm Establishment. *Microbiology* 156, 431–441. doi:10.1099/mic.0.033290-0
- Navratilova, I., and Hopkins, A. L. (2010). Fragment Screening by Surface Plasmon Resonance. *ACS Med. Chem. Lett.* 1, 44–48. doi:10.1021/ml900002k
- Nelson, S. K., Kelleher, A., Robinson, G., Reiling, S., and Asojo, O. A. (2013). Structure of 2-Keto-3-Deoxy-D-Manno-Octulosonate-8-Phosphate Synthase

- from *Pseudomonas Aeruginosa*. *Acta Cryst. Sect F* 69, 1084–1088. doi:10.1107/S1744309113023993
- Neumann, T., Junker, H.-D., Schmidt, K., and Sekul, R. (2007). SPR-based Fragment Screening: Advantages and Applications. *Ctmc* 7, 1630–1642. doi:10.2174/156802607782341073
- Niesen, F. H., Berglund, H., and Vedadi, M. (2007). The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions that Promote Protein Stability. *Nat. Protoc.* 2, 2212–2221. doi:10.1038/nprot.2007.321
- Nikiforov, P. O., Surade, S., Blaszczyk, M., Delorme, V., Brodin, P., Baulard, A. R., et al. (2016). A Fragment Merging Approach towards the Development of Small Molecule Inhibitors of *Mycobacterium tuberculosis* EthR for Use as Ethionamide Boosters. *Org. Biomol. Chem.* 14, 2318–2326. doi:10.1039/c5ob02630j
- Nishino, N., and Powers, J. C. (1980). *Pseudomonas aeruginosa* Elastase. Development of a New Substrate, Inhibitors, and an Affinity Ligand. *J. Biol. Chem.* 255, 3482–3486. doi:10.1016/s0021-9258(19)85724-1
- Norambuena, J., Flores, R., Cárdenas, J. P., Quatrini, R., Chávez, R., and Levican, G. (2012). Thiol/Disulfide System Plays a Crucial Role in Redox Protection in the Acidophilic Iron-Oxidizing Bacterium *Leptospirillum Ferriphilum*. *PLoS One* 7, e44576. doi:10.1371/journal.pone.0044576
- O'Loughlin, C. T., Miller, L. C., Sityaporn, A., Drescher, K., Semmelhack, M. F., and Bassler, B. L. (2013). A Quorum-sensing Inhibitor Blocks *Pseudomonas aeruginosa* Virulence and Biofilm Formation. *Proc. Natl. Acad. Sci.* 110, 17981–17986. doi:10.1073/pnas.1316981110
- Pan, P., Tonge, P., and Pan, P. (2012). Targeting InhA, the FASII Enoyl-ACP Reductase: SAR Studies on Novel Inhibitor Scaffolds. *Ctmc* 12, 672–693. doi:10.2174/156802612799984535
- Pandurangan, A. P., and Blundell, T. L. (2020). Prediction of Impacts of Mutations on Protein Structure and Interactions: SDM, a Statistical Approach, and mCSM, Using Machine Learning. *Protein Sci.* 29, 247–257. doi:10.1002/pro.3774
- Pandurangan, A. P., Ochoa-Montañón, B., Ascher, D. B., and Blundell, T. L. (2017). SDM: A Server for Predicting Effects of Mutations on Protein Stability. *Nucleic Acids Res.* 45, W229–W235. doi:10.1093/nar/gkx439
- Pang, Z., Raudonis, R., Glick, B. R., Lin, T.-J., and Cheng, Z. (2019). Antibiotic Resistance in *Pseudomonas aeruginosa*: Mechanisms and Alternative Therapeutic Strategies. *Biotechnol. Adv.* 37, 177–192. doi:10.1016/j.biotechadv.2018.11.013
- Papaioannou, E., Utari, P., and Quax, W. (2013). Choosing an Appropriate Infection Model to Study Quorum Sensing Inhibition in *Pseudomonas* Infections. *Ijms* 14, 19309–19340. doi:10.3390/ijms140919309
- Parsons, J. B., and Rock, C. O. (2011). Is Bacterial Fatty Acid Synthesis a Valid Target for Antibacterial Drug Discovery? *Curr. Opin. Microbiol.* 14, 544–549. doi:10.1016/j.mib.2011.07.029
- Patel, D., Bauman, J. D., and Arnold, E. (2014). Advantages of Crystallographic Fragment Screening: Functional and Mechanistic Insights from a Powerful Platform for Efficient Drug Discovery. *Prog. Biophys. Mol. Biol.* 116, 92–100. doi:10.1016/j.pbiomolbio.2014.08.004
- Pearce, N. M., Krojer, T., Bradley, A. R., Collins, P., Nowak, R. P., Talon, R., et al. (2017). A Multi-crystal Method for Extracting Obscured Crystallographic States from Conventionally Uninterpretable Electron Density. *Nat. Commun.* 8, 24–29. doi:10.1038/ncomms15123
- Perera, T. P. S., Jovcheva, E., Mevellec, L., Vialard, J., De Lange, D., Verhulst, T., et al. (2017). Discovery and Pharmacological Characterization of JNJ-42756493 (Erdafitinib), a Functionally Selective Small-Molecule FGFR Family Inhibitor. *Mol. Cancer Ther.* 16, 1010–1020. doi:10.1158/1535-7163.MCT-16-0589
- Perumal, D., Lim, C. S., and Sakharkar, M. K. (2007). “In Silico Identification of Putative Drug Targets in *Pseudomonas aeruginosa* through Metabolic Pathway Analysis,” in *Pattern Recognition in Bioinformatics. PRIB 2007. Lecture Notes in Computer Science*. Editors J. C. Rajapakse, B. Schmidt, and G. Volkert (Berlin, Heidelberg: Springer), 4774, 323–336. doi:10.1007/978-3-540-75286-8\_31
- Perumal, D., Sakharkar, K. R., Tang, T. H., Chow, V. T. K., Lim, C. S., Samal, A., et al. (2010). Cloning and Targeted Disruption of Two Lipopolysaccharide Biosynthesis Genes, *kdsA* and *waaG*, of *Pseudomonas aeruginosa* PAO1 by Site-Directed Mutagenesis. *J. Mol. Microbiol. Biotechnol.* 19, 169–179. doi:10.1159/000322157
- Pires, D. E. V., Ascher, D. B., and Blundell, T. L. (2014). MCSM: Predicting the Effects of Mutations in Proteins Using Graph-Based Signatures. *Bioinformatics* 30, 335–342. doi:10.1093/bioinformatics/btt691
- Potron, A., Poirel, L., and Nordmann, P. (2015). Emerging Broad-Spectrum Resistance in *Pseudomonas aeruginosa* and *Acinetobacter Baumannii*: Mechanisms and Epidemiology. *Int. J. Antimicrob. Agents* 45, 568–585. doi:10.1016/j.ijantimicag.2015.03.001
- Radoux, C. J., Olsson, T. S. G., Pitt, W. R., Groom, C. R., and Blundell, T. L. (2016). Identifying Interactions that Determine Fragment Binding at Protein Hotspots. *J. Med. Chem.* 59, 4314–4325. doi:10.1021/acs.jmedchem.5b01980
- Raetz, C. R. H., and Whitfield, C. (2002). Lipopolysaccharide Endotoxins. *Annu. Rev. Biochem.* 71, 635–700. doi:10.1146/annurev.biochem.71.110601.135414
- Rainard, J. M., Pandarakalam, G. C., and McElroy, S. P. (2018). Using Microscale Thermophoresis to Characterize Hits from High-Throughput Screening: A European Lead Factory Perspective. *SLAS DISCOVERY: Advancing Sci. Drug Discov.* 23, 225–241. doi:10.1177/2472555217744728
- Ranjbar, M., Behrouz, B., Norouzi, F., and Mousavi Gargari, S. L. (2019). Anti-PcrV IgY Antibodies Protect against *Pseudomonas aeruginosa* Infection in Both Acute Pneumonia and Burn Wound Models. *Mol. Immunol.* 116, 98–105. doi:10.1016/j.molimm.2019.10.005
- Roberts, R. J. (2005). How Restriction Enzymes Became the Workhorses of Molecular Biology. *Proc. Natl. Acad. Sci.* 102, 5905–5908. doi:10.1073/pnas.0500923102
- Rock, C. O., and Jackowski, S. (2002). Forty Years of Bacterial Fatty Acid Synthesis. *Biochem. Biophysical Res. Commun.* 292, 1155–1166. doi:10.1006/bbrc.2001.2022
- Rodriguez, E. L., Poddar, S., Iftekhar, S., Suh, K., Woolfork, A. G., Ovbude, S., et al. (2020). Affinity Chromatography: A Review of Trends and Developments over the Past 50 Years. *J. Chromatogr. B* 1157, 122332. doi:10.1016/j.jchromb.2020.122332
- Rotella, D. P. (1997). Antibacterial Agents that Inhibit Lipid A Biosynthesis. *Chemtracts* 10, 665–668.
- Saint-Criq, V., Villeret, B., Bastaert, F., Kheir, S., Hatton, A., Cazes, A., et al. (2018). *Pseudomonas aeruginosa* LasB Protease Impairs Innate Immunity in Mice and Humans by Targeting a Lung Epithelial Cystic Fibrosis Transmembrane regulator-IL-6-antimicrobial-repair Pathway. *Thorax* 73, 49–61. doi:10.1136/thoraxjnl-2017-210298
- Šali, A., and Blundell, T. L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* 234, 779–815.
- Schnell, R., Oehlmann, W., Sandalova, T., Braun, Y., Huck, C., Maringer, M., et al. (2012). Tetrahydridipicolinate N-Succinyltransferase and Dihydridipicolinate Synthase from *Pseudomonas aeruginosa*: Structure Analysis and Gene Deletion. *PLoS One* 7, e31133. doi:10.1371/journal.pone.0031133
- Schuffenhauer, A., Ruedisser, S., Marzinzik, A., Jahnke, W., Selzer, P., and Jacoby, E. (2005). Library Design for Fragment Based Screening. *Ctmc* 5, 751–762. doi:10.2174/1568026054637700
- Scoffin, R., and Slater, M. (2015). The Virtual Elaboration of Fragment Ideas : Growing , Merging and Linking Fragments with Realistic Chemistry Drug Discovery , Development & Delivery. *Int. Pharm. Ind.* 7, 2–5.
- Scott, D. E., Coyne, A. G., Hudson, S. A., and Abell, C. (2012). Fragment-based Approaches in Drug Discovery and Chemical Biology. *Biochemistry* 51, 4990–5003. doi:10.1021/bi3005126
- Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., et al. (2020). Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* 577, 706–710. doi:10.1038/s41586-019-1923-7
- Senisterra, G., Chau, I., and Vedadi, M. (2012). Thermal Denaturation Assays in Chemical Biology. *ASSAY Drug Development Tech.* 10, 128–136. doi:10.1089/adt.2011.0390
- Senturk, S., Shirole, N. H., Nowak, D. G., Corbo, V., Pal, D., Vaughan, A., et al. (2017). Rapid and Tunable Method to Temporally Control Gene Editing Based on Conditional Cas9 Stabilization. *Nat. Commun.* 8, 1–10. doi:10.1038/ncomms14370
- Shao, X., Xie, Y., Zhang, Y., Liu, J., Ding, Y., Wu, M., et al. (2020). Novel Therapeutic Strategies for Treating *Pseudomonas aeruginosa* Infection. *Expert Opin. Drug Discov.* 15, 1403–1423. doi:10.1080/17460441.2020.1803274
- Shen, S., and Kozikowski, A. P. (2016). Why Hydroxamates May Not Be the Best Histone Deacetylase Inhibitors-What Some May Have Forgotten or Would rather Forget? *ChemMedChem* 11, 15–21. doi:10.1002/cmdc.201500486
- Sheremet, A. B., Zigangirova, N. A., Zayakin, E. S., Luyksaar, S. I., Kapotina, L. N., Nesterenko, L. N., et al. (2018). Small Molecule Inhibitor of Type Three Secretion System Belonging to a Class 2,4-Disubstituted-4h-[1,3,4]-Thiadiazine-5-Ones Improves Survival and Decreases Bacterial Loads in an Airway *Pseudomonas aeruginosa* Infection in Mice. *Biomed. Res. Int.* 2018, 1–13. doi:10.1155/2018/5810767
- Shrivastava, S., Shrivastava, P., and Ramasamy, J. (2018). World Health Organization Releases Global Priority List of Antibiotic-Resistant Bacteria to

- Guide Research, Discovery, and Development of New Antibiotics. *J. Med. Soc.* 32, 76–77. doi:10.4103/jms.jms.25.17
- Silver, L. L. (2007). Multi-targeting by Monotherapeutic Antibacterials. *Nat. Rev. Drug Discov.* 6, 41–55. doi:10.1038/nrd2202
- Singh, M., Tam, B., and Akabayov, B. (2018). NMR-fragment Based Virtual Screening: A Brief Overview. *Molecules* 23, 233. doi:10.3390/molecules23020233
- Skurnik, D., Roux, D., Aschard, H., Cattoir, V., Yoder-Himes, D., Lory, S., et al. (2013). A Comprehensive Analysis of *In Vitro* and *In Vivo* Genetic Fitness of *Pseudomonas aeruginosa* Using High-Throughput Sequencing of Transposon Libraries. *Plos Pathog.* 9, e1003582. doi:10.1371/journal.ppat.1003582
- Sohn, M.-J., Zheng, C.-J., and Kim, W.-G. (2008). Macrolactin S, a New Antibacterial Agent with Fab G-Inhibitory Activity from *Bacillus* Sp. AT28. *J. Antibiot.* 61, 687–691. doi:10.1038/ja.2008.98
- Souers, A. J., Levenson, J. D., Boghaert, E. R., Ackler, S. L., Catron, N. D., Chen, J., et al. (2013). ABT-199, a Potent and Selective BCL-2 Inhibitor, Achieves Antitumor Activity while Sparing Platelets. *Nat. Med.* 19, 202–208. doi:10.1038/nm.3048
- Srivastava, V. K., and Yadav, R. (2019). Isothermal Titration Calorimetry. *Elsevier Inc.* 125–137. doi:10.1016/B978-0-12-816548-5.00009-5
- Storz, M. P., Maurer, C. K., Zimmer, C., Wagner, N., Brengel, C., De Jong, J. C., et al. (2012). Validation of PqsD as an Anti-biofilm Target in *Pseudomonas aeruginosa* by Development of Small-Molecule Inhibitors. *J. Am. Chem. Soc.* 134, 16143–16146. doi:10.1021/ja3072397
- Tap, W. D., Wainberg, Z. A., Anthony, S. P., Ibrahim, P. N., Zhang, C., Healey, J. H., et al. (2015). Structure-Guided Blockade of CSF1R Kinase in Tenosynovial Giant-Cell Tumor. *N. Engl. J. Med.* 373, 428–437. doi:10.1056/NEJMoa1411366
- Tasdemir, D., Lack, G., Brun, R., Rüedi, P., Scapozza, L., and Perozzo, R. (2006). Inhibition of *Plasmodium falciparum* Fatty Acid Biosynthesis: Evaluation of FabG, FabZ, and FabI as Drug Targets for Flavonoids. *J. Med. Chem.* 49, 3345–3353. doi:10.1021/jm0600545
- Thomann, A., De Mello Martins, A. G. G., Brengel, C., Empting, M., and Hartmann, R. W. (2016). Application of Dual Inhibition Concept within Looped Autoregulatory Systems toward Antiviral Agents against *Pseudomonas aeruginosa* Infections. *ACS Chem. Biol.* 11, 1279–1286. doi:10.1021/acschembio.6b00117
- Turner, K. H., Wessel, A. K., Palmer, G. C., Murray, J. L., and Whiteley, M. (2015). Essential Genome of *Pseudomonas aeruginosa* in Cystic Fibrosis Sputum. *Proc. Natl. Acad. Sci. USA* 112, 4110–4115. doi:10.1073/pnas.1419677112
- UK Cystic Fibrosis Registry 2020 Annual Data Report (2021). *UK Cystic Fibrosis Registry 2020 Annual Data Report. Registry.*
- Urban, A., Leipelt, M., Eggert, T., and Jaeger, K.-E. (2001). DsbA and DsbC Affect Extracellular Enzyme Formation in *Pseudomonas aeruginosa*. *J. Bacteriol.* 183, 587–596. doi:10.1128/JB.183.2.587-596.2001
- Van Delden, C., and Igilewski, B. H. (1998). <http://www.cdc.gov/ncidod/eid/vol4no4/vandelden.htm>. *Emerg. Infect. Dis.* 4, 551–560. doi:10.3201/eid0404.980405
- Van Opijnen, T., and Camilli, A. (2013). Transposon Insertion Sequencing: A New Tool for Systems-Level Analysis of Microorganisms. *Nat. Rev. Microbiol.* 11, 435–442. doi:10.1038/nrmicro3033
- Walls, D., and Walker, J. M. (2017). Protein Chromatography. *Protein Chromatogr.* 1485, 423. doi:10.1007/978-1-4939-6412-3
- Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., et al. (2015). Identification and Characterization of Essential Genes in the Human Genome. *Science* 350, 1096–1101. doi:10.1126/science.aac7041
- Wang, X., Choe, Y., Craik, C. S., and Ellman, J. A. (2002). Design and Synthesis of Novel Inhibitors of Gelatinase B. *Bioorg. Med. Chem. Lett.* 12, 2201–2204. doi:10.1016/S0960-894X(02)00365-7
- Wartchow, C. A., Podlaski, F., Li, S., Rowan, K., Zhang, X., Mark, D., et al. (2011). Biosensor-based Small Molecule Fragment Screening with Biolayer Interferometry. *J. Comput. Aided. Mol. Des.* 25, 669–676. doi:10.1007/s10822-011-9439-8
- Wei, W., Cherukupalli, S., Jing, L., Liu, X., and Zhan, P. (2020). Fsp3: A New Parameter for Drug-Likeness. *Drug Discov. Today* 25, 1839–1845. doi:10.1016/j.drudis.2020.07.017
- Who (2017). WHO Priority Pathogens List for R&D of New Antibiotics. Available at: <https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed>.
- Wickramasinghe, S. R., Inglis, K. A., Urch, J. E., Müller, S., van Aalten, D. M. F., and Fairlamb, A. H. (2006). Kinetic, Inhibition and Structural Studies on 3-Oxoacyl-ACP Reductase from *Plasmodium falciparum*, a Key Enzyme in Fatty Acid Biosynthesis. *Biochem. J.* 393, 447–457. doi:10.1042/BJ20050832
- Williams, A. H., Immormino, R. M., Gewirth, D. T., and Raetz, C. R. H. (2006). Structure of UDP-N-Acetylglucosamine Acyltransferase with a Bound Antibacterial Pentadecapeptide. *Proc. Natl. Acad. Sci.* 103, 10877–10882. doi:10.1073/pnas.0604465103
- Winter, G. (2010). Xia2: An Expert System for Macromolecular Crystallography Data Reduction. *J. Appl. Cryst.* 43, 186–190. doi:10.1107/S0021889809045701
- Wretling, B., and Pavlovskis, O. R. (1983). *Pseudomonas aeruginosa* Elastase and its Role in *Pseudomonas* Infections. *Rev. Infect. Dis.* 5 (Suppl. 5), S998–S1004. doi:10.1093/clinids/5.supplement\_5.s998
- Yadav, S. P., Bergqvist, S., Doyle, M. L., Neubert, T. A., and Yamniuk, A. P. (2012). MIRG Survey 2011: Snapshot of Rapidly Evolving Label-free Technologies Used for Characterizing Molecular Interactions. *J. Biomol. Tech.* 23, 94–100. doi:10.7171/jbt.12-2303-002
- Yamada, Y., Takashima, H., Walmsley, D. L., Ushiyama, F., Matsuda, Y., Kanazawa, H., et al. (2020). Fragment-Based Discovery of Novel Non-hydroxamate LpxC Inhibitors with Antibacterial Activity. *J. Med. Chem.* 63, 14805–14820. doi:10.1021/acs.jmedchem.0c01215
- Yamaotsu, N., and Hirono, S. (2018). In Silico fragment-mapping Method: a New Tool for Fragment-Based/structure-Based Drug Discovery. *J. Comput. Aided. Mol. Des.* 32, 1229–1245. doi:10.1007/s10822-018-0160-8
- Yang, C.-Y., and Wang, S. (2010). Computational Analysis of Protein Hotspots. *ACS Med. Chem. Lett.* 1, 125–129. doi:10.1021/ml100026a
- Yang, L., Nilsson, M., Gjermansen, M., Givskov, M., and Tolker-Nielsen, T. (2009). Pyoverdine and PQS Mediated Subpopulation Interactions Involved in *Pseudomonas aeruginosa* Biofilm Formation. *Mol. Microbiol.* 74, 1380–1392. doi:10.1111/j.1365-2958.2009.06934.x
- You, Y., Ramachandra, S. G., and Jin, T. (2020). A CRISPR-Based Method for Testing the Essentiality of a Gene. *Sci. Rep.* 10, 1–8. doi:10.1038/s41598-020-71690-8
- Zelikman, S., Aizler, Y., Dudkevitch, R., Shoshani, S., Korenblit, S., Burstein, D., et al. (2020). Identification and Characterization of New *Pseudomonas aeruginosa* Type III Secretion System Effectors. 9th ILANIT/FISEB Conf, Eilat, Israel, February 17–20, 2020. Available at: <https://program.eventact.com/Agenda/Lecture/206800?code=4266825>
- Zhang, F., Luo, S.-Y., Ye, Y.-B., Zhao, W.-H., Sun, X.-G., Wang, Z.-Q., et al. (2008). The Antibacterial Efficacy of an Aceraceous Plant [Shantung maple (*Acer truncatum* Bunge)] May Be Related to Inhibition of Bacterial  $\beta$ -oxoacyl-acyl Carrier Protein Reductase (FabG). *Biotechnol. Appl. Biochem.* 51, 73. doi:10.1042/ba20070255
- Zhang, Y.-M., Lu, Y.-J., and Rock, C. O. (2004). The Reductase Steps of the Type II Fatty Acid Synthase as Antimicrobial Targets. *Lipids* 39, 1055–1060. doi:10.1007/s11745-004-1330-3
- Zhang, Y.-M., and Rock, C. O. (2004). Evaluation of Epigallocatechin Gallate and Related Plant Polyphenols as Inhibitors of the FabG and FabI Reductases of Bacterial Type II Fatty-Acid Synthase. *J. Biol. Chem.* 279, 30994–31001. doi:10.1074/jbc.M403697200
- Zhu, T., Cao, S., Su, P.-C., Patel, R., Shah, D., Chokshi, H. B., et al. (2013). Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis. *J. Med. Chem.* 56, 6560–6572. doi:10.1021/jm301916b

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

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

Copyright © 2022 Arif, Floto and Blundell. 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.





# Recent Approaches to the Identification of Novel Microtubule-Targeting Agents

Susanna Eli<sup>1†</sup>, Rossella Castagna<sup>2†</sup>, Marina Mapelli<sup>1\*</sup> and Emilio Parisini<sup>2\*</sup>

<sup>1</sup>IEO, European Institute of Oncology IRCCS, Milan, Italy, <sup>2</sup>Latvian Institute of Organic Synthesis, Aizkraukles Iela 21, Riga, Latvia

## OPEN ACCESS

### Edited by:

Marco Nardini,  
University of Milan, Italy

### Reviewed by:

Ranieri Bizzarri,  
University of Pisa, Italy  
Pietro Roversi,  
Italian National Research Council, Italy

### \*Correspondence:

Marina Mapelli  
marina.mapelli@ieo.it  
Emilio Parisini  
emilio.parisini@osi.lv

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

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

**Received:** 22 December 2021

**Accepted:** 21 February 2022

**Published:** 30 March 2022

### Citation:

Eli S, Castagna R, Mapelli M and  
Parisini E (2022) Recent Approaches to  
the Identification of Novel Microtubule-  
Targeting Agents.  
Front. Mol. Biosci. 9:841777.  
doi: 10.3389/fmolb.2022.841777

Microtubules are key components of the eukaryotic cytoskeleton with essential roles in cell division, intercellular transport, cell morphology, motility, and signal transduction. They are composed of protofilaments of heterodimers of  $\alpha$ -tubulin and  $\beta$ -tubulin organized as rigid hollow cylinders that can assemble into large and dynamic intracellular structures. Consistent with their involvement in core cellular processes, affecting microtubule assembly results in cytotoxicity and cell death. For these reasons, microtubules are among the most important targets for the therapeutic treatment of several diseases, including cancer. The vast literature related to microtubule stabilizers and destabilizers has been reviewed extensively in recent years. Here we summarize recent experimental and computational approaches for the identification of novel tubulin modulators and delivery strategies. These include orphan small molecules, PROTACs as well as light-sensitive compounds that can be activated with high spatio-temporal accuracy and that represent promising tools for precision-targeted chemotherapy.

**Keywords:** tubulin drugs, microtubule drugs, chemotherapeutic agents, photopharmacology, photocaged, photoswitch, PROTAC, artificial intelligence

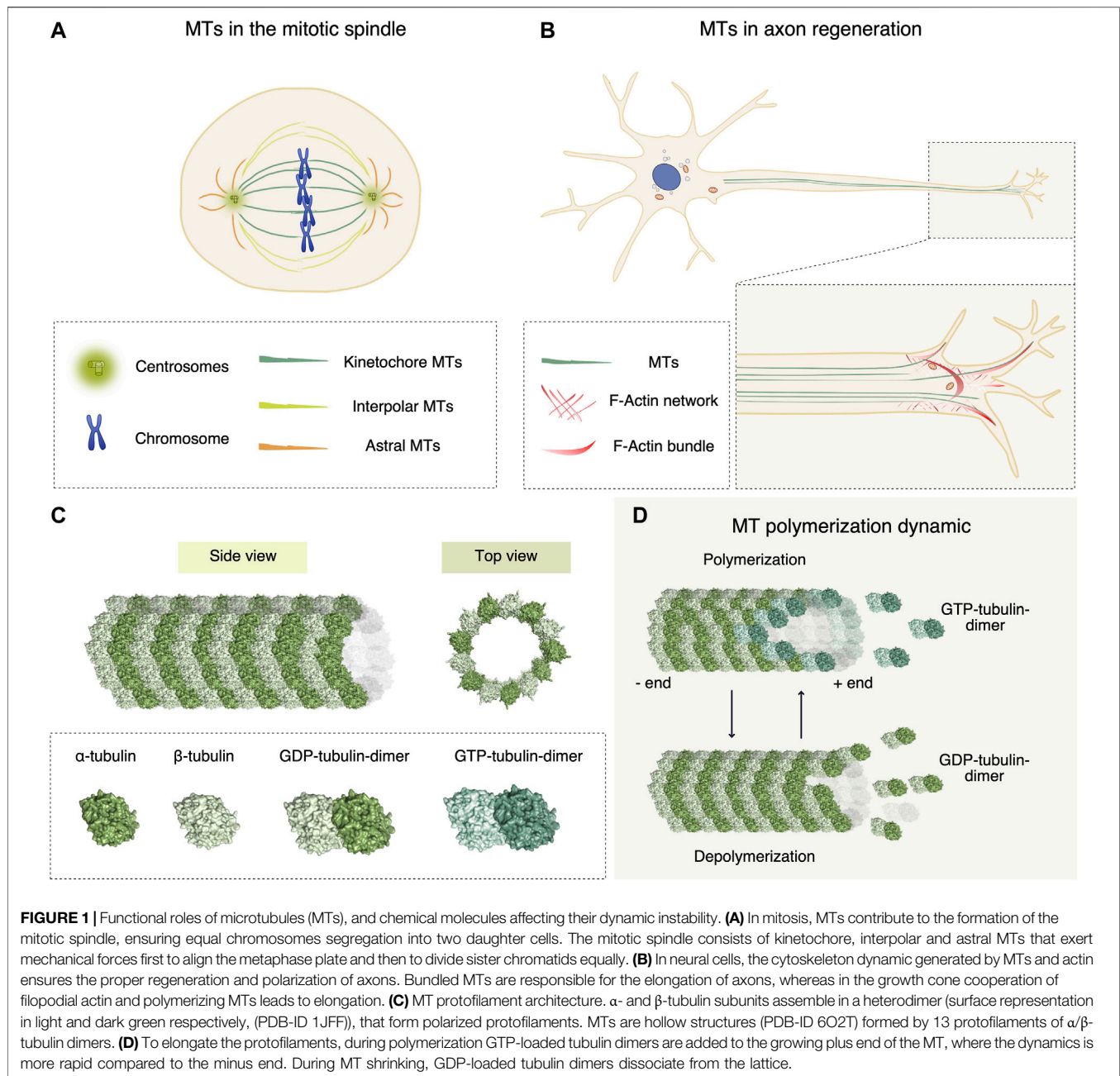
## INTRODUCTION

Microtubules (MTs) play a central role in many biological processes, ranging from cell signaling, cell morphology, to cell movement and division (Dustin, 1978). MTs organize networks that provide the structural scaffolding for vesicular trafficking and organelle positioning (Bonifacino and Neefjes, 2017) by acting as platforms for the motors kinesins and dyneins (Gennerich and Vale, 2009) (Figures 1A,B). In dividing cells, MTs directly contribute to the formation of the mitotic spindle ensuring equal chromosomes segregation in a highly orchestrated process (Rieder and Salmon, 1998).

The modulation of MT polymerization dynamics, also known as dynamic instability, is the major instrument that cells use to regulate all these biological functions. MTs are hollow structures formed by 13 polarized protofilaments of  $\alpha/\beta$ -tubulin dimers (Figure 1C). MT polymerization consists in the addition of  $\alpha/\beta$ -tubulin dimers at the MT ends (Figure 1D) (Waterman-Storer and Salmon, 1997), with rates in the order of minutes in interphase to seconds in mitosis (Saxton et al., 1984). A plethora of MT-binding proteins is crucial for the spatial and temporal regulation of MT dynamics. For this reason, proteins affecting MT organization have been studied as chemotherapeutic agents (Zhou and Giannakakou, 2005).

The first MT regulators have been identified in neurons (Cassimeris and Spittle, 2001). Indeed, the cytoskeleton dynamic generated by MTs and actin ensures the proper growth of neurons (Capizzi et al., 2021; Alushin et al., 2014) and polarization of axons (Blanquie and Bradke, 2018) (Figure 1B). The pharmacological modulation of cytoskeleton dynamics in injured axons may overturn a dystrophic structure to a regenerative one. For example, the MT stabilizing agent Taxol was





proven to facilitate axonal regeneration after spinal cord injury by stimulating axonal growth and reducing fibrotic scarring (Hellal et al., 2011).

MT-targeting agents (MTAs) are valuable effectors for cancer therapy because they interfere with MT polymerization dynamics leading to mitotic arrest and apoptosis (Giannakakou et al., 2000). Many MTAs used in clinics were discovered by large-scale screening of natural compounds derived from bacteria, plants, fungi and sponges (Miller et al., 2018). Mimicking their function, synthetic inhibitors orchestrate MT filament dynamics by binding to tubulin according to their target binding site (Dumontet and Jordan, 2010). At high concentrations MTAs

affect net MTs polymer mass, while at low doses they modulate dynamic instability (Wilson et al., 1999) causing a delay in mitotic entry or a mitotic block, with consequent induction of apoptosis (Steinmetz and Protá, 2018). Interestingly, many pieces of evidence correlate the lethality of MTAs with disruption of mitosis-independent functions such as cell signaling and vesicular trafficking (Komlodi-Pasztor et al., 2011). MTAs may also act synergically with DNA-damaging agents in cancer treatment by inhibiting the correct transport of DNA repair proteins into the nucleus (Poruchynsky et al., 2015). Thus, the molecular effects of each MTA may be very complex to dissect.

## MTAs AND THEIR USE IN THE CLINICS

MTAs are usually classified in two groups, according to their mechanism of action: the microtubule-stabilizing agents (MSAs), which promote their assembly (Dumontet and Jordan, 2010), and the microtubule-destabilizing agents (MDAs), which trigger MTs disassembly in  $\alpha/\beta$ -tubulin oligomers or dimers. The specific binding site and the corresponding mechanism of action of each compound was identified and investigated at the atomic level by crystallographic studies (for a review see (Steinmetz and Prota, 2018)), recently integrated with cryo-EM analysis of entire MT filaments (Alushin et al., 2014). Six different MTA binding sites have been characterized on the  $\alpha/\beta$ -tubulin dimer. The  $\beta$ -tubulin binding sites are the Taxane, Laulimalide, Colchicine, Vinca and Maytansine site. The Pironetin site is the only one located on the  $\alpha$ -tubulin subunit (Usui et al., 2004) (Figure 2A). In the following paragraphs, we will survey the different MTAs, their binding site, mechanism of action and current clinical applications.

### Microtubule-Stabilizing Agents—MSAs Taxane-Site

Since their discovery in the seventies (Wani et al., 1971), Taxanes have been studied extensively. This category includes first-generation anticancer drugs, such as paclitaxel and docetaxel, and new generation molecules, as cabazitaxel that challenge drug resistance (de Bono et al., 2010). The cause of Taxane-resistance has been identified in the overexpression of class III  $\beta$ -tubulin (Mozzetti et al., 2005) and in its affinity for multidrug-resistance proteins, such as P-glycoprotein (Goldstein, 1996). Taxanes bind to  $\beta$ -tubulin in a MT luminal side pocket with polar and hydrophobic interactions, resulting in non-functional MTs that prevent cells from proceeding into anaphase (Figure 2B). They promote MT polymerization by binding to tubulin filaments with particularly high affinity for regions that are relevant for inter-protofilament contacts. Taxanes are widely used for the treatment of ovarian, breast and lung cancer (Ojima et al., 2015). To reduce the toxicity derived from formulations needed to solubilize paclitaxel, the use of nanoparticles of albumin-bound Taxanes has been studied (Stinchcombe, 2007). In addition, Taxanes conjugated with tumor-targeting antibodies (Guillemand and Saragovi, 2001) or vitamins (Russell-Jones et al., 2004) are under development in order to redirect the effect of the drug specifically to the cancer stem cells. Local delivery of Taxol may enhance also the regenerative capacity of injured neurons by inducing efficient axon regrowth in the optic nerve (Sengottuvel et al., 2011).

### Laulimalide and Peloruside-A

Laulimalide and Peloruside-A are part of the MT-stabilizing family. They bind to  $\beta$ -tubulin, facing the lateral interface between tubulin protofilaments, thus inhibiting their disassembly (Figure 2B). These compounds act via two mechanisms: they inhibit interactions between  $\alpha/\beta$ -tubulin dimers and at the same time they strengthen the lateral contacts between protofilaments by stabilizing the Taxane-site M-loop (Prota et al., 2014b). This crosstalk opens the possibility for combinatorial strategies of Laulimalide/Peloruside with

Taxane-site ligands (Khrapunovich-Baine et al., 2011). Interestingly, Laulimalide and Peloruside were found effective against paclitaxel multidrug-resistant cancer cell lines with mutations in the Taxane site and overexpression of P-glycoprotein (Liu et al., 2007).

### Microtubule-Destabilizing Agents - MDAs Colchicine-Site

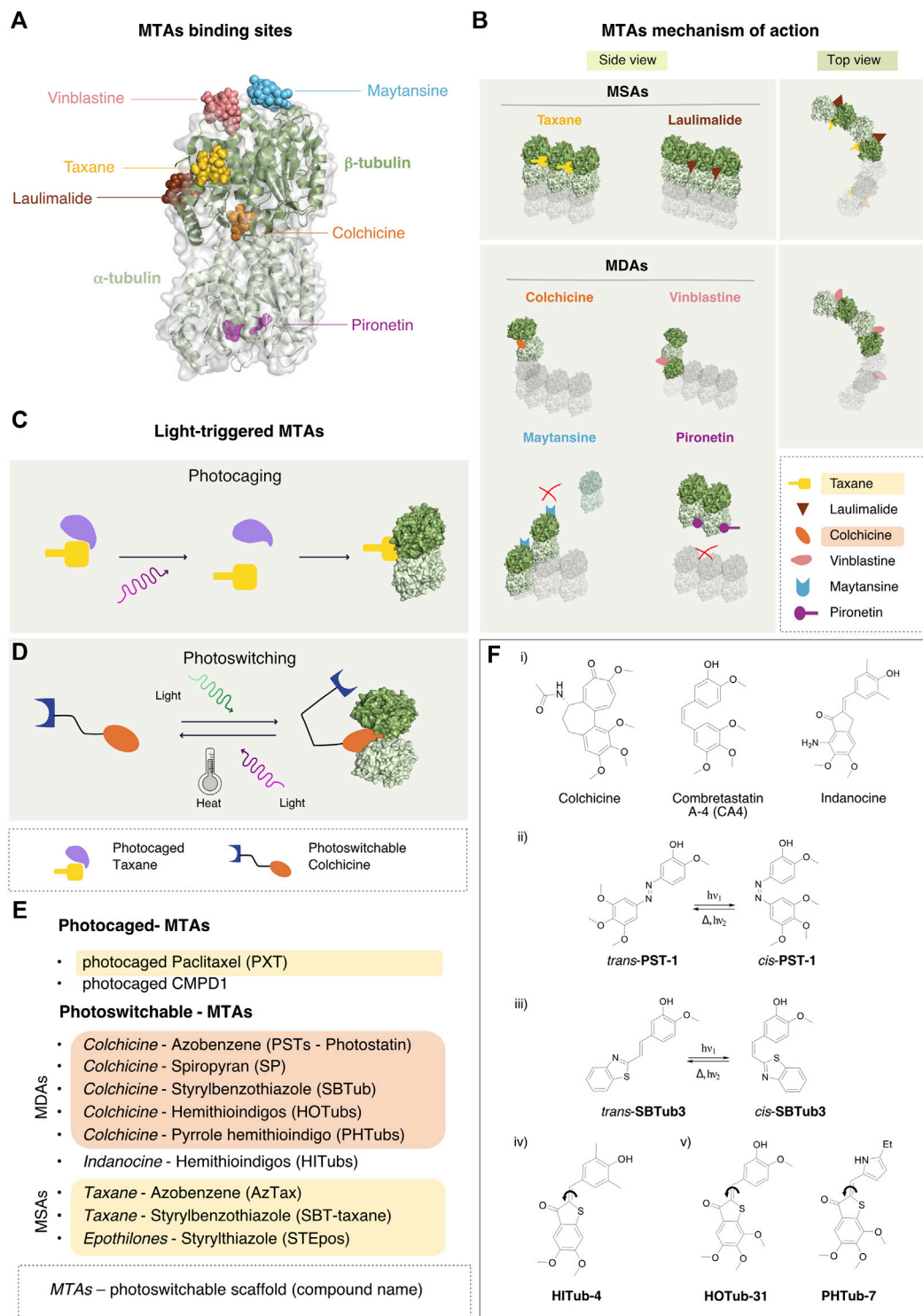
Colchicine was the first MT-destabilizing agent to be discovered (Borisy and Taylor, 1967). It strongly impairs mitotic progression but its use in clinics is limited due to its high toxicity and low therapeutic index. Since the structural characterization of its binding mode (Ravelli et al., 2004), many colchicine-binding site inhibitors (CBSIs) have been developed as anti-inflammatory agents that primarily prevent the activation of the inflammatory bodies (Gasparyan et al., 2015). The binding pocket of these compound is located in the intermediate domain near the  $\alpha/\beta$ -dimer interface (Figures 2A,B). When MTs assemble,  $\alpha/\beta$ -tubulin dimers alter their architecture from a curved to a straight conformation (Nogales et al., 1998). Inhibitors of the colchicine family prevent this conformational change. Podophyllotoxin and Combretastatin bind in the same pocket as colchicine and are past phase II in different clinical trials (Young and Chaplin, 2004). Nocodazole is a reversible inhibitor of this pocket widely used in research as synchronization and anti-mitotic agent.

### Vinca-Site

Compounds that target the Vinca site comprise different molecules of both natural and synthetic origin that destabilize MTs. They include the Vinca alkaloids, vinblastine, vincristine, eribulin, diazonamides and trizolopyrimidines (Moudi et al., 2013), which all bind between two  $\alpha/\beta$ -tubulin dimers, at the longitudinal interface of protofilaments. Their binding site consist in a core pocket extending with additional contacts toward the GTP-binding site of  $\beta$ -tubulin at the plus end of the filament (Ranaivoson et al., 2012) (Figure 2C). Vinca-site compounds hamper the conformational changes of  $\alpha/\beta$ -tubulin dimers necessary for their correct incorporation into MT filaments, and sequester them in ring-like oligomers incompatible with the straight protofilaments lattice (Na and Timasheff, 1980). In addition, Vinca drugs inhibit the hydrolysis of exchangeable GTP, bound to  $\beta$ -tubulin (David-Pfeuty et al., 1979). The working principles of these agents highlight the importance of the correct longitudinal curvature of the  $\alpha/\beta$ -tubulin dimers for MT dynamics, as an additional possible target for further destabilizing mechanisms. Vinca alkaloids were first used in the treatment of childhood hematologic malignancies and after for solid and adult hematologic malignancies (Bohannon et al., 1963). Many semisynthetic analogs as vindesine (Gökbuget and Hoelzer, 1997), vinorelbine and vinflunine, were developed to overcome side effects such as neuropathy and myelosuppression, likely linked to altered axonal transport.

### Maytansine-Site

Maytansine binds on the exposed side of the  $\beta$ -tubulin subunit (Prota et al., 2014a). It works by blocking the longitudinal



**FIGURE 2 |** MTAs binding sites, mechanism of action and the photo-pharmacological approach. **(A)** Binding sites of the major classes of MT stabilizers and destabilizers. Cartoon and surface representation of the  $\alpha/\beta$ -tubulin dimer with the six major sites targeted by MT binding molecules. Most molecules associate with pockets on the  $\beta$ -tubulin subunit, including maytansine (PDB-ID 4TV8, light blue), vinblastine (PDB-ID 5J2T, pink), the Taxane vincadernolide (PDB-ID 5LXT, gold), laulimalide (PDB-ID 4O4H, brown). Colchicine fits at the interface between  $\alpha$ -tubulin and  $\beta$ -tubulin subunits of the hetero-dimer (PDB-ID 4O2B, orange). The only characterized MT destabilizer binding to  $\alpha$ -tubulin is pironetin (PDB-ID 5LA6, purple). **(B)** Mechanism of action of the microtubule-binding agents. Diagram of the mechanisms of action of the MSAs Taxane (gold) and laulimalide (brown): these ligands bind between two adjacent  $\beta$ -tubulin subunits stabilizing the lateral interaction between protofilaments in the MT. In the MDAs category, colchicine (orange) inhibits MT growth by preventing the conformational change of the  $\alpha/\beta$ -tubulin dimer (Continued)

**FIGURE 2 |** required for MT-lattice formation. Vinca-site agents (pink) intercalate between two longitudinally-aligned  $\alpha/\beta$ -tubulin dimers along the MT lattice leading to MT destabilization. Also maytansine agents (light blue) fit on top of the  $\beta$ -tubulin subunit, constraining longitudinal interaction between two  $\alpha/\beta$ -tubulin dimers. Pironetin agents (purple) bind to the  $\alpha$ -tubulin subunit and prevent its binding to the  $\alpha/\beta$ -tubulin dimer in the same protofilament. **(C)** Activation of photocaged compounds. Prior to illumination, the pharmacologically active moiety of the drug is shielded by a photolabile element, which acts as a hindering functionality. Upon illumination, cleavage of the photolabile moiety occurs and the drug becomes irreversibly activated and primed for binding to the target. **(D)** A photoswitchable element in the scaffold of the drug allows a reversible conformational change upon illumination with a specific wavelength of light, thus causing drug activation. The back reaction can either occur spontaneously upon light cessation (thermal activation), or by illumination with a different wavelength of light. **(E)** Existing photoswitchable versions of known MTAs. The name of the original MTA drug is indicated first, followed by the type of scaffold used for the synthesis of the photoswitchable analog and by the name of the corresponding photoswitchable MTA thus obtained. **(F)** Selected examples of microtubule-targeting agents. i) Lead structures of selected microtubule-destabilizing agents used for the development of photoswitchable analogues. ii) *Trans-cis* photoisomerization reaction of photostatin (PST-1) iii) *Trans-cis* photoisomerization reaction of styrylbenzothiazole (SBTub3) iv) Chemical structure of a hemithioindigo-based indanone-like tubulin inhibitor (HITub-4) v) Chemical structure of a hemithioindigo colchicinoid tubulin inhibitor (HOTub-31) and a pyrrole hemithioindigo tubulin inhibitor (PHTub-7).

protofilament interaction by either inhibiting the addition of dimers or by the formation of unfunctional complexes with  $\beta$ -tubulin (**Figures 2A,B**). Differently from Vinca alkaloids, maytansine prevents the correct turn-over of GTP-loaded  $\beta$ -subunits at the MT minus-ends (Lopus et al., 2010). Many companies are focusing on the development of antibody-maytansinoids conjugate (ADC) therapeutics, which suppress MT dynamic instability in mitosis (Oroudjev et al., 2010). Maytansine derivatives are currently in use for metastatic breast cancer (Krop et al., 2012).

### Pironetin-Site

Pironetin and its dimethyl derivatives destabilize MT polymerization and preclude mitotic progression. These ligands fit into a hydrophobic pocket in the  $\alpha$ -tubulin subunit, adjacent to the  $\alpha/\beta$ -tubulin dimers interface (Yang et al., 2016) (**Figures 2A,B**). In this way they promote the disassembly of the MT protofilaments by meddling the longitudinal contacts between dimers. Because of their peculiar binding-site, these compounds may be effective for the treatments of tumors with resistance against  $\beta$ -tubulin-targeting drugs.

Lately, some innovative approaches have provided interesting results and indicated possible ways forward in the discovery of new MTAs. Here, we will briefly review some of the most recent developments in the field.

## PHOTOPHARMACOLOGY: LIGHT-TRIGGERED ANALOGUES OF MTAs

The use of photosensitive drugs, which are activated by specific wavelengths of light, can provide good spatio-temporal control over drug action (Hoorens and Szymanski, 2018; Matera et al., 2018; Fuchter, 2020). Although still at the preclinical stage, photopharmacology may indeed help overcome crucial pharmacological problems in the future and provide alternative therapeutic strategies. Interestingly, existing clinically approved photodynamic therapies, which are based on the use of light in combination with photosensitizing agents to promote the formation of cytotoxic reactive oxygen species (ROS), are considered less-invasive than traditional pharmacological approaches. As such, they are widely

investigated for cancer treatment (Huang, 2005), although they are occasionally affected by such limitations as inefficient deep-tissue penetration, incomplete tumor suppression and the insurgence of side effects from the patient's accidental exposure to strong Sun/indoor light after treatment.

The design of both light-activated compounds and photopharmacological agents for targeted cancer therapies has recently been extensively reviewed with respect to their clinical potential (Vickerman et al., 2021) as well as their use in photodynamic therapy (Zhao et al., 2021) and in photopharmacology (Dunkel and Ilaš, 2021). There are two different types of light-triggerable bioactive compounds: photocaged and photoswitchable. The former are molecules in which the pharmacologically active moiety is shielded by a photolabile element that can be cleaved upon illumination, thus activating the drug irreversibly (**Figure 2C**), while the latter are drugs whose chemical scaffold accommodates a photoswitchable element, whereby the conformation and activity of the drug can be controlled in a reversible manner using light (**Figure 2D**).

### Photocaged Compounds

In photocaged compounds, the covalent attachment of the photolabile element in the starting form of the drug renders the compound inactive. However, when the hindering functionality is cleaved by irradiation with a specific wavelength of light, the drug is primed for binding to its target and exerts its inhibitory function. This activation process, albeit irreversible, allows a good degree of spatio-temporal control of the drug.

The first light-triggered microtubule-targeting agent ever developed was a photocaged version of paclitaxel, activated by nitrogen-pulsed laser in the ultraviolet (UV wavelength at 337 nm) (Buck and Zheng, 2002). Later, to improve cleavage efficiency and to obtain a completely inactive caged paclitaxel that could exert its activity exclusively upon illumination (UV-activated at 360 nm), a double caging strategy was developed (Gropeanu et al., 2012).

A visible-light-activated (430 nm) caged paclitaxel was subsequently developed using a 7-N,N-diethylamino-4-hydroxymethyl coumarin photolabile group (Skwarczynski et al., 2006). Further optimization of this class of photolabile-



paclitaxel (phototaxel, UV-activated at 355 nm) led to compounds that feature greater water solubility (>100 mg/ml) and stability in physiological conditions (Noguchi et al., 2008).

More recently, a photocaged version of the small molecule CMPD1, another tubulin polymerization inhibitor, was synthesized using 4,5-dimethoxy-2-nitrobenzyl (DMNB) as photolabile protecting group (UV-activated at 365 nm). The tubulin inhibitor thus released was shown to be toxic for glioblastoma cells and to lead to apoptotic cell death (Döbber et al., 2017).

## Photoswitchable Compounds

Photoswitchable compounds undergo a conformational change when illuminated by proper wavelengths of light in the UV-vis spectra. However, upon switching conformation, the molecule becomes thermodynamically less stable and the compound is converted back to its thermally stable geometry. This can occur either by spontaneous back reaction on a time scale that ranges from fractions of a second to years, or by illumination with different wavelengths of light.

Photoswitchable drugs are typically designed via a bioisosteric replacement approach, whereby an azobenzene moiety is introduced in the chemical structure of the starting drug to act as photoswitchable unit. This *azologization* process is usually considered a mildly invasive chemical change in the drug scaffold. Alternatively, the azobenzene moiety can be attached to the drug in a lateral position and this approach is called *azoextension*. The aim of both approaches is to favor or to disturb the pharmacological activity of the compound by switching its conformation between the two isomers. Ideally, the goal is to obtain a compound that is pharmacologically inactive in the *trans* form and becomes activated by switching to the *cis* conformation. By implementing an *azologization* strategy on the structure of combretastatin A-4 (CA4), one of the most prominent colchicine binding site inhibitors, Borowiak et al. developed a class of photoswitchable analogs of CA4 named photostatin (PSTs), which are 250 times more active in their *cis* form (obtained by illumination in the range 370–430 nm) than in their *trans* conformation (Borowiak et al., 2015). Further *in vitro* (Engdahl et al., 2015) and *in vivo* (Zenker et al., 2017) studies have confirmed that PSTs are a robust chemical tool for exerting a photoswitchable cytotoxic effect with excellent optical control over microtubule dynamics at the single cell level.

The success of the photoswitchable design of PSTs relies on the structural features of CA4, a stilbenoid molecule that is well suited for an *azologization* approach. Indeed, unlike other CA4 analogues, it retains solubility and biological activity over tubulin upon *azologization* (Tron et al., 2006; Gossiau et al., 2008; Rastogi et al., 2018). Moreover, CA4 binds to the colchicine binding site in a *cis*-like conformation (PDB 5LYJ), which was shown to provide greater stabilization energy than the *trans* form (Gaspari et al., 2017).

Incidentally, azobenzene is not the only scaffold used in photopharmacology. The styrylbenzothiazole (SBT) moiety has been recently used as a scaffold for the design of photoswitchable microtubule inhibitors (Gao et al., 2021b). These compounds have been shown to be metabolically stable and to allow the photocontrol of microtubule dynamics with sub-cellular

precision, photoswitched either by one-photon excitation (LED at 360 nm), or by two-photon excitation (laser at 780 nm). Moreover, they bind to the colchicine binding site in a *cis*-like conformation, similarly to CA4, as shown in the crystal structures of Z-SBTub2 (PDB: 6ZWC) and Z-SBTub3 (PDB: 6ZWB) bound to the tubulin-DARPin D1 complex.

The fact that the patented tubulin-binding anti-cancer drugs are structurally extremely diverse (Haider et al., 2019) provides many interesting opportunities for photopharmacology. Hemithioindigos (HTI) have been used as photoswitches in the design of microtubule binders based on the colchicoid pharmacophore. These compounds, which are cell-compatible and show antimitotic photoswitchable bioactivity, have been named hemithioindigo-colchicinoid tubulin binders (HOTubs, 450 nm) (Sailer et al., 2019). Likewise, the microtubule inhibitor colchicine served as template for the design of pyrrole hemithioindigo (PHTubs) compounds (photoswitched in the range 435–450 nm) (Sailer et al., 2021). Furthermore, colchicine has been used as structural reference for the development of photoswitchable cytotoxic compounds based on spiropyran (SP, exposed to UV light at 365 nm) (Rastogi et al., 2021).

By further optimization of the HTI scaffold, a photoswitchable class of compounds based on tubulin-inhibiting indanones have also been proposed (Sailer et al., 2020). The hemithioindigo-based indanone-like tubulin inhibitors (HITubs) show improved cellular potency relative to HOTubs as well as antimitotic photo-modulated activity (photoswitched with visible light at 450 and 530 nm).

Besides the generation of photoswitchable versions of microtubule destabilizers based on colchicine analogues, the photopharmacology community has also focused on the development of photoswitchable versions of microtubule stabilizers compounds (Figures 2E,F). A photoswitchable version of the microtubule stabilizing drug paclitaxel has been developed using an *azoextension* approach. The 3'-azobenzamide-Taxanes (AzTax) series of compounds proposed allowed the spatio-temporal control of microtubule stabilization in living cells (photoswitched in the range 360–530 nm) (Müller-Deku et al., 2020).

The possibility to use the styrylbenzothiazole (SBT) scaffold to obtain an *SBT-extension* of the Taxane drug was recently investigated. Unfortunately, the SBT-Taxane (SBTax) compounds show poor solubility, low bioactivity and modest photocontrol on the biological process. The *extension* approach has been also applied with styrylthiazole (ST) to epothilones (STEpOs, photoactivated at the 405 nm laser line), which are a class of MSA targeting the same Taxane binding site (Gao et al., 2021a).

Recently, a rationalization of the protocol for handling photosensitive drugs to photocontrol microtubule dynamics in biological assays has been proposed (Thorn-Seshold and Meiring, 2021).

## PROTACs

PROteolysis TArgeting Chimeras (PROTACs) is a technology that holds great promise for overcoming drug resistance problems as it allows the inactivation of the target protein by

inducing its complete degradation rather than its sheer inhibition (Garber, 2021). PROTACs are bifunctional molecules featuring an E3 ubiquitin ligase moiety tethered to a ligand of the target protein of interest via a linker of optimal length. The association of the ligand moiety of the PROTAC with the protein of interest promotes ubiquitination of the target protein and its degradation by the ubiquitin proteasome system (UPS). Recently, an attempt to develop the first tubulin-targeting PROTAC has been made (Gasic et al., 2020). The validity of this approach is corroborated by the observation that a number of compounds that are known to bind covalently to different cysteines on  $\beta$ -tubulin promote tubulin degradation (Yang et al., 2019). The authors designed different degrader molecules, all of them featuring a E3 ubiquitin ligase Cereblon (CRBN) moiety and either a monomethyl auristatin E (MMAE) scaffold or a combretastatin A-4 (CA4) scaffold as tubulin-binding moiety, the former binding at the interface between  $\alpha$  and  $\beta$ -tubulin, the latter binding only to  $\beta$ -tubulin. However, neither strategies resulted in tubulin degradation. While the authors conclude that tubulin may be resistant to degradation by CRBN-recruiting PROTACs, they also suggest that the use of other E3 ligases or the use of a different tubulin-binding moiety may eventually lead to a successful tubulin degradation by different PROTACs. It is also conceivable that by combining photoactivation with protein degradation, the use of PHOTACs (PHOTOchemically TArgeting Chimeras) may enable a precise spatio-temporal control of degraders by light (Reynders et al., 2020) and could represent a promising approach to high precision modulation of microtubule stability.

## ARTIFICIAL INTELLIGENCE

Artificial intelligence has established itself as the most effective way to explore the chemical space in drug discovery. Recently, a drug-target prediction platform called BANDIT, which allows the integration of different chemical, genomic, clinical and pharmacological data types in a single multi-dimensional screening approach has been developed and used successfully for the identification of novel tubulin interactors (Madhukar et al., 2019). BANDIT integrates more than 20 million data points from such diverse data types as drug efficacy, post-treatment transcriptional responses, drug structures, reported adverse effects, bioassay results, and known targets. This integrated approach allows the identification of drugs that share the same target much more accurately than approaches that use single data types.

## REFERENCES

- Alushin, G. M., Lander, G. C., Kellogg, E. H., Zhang, R., Baker, D., and Nogales, E. (2014). High-Resolution Microtubule Structures Reveal the Structural Transitions in  $\alpha\beta$ -Tubulin upon GTP Hydrolysis. *Cell* 157, 1117–1129. doi:10.1016/j.cell.2014.03.053
- Blanquie, O., and Bradke, F. (2018). Cytoskeleton Dynamics in Axon Regeneration. *Curr. Opin. Neurobiol.* 51, 60–69. doi:10.1016/j.conb.2018.02.024
- Bohannon, R. A., Miller, D. G., and Diamond, H. D. (1963). Vincristine in the Treatment of Lymphomas and Leukemias. *Cancer Res.* 23, 613–621.
- Bonifacino, J. S., and Neefjes, J. (2017). Moving and Positioning the Endolysosomal System. *Curr. Opin. Cell Biol.* 47, 1–8. doi:10.1016/j.jceb.2017.01.008
- Borisy, G. G., and Taylor, E. W. (1967). The Mechanism of Action of Colchicine. *J. Cell Biol* 34, 525–533. doi:10.1083/jcb.34.2.525
- Borowiak, M., Nahaboo, W., Reynders, M., Nekolla, K., Jalinot, P., Hasseroth, J., et al. (2015). Photoswitchable Inhibitors of Microtubule Dynamics Optically
- Using this approach on microtubules, the authors initially identified a set of 24 structurally diverse orphan small molecules, which they tested experimentally on breast cancer cells. Of these 24 compounds, 14 were active on microtubules. Interestingly, only nine of these 14 compounds were also identified using structure-based only prediction methods. Tested on an ovarian carcinoma cell line that is resistant to Eribulin, an FDA approved MDAs, the authors identified three compounds that show good microtubule depolymerization activity against these cells, thus overcoming the Eribulin resistance problem.

## CONCLUSION

Because of its essential role in mitosis, tubulin is a fundamental target in drug discovery and an important benchmark for testing new cancer therapeutic approaches and ideas. This continuous quest for novel microtubule interactors is also justified by the drug resistance problem, which is often hampering the clinical efficacy of the current gold standard MTAs such as paclitaxel and vinblastine. Here we reported some of the most interesting and innovative recent approaches to MT modulation. It is likely that some of them will continue to be explored to identify new MTAs that may eventually provide viable alternatives to the current therapeutic protocols.

## AUTHOR CONTRIBUTIONS

SE, RC MM and EP wrote the manuscript. SE illustrated the concepts covered by the review.

## FUNDING

SE is a PhD student within the European School of Molecular Medicine (SEMM). RC and EP acknowledge the European Regional Development Fund (ERDF) project PhOAMS (No. 1.1.1.5/21/A/003), and ERDF project BioDrug (No. 1.1.1.5/19/A/004) for financial support. EP also acknowledges the Latvian Council of Science (grant No. lzp-2020/2–0013) for funding. This work was supported also by a grant to M.M. from the Italian Association for Cancer Research (AIRC) (IG 2020 ID 25098) and partially supported by the Italian Ministry of Health with Ricerca Corrente and 5x1000 funds.

- Control Mitosis and Cell Death. *Cell* 162, 403–411. doi:10.1016/j.cell.2015.06.049
- Buck, K. B., and Zheng, J. Q. (2002). Growth Cone Turning Induced by Direct Local Modification of Microtubule Dynamics. *J. Neurosci.* 22, 9358–9367. doi:10.1523/jneurosci.22-21-09358.2002
- Capizzi, M., Carpentier, R., Denarier, E., Adrait, A., Kassem, R., Mapelli, M., et al. (2021). Developmental Defects in Huntington's Disease Show that Axonal Growth and Microtubule Reorganization Require NUMA1. *Neuron* 110 (1), 36–50. e5. doi:10.1016/j.neuron.2021.10.033
- Cassimeris, L., and Spittle, C. (2001). Regulation of Microtubule-Associated Proteins. *Int. Rev. Cytol.* 210, 163–226. doi:10.1016/s0074-7696(01)10006-9
- David-Pfeuty, T., Simon, C., and Pantaloni, D. (1979). Effect of Antimitotic Drugs on Tubulin GTPase Activity and Self-Assembly. *J. Biol. Chem.* 254, 11696–11702. doi:10.1016/s0021-9258(19)86539-0
- de Bono, J. S., Oudard, S., Ozguroglu, M., Hansen, S., Machiels, J.-P., Kocak, I., et al. (2010). Prednisone Plus Cabazitaxel or Mitoxantrone for Metastatic Castration-Resistant Prostate Cancer Progressing after Docetaxel Treatment: a Randomised Open-Label Trial. *The Lancet* 376, 1147–1154. doi:10.1016/s0140-6736(10)61389-x
- Döbber, A., Phoa, A. F., Abbassi, R. H., Stringer, B. W., Day, B. W., Johns, T. G., et al. (2017). Development and Biological Evaluation of a Photoactivatable Small Molecule Microtubule-Targeting Agent. *ACS Med. Chem. Lett.* 8, 395–400. doi:10.1021/acsmmedchemlett.6b00483
- Dumontet, C., and Jordan, M. A. (2010). Microtubule-binding Agents: a Dynamic Field of Cancer Therapeutics. *Nat. Rev. Drug Discov.* 9, 790–803. doi:10.1038/nrd3253
- Dunkel, P., and Ilaš, J. (2021). Targeted Cancer Therapy Using Compounds Activated by Light. *Cancers (Basel)* 13, 1. doi:10.3390/cancers13133237
- Dustin, P. (1978). *Microtubules Berlin Heidelberg*. Imprint: Springer.
- Engdahl, A. J., Torres, E. A., Lock, S. E., Engdahl, T. B., Mertz, P. S., and Streu, C. N. (2015). Synthesis, Characterization, and Bioactivity of the Photoisomerizable Tubulin Polymerization Inhibitor Azo-Combretastatin A4. *Org. Lett.* 17, 4546–4549. doi:10.1021/acs.orglett.5b02262
- Fuchter, M. J. (2020). On the Promise of Photopharmacology Using Photoswitches: A Medicinal Chemist's Perspective. *J. Med. Chem.* 63, 11436–11447. doi:10.1021/acs.jmedchem.0c00629
- Gao, L., Meiring, J. C. M., Heise, C., Rai, A., Müller-Deku, A., Akhmanova, A., et al. (2021a). Photoswitchable Epothilone-Based Microtubule Stabilisers Allow GFP-Imaging-Compatible, Optical Control over the Microtubule Cytoskeleton. *Angewandte Chemie International Edition*, n/a.
- Gao, L., Meiring, J. C. M., Kraus, Y., Wranik, M., Weinert, T., Pritzl, S. D., et al. (2021b). A Robust, GFP-Orthogonal Photoswitchable Inhibitor Scaffold Extends Optical Control over the Microtubule Cytoskeleton. *Cell Chem. Biol.* 28, 228–241. e6. doi:10.1016/j.chembiol.2020.11.007
- Garber, K. (2021). The PROTAC Gold rush. *Nat. Biotechnol.* 40 (1), 12–16. doi:10.1038/s41587-021-01173-2
- Gasic, I., Groendyke, B. J., Nowak, R. P., Yuan, J. C., Kalabathula, J., Fischer, E. S., et al. (2020). Tubulin Resists Degradation by Cereblon-Recruiting PROTACs. *Cells* 9, 1. doi:10.3390/cells9051083
- Gaspari, R., Prota, A. E., Bargsten, K., Cavalli, A., and Steinmetz, M. O. (2017). Structural Basis of Cis- and Trans-combretastatin Binding to Tubulin. *Chem 2*, 102–113. doi:10.1016/j.chempr.2016.12.005
- Gasparyan, A. Y., Ayyavazyan, L., Yessirkepov, M., and Kitas, G. D. (2015). Colchicine as an Anti-inflammatory and Cardioprotective Agent. *Expert Opin. Drug Metab. Toxicol.* 11, 1781–1794. doi:10.1517/17425255.2015.1076391
- Gennerich, A., and Vale, R. D. (2009). Walking the Walk: How Kinesin and Dynein Coordinate Their Steps. *Curr. Opin. Cell Biol.* 21, 59–67. doi:10.1016/j.ccb.2008.12.002
- Giannakakou, P., Sackett, D., and Fojo, T. (2000). Tubulin/microtubules: Still a Promising Target for New Chemotherapeutic Agents. *J. Natl. Cancer Inst.* 92, 182–183. doi:10.1093/jnci/92.3.182
- Gökbuget, N., and Hoelzer, D. (1997). Vindesine in the Treatment of Leukaemia. *Leuk. Lymphoma* 26, 497–506. doi:10.3109/10428199709050886
- Goldstein, L. J. (1996). MDR1 Gene Expression in Solid Tumours. *Eur. J. Cancer* 32, 1039–1050. doi:10.1016/0959-8049(96)00100-1
- Gossiau, A., Pabbaraja, S., Knapp, S., and Chen, K. Y. (2008). Trans- and Cis-Stilbene Polyphenols Induced Rapid Perinuclear Mitochondrial Clustering and P53-independent Apoptosis in Cancer Cells but Not normal Cells. *Eur. J. Pharmacol.* 587, 25–34. doi:10.1016/j.ejphar.2008.03.027
- Gropeanu, R. A., Baumann, H., Ritz, S., Mailänder, V., Surrey, T., and Del Campo, A. (2012). Phototriggerable 2',7-Caged Paclitaxel. *PLOS ONE* 7, e43657. doi:10.1371/journal.pone.0043657
- Guillemard, V., and Saragovi, H. U. (2001). Taxane-antibody Conjugates Afford Potent Cytotoxicity, Enhanced Solubility, and Tumor Target Selectivity. *Cancer Res.* 61, 694–699.
- Haider, K., Rahaman, S., Yar, M. S., and Kamal, A. (2019). Tubulin Inhibitors as Novel Anticancer Agents: an Overview on Patents (2013–2018). *Expert Opin. Ther. Patents* 29, 623–641. doi:10.1080/13543776.2019.1648433
- Hellal, F., Hurtado, A., Ruschel, J., Flynn, K. C., Laskowski, C. J., Umlauf, M., et al. (2011). Microtubule Stabilization Reduces Scarring and Causes Axon Regeneration after Spinal Cord Injury. *Science* 331, 928–931. doi:10.1126/science.1201148
- Hoorens, M. W. H., and Szymanski, W. (2018). Reversible, Spatial and Temporal Control over Protein Activity Using Light. *Trends Biochem. Sci.* 43, 567–575. doi:10.1016/j.tibs.2018.05.004
- Huang, Z. (2005). A Review of Progress in Clinical Photodynamic Therapy. *Technol. Cancer Res. Treat.* 4, 283–293. doi:10.1177/153303460500400308
- Khrapunovich-Baine, M., Menon, V., Yang, C.-P. H., Northcote, P. T., Miller, J. H., Angeletti, R. H., et al. (2011). Hallmarks of Molecular Action of Microtubule Stabilizing Agents. *J. Biol. Chem.* 286, 11765–11778. doi:10.1074/jbc.m110.162214
- Komlodi-Pasztor, E., Sackett, D., Wilkerson, J., and Fojo, T. (2011). Mitosis Is Not a Key Target of Microtubule Agents in Patient Tumors. *Nat. Rev. Clin. Oncol.* 8, 244–250. doi:10.1038/nrclinonc.2010.228
- Krop, I. E., Lorusso, P., Miller, K. D., Modi, S., Yardley, D., Rodriguez, G., et al. (2012). A Phase II Study of Trastuzumab Emtansine in Patients with Human Epidermal Growth Factor Receptor 2-positive Metastatic Breast Cancer Who Were Previously Treated with Trastuzumab, Lapatinib, an Anthracycline, a Taxane, and Capecitabine. *Jco* 30, 3234–3241. doi:10.1200/jco.2011.40.5902
- Liu, J., Towle, M. J., Cheng, H., Saxton, P., Reardon, C., Wu, J., et al. (2007). *In Vitro* and *In Vivo* Anticancer Activities of Synthetic (-)-Laulimalide, a Marine Natural Product Microtubule Stabilizing Agent. *Anticancer Res.* 27, 1509–1518.
- Lopus, M., Oroudjev, E., Wilson, L., Wilhelm, S., Widdison, W., Chari, R., et al. (2010). Maytansine and Cellular Metabolites of Antibody-Maytansinoid Conjugates Strongly Suppress Microtubule Dynamics by Binding to Microtubules. *Mol. Cancer Ther.* 9, 2689–2699. doi:10.1158/1535-7163.mct-10-0644
- Madhukar, N. S., Khade, P. K., Huang, L., Gayvert, K., Galletti, G., Stogniew, M., et al. (2019). A Bayesian Machine Learning Approach for Drug Target Identification Using Diverse Data Types. *Nat. Commun.* 10, 5221. doi:10.1038/s41467-019-12928-6
- Matera, C., Gomila, A. M. J., Camarero, N., Libergoli, M., Soler, C., and Gorostiza, P. (2018). Photoswitchable Antimetabolite for Targeted Photoactivated Chemotherapy. *J. Am. Chem. Soc.* 140, 15764–15773. doi:10.1021/jacs.8b08249
- Miller, J. H., Field, J. J., Kanakanthara, A., Owen, J. G., Singh, A. J., and Northcote, P. T. (2018). Marine Invertebrate Natural Products that Target Microtubules. *J. Nat. Prod.* 81, 691–702. doi:10.1021/acs.jnatprod.7b00964
- Moudi, M., Go, R., Yien, C. Y., and Nazre, M. (2013). Vinca Alkaloids. *Int. J. Prev. Med.* 4, 1231–1235.
- Mozzetti, S., Ferlini, C., Concolino, P., Filippetti, F., Raspaglio, G., Prislei, S., et al. (2005). Class III Beta-Tubulin Overexpression Is a Prominent Mechanism of Paclitaxel Resistance in Ovarian Cancer Patients. *Clin. Cancer Res.* 11, 298–305.
- Müller-Deku, A., Meiring, J. C. M., Loy, K., Kraus, Y., Heise, C., Bingham, R., et al. (2020). Photoswitchable Paclitaxel-Based Microtubule Stabilisers Allow Optical Control over the Microtubule Cytoskeleton. *Nat. Commun.* 11, 4640. doi:10.1038/s41467-020-18389-6
- Na, G. C., and Timasheff, S. N. (1980). Thermodynamic Linkage between Tubulin Self-Association and the Binding of Vinblastine. *Biochemistry* 19, 1355–1365. doi:10.1021/bi00548a014
- Nogales, E., Wolf, S. G., and Downing, K. H. (1998). Structure of the  $\alpha\beta$  Tubulin Dimer by Electron Crystallography. *Nature* 391, 199–203. doi:10.1038/34465
- Noguchi, M., Skwarczynski, M., Prakash, H., Hirota, S., Kimura, T., Hayashi, Y., et al. (2008). Development of Novel Water-Soluble Photocleavable Protective Group and its Application for Design of Photoresponsive Paclitaxel Prodrugs. *Bioorg. Med. Chem.* 16, 5389–5397. doi:10.1016/j.bmc.2008.04.022

- Ojima, I., Lichtenthal, B., Lee, S., Wang, C., and Wang, X. (2015). Taxane Anticancer Agents: A Patent Perspective. *Expert Opin. Ther. Patents* 26, 1–20. doi:10.1517/13543776.2016.1111872
- Oroudjev, E., Lopus, M., Wilson, L., Audette, C., Provenzano, C., Erickson, H., et al. (2010). Maytansinoid-antibody Conjugates Induce Mitotic Arrest by Suppressing Microtubule Dynamic Instability. *Mol. Cancer Ther.* 9, 2700–2713. doi:10.1158/1535-7163.mct-10-0645
- Poruchynsky, M. S., Komlodi-Pasztor, E., Trostel, S., Wilkerson, J., Regairaz, M., Pommier, Y., et al. (2015). Microtubule-targeting Agents Augment the Toxicity of DNA-Damaging Agents by Disrupting Intracellular Trafficking of DNA Repair Proteins. *Proc. Natl. Acad. Sci. USA* 112, 1571–1576. doi:10.1073/pnas.1416418112
- Prota, A. E., Bargsten, K., Diaz, J. F., Marsh, M., Cuevas, C., Liniger, M., et al. (2014a). A New Tubulin-Binding Site and Pharmacophore for Microtubule-Destabilizing Anticancer Drugs. *Proc. Natl. Acad. Sci.* 111, 13817–13821. doi:10.1073/pnas.1408124111
- Prota, A. E., Bargsten, K., Northcote, P. T., Marsh, M., Altmann, K.-H., Miller, J. H., et al. (2014b). Structural Basis of Microtubule Stabilization by Laulimalide and Peloruside A. *Angew. Chem. Int. Ed.* 53, 1621–1625. doi:10.1002/anie.201307749
- Ranaivoson, F. M., Gigant, B., Berritt, S., Joullie, M., and Knossow, M. (2012). Structural Plasticity of Tubulin Assembly Probed by Vinca-Domain Ligands. *Acta Crystallogr. D Biol. Cryst.* 68, 927–934. doi:10.1107/s0907444912017143
- Rastogi, S. K., Zhao, Z., Barrett, S. L., Shelton, S. D., Zafferani, M., Anderson, H. E., et al. (2018). Photoresponsive Azo-Combretastatin A-4 Analogues. *Eur. J. Med. Chem.* 143, 1–7. doi:10.1016/j.ejmech.2017.11.012
- Rastogi, S. K., Zhao, Z., Gildner, M. B., Shoulders, B. A., Velasquez, T. L., Blumenthal, M. O., et al. (2021). Synthesis, Optical Properties and *In Vitro* Cell Viability of Novel Spiropyrans and Their Photostationary States. *Tetrahedron* 80, 131854. doi:10.1016/j.tet.2020.131854
- Ravelli, R. B. G., Gigant, B., Curmi, P. A., Jourdain, I., Lachkar, S., Sobel, A., et al. (2004). Insight into Tubulin Regulation from a Complex with Colchicine and a Stathmin-like Domain. *Nature* 428, 198–202. doi:10.1038/nature02393
- Reynders, M., Matsuura, B. S., Bérouti, M., Simoneschi, D., Marzio, A., Pagano, M., et al. (2020). PHOTACs Enable Optical Control of Protein Degradation. *Sci. Adv.* 6, eaay5064. doi:10.1126/sciadv.aay5064
- Rieder, C. L., and Salmon, E. D. (1998). The Vertebrate Cell Kinetochore and its Roles during Mitosis. *Trends Cell Biol.* 8, 310–318. doi:10.1016/s0962-8924(98)01299-9
- Russell-Jones, G., Mctavish, K., Mcewan, J., Rice, J., and Nowotnik, D. (2004). Vitamin-mediated Targeting as a Potential Mechanism to Increase Drug Uptake by Tumours. *J. Inorg. Biochem.* 98, 1625–1633. doi:10.1016/j.jinorgbio.2004.07.009
- Sailer, A., Ermer, F., Kraus, Y., Bingham, R., Lutter, F. H., Ahlfeld, J., et al. (2020). Potent Hemithioindigo-Based Antimitotics Photocontrol the Microtubule Cytoskeleton in Cellulo. *Beilstein J. Org. Chem.* 16, 125–134. doi:10.3762/bjoc.16.14
- Sailer, A., Ermer, F., Kraus, Y., Lutter, F. H., Donau, C., Bremerich, M., et al. (2019). Hemithioindigos for Cellular Photopharmacology: Desymmetrised Molecular Switch Scaffolds Enabling Design Control over the Isomer-Dependency of Potent Antimitotic Bioactivity. *ChemBioChem* 20, 1305–1314. doi:10.1002/cbic.201800752
- Sailer, A., Meiring, J. C. M., Heise, C., Pettersson, L. N., Akhmanova, A., Thorn-Seshold, J., et al. (2021). Pyrrole Hemithioindigo Antimitotics with Near-Quantitative Bidirectional Photoswitching that Photocontrol Cellular Microtubule Dynamics with Single-Cell Precision\*\*. *Angew. Chem. Int. Ed.* 60, 23695–23704. doi:10.1002/anie.202104794
- Saxton, W. M., Stemple, D. L., Leslie, R. J., Salmon, E. D., Zavortink, M., and McIntosh, J. R. (1984). Tubulin Dynamics in Cultured Mammalian Cells. *J. Cel. Biol.* 99, 2175–2186. doi:10.1083/jcb.99.6.2175
- Sengottuvel, V., Leibinger, M., Pfeimer, M., Andreadaki, A., and Fischer, D. (2011). Taxol Facilitates Axon Regeneration in the Mature CNS. *J. Neurosci.* 31, 2688–2699. doi:10.1523/jneurosci.4885-10.2011
- Skwarczynski, M., Noguchi, M., Hirota, S., Sohma, Y., Kimura, T., Hayashi, Y., et al. (2006). Development of First Photoresponsive Prodrug of Paclitaxel. *Bioorg. Med. Chem. Lett.* 16, 4492–4496. doi:10.1016/j.bmcl.2006.06.030
- Steinmetz, M. O., and Prota, A. E. (2018). Microtubule-Targeting Agents: Strategies to Hijack the Cytoskeleton. *Trends Cell Biol.* 28, 776–792. doi:10.1016/j.tcb.2018.05.001
- Stinchcombe, T. E. (2007). Nanoparticle Albumin-Bound Paclitaxel: a Novel Cremphor-EL-free Formulation of Paclitaxel. *Nanomedicine* 2, 415–423. doi:10.2217/17435889.2.4.415
- Thorn-Seshold, O., and Meiring, J. (2021). *Photocontrolling Microtubule Dynamics with Photoswitchable Chemical Reagents*. Cambridge University Press.
- Tron, G. C., Piral, T., Sorba, G., Pagliai, F., Busacca, S., and Genazzani, A. A. (2006). Medicinal Chemistry of Combretastatin A4: Present and Future Directions. *J. Med. Chem.* 49, 3033–3044. doi:10.1021/jm0512903
- Usui, T., Watanabe, H., Nakayama, H., Tada, Y., Kanoh, N., Kondoh, M., et al. (2004). The Anticancer Natural Product Pironetin Selectively Targets Lys352 of  $\alpha$ -Tubulin. *Chem. Biol.* 11, 799–806. doi:10.1016/j.chembiol.2004.03.028
- Vickerman, B. M., Zywo, E. M., Tarrant, T. K., and Lawrence, D. S. (2021). Taking Phototherapeutics from Concept to Clinical Launch. *Nat. Rev. Chem.* 5, 816–834. doi:10.1038/s41570-021-00326-w
- Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P., and Mcphail, A. T. (1971). Plant Antitumor Agents. VI. Isolation and Structure of Taxol, a Novel Antileukemic and Antitumor Agent from *Taxus Brevifolia*. *J. Am. Chem. Soc.* 93, 2325–2327. doi:10.1021/ja00738a045
- Waterman-Storer, C. M., and Salmon, E. D. (1997). Microtubule Dynamics: Treadmilling Comes Around Again. *Curr. Biol.* 7, R369–R372. doi:10.1016/s0960-9822(06)00177-1
- Wilson, L., Panda, D., and Ann Jordan, M. (1999). Modulation of Microtubule Dynamics by Drugs. A Paradigm for the Actions of Cellular Regulators. *Cell Struct. Funct.* 24, 329–335. doi:10.1247/csf.24.329
- Yang, J., Li, Y., Yan, W., Li, W., Qiu, Q., Ye, H., et al. (2019). Covalent Modification of Cys-239 in  $\beta$ -tubulin by Small Molecules as a Strategy to Promote Tubulin Heterodimer Degradation. *J. Biol. Chem.* 294, 8161–8170. doi:10.1074/jbc.ra118.006325
- Yang, J., Wang, Y., Wang, T., Jiang, J., Botting, C. H., Liu, H., et al. (2016). Pironetin Reacts Covalently with Cysteine-316 of  $\alpha$ -tubulin to Destabilize Microtubule. *Nat. Commun.* 7, 12103. doi:10.1038/ncomms12103
- Young, S. L., and Chaplin, D. J. (2004). Combretastatin A4 Phosphate: Background and Current Clinical Status. *Expert Opin. Invest. Drugs* 13, 1171–1182. doi:10.1517/13543784.13.9.1171
- Zenker, J., White, M. D., Templin, R. M., Parton, R. G., Thorn-Seshold, O., Bissiere, S., et al. (2017). A Microtubule-Organizing center Directing Intracellular Transport in the Early Mouse Embryo. *Science* 357, 925–928. doi:10.1126/science.aam9335
- Zhao, X., Liu, J., Fan, J., Chao, H., and Peng, X. (2021). Recent Progress in Photosensitizers for Overcoming the Challenges of Photodynamic Therapy: from Molecular Design to Application. *Chem. Soc. Rev.* 50, 4185–4219. doi:10.1039/d0cs00173b
- Zhou, J., and Giannakakou, P. (2005). Targeting Microtubules for Cancer Chemotherapy. *Cmcaca* 5, 65–71. doi:10.2174/1568011053352569

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

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

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





# Structural Basis of Human Dimeric $\alpha$ -Amino- $\beta$ -Carboxymuconate- $\epsilon$ -Semialdehyde Decarboxylase Inhibition With TES-1025

Michele Cianci<sup>1\*†</sup>, Nicola Giacchè<sup>2\*†</sup>, Lucia Cialabrini<sup>1</sup>, Andrea Carotti<sup>3</sup>, Paride Liscio<sup>2</sup>, Emiliano Rosatelli<sup>2</sup>, Francesca De Franco<sup>2</sup>, Massimiliano Gasparrini<sup>1</sup>, Janet Robertson<sup>2</sup>, Adolfo Amici<sup>4</sup>, Nadia Raffaelli<sup>1</sup> and Roberto Pellicciari<sup>2</sup>

<sup>1</sup>Biochemistry and Structural Biology Laboratory, Department of Agricultural, Food and Environmental Sciences, Polytechnic University of Marche, Ancona, Italy, <sup>2</sup>TES Pharma S.r.l., Perugia, Italy, <sup>3</sup>Department of Pharmaceutical Sciences, University of Perugia, Perugia, Italy, <sup>4</sup>Department of Clinical Sciences DISCO, Section of Biochemistry, Polytechnic University of Marche, Ancona, Italy

## OPEN ACCESS

### Edited by:

Gianluca Molla,  
University of Insubria, Italy

### Reviewed by:

Giorgio Giardina,  
Sapienza University of Rome, Italy  
John Tanner,  
University of Missouri, United States

### \*Correspondence:

Michele Cianci  
m.cianci@univpm.it  
Nicola Giacchè  
ngiacche@tespharma.com

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

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

Received: 13 December 2021

Accepted: 24 February 2022

Published: 07 April 2022

### Citation:

Cianci M, Giacchè N, Cialabrini L,  
Carotti A, Liscio P, Rosatelli E,  
De Franco F, Gasparrini M,  
Robertson J, Amici A, Raffaelli N and  
Pellicciari R (2022) Structural Basis of  
Human Dimeric  $\alpha$ -Amino- $\beta$ -  
Carboxymuconate- $\epsilon$ -Semialdehyde  
Decarboxylase Inhibition With TES-  
1025.  
Front. Mol. Biosci. 9:834700.  
doi: 10.3389/fmolb.2022.834700

Human  $\alpha$ -amino- $\beta$ -carboxymuconate- $\epsilon$ -semialdehyde decarboxylase (ACMSD) stands at a branch point of the *de novo* NAD<sup>+</sup> synthesis pathway and plays an important role in maintaining NAD<sup>+</sup> homeostasis. It has been recently identified as a novel therapeutic target for a wide range of diseases, including inflammatory, metabolic disorders, and aging. So far, in absence of potent and selective enzyme inhibitors, only a crystal structure of the complex of human dimeric ACMSD with pseudo-substrate dipicolinic acid has been resolved. In this study, we report the crystal structure of the complex of human dimeric ACMSD with TES-1025, the first nanomolar inhibitor of this target, which shows a binding conformation different from the previously published predicted binding mode obtained by docking experiments. The inhibitor has a  $K_i$  value of  $0.85 \pm 0.22$  nM and binds in the catalytic site, interacting with the Zn<sup>2+</sup> metal ion and with residues belonging to both chains of the dimer. The results provide new structural information about the mechanism of inhibition exerted by a novel class of compounds on the ACMSD enzyme, a novel therapeutic target for liver and kidney diseases.

**Keywords:** ACMSD, X-ray crystallography, TES-1025, decarboxylase, drug discovery, *de novo* NAD<sup>+</sup> synthesis

## INTRODUCTION

Human  $\alpha$ -amino- $\beta$ -carboxymuconate- $\epsilon$ -semialdehyde decarboxylase (ACMSD, EC 4.1.1.45) (Fukuoka et al., 2002) stands at a branch point of the *de novo* NAD<sup>+</sup> synthesis pathway, starting from the essential amino acid tryptophan, and plays an important role in maintaining NAD<sup>+</sup> homeostasis. Given the beneficial effects of replenished NAD<sup>+</sup> pools, there is an intense search for strategies to increase intracellular NAD<sup>+</sup> by limiting NAD<sup>+</sup> consumption or increasing NAD<sup>+</sup> production (Katsyuba and Auwerx, 2017; Katsyuba et al., 2020). In this view, ACMSD inhibition is emerging as a potent strategy to replenish NAD<sup>+</sup> levels by improving the coenzyme's production (Yoshino, 2019).

In detail, ACMSD catalyzes the decarboxylation of 2-amino 3-carboxymuconate 6-semialdehyde (ACMS), an intermediate in the *de novo* NAD<sup>+</sup> synthesis pathway, to 2-aminomuconate-6-semialdehyde (AMS), through a metal-mediated, O<sub>2</sub>-independent, non-

oxidative decarboxylation reaction (Li et al., 2006), which proceeds through a metal-bound hydroxide (Huo et al., 2013, 2015). AMS can either undergo spontaneous cyclization of the pyridine ring to form picolinic acid (PIC) or be oxidized to 2-aminomuconate, which is further metabolized so that it can enter the tricarboxylic acid (TCA) cycle. Otherwise ACMS, if not metabolized by ACMSD, can cyclize spontaneously to quinolinic acid (QUIN), which is further converted to the coenzyme NAD<sup>+</sup> (Figure 1). Because the cyclization of ACMS into QUIN is a spontaneous reaction, the amount of ACMS undergoing this conversion and therefore leading to the production of NAD<sup>+</sup> is primarily determined by the activity of ACMSD (Fukuoka et al., 2002). Thus, inhibition of ACMSD, which is primarily and highly expressed in the liver and kidneys (Pucci et al., 2007), would channel ACMS toward *de novo* NAD<sup>+</sup> biosynthesis, providing a novel way to replenish NAD<sup>+</sup> levels and re-establish NAD<sup>+</sup> homeostasis in pathological conditions, particularly in liver- and kidney-associated diseases.

ACMSD is catalytically inactive in the monomeric form and active in the homodimeric form since the neighboring subunit contributes with one of the two substrate-binding arginine residues (Huo et al., 2013). Furthermore, recent experiments using size-exclusion chromatography coupled with small-angle X-ray scattering (SEC-SAXS) analysis have evidenced a protein concentration-dependent activity of the enzyme, revealing that its quaternary structure is in a dynamic equilibrium among the monomeric, dimeric, and higher-order oligomeric states (Yang et al., 2019).

The first small molecule inhibitors of ACMSD to be identified were the anti-tuberculosis drug pyrazinamide (Saito et al., 2000) and the phthalate monoester, such as mono (2-ethylhexyl) phthalate (MEHP) (Fukuwatari et al., 2004) with weak and nonselective activity. Subsequent efforts in understanding the mechanism of recognition and binding of ligands to the active site of the human enzyme resulted in the release of the first co-crystal complex of ACMSD (PDB code 2WM1) with the inhibitor 1,3-dihydroxyacetonephosphate (DHAP, 1) (Garavaglia et al., 2009) (Figure 2).

Subsequently, the structure of the human recombinant ACMSD complex with the competitive inhibitor pyridine-2,6-dicarboxylic acid (PDC, 2) (PDB code 4IH3) was published, refining the previous findings (Huo et al., 2015) (Figure 2).

Recently, the salicylic-derivative, nonsteroidal anti-inflammatory drug (NSAID) and FDA-approved drug diflunisal (3) was identified to inhibit ACMSD with an IC<sub>50</sub> of 13.5 μM, and its complex structure in the *Pseudomonas fluorescens* ACMSD has been resolved and published (PDB code 7K12) (Yang et al., 2021) (Figure 2). A chronological summary of all ACMSD structures that have been resolved and published is reported in Table 1.

TES-1025 (CAS. 1883602-21-8, 2-[3-[(5-cyano-6-oxo-4-thiophen-2-yl-1H-pyrimidin-2-yl) sulfanylmethyl]phenyl]acetic acid, PubChem CID: 137142885) is a potent and selective human α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (ACMSD) inhibitor with an

IC<sub>50</sub> of 13 nM. The compound has been selected as the first low-nanomolar inhibitor of human ACMSD with a suitable overall balance of good physicochemical properties and *in vitro* safety profile, identified after the discovery, synthesis, and biological evaluation of a series of 2-thiopyrimidone-5-carbonitriles as the first class of small-molecule drug-like ACMSD inhibitors. Proof-of-concept studies for the first time revealed that the inhibition of ACMSD by TES-1025 led to the modulation of intracellular NAD<sup>+</sup> levels with consequent *in vivo* enhancement of *de novo* NAD<sup>+</sup> biosynthesis via ACMSD target engagement (Pellicciari et al., 2018).

On the basis of the discovery of TES-1025 and related analogs, we have established valuable tools for a better understanding of the therapeutic applications of ACMSD inhibitors for disorders such as mitochondrial dysfunctions and metabolic and renal diseases, associated with the dysregulation or reduced NAD<sup>+</sup> levels. *In vivo* efficacy data obtained with TES-1025 in preclinical murine models of liver and kidney diseases (Katsyuba et al., 2018) suggested ACMSD as a promising novel therapeutic target to improve health in pathological settings such as that of acute kidney injury (AKI) (Kellum and Prowle, 2018; Manrique-Caballero et al., 2021).

In this study, we report the crystal structure of the complex of human dimeric ACMSD with TES-1025 (4), (Figure 2) the first potent and selective ACMSD inhibitor.

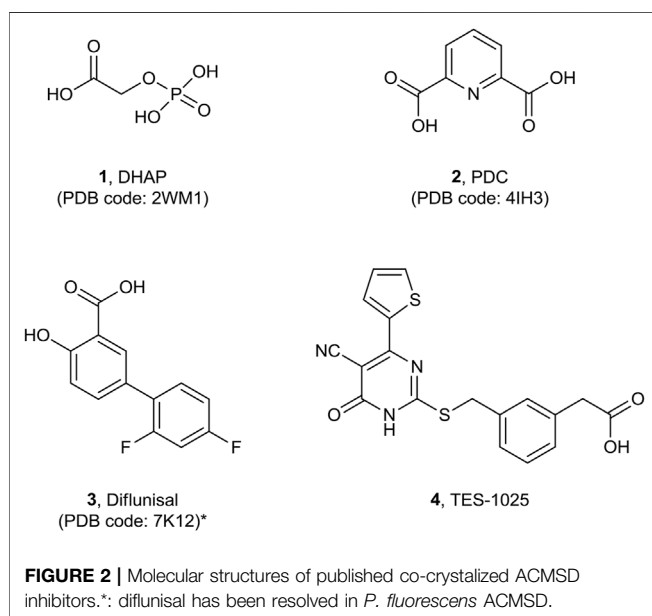
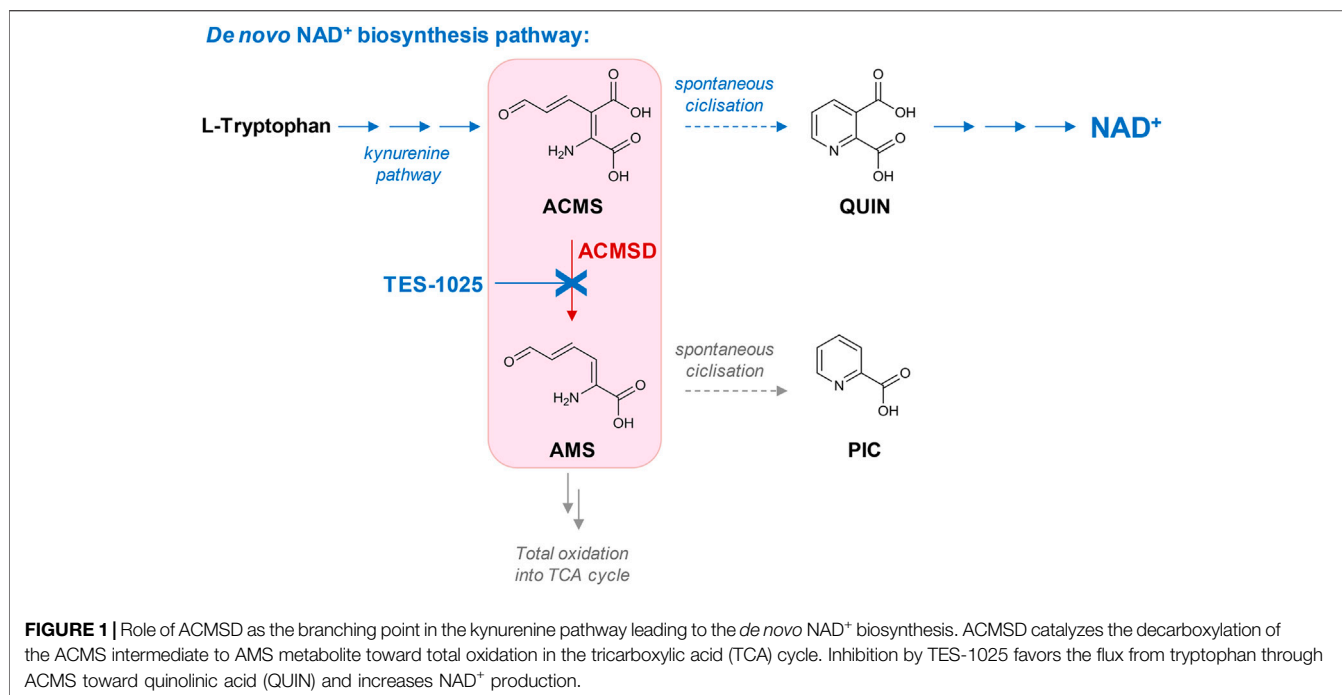
## MATERIALS AND METHODS

### Expression and Purification of hACMSD

Expression of the recombinant protein was achieved as described previously (Pucci et al., 2007). Purification was performed as described previously (Garavaglia et al., 2009), with some modifications. In detail, a pellet of *Pichia pastoris* cells expressing the enzyme and derived from 400 ml culture was resuspended in 80 ml of lysis buffer consisting of 10 mM potassium phosphate, pH 7.0, 50 mM NaCl, 1 mM DTT, 5 mM 2-mercaptoethanol, and 1 mM PMSF and aprotinin, leupeptin, chymostatin, pepstatin, and antipain at 0.002 mg/ml each. After disruption by two cycles of French press (SML-Aminco, Urbana, IL, United States) at 1,000 psi, the suspension was centrifuged at 40,000 × g, for 30 min at 4°C. The supernatant was made to 10 mg/ml by dilution with lysis buffer, and streptomycin sulfate was added dropwise at a final concentration of 1%. After 30 min stirring on ice, the sample was centrifuged at 20,000 × g for 10 min at 4°C, and the supernatant was applied to 4 ml TALON Superflow resin (Cytiva, United States) equilibrated with 50 mM potassium phosphate pH 7.4 and 50 mM NaCl (buffer A). After washing with 20 mM imidazole in buffer A, elution was performed with 350 mM imidazole in buffer A.

### Kinetics Studies

The inhibitory effect of TES-1025 on ACMSD activity was determined by the coupled spectrophotometric assay (Pucci et al., 2007). Briefly, pre-assay mixtures consisting of different



concentrations of hydroxyanthranilic acid (from 5 to 20  $\mu\text{M}$ ) and an excess amount of recombinant *R. metallidurans* hydroxyanthranilic acid dioxygenase, in 50 mM 4-morpholinepropanesulfonic acid, pH 6.5 and 100 mM ammonium iron sulfate, were incubated at 37°C, with monitoring ACMS formation at 360 nm. After the reaction was complete, 30 nM ACMSD and TES-1025 (from 0.5 to 40 nM) were added. The enzyme activity was calculated by the initial rate of the absorbance decrease subtracted from that of a control mixture in the absence of ACMSD. The  $K_i$  value was

calculated from the initial velocity data using the Dixon equation for tightly bound competitive inhibitor (Segel, 1993):

$$[I_t] = \left( K_i \left( 1 + \frac{[S]}{K_m} \right) \frac{V_0}{V_i} + [E_t] \right) \left( 1 - \frac{V_i}{V_0} \right), \quad (1)$$

where  $V_i$  is the initial velocity at a given  $[S]$  in the presence of the inhibitor,  $V_0$  is the initial velocity at the same  $[S]$  in the absence of the inhibitor,  $K_m$  is the Michaelis–Menten constant for the substrate,  $[I_t]$  and  $[E_t]$  are the total amount of the inhibitor and enzyme, respectively, and  $K_i$  is the apparent  $K_i$  value. The intercept on the y-axis of the replot of apparent  $K_i$  values against  $[S]$  gives the  $K_i$  value.

## Protein Crystallization and Data Collection

TES-1025 was developed by TES Pharma as reported by Pellicciari et al., (2018). For the crystallization trials, the TALON pool containing the purified protein was diluted ten-fold with 50 mM potassium phosphate and 5 mM 2-mercaptoethanol and then concentrated by ultrafiltration, using an Amicon Ultra Centrifugal Filter (cutoff 10 kDa, Merck, Millipore), at 4°C, to a final protein concentration of 4.5 mg/ml. For all crystallization trials, the sitting drop vapor diffusion method was applied. The concentrated enzyme was incubated at the 1:10 molar ratio with a stock solution of 50 mM TES-1025 in dimethyl sulfoxide, for 1 hour at room temperature, and 1.25  $\mu\text{L}$  of protein–ligand solution was mixed with an equal volume of reservoir solution, and it was equilibrated against 100  $\mu\text{L}$  of the reservoir solution. The best crystals were obtained in reservoir solution containing 100 mM Na(CH<sub>3</sub>COO), pH 5.7, 22% (w/v) PEG 4000, and they grew to their final size in few weeks at 18°C.

**TABLE 1** | Summary of all ACMSD X-ray structures deposited in the Protein Data Bank (PDB) (order for date of release).

PDB code	Date of release	Resolution (Å)	Organism	Conformation	Mutation	Metal cofactor	Ligand	References
2HBX	2006-09-19	2.50	<i>Pseudomonas fluorescens</i>	Dimer	No	Co <sup>2+</sup>	—	Martynowski et al. (2006)
2HBV	2006-09-19	1.65	<i>Pseudomonas fluorescens</i>	Dimer	No	Mg <sup>2+</sup> , Zn <sup>2+</sup>	—	Martynowski et al. (2006)
2WM1	2009-11-03	2.01	<i>Homo sapiens</i>	Monomer (but dimer in the lattice)	No	Zn <sup>2+</sup>	DHAP	Garavaglia et al. (2009)
4EPK	2012-08-22	2.60	<i>Pseudomonas fluorescens</i>	Dimer	H228G	Mg <sup>2+</sup> , Zn <sup>2+</sup>	—	Huo et al. (2012)
4ERI	2012-08-22	2.00	<i>Pseudomonas fluorescens</i>	Dimer	H228Y	Mg <sup>2+</sup> , Zn <sup>2+</sup>	—	Huo et al. (2012)
4ERA	2012-08-22	2.40	<i>Pseudomonas fluorescens</i>	Dimer	H228Y	Co <sup>2+</sup>	—	Huo et al. (2012)
4ERG	2012-08-22	2.79	<i>Pseudomonas fluorescens</i>	Dimer	—	Fe <sup>3+</sup>	—	Huo et al. (2012)
4IGM	2014-05-07	2.39	<i>Homo sapiens</i>	Dimer	No	Zn <sup>2+</sup>	—	No references
4IGN	2014-05-07	2.33	<i>Homo sapiens</i>	Dimer	R47A	Zn <sup>2+</sup>	—	Huo et al. (2015)
4IH3	2014-05-21	2.49	<i>Homo sapiens</i>	Dimer	No	Zn <sup>2+</sup>	PDC	Huo et al. (2015)
4OFC	2014-11-19	1.99	<i>Homo sapiens</i>	Dimer	No	Zn <sup>2+</sup>	—	Huo et al. (2015)
6MGS	2019-06-19	3.13	<i>Pseudomonas fluorescens</i>	Dimer	No	Co <sup>2+</sup>	—	Yang et al. (2019)
6MGT	2019-06-19	2.77	<i>Pseudomonas fluorescens</i>	Dimer	H110A	Co <sup>2+</sup>	—	Yang et al. (2019)
7K12	2021-01-13	2.17	<i>Pseudomonas fluorescens</i>	Dimer	No	Zn <sup>2+</sup>	Diflunisal	Yang et al. (2021)
7K13	2021-01-13	1.83	<i>Pseudomonas fluorescens</i>	Dimer	No	Zn <sup>2+</sup>	Diflunisal-derivative	Yang et al. (2021)

Crystals were transported to the synchrotron in plates, mounted in nylon loops, and flash-frozen directly at 100 K in a nitrogen gas stream. Diffraction data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) at beamline ID30A-3 (MASSIF-3) (von Stetten et al., 2020).

## Structure Determination, Refinement, and Analysis

The diffraction data were integrated and scaled with the XDS/XSCALE program package (Kabsch, 2010). The crystals belong to the space group P2<sub>1</sub>2<sub>1</sub>2, with unit cell a = 153.3 Å, b = 92.5 Å, and c = 103.9 Å. Starting phases for solving the crystal structure were obtained with molecular replacement using PHASER (Adams et al., 2002) with monomer A of the hACMSD structure as a starting model reported by Huo et al. (2015) (PDB code 4IH3) after atom randomization to avoid any bias as a search model. Automated model building was accomplished by the PHENIX (Adams et al., 2002) suite, followed by manual fitting of the side chains and solvent molecules into electron density maps performed using COOT (Emsley and Cowtan, 2004) and PHENIX suite (Adams et al., 2002), while monitoring R<sub>work</sub>, R<sub>free</sub>, and Ramachandran plot with PROCHECK (Laskowski et al., 1993) and related geometrical parameters. The Fourier difference electron density OMIT maps at 3σ were inspected to verify the presence of TES-1025. The models were checked with the PDB REDO web server (Joosten et al., 2014). Model coordinates and structure factors of the X-ray crystal structure

of ACMSD co-crystallized in the presence of TES-1025 were deposited in the Protein Data Bank (PDB) under the accession code: 7PWY. Data collection, processing, and final refinement statistics are given in **Table 2**. Ligand interaction diagram is generated by the tool of Maestro of the Schrodinger suite 2017–1 (<https://www.schrodinger.com>). The images produced in this article were generated using CCP4mg (McNicholas et al., 2011) and PyMOL software (<https://www.pymol.org>).

## RESULTS

The human ACMSD crystal structure in complex with TES-1025 at 2.5 Å resolution was refined to final R<sub>work</sub> and R<sub>free</sub> values of 0.210 and 0.252, respectively. The Ramachandran plot shows more than 96% of the residues in the favored regions and 4% in the allowed regions (**Table 2**). The molecular replacement solution of the hACMSD structure in complex with TES-1025 comprised four monomers in the asymmetric unit to form two homodimers. The average RMSD, calculated with SUPERPOSE (Winn et al., 2011) over 320 residues of each monomer, against the starting model, is 0.45 Å, to confirm that the overall fold of the enzyme is maintained. In brief, hACMSD shows a molecular architecture comprising of 12 α-helices, 11 β-strands, and the connecting loops. Residues 14–48 form the small insertion domain that comprises a short α-helix and a three-stranded anti-parallel β-sheet; the remaining protein residues form a (α/β)<sub>8</sub> barrel domain and a two-α-helices C-terminal extension.



**TABLE 2 |** Data collection and refinement statistics.

Wavelength (Å)	0.967
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2
Cell parameters (a, b, and c, Å)	153.4, 92.6, 103.9
Resolution range (Å)	45.89–2.50 (2.50–2.58) <sup>a</sup>
Total reflections	395,775 (30,276) <sup>a</sup>
Unique reflections	51,945 (4,423) <sup>a</sup>
Redundancy	7.6 (6.8) <sup>a</sup>
Completeness (%)	99.9 (100.0) <sup>a</sup>
Mean I/sigma(I)	11.5 (1.6) <sup>a</sup>
R <sub>merge</sub> <sup>b</sup>	0.11 (11.1) <sup>a</sup>
R <sub>pim</sub> <sup>c</sup>	0.063 (0.688) <sup>a</sup>
CC1/2	0.999 (0.766) <sup>a</sup>
CC*	1.00 (0.926) <sup>a</sup>
Reflections used in refinement	51,611 (5,113) <sup>a</sup>
Reflections used for R <sub>free</sub>	2,565 (292) <sup>a</sup>
Wilson B-factor (Å <sup>2</sup> )	52.35
R <sub>work</sub> <sup>d</sup>	0.210 (0.310) <sup>a</sup>
R <sub>free</sub> <sup>d</sup>	0.252 (0.358) <sup>a</sup>
Total no. of atoms <sup>d</sup>	10,647
Macromolecules	10,296
Ligands	62
Water molecules	289
Protein residues	1,289
RMSD <sup>d</sup>	—
Bond length (Å)	0.003
Angles (°)	0.55
Ramachandran <sup>d</sup>	—
Favored (%)	96.54
Allowed (%)	3.46
Outliers (%)	0.00
Average B-factor <sup>d</sup>	68.28
Macromolecules	68.48
Ligands	73.64
Solvent	59.83

<sup>a</sup>Values in the highest resolution shell.

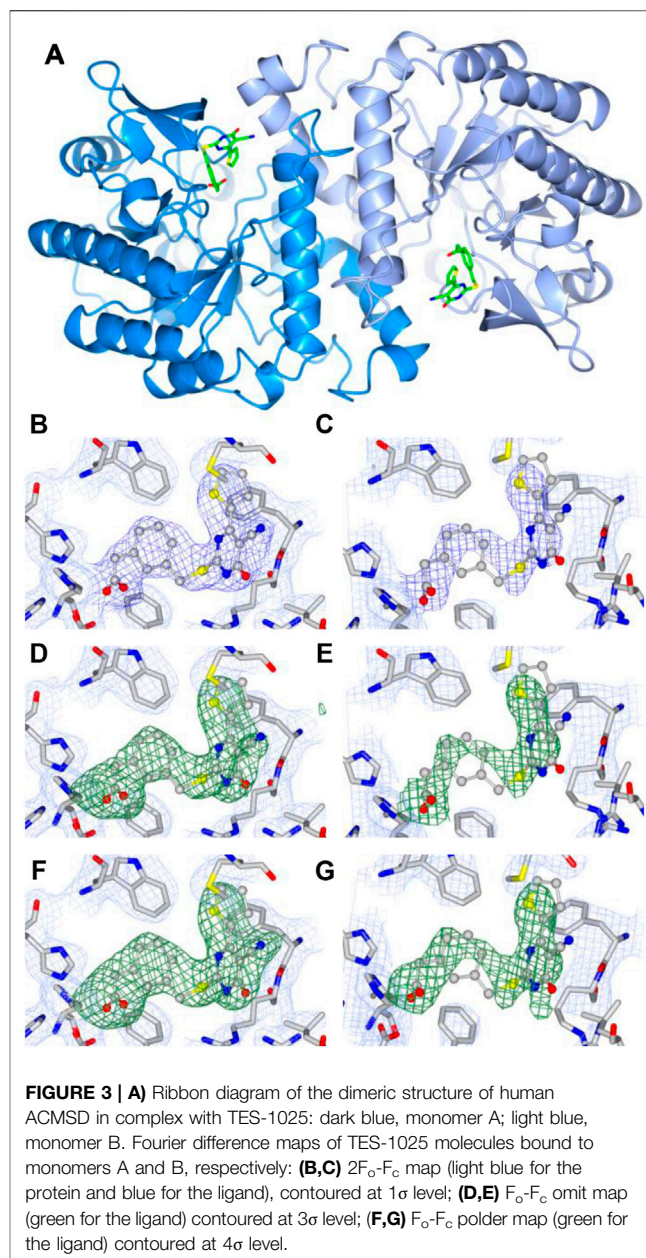
<sup>b</sup> $R_{\text{merge}} = \sum_{hkl} \sum_j |I_j - \bar{I}| / \sum_{hkl} \sum_j I_j$ , where  $I$  is the intensity of a reflection, and  $\bar{I}$  is the mean intensity of all symmetry-related reflections  $j$ .

<sup>c</sup> $R_{\text{pim}} = \sum_{hkl} \{ [1/(N-1)]^{1/2} \sum_j |I_j - \bar{I}| / \sum_j I_j \}$ , where  $I$  is the intensity of a reflection, and  $\bar{I}$  is the mean intensity of all symmetry-related reflections  $j$ , and  $N$  is the multiplicity (Weiss, 2001).

<sup>d</sup>Calculated with PHENIX suite (Adams et al., 2002),  $R_{\text{free}}$  is calculated using 5% of the total reflections that were randomly selected and excluded from refinement.

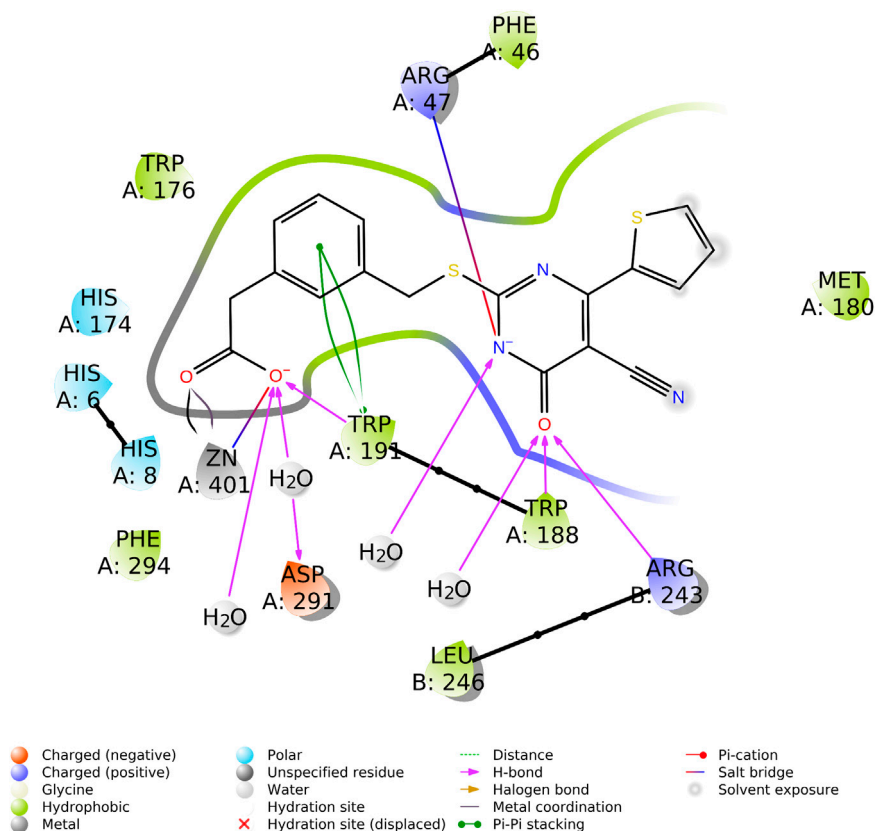
A total of two homodimers are observed in the asymmetric unit. Analysis of interface area, solvation energy gain upon interface formation, and the total binding energy of the interface, calculated using PISA (Krissinel and Henrick, 2007), confirms homodimers A, B and C, D as biological units. hACMSD homodimers have been reported previously (Garavaglia et al., 2009; Huo et al., 2015) (Figure 3A). In homodimer AB, there are no disordered regions, while in homodimer CD there are disordered regions namely in chain C between residues 30–37, 178–185, and 242–253 and in chain D between residues 242–253.

Similar to previously published X-ray structures (Garavaglia et al., 2009; Huo et al., 2015), the active site of hACMSD is located at the C-terminal opening of the  $\beta$ -barrel and characterized by a  $\text{Zn}^{2+}$  metal ion coordinated with a distorted trigonal bipyramid geometry, by residues His6, His8, His174, His 224, the moiety of Asp291, and a conserved water molecule. A water molecule bridges the  $\text{Zn}^{2+}$  metal ion with the guanidino moiety of Arg235 belonging to the neighboring subunit (chain B) of the hACMSD dimer.



**FIGURE 3 |** (A) Ribbon diagram of the dimeric structure of human ACMSD in complex with TES-1025: dark blue, monomer A; light blue, monomer B. Fourier difference maps of TES-1025 molecules bound to monomers A and B, respectively: (B,C)  $2F_o - F_c$  map (light blue for the protein and blue for the ligand), contoured at  $1\sigma$  level; (D,E)  $F_o - F_c$  omit map (green for the ligand) contoured at  $3\sigma$  level; (F,G)  $F_o - F_c$  polder map (green for the ligand) contoured at  $4\sigma$  level.

The presence and positioning of the TES-1025 molecules was indicated by the Fourier difference  $2F_o - F_c$  maps at  $1\sigma$ , the Fourier difference  $F_o - F_c$  OMIT maps at  $3\sigma$  level, and the polder maps at  $4\sigma$  level in each active site of hACMSD homodimer AB (Figure 3). In homodimer CD, no electron density was observed that could be attributed to a TES-1025 molecule. The molecule TES-1025 located in monomer A was refined with full occupancy, and the molecule located in monomer B was refined with a partial occupancy of 0.80. The correlation coefficients for the polder map (Liebschner et al., 2017), calculated by omitting both TES-1025 molecules were  $\text{CC}(1, 3) = 0.89$ , i.e., larger than  $\text{CC}(1, 2) = 0.54$  and  $\text{CC}(2, 3) = 0.56$ . When omitting only TES-1025 molecule in chain A, we obtained  $\text{CC}(1, 3) = 0.91$ , i.e., larger than  $\text{CC}(1, 2) = 0.62$  and  $\text{CC}(2, 3) = 0.55$ . When omitting only



**FIGURE 4** | Ligand interaction diagram of TES-1025 in the ACMSD active site of monomer A, reporting the ligand–protein type of interactions and involved residues (diagram calculated and generated by Maestro, Schrodinger suite).

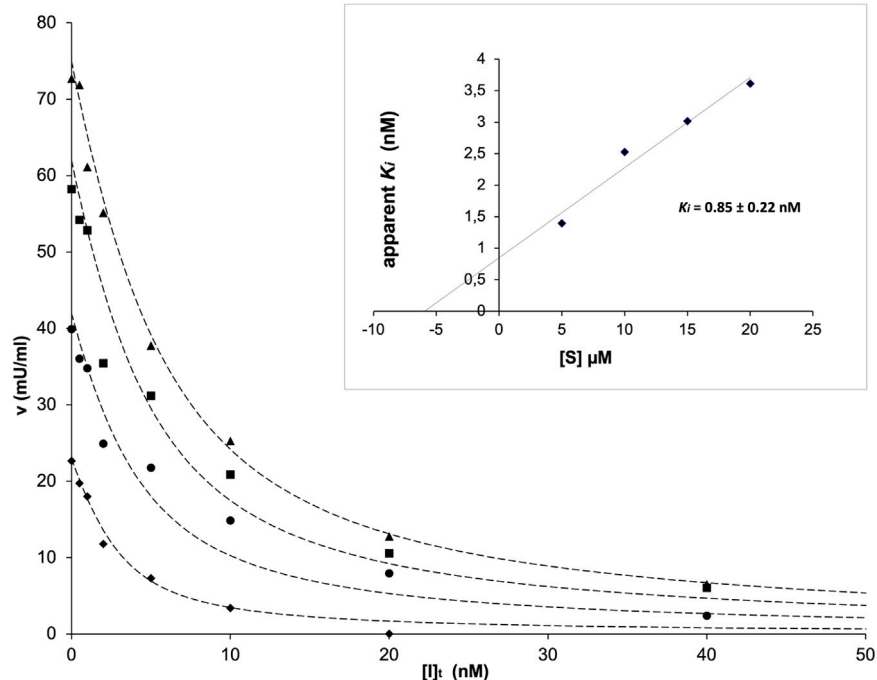
TES-1025 molecule in chain B, we obtained  $CC(1, 3) = 0.84$ , i.e., larger than  $CC(1, 2) = 0.59$  and  $CC(2, 3) = 0.57$ . Since the  $CC(1, 3)$  is larger than  $CC(1, 2)$  and  $CC(2, 3)$ , then the density observed corresponds to the atomic features of the TES-1025 molecules.

The active site of each subunit binds one TES-1025 molecule with the same set of interactions. The carboxylic moiety of the ligand coordinates the  $Zn^{2+}$  metal ion ( $Zn \cdots O$  distance is 2.5 Å) and establishes an H-bond with the indole moiety of Trp191 residue (the  $H \cdots O$  distance is 2.6 Å, and the  $N-H \cdots O$  angle is 135.8°) while interacts with Asp291 through a water molecule. The pyrimidine ring interacts through the carbonyl group with Arg243 (the  $H \cdots O$  distance is 2.5 Å, and the  $N-H \cdots O$  angle is 143.6°), belonging to the neighboring subunit (chain B) of the functional dimer of ACMSD, while a second H-bond is established with the catalytic residue Trp188 (the  $H \cdots O$  distance is 2.7 Å, and the  $N-H \cdots O$  angle is 132.5°). The negatively charged nitrogen of the TES-1025 pyrimidine ring engages a charge–charge interaction with the positively charged nitrogen of the Arg47 (the  $N \cdots N$  distance is 3.6 Å). Also, further interactions with solvent water molecules are defined. The 2-thiophene ring fits into a hydrophobic cavity generated by Trp176, Phe46, Met180, and Trp191; this latter residue also makes Pi–Pi stacking with the phenyl ring of TES-1025. (Figure 4).

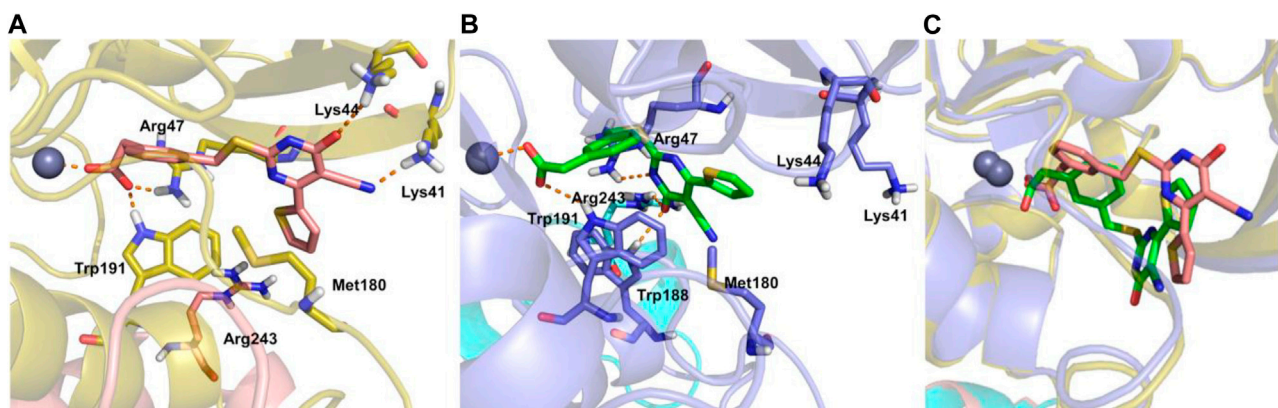
The kinetic analysis of the inhibition exerted by TES-1025 on ACMSD was performed by assaying the enzyme activity in the presence of varying concentrations of the inhibitor and substrate. The  $IC_{50}$  value of the inhibitor at 10  $\mu M$  substrate concentration is reported to be about 13 nM (Pellicciari et al., 2018). This value is very close to the concentration of ACMSD which is used in the activity assay, implying that a significant portion of the total inhibitor in the assay mixture is enzyme-bound. Therefore, to investigate the inhibition kinetics, the Dixon method (Segel, 1993) was used as described in Materials and Methods. The best fit with the experimental data was obtained by using the Dixon Equation 1 for a tightly bound competitive inhibitor (Figure 5). A  $K_i$  value of  $0.85 \pm 0.22$  nM was calculated (inset in Figure 5).

## DISCUSSION

The structural basis of human dimeric  $\alpha$ -amino- $\beta$ -carboxymuconate- $\epsilon$ -semialdehyde decarboxylase inhibition with TES-1025 is severalfold. Overall, the X-ray structure of the ACMSD–TES-1025 complex described in this study experimentally confirmed the competitive inhibition mode displayed by the ligand, in which the catalytic Arg47 and the  $Zn^{2+}$  metal ion together with Trp191 and chain B are



**FIGURE 5 |** ACMSD activity as a function of inhibitor concentration in the presence of fixed concentrations of the substrate traced with diamonds (5  $\mu$ M), circles (10  $\mu$ M), squares (15  $\mu$ M), and triangles (20  $\mu$ M). In the inset, the replot of apparent  $K_i$  versus  $[S]$  is shown.



**FIGURE 6 |** TES-1025-binding (pink sticks) modes in the docking model **(A)**, TES-1025 (green sticks) in the crystal structure **(B)**, and superposition of the two ligand poses **(C)**, together with the protein ribbons and  $\text{Zn}^{2+}$  ions. ACMSD chains A and B are shown in yellow and pink ribbons for the induced fit result, while are in blue and cyan ribbons for the X-ray image, respectively. The  $\text{Zn}^{2+}$  metal ion is displayed as a gray ball. The relevant residues of the binding site are labeled and shown as sticks. Hydrogen bond and salt bridge interactions are indicated with orange dashed lines.

engaged in unique specific interactions with the meta carboxylate group and the pyrimidine ring of the ligand. These two interactions lock the head and the tail of the ligand, respectively. Moreover, the pyrimidine ring strongly interacts with the pocket residues side chains through a network of hydrogen bonds mediated by water molecules, thus stabilizing the central core of the ligand. This strong network of interactions generated by TES-1025 ensured the low nanomolar potency inhibition of the enzyme and

confirmed the previously published data of structure and activity relationship (Pellicciari et al., 2018). The kinetic analysis confirmed that TES-1025 is a competitive inhibitor with a  $K_i$  value in the low nanomolar range.

In the article by Pellicciari et al. (2018), the investigation of the induced-fit docking pose of TES-1025 into the hACMSD catalytic site coming from the 4IH3 X-ray (Huo et al., 2015) was presented. The results indicated that the majority of the binding affinity was due to the ionic interaction between the m-carboxy group of the



ligand with the  $\text{Zn}^{2+}$  metal ion and the catalytic residues Arg47 and Trp191. Meanwhile, the pyrimidine ring was strongly anchored to Lys41 and Lys44 residues by hydrogen bond and ionic interactions. Furthermore, additional hydrophobic contacts were also established between the thiophene and Met180. In this view, the current X-ray structure of ACMSD in complex with TES-1025 reports a different ligand disposition with respect to the putative binding mode predicted by docking computation (ligand RMSD calculated on non-hydrogen atoms is 6.42 Å when using the protein backbone for the alignment). Indeed, upon TES-1025 binding, the side chain of Arg47, as resolved in the crystal structure, is oriented to the middle of the pocket driven by the interaction with the acidic oxygen of the pyrimidine ring of the ligand. It is worth noting that this Arg47 orientation was observed neither in the 4IH3 complex used for the *in silico* studies nor in other released ACMSD X-ray structures (PDB codes: 2WM1, 4IGM, 4IGN, and 4OFC) (Garavaglia et al., 2009; Huo et al., 2015). Moreover, the carbonyl moiety of the pyrimidine ring of TES-1025 engages an H-bond with Arg243, belonging to the neighboring subunit (chain B) of the functional dimer of ACMSD. The crystallographic disposition showed a 180° rotation of the pyrimidine ring around the  $\text{Csp}^2\text{-S}$  bond that completely abolishes the interactions present in the docking pose with Lys44 and Lys41 (Figure 6).

The previous detailed works of the Aimin Liu lab (Huo et al., 2013; Huo et al., 2015) confirmed the metal ion ( $\text{Zn}^{2+}$ ) dependence of the human ACMSD for its catalytic activity and the active role of Arg47 and Arg235 of the chain B for the interaction of the natural substrate ACMS in the homodimeric functional complex. TES-1025 was able to interact with all the essential catalytic features of the site, with only the substitution of Arg235 with Arg243, stabilizing a homodimeric inactive complex and confirming the high affinity and potency of the ligand. Moreover, the preferential recognition and binding of the pyrimidine moiety of TES-1025 in ACMSD supported the subfamilial similarities with the related enzyme 5-carboxyl-uracil decarboxylase (IDCase) (Huo et al., 2015).

The asymmetric unit consists of two homodimers, namely AB and CD. In AB, two TES-1025 molecules were observed, while in CD no electron density was observed that could be attributed to a TES-1025 molecule. The two homodimers display high dynamics of the different chains too. Indeed, in homodimer AB, there are no disordered loop regions, while in homodimer CD there are disordered regions. Considering that upon ligand binding,

Arg243 interacts with the pyrimidine ring of TES-1025, and the presence of the ligand is reflected in the folding of the residue region 240–253 to an  $\alpha$ -helix structure.

In summary, the determination of the crystal structure of the human ACMSD homodimer with the first nanomolar and selective inhibitor TES-1025 reveals unforeseen interactions of the functional groups of the small molecule with the catalytic side chains and the metal ion within the active site, elucidating the principles of its potent inhibitory mechanism. These results further validate the selectivity of TES-1025 for the enzyme and consolidate the knowledge about ACMSD as a promising therapeutic target (Katsyuba et al., 2018; Kellum and Prowle, 2018; Kellum et al., 2020; Manrique-Caballero et al., 2021) for the recovery of the *de novo*  $\text{NAD}^+$  biosynthesis pathway and the maintaining of  $\text{NAD}^+$  homeostasis impaired in hepatic (Zhou et al., 2016) and renal diseases (Poyan Mehr et al., 2018).

## DATA AVAILABILITY STATEMENT

Structure factors and coordinates have been deposited in the Protein Data Bank with the PDB code: 7PWY.

## AUTHOR CONTRIBUTIONS

NG designed the research and wrote the manuscript. MC designed the crystallographic experiments, collected and analyzed data, and wrote the manuscript. AA, LC, FD, and MG performed protein expression and purification, prepared crystallographic samples, and performed kinetics experiments. AC performed *in silico* ligand interaction analysis and calculations and wrote the manuscript. PL and ER prepared TES-1025. JR and AA edited the manuscript. NR and RP designed the research and edited the manuscript.

## ACKNOWLEDGMENTS

We thank the research staff of ESRF (Grenoble, France) for the technical support. The synchrotron MX data were collected at ESRF (Grenoble, France) under the beam time award MX-1949.

## REFERENCES

- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., et al. (2002). PHENIX: Building New Software for Automated Crystallographic Structure Determination. *Acta Crystallogr. D Biol. Cryst.* 58, 1948–1954. doi:10.1107/S0907444902016657
- Emsley, P., and Cowtan, K. (2004). Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. D Biol. Cryst.* 60, 2126–2132. doi:10.1107/S0907444904019158
- Fukuoka, S.-I., Ishiguro, K., Yanagihara, K., Tanabe, A., Egashira, Y., Sanada, H., et al. (2002). Identification and Expression of a cDNA Encoding Human  $\alpha$ -Amino- $\beta$ -carboxymuconate- $\epsilon$ -semialdehyde Decarboxylase (ACMSD). *J. Biol. Chem.* 277, 35162–35167. doi:10.1074/jbc.M200819200
- Fukuwatari, T., Ohsaki, S., Fukuoka, S. I., Sasaki, R., and Shibata, K. (2004). Phthalate Esters Enhance Quinolinate Production by Inhibiting  $\alpha$ -Amino- $\beta$ -Carboxymuconate- $\epsilon$ -Semialdehyde Decarboxylase (ACMSD), a Key Enzyme of the Tryptophan Pathway. *Toxicol. Sci.* 81, 302–308. doi:10.1093/toxsci/kfh204
- Garavaglia, S., Perozzi, S., Galeazzi, L., Raffaelli, N., and Rizzi, M. (2009). The crystal Structure of Human  $\alpha$ -amino- $\beta$ -carboxymuconate- $\epsilon$ -semialdehyde Decarboxylase in Complex with 1,3-dihydroxyacetonephosphate Suggests a Regulatory Link between NAD Synthesis and Glycolysis. *FEBS J.* 276, 6615–6623. doi:10.1111/j.1742-4658.2009.07372.x
- Huo, L., Davis, I., Chen, L., and Liu, A. (2013). The Power of Two. *J. Biol. Chem.* 288, 30862–30871. doi:10.1074/jbc.M113.496869
- Huo, L., Fielding, A. J., Chen, Y., Li, T., Iwaki, H., Hosler, J. P., et al. (2012). Evidence for a Dual Role of an Active Site Histidine in  $\alpha$ -Amino- $\beta$ -



- carboxymuconate- $\epsilon$ -semialdehyde Decarboxylase. *Biochemistry* 51, 5811–5821. doi:10.1021/bi300635b
- Huo, L., Liu, F., Iwaki, H., Li, T., Hasegawa, Y., and Liu, A. (2015). Human  $\alpha$ -amino- $\beta$ -carboxymuconate- $\epsilon$ -semialdehyde Decarboxylase (ACMSD): A Structural and Mechanistic Unveiling. *Proteins* 83, 178–187. doi:10.1002/prot.24722
- Joosten, R. P., Long, F., Murshudov, G. N., and Perrakis, A. (2014). The PDB\_REDO Server for Macromolecular Structure Model Optimization. *Int. Union Crystallogr. J.* 1, 213–220. doi:10.1107/S2052252514009324
- Kabsch, W. (2010). Xds. *Acta Crystallogr. D Biol. Cryst.* 66, 125–132. doi:10.1107/S0907444909047337
- Katsyuba, E., and Auwerx, J. (2017). Modulating NAD<sup>+</sup> Metabolism, from Bench to Bedside. *EMBO J.* 36, 2670–2683. doi:10.15252/embj.201797135
- Katsyuba, E., Mottis, A., Zietak, M., De Franco, F., van der Velpen, V., Gariani, K., et al. (2018). De Novo NAD<sup>+</sup> Synthesis Enhances Mitochondrial Function and Improves Health. *Nature* 563, 354–359. doi:10.1038/s41586-018-0645-6
- Katsyuba, E., Romani, M., Hofer, D., and Auwerx, J. (2020). NAD<sup>+</sup> Homeostasis in Health and Disease. *Nat. Metab.* 2, 9–31. doi:10.1038/s42255-019-0161-5
- Kellum, J. A., and Prowle, J. R. (2018). Paradigms of Acute Kidney Injury in the Intensive Care Setting. *Nat. Rev. Nephrol.* 14, 217–230. doi:10.1038/nrneph.2017.184
- Kellum, J. A., van Till, J. W. O., and Mulligan, G. (2020). Targeting Acute Kidney Injury in COVID-19. *Nephrol. Dial. Transpl.* 35, 1652–1662. doi:10.1093/ndt/gfaa231
- Krissinel, E., and Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* 372, 774–797. doi:10.1016/j.jmb.2007.05.022
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Cryst.* 26, 283–291. doi:10.1107/S0021889892009944
- Li, T., Iwaki, H., Fu, R., Hasegawa, Y., Zhang, H., and Liu, A. (2006).  $\alpha$ -Amino- $\beta$ -carboxymuconic- $\epsilon$ -semialdehyde Decarboxylase (ACMSD) Is a New Member of the Amidohydrolase Superfamily. *Biochemistry* 45, 6628–6634. doi:10.1021/bi060108c
- Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., et al. (2017). Polder Maps: Improving OMIT Maps by Excluding Bulk Solvent. *Acta Cryst. Sect D Struct. Biol.* 73, 148–157. doi:10.1107/S2059798316018210
- Manrique-Caballero, C. L., Kellum, J. A., Gómez, H., De Franco, F., Giacchè, N., and Pellicciari, R. (2021). Innovations and Emerging Therapies to Combat Renal Cell Damage: NAD<sup>+</sup> as a Drug Target. *Antioxid. Redox Signaling* 35, 1449–1466. doi:10.1089/ars.2020.8066
- Martynowski, D., Eyobo, Y., Li, T., Yang, K., Liu, A., and Zhang, H. (2006). Crystal Structure of  $\alpha$ -Amino- $\beta$ -carboxymuconate- $\epsilon$ -semialdehyde Decarboxylase: Insight into the Active Site and Catalytic Mechanism of a Novel Decarboxylation Reaction. *Biochemistry* 45, 10412–10421. doi:10.1021/bi060903q
- McNicholas, S., Potterton, E., Wilson, K. S., and Noble, M. E. M. (2011). Presenting Your Structures: the CCP 4 mg Molecular-Graphics Software. *Acta Crystallogr. D Biol. Cryst.* 67, 386–394. doi:10.1107/S0907444911007281
- Pellicciari, R., Liscio, P., Giacchè, N., De Franco, F., Carotti, A., Robertson, J., et al. (2018).  $\alpha$ -Amino- $\beta$ -carboxymuconate- $\epsilon$ -semialdehyde Decarboxylase (ACMSD) Inhibitors as Novel Modulators of De Novo Nicotinamide Adenine Dinucleotide (NAD<sup>+</sup>) Biosynthesis. *J. Med. Chem.* 61, 745–759. doi:10.1021/acs.jmedchem.7b01254
- Poyan Mehr, A., Tran, M. T., Ralto, K. M., Leaf, D. E., Washco, V., Messmer, J., et al. (2018). De Novo NAD<sup>+</sup> Biosynthetic Impairment in Acute Kidney Injury in Humans. *Nat. Med.* 24, 1351–1359. doi:10.1038/s41591-018-0138-z
- Pucci, L., Perozzi, S., Cimadamore, F., Orsomando, G., and Raffaelli, N. (2007). Tissue Expression and Biochemical Characterization of Human 2-amino 3-carboxymuconate 6-semialdehyde Decarboxylase, a Key Enzyme in Tryptophan Catabolism. *FEBS J.* 274, 827–840. doi:10.1111/j.1742-4658.2007.05635.x
- Saito, K., Fujigaki, S., Heyes, M. P., Shibata, K., Takemura, M., Fujii, H., et al. (2000). Mechanism of Increases in L-Kynurenine and Quinolinic Acid in Renal Insufficiency. *Am. J. Physiology-Renal Physiol.* 279, F565–F572. doi:10.1152/ajprenal.2000.279.3.F565
- Segel, I. H. (1993). *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*. John Wiley & Sons.
- von Stetten, D., Carpentier, P., Flot, D., Beteva, A., Caserotto, H., Dobias, F., et al. (2020). ID30A-3 (MASSIF-3) - a Beamline for Macromolecular Crystallography at the ESRF with a Small Intense Beam. *J. Synchrotron Radiat.* 27, 844–851. doi:10.1107/S1600577520004002
- Weiss, M. S. (2001). Global Indicators of X-ray Data Quality. *J. Appl. Cryst.* 34, 130–135. doi:10.1107/S0021889800018227
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., et al. (2011). Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr. D Biol. Cryst.* 67, 235–242. doi:10.1107/S0907444910045749
- Yang, Y., Borel, T., De Azambuja, F., Johnson, D., Sorrentino, J. P., Udokwu, C., et al. (2021). Diflunisal Derivatives as Modulators of ACMS Decarboxylase Targeting the Tryptophan-Kynurenine Pathway. *J. Med. Chem.* 64, 797–811. doi:10.1021/acs.jmedchem.0c01762
- Yang, Y., Davis, I., Matsui, T., Rubalcava, I., and Liu, A. (2019). Quaternary Structure of  $\alpha$ -amino- $\beta$ -carboxymuconate- $\epsilon$ -semialdehyde Decarboxylase (ACMSD) Controls its Activity. *J. Biol. Chem.* 294, 11609–11621. doi:10.1074/jbc.RA119.009035
- Yoshino, J. (2019). ACMSD: A Novel Target for Modulating NAD<sup>+</sup> Homeostasis. *Trends Endocrinol. Metab.* 30, 229–232. doi:10.1016/j.tem.2019.02.002
- Zhou, C.-C., Yang, X., Hua, X., Liu, J., Fan, M.-B., Li, G.-Q., et al. (2016). Hepatic NAD<sup>+</sup> deficiency as a Therapeutic Target for Non-alcoholic Fatty Liver Disease in Ageing. *Br. J. Pharmacol.* 173, 2352–2368. doi:10.1111/bph.13513

**Conflict of Interest:** NG, PL, ER, FD, and JR are employees of TES Pharma S.r.l.; RP is President and CEO of TES Pharma S.r.l.; AC is a consultant of TES Pharma S.r.l.; MC, LC, MG, and NR have a research collaboration contract with TES Pharma S.r.l.

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

Copyright © 2022 Cianci, Giacchè, Cialabrini, Carotti, Liscio, Rosatelli, De Franco, Gasparrini, Robertson, Amici, Raffaelli and Pellicciari. 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.



# Experiences From Developing Software for Large X-Ray Crystallography-Driven Protein-Ligand Studies

Nicholas M. Pearce<sup>1\*</sup>, Rachael Skyner<sup>2</sup> and Tobias Krojer<sup>3</sup>

<sup>1</sup>Department of Chemistry and Pharmaceutical Sciences, VU University Amsterdam, Amsterdam, Netherlands, <sup>2</sup>OMass Therapeutics, The Oxford Science Park, Oxford, United Kingdom, <sup>3</sup>MAX IV Laboratory, Lund University, Lund, Sweden

## OPEN ACCESS

### Edited by:

Pietro Roversi,  
Italian National Research Council, Italy

### Reviewed by:

David Ryan Koes,  
University of Pittsburgh, United States  
Marius Schmidt,  
University of Wisconsin–Milwaukee,  
United States

### \*Correspondence:

Nicholas M. Pearce  
n.m.pearce@vu.nl

### Specialty section:

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

**Received:** 24 January 2022

**Accepted:** 14 March 2022

**Published:** 11 April 2022

### Citation:

Pearce NM, Skyner R and Krojer T  
(2022) Experiences From Developing  
Software for Large X-Ray  
Crystallography-Driven Protein-  
Ligand Studies.  
Front. Mol. Biosci. 9:861491.  
doi: 10.3389/fmolb.2022.861491

The throughput of macromolecular X-ray crystallography experiments has surged over the last decade. This remarkable gain in efficiency has been facilitated by increases in the availability of high-intensity X-ray beams, (ultra)fast detectors and high degrees of automation. These developments have in turn spurred the development of several dedicated centers for crystal-based fragment screening which enable the preparation and collection of hundreds of single-crystal diffraction datasets per day. Crystal structures of target proteins in complex with small-molecule ligands are of immense importance for structure-based drug design (SBDD) and their rapid turnover is a prerequisite for accelerated development cycles. While the experimental part of the process is well defined and has by now been established at several synchrotron sites, it is noticeable that software and algorithmic aspects have received far less attention, as well as the implications of new methodologies on established paradigms for structure determination, analysis, and visualization. We will review three key areas of development of large-scale protein-ligand studies. First, we will look into new software developments for batch data processing, followed by a discussion of the methodological changes in the analysis, modeling, refinement and deposition of structures for SBDD, and the changes in mindset that these new methods require, both on the side of depositors and users of macromolecular models. Finally, we will highlight key new developments for the presentation and analysis of the collections of structures that these experiments produce, and provide an outlook for future developments.

**Keywords:** macromolecular crystallography, fragment screening, data management, multi-state modelling, data presentation and analysis

## INTRODUCTION

Modern drug development is an intensely multi-disciplinary exercise that relies on expertise ranging from fundamental biophysics to clinical trials (Kiriiri et al., 2020). Especially in the early stages of small molecule development, structural biology has been instrumental in guiding the rational development of numerous novel small molecule drugs (Maveyraud and Mourey, 2020). Structural knowledge of the interaction of a protein with varied small molecule ligands is used to inform the design process of compounds with improved binding affinities (Hughes et al., 2011). Determination of protein-ligand structures is done by X-ray crystallography (Maveyraud and Mourey, 2020), NMR

(Nitsche and Otting, 2018), Cryo-EM (Renaud et al., 2018), and even recently with MicroED (Clabbers et al., 2020). However, despite recent breakthroughs in other methodologies (most notably cryoEM), X-ray crystallography remains the workhorse for structure-based drug design (SBDD), at least for the time being.

The prevalence of crystallography in these efforts is a testament to the platform methodologies that have been developed to enable routine crystal structure determination. Over the last decade, macromolecular crystallography has seen a remarkable gain in efficiency and throughput driven by improvements in beam intensity and X-ray detectors, the availability of fast and reliable sample changers, and advances in software for data acquisition and analysis (Owen et al., 2016; Förster and Schulze-Briese, 2019). This has culminated in the development of fully automated beamlines that reduce interaction of the scientist with the diffraction experiment to the barest minimum (Svensson et al., 2019). Moreover, so-called “unattended” data collection in combination with automated data processing (Winter, 2010; Vonrhein et al., 2011; Sparta et al., 2016) and refinement pipelines (Sharff et al., 2011; Wojdyr et al., 2013; Echols et al., 2014; Schiebel et al., 2016) running on high-performance computing systems means that it only takes minutes to get from data collection to high-quality electron density maps. Often overlooked in enabling these achievements is the role and importance of sophisticated and robust systems for managing sample logistics, which are now available at most synchrotron sites (Delagenière et al., 2011). When combined with the technical experimental advances, it is these information management systems which truly enable the establishment of routine high-throughput crystallographic experiments (Zheng et al., 2014).

These advancements have culminated in the establishment of several publicly accessible centers for crystal-based fragment screening (Lima et al., 2020; Cornaciu et al., 2021; Douangamath et al., 2021; Wollenhaupt et al., 2021; Kaminski et al., 2022). The extremes of these setups have now transformed protein crystallography from a structure determination method into another biophysical screening assay technique (Douangamath et al., 2020; Schuller et al., 2021; Günther et al., 2021). These facilities have developed several bespoke software solutions for data capture, processing and deposition, incorporating new software packages for restraints generation for new small molecule compounds, model building and refinement (Sparta et al., 2016; Krojer et al., 2017; Long et al., 2017; Pearce et al., 2017c; Cornaciu et al., 2021; Lima et al., 2021), as well as significant advancements in algorithms for detecting (weakly) binding ligands (Pearce et al., 2017b) and analyzing the output chemical information (Deane et al., 2017). Some of these solutions are still confined to specialized screening setups and are thus only used by a small number of protein crystallographers—some because they are not generic enough and some because they are simply unknown.

Here, we provide an overview of the recent history of software tools for large-scale structure determination and data analysis. Based on experiences from fragment screening, we identify a subset of these approaches which we think deserve closer attention and broader awareness within the structure

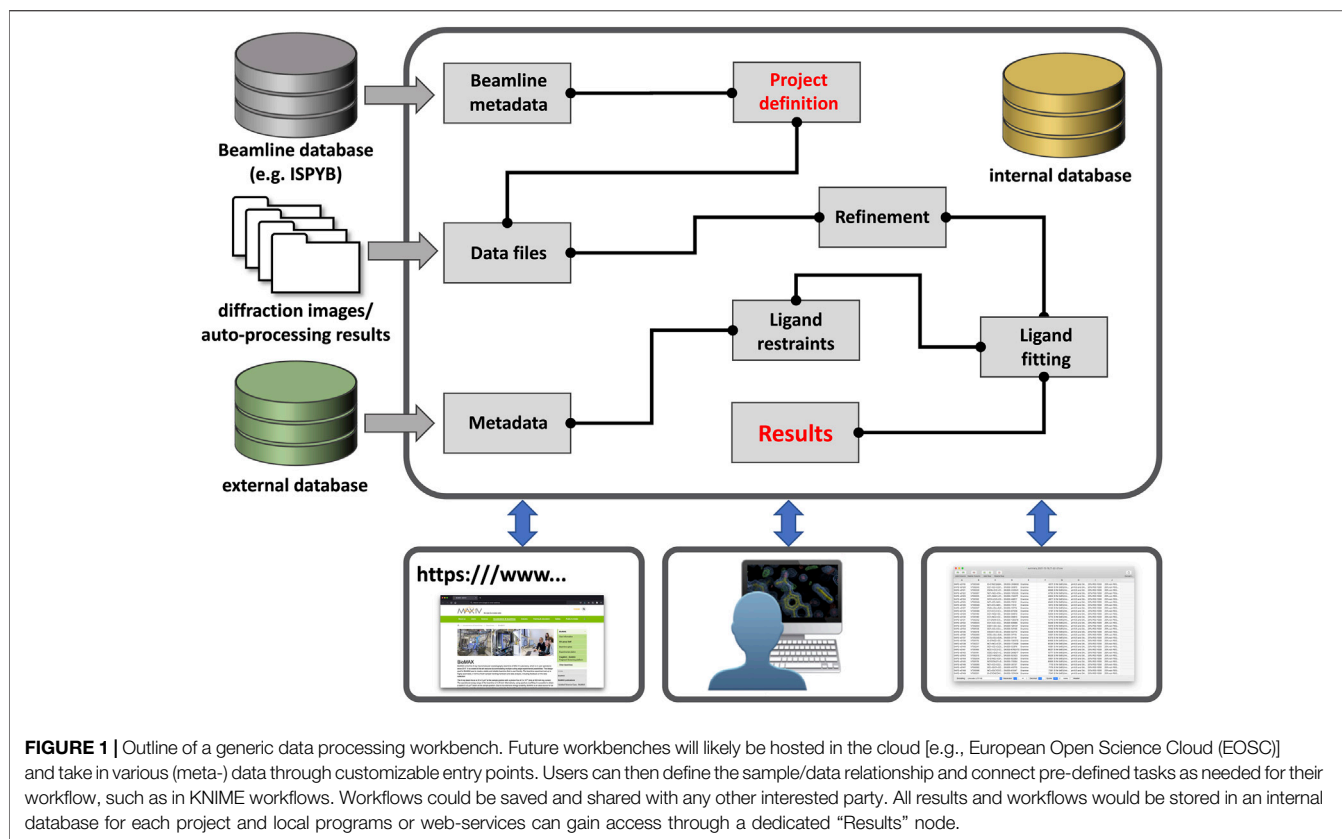
determination community, and ultimately should be incorporated into the main-stream crystallographic toolbox. Furthermore, we highlight recent developments for analyzing protein-ligand structures that bring together structural biologists and computational chemists.

## BATCH DATA PROCESSING AND REFINEMENT TOOLS

Currently, the majority of crystallographic structure determination is done through the graphical interfaces of the CCP4 and PHENIX packages (Echols et al., 2012; Potterton et al., 2018). These offer user-friendly interfaces that guide newcomers as well as experienced users through the structure determination process, and both have large (and overlapping) user bases. However, these fundamentally adhere to a “one structure per project” paradigm, where one data set leads to one atomic model which leads to one deposition in the Protein Data Bank (PDB) (Berman et al., 2000). This is suitable for the determination of (a small number of) novel crystal structures, where each chemically-distinct structure becomes its own project, but far less so for the determination of many related-but-distinct protein-ligand complexes, which can now involve the simultaneous determination of hundreds of structures (Schuller et al., 2021). Moreover, the graphical user interfaces have historically lacked adequate meta-data-tracking functionality and provide no direct connections between the various experimental and computational stages of the experiment—neither cloning, expression, purification, crystallization nor data collection. In reality, and in the absence of established, widely-used and integrated solutions that remove the burden of record keeping, many—if not most—practitioners still record and track their experiments in ad-hoc electronic spreadsheets and physical notebooks.

The lack of suitable large-scale (publicly-available) processing tools became painfully apparent with the establishment of dedicated synchrotron-based centers for crystallographic fragment screening where hundreds of crystals of the same protein are determined in complex with different compounds (Collins et al., 2017; Lima et al., 2020; Cornaciu et al., 2021; Douangamath et al., 2021; Sharpe and Wojdyla, 2021; Wollenhaupt et al., 2021; Kaminski et al., 2022). As a result, three novel programs—CRIMS, FragMAXapp, and XChemExplorer—became available, which are able to process hundreds of related datasets as part of a single session (Krojer et al., 2017; Cornaciu et al., 2021; Lima et al., 2021; Wollenhaupt et al., 2021). These software differ in design, layout, and scope, but all facilitate batch data processing and record a large set of meta-data in a dedicated database system which facilitates project tracking, thereby enabling posterior analysis as well as simplifying PDB deposition (Figure 1).

XChemExplorer is a standalone program developed at the XChem facility at Diamond Light Source for use with fragment screening experiments (Krojer et al., 2017). It enables the selection of auto-processing results, initial refinement, dataset annotation, interactive model building with Coot (Emsley et al.,



2010) and refinement with Refmac (Murshudov et al., 2011) or BUSTER (Bricogne et al., 2017). FragMAXapp was jointly developed between the FragMAX facility at the MAX IV synchrotron and the fragment screening facility at Helmholtz-Zentrum Berlin (HZB) and the BESSY synchrotron (Lima et al., 2021; Wollenhaupt et al., 2021); FragMAXapp offers a web-based platform with similar functionalities to XChemExplorer, but with more customization options for diffraction data processing and initial refinement. Both of these software solutions allow a series of data sets to be consolidated and organized into a single project, and provide launchpads for additional processing with programs such as PanDDA (Pearce et al., 2017b). Finally, CRIMS is a web-based platform developed at the European Synchrotron Radiation Facility (ESRF) and is probably the most comprehensive of the three in terms of meta-data tracking and database integration (Cornaciu et al., 2021); this takes advantage of the extensive database infrastructure at the ESRF High-Throughput Crystallization (HTX) facility and can also directly communicate with the ISPYB laboratory information management system that is widely used at synchrotron beamlines (Delagenière et al., 2011). As a demonstration of the importance of these tools, and their future development, they were instrumental in facilitating the rapid solution and availability of numerous protein-ligand structures of several proteins from Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in the wake of the coronavirus disease (COVID-19) pandemic (Douangamath et al., 2020; Günther et al., 2021; Newman et al., 2021; Schuller et al.,

2021; Kozielski et al., 2022). However, availability of these programs is currently restricted to certain synchrotron sites and they do not allow for customized workflow configuration or provide interfaces for definition of new experiments (Moreno-Chicano et al., 2019; Brändén and Neutze, 2021; Schulz et al., 2022).

## IDENTIFICATION, MODELING AND REFINEMENT OF BINDING MOLECULES

The second problem that is encountered during crystallographic fragment screening experiments—after the processing of the crystallographic data—is the identification of binding molecules. In a non-focussed fragment screen, only a small percentage of fragments are expected to bind at a particular location on the protein surface, but since fragments may bind at any location on the protein surface, manual inspection of the data quickly becomes infeasible, meaning that automated methods are required (Pearce et al., 2017b).

The most popular crystallographic methods for identifying and validating the presence of bound ligands revolve around internal consistency metrics, i.e., metrics which compare the atomic model and the experimental electron density (Echols et al., 2014; Liebschner et al., 2017). These are used both to identify interesting areas of the electron density map, and to validate the atomic models produced: First, “blobs” are identified by calculating one of a variety of difference maps, and then,



constructed ligand models are validated by comparing the model to the experimental electron density. For the purposes of modeling, most methods seek to generate an electron density map that is minimally biased towards the refined atomic model, and which is quantified in terms of a robust signal-to-noise ratio to indicate the importance of an electron density feature. Most approaches therefore utilize different flavors of OMIT maps—in particular OMIT difference maps—which provide a measure of the electron density in a region in terms of the global model error (in units of the rmsd-value of the difference map). This is typically done at the 3-rmsd level (also called the 3-sigma level). Atoms in a selected region must be removed to calculate the difference map, and are either replaced by bulk solvent (Bhat and Cohen, 1984; Terwilliger et al., 2008; Pražnikar et al., 2009) or by vacuum (e.g., polder maps) (Vonnrhein and Bricogne, 2005; Liebschner et al., 2017).

It is vitally important, and perhaps underappreciated, to understand exactly what the hereby-identified features in difference maps represent. For the interrogated region, these difference maps measure the density at a particular site relative to either vacuum (polder maps) or bulk solvent (other OMIT maps) in units of the error in your model: significant differences (typically those above 3-rmsd) therefore indicate electron density features that very likely are not experimental noise. However, this indication is different to a measure of signal (such as the presence of an unmodelled bound ligand), since density from any molecule(s), e.g., semi-ordered water molecules, may also produce electron density above this noise level—i.e., it is not only ligand binding that produces electron density in these maps. We should generally expect a map such as a polder map to show a significant amount of density for most sites on the protein surface (except in very poorly refined models or at lower resolution), since even semi-ordered solvent is likely more than three noise units higher than vacuum for high resolution data. Care must therefore be taken when using OMIT methods so as not to overinterpret the maps, whose purpose is only to quantify and show clear unbiased density for a region.

However it is common for the presence of any difference density in an OMIT map to be presented as (incontrovertible) evidence for the presence of a ligand, with even clear cases of mismodeling being difficult to refute (Stanfield et al., 2016a; Stanfield et al., 2016b). The presence of difference density in a particular map—of the appropriate shape and size—may even be misconstrued as evidence for binding by even an experienced user, especially when solving multiple structures with a set of small molecules, since at least one molecule is likely to match a putative blob in a binding site purely by chance, due to the molecules' small sizes and simple shapes.

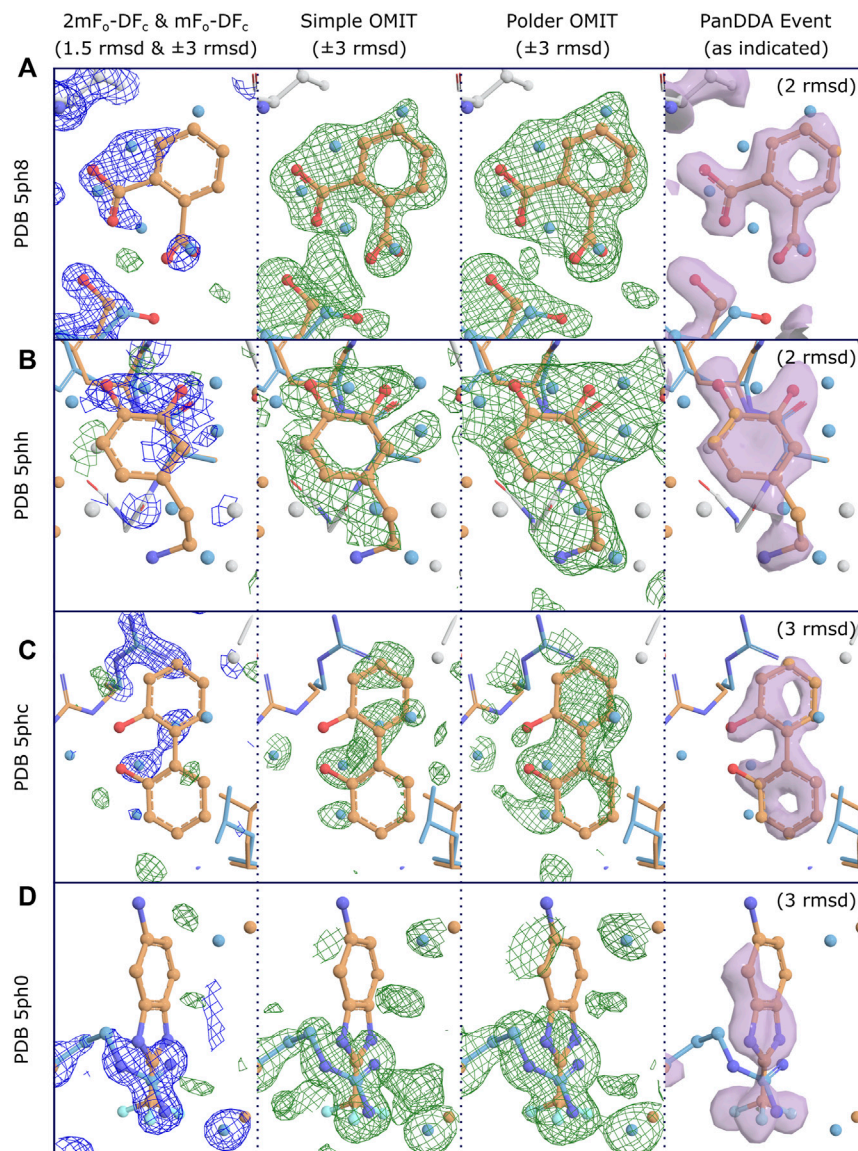
The reason for disagreement about interpretation stems from the nature of OMIT maps. For large strongly-binding ligands, and given medium- to high-resolution data, difference map methods will work well for accurately identifying binding poses, since the correct solution will be well-defined and unambiguous. However, for weakly-binding or smaller compounds, in part due to the reasons described above, these methods are worryingly vulnerable to false positives (incorrectly modeling a ligand which did not bind and is not present in the electron density),

as there is no objective approach for determining whether a blob is due to a bound ligand, or due to (semi-) ordered solvent or other molecules in the binding site (Stanfield et al., 2016a; Stanfield et al., 2016b), even when the difference OMIT map shows electron density above a certain threshold. Without robust metrics, what constitutes unambiguity becomes a subjective measure. Opportunities for map misinterpretation are further exacerbated by partial-occupancy ligands, as will be discussed in detail below (**Figure 2**). Conversely, because the noise level of difference maps is related to the quality of the model, and typical macromolecular atomic models are still generally rather poor (*R*-values of greater than 20% are still the norm), difference maps are not able to identify more weakly binding ligands, since these will fall below the “noise level” of a difference map, leading to false negatives (failing to identify a ligand which did bind).

The large-scale availability of related crystallographic datasets from fragment screening experiments enabled the development of a data-driven multi-dataset ligand identification method: PanDDA (Pearce et al., 2017b). This approach aligns and compares electron density maps from different datasets and identifies local outliers in datasets which deviate from the population of electron density maps; this can be thought of as the multi-dataset generalization of isomorphous-difference (Fo-Fo) maps (Rould and Carter, 2003). For this to work, a number of datasets must not contain binders (i.e., be APO or “ground-state” datasets) against which putative binding datasets can be contrasted. For unfocussed fragment screening experiments, most datasets do not contain a binding fragment, which then constitute APO datasets. Alternately, when performing a fragment screen, a series of true APO datasets can also be collected and used, as is regularly performed at the XChem facility (Douangamath et al., 2021). For datasets with identified outliers, these outliers indicate an “event” at that location in that dataset, which gives strong evidence that a change has occurred in this dataset, and can—in contrast to OMIT maps—be used as a measure of signal. In fragment screening data, events are generally binding ligands, though some random structural changes and processing artifacts can also occur.

An additional insight while developing the PanDDA approach was that the majority of the identified binding events were partial-occupancy features, meaning that the fragment was only bound to a fraction of the protein molecules in the crystal—this is to be routinely expected for fragments, which have low binding affinities. In these cases, which constitute the large majority of identified fragments (Pearce et al., 2017b), the observed electron density is a weighted summation of the electron density for the fragment and also the “apo” or “ground-state” of the crystal, e.g., (dis)ordered water or other solvent molecules. Interpreting the raw experimental electron density map (or any kind of OMIT map) is not feasible in these cases, and indeed should be actively avoided, for fear of misinterpreting the superposed solvent density as part of the ligand density (**Figure 2**); this could cause mismodelling of the ligand, and would thereby mislead downstream medicinal chemists.

The real power of the PanDDA approach is the overcoming of the partial-occupancy obstacle by estimating the occupancy of the



**FIGURE 2 |** Conventional OMIT maps do not generally produce clear unambiguous evidence of binding for low occupancy ligands, even at high resolution.

Examples for four binding fragments showing different levels of support for binding from OMIT maps, where clear evidence of binding is shown by PanDDA event maps. Map coloring:  $2mF_o-DF_c$  maps are shown as blue mesh and all difference maps are shown as green/red mesh; associated PanDDA event maps are shown as purple surfaces. Map type and contour are as indicated. Both types of OMIT maps are produced by phenix.polder. Maps are truncated (carved) at  $3\text{\AA}$  around relevant residues and ligands for clarity; PanDDA maps are carved at  $2\text{\AA}$ . Model coloring: To distinguish alternate conformations, carbon atoms for ligand-associated conformations are coloured orange and non-ligand-associated conformations are coloured blue; main-conformation (full-occupancy) atoms are coloured light gray; all other atoms are colored by element except waters, which are coloured as per carbon atoms. Resolutions: **(A)**  $1.40\text{\AA}$ , **(B)**  $1.60\text{\AA}$ , **(C)**  $1.29\text{\AA}$ , **(D)**  $1.34\text{\AA}$ . Refined ligand occupancies: **(A)** 0.41, **(B)** 0.50, **(C)** 0.38, **(D)** 0.22. PanDDA event map pseudo-occupancies (1-BDC): **(A)** 0.15, **(B)** 0.17, **(C)** 0.10, **(D)** 0.11. **(A)** Binding is not evident in the  $2mF_o-DF_c$  maps at a moderate contour level, but is clearly supported by both types of OMIT map, especially when considered in combination with the extra density from the superposed water molecules from non-ligand-associated conformations, as modeled. It is debatable whether the OMIT maps alone would provide strong enough evidence to support modeling of the ligand, but the single ligand conformation is clearly evidenced in the PanDDA event map, preventing potential misinterpretation of the OMIT map as multiple conformations of the ligand. **(B)** Similar to **(A)**, but with less evidence in the simple OMIT map. The polder OMIT map provides an envelope which fits well with the envelope provided by the ligand and the superposed water molecules, as modeled. It is unlikely either OMIT map would be accepted as evidence of binding, but once more, the ligand conformation is clearly identified in the event map. **(C)** OMIT maps show mostly features which correspond to superposed (not-ligand-associated) waters, and do not present evidence for the bound ligand, unlike the event map. **(D)** Ligand binding coincides with an alternate conformation of an arginine residue, which dominates the refined maps and OMIT maps.

superposed ground state, and subsequently subtracting the appropriate fraction of the ground-state density, which is derived from the analysis of the ground-state (APO) datasets.

This subtraction reveals an approximation to the experimental density for the bound state only, i.e., what would be obtained if the ligand was bound at full occupancy. This “event map” can

then be used for ligand modeling (**Figure 2**). Several examples in the original PanDDA manuscript show examples of where interpreting the standard electron density maps lead to misinterpretation, as the standard experimental maps shows diffuse density that could be interpreted as the ligand in multiple conformations; however after application of the PanDDA approach, the ligand is clearly in one conformation, with a superposed solvent state (Pearce et al., 2017a; Pearce et al., 2017b; Pearce et al., 2017c).

In the original PanDDA implementation, no assumptions are made about whether datasets are comparable, and it has since been shown that pre-processing with methods such as cluster4x can dramatically increase the sensitivity of the method even further (Ginn, 2020). Combined, these approaches enable the identification and modeling of very low-affinity compounds which would not have been identified previously, and in doing so have greatly increased the amount of chemical matter available for ligand binding studies. However, the identification of weaker- and weaker-binding ligands have revealed significant weaknesses in currently available approaches for structure determination and usage.

In light of the experiences in fragment screening, it is clear that there is a large and underappreciated potential for misinterpreting electron density when modeling ligands, either because the ligand is not there at all (Stanfield et al., 2016b), or because the ligand binds at subunitary occupancy. Since we cannot know the occupancy of the ligand a priori, we may unwittingly be misinterpreting density for another molecule as density for the ligand, resulting in the wrong pose for (parts of) the ligand (Pearce et al., 2017b). Ligands in the PDB are routinely modeled at full occupancy (Pearce et al., 2017c), showing that partial occupancy is rarely considered. Our experiences with fragment screening have provided numerous cases where the outcome of a PanDDA analysis is very different to what might have been created using traditional approaches (Pearce et al., 2017a; Pearce et al., 2017b; Pearce et al., 2017c), including cases where the ligand was initially perceived to be in multiple conformations, but was in fact in one conformation, and vice versa. The exact prevalence of such mismodelling in the PDB is unknown, and though errors may be minor, inaccurate binding poses of important functional groups could seriously mislead downstream applications such as structure-based drug design.

Since sub-unitary occupancy became an inherent feature of the crystallographic data, this spurred the development of methods for the generation and refinement of multi-state models for superpositions of ligand-bound and ground-state states (Pearce et al., 2017b; Pearce et al., 2017c). Combining the ligand-bound model (derived from the event map) with the ground-state model (derived from the APO data sets) generates a multi-state model for refinement (Pearce et al., 2017c), and combined with appropriate occupancy restraints, this can produce relatively high-quality models for even low-occupancy ligands (Pearce et al., 2017b), although of course these models are lower quality than one would expect from stoichiometric binders. Visual inspection of the refined electron density maps becomes less useful for low-occupancy ligands, since all refined electron density maps will continue to contain multi-state superpositions,

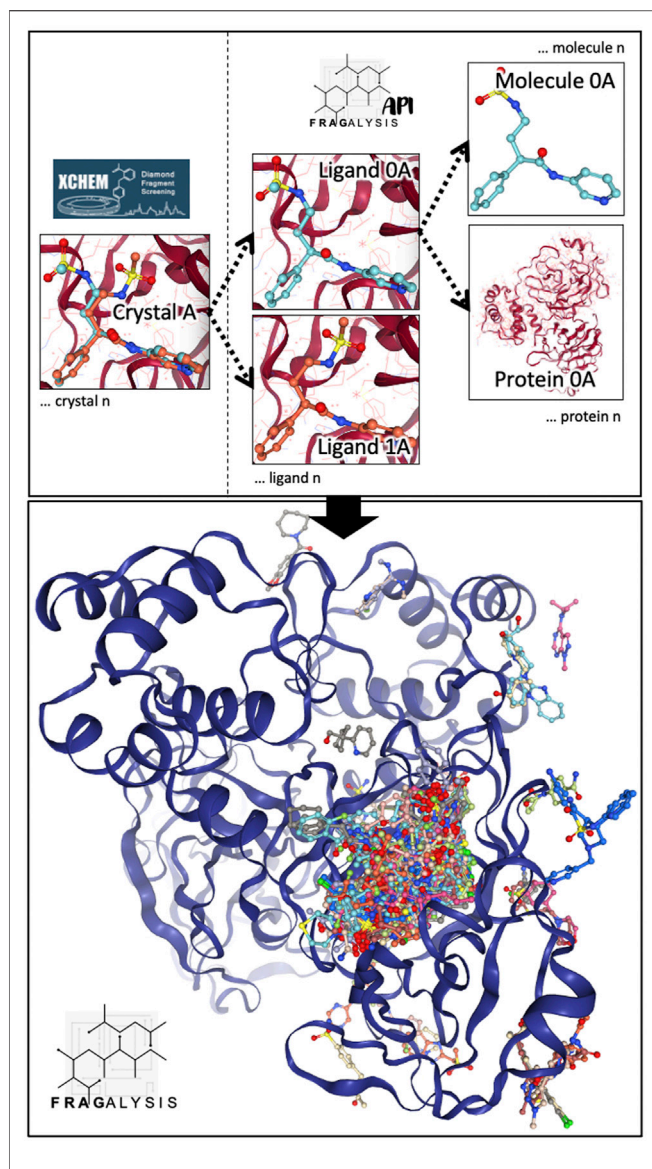
and so multiple validation metrics are useful for identifying errant parametric features of the models, such as inappropriate occupancies or B-factors (Pearce et al., 2017b; Pearce et al., 2017c). Iterative modeling using this multi-state approach requires a new mindset for the crystallographer, since it requires the different states of the model to be inspected separately and modeled into different electron density maps (i.e., the ground-state conformation into a ground-state map and the bound-state conformation into an event map), before being recombined for refinement against the original experimental data. This is currently technically difficult, and tools need to be further developed before this can be routinely applied by non-experts. However, routinely ignoring the partial-occupancy nature of ligands in crystallographic models is a significant oversight within the field, the current extent and effect of which is simply unknown.

## STRUCTURE PRESENTATION AND ANALYSIS

With the influx of large volumes of structural data, such as that generated by synchrotron fragment-screening facilities, application-specific platforms for exploring and exploiting this data become a critical component in the context of fragment-based lead design (FBLD) (Bradley et al., 2015). Popular desktop-based software for the visualizing and analyzing crystal structures (e.g., PyMOL) are only tractable for a handful of structures, and are ineffective in identifying trends in data which could inform further FBLD (Deane et al., 2017). Curating and understanding of FBLD outputs is time-consuming due to the vast number of structures, and often rely heavily on the expertise of the researcher to keep track of interesting and potentially novel features from their analysis (de Souza Neto et al., 2020).

The usage of multi-state models introduces an additional difficulty when using these models in downstream applications. Users of the PDB are mostly not trained crystallographers, and can be confused by the presence of multiple superposed conformations in an atomic model—in the best case scenario, they are simply a nuisance artifact to be removed. Therefore, it is preferable to remove the superposed ground-state conformations from models, and present only the scientifically-interesting bound state. However, the PDB currently has no mechanism for presenting different states of the model, and thus depositors are left with a choice: deposit the full multi-state crystallographic model or deposit only the bound-state model. While the second option—which is the one that has been adopted by the XChem facility—is beneficial to the users of the PDB (who are after all the intended audience), this practice introduces problems for those that wish to reproduce model refinements, since only part of the model has been deposited; this has caused these models to be accurately identified as being poor and unreliable by validation efforts (Wlodawer et al., 2020). Now that hundreds of such models are being deposited in the PDB every year, there is urgently-needed functionality in the PDB for presenting the different components of multi-state models to end users, so that the full crystallographic model can be deposited.





**FIGURE 3 |** Fragalysis aims to provide immediate access to ligand-protein information without confounding crystallographic artifacts. To achieve this, a given crystal structure (top left—Crystal A) is inspected to find all of the individual ligands. These ligands are then separated into separate bound-state entities (top right—Ligand 0A and Ligand 1A) using the Fragalysis API. Ligands are subsequently separated from their respective protein (top-right Ligand 0A and Protein 0A), and presented in Fragalysis as part of an ensemble of all ligands and proteins in the same reference frame (bottom).

The development of routines which further allow the reproduction of PanDDA results are another critical area of required development to ensure that refinements of these models are reproducible and that the models can be validated.

The need for rapid access to 3D structural data necessary for competitive design of lead series in FBLD was brought into sharp focus by the coronavirus pandemic—even once the structures are processed and aligned, annotation and interpretation of this vast amount of information remains a formidable task. A plethora of structural data concerning SARS-CoV-2 and related proteins has

been generated, but required a specialized data portal (PDBe-KB Covid-19 portal; <https://www.ebi.ac.uk/pdbe/covid-19>) (PDBe-KB consortium, 2022) to present data and findings in a coordinated way, including such information as 1) biological function, 2) display of a representative structure, 3) 3D superposed views of the structures and ligands, and 4) lists of relevant publications. Some generic online resources, for example michelaNGLo (Ferla et al., 2020), allow scientists to collaboratively annotate structural data with additional scientific context through an open-source web-based application that allows the creation and sharing of interactive pages containing interactive 3D representations of macromolecular data. Whilst this is extremely powerful, visitors to such pages should be aware that the structural data has often not been officially peer reviewed (though we must also note that peer review of structures in the PDB is not universal). However, since it is well-known that many structures in the PDB are not well-refined or have other serious problems—and that structures are often only made available to reviewers upon request—in both cases it is ultimately still up to individual users to appropriately interpret the data.

The Fragalysis platform combines the collaborative nature of an online discussion tool with the abilities to interrogate hundreds of crystal structures simultaneously. Fragalysis (<https://fragalysis.diamond.ac.uk>) is an open-source web-based application which was designed for the dissemination, evaluation and elaboration of fragment screening results from the XChem screening facility (<https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening.html>), and is aimed at the non-expert user to facilitate the progression of initial fragment hits from the beamline to more potent protein inhibitors. This is achieved by providing context to crystallographic data: All ligands in each crystal structure are treated as an individual entity, and for each ligand, the protein and ligand can be inspected or analyzed in isolation or together as part of an ensemble (Figure 3). This overlay of structural information provides the user with a starting point to consider how they might apply the existing data to the elaboration and prioritization of new molecules that aim to bind more potently.

In the recent development of Fragalysis, emphasis has been put onto how crystallographers can best communicate the key features and limitations of their crystallographic models to non-expert users, including the presentation of electron density maps—PanDDA event maps, refined  $2mF_o-DF_c$  maps and  $mF_o-DF_c$  maps—for newer public datasets (e.g., SARS-CoV-2 Main Protease: <https://fragalysis.diamond.ac.uk/viewer/react/preview/target/Mpro>), and a new paradigm whereby atoms and bonds in ligands can be highlighted and commented upon. This both allows experts to annotate the structures for non-expert users, but also allows users to interrogate the experimental data themselves. These features will be presented fully in future publications. However, it is important to highlight that such developments are imminently necessary more widely in macromolecular crystallography, and structural biology in general. With such a massive amount of data available, and with so many computational methods in FBLD starting from crystallographic models, it is important that the structure-determination community takes responsibility for properly



communicating and annotating their data and aiding the wider structural community in how to best interpret and make use of the data beyond crystallographic modeling. Of particular importance is the communication of quality considerations; here, several articles and reviews describe valuable tools for evaluation of raw experimental data and solved macromolecular structures (Kleywegt, 2000; Gore et al., 2012; Adams et al., 2016; Wlodawer, 2017).

Of course, this accumulation and digestion of populations of structures will likely be greatly affected in the near-term by the rapid advances of deep-learning structure-prediction methods such as AlphaFold (Jumper et al., 2021), and RoseTTAFold (Baek et al., 2021). Neither AlphaFold nor RoseTTAFold (in their current released versions) incorporate the ability to add native ligands/cofactors, although a cavity which could accommodate one is often observed, nor allow the predicted binding poses of arbitrary ligands, but the development of these functionalities is inevitable and much anticipated.

## DISCUSSION AND CONCLUSION

Despite the gains in efficiency in crystallographic experiments at synchrotrons, driven in part by automated fragment-screening beamlines, there has been no real corresponding increase of X-ray crystal structures in the PDB. The number of yearly released X-ray structures has remained largely the same since 2016 (with a COVID-19-driven exception for 2020). The number of structures that are deposited in the PDB is almost certainly significantly smaller than the real number of datasets collected or structures determined every year, especially since only a small fraction of the vast number of structures determined by pharmaceutical companies ends up in the public domain (Mullard, 2021). Nevertheless, it still takes a remarkably long time to get structures “deposition ready” and then into the PDB. There are two main problems: 1) Structure refinement does not have a clear endpoint and crystallographers often go through many refinement and model rebuilding iterations before they are comfortable depositing their data, and 2) there are hardly any tools available that help with data organization, metadata capture and large-scale PDB deposition. This leads to a situation where crystallographers tend to spend a disproportionate amount of time on finalizing what are usually considered “simple” protein-ligand structures. Consequently, this time is not available for actual structure analysis, which is further exacerbated by a lack of tools for parallel and comparative analysis of related crystal structures, combined with suitable graphical presentation.

Structural biology needs more dedicated and integrated tools for batch data processing, otherwise the gains in experimental efficiency will not result in a corresponding explosion of structural information. CRIMS, FragMAXapp and XChemExplorer begin to tackle these data-organization problems, but they are specialized for the environment where they have been developed, lack flexible workflow configuration, and are restricted to the determination of protein-ligand complexes by X-ray crystallography. While the details still need to be worked out, we would like to outline what such a novel and modular batch-processing workbench could look like.

Such a platform should allow flexible and abstract workflow configuration with dedicated APIs for interaction with external databases or processing tools (Figure 1). It should facilitate direct usage of auto-processing results obtained at different synchrotrons and the meta-data stored in databases like ISPYB. This would significantly speed up the process, reduce errors and unburden users from the tedious task of data capture and management of vast amounts of raw diffraction data. Such a tool would not only be useful for single crystal diffraction experiments, but for any other multi-dataset experiment, e.g., serial synchrotron crystallography (SSX) data collections, whether for protein-ligand or time-resolved studies.

Furthermore, there are several paradigmatic lessons for routine macromolecular crystallography to adopt from fragment screening experiments. Our models need to become much more complex than they have been historically, and the databases and visualization methods for presenting these models need to develop quickly to account for this. In the short term, there is still much to be done to convince the community that more complex approaches—i.e., multi-state superpositions—are necessary, but done correctly, more complex structure-determination paradigms should allow for much more robust validation protocols than the current self-consistency metric paradigms. These problems in structure determination and validation are further exacerbated by the sheer number of structures that can be determined, and analytical and visualization methods for identifying trends and common features in these structures are due for an overhaul. Lastly, arguably the greatest responsibility of the macromolecular crystallographic community is to ensure that their data, especially when not proprietary, are released to the public in a timely fashion, and with clear and concise context that make them interpretable to observers and users who may or may not have the necessary expertise to analyze them critically themselves.

These changes will be a great challenge for the developers of databases and software for the deposition, presentation, interpretation and analysis of crystallographic data. These changes should be seen as an opportunity for those developers to redefine how we think about crystallographic data, and to treat this challenge as an exciting new area of scientific research. It is probable that the majority of applications will continue to move towards being web-based, and that more remote computing resources will need to be made available on a global level through academic funding routes. It should also be (made) apparent to funding bodies that disciplines like crystallography don't function without robust, scalable, and sustainable software behind the scenes, and that more needs to be done to ensure secure medium- and long-term funding for software so that it can be developed and maintained appropriately. This must include funding streams to support individual research software development.

Conversely, it is the responsibility of computational scientists who develop new methods and algorithms to properly explain the relevant applications of their methods, and to ensure that those methods are made available where they are not proprietary. This includes publishing them in a version-controlled environment such as GitHub, and publishing links to these repositories within

their manuscripts. It is our opinion that making data and algorithms ‘available upon request’ is no longer good enough, given the availability of a number of easy-to-use public repositories such as Zenodo (European Organization For Nuclear Research and OpenAIRE, 2013), where arbitrary files can be uploaded; to this point, the data and scripts for reproducing the OMIT maps for **Figure 2** have been uploaded to Zenodo (<https://doi.org/10.5281/zenodo.6334726>).

Though the experiences and methods highlighted in this manuscript primarily arose from highly specialized experiments in fragment-based discovery, we believe many of the approaches should inspire developments in other areas of macromolecular crystallography. For instance, the multi-dataset-management approach to experimental crystallography is one that might greatly smooth the process of determining a (set of) typical crystallographic structure(s), and remove much of the tedious meta-data tracking, whilst more technical model-building aspects also have clear applications in more niche crystallographic experiments such as time-resolved crystallography, Laue crystallography, and X-ray Free Electron Laser (XFEL) experiments.

## REFERENCES

- Adams, P. D., Aertgeerts, K., Bauer, C., Bell, J. A., Berman, H. M., Bhat, T. N., et al. (2016). Outcome of the First wwPDB/CCDC/D3R Ligand Validation Workshop. *Structure* 24, 502–508. doi:10.1016/j.str.2016.02.017
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* 373, 871–876. doi:10.1126/science.abj8754
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242. doi:10.1093/nar/28.1.235
- Bhat, T. N., and Cohen, G. H. (1984). OMITMAP: An Electron Density Map Suitable for the Examination of Errors in a Macromolecular Model. *J. Appl. Cryst.* 17, 244–248. doi:10.1107/S0021889884011456
- Bradley, A. R., Wall, I. D., von Delft, F., Green, D. V. S., Deane, C. M., and Marsden, B. D. (2015). WONKA: Objective Novel Complex Analysis for Ensembles of Protein-Ligand Structures. *J. Comput. Aided Mol. Des.* 29, 963–973. doi:10.1007/s10822-015-9866-z
- Brändén, G., and Neutze, R. (2021). Advances and Challenges in Time-Resolved Macromolecular Crystallography. *Science* 373. doi:10.1126/science.aba0954
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., et al. (2017). *BUSTER*. Cambridge, United Kingdom: Global Phasing Ltd.
- Clabbers, M. T. B., Fisher, S. Z., Coinçon, M., Zou, X., and Xu, H. (2020). Visualizing Drug Binding Interactions Using Microcrystal Electron Diffraction. *Commun. Biol.* 3, 417. doi:10.1038/s42003-020-01155-1
- Collins, P. M., Ng, J. T., Talon, R., Nekrosiute, K., Krojer, T., Douangamath, A., et al. (2017). Gentle, Fast and Effective crystal Soaking by Acoustic Dispensing. *Acta Cryst. Sect. D Struct. Biol.* 73, 246–255. doi:10.1107/S205979831700331X
- Cornaci, I., Bourgeois, R., Hoffmann, G., Dupeux, F., Humm, A. S., Mariaule, V., et al. (2021). The Automated Crystallography Pipelines at the EMBL HTX Facility in Grenoble. *Jove-Journal of Visualized Experiments* 5.
- de Souza Neto, L. R., Moreira-Filho, J. T., Neves, B. J., Maidana, R. L. B. R., Guimarães, A. C. R., Furnham, N., et al. (2020). In Silico Strategies to Support Fragment-To-Lead Optimization in Drug Discovery. *Front. Chem.* 8, 93. doi:10.3389/fchem.2020.00093
- Deane, C. M., Wall, I. D., Green, D. V. S., Marsden, B. D., and Bradley, A. R. (2017). WONKAandOOMPPAA: Analysis of Protein-Ligand Interaction Data to

## DATA AVAILABILITY STATEMENT

OMIT map figure data has been uploaded to Zenodo (<https://doi.org/10.5281/zenodo.6334726>). Further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

NP, RS, and TK conceptualised and wrote the manuscript.

## ACKNOWLEDGMENTS

TK acknowledges funding from the Swedish Research Council (Grant No. 2018-06454) and support by iNEXT-Discovery, project number 871037, funded by the Horizon 2020 program of the European Commission. Nicholas Pearce acknowledges funding from a Veni Fellowship, fellowship number VI.Veni.192.143 from the Dutch Research Council (NWO).

- Direct Structure-Based Drug Design. *Acta Cryst. Sect. D Struct. Biol.* 73, 279–285. doi:10.1107/S2059798316009529
- Delageniere, S., Brenchereau, P., Launer, L., Ashton, A. W., Leal, R., Veyrier, S., et al. (2011). ISPyB: an Information Management System for Synchrotron Macromolecular Crystallography. *Bioinformatics* 27, 3186–3192. doi:10.1093/bioinformatics/btr535
- Douangamath, A., Fearon, D., Gehrtz, P., Krojer, T., Lukacik, P., Owen, C. D., et al. (2020). Crystallographic and Electrophilic Fragment Screening of the SARS-CoV-2 Main Protease. *Nat. Commun.* 11, 5047. doi:10.1038/s41467-020-18709-w
- Douangamath, A., Powell, A., Fearon, D., Collins, P. M., Talon, R., Krojer, T., et al. (2021). Achieving Efficient Fragment Screening at Xchem Facility at diamond Light Source. *JoVE*. doi:10.3791/62414
- Echols, N., Grosse-Kunstleve, R. W., Afonine, P. V., Bunkóczi, G., Chen, V. B., Headd, J. J., et al. (2012). Graphical Tools for Macromolecular Crystallography in PHENIX. *J. Appl. Cryst.* 45, 581–586. doi:10.1107/S0021889812017293
- Echols, N., Moriarty, N. W., Klei, H. E., Afonine, P. V., Bunkóczi, G., Headd, J. J., et al. (2014). Automating Crystallographic Structure Solution and Refinement of Protein-Ligand Complexes. *Acta Cryst. D Biol. Crystallogr.* 70, 144–154. doi:10.1107/S139900471302748X
- Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and Development of Coot. *Acta Crystallogr. D Biol. Cryst.* 66, 486–501. doi:10.1107/S0907444910007493
- European Organization For Nuclear Research, and OpenAIRE (2013). *Zenodo*. CERN. doi:10.25495/7gkx-rd71
- Ferla, M. P., Pagnamenta, A. T., Damerell, D., Taylor, J. C., and Marsden, B. D. (2020). MichelaNglo: Sculpting Protein Views on Web Pages without Coding. *Bioinformatics* 36, 3268–3270. doi:10.1093/bioinformatics/btaa104
- Förster, A., and Schulze-Bries, C. (2019). A Shared Vision for Macromolecular Crystallography over the Next Five Years. *Struct. Dyn.* 6, 064302. doi:10.1063/1.5131017
- Ginn, H. M. (2020). Pre-clustering Data Sets Using Cluster4x Improves the Signal-To-Noise Ratio of High-Throughput Crystallography Drug-Screening Analysis. *Acta Cryst. Sect. D Struct. Biol.* 76, 1134–1144. doi:10.1107/S2059798320012619
- Gore, S., Velankar, S., and Kleywegt, G. J. (2012). Implementing an X-ray Validation Pipeline for the Protein Data Bank. *Acta Crystallogr. D Biol. Cryst.* 68, 478–483. doi:10.1107/S0907444911050359
- Günther, S., Reinke, P. Y. A., Fernández-García, Y., Lieske, J., Lane, T. J., Ginn, H. M., et al. (2021). X-ray Screening Identifies Active Site and Allosteric Inhibitors

- of SARS-CoV-2 Main Protease. *Science* 372, 642–646. doi:10.1126/science.abf7945
- Hughes, J., Rees, S., Kalindjian, S., and Philpott, K. (2011). Principles of Early Drug Discovery. *Br. J. Pharmacol.* 162, 1239–1249. doi:10.1111/j.1476-5381.2010.01127.x
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2
- Kaminski, J. W., Vera, L., Stegmann, D. P., Vering, J., Eris, D., Smith, K. M. L., et al. (2022). Fast Fragment- and Compound-Screening Pipeline at the Swiss Light Source. *Acta Cryst. Sect D Struct. Biol.* 78, 328–336. doi:10.1107/S2059798322000705
- Kiriiri, G. K., Njogu, P. M., and Mwangi, A. N. (2020). Exploring Different Approaches to Improve the success of Drug Discovery and Development Projects: a Review. *Futur. J. Pharm. Sci.* 6, 27. doi:10.1186/s43094-020-00047-9
- Kleywegt, G. J. (2000). Validation of Protein crystal Structures. *Acta Crystallogr. D Biol. Cryst.* 56, 249–265. doi:10.1107/s0907444999016364
- Kozielski, F., Sele, C., Talibov, V. O., Lou, J., Dong, D., Wang, Q., et al. (2022). Identification of Fragments Binding to SARS-CoV-2 Nsp10 Reveals Ligand-Binding Sites in Conserved Interfaces between Nsp10 and Nsp14/nsp16. *RSC Chem. Biol.* 3, 44–55. doi:10.1039/d1cb00135c
- Krojer, T., Talon, R., Pearce, N., Collins, P., Douangamath, A., Brandao-Neto, J., et al. (2017). TheXChemExplorer graphical Workflow Tool for Routine or Large-Scale Protein-Ligand Structure Determination. *Acta Cryst. Sect D Struct. Biol.* 73, 267–278. doi:10.1107/S2059798316020234
- Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., et al. (2017). Polder Maps: Improving OMIT Maps by Excluding Bulk Solvent. *Acta Cryst. Sect D Struct. Biol.* 73, 148–157. doi:10.1107/S2059798316018210
- Lima, G. M. A., Jagudin, E., Talibov, V. O., Benz, L. S., Marullo, C., Barthel, T., et al. (2021). FragMAXapp: Crystallographic Fragment-Screening Data Analysis and Project-Management System. *Acta Cryst. Sect D Struct. Biol.* 77, 799–808. doi:10.1107/S2059798321003818
- Lima, G. M. A., Talibov, V. O., Jagudin, E., Sele, C., Nyblom, M., Knecht, W., et al. (2020). FragMAX: the Fragment-Screening Platform at the MAX IV Laboratory. *Acta Cryst. Sect D Struct. Biol.* 76, 771–777. doi:10.1107/S205979832000889X
- Long, F., Nicholls, R. A., Emsley, P., Gražulis, S., Merkys, A., Vaitkus, A., et al. (2017). AceDRG: a Stereochemical Description Generator for Ligands. *Acta Cryst. Sect D Struct. Biol.* 73, 112–122. doi:10.1107/S2059798317000067
- Maveyraud, L., and Mourey, L. (2020). Protein X-ray Crystallography and Drug Discovery. *Molecules* 25, 1030. doi:10.3390/molecules25051030
- Moreno-Chicano, T., Ebrahim, A., Axford, D., Appleby, M. V., Beale, J. H., Chaplin, A. K., et al. (2019). High-throughput Structures of Protein-Ligand Complexes at Room Temperature Using Serial Femtosecond Crystallography. *Int. Union Crystallogr. J.* 6, 1074–1085. doi:10.1107/S2052252519011655
- Mullard, A. (2021). What Does AlphaFold Mean for Drug Discovery? *Nat. Rev. Drug Discov.* 20, 725–727. doi:10.1038/d41573-021-00161-0
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., et al. (2011). REFMAC5 for the Refinement of Macromolecular crystal Structures. *Acta Crystallogr. D Biol. Cryst.* 67, 355–367. doi:10.1107/S0907444911001314
- Newman, J. A., Douangamath, A., Yadzani, S., Yosaatmadja, Y., Aimon, A., Brandão-Neto, J., et al. (2021). Structure, Mechanism and Crystallographic Fragment Screening of the SARS-CoV-2 NSP13 Helicase. *Nat. Commun.* 12, 4848. doi:10.1038/s41467-021-25166-6
- Nitsche, C., and Otting, G. (2018). NMR Studies of Ligand Binding. *Curr. Opin. Struct. Biol.* 48, 16–22. doi:10.1016/j.sbi.2017.09.001
- Owen, R. L., Juanhuix, J., and Fuchs, M. (2016). Current Advances in Synchrotron Radiation Instrumentation for MX Experiments. *Arch. Biochem. Biophys.* 602, 21–31. doi:10.1016/j.abb.2016.03.021
- PDBE-Kb consortium Varadi, M., Anyango, S., Armstrong, D., Berrisford, J., Choudhary, P., et al. (2022). PDBE-KB: Collaboratively Defining the Biological Context of Structural Data. *Nucleic Acids Res.* 50, D534–D542. doi:10.1093/nar/gkab988
- Pearce, N. M., Bradley, A. R., Krojer, T., Marsden, B. D., Deane, C. M., and von Delft, F. (2017a). Partial-occupancy Binders Identified by the Pan-Dataset Density Analysis Method Offer New Chemical Opportunities and Reveal Cryptic Binding Sites. *Struct. Dyn.* 4, 032104. doi:10.1063/1.4974176
- Pearce, N. M., Krojer, T., Bradley, A. R., Collins, P., Nowak, R. P., Talon, R., et al. (2017b). A Multi-crystal Method for Extracting Obscured Crystallographic States from Conventionally Uninterpretable Electron Density. *Nat. Commun.* 8, 15123. doi:10.1038/ncomms15123
- Pearce, N. M., Krojer, T., and von Delft, F. (2017c). Proper Modelling of Ligand Binding Requires an Ensemble of Bound and Unbound States. *Acta Cryst. Sect D Struct. Biol.* 73, 256–266. doi:10.1107/S2059798317003412
- Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., et al. (2018). CCP4i2: the New Graphical User Interface to the CCP4 Program Suite. *Acta Cryst. Sect D Struct. Biol.* 74, 68–84. doi:10.1107/S2059798317016035
- Pražnikar, J., Afonine, P. V., Gunčar, G., Adams, P. D., and Turk, D. (2009). Averaged Kick Maps: Less Noise, More Signal...and Probably Less Bias and Probably Less Bias. *Acta Crystallogr. D Biol. Cryst.* 65, 921–931. doi:10.1107/S09074449090021933
- Renaud, J.-P., Chari, A., Ciferri, C., Liu, W.-T., Rémy, H.-W., Stark, H., et al. (2018). Cryo-EM in Drug Discovery: Achievements, Limitations and Prospects. *Nat. Rev. Drug Discov.* 17, 471–492. doi:10.1038/nrd.2018.77
- Rould, M. A., and Carter, C. W. (2003). “Isomorphous Difference Methods,” in *Macromolecular Crystallography, Part D Methods in Enzymology* (Amsterdam, Netherlands: Elsevier), 145–163. doi:10.1016/S0076-6879(03)74007-5
- Schiebel, J., Krimmer, S. G., Röwer, K., Knörlein, A., Wang, X., Park, A. Y., et al. (2016). High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure* 24, 1398–1409. doi:10.1016/j.str.2016.06.010
- Schuller, M., Correy, G. J., Gahbauer, S., Fearon, D., Wu, T., Díaz, R. E., et al. (2021). Fragment Binding to the Nsp3 Macrodome of SARS-CoV-2 Identified through Crystallographic Screening and Computational Docking. *Sci. Adv.* 7. doi:10.1126/sciadv.abf8711
- Schulz, E. C., Yorke, B. A., Pearson, A. R., and Mehrabi, P. (2022). Best Practices for Time-Resolved Serial Synchrotron Crystallography. *Acta Cryst. Sect D Struct. Biol.* 78, 14–29. doi:10.1107/S2059798321011621
- Sharff, A., Keller, P., Vonrhein, C., Smart, O., Womack, T., Flensburg, C., et al. (2011). *Pipedream*. Cambridge, United Kingdom: Global Phasing Ltd.
- Sharpe, M. E., and Wojdyla, J. A. (2021). Fragment-Screening and Automation at the Swiss Light Source Macromolecular Crystallography Beamlines. *J. Crystallogr. Soc. Jpn.* 63, 232–235. doi:10.5940/jcrsj.63.232
- Sparta, K. M., Krug, M., Heinemann, U., Mueller, U., and Weiss, M. S. (2016). XDSAPP2.0. *J. Appl. Cryst.* 49, 1085–1092. doi:10.1107/S1600576716004416
- Stanfield, R., Pozharski, E., and Rupp, B. (2016a). Additional Comment on Three X-ray Crystal Structure Papers. *J.I.* 196, 528–530. doi:10.4049/jimmunol.1502281
- Stanfield, R., Pozharski, E., and Rupp, B. (2016b). Comment on Three X-ray Crystal Structure Papers. *J.I.* 196, 1–524. doi:10.4049/jimmunol.1501343
- Svensson, O., Gilski, M., Nurizzo, D., and Bowler, M. W. (2019). A Comparative Anatomy of Protein Crystals: Lessons from the Automatic Processing of 56 000 Samples. *Int. Union Crystallogr. J.* 6, 822–831. doi:10.1107/S2052252519008017
- Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Adams, P. D., Read, R. J., et al. (2008). Iterative-build OMIT Maps: Map Improvement by Iterative Model Building and Refinement without Model Bias. *Acta Crystallogr. D Biol. Cryst.* 64, 515–524. doi:10.1107/S0907444908004319
- Vonrhein, C., and Bricogne, G. (2005). Automated Structure Refinement for High-Throughput Ligand Detection with BUSTER-TNT. *Acta Cryst. Sect A.* 61, c248. doi:10.1107/S0108767305089415
- Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., et al. (2011). Data Processing and Analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Cryst.* 67, 293–302. doi:10.1107/S0907444911007773
- Winter, G. (2010). xia2: an Expert System for Macromolecular Crystallography Data Reduction. *J. Appl. Cryst.* 43, 186–190. doi:10.1107/S0021889809045701
- Wlodawer, A., Dauter, Z., Shabalin, I. G., Gilski, M., Brzezinski, D., Kowiel, M., et al. (2020). Ligand-centered Assessment of SARS-CoV-2 Drug Target Models in the Protein Data Bank. *FEBS J.* 287, 3703–3718. doi:10.1111/febs.15366
- Wlodawer, A. (2017). Stereochemistry and Validation of Macromolecular Structures. *Methods Mol. Biol.* 1607, 595–610. doi:10.1007/978-1-4939-7000-1\_24

- Wojdyr, M., Keegan, R., Winter, G., and Ashton, A. (2013). DIMPLe- a Pipeline for the Rapid Generation of Difference Maps from Protein Crystals with Putatively Bound Ligands. *Acta Cryst. Sect A*. 69, s299. doi:10.1107/S0108767313097419
- Wollenhaupt, J., Barthel, T., Lima, G. M. A., Metz, A., Wallacher, D., Jagudin, E., et al. (2021). Workflow and Tools for Crystallographic Fragment Screening at the Helmholtz-Zentrum Berlin. *JoVE* 3. doi:10.3791/62208
- Zheng, H., Hou, J., Zimmerman, M. D., Wlodawer, A., and Minor, W. (2014). The Future of Crystallography in Drug Discovery. *Expert Opin. Drug Discov.* 9, 125–137. doi:10.1517/17460441.2014.872623

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

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

Copyright © 2022 Pearce, Skyner and Krojer. 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.





# High-Throughput Native Mass Spectrometry Screening in Drug Discovery

Agni F. M. Gavriilidou, Kleitos Sokratous, Hsin-Yung Yen and Luigi De Colibus\*

OMass Therapeutics, Oxford, United Kingdom

## OPEN ACCESS

### Edited by:

Pietro Roversi,  
Italian National Research Council, Italy

### Reviewed by:

Daniel Deredge,  
University of Maryland, United States  
Keith Robert Willison,  
Imperial College London,  
United Kingdom

### \*Correspondence:

Luigi De Colibus  
luigi.decolibus@omass.com

### Specialty section:

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

**Received:** 17 December 2021

**Accepted:** 15 February 2022

**Published:** 14 April 2022

### Citation:

Gavriilidou AFM, Sokratous K,  
Yen H-Y and De Colibus L (2022) High-  
Throughput Native Mass  
Spectrometry Screening in  
Drug Discovery.  
Front. Mol. Biosci. 9:837901.  
doi: 10.3389/fmolb.2022.837901

The design of new therapeutic molecules can be significantly informed by studying protein-ligand interactions using biophysical approaches directly after purification of the protein-ligand complex. Well-established techniques utilized in drug discovery include isothermal titration calorimetry, surface plasmon resonance, nuclear magnetic resonance spectroscopy, and structure-based drug discovery which mainly rely on protein crystallography and, more recently, cryo-electron microscopy. Protein-ligand complexes are dynamic, heterogeneous, and challenging systems that are best studied with several complementary techniques. Native mass spectrometry (MS) is a versatile method used to study proteins and their non-covalently driven assemblies in a native-like folded state, providing information on binding thermodynamics and stoichiometry as well as insights on ternary and quaternary protein structure. Here, we discuss the basic principles of native mass spectrometry, the field's recent progress, how native MS is integrated into a drug discovery pipeline, and its future developments in drug discovery.

**Keywords:** Native Mass Spectrometry, Structural Biology, Drug Discovery, Screening, Affinity

## INTRODUCTION

The investigation of non-covalent interactions between a biological macromolecule and small molecule, which are driven by a myriad of forces including hydrogen bonds, Van der Waals forces, electrostatic and hydrophobic interactions, play a crucial role in the development of drug candidates. A variety of analytical methods are utilized to identify and quantify protein-ligand interactions, including isothermal titration calorimetry (ITC) (de Azevedo and Dias, 2008), surface plasmon resonance (SPR) spectroscopy (Cooper, 2002), nuclear magnetic resonance (NMR) spectroscopy (Meyer and Peters, 2003), frontal affinity chromatography combined with mass spectrometry (Slonuskiewicz et al., 2005), ThermoFluor and enzyme-linked immunosorbent assays (Benesch and Ruotolo, 2011; Aebersold and Mann, 2016).

The mass spectrometry (MS) analytical toolbox contains numerous biophysical techniques (Benesch and Ruotolo, 2011), but only a few are used in a high-throughput manner for drug discovery. One example is the analysis of biomolecules and their assemblies using MS-based proteomics approaches, which can provide means for proteome-wide quantitation of proteins, monitor their levels, and characterize protein-protein interactions and post-translational modifications (Aebersold and Mann, 2016). Another technique that has become a valuable complement to X-ray crystallography in determining protein structure, dynamics and identification of small molecule binding sites is hydrogen/deuterium exchange (HDX) coupled with mass spectrometry (Smith et al., 1997; Chalmers et al., 2011; Konermann et al., 2011; Engen and

Wales, 2015; Masson et al., 2019). Both techniques are widely used for the study of proteins and biomolecules in final denaturing conditions. Native MS, the core focus of this review, represents an addition to the analytical toolbox of mass spectrometry and has, over the past decade, experienced immense growth. It is used for studying intact proteins and their complexes, including interactions with small-molecule drugs, in a native-like folded state. Unlike the analytical tools outlined above, native MS enables the investigation of non-covalent interactions, without the need for labelling or crosslinking, using just picomoles of material (Laganowsky et al., 2013), while at the same time offering high-resolution (Rose et al., 2012; Gault et al., 2016) and a speed of analysis in the timescale of milliseconds (Breuker and McLafferty, 2008).

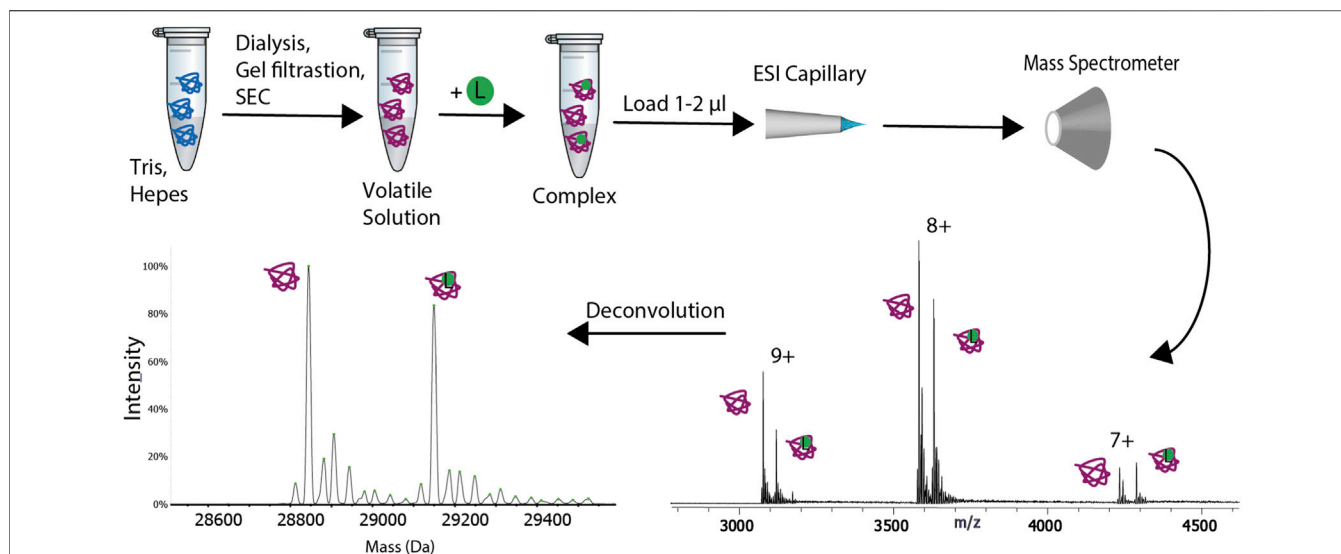
Native MS exploits the gentle nature of electrospray ionization (ESI) (Fenn et al., 1989) to transfer proteins and their non-covalent complexes from volatile buffered aqueous solutions into the gas-phase. Under controlled operating conditions within the mass spectrometer the biomolecules can retain a near native-like structure and quaternary non-covalent interactions can be preserved (Khristenko et al., 2019). Shortly after the development of ESI-MS in the late 1980s (Fenn et al., 1989), successful applications demonstrating its potential use as a tool for probing non-covalent protein complexes were reported. Early examples include the detection of the intact heme-myoglobin complex in 1991 (Katta and Chait, 1991) and the non-covalent complex between the cytoplasmic receptor FKBP with the immunosuppressive agents FK506 and rapamycin (Ganem et al., 1991), in the same year. The initial studies demonstrating the capability of ESI-MS to preserve protein-ligand interactions, were followed by studies utilizing ESI-MS for the quantification of protein-ligand binding affinities. In 1993, Loo and co-workers reported the enthalpy of dissociation ( $\Delta H$ ) of the ribonuclease S-protein—S-peptide complex, and the values of the dissociation constant ( $K_d$ ) of the same complex over a range of temperatures (Loo et al., 1993). Furthermore, the ability of ESI-MS to measure non-covalently bound protein assemblies, like the ribosome, a 2 megaDalton (MDa) protein-RNA complex (McKay et al., 2006), the tetradecameric GroEL (Sobott and Robinson, 2006) and the 20S proteasome, a 690 kiloDalton (kDa) 28-mer (Loo et al., 2005) has also been successfully demonstrated during the years. Although these examples for the preservation of non-covalent interactions during ESI (Fenn et al., 1989) were received with great enthusiasm, they raised at the same time the question of whether or not ESI (Fenn et al., 1989) could preserve the native solution structure of biomolecules and their assemblies during transfer into the gas-phase (Wolynes, 1995). Even almost 30 years after these initial demonstrations, there is much debate on the structure of protein ions in gas-phase. Clearly, the new gaseous environment could cause dramatic structural rearrangements; however, the question is how long it takes to transform solution structure to solvent-free structure and whether or not this is within the time frame of the ESI process and MS analysis. Whilst it appears probable that surface side-chain collapse occurs within picoseconds of dehydration, some elements of gross structural rearrangement may require milliseconds or even more (Breuker and McLafferty, 2008). For example, an ion mobility

spectrometry (IMS) study showed cytochrome C gas-phase ions' unfolding in the milliseconds region (Badman et al., 2005). Another IMS study by Wyttenbach and Bowers has investigated the structural stability of ubiquitin during the transition from solution to gas-phase. The authors concluded that during the ESI process, the native state of ubiquitin is preserved. Moreover, ubiquitin can survive for more than 100 milliseconds in a 294 K solvent-free environment (Wyttenbach and Bowers, 2011). Therefore, there may be a window for observing gas-phase ion species, which are relatively similar to their solution structures. Several other studies have also suggested that gas-phase ions generated during the ESI process can retain significant aspects of their solution structures (Ruotolo and Robinson, 2006; Koeniger and Clemmer, 2007; Bernstein et al., 2009; Grabenauer et al., 2010; Breuker et al., 2011).

Over the last 30 years, numerous examples have proven the analytical advantages of native MS (Fenn et al., 1989) in addressing biological questions such as stoichiometry determination (Rostom and Robinson, 1999; Rostom et al., 2000; Geels et al., 2006), oligomeric state formation (Hernández et al., 2006; Zhou et al., 2008), allostery (Dyachenko et al., 2013; Gavriilidou et al., 2018a), application to amyloids (Santambrogio et al., 2011; Bleiholder et al., 2013), antibodies (Thompson et al., 2014; Terral et al., 2016) and membrane proteins (Gupta et al., 2018; Robinson, 2019; Keener et al., 2021). Here, we will discuss the current application of native MS in ligand/drug screening, its latest advances, especially in the field of membrane proteins, which represent the majority of current pharmacological targets and the future option to integrate native MS with other structural biology techniques to shed light on the mechanism of action of a drug candidate *in situ*.

## INSTRUMENTATION OF NATIVE MASS SPECTROMETRY

ESI (Fenn et al., 1989) is the consequence of the application of electricity to an ion-containing liquid that undergoes a series of subsequent evaporation and droplet shrinking events, which in turn lead to its dispersion into a fine jet. The fundamental principles of ESI have been extensively studied to date (Kearle and Verkerk, 2009; Konermann et al., 2013). Briefly, the three prevailing models are the charged-residue model (CRM) (Kearle and Peschke, 2000), the ion evaporation model (IEM) (Iribarne and Thomson, 1976; Thomson and Iribarne, 1979) and the chain ejection model (CEM) (Konermann et al., 2013). Evidence from ESI-MS results obtained on a variety of globular proteins studied under native conditions support that the CRM mechanism is followed in the case of native MS and it is currently the most widely accepted framework for the modelling of ESI (Fernandez De la Mora, 2000; Heck and van den Heuvel, 2004; Nesatyy and Suter, 2004). The CEM applies to unfolded proteins that are particularly hydrophobic and also capable of accommodating excess charges, whereas the IEM mechanism has been suggested to be followed in the case of small molecules such as peptide ions.



**FIGURE 1** | Schematic representation of the experimental procedure for native mass spectrometry. The sample is transferred from its buffer to a volatile solution. Buffer exchange can be performed *via* one of the following methods: gel filtration, size exclusion chromatography (SEC) or dialysis. The sample is then loaded on a ESI or nanoESI capillary and it is transferred to the gas-phase. The ions are subsequently analyzed with the mass spectrometer. A ligand (L), represented with a green circle can be added to the buffer exchanged sample. The deconvolution of the spectra will result in two peaks, one for the apo protein and the other for the complex.

The ionization of proteins and protein assemblies by ESI (Fenn et al., 1989) generates multiple charge states, which are visualized in a mass spectrum by a series of peaks, each of a specific mass-to-charge ratio ( $m/z$ ). The collection of charge states that represent a single protein moiety, is referred as the charge state distribution (CSD) of gas-phase protein ions, and usually resembles one or more Gaussian distributions (Borysik et al., 2004) (**Figure 1**). The extent of CSD observed in an ESI spectrum depends on the solvent-exposed surface area of the protein, with more compact structures, such as those that can be obtained in a native MS spectrum, acquiring fewer charges (Konermann and Douglas, 1998; Fenselau et al., 2000). Multiple Gaussian distributions for a single species may indicate the presence of more than one solution conformation (Chowdhury et al., 1990; Dobo and Kaltashov, 2001). Overall, solution conditions and experimental parameters have to be tuned carefully, and it has been shown that they affect how molecules are transferred to the gas-phase (Benkestock et al., 2004; Peschke et al., 2004). For example, protein unfolding during the ESI (Fenn et al., 1989) process has been related to increased charge states (Shelimov et al., 1997). The unfolding at higher charge states is attributed to increased intra- and intermolecular Coulomb repulsion within a protein ion. It can be assumed that lower charge states minimize structure unfolding, leading to the formation of more native-like species. Moreover, solution additives or impurities could lead to the formation of adducts with alkali metal cations, such as sodium and potassium, which can destabilize a protein-ligand system during the ESI process by lowering the activation barrier to dissociation (Hopper and Oldham, 2011). However, other studies have suggested that alkali metals can actually stabilize gas-phase protein structures *via* the formation of additional interactions (Wu et al., 1999; Rožman and Gaskell, 2010).

A miniaturized version of ESI (Fenn et al., 1989), known as nano-ESI, was introduced in 1994 by Wilm and Mann (Wilm and Mann, 1994). Its advantages over conventional ESI are higher sensitivity, better resolution, increased tolerance to non-volatile salts (Lössl et al., 2014; Susa et al., 2017) and low sample consumption (Juraschek et al., 1999; Kebarle and Verkerk, 2009). Moreover, nano-ESI (Wilm and Mann, 1994) is also more sensitive and tolerant to buffer contaminants than conventional ESI (Wilm and Mann, 1996), and since the droplets formed by nano-ESI are smaller, low source/desolvation temperatures can be used, enabling better preservation of non-covalent interactions, which may be destabilized as a result of heating (Hernández and Robinson, 2007). The most commonly used volatile buffer is ammonium acetate which has a pH range of 6–8 and evaporates readily during ionization (Hernández and Robinson, 2007; Gavrilidou et al., 2015).

Nano-ESI has also been shown to be applicable to the analysis of membrane proteins. However, when working with membrane proteins, the sample must be supplemented with detergent or transferred to the gas-phase *via* lipid-based vehicles such as nanodiscs, amphipols or bicelles (Hopper et al., 2013; Laganowsky et al., 2013). For a more in depth focus on new analytical advances and application of native MS on membrane proteins we direct the readers to a recent review by Keener et al. (2021).

Sample ionization followed by separation of ions according to their  $m/z$  ratio and finally, ion detection are the three key steps that are followed in every MS experiment.

Several types of MS instruments are currently available, offering analytical sensitivity, specificity and speed in the analysis of mainly small molecules, such as drugs or peptides,

as well as analysis of large therapeutic molecules (Rathore et al., 2018) or even larger macromolecules, such as viral capsids, with molecular weights in the MDa mass range (Snijder et al., 2013).

As mentioned above, native MS analysis results in a relatively narrow charge state distribution of gas-phase ions with fewer charges than the same system would attain under denaturing conditions (Konermann and Douglas, 1998). Hence, native MS analysis of proteins and protein assemblies requires mass analyzers that can operate at a higher  $m/z$  range and at the same time be optimized for the efficient transmission of large molecular ions carrying a relatively low number of charges. The first mass analyzer adapted for native MS experiments was the time of flight (ToF) (Mirgorodskaya et al., 1994) mass analyzer, which has a theoretically unlimited  $m/z$  range. Modified ToF-based instruments dominated the field of native MS for over 2 decades. However, more recently, Fourier transform ion cyclotron resonance (FT-ICR) and Orbitrap mass analyzers, have also been adapted for native MS (Zhang et al., 2011; Rose et al., 2012).

Many modern MS-instruments used for native MS combine these mass analyzers with either a quadrupole or an ion trap mass analyzer, in a single configuration. One of the most common hybrid instruments extensively used in studies of biomolecular assemblies is the Quadrupole Time-of-flight (QToF) (Morris et al., 1996), which consists of a quadrupole filter, a collision cell and a ToF analyzer. Recently, an Orbitrap mass analyzer with ultrahigh mass range (UHMR) featured with a quadrupole analyzer has been developed for native MS (Van De Waterbeemd et al., 2017; Fort et al., 2018). An Orbitrap-based trihybrid instrument featuring both a quadrupole and a linear ion trap mass analyzer has also been added to MS instruments capable of native MS studies (Gault et al., 2020).

## DETERMINATION OF PROTEIN-DRUG INTERACTIONS

Native MS is used to study a wide diversity of biological samples that differ in mass, polydispersity, symmetry and dynamic flexibility (Kaur et al., 2019). This brings tremendous analytical advantages in interrogating protein-drug interactions. Different oligomeric states can be investigated simultaneously with no need for labelling or crosslinking. Specific information is obtained for each individual species present, without data being averaged over different species. Native MS can distinguish by mass and thus reveal the entire distribution of ligand-bound states. Therefore, the dynamics of quaternary structure can be studied in real time (Yee et al., 2019). Gavriilidou et al. (2018a) demonstrated this advantage of native MS in the study of the dimer–tetramer equilibrium of M2 pyruvate kinase (PKM2), a regulatory enzyme that is often inactive in the glycolytic pathway in tumor cells. An allosteric activator, fructose-1,6-bisphosphate (FBP), was found to shift the dimer–tetramer equilibrium toward the active tetramer, with the 4:4 stoichiometry of FBP binding to the tetramer only. Other than revealing multimeric concomitant binding, native MS helps distinguish allosteric mechanisms as has been shown by the

Sharon group (Dyachenko et al., 2013) for the ligation pathway of ATP to GroEL. Moreover, their approach was able to discriminate between the Monod–Wyman–Changeux (Jacques et al., 1965) and the Koshland–Némethy–Filmer (KNF) allosteric models (Koshland et al., 1965). Native MS showed its strength as well in analyzing complexes of proteins with covalently bound molecules (Lu et al., 2021). For example, Douangamath et al. conducted a combined mass spectrometry and high throughput crystallographic fragment screen against SARS-CoV-2 main protease ( $M^{\text{pro}}$ ), using over 1,250 fragments from a compound library that yielded 48 high-value covalent fragments (Douangamath et al., 2020).

The typical native MS analysis for drug-binding requires only 1–2  $\mu\text{l}$  of protein of interest at low-micromolar concentration. Given the broad dynamic range of signal detection and uniform ionization efficiencies between protein and protein-drug complexes (Peschke et al., 2004; Mehmood et al., 2015), drug-binding affinity can be unbiasedly determined. By looking directly at the spectra (Figure 2), the affinity of the protein–ligand interaction can be quickly evaluated without further analysis. The intensity of the peak corresponding to the protein–drug complex will increase with increasing protein–drug affinity.

Quantitative measurement of affinities requires the measurement of the drug  $K_d$ , which can be performed by the titration approach, studied by Daniel et al. (Daniel et al., 2002), and fitting the data to the equation:

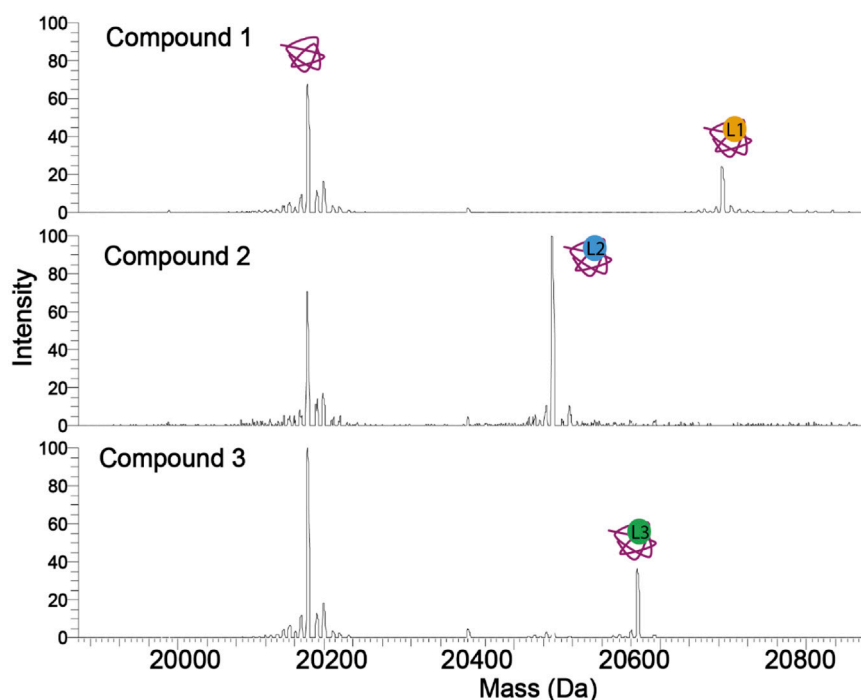
$$\frac{I(\text{PL})}{I(\text{P})} = \frac{1}{2} \left( -1 - \frac{[\text{P}]_0}{K_d} + \frac{[\text{L}]_0}{K_d} + \sqrt{4 \frac{[\text{L}]_0}{K_d} + \left( \frac{[\text{L}]_0}{K_d} - \frac{[\text{P}]_0}{K_d} - 1 \right)^2} \right)$$

A different approach was applied by Liu et al. to quantify the interactions between bovine b-lactoglobulin and a series of fatty acids by direct ESI-MS assay (Bagal et al., 2009; Liu et al., 2011). The equation of this assay is:

$$K_a = \frac{1}{K_d} = \frac{R}{[\text{L}]_0 - \frac{R}{1+R}[\text{P}]_0}, \quad R = \frac{I(\text{PL})}{I(\text{P})}$$

where  $I(\text{PL})$  is the complex and  $I(\text{P})$  the apoprotein peak intensity, respectively,  $R$  is their ratio,  $[\text{P}]_0$  and  $[\text{L}]_0$  are the concentrations of the protein and the ligand, and  $K_a$  is the association constant. Both approaches provide consistent results (Figure 3). However, the latter approach requires only a single concentration point for  $K_d$  measurement. Therefore, due to its speed and low material consumption, native MS has the potential to be the primary choice to acquire the affinity information of drug-binding in high throughput screening as it has been shown in various studies discussed in *High-Definition Screening by Native MS* (Maple et al., 2012; Woods et al., 2016; Nguyen et al., 2021). This is a unique feature of native MS compared to other biophysical methods. The dissociation constants of thousands of compounds against a protein target can be reliably ranked using this assay, which will mandate the





**FIGURE 2** | Spectra of a protein in complex with three compounds (ligands). The tighter binding compound yields a more intense complex peak. Therefore, without calculating the  $K_d$ , the affinity of a compound against a protein can be estimated. Compound 2 has the highest occupancy to the protein, followed by compounds 3 and 1.

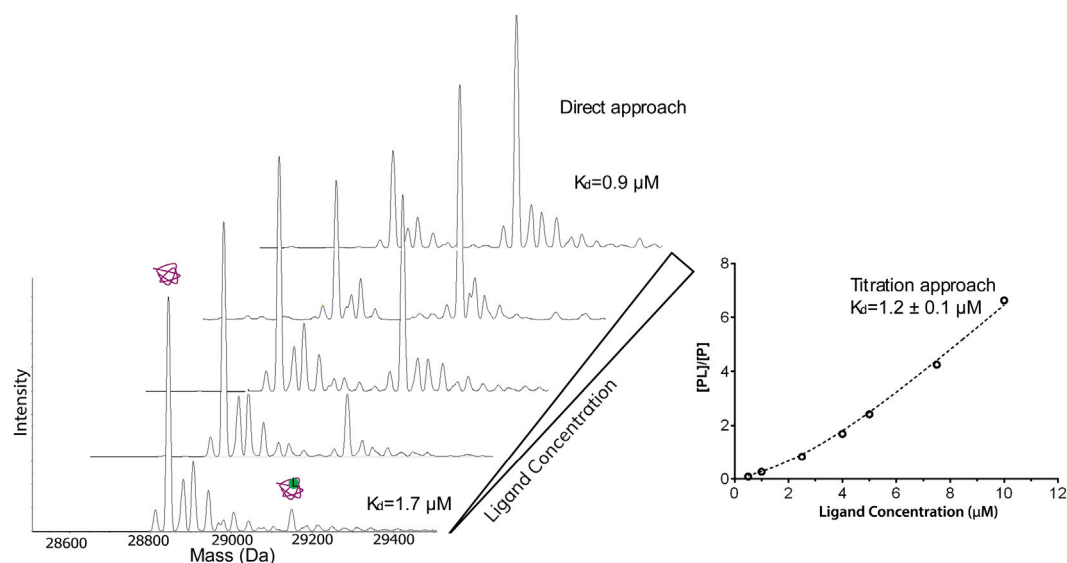
selection of those that will move to the next step of the screening process.

The ion intensity ratio between the free protein and the protein–ligand complex is measured for affinity measurements, and ideally, the target protein concentration should be below the expected  $K_d$  value. If the protein concentration is very low, such as that it appears in the noise level of the spectrum, the measurement of the intensity of the peaks may be unreliable. Therefore, using a protein concentration that will yield a high-intensity protein peak is crucial during a screening experiment, and that will allow ranking the affinities of the compounds against that protein reliably. Moreover, native MS experiments can be configured in a competitive format. Competition among different binding partners of a target molecule for its binding site can provide information about the binding affinities and specificity of host–guest complexes binding. For example, if a known ligand (hot ligand, HL) with a known binding site is available, this can be used to quickly assess the specificity for that site for a different ligand (L). Native MS competitive binding experiments consist in keeping the protein and L concentrations constant, while increasing the HL concentrations. The relative abundances of the different species in the mass spectra allow assessing whether the displacement of L with HL is competitive or not. If the HL is competing with the L for the same site, the fraction of the complex with the L decreases and the fraction of the complex with the HL increases as the HL concentration is increased as shown in **Figure 4A**. If the HL is not binding in the same site as the L, as the HL concentration increases and the protein becomes

saturated, the peak of the protein with the L will decrease, and an additional HL + L peak will appear (**Figure 4B**). Jørgensen et al. (1998) used the competition approach to study the specificity of the interactions between the glycopeptide antibiotics (vancomycin, ristocetin) and several peptide and also measure their affinities. Wortmann et al. (2008), in a study of kinase inhibitors, followed the disappearance of the peak of a test ligand relative to a reference ligand in the low  $m/z$  range of the mass spectrometer, allowing them to determine high-affinity binding constants in the picomolar range.

Additional to affinity measurements, biomolecules' thermodynamic and kinetic properties can be investigated by following the abundance of different species as a function of temperature and over a time course (Gülbakan et al., 2015). Thermodynamics are measured using custom variable-temperature ESI sources that enable precise temperature control of the analyte solution prior to ion formation.

Cong et al. (2016) investigated the thermodynamics of lipid binding to AmtB, an integral membrane protein of *Escherichia coli*. Their approach allowed them to determine the thermodynamics of individual binding events for lipids with variable chain lengths, resolving unique thermodynamic properties. Another study by Marchand et al. (2018) determined the entropic and enthalpic contributions to the binding equilibrium of G-quadruplex nucleic acid structures and their ligands using a temperature-controlled nano-ESI (Wilm and Mann, 1994) source. The Klassen group measured biotin's dissociation rate ( $k_{off}$ ) from the tetrameric protein



**FIGURE 3 |** Native ESI-MS spectra and the fitted curve of titration of dichlorphenamide against 4  $\mu\text{M}$  carbonic anhydrase I in 50 mM ammonium acetate, 2% dimethylsulfoxide. The concentration of dichlorphenamide was varied. The measured  $K_d$  from the titration curve was 1.2  $\mu\text{M}$  which is in good agreement with the values calculated based on a single spectrum using the direct approach (1.7 and 0.9  $\mu\text{M}$ ) and published values.

streptavidin (Deng et al., 2013). Marchand et al. (2020) in 2020 developed a temperature-jump electrospray source for mass spectrometry that allows performing fast kinetics experiments (0.16–32 s) at different temperatures (10–90°C).

## Membrane Proteins

There are several analytical challenges in studying membrane proteins because of the complex interactions and environmental constraints that accompany their amphipathic nature. The natural abundance of membrane proteins is typically low, and overexpression and purification of membrane proteins in high yields can be challenging (Wagner et al., 2007; Zweers et al., 2009; Gubellini et al., 2011).

Native MS has become an emerging technique in recent years for investigating the structure, dynamics, and interactions of membrane proteins (Barrera et al., 2009; Laganowsky et al., 2013; Gupta et al., 2018). The sensitivity and resolution of native MS render it a powerful tool to investigate the membrane proteins and various aspects of their structure and function, such as macromolecular assemblies, lipid/ligand interactions, post-translational modifications and, most importantly, the interplay between them. Usually, membrane proteins are studied in their natural membrane or encapsulated in a membrane mimetic to solubilize the protein prior to analysis (Strop and Brunger, 2005). Different reconstitution systems, such as detergents (Laganowsky et al., 2013), nanodiscs (Marty et al., 2016) and amphipathic polymers or amphipols (Calabrese et al., 2015) have been tested and sometimes proved critical in preserving subunit and lipid interactions of membrane proteins (Hopper et al., 2013).

Native MS has been applied for analyzing various membrane mimetics, which cause significant effects on the quality of mass spectra. For example, detergent micelles cause a huge mass heterogeneity which hampers the accurate mass measurement of membrane proteins, and therefore detergent adducts need to be removed by activation processes in the mass spectrometer (Figure 5). The activation energy applied may cause protein unfolding and disruption of ligand interactions in the gas-phase, hence care must be devoted to selecting a detergent which is able to preserve the native state of the membrane protein in solution and optimize the quality of mass spectra. It has been shown that the chemical properties of the detergents mediate the charge state, both during ionization and detergent removal in the mass spectrometer (Reading et al., 2015). Therefore, screening of different detergents may be required to find a suitable one (Yen et al., 2017). Various studies have shown the capability of native MS in interrogating membrane protein interactions. Binding and affinity measurement of small-molecules to membrane proteins has been investigated for the first time in the Robinson group (Marcoux et al., 2013). In a 2013, the strength of the interaction between a 17-residue peptide and the homotrimeric *E. coli* outer membrane porin F (OmpF) was quantified via the titration approach (Housden et al., 2013) which enabled the observation of OmpF bound to up to three peptides. The resulting  $K_d$  value agreed with that derived from ITC. In the same year, measuring the rates of lipid binding and calculating the  $K_d$  values showed that the ATP-binding cassette transporter P-glycoprotein binds diacylglycerides more tightly than zwitterionic lipids (Marcoux et al., 2013).

The application of native MS has also been demonstrated for GPCRs, one of the most important protein families for drug discovery. In a study by Hsin-Yung Yen et al. (2017), the ligand/drug-binding of a human purinergic receptor P2Y<sub>1</sub>R can be preserved in a high-resolution mass spectrometer. Intriguingly, the resolution of mass spectrometry revealed the impact of receptor phosphorylation on attenuating the binding of the drug MRS2500 but not that of ATP. The importance of solution conditions in maintaining native GPCR interactions in MS has recently been highlighted by an investigation into the effects of nonvolatile salts on GPCR–small molecule interactions (Agasid et al., 2021). The inclusion of Tris and NaCl in the ESI solution was demonstrated to significantly strengthen the noncovalent interaction of the receptor to the endogenous ligand glucagon. In addition to ligand interactions of GPCRs, Gavriilidou et al. (2019) characterized turkey  $\beta$ 1 adrenergic receptor in complex with mini-Gs, an engineered G<sub>as</sub> subunit, and the impact of ligands on complex stability. A full agonist (isoprenaline) stabilizes complex formation whereas an inverse agonist (S32212) disrupts the receptor-mini-Gs complex, showing the potential of native MS in investigating the effect of small molecules on GPCR coupling. By exploring these effects with experimental GPCR-targeting drugs, it may be possible to examine bias toward particular signaling pathways, thereby facilitating the development of highly selective therapeutic agents. Native MS has shown a significant advantage in studying membrane protein-lipid interactions. The resolution of this technique enables the detailed interrogations of a wide range of membrane proteins bound to different lipids, and the impact of acyl chain lengths and degrees of saturation on lipid-binding (Gault et al., 2016). In 2014, Bechara et al. (2015) published a protocol for identifying lipids to ABC transporter TmrAB *via* a progressive delipidation approach. By controlling the extent of protein delipidation *via* timed exposure to detergent, the preferential binding of TmrAB to negatively charged phosphatidylglycerol was revealed, and the potential role of lipids in modulating glycolipid translocation of TmrAB was proposed. In a study by the Robinson group, the specific effect of phosphatidylinositol-4,5-bisphosphate PIP<sub>2</sub> on G protein-coupling of three class A GPCRs was unveiled (Yen et al., 2018). The native MS results showed that PIP<sub>2</sub> significantly increased mini-Gs' coupling to the  $\beta$ 1 adrenergic receptor, whereas other phospholipids did not possess a similar effect. Joint use of native MS, computational modelling and site-directed mutagenesis enabled elucidation of the mechanism of PIP<sub>2</sub> in bridging the receptor and the G proteins, highlighting the structural specificity of this phenomenon. A recently developed approach, nativeomics (Gault et al., 2020), aims at combining native MS with small-molecule fragmentation to directly identify bound molecules ejected after native MS. The mass spectrometer is optimized and fine-tuned in detecting both intact protein–ligand complexes at the high *m/z* range and fragmented ligands at the low *m/z* range. The strength of this approach is that it enables determination of the chemical

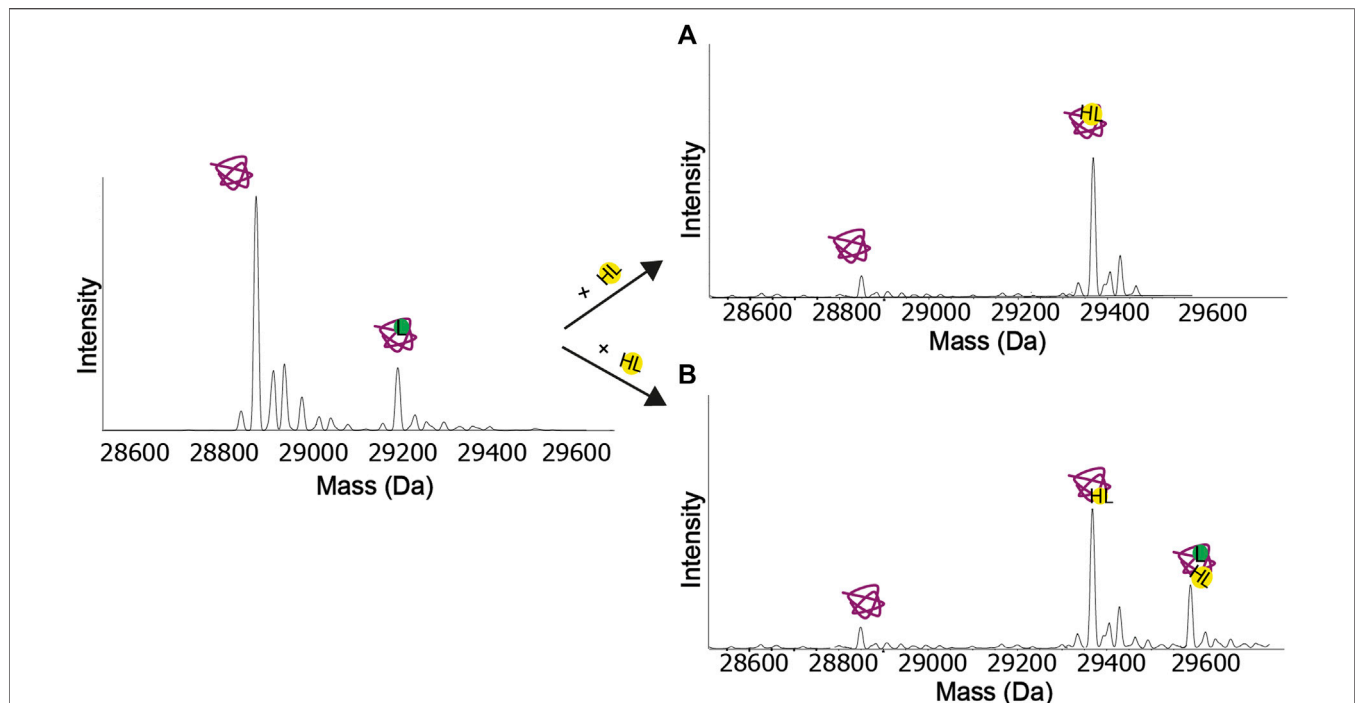
identity of endogenous ligands or lipids and drugs of unknown identity bound to membrane proteins. In order to achieve this, multiple rounds of fragmentation are applied to progressively dissociate the protein–ligand assembly and yield ligand partners for fragmentation.

A revolutionary new method that allows membrane proteins to be directly studied from native membranes was recently developed by Robinson's group and was named SoLVe-MS (sonication of lipid Vesicles for MS) (Chorev et al., 2018). To enable MS analysis, large membrane fragments isolated from cells are sonicated to produce smaller liposomes. This technique provided a breakthrough in studying endogenously expressed membrane proteins and their associations with native environment such as lipids and ligands, without purifying the proteins.

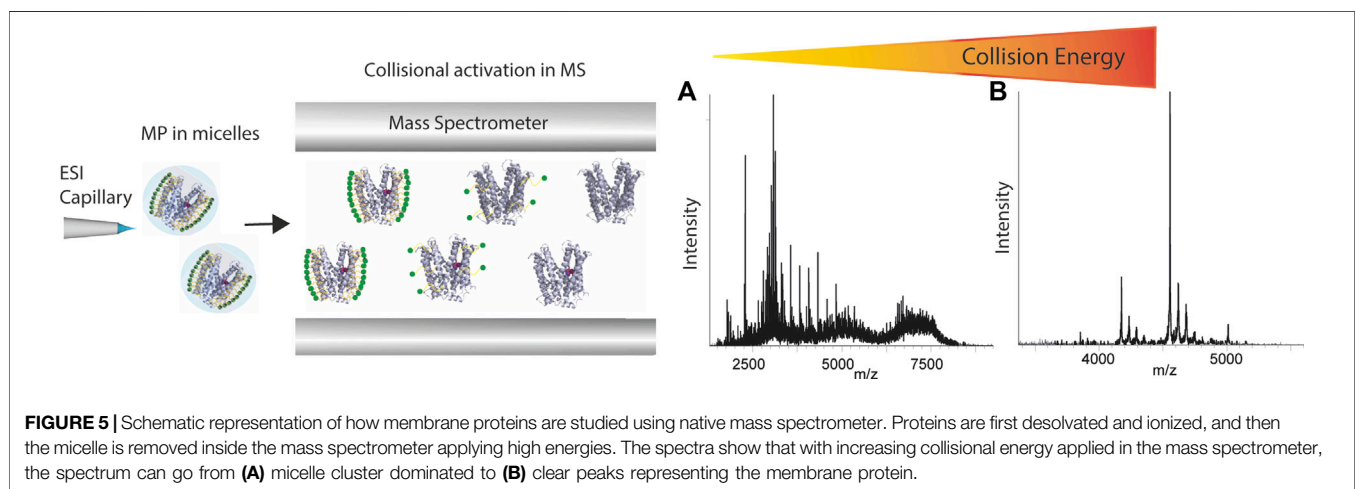
## HIGH-DEFINITION SCREENING BY NATIVE MS

The application of native MS to compound library screening has only recently emerged, as outlined in the following examples. The native MS screening approach shows a particular potential for fragment-based drug discovery (FBDD). The ability to capture the weak binding of fragments at a millimolar range of affinity and the low sample consumption widens the accessibility even for compounds with poor solubility (Vivat Hannah et al., 2010; Woods et al., 2016; Gavriilidou et al., 2018b). In order to improve the throughput, an automated electrospray platform, the NanoMate, was introduced (Zhang et al., 2003). This chip-based nano-ESI (Wilm and Mann, 1994) system provides consistent electrospray conditions across each analytical run, improving reproducibility. NanoMate automated sampling also significantly increases the analytical throughput compared to manual sample manipulation. The proof-of-principle study of Maple et al. has demonstrated the native MS screening of a fragment library consisting of 157 compounds against an apoptotic protein target within 6 h, using the NanoMate system (Maple et al., 2012). The throughput and results are comparable to those that NMR or ITC-based library screening approaches can obtain. In another study by Woods et al., a 720-member fragment library was screened, and ESI-MS affinity measurements correlated with the ones obtained from SPR when followed-up by X-ray crystallography (Woods et al., 2016). A recent study showed protein–small molecule interactions from mixtures containing up to ~8,900 potential small molecule ligands in a single manual measurement using nano-ESI (Wilm and Mann, 1994) emitters in combination with a rapid, low-volume gel filtration step to remove unbound molecules (Nguyen et al., 2021).

Screening with native MS increases the throughput by allowing drug multiplexing per well. The sample consumption is very low compared to other methods and has a wide dynamic range. The number of compounds per well is mandated by the resolution needed to separate protein complexes with compounds with a slight mass difference. With



**FIGURE 4 |** Competition experiments. **(A)** Both the ligand (L) and the hot ligand (HL) bind at the same binding site. When the concentration of the hot ligand increases, the protein-ligand peak eventually disappears. **(B)** The ligand and the hot ligand do not bind at the same binding site. When the concentration of the hot ligand increases and the protein gets saturated, the protein-ligand peak eventually disappears, and a new peak appears, corresponding to the protein with both ligands bound.



**FIGURE 5 |** Schematic representation of how membrane proteins are studied using native mass spectrometer. Proteins are first desolvated and ionized, and then the micelle is removed inside the mass spectrometer applying high energies. The spectra show that with increasing collisional energy applied in the mass spectrometer, the spectrum can go from **(A)** micelle cluster dominated to **(B)** clear peaks representing the membrane protein.

the current technology and using multiplexed libraries, it is possible to screen over 50 k compounds in less than 8 h. At the end of the screening, a list of compounds and their affinities to the protein is generated. Based on the  $K_d$  values, compounds with the desirable affinity, for example,  $K_d < 50 \mu\text{M}$ , are chosen for further validation. The validation will include the specificity of binding to the protein-target by screening these compounds to a different protein or in a competition format by using a hot ligand. Moreover, the  $K_d$  values measured with the direct approach are confirmed with the titration method (Figure 6).

Native MS is a very advantageous method during screening due to its supreme speed, selectivity, sensitivity, and quantitative capability.

Furthermore, mass spectrometry combined with ion mobility can reveal conformational changes of proteins upon binding to small molecules. This is a unique advantage over the other biophysical methods, as in a single ion mobility MS experiment, the binding affinity and the conformational impact of a compound to a protein can be measured. Ion mobility spectrometry (IMS) (Zhong et al., 2012) is the analytical technique that separates gas-phase ions based on



their different mobility through a buffer gas at an applied electric field. The mobility of each ion in the gas will define its travel or drift time (**Figure 7**). The measured drift time is proportional to the collisional cross-section (CCS) of the ion, which is a physical property of the molecule and depends on its shape. Once the CCS is determined, it can be related to its quaternary structure (Ruotolo et al., 2008; Bush et al., 2010; Jurneczko and Barran, 2011; Lanucara et al., 2014). In contrast to MS, the separation of gas-phase ions in IMS is based on charge and shape rather than their  $m/z$  ratio.

The combination of IMS with MS, commonly referred to as IM-MS, affords 2D separation of analytes on the size-to-charge and mass-to-charge axes, respectively (**Figure 7**). Several studies on highly complex mixtures have shown that IM-MS gives far greater resolution and information than is possible by either method alone (Ruotolo et al., 2007; Duijn et al., 2009; Hopper and Oldham, 2009; Smith et al., 2010; Bleiholder et al., 2011; Bleiholder et al., 2013). Studies to investigate changes in protein structure that are driven by ligand binding have been carried out. For example, Lai et al. used native IM-MS to characterize large-scale conformational shifts of the *Escherichia coli* molecular chaperone DnaK in response to nucleotide and substrate binding. Unique conformational states that arose due to the allosteric effects of small molecule interactions were identified and combined with results from double electron-electron resonance spectroscopy to confirm structural data derived from both NMR and X-crystallography experiments (Lai et al., 2017). Ashcroft's group screened small molecule inhibitors against amyloid precursors, identifying the interacting protein species and defining the mode of inhibition. They were able to classify a variety of small molecules that are potential inhibitors of human islet amyloid polypeptide (hIAPP) aggregation or amyloid-beta 1–40 aggregation as specific, nonspecific, colloidal or non-interacting (Young et al., 2014). Small conformational changes in globular proteins that occur upon ligand binding can also be observed through slight changes in the arrival time distribution (Atmanene et al., 2012; Stojko et al., 2015). The Klassen group has combined multistage ion activation and IM to determine the identity of bound ligands during a screening, an approach they have termed "Catch-and-Release" (CaR) ESI-MS (El-Hawiet et al., 2012; Rezaei Darestani et al., 2016; El-Hawiet et al., 2018). In this approach, a protein of interest is mixed with a library of ligands and the resulting mixture is analyzed by native MS. Multiple stages of ion activation and fragmentation are applied, in which the ligands are fragmented. By combining this process with IM, both the arrival time distributions and the fragmentations patterns of the ligands can be used to ascertain their identity.

Collision induced unfolding (CIU) is a collisional activation method in which the unfolding of protein complexes in the gas-phase is monitored with IM-MS. During IM-MS screening, CIU fingerprints can be uniquely related to specific protein-ligand binding modes (Hopper and Oldham, 2009; Hyung et al., 2009; Niu et al., 2013). The

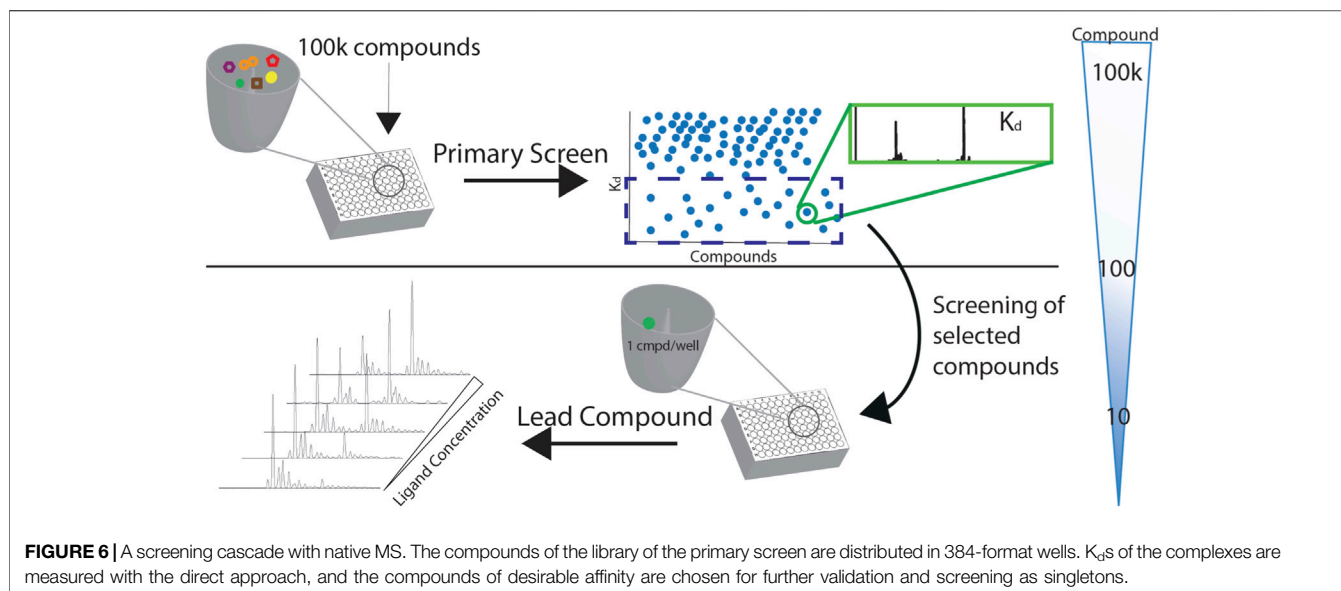
unfolding pathway of the protein can be followed in detail to allow the comparison between different conformational families of the protein (Han et al., 2011; Han et al., 2012; Han and Ruotolo, 2013). Ruotolo's group developed a data analysis workflow to remove chemical noise patterns caused by ionized surfactants during studies of membrane proteins (Fantin et al., 2019). Following the denoising protocol, separate gas-phase unfolding signatures with CIU for lipid and protoporphyrin binding to the dimer of translocator protein (TSPO) were generated. Complexes containing ligands known to bind at two separate sites were detected as possessing differential stabilities using CIU, where protoporphyrin IX binding provided a greater degree of gas-phase stabilization for TSPO than any lipids assessed. These data were combined with liquid-liquid extracts to propose and identify unknown endogenous TSPO ligands.

It is expected that as the resolution of IM-MS methods in both the mobility and  $m/z$  dimensions and the accuracy of computational models of protein structure and dynamics increase the use of IM-MS for studying the impact of ligand binding on protein structures and assemblies will also significantly increase.

## FUTURE PERSPECTIVE OF NATIVE MS IN STRUCTURE-BASED DRUG LEADS DISCOVERY

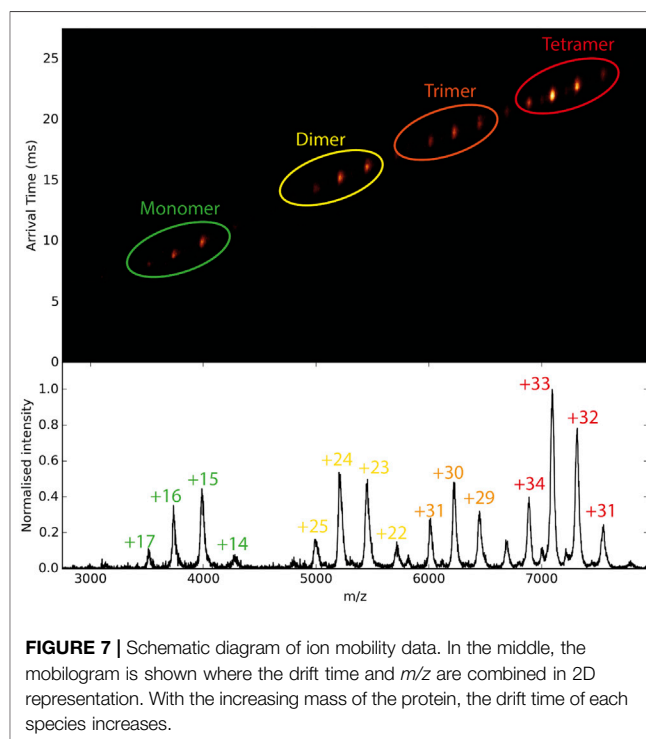
Technological advances continue to emerge, and the recent introduction of orbitrap mass analyzers modified for protein assemblies affords excellent resolution and sensitivity, enabling facile definition of concurrent binding of small molecules in a MDa complex (Rose et al., 2012; Gault et al., 2016). Taking advantage of the high resolution, native MS is pacing towards the analysis of proteins and protein complexes directly from cells, characterizing their heterogeneity and flexibility in real time (Gan et al., 2017; Chorev et al., 2018). Intact assemblies from membranes, without chemical disruption, can be analyzed using mass spectrometry to define their composition and characterize any endogenous ligand or lipid binding. This development is analogous to the Cryo-EM structure determination of MDa complexes from native cell extracts (Kyriakis et al., 2021). However, improvements in data analysis are still required for this field to bloom, as the spectra generated from these studies can be difficult and time-consuming to annotate. The improvement of data analysis software will also allow native MS to be integrated as a routine method in the pharmaceutical sector.

The current mass spectrometers allow high throughput screening of small molecules for binding against a protein target and determining the dissociation constant. As shown and discussed above, the thermodynamic and kinetic properties of biomolecules can also be measured with MS using modified instrumentation (Gülbakan et al., 2015). Therefore, the commercialization of these novel sources for mass spectrometers that will vary the temperature in a controlled way over a time course will allow for the detailed thermodynamic study of protein-ligand complex.



3D models of macromolecular complexes have been deduced by combining MS experiments with data from crystallography, NMR, small-angle X-ray scattering and EM (Lebrette et al., 2015; Nematollahi et al., 2015) and it is expected that in the future more structures will be solved with the help of mass spectrometry. Structure-based drug discovery (SBB) has also benefited from the introduction of mass spectrometry as a complementary tool accelerating the sample screening workflow (Olinares et al., 2021). Based on a combinatorial study of native MS and other techniques, Heck's group (Snijder et al., 2017) revealed details of the interactions between the Kai proteins, a system that cyanobacteria use as a circadian oscillator. The stoichiometry of the different Kai proteins was monitored by native MS, allowing for structural characterization by single-particle cryo-electron microscopy (cryo-EM) and MS. Pseudoatomic models of biomolecular assemblies have been generated with computational methods based on data from native mass spectrometry (Marklund and Benesch, 2019). Politis et al. show many examples implementing this approach (Politis et al., 2014). They describe a method for the characterization of protein assemblies structures integrating results derived from different MS-based techniques with modeling data. They encoded results from native MS, bottom-up proteomics, IM-MS and chemical cross-linking MS into modeling restraints to compute the most likely structures. Esser et al. (2021) presented native electrospray ion-beam deposition (native ES-IBD) for the preparation of extremely high-purity cryo-EM samples, based on mass selection in the gas-phase. Folded protein ions generated by native MS were mass-filtered with fine tuning of the mass spectrometer, and gently deposited on cryo-EM grids, and subsequently frozen in liquid nitrogen. Single particle analysis revealed that they remain structurally intact.

Notably, native MS allows the identification of a protein bound ligand when it is featureless in cryo-EM and X-ray maps. In fact, in the specific case of membrane proteins, known to be valuable therapeutic targets, bound detergents



or lipids cannot be identified completely by cryo-EM and X-ray crystallography unless their hydrophobic tails are constrained in specific places. Conversely, the hydrophilic heads tend to be defined more accurately due to the electrostatic interactions with the protein partner. Moreover, multiple chemical species bound to membrane proteins, such as a mixture of lipids present at partial occupancy, produce poorly defined electron or cryo-EM densities, making it hard to assign chemical entities. Native

MS can resolve lipid mixture bound to protein and provide the relative abundance of each component. Additionally, large multimeric targets could lose ancillary or weakly attached subunits when extracted and purified from their native environment to perform SBBDD studies. We envisage that in the not-too-distant future, *in situ* native MS will add another dimension to the 3D views of large macromolecular assemblies currently imaged in their native frozen-hydrated state by electron cryotomography (ECT) (Oikonomou and Jensen, 2017). The so-called mass spectrometry imaging (MSI) (Griffiths et al., 2019) will offer spatial information about drug distribution directly at the cellular level, evaluation of druggability *in situ*, and provide crucial insights about the ripple effects of a drug candidate on whole cellular physiology speeding up the development of a drug considerably.

## CONCLUSION

In the last few years, native MS has become a well-established technique for drug discovery. Its high sensitivity, simplicity, speed, wide dynamic range, low protein and ligand consumption, and the possibility of automation and high throughput makes it an integral component of the biophysical toolkit commonly used for primary screening, adding to techniques such as NMR and SPR. Moreover, the ability to explore all the aspects of protein-ligand interactions and dynamics is the basis of the unique potential of native MS in the fragment hit identification. The majority of published applications of native MS in drug discovery are on soluble proteins. Nevertheless, in recent years the frontier conquered by native MS in membrane proteins broadens the repertoire of protein targets screened by this technique.  $K_d$  determinations by direct-ESI assay or titration experiments have provided accurate values and can be used to

quickly assess the compounds' affinity to the target protein during a screening campaign. Native MS can also efficiently assess compound specificity for a particular binding site in competitive binding experiments. It is possible to distinguish multiple binding sites with the appropriate instrumental parameters, revealing complex allosteric mechanisms. In addition, IMS studies during screening have provided insight into conformational changes of a protein upon binding to a compound. The continuous improvements in mass spectrometry hardware and software are expanding the limits of native MS applications. In addition, native MS is expected to be fully integrated with other structural biology techniques, such as X-ray crystallography, cryo-EM, and in a not distant future cryo-ET, in the drug discovery pipeline, providing unprecedented insights on protein-ligand binding and ligands screening that could significantly impact the drug discovery process.

## AUTHOR CONTRIBUTIONS

AFMG drafted part of the introduction, the sections on the protein-ligand interactions and the high-definition screening, contributed to the sections on the future perspective and conclusion. AFMG generated all the figures and analysed data. KS drafted the section on the instrumentation, generated and analysed data. HYY drafted the membrane proteins section. LDC drafted the abstract and parts of the manuscript, assembled and revised it.

## ACKNOWLEDGMENTS

We would like to thank Dr Ali Jazayeri, Dr Idilr Liko, Dr Jonathan Hopper and Professor Dame Carol Robinson for the critical reading of the manuscript.

## REFERENCES

- Aebersold, R., and Mann, M. (2016). Mass-spectrometric Exploration of Proteome Structure and Function. *Nature* 537 (7620), 347–355. doi:10.1038/nature19949
- Agasid, M. T., Sørensen, L., Urner, L. H., Yan, J., and Robinson, C. V. (2021). The Effects of Sodium Ions on Ligand Binding and Conformational States of G Protein-Coupled Receptors-Insights from Mass Spectrometry. *J. Am. Chem. Soc.* 143 (11), 4085–4089. doi:10.1021/jacs.0c11837
- Atmanene, C., Petiot-Bécard, S., Zeyer, D., Van Dorsselaer, A., Vivat Hannah, V., and Sanglier-Cianféran, S. (2012). Exploring Key Parameters to Detect Subtle Ligand-Induced Protein Conformational Changes Using Traveling Wave Ion Mobility Mass Spectrometry. *Anal. Chem.* 84 (11), 4703–4710. doi:10.1021/ac203223h
- Badman, E. R., Myung, S., and Clemmer, D. E. (2005). Evidence for Unfolding and Refolding of Gas-phase Cytochrome C Ions in a Paul Trap. *J. Am. Soc. Mass Spectrom.* 16 (9), 1493–1497. doi:10.1016/j.jasms.2005.04.013
- Bagal, D., Kitova, E. N., Liu, L., El-Hawiet, A., Schnier, P. D., and Klassen, J. S. (2009). Gas Phase Stabilization of Noncovalent Protein Complexes Formed by Electrospray Ionization. *Anal. Chem.* 81 (18), 7801–7806. doi:10.1021/ac900611a
- Barrera, N. P., Isaacson, S. C., Zhou, M., Bavro, V. N., Welch, A., Schaedler, T. A., et al. (2009). Mass Spectrometry of Membrane Transporters Reveals Subunit Stoichiometry and Interactions. *Nat. Methods* 6 (8), 585–587. doi:10.1038/nmeth.1347
- Bechara, C., Nöll, A., Morgner, N., Degiacomi, M. T., Tampé, R., and Robinson, C. V. (2015). A Subset of Annular Lipids Is Linked to the Flippase Activity of an ABC Transporter. *Nat. Chem* 7 (3), 255–262. doi:10.1038/nchem.2172
- Benesch, J. L., and Ruotolo, B. T. (2011). Mass Spectrometry: Come of Age for Structural and Dynamical Biology. *Curr. Opin. Struct. Biol.* 21 (5), 641–649. doi:10.1016/j.sbi.2011.08.002
- Benkestock, K., Sundqvist, G., Edlund, P.-O., and Roeraade, J. (2004). Influence of Droplet Size, Capillary-Cone Distance and Selected Instrumental Parameters for the Analysis of Noncovalent Protein-Ligand Complexes by Nano-Electrospray Ionization Mass Spectrometry. *J. Mass Spectrom.* 39 (9), 1059–1067. doi:10.1002/jms.685
- Bernstein, S. L., Dupuis, N. F., Lazo, N. D., Wyttenbach, T., Condron, M. M., Bitan, G., et al. (2009). Amyloid- $\beta$  Protein Oligomerization and the Importance of Tetramers and Dodecamers in the Aetiology of Alzheimer's Disease. *Nat. Chem* 1 (4), 326–331. doi:10.1038/nchem.247
- Bleiholder, C., Do, T. D., Wu, C., Economou, N. J., Bernstein, S. S., Buratto, S. K., et al. (2013). Ion Mobility Spectrometry Reveals the Mechanism of Amyloid Formation of A $\beta$ (25–35) and its Modulation by Inhibitors at the Molecular Level: Epigallocatechin Gallate and Scyllo-Inositol. *J. Am. Chem. Soc.* 135 (45), 16926–16937. doi:10.1021/ja406197f
- Bleiholder, C., Dupuis, N. F., Wyttenbach, T., and Bowers, M. T. (2011). Ion Mobility-Mass Spectrometry Reveals a Conformational Conversion from Random Assembly to  $\beta$ -sheet in Amyloid Fibril Formation. *Nat. Chem* 3 (2), 172–177. doi:10.1038/nchem.945

- Borysik, A. J. H., Radford, S. E., and Ashcroft, A. E. (2004). Co-populated Conformational Ensembles of  $\beta$ 2-Microglobulin Uncovered Quantitatively by Electrospray Ionization Mass Spectrometry. *J. Biol. Chem.* 279 (26), 27069–27077. doi:10.1074/jbc.M401472200
- Breuker, K., Brüscheiler, S., and Tollinger, M. (2011). Electrostatic Stabilization of a Native Protein Structure in the Gas Phase. *Angew. Chem. Int. Ed.* 50 (4), 873–877. doi:10.1002/anie.201005112
- Breuker, K., and McLafferty, F. W. (2008). Stepwise Evolution of Protein Native Structure with Electrospray into the Gas Phase, 10–12 to 102 S. *Proc. Natl. Acad. Sci.* 105 (47), 18145–18152. doi:10.1073/pnas.0807005105
- Bush, M. F., Hall, Z., Giles, K., Hoyes, J., Robinson, C. V., and Ruotolo, B. T. (2010). Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-phase Structural Biology. *Anal. Chem.* 82 (22), 9557–9565. doi:10.1021/ac1022953
- Calabrese, A. N., Watkinson, T. G., Henderson, P. J. F., Radford, S. E., and Ashcroft, A. E. (2015). Amphipols Outperform Dodecylmaltoside Micelles in Stabilizing Membrane Protein Structure in the Gas Phase. *Anal. Chem.* 87 (2), 1118–1126. doi:10.1021/ac5037022
- Chalmers, M. J., Busby, S. A., Pascal, B. D., West, G. M., and Griffin, P. R. (2011). Differential Hydrogen/deuterium Exchange Mass Spectrometry Analysis of Protein-Ligand Interactions. *Expert Rev. Proteomics* 8 (1), 43–59. doi:10.1586/ep.10.109
- Chorev, D. S., Baker, L. A., Wu, D., Beilstein-Edmands, V., Rouse, S. L., Zeev-Ben-Mordehai, T., et al. (2018). Protein Assemblies Ejected Directly from Native Membranes Yield Complexes for Mass Spectrometry. *Science* 362 (6416), 829–834. doi:10.1126/science.aau0976
- Chowdhury, S. K., Katta, V., and Chait, B. T. (1990). Probing Conformational Changes in Proteins by Mass Spectrometry. *J. Am. Chem. Soc.* 112 (24), 9012–9013. doi:10.1021/ja00180a074
- Cong, X., Liu, Y., Liu, W., Liang, X., Russell, D. H., and Laganowsky, A. (2016). Determining Membrane Protein-Lipid Binding Thermodynamics Using Native Mass Spectrometry. *J. Am. Chem. Soc.* 138 (13), 4346–4349. doi:10.1021/jacs.6b01771
- Cooper, M. A. (2002). Optical Biosensors in Drug Discovery. *Nat. Rev. Drug Discov.* 1 (7), 515–528. doi:10.1038/nrd838
- Daniel, J. M., Friess, S. D., Rajagopalan, S., Wendt, S., and Zenobi, R. (2002). Quantitative Determination of Noncovalent Binding Interactions Using Soft Ionization Mass Spectrometry. *Int. J. Mass Spectrom.* 216 (1), 1–27. doi:10.1016/s1387-3806(02)00585-7
- de Azevedo, W. F., and Dias, R. (2008). Experimental Approaches to Evaluate the Thermodynamics of Protein-Drug Interactions. *Curr. Drug Targets* 9 (12), 1071–1076. doi:10.2174/138945008786949441
- Deng, L., Kitova, E. N., and Klassen, J. S. (2013). Dissociation Kinetics of the Streptavidin-Biotin Interaction Measured Using Direct Electrospray Ionization Mass Spectrometry Analysis. *J. Am. Soc. Mass Spectrom.* 24 (1), 49–56. doi:10.1007/s13361-012-0533-5
- Dobo, A., and Kaltashov, I. A. (2001). Detection of Multiple Protein Conformational Ensembles in Solution via Deconvolution of Charge-State Distributions in ESI MS. *Anal. Chem.* 73 (20), 4763–4773. doi:10.1021/ac010713f
- Douangamath, A., Fearon, D., Gehrtz, P., Krojer, T., Lukacik, P., Owen, C. D., et al. (2020). Crystallographic and Electrophilic Fragment Screening of the SARS-CoV-2 Main Protease. *Nat. Commun.* 11, 5047. doi:10.1038/s41467-020-18709-w
- Duijn, E. v., Barendregt, A., Synowsky, S., Versluis, C., and Heck, A. J. R. (2009). Chaperonin Complexes Monitored by Ion Mobility Mass Spectrometry. *J. Am. Chem. Soc.* 131 (12), 1452–1459. doi:10.1021/ja8055134
- Dyachenko, A., Gruber, R., Shimon, L., Horovitz, A., and Sharon, M. (2013). Allosteric Mechanisms Can Be Distinguished Using Structural Mass Spectrometry. *Proc. Natl. Acad. Sci.* 110, 7235–7239. doi:10.1073/pnas.1302395110/-/DCSupplemental/pnas.201302395SI.pdf
- El-Hawiet, A., Chen, Y., Shams-Ud-Doha, K., Kitova, E. N., Kitov, P. I., Bode, L., et al. (2018). Screening Natural Libraries of Human Milk Oligosaccharides against Lectins Using CaR-ESI-MS. *Analyst* 143 (2), 536–548. doi:10.1039/C7AN01397C
- El-Hawiet, A., Shoemaker, G. K., Daneshfar, R., Kitova, E. N., and Klassen, J. S. (2012). Applications of a Catch and Release Electrospray Ionization Mass Spectrometry Assay for Carbohydrate Library Screening. *Anal. Chem.* 84 (1), 50–58. doi:10.1021/ac202760e
- Engen, J. R., and Wales, T. E. (2015). Analytical Aspects of Hydrogen Exchange Mass Spectrometry. *Annu. Rev. Anal. Chem.* 8, 127–148. doi:10.1146/annurev-anchem-062011-143113
- Esser, T. K., Boehning, J., and Fremdling, P. (2021). *Mass-selective and Ice-free Cryo-EM Protein Sample Preparation via Native Electrospray Ion-Beam Deposition*. bioRxiv.
- Fantini, S. M., Parson, K. F., Niu, S., Liu, J., Polasky, D. A., Dixit, S. M., et al. (2019). Collision Induced Unfolding Classifies Ligands Bound to the Integral Membrane Translocator Protein. *Anal. Chem.* 91 (24), 15469–15476. doi:10.1021/acs.analchem.9b03208
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989). Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* 246, 64–71. doi:10.1126/science.2675315
- Fenselau, C., Szilágyi, Z., and Williams, T. (2000). Intercharge Distances in Zn7-Metallothionein Analyzed by Nanospray on a Quadrupole Ion Trap and Molecular Modeling. *J. Mass. Spectrom. Soc. Jpn.* 48 (1), 23–25. doi:10.5702/masspec.48.23
- Fernandez De la Mora, J. (2000). Electrospray Ionization of Large Multiply Charged Species Proceeds via Dole's Charged Residue Mechanism. *Analytica Chim. Acta* 406 (1), 93–104. doi:10.1016/S0003-2670(99)00601-7
- Fort, K. L., van de Waterbeemd, M., Boll, D., Reinhardt-Szyba, M., Belov, M. E., Sasaki, E., et al. (2018). Expanding the Structural Analysis Capabilities on an Orbitrap-Based Mass Spectrometer for Large Macromolecular Complexes. *Analyst* 143 (1), 100–105. doi:10.1039/C7AN01629H
- Gan, J., Ben-nissan, G., Arkind, G., Tarnavsky, M., Trudeau, D., Noda Garcia, L., et al. (2017). Native Mass Spectrometry of Recombinant Proteins from Crude Cell Lysates. *Anal. Chem.* 89, 4398–4404. doi:10.1021/acs.analchem.7b00398
- Ganem, B., Li, Y. T., and Henion, J. D. (1991). Detection of Noncovalent Receptor-Ligand Complexes by Mass Spectrometry. *J. Am. Chem. Soc.* 113 (16), 6294–6296. doi:10.1021/ja00016a069
- Gault, J., Donlan, J. A. C., Liko, I., Hopper, J. T. S., Gupta, K., Housden, N. G., et al. (2016). High-resolution Mass Spectrometry of Small Molecules Bound to Membrane Proteins. *Nat. Methods* 13 (4), 333–336. doi:10.1038/nmeth.3771
- Gault, J., Liko, I., Landreh, M., Shutin, D., Bolla, J. R., Jefferies, D., et al. (2020). Combining Native and 'omics' Mass Spectrometry to Identify Endogenous Ligands Bound to Membrane Proteins. *Nat. Methods* 17 (5), 505–508. doi:10.1038/s41592-020-0821-0
- Gavriilidou, A. F. M., Gülbakan, B., and Zenobi, R. (2015). Influence of Ammonium Acetate Concentration on Receptor-Ligand Binding Affinities Measured by Native Nano ESI-MS: A Systematic Study. *Anal. Chem.* 87, 10378–10384. doi:10.1021/acs.analchem.5b02478
- Gavriilidou, A. F. M., Holding, F. P., Coyle, J. E., and Zenobi, R. (2018). Application of Native ESI-MS to Characterize Interactions between Compounds Derived from Fragment-Based Discovery Campaigns and Two Pharmaceutically Relevant Proteins. *SLAS DISCOVERY: Advancing Sci. Drug Discov.* 23 (9), 951–959. doi:10.1177/2472555218775921
- Gavriilidou, A. F. M., Holding, F. P., Mayer, D., Coyle, J. E., Veprintsev, D. B., and Zenobi, R. (2018). Native Mass Spectrometry Gives Insight into the Allosteric Binding Mechanism of M2 Pyruvate Kinase to Fructose-1,6-Bisphosphate. *Biochemistry* 57 (11), 1685–1689. doi:10.1021/acs.biochem.7b01270
- Gavriilidou, A. F. M., Hunziker, H., Mayer, D., Vuckovic, Z., Veprintsev, D. B., and Zenobi, R. (2019). Insights into the Basal Activity and Activation Mechanism of the  $\beta$ 1 Adrenergic Receptor Using Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 30 (3), 529–537. doi:10.1007/s13361-018-2110-z
- Geels, R. B. J., Van Der Vies, S. M., Heck, A. J. R., and Heeren, R. M. A. (2006). Electron Capture Dissociation as Structural Probe for Noncovalent Gas-phase Protein Assemblies. *Anal. Chem.* 78 (20), 7191–7196. doi:10.1021/ac060960p
- Grabenaus, M., Wyttenbach, T., Sanghera, N., Slade, S. E., Pinheiro, T. J. T., Scrivens, J. H., et al. (2010). Conformational Stability of Syrian Hamster Prion Protein PrP(90–231). *J. Am. Chem. Soc.* 132 (26), 8816–8818. doi:10.1021/ja100243h
- Griffiths, R. L., Sisley, E. K., Lopez-Clavijo, A. F., Simmonds, A. L., Styles, I. B., and Cooper, H. J. (2019). Native Mass Spectrometry Imaging of Intact Proteins and Protein Complexes in Thin Tissue Sections. *Int. J. Mass Spectrom.* 437, 23–29. doi:10.1016/j.ijms.2017.10.009
- Gubellini, F., Verdon, G., Karpowich, N. K., Luff, J. D., Boël, G., Gauthier, N., et al. (2011). Physiological Response to Membrane Protein Overexpression in *E. coli*. *Mol. Cell Proteomics* 10 (10), 007930. doi:10.1074/mcp.M111.007930



- Gülbakan, B., Barylyuk, K., and Zenobi, R. (2015). Determination of Thermodynamic and Kinetic Properties of Biomolecules by Mass Spectrometry. *Curr. Opin. Biotechnol.* 31, 65–72. doi:10.1016/j.copbio.2014.08.003
- Gupta, K., Li, J., Liko, I., Gault, J., Bechara, C., Wu, D., et al. (2018). Identifying Key Membrane Protein Lipid Interactions Using Mass Spectrometry. *Nat. Protoc.* 13 (5), 1106–1120. doi:10.1038/nprot.2018.014
- Han, L., Hyung, S.-J., Mayers, J. J. S., and Ruotolo, B. T. (2011). Bound Anions Differentially Stabilize Multiprotein Complexes in the Absence of Bulk Solvent. *J. Am. Chem. Soc.* 133 (29), 11358–11367. doi:10.1021/ja203527a
- Han, L., Hyung, S.-J., and Ruotolo, B. T. (2012). Bound Cations Significantly Stabilize the Structure of Multiprotein Complexes in the Gas Phase. *Angew. Chem. Int. Ed.* 51 (23), 5692–5695. doi:10.1002/anie.201109127
- Han, L., and Ruotolo, B. T. (2013). Traveling-wave Ion Mobility-Mass Spectrometry Reveals Additional Mechanistic Details in the Stabilization of Protein Complex Ions through Tuned Salt Additives. *Int. J. Ion Mobil. Spec.* 16 (1), 41–50. doi:10.1007/s12127-013-0121-9
- Heck, A. J. R., and van den Heuvel, R. H. H. (2004). Investigation of Intact Protein Complexes by Mass Spectrometry. *Mass. Spectrom. Rev.* 23 (5), 368–389. doi:10.1002/mas.10081
- Hernández, H., Dziembowski, A., Taverner, T., Séraphin, B., and Robinson, C. V. (2006). Subunit Architecture of Multimeric Complexes Isolated Directly from Cells. *EMBO Rep.* 7 (6), 605–610. doi:10.1038/sj.embor.7400702
- Hernández, H., and Robinson, C. V. (2007). Determining the Stoichiometry and Interactions of Macromolecular Assemblies from Mass Spectrometry. *Nat. Protoc.* 2 (3), 715–726. doi:10.1038/nprot.2007.73
- Hopper, J. T. S., and Oldham, N. J. (2011). Alkali Metal Cation-Induced Destabilization of Gas-phase Protein-Ligand Complexes: Consequences and Prevention. *Anal. Chem.* 83 (19), 7472–7479. doi:10.1021/ac201686f
- Hopper, J. T. S., and Oldham, N. J. (2009). Collision Induced Unfolding of Protein Ions in the Gas Phase Studied by Ion Mobility-Mass Spectrometry: The Effect of Ligand Binding on Conformational Stability. *J. Am. Soc. Mass. Spectrom.* 20 (10), 1851–1858. doi:10.1016/j.jasms.2009.06.010
- Hopper, J. T. S., Yu, Y. T.-C., Li, D., Raymond, A., Bostock, M., Liko, I., et al. (2013). Detergent-free Mass Spectrometry of Membrane Protein Complexes. *Nat. Methods* 10 (12), 1206–1208. doi:10.1038/nmeth.2691
- Housden, N. G., Hopper, J. T. S., Lukyanova, N., Rodriguez-Larrea, D., Wojdyla, J. A., Klein, A., et al. (2013). Intrinsically Disordered Protein Threads through the Bacterial Outer-Membrane Porin OmpF. *Science* 340 (6140), 1570–1574. doi:10.1126/science.1237864
- Hyung, S.-J., Robinson, C. V., and Ruotolo, B. T. (2009). Gas-Phase Unfolding and Disassembly Reveals Stability Differences in Ligand-Bound Multiprotein Complexes. *Chem. Biol.* 16 (4), 382–390. doi:10.1016/j.chembiol.2009.02.008
- Iribarne, J. V., and Thomson, B. A. (1976). On the Evaporation of Small Ions from Charged Droplets. *J. Chem. Phys.* 64 (6), 2287. doi:10.1063/1.432536
- Jacques, M., Jeffries, W., and Jean-Pierre, C. (1965). On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* 12, 88–118.
- Jørgensen, T. J. D., Roepstorff, P., and Heck, A. J. R. (1998). Direct Determination of Solution Binding Constants for Noncovalent Complexes between Bacterial Cell Wall Peptide Analogues and Vancomycin Group Antibiotics by Electrospray Ionization Mass Spectrometry. *Anal. Chem.* 70 (20), 4427–4432. doi:10.1021/ac980563h
- Juraschek, R., Dülcks, T., and Karas, M. (1999). Nanoelectrospray-More Than Just a Minimized-Flow Electrospray Ionization Source. *J. Am. Soc. Mass. Spectrom.* 10, 300–308. doi:10.1016/S1044-0305(98)00157-3
- Jurneczko, E., and Barran, P. E. (2011). How Useful Is Ion Mobility Mass Spectrometry for Structural Biology? the Relationship between Protein crystal Structures and Their Collision Cross Sections in the Gas Phase. *Analyst* 136 (1), 20–28. doi:10.1039/c0an00373e
- Katta, V., and Chait, B. T. (1991). Observation of the Heme-Globin Complex in Native Myoglobin by Electrospray-Ionization Mass Spectrometry. *J. Am. Chem. Soc.* 113 (13), 8534–8535. doi:10.1021/ja00022a058
- Kaur, U., Johnson, D. T., Chea, E. E., Deredge, D. J., Espino, J. A., and Jones, L. M. (2019). Evolution of Structural Biology through the Lens of Mass Spectrometry. *Anal. Chem.* 91 (1), 142–155. doi:10.1021/acs.analchem.8b05014
- Kebarle, P., and Peschke, M. (2000). On the Mechanisms by Which the Charged Droplets Produced by Electrospray lead to Gas Phase Ions. *Analytica Chim. Acta* 406, 11–35. doi:10.1016/S0003-2670(99)00598-x
- Kebarle, P., and Verkerk, U. H. (2009). Electrospray: from Ions in Solution to Ions in the Gas Phase, what We Know Now. *Mass. Spectrom. Rev.* 28 (6), 898–917. doi:10.1002/mas.20247/full
- Keener, J. E., Zhang, G., and Marty, M. T. (2021). Native Mass Spectrometry of Membrane Proteins. *Anal. Chem.* 93 (1), 583–597. doi:10.1021/acs.analchem.0c04342
- Khristenko, N., Amato, J., Livet, S., Pagano, B., Randazzo, A., and Gabelica, V. (2019). Native Ion Mobility Mass Spectrometry: When Gas-phase Ion Structures Depend on the Electrospray Charging Process. *J. Am. Soc. Mass. Spectrom.* 30 (6), 1069–1081. doi:10.1007/s13361-019-02152-3
- Koeniger, S. L., and Clemmer, D. E. (2007). Resolution and Structural Transitions of Elongated States of Ubiquitin. *J. Am. Soc. Mass. Spectrom.* 18 (2), 322–331. doi:10.1016/j.jasms.2006.09.025
- Konermann, L., Ahadi, E., Rodriguez, A. D., and Vahidi, S. (2013). Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* 85 (1), 2–9. doi:10.1021/ac302789c
- Konermann, L., and Douglas, D. J. (1998). Unfolding of Proteins Monitored by Electrospray Ionization Mass Spectrometry: A Comparison of Positive and Negative Ion Modes. *J. Am. Soc. Mass. Spectrom.* 9 (12), 1248–1254. doi:10.1016/S1044-0305(98)00103-2
- Konermann, L., Pan, J., and Liu, Y.-H. (2011). Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* 40 (3), 1224–1234. doi:10.1039/c0cs00113a
- Koshland, D. E., Némethy, G., and Filmer, D. (1965). Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* 5 (1), 365–385. doi:10.1021/bi00865a047
- Kyrilis, F. L., Semchonok, D. A., Skolidis, I., Tüting, C., Hamdi, F., O'Reilly, F. J., et al. (2021). Integrative Structure of a 10-megadalton Eukaryotic Pyruvate Dehydrogenase Complex from Native Cell Extracts. *Cel Rep.* 34 (6), 108727. doi:10.1016/j.celrep.2021.108727
- Laganowsky, A., Reading, E., Hopper, J. T. S., and Robinson, C. V. (2013). Mass Spectrometry of Intact Membrane Protein Complexes. *Nat. Protoc.* 8 (4), 639–651. doi:10.1038/nprot.2013.024
- Lai, A. L., Clerico, E. M., Blackburn, M. E., Patel, N. A., Robinson, C. V., Borbat, P. P., et al. (2017). Key Features of an Hsp70 Chaperone Allosteric Landscape Revealed by Ion-Mobility Native Mass Spectrometry and Double Electron-Electron Resonance. *J. Biol. Chem.* 292 (21), 8773–8785. doi:10.1074/jbc.M116.770404
- Lanucara, F., Holman, S. W., Gray, C. J., and Evers, C. E. (2014). The Power of Ion Mobility-Mass Spectrometry for Structural Characterization and the Study of Conformational Dynamics. *Nat. Chem* 6 (4), 281–294. doi:10.1038/nchem.1889
- Lebrette, H., Borezée-Durant, E., Martin, L., Richaud, P., Boeri Erba, E., and Cavazza, C. (2015). Novel Insights into Nickel Import in *Staphylococcus aureus*: the Positive Role of Free Histidine and Structural Characterization of a New Thiazolidine-type Nickel Chelator. *Metallomics* 7, 613–621. doi:10.1039/c4mt00295d
- Liu, L., Kitova, E. N., and Klassen, J. S. (2011). Quantifying Protein-Fatty Acid Interactions Using Electrospray Ionization Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* 22 (2), 310–318. doi:10.1007/s13361-010-0032-5
- Loo, J. A., Berhane, B., Kaddis, C. S., Wooding, K. M., Xie, Y., Kaufman, S. L., et al. (2005). Electrospray Ionization Mass Spectrometry and Ion Mobility Analysis of the 20S Proteasome Complex. *J. Am. Soc. Mass. Spectrom.* 16 (7), 998–1008. doi:10.1016/j.jasms.2005.02.017
- Loo, R. R. O., Goodlett, D. R., Smith, R. D., and Loo, J. A. (1993). Observation of a Noncovalent Ribonuclease S-Protein/S-Peptide Complex by Electrospray Ionization Mass Spectrometry. *J. Am. Chem. Soc.* 115 (10), 4391–4392. doi:10.1021/ja00063a079
- Lössl, P., Snijder, J., and Heck, A. J. R. (2014). Boundaries of Mass Resolution in Native Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* 25 (6), 906–917. doi:10.1007/s13361-014-0874-3
- Lu, W., Kostic, M., Zhang, T., Che, J., Patricelli, M. P., Jones, L. H., et al. (2021). Fragment-based Covalent Ligand Discovery. *RSC Chem. Biol.* 2 (2), 354–367. doi:10.1039/d0cb00222d
- Maple, H. J., Garlish, R. A., Rigau-Roca, L., Porter, J., Whitcombe, I., Prosser, C. E., et al. (2012). Automated Protein-Ligand Interaction Screening by Mass Spectrometry. *J. Med. Chem.* 55 (2), 837–851. doi:10.1021/jm201347k
- Marchand, A., Czar, M. F., Eggel, E. N., Kaeslin, J., and Zenobi, R. (2020). Studying Biomolecular Folding and Binding Using Temperature-Jump Mass Spectrometry. *Nat. Commun.* 11 (1), 1. doi:10.1038/s41467-019-14179-x

- Marchand, A., Rosu, F., Zenobi, R., and Gabelica, V. (2018). Thermal Denaturation of DNA G-Quadruplexes and Their Complexes with Ligands: Thermodynamic Analysis of the Multiple States Revealed by Mass Spectrometry. *J. Am. Chem. Soc.* 140 (39), 12553–12565. doi:10.1021/jacs.8b07302
- Marcoux, J., Wang, S. C., Politis, A., Reading, E., Ma, J., Biggin, P. C., et al. (2013). Mass Spectrometry Reveals Synergistic Effects of Nucleotides, Lipids, and Drugs Binding to a Multidrug Resistance Efflux Pump. *Proc. Natl. Acad. Sci.* 110 (24), 9704–9709. doi:10.1073/pnas.1303888110
- Marklund, E. G., and Benesch, J. L. (2019). Weighing-up Protein Dynamics: the Combination of Native Mass Spectrometry and Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* 54, 50–58. doi:10.1016/j.sbi.2018.12.011
- Marty, M. T., Hoi, K. K., Gault, J., and Robinson, C. V. (2016). Probing the Lipid Annular Belt by Gas-phase Dissociation of Membrane Proteins in Nanodiscs. *Angew. Chem. Int. Ed.* 55 (2), 550–554. doi:10.1002/anie.201508289
- Masson, G. R., Burke, J. E., Ahn, N. G., Anand, G. S., Borchers, C., Brier, S., et al. (2019). Recommendations for Performing, Interpreting and Reporting Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) Experiments. *Nat. Methods* 16 (7), 595–602. doi:10.1038/s41592-019-0459-y
- McKay, A. R., Ruotolo, B. T., Ilag, L. L., and Robinson, C. V. (2006). Mass Measurements of Increased Accuracy Resolve Heterogeneous Populations of Intact Ribosomes. *J. Am. Chem. Soc.* 128 (35), 11433–11442. doi:10.1021/ja061468q
- Mehmood, S., Allison, T. M., and Robinson, C. V. (2015). Mass Spectrometry of Protein Complexes: From Origins to Applications. *Annu. Rev. Phys. Chem.* 66 (1), 453–474. doi:10.1146/annurev-physchem-040214-121732
- Meyer, B., and Peters, T. (2003). NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. *Angew. Chem. Int. Ed.* 42 (8), 864–890. doi:10.1002/anie.200390233
- Mirgorodskaya, O. A., Shevchenko, A. A., Chernushevich, I. V., Dodonov, A. F., and Miroshnikov, A. I. (1994). Electrospray-Ionization Time-of-Flight Mass Spectrometry in Protein Chemistry. *Analytical Chemistry* 66.1, 99–107.
- Morris, H. R., Paxton, T., Dell, A., Langhorne, J., Berg, M., and Bordoli, R. S. (1996). High Sensitivity Collisionally-Activated Decomposition Tandem Mass Spectrometry on a Novel Quadrupole/Orthogonal-Acceleration Time-of-Flight Mass Spectrometer. *Rapid Communications in Mass Spectrometry* 10.0, 889–896.
- Nematollahi, L. A., Garza-Garcia, A., Bechara, C., Esposito, D., Morgner, N., Robinson, C. V., et al. (2015). Flexible Stoichiometry and Asymmetry of the PIDDosome Core Complex by Heteronuclear NMR Spectroscopy and Mass Spectrometry. *J. Mol. Biol.* 427 (4), 737–752. doi:10.1016/j.jmb.2014.11.021
- Nesatyy, V. J., and Suter, M. J.-F. (2004). On the Conformation-dependent Neutralization Theory and Charging of Individual Proteins and Their Non-covalent Complexes in the Gas Phase. *J. Mass. Spectrom.* 39 (1), 93–97. doi:10.1002/jms.522
- Nguyen, G. T. H., Bennett, J. L., Liu, S., Hancock, S. E., Winter, D. L., Glover, D. J., et al. (2021). Multiplexed Screening of Thousands of Natural Products for Protein-Ligand Binding in Native Mass Spectrometry. *J. Am. Chem. Soc.* 143, 21379–21387. doi:10.1021/jacs.1c10408
- Niu, S., Rabuck, J. N., and Ruotolo, B. T. (2013). Ion Mobility-Mass Spectrometry of Intact Protein-Ligand Complexes for Pharmaceutical Drug Discovery and Development. *Curr. Opin. Chem. Biol.* 17, 809–817. doi:10.1016/j.cbpa.2013.06.019
- Oikonomou, C. M., and Jensen, G. J. (2017). Cellular Electron Cryotomography: Toward Structural Biology *In Situ*. *Annu. Rev. Biochem.* 86, 873–896. doi:10.1146/annurev-biochem-061516-044741
- Olinares, P. D. B., Kang, J. Y., Llewellyn, E., Chiu, C., Chen, J., Malone, B., et al. (2021). Native Mass Spectrometry-Based Screening for Optimal Sample Preparation in Single-Particle Cryo-EM. *Structure* 29 (2), 186–195. doi:10.1016/j.str.2020.11.001
- Peschke, M., Verkerk, U. H., and Kebarle, P. (2004). Features of the ESI Mechanism that Affect the Observation of Multiply Charged Noncovalent Protein Complexes and the Determination of the Association Constant by the Titration Method. *J. Am. Soc. Mass. Spectrom.* 15 (10), 1424–1434. doi:10.1016/j.jasms.2004.05.005
- Politis, A., Stengel, F., Hall, Z., Hernández, H., Leitner, A., Walzthoeni, T., et al. (2014). A Mass Spectrometry-Based Hybrid Method for Structural Modeling of Protein Complexes. *Nat. Methods* 11, 403–406. doi:10.1038/nmeth.2841
- Rathore, D., Faustino, A., Schiel, J., Pang, E., Boyne, M., and Rogstad, S. (2018). The Role of Mass Spectrometry in the Characterization of Biologic Protein Products. *Expert Rev. Proteomics* 15 (5), 431–449. doi:10.1080/14789450.2018.1469982
- Reading, E., Liko, I., Allison, T. M., Benesch, J. L. P., Laganowsky, A., and Robinson, C. V. (2015). The Role of the Detergent Micelle in Preserving the Structure of Membrane Proteins in the Gas Phase. *Angew. Chem.* 127, 4660–4664. doi:10.1002/anie.20141162210.1002/ange.201411622
- Rezaei Darestani, R., Winter, P., Kitova, E. N., Tuszyński, J. A., and Klassen, J. S. (2016). Screening Anti-cancer Drugs against Tubulin Using Catch-And-Release Electrospray Ionization Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* 27 (5), 876–885. doi:10.1007/s13361-016-1360-x
- Robinson, C. V. (2019). Mass Spectrometry: From Plasma Proteins to Mitochondrial Membranes. *Proc. Natl. Acad. Sci. USA* 116 (8), 2814–2820. doi:10.1073/pnas.1820450116
- Rose, R. J., Damoc, E., Denisov, E., Makarov, A., and Heck, A. J. R. (2012). High-sensitivity Orbitrap Mass Analysis of Intact Macromolecular Assemblies. *Nat. Methods* 9 (11), 1084–1086. doi:10.1038/nmeth.2208
- Rostom, A. A., and Robinson, C. V. (1999). Detection of the Intact GroEL Chaperonin Assembly by Mass Spectrometry. *J. Am. Chem. Soc.* 121 (7), 4718–4719. doi:10.1021/ja990238r
- Rostom, A. A., Fucini, P., Benjamin, D. R., Juenemann, R., Nierhaus, K. H., Hartl, F. U., et al. (2000). Detection and Selective Dissociation of Intact Ribosomes in a Mass Spectrometer. *Proc. Natl. Acad. Sci.* 97 (10), 5185–5190. doi:10.1073/pnas.97.10.5185
- Rožman, M., and Gaskell, S. J. (2010). Non-covalent Interactions of Alkali Metal Cations with Singly Charged Tryptic Peptides. *J. Mass. Spectrom.* 45 (12), 1409–1415. doi:10.1002/jms.1856
- Ruotolo, B., and Robinson, C. (2006). Aspects of Native Proteins Are Retained in Vacuum. *Curr. Opin. Chem. Biol.* 10 (5), 402–408. doi:10.1016/j.cbpa.2006.08.020
- Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S.-J., and Robinson, C. V. (2008). Ion Mobility-Mass Spectrometry Analysis of Large Protein Complexes. *Nat. Protoc.* 3 (7), 1139–1152. doi:10.1038/nprot.2008.78
- Ruotolo, B. T., Hyung, S.-J., Robinson, P. M., Giles, K., Bateman, R. H., and Robinson, C. V. (2007). Ion Mobility-Mass Spectrometry Reveals Long-Lived, Unfolded Intermediates in the Dissociation of Protein Complexes. *Angew. Chem. Int. Ed.* 46 (42), 8001–8004. doi:10.1002/anie.200702161
- Santambrogio, C., Ricagno, S., Sobott, F., Colombo, M., Bolognesi, M., and Grandori, R. (2011). Characterization of  $\beta$ 2-microglobulin Conformational Intermediates Associated to Different Fibrillation Conditions. *J. Mass. Spectrom.* 46 (8), 734–741. doi:10.1002/jms.1946
- Shelimov, K. B., Clemmer, D. E., Hudgins, R. R., and Jarrold, M. F. (1997). Protein Structure In Vacuo: Gas-phase Conformations of BPTI and Cytochrome C. *J. Am. Chem. Soc.* 119 (9), 2240–2248. doi:10.1021/ja9619059
- Slonusakiewicz, J., Ng, W., Dai, J., Pasternak, A., and Redden, P. (2005). Frontal Affinity Chromatography with MS Detection (FAC-MS) in Drug Discovery. *Drug Discov. Today* 10 (6), 409–416. doi:10.1016/s1359-6446(04)03360-4
- Smith, D. L., Deng, Y., and Zhang, Z. (1997). Probing the Non-covalent Structure of Proteins by Amide Hydrogen Exchange and Mass Spectrometry. *J. Mass. Spectrom.* 32 (2), 135–146. doi:10.1002/(SICI)1096-9888(199702)32:2<135::AID-JMS486>3.0.CO;2-M
- Smith, D. P., Radford, S. E., and Ashcroft, A. E. (2010). Elongated Oligomers in 2-microglobulin Amyloid Assembly Revealed by Ion Mobility Spectrometry-Mass Spectrometry. *Proc. Natl. Acad. Sci.* 107 (15), 6794–6798. doi:10.1073/pnas.0913046107
- Snijder, J., Rose, R. J., Veasler, D., Johnson, J. E., and Heck, A. J. R. (2013). Studying 18 MDa Virus Assemblies with Native Mass Spectrometry. *Angew. Chem. Int. Ed.* 52 (14), 4020–4023. doi:10.1002/anie.201210197
- Snijder, J., Schuller, J. M., Wiegand, A., Lössl, P., Schmelling, N., Axmann, I. M., et al. (2017). Structures of the Cyanobacterial Circadian Oscillator Frozen in a Fully Assembled State. *Science* 355 (6330), 1181–1184. doi:10.1126/science.aag3218
- Sobott, F., and Robinson, C. V. (2006). Understanding Protein Interactions and Their Representation in the Gas Phase of the Mass Spectrometer. *Princ. Mass. Spectrom. Appl. Biomol.* 2006, 147–175. doi:10.1002/047005042X.ch4
- Stojko, J., Fieulaine, S., Petiot-Bécard, S., Van Dorsselaer, A., Meinel, T., Giglione, C., et al. (2015). Ion Mobility Coupled to Native Mass Spectrometry as a Relevant Tool to Investigate Extremely Small Ligand-Induced Conformational Changes. *Analyst* 140 (21), 7234–7245. doi:10.1039/C5AN01311A

- Strop, P., and Brunger, A. T. (2005). Refractive index-based Determination of Detergent Concentration and its Application to the Study of Membrane Proteins. *Protein Sci.* 14 (8), 2207–2211. doi:10.1110/ps.051543805
- Susa, A. C., Xia, Z., and Williams, E. R. (2017). Small Emitter Tips for Native Mass Spectrometry of Proteins and Protein Complexes from Nonvolatile Buffers that Mimic the Intracellular Environment. *Anal. Chem.* 89 (5), 3116–3122. doi:10.1021/acs.analchem.6b04897
- Terral, G., Beck, A., and Cianférani, S. (2016). Insights from Native Mass Spectrometry and Ion Mobility-Mass Spectrometry for Antibody and Antibody-Based Product Characterization. *J. Chromatogr. B* 1032, 79–90. doi:10.1016/j.jchromb.2016.03.044
- Thompson, N. J., Rosati, S., and Heck, A. J. R. (2014). Performing Native Mass Spectrometry Analysis on Therapeutic Antibodies. *Methods* 65 (1), 11–17. doi:10.1016/j.ymeth.2013.05.003
- Thomson, B. A., and Iribarne, J. V. (1979). Field Induced Ion Evaporation from Liquid Surfaces at Atmospheric Pressure. *J. Chem. Phys.* 71 (11), 4451–4463. doi:10.1063/1.438198
- Van De Waterbeemd, M., Fort, K. L., Boll, D., Reinhardt-Szyba, M., Routh, A., Makarov, A., et al. (2017). High-fidelity Mass Analysis Unveils Heterogeneity in Intact Ribosomal Particles. *Nat. Methods* 14 (3), 283–286. doi:10.1038/nmeth.4147
- Vivat Hannah, V., Atmanene, C., Zeyer, D., Van Dorsselaer, A., and Sanglier-Cianférani, S. (2010). Native MS: an 'ESI' Way to Support Structure- and Fragment-Based Drug Discovery. *Future Med. Chem.* 2 (1), 35–50. doi:10.4155/fmc.09.141
- Wagner, S., Baars, L., Ytterberg, A. J., Klussmeier, A., Wagner, C. S., Nord, O., et al. (2007). Consequences of Membrane Protein Overexpression in *Escherichia coli*. *Mol. Cell Proteomics* 6 (9), 1527–1550. doi:10.1074/mcp.M600431-MCP200
- Wilm, M., and Mann, M. (1996). Analytical Properties of the Nanoelectrospray Ion Source. *Anal. Chem.* 68, 1–8. doi:10.1021/ac9509519
- Wilm, M. S., and Mann, M. (1994). Electrospray and Taylor-Cone Theory, Dole's Beam of Macromolecules at Last? *Int. J. Mass. Spectrom. Ion Process.* 136 (2), 167–180. doi:10.1016/0168-1176(94)04024-9
- Wolynes, P. G. (1995). Biomolecular Folding In Vacuo!!!(?). *Proc. Natl. Acad. Sci.* 92 (7), 2426–2427. doi:10.1073/pnas.92.7.2426
- Woods, L. A., Dolezal, O., Ren, B., Ryan, J. H., Peat, T. S., and Poulsen, S. A. (2016). Native State Mass Spectrometry, Surface Plasmon Resonance, and X-ray Crystallography Correlate Strongly as a Fragment Screening Combination. *J. Med. Chem.* 59, 2192–2204. doi:10.1021/acs.jmedchem.5b01940
- Wortmann, A., Jecklin, M. C., Touboul, D., Badertscher, M., and Zenobi, R. (2008). Binding Constant Determination of High-Affinity Protein-Ligand Complexes by Electrospray Ionization Mass Spectrometry and Ligand Competition. *J. Mass. Spectrom.* 43 (5), 600–608. doi:10.1002/jms.1355
- Wu, C., Klasmeier, J., and Hill, H. H. (1999). Atmospheric Pressure Ion Mobility Spectrometry of Protonated and Sodiated Peptides. *Rapid Commun. Mass. Spectrom.* 13 (12), 1138–1142. doi:10.1002/(sici)1097-0231(19990630)13:12<1138::aid-rcm625>3.0.co;2-8
- Wyttenbach, T., and Bowers, M. T. (2011). Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-free Ion Mobility-Mass Spectrometry Environment. *J. Phys. Chem. B* 115 (42), 12266–12275. doi:10.1021/jp206867a
- Yee, A. W., Aldeghi, M., Blakeley, M. P., Ostermann, A., Mas, P. J., Moulin, M., et al. (2019). A Molecular Mechanism for Transthyretin Amyloidogenesis. *Nat. Commun.* 10 (1), 1–10. doi:10.1038/s41467-019-08609-z
- Yen, H.-Y., Hoi, K. K., Liko, I., Hedger, G., Horrell, M. R., Song, W., et al. (2018). PtdIns(4,5)P2 Stabilizes Active States of GPCRs and Enhances Selectivity of G-Protein Coupling. *Nature* 559, 423–427. doi:10.1038/s41586-018-0325-6
- Yen, H. Y., Hopper, J. T. S., Liko, I., Allison, T. M., Zhu, Y., Wang, D., et al. (2017). Ligand Binding to a G Protein-Coupled Receptor Captured in a Mass Spectrometer. *Sci. Adv.* 3 (6), e1701016–6. doi:10.1126/sciadv.1701016
- Young, L. M., Saunders, J. C., Mahood, R. A., Revell, C. H., Foster, R. J., Tu, L.-H., et al. (2014). Screening and Classifying Small-Molecule Inhibitors of Amyloid Formation Using Ion Mobility Spectrometry-Mass Spectrometry. *Nat. Chem* 7 (1), 73–81. doi:10.1038/nchem.2129
- Zhang, H., Cui, W., Wen, J., Blankenship, R. E., and Gross, M. L. (2011). Native Electrospray and Electron-Capture Dissociation FTICR Mass Spectrometry for Top-Down Studies of Protein Assemblies. *Anal. Chem.* 83 (14), 5598–5606. doi:10.1021/ac200695d
- Zhang, S., Van Pelt, C. K., and Wilson, D. B. (2003). Quantitative Determination of Noncovalent Binding Interactions Using Automated Nanoelectrospray Mass Spectrometry. *Anal. Chem.* 75 (13), 3010–3018. doi:10.1021/ac034089d
- Zhong, Y., Hyung, S. J., and Ruotolo, B. T. (2012). Ion Mobility-Mass Spectrometry for Structural Proteomics. *Expert Rev. Proteomics* 9, 47–58.
- Zhou, M., Sandercock, A. M., Fraser, C. S., Ridlova, G., Stephens, E., Schenauer, M. R., et al. (2008). Mass Spectrometry Reveals Modularity and a Complete Subunit Interaction Map of the Eukaryotic Translation Factor eIF3. *Proc. Natl. Acad. Sci.* 105 (47), 18139–18144. doi:10.1073/pnas.0801313105
- Zweers, J. C., Wiegert, T., and van Dijk, J. M. (2009). Stress-responsive Systems Set Specific Limits to the Overproduction of Membrane Proteins in *Bacillus subtilis*. *Appl. Environ. Microbiol.* 75 (23), 7356–7364. doi:10.1128/AEM.01560-09

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

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

Copyright © 2022 Gavriilidou, Sokratous, Yen and De Colibus. 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.



# Targeting the Essential Transcription Factor HP1043 of *Helicobacter pylori*: A Drug Repositioning Study

Federico Antoniciello<sup>1</sup>, Davide Roncarati<sup>1</sup>, Annamaria Zannoni<sup>1†</sup>, Elena Chiti<sup>1</sup>, Vincenzo Scarlato<sup>1</sup> and Federica Chiappori<sup>2\*</sup>

<sup>1</sup>Department of Pharmacy and Biotechnology (FaBiT), University of Bologna, Bologna, Italy, <sup>2</sup>Istituto di Tecnologie Biomediche–Consiglio Nazionale Delle Ricerche (ITB-CNR), Segrate (MI), Italy

## OPEN ACCESS

### Edited by:

Marco Nardini,  
University of Milan, Italy

### Reviewed by:

Giuseppe Zanotti,  
University of Padua, Italy  
Hyouon Sook Kim,  
National Cancer Center, South Korea

### \*Correspondence:

Federica Chiappori  
federica.chiappori@cnr.it

### <sup>†</sup>Present address:

Annamaria Zannoni,  
Institute for Molecular Infection  
Biology, University of Würzburg,  
Germany

### Specialty section:

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

**Received:** 01 March 2022

**Accepted:** 07 April 2022

**Published:** 11 May 2022

### Citation:

Antoniciello F, Roncarati D, Zannoni A,  
Chiti E, Scarlato V and Chiappori F  
(2022) Targeting the Essential  
Transcription Factor HP1043 of  
*Helicobacter pylori*: A Drug  
Repositioning Study.  
Front. Mol. Biosci. 9:887564.  
doi: 10.3389/fmolb.2022.887564

Antibiotic-resistant bacterial pathogens are a very challenging problem nowadays. *Helicobacter pylori* is one of the most widespread and successful human pathogens since it colonizes half of the world population causing chronic and atrophic gastritis, peptic ulcer, mucosa-associated lymphoid tissue-lymphoma, and even gastric adenocarcinoma. Moreover, it displays resistance to numerous antibiotics. One of the *H. pylori* pivotal transcription factors, HP1043, plays a fundamental role in regulating essential cellular processes. Like other bacterial transcription factors, HP1043 does not display a eukaryote homolog. These characteristics make HP1043 a promising candidate to develop novel antibacterial strategies. Drug repositioning is a relatively recent strategy employed in drug development; testing approved drugs on new targets considerably reduces the time and cost of this process. The combined computational and *in vitro* approach further reduces the number of compounds to be tested *in vivo*. Our aim was to identify a subset of known drugs able to prevent HP1043 binding to DNA promoters. This result was reached through evaluation by molecular docking the binding capacity of about 14,350 molecules on the HP1043 dimer in both conformations, bound and unbound to the DNA. Employing an ad hoc pipeline including MMGBSA molecular dynamics, a selection of seven drugs was obtained. These were tested *in vitro* by electrophoretic mobility shift assay to evaluate the HP1043–DNA interaction. Among these, three returned promising results showing an appreciable reduction of the DNA-binding activity of HP1043. Overall, we applied a computational methodology coupled with experimental validation of the results to screen a large number of known drugs on one of the *H. pylori* essential transcription factors. This methodology allowed a rapid reduction of the number of drugs to be tested, and the drug repositioning approach considerably reduced the drug design costs. Identified drugs do not belong to the same pharmaceutical category and, by computational analysis, bound different cavities, but all display a reduction of HP1043 binding activity on the DNA.

**Keywords:** HP1043, *Helicobacter pylori*, antibiotic resistance, drug repositioning, essential transcription factor, virtual high-throughput screening (vHTS), MMGBSA



# 1 INTRODUCTION

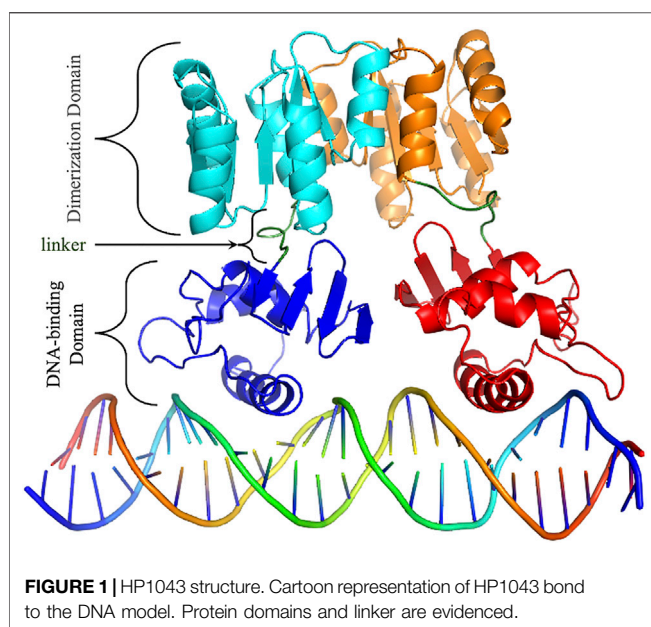
*Helicobacter pylori* is one of the most widespread and successful human pathogens since it colonizes half of the world population (Hooi et al., 2017). Infected people carry this bacterium for decades or even for life; if untreated, *H. pylori* can remain clinically silent for a long time due to the dynamic equilibrium between the bacterium and the host or evolve into chronic gastritis or even more severe diseases such as atrophic gastritis, peptic ulcer, MALT-lymphoma, or gastric adenocarcinoma. Despite its declining incidence rate, gastric cancer remains the fifth most common malignancy in the world and the third leading cause of cancer-related death (Sugano et al., 2015). *H. pylori* infections are currently treated with a combination of a proton pump inhibitor and different antibiotics; unfortunately, the available therapies are losing efficacy because of the antibiotic resistance insurgence, with eradication cure rates lower than 70% (Vianna et al., 2016). For instance, due to a constant increase in *H. pylori* resistance to clarithromycin, the triple clarithromycin-based treatment has become progressively less efficacious (Jenks and Edwards, 2002). For this reason, the World Health Organization (WHO) has included *H. pylori* in a global priority list of 12 antibiotic-resistant bacterial pathogens to help the scientific research prioritize the discovery and development of new antibiotics (WHO, 2017). Alongside multidrug-resistant bacteria, antibiotic-resistant *H. pylori* strains pose a major public health issue, and novel antibacterial strategies against *H. pylori* persistent infection are overdue.

Bacterial pathogens sense the host environment and respond with the expression of gene products required to adapt to a particular niche. These adaptive responses rely on transcriptional regulatory circuits that control the coordinated expression of several proteins, including virulence factors, in space and time (Seshasayee et al., 2006). *H. pylori* makes no exception to this

paradigm, despite a remarkable paucity of annotated transcriptional regulators. To date, only 17 transcriptional regulators have been identified and characterized to different extents (Danielli and Scarlato, 2010). Among them, the HP1043 regulator seems to play an essential role. The *hp1043* gene cannot be deleted, nor the amount of protein modulated, supporting the hypothesis that HP1043 could be involved in the regulation of fundamental cellular processes (Schär et al., 2005). To shed light on this possibility, by chromatin immunoprecipitation-sequencing (ChIP-seq), Pellicciari et al. (2017) comprehensively identified genome-wide the HP1043 *in vivo* binding sites.

HP1043 is a 223 amino acid long protein, composed by a dimerization domain (or response regulatory, residues 2–112) and a DNA-binding domain (OmpR/PhoB-type, residues 118–216) connected by a short 5 amino acid linker (residues 113–117) (Figure 1).

Intriguingly, the study showed that HP1043 dimerizes and binds *in vivo* the promoter of genes involved in all the fundamental processes of the bacterial life cycle, namely, DNA replication, RNA transcription and translation, and energy production and conversion (Pellicciari et al., 2017). The resulting head-to-head dimer conformation, different from the canonical head-to-tail of the OmpR/PhoB subfamily, was theoretically obtained, based on experimental mutagenesis data and inter-domain linker flexibility evaluation (Zannoni et al., 2021). HP1043 appears to be fundamental for the fitness of the bacterium, a prerequisite for a successful infection, and it is a pivotal regulator on which *H. pylori* relies to modulate its metabolism and growing behavior. For these reasons, HP1043 makes a promising candidate to develop novel antibacterial strategies. Essential transcriptional regulators are appealing targets for the development of new antibiotics (González et al., 2018). Indeed, the inhibition of such regulators leads to the altered expression of crucial genes for cell viability. These regulators have no counterpart in humans while being specific for a particular microorganism. They are usually small soluble proteins that are easy to purify (Roncarati et al., 2022). In particular, the ease of handling makes the transcriptional regulators appropriate for experimental approaches such as *in vitro* binding assay, inhibition tests, and co-structural analysis. Moreover, they are ideal proteins for *in silico* approaches, such as structure-based virtual screening for compounds that are able to bind and hinder the regulatory function. The recombinant purified HP1043 of the *H. pylori* 26695 strain has been used to set up a fluorescence-based thermal shift assay to identify HP1043 binding molecules. This approach has been used to carry out a high-throughput screening of 1120 small molecules FDA-approved for human use and off-patent which resulted in the identification of 14 compounds that bind to the native state of HP1043. Notably, seven identified HP1043 binders were natural flavonoids interacting with the helix-turn-helix DNA-binding motif. These natural compounds inhibited the *in vitro* DNA-binding activity of HP1043, and four of them showed bactericidal activity against *H. pylori* (González et al., 2019b). The same screening led to the identification of other FDA-approved drugs that can form stable complexes with HP1043, including the 1,4-dihydropyridine



calcium channel blocker nicardipine (González et al., 2019a), opening the possibility of a drug-repurposing approach for the treatment of *H. pylori* infection.

An alternative approach in the field of drug discovery is based on the generation of high-resolution structural data of the selected target followed by a repositioning strategy (Pushpakom et al., 2019) that might potentially hinder the target's function. To this aim, our group has recently combined biochemical and computational approaches to characterize the binding architecture of the HP1043 regulator of the *H. pylori* G27 strain to a selected target promoter,  $P_{hp1227}$ . Experimental data combined with the available HP1043 NMR structure were used as restraints to guide an *in silico* protein–DNA docking. The generated model shows an HP1043 dimer interacting in a head-to-head conformation with both the major and minor grooves of a target DNA sequence (Zannoni et al., 2021). Moreover, the dimer conformation is known to increase the binding to the promoter region on DNA (Simonovic and Volz, 2001). In this study, we screened a collection of about 14,350 small molecules, composed of approved drugs (FDA approved, NIH clinical collection, Drug Bank approved, and Therapeutics Target Database approved) and “substances with the main target as transcription factors” included in the ZINC15 database. Virtual high-throughput screening (VHTS) was performed on both complexes, HP1043 dimer free, to identify molecules that can bind to the protein–DNA interface, and HP1043 dimer bound to DNA, to identify molecules that can bind the complex and potentially induce DNA unbinding. Promising complexes were analyzed by molecular dynamics as obtained by docking to evaluate the drug-binding effect dynamically. This work aimed to evaluate the reposition of approved drugs and transcription factors known as ligands to the HP1043 *H. pylori* transcription factor, selecting by computational methods a subset of molecules able to prevent DNA binding. Selected drugs were afterward experimentally tested; and among these, three returned promising results showing, *in vitro*, an appreciable reduction of the DNA-binding activity of HP1043.

## 2 MATERIALS AND METHODS

### 2.1 Molecular Docking

The previously obtained molecular models of dimeric HP1043 alone and bound to DNA were employed (Zannoni et al., 2021). HP1043 model dimer bound to the DNA was obtained by a data-driven docking, employing NMR data, and domain X-ray deposited data, and the residues involved in DNA binding were experimentally validated (for details see (Zannoni et al., 2021)). Both complexes were relaxed through 100 ns of molecular dynamic simulation and the most representative conformations obtained from cluster analysis were employed for the virtual screening step.

We used the ZINC database (<http://zinc15.docking.org/>) (Sterling and Irwin, 2015) to select a ligand dataset of about 14,350 molecules from five different datasets. In detail, we selected substances that can target a transcription factor (TF),

Drug Bank-approved molecules (DBa), Food and Drug Administration-approved drugs (FDAa), Therapeutic Target Database-approved drugs (TTDa), and NIH clinical collection (Ncc).

Docking simulations were performed using Autodock 4.2 and AutoDockTool (ADT) interface (Morris et al., 2009). An in-house Perl script pipeline was employed for ligand preparation, parameter file for docking simulation, Slurm job-scheduling input file, and docking results analysis, while the protein structure was prepared with the ADT interface, a docking box of  $126 \times 126 \times 126$  points, with a spacing of 0.375 Å, was centered on the protein–DNA complex, including almost the whole complex, excluding the DNA underside. Considering the unbound complex, two docking boxes were prepared with the same characteristics as the previous one, both centered on the HP1043 DNA-binding domain of each protein chain, since a single box cannot include the entire docking surface.

The affinity maps for all the atom types available in AutoGrid were pre-calculated. Docking simulations were performed by treating the protein as rigid, ligands as flexible, and 50 runs for each simulation of the Lamarckian genetic algorithm were performed using the AutoDock 4.2.6 suite (Morris et al., 2009). Selected molecules were subjected to a deep re-docking screening of 200 runs per simulation, maintaining the other conditions.

The pipeline parses the docking results to identify, for each simulation, the most representative conformation, that is, the best energy conformation and all the conformations with a docking energy within 1 kcal/mol from the first ranked solution. Docking energy, cluster population, estimated  $K_i$  value, and atomic coordinates of each selected solution are extracted. The pipeline reconstructs the coordinate files of protein–ligand complexes. Moreover, final molecules were selected based on the  $K_i$  value, in the order of size of pM or lower and only the re-docking results, on cluster numerosity, greater than 20 units.

### 2.2 Molecular Dynamics

The molecular dynamics of the HP1043 bound or unbound to DNA molecule, as obtained by molecular docking, and in complex with a ligand selection was performed using AMBER 18 (Salomon-Ferrer et al., 2013). To parameterize the complex, ff14SB (Maier et al., 2015) was employed for the protein, OL15 (Cheatham and Case, 2013) force field for DNA, and water was treated as an optimal point charge. The total charge of each complex was balanced with  $\text{Na}^+$  counter ions, and the solution molarity was set to 150 mM adding  $\text{Na}^+$  and  $\text{Cl}^-$  ions. Solvated complexes were minimized for 1,000 steps, and heated until 300 K in 100 ps followed by 50 ps of NPT equilibration. Ten simulations of 10 ns each were performed using the molecular mechanics generalized Born surface area (MMGBSA) protocol, employing periodic boundary conditions. Trajectory and energetic analyses were performed using the cpptraj and MMBPSA.py tools (Miller et al., 2012). In detail, Ca root mean square deviation (RMSD), per-residue root mean square fluctuation (RMSF), the distance between protein interface residues and specific DNA nucleotides, and Ca hierarchical agglomerative clustering analysis were evaluated using cpptraj. To perform binding energy analysis on protein–ligand and

protein (ligand)–DNA complexes, the tool is MMPBSA.py was applied on 50 equally distributed frames along the joint trajectory, and the solvation free energy was evaluated using the modified generalized Born (GB) model (Onufriev et al., 2004) using 1.0, 0.8, and 4.85 values for  $\alpha$ ,  $\beta$ , and  $\gamma$ , respectively, and ions concentration was set at 0.150 M.

LigPlot+ (Laskowski, 2011) was employed for the identification of interacting residues and its classification as contacts or residues establishing hydrogen bonds.

## 2.3 Chemicals

Plerixafor and ribociclib were purchased from AdooQ® Biosciences (United States), and oxcarbazepine and temoporfin were purchased from Cayman Chemical (United States), while dihydroergotoxine, tetraethylenepentamine, and trientine were purchased from Sigma-Aldrich (United States). All drugs were properly stored at  $-20^{\circ}\text{C}$  according to the manufacturer's indications. For each compound, stock solutions of 16.67 mM were prepared by dissolving the powder in either  $\text{H}_2\text{O}$  or 100% DMSO. The pH of aqueous solutions was adjusted when needed.

## 2.4 Overexpression and Purification of Recombinant His6-HP1043

Recombinant N-terminal His-tagged HP1043 wild-type and mutant proteins were overexpressed in *Escherichia coli* BL21 DE3 cells transformed with plasmid pTrc::1043 (Supplementary Table S1). For electrophoretic mobility shift assays (EMSA), HP1043 was purified as previously described (Zannoni et al., 2021). Briefly, the cells were incubated with lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, and 10 mM imidazole), lysozyme (0.5 mg/ml), and  $1\times$  cComplete™ Protease Inhibitor Cocktail (Roche, Basel, Switzerland) for 1 h at  $4^{\circ}\text{C}$ . Afterward, the mix was sonicated and centrifuged at  $22,000\times g$  for 30 min at  $4^{\circ}\text{C}$ . The supernatant was collected and incubated in a batch with  $\text{Ni}^{2+}$ -NTA resin for His-tagged purification. The bound protein was eluted using an imidazole gradient (10–250 mM) and dialyzed twice against the store buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, and 10% glycerol). Protein purity was assessed by SDS-PAGE and its concentration was determined by measuring the absorbance at 280 nm using a NanoDrop® spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States).

## 2.5 Electrophoretic Mobility Shift Assays

The DNA-binding activity of the recombinant HP1043 in the presence of putative inhibitors was assessed by EMSAs as already described (Zannoni et al., 2021). In brief, a 190-base pair (bp) promoter region of *hp1227* (strain 26695 annotation;  $P_{hp1227}$ ) (Tomb et al., 1997) including the HP1043 binding consensus sequence was amplified from pGEMT-P1227\_WT (Supplementary Table S1) by PCR with oligos Php1227\_EMSA\_F and Php1227\_EMSA\_R and used as a target sequence of HP1043 (Supplementary Table S2). A 127-bp sequence obtained through PCR amplification with oligos 16S\_RTF and 16S\_RTR was used as a non-specific control. The recombinant HP1043 protein (4  $\mu\text{M}$ ) was mixed with

approximately 10 ng of target promoter probe and 10 ng of non-specific DNA probe in a 25  $\mu\text{l}$  reaction volume containing  $1\times$  binding buffer (10 mM Tris-HCl pH 7.5, 40 mM KCl, 100  $\mu\text{g}/\text{ml}$  BSA, 1 mM DTT, and 5% glycerol). For binding interference assays, putative binders were added to final concentrations of 1, 0.5, 0.2, 0.1 and 0.05 mM to the mixtures of protein and DNA. Single ready-to-use aliquots were thawed and immediately diluted to the desired concentration and thereby use the same volume of ligand solution in EMSA analysis. Binding assays with DMSO (6% (v/v)) or  $\text{H}_2\text{O}$  instead of inhibitors and  $1\times$  binding buffer instead of the protein were included as controls. The reactions were incubated at room temperature for 30 min and then separated on a 6% non-denaturing polyacrylamide gel  $0.5\times$  running buffer (45 mM Tris pH 7.5 and 45 mM orthoboric acid) at 90 V for 75 min, using a Mini Gel Tank apparatus (Invitrogen, Waltham, MA, United States). DNA bands were stained with  $1\times$  ethidium bromide and visualized using a Gel Doc XR+ image analyzer (BioRad, Hercules, CA, United States).

## 3 RESULTS

### 3.1 Repositioning VHTS

#### 3.1.1 HP1043–DNA Complex

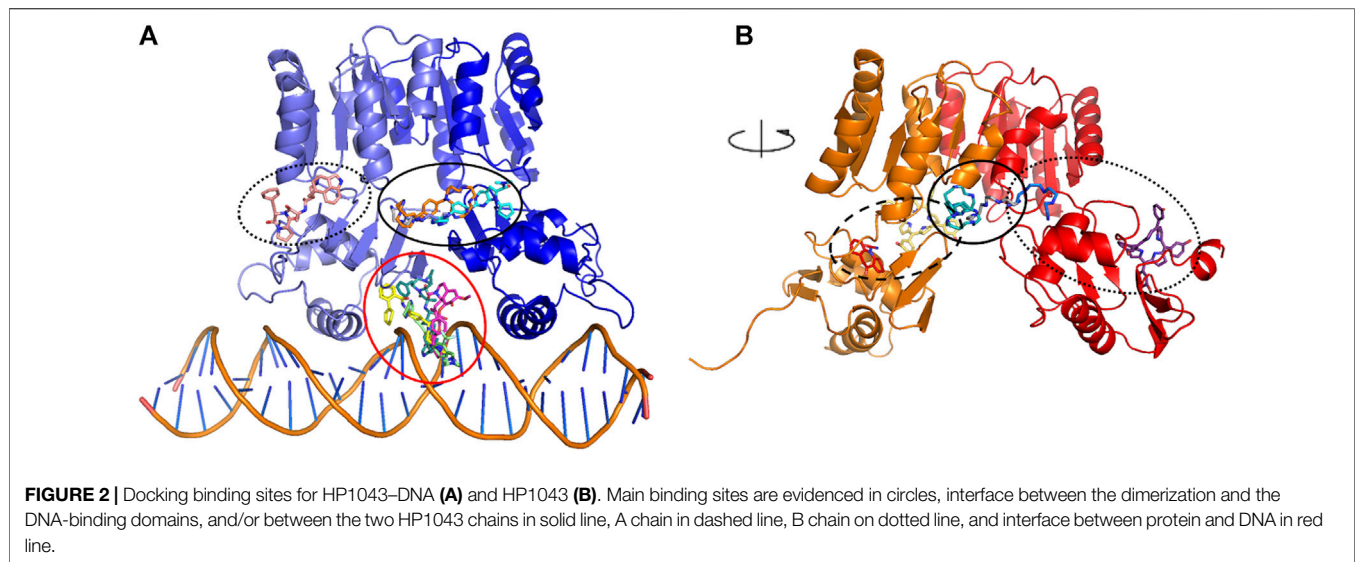
To identify the possible ligands of the HP1043–DNA complex, a dataset of 14,350 approved molecules was screened. The first docking screening returned 180 ligands with a  $K_i$  value in the order of magnitude of pM. These were subjected to a re-docking procedure to identify a subset of the best candidates. The re-docking simulation returned 323 conformations (multiple poses of the same ligand were allowed) with a  $K_i$  value in the pM order of magnitude. Among these, 50 conformations corresponding to 41 different ligands displayed cluster numerosity greater than 20 units, for these reasons, were considered interesting candidates.

A subset of eight ligands bound at the interface between the dimerization and the DNA-binding domains, and/or between the two HP1043 chains and one docked at the B chain external surface, the remaining ligands bound at the interface between the protein and DNA (Figure 2A). The latter can be considered ligands with lower specificity for the free form of HP1043; instead, the first eight ligands can be regarded as promising candidates for targeting HP1043, also in free form.

#### 3.1.2 HP1043 Complex

We applied the aforementioned protocol to the HP1043 dimer unbound to DNA to identify molecules able to bind the transcription factor before binding to the DNA. The first docking screening returned 155 ligands, corresponding to 161 different poses, with an estimated  $K_i$  in the fM and pM order of magnitude and sufficiently represented (see Materials and Methods for details). These were re-docked as previously described, and only two ligands in six different conformations satisfied the  $K_i$  and representative restraints. Here, we also considered ligands with the nM  $K_i$  value with cluster numerosity greater than 100 units, returning overall 19 conformations, corresponding to 15 different ligands or same ligands but localized in different binding sites. Among these,





seven localized at the interface between the two chains, one bound the DNA-binding domain of the A chain at the internal surface. In contrast, nine localized at the external surface, two at the interface between the two domains of the B chain, and the last one bound to the external surface of the B chain (**Figure 2B**).

To compare docking results with the literature, available data (González et al., 2019a; González et al., 2019b) were analyzed with our docking protocol. Different from the literature, we docked the small molecules to HP1043 dimeric conformation; both in the DNA bound and unbound form. Even so, the obtained computational binding energy was comparable to the published results (**Supplementary Table S3** compared to docking binding energy in (González et al., 2019a; 2019b)). Binding energies ranged from  $-5.03$  to  $-10.07$  kcal/mol for the DNA-bound complex and from  $-5.83$  to  $-10.10$  kcal/mol for the unbound complex. Therefore, we selected our screening molecules with a computational binding energy lower than about 10 kcal/mol.

### 3.2 Dynamical Evaluation of Selected Drugs

Among the 56 docking results selected with VHTS (listed in **Supplementary Table S4**), “Not for sale” or too expensive compounds and ligands bound only to DNA molecule were excluded. A subset of 13 compounds (**Table 1**) corresponding to 17 different binding poses was submitted to MMGBSA molecular dynamics simulation to obtain an accurate evaluation of binding mode and energy (**Table 2**) in the same complex conformation as docking results. Otherwise, the first 11 complexes listed in **Table 2** (HP1043–DNA) were simulated and bound to the HP1043 dimer in the complex with the DNA molecule, and the last six complexes listed in **Table 2** (HP1043) were simulated and bound to the HP1043 dimer free form. The selected molecules are mainly representative of the different binding sites. One at the dimer cleft (in green in **Figure 3**, P–P in **Table 2**), involving the residues of both chains (single underlined in **Table 2**), one at the dimerization domain interface between the two chains (in yellow in **Figure 3**, dotted-

underlined in **Table 2**) and one with a lower specificity at the protein–DNA interface (in red in **Figure 3**, P–DNA in **Table 2**), which involves, in addition to the residues of the A chain, the DNA molecule (double-underlined in **Table 2**). Three other binding sites were identified. One localized at the domain interface of the B chain (Bi in **Table 2**) and partially involved residues of the dimer cleft, one localized on the external surface of the A chain (A ext in **Table 2**) and the other on the external surface of the B chain (B ext in **Table 2**).

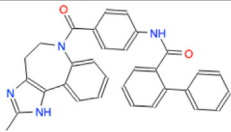
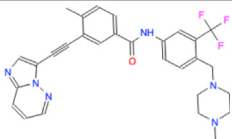
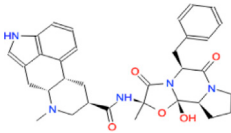
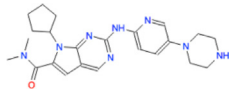
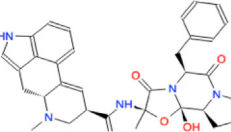
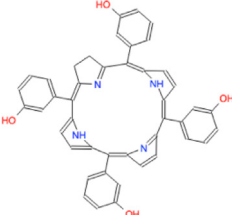
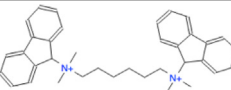
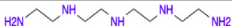
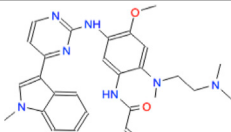

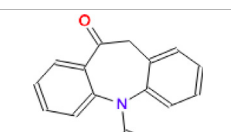
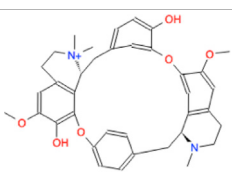
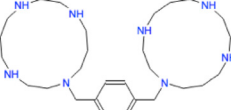
Molecular dynamics stability was evaluated based on backbone RMSD, and all complexes reached equilibrium during the simulation (**Supplementary Figure S1**). Two complexes, namely, HP1043\_DNA-tetraethylenepentamine and HP1043\_DNA-dihydroergotoxine, displayed the expulsion of the ligand during the simulation (**Figures 4A,B**), suggesting a lower affinity of these drugs for the HP1043\_DNA complex, despite the negative binding energy between the drug and HP1043.

Considering ligand binding, as confirmed by interacting residue analysis, most complexes maintained a stable position (see below) except for HP1043\_DNA-ponatinib, HP1043\_DNA-osimertinib, and HP1043\_DNA-tubocurarine complexes, where the three drugs showed a higher affinity for the DNA molecule than the transcription factor. Thus, they maintained only interactions with DNA, suggesting a low specificity of these drugs for HP1043. In addition, ligand tetraethylenepentamine bound to the A chain of HP1043 moved to the P–P interface during the simulation, and this result superimposed to the binding position of the complex with tetraethylenepentamine bound to the P–P interface of HP1043 (**Figure 4C**). Finally, ligand ergotamine, bound to HP1043–DNA, moved toward the A chain, reducing the interaction interface between protein and DNA and causing the displacement of helix H8, necessary for DNA interaction, a partial unfolding of C-terminal domain of the A chain and increasing the DNA bending (**Figure 4D**).

A subset of tested drugs induced movements of the C-terminal domains. Complexes bound to hexafluronium and ribociclib



**TABLE 1** | Selected drugs.

Ligand	Molecule name	Molecule sketch	Ligand	Molecule name	Molecule sketch
ZINC000012503187 (FDAa, DBa, and TTDa)	conivaptan		ZINC000036701290 (FDAa and DBa)	ponatinib	
ZINC000014880002 (TTDa)	dihydroergotaxine		ZINC000072316335 (FDAa and DBa)	ribociclib	
ZINC000052955754 (FDAa, DBa, and TTDa)	ergotamine		ZINC000003934128 (DBa)	temoporfin	
ZINC000001566899 (DBa and TTDa)	hexafluronium		ZINC000019363,537 (FDAa and DBa)	tetraethylenepentamine	
ZINC000098023177 (FDAa and DBa)	osimertinib		ZINC000019364225 (FDAa, DBa, and TTDa)	trientine	
ZINC000000004724 (FDAa, DBa, and TTDa)	oxcarbazepine		ZINC000003978083 (DBa and TTDa)	tubocurarin	
ZINC000022443609 (FDAa, DBa, and TTDa)	plerixafor		-		

FDAa = Food and Drug Administration approved; DBa = Drug Bank approved; TTDa = Therapeutics Target Database approved

displayed a mutual rotation of the C-terminal domains that increased its distance (**Figures 4E,F**), while plerixafor, oxcarbazepine, and both complexes with tetraethylenepentamine bound to HP1043, without DNA, displayed a reduction of domain distance (**Figure 5**). The remaining complexes maintained the conformation of the C-terminal domain during the whole simulation.

Considering the residue fluctuation of complexes without DNA, residues belonging to the DNA-binding domain displayed major

flexibility (**Supplementary Figure S2**). On the other hand, all complexes presented reduced flexibility of A chain residues 120–145 compared to HP1043 unbound to ligands (red line in **Supplementary Figure S2**). These residues belonged to the DNA-binding domain and localized at the domain interface with the dimerization domain. Residues 170–190, belonging to the A chain and composing the interface to DNA (green line in **Supplementary Figure S2**), also displayed a reduced flexibility for all bound complexes except for both complexes bound to tetraethylenepentamine. Overall,

**TABLE 2 |** HP1043 drug docking and MMGBSA.

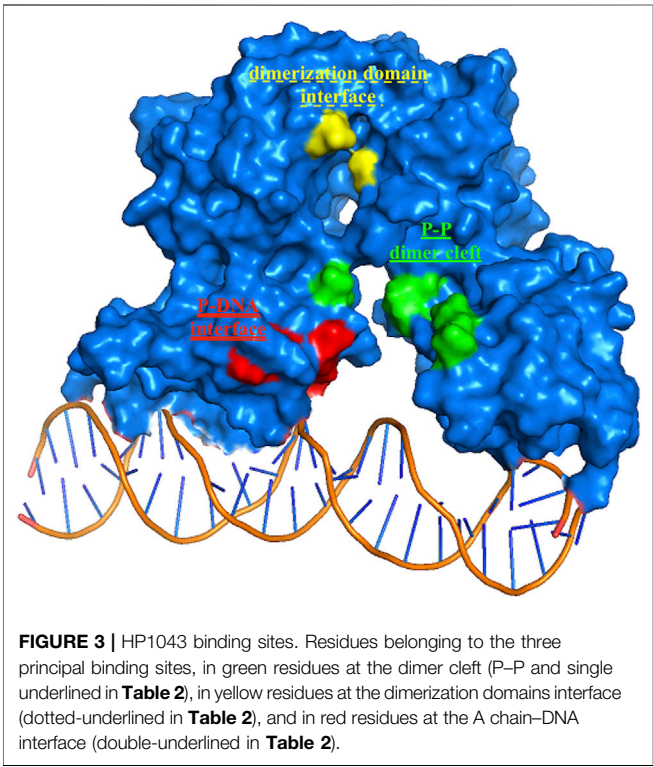
	Binding Site	Molecule name	Binding energy (kcal/mol)	# cluster	Ki	Docking interaction residues (LIGPLOT)	MMGBSA P-L (kcal/mol) (SD)	MD interaction residues (Ligplot)
HP1043-DNA	P-DNA	hexafluronium	-15.41	40	5.06 pM	Contacts: Y360, V365, F372, K417, M418, P421, DNA	-20.1 (±6.93)	Contacts: Y360, R363, E364, V365, K417, DNA
		ponatinib	-13.08	26	256.40 pM	Contacts: Y360, F372, DNA Hbonds: DNA	-35.27 (±12.22)	Contacts: DNA
		osimertinib	-12.89	22	356.10 pM	Contacts: DNA Hbonds: DNA	-10.56 (±4.78)	Contacts: DNA
		tubocurarin	-12.46	51	737.90 pM	Contacts: P148, F149, V365, DNA Hbonds: K194, DNA	-20.99 (±8.11)	Contacts: DNA
		ergotamine	-12.46	26	736.10 pM	Contacts: Y360, V365, E366, V367, F372, DNA Hbonds: DNA	-20.62 (±5.61)	Contacts: L126, I135, Y137, DNA
		conivaptan	-12.42	35	791.50 pM	Contacts: I358, V365, M418, P421, DNA	-34.95 (±6.76)	Hbonds: DNA contacts: L349, Y360, R363, V365, L375, I414, K417, M418, P421, L422, DNA
	P-P	<b>plerixafor</b>	-23.84	54	3.37 aM	Hbonds: Y360, DNA Contacts: R114, E150, S290, K357, I359 Hbonds: E133, K145, E355, E364	-75.23 (±11.93)	Hbonds: DNA Contacts: E174, S290, E364 Hbonds: D131, E133, D354, E355
		<b>tetraethylenepentamine</b>	-15.57	34	3.86 pM	Contacts: S291, S352 Hbonds: E133, K145, E355, E364	-52.48 (±12.59)	Hbonds: W173, E174, E175, P176, E177 (ligand is leaving the complex)
		<b>trientine</b>	-13.84	51	71.87 pM	Contacts: E133, K145, E364 Hbonds: E355	-44.38 (±12.42)	Contacts: E133 Hbonds: D131, E132, E355, E364
		<b>ribociclib</b>	-12.7	25	487.80 pM	Contacts: R114, F115, P130, K145, K147, T153, H154, R157 Hbonds: W116, E133, E150, E364	-24.23 (±5.23)	Contacts: A111, L113, F115, P130, F149, E150, T153 Hbonds: E110, R114
	B ext	<b>dihydroergotoxine</b>	-12.39	52	827.20 pM	Contacts: D264, I265, K288, Y439, K441, P442, A443, E446 Hbonds: H289, E445	-6.65 (±9.33)	Contacts: M263, D264, I265, R266, N267, K288 (ligand is leaving the complex)
HP1043	P-P	<b>plerixafor</b>	-17.65	35	116.01 fM	Contacts: K286, F310, Q312, G313 Hbonds: E175, E287, E311, A314, D315, D354	-84.34 (±18.64)	Contacts: A111, L113, E174, P176, K286, T350, S352 Hbonds: E175, E177, E287, S290, D354
		<b>tetraethylenepentamine</b>	-12.86	20	373.37 pM	Hbonds: E174, E175, E311	-58.09 (±13.58)	Contacts: A111 Hbonds: E174, E175, E287, E311
	Bi	<b>oxcarbazepine</b>	-9.94	103	51.39 nM	Contacts: S341, V343, I344, I351, L375, T376, A379, R382 Hbonds: N342, R380	-22.92 (±3.00)	Contacts: V343, I344, I351, I358, L375, T376, A379, R380
		<b>temoporfin</b>	-12.23	146	1.09 nM	Contacts: E333, L336, F338, W339, N342, P353, E356, V367, T376, H381 Hbonds: R337, K368	-29.49 (±5.45)	Contacts: D92, M224, L336, F338, W339, P353, E356, T376
	A ext	<b>tetraethylenepentamine</b>	-13.68	21	94.13 pM	Hbonds: D170, W173, E174, E175	-49.66 (±12.8)	Hbonds: R337 Hbonds: E174, E175, E311
		<b>temoporfin</b>	-11.18	102	6.41 nM	Contacts: I121, E122, G124, D160, Q161,	-36.01 (±5.20)	Contacts: I121, E122, I123, G124, Q161, I162, M195, L199, I201, S202, T203,

(Continued on following page)

TABLE 2 | (Continued) HP1043 drug docking and MMGBSA.

Binding Site	Molecule name	Binding energy (kcal/mol)	# cluster	Ki	Docking interaction residues (LIGPLOT)	MMGBSA P-L (kcal/mol) (SD)	MD interaction residues (Ligplot)
					I201, T203, F214, Y216, P217, C221 Hbonds: I162, S202		F214, Y216, P217, P219, A220, E223 Hbonds: C215

P-DNA = HP1043–DNA interface; P–P = HP1043 dimer cleft (A chain–B chain); Bi = B chain domain interface; A ext = external surface of the A chain; B ext = external surface of the B chain; contacts = hydrophobic contacts; Hbonds = hydrogen bonds; in bold: experimentally tested molecules.



HP1043–tetraethylenepentamine\_Aext complex displayed reduced flexibility all along the structure, compared to the other complexes. Analyzing HP1043 complexes bound to DNA, the residue flexibility was generally increased compared to the unbound complex. Above all, HP1043\_DNA bound to the dihydroergotoxine molecule, as previously mentioned, displayed the unbinding of C-terminal domain from the DNA molecule. Differences in local flexibility can be evidenced for residues 188–200 (A chain) (yellow line in **Supplementary Figure S2**), matching to helix H8 involved in DNA binding, belonging to complex HP1043\_DNA bound to hexafluronium, ergotamine, and tetraethylenepentamine. Residues 389–390 (166–168 B chain) corresponding to the domain interface loop (blue arrow in **Supplementary Figure S2**), showed higher flexibility in complexes bound to dihydroergotoxine, hexafluronium, ponatinib, and plerixafor molecules. Finally, B chain helix H8 (gray line in **Supplementary Figure S2**) displayed

increased flexibility in complexes HP1043\_DNA–dihydroergotoxine, HP1043\_DNA–hexafluronium, and HP1043\_DNA–ponatinib.

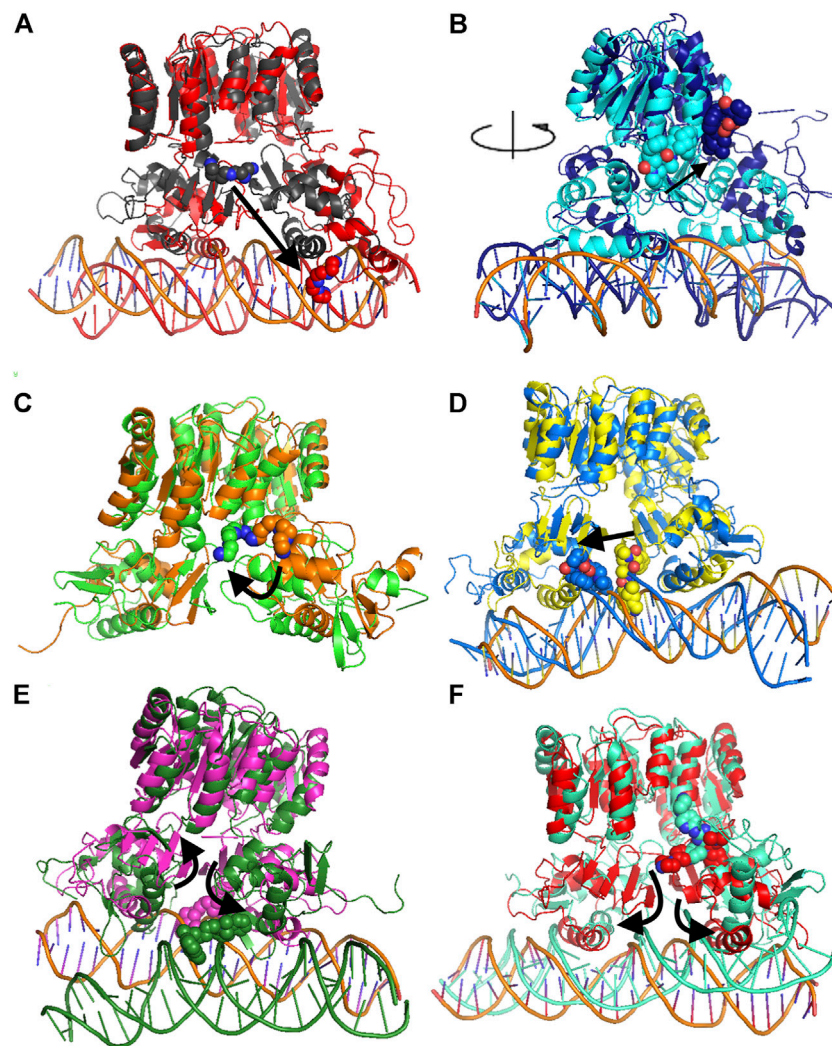
3.1.3 Drug Binding

The MMGBSA approach was applied to estimate the binding energy of drug ligands to the HP1043 transcription factor with results reported in **Table 2**. Binding free energy values ranged between –6.65 and –84.34 kcal/mol. Ligands bound at the protein–DNA interface (molecules in P-DNA–binding site in **Table 2**) displayed a lower affinity (–35.27 to –10.56 kcal/mol) compared to drugs localized at the P–P binding site (–75.23 to –24.23), due to fewer contacts/hydrogen bonds (listed in “MD interaction residues” of **Table 2**). Moreover, these drugs principally bound the DNA molecule, showing a low specificity for the HP1043 transcription factor. Considering the drugs bound only to the transcription factor, two left the complex during the simulation (tetraethylenepentamine and dihydroergotoxine), and thus, their binding energy estimation cannot be considered a reliable value. Among the others, plerixafor and trientine showed estimated binding energy of –75.23 and –44.38 kcal/mol, respectively. Drugs evaluated on the HP1043 free complex presented similar conditions, and molecules bound to a single domain or on the external surface displayed lower affinity, except for tetraethylenepentamine. Instead, drugs bound at the dimer cleft (plerixafor and tetraethylenepentamine) showed an estimated binding energy of –84.34 and –58.09 kcal/mol, respectively, thus, resulting in a higher affinity for HP1043.

The binding cavities identified during docking simulations were partially maintained. From the interacting residue analysis, three main residue groups were identified: Tyr360 (Tyr137 B), Glu364 (Glu141 B), Val365 (Val142 B), Phe372 (Phe149 B), and Thr376 (Thr153 B) at the B chain–DNA interface (red in **Figure 3** and double-underlined in **Table 2**); Glu133, Lys145, Glu174, Glu175, and Glu355 (Glu132 B) at the dimer cleft (P–P) involving the residues of DNA-binding domain of both chains (green in **Figure 3** and single underlined in **Table 2**); and Ala111, Glu311 (Glu88 B) at the dimerization domain interface (yellow in **Figure 3** and dotted-underlined in **Table 2**). All the listed residues are involved both in hydrophobic contacts and hydrogen bonds with analyzed drugs (see “MD interaction residues” in **Table 2** for details).

3.1.4 DNA Binding

We estimated the binding energy between the DNA molecule and HP1043 transcription factor bound to tested drugs with values reported in **Table 3**. All complexes displayed remarkable negative binding energy, except for HP1043\_DNA bound to hexafluronium



**FIGURE 4 |** Comparison of MD representative conformation and docking pose. HP1043\_DNA–tetraethylenepentamine (A), HP1043\_DNA–dihydroergotamine (B), HP1043–tetraethylenepentamine (C), HP1043\_DNA–ergotamine (D), HP1043\_DNA–hexafluronium (E), and HP1043\_DNA–riboiclib (F). Ligand and domain movements are evidenced by black arrows. Docking conformation are in gray, light blue, orange, yellow, pink, and light red; representative MD conformation are colored as DNA in red, dark blue, green, blue, dark green, and aquamarine.

ligand, and this value corresponds to an increased distance between the A chain, or B chain, and the DNA molecule, evidenced in **Figure 6**. Also, complex HP1043\_DNA–dihydroergotamine showed no binding of the A chain to DNA (**Figure 6**). A distance increment between the B chain and the DNA molecule, but for partial simulation time, was also displayed by HP1043\_DNA–dihydroergotamine and HP1043\_DNA–trientine complexes, which presented a reduced affinity compared to the other complexes for DNA.

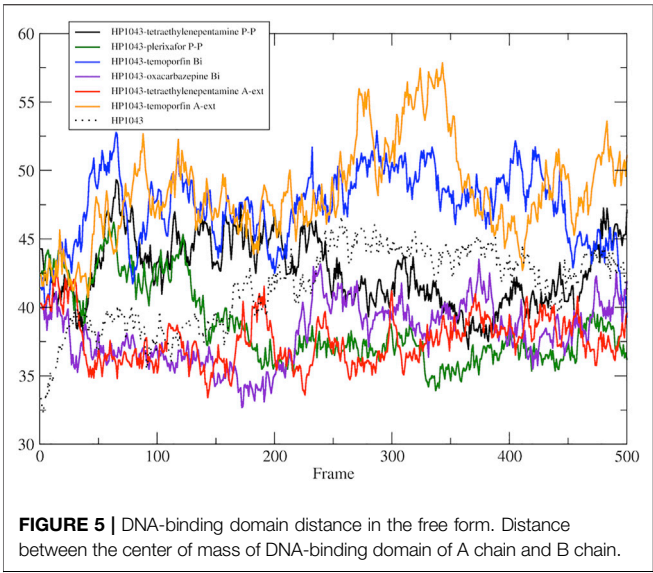
The distances between key residues for DNA binding (Zannoni et al., 2021) and DNA molecules were observed during trajectories, whose results are reported in **Table 3**. These data confirmed no DNA binding in complexes HP1043\_DNA–hexafluronium and HP1043\_DNA–dihydroergotamine since both chains showed a distance increase, and the displacement of the A chain in complex

HP1043\_DNA–ergotamine, and of the B chain in complex HP1043\_DNA–trientine.

### 3.3 *In Vitro* Inhibition of HP1043 DNA-Binding Activity

Previous studies have documented the consistent ability of HP1043 to bind specific sequences embedded in several promoters, proving its crucial role in the viability of *H. pylori* (Pellicciari et al., 2017; Zannoni et al., 2021). The EMSA is a versatile and sensitive tool for detecting protein–nucleic acid interaction and its inhibition. To determine *in vitro* whether the recombinant HP1043 protein retained its biological activity to bind the target promoter  $P_{hp1227}$ , we evaluated the shift in electrophoretic mobility of the DNA probe after protein binding on a polyacrylamide gel (see Materials and Methods





**Section 2.4).** EMSA analyses showed a decrease of the unbound DNA in a concentration-dependent manner (**Figure 7**). We chose a concentration of 4  $\mu$ M HP1043 monomer for subsequent tests on putative selected inhibitors, a non-completely saturating binding condition. Moreover, a 127-bp probe from the 16S gene was used as a non-specific probe for HP1043 binding in each reaction.

Among ligands previously selected by molecular dockings and dynamics, those bound to DNA molecules were excluded to prevent non-specific interactions. The seven ligands available for sale (plerixafor, tetraethylenepentamine, trientine, ribociclib, dihydroergotoxine, oxcarbazepine, and temoporfin) were tested to evaluate inhibition properties through EMSA. In particular, the recombinant HP1043 protein was incubated with the validated target  $P_{hp1227}$  in the presence of decreasing concentrations of ligand (from 1 mM to 50  $\mu$ M). DNA-binding inhibition could indicate whether dynamic interactions between the compound and HP1043 hinder the formation of the protein-promoter complex. Since the compounds were diluted in either H<sub>2</sub>O or DMSO, negative control reactions were included in EMSA analysis, where equivalent volumes of solvent were added to the protein–DNA mixes replacing the ligands. In addition, we prepared a second set of negative controls to remove the hypothesis that the ligand itself might induce a mobility shift of DNA probes. For these reactions, 1  $\times$  binding buffer was used instead of the HP1043 dilution. The magnitude of the inhibitory effect was deduced from the optical density of the free DNA bands, and the ligands were considered inhibitors when capable of interfering with the protein-dependent specific shift of  $P_{hp1227}$ .

As shown in **Figure 8A**, the sharp inhibitory effect for temoporfin was detected even at 50  $\mu$ M (lane h, **Figure 8A**). Such concentrations correspond to a mole monomeric-HP1043: ligand ratio of 1:12. At higher concentrations (lane d, e, f, and g; **Figure 8A**), reaching a mole ratio of 1:250, temoporfin significantly reduced the mobility shift of the specific DNA probe. Also, trientine (**Figure 8B**) and tetraethylenepentamine (**Figure 8C**)

TABLE 3   HP1043-DNA-binding energy estimation (MMGBSA); distance of DNA key residues compared to HP1043_DNA.																		
Binding site	Molecule name	MMGBSA P-DNA (kcal/mol) (SD)	Helix $\alpha 8$						$\beta 11$ – $\beta 12$				Helix $\alpha 8$				$\beta 11$ – $\beta 12$	
			I188	N189	R192	Q193	D196	K197	T206	R208	I411	N412	R415	Q416	D419	K420	T429	R431
HP1043-DNA	P-DNA																	
	hexafluorinium	–9.95 ( $\pm 21.07$ )	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	ponatinib	–62.08 ( $\pm 20.49$ )	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	osimertinib	–52.86 ( $\pm 20.07$ )																
	tubocurarin	–63.51 ( $\pm 16.27$ )																
P-P	ergotamine	–62.99 ( $\pm 19.32$ )	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	conivaptan	–57.33 ( $\pm 20.8$ )																
	plerixafor	–47.72 ( $\pm 21.7$ )																
	tetraethylenepentamine	–56.29 ( $\pm 17.28$ )																
	trientine	–27.47 ( $\pm 26.67$ )																
B ext	ribociclib	–81.81 ( $\pm 17.17$ )	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	dihydroergotoxine	–30.64 ( $\pm 20.54$ )	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

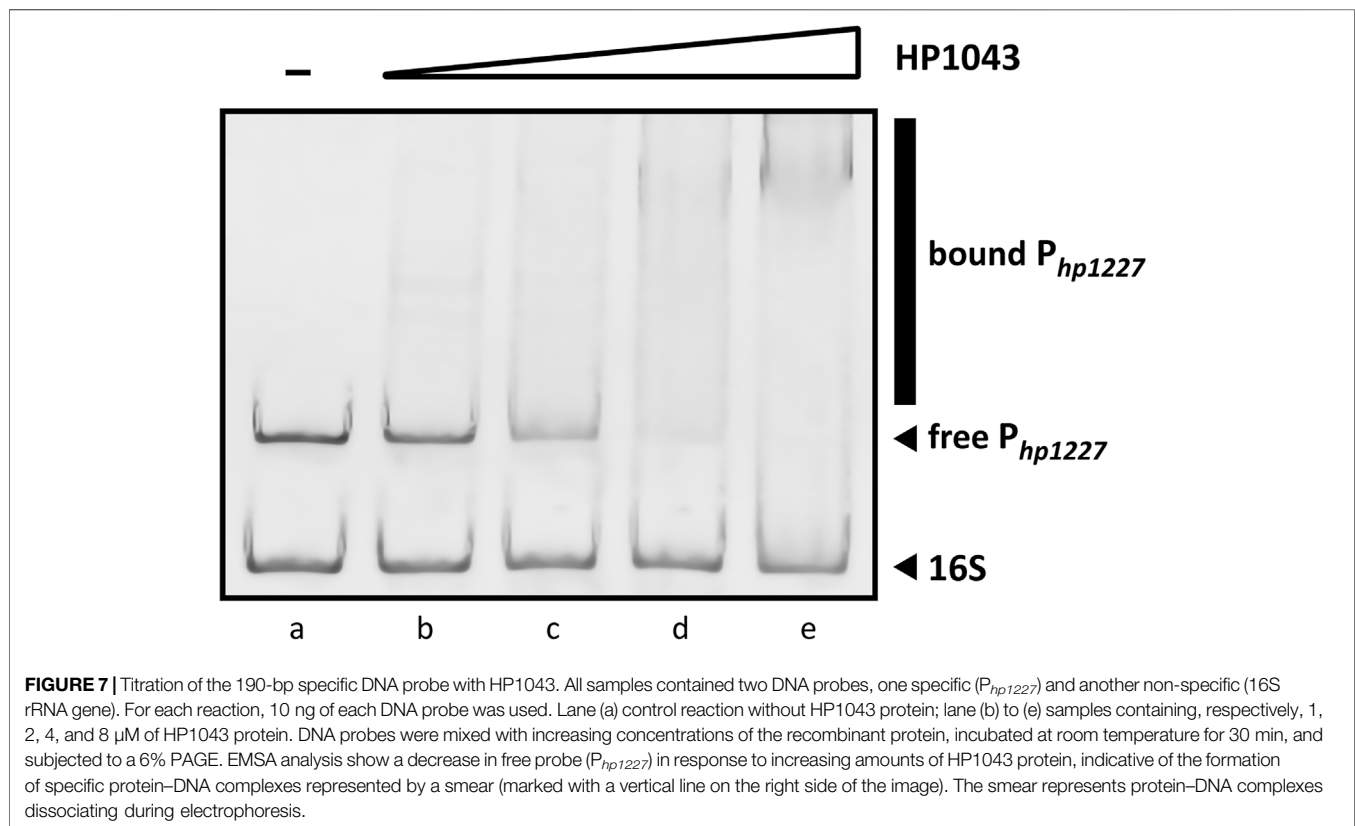
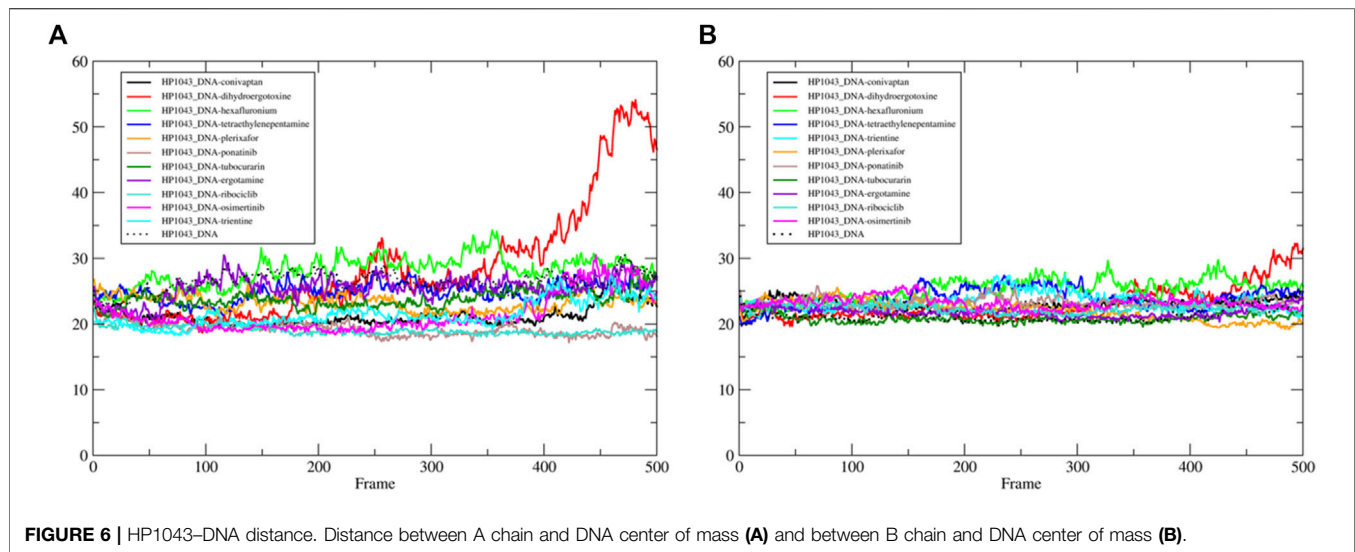
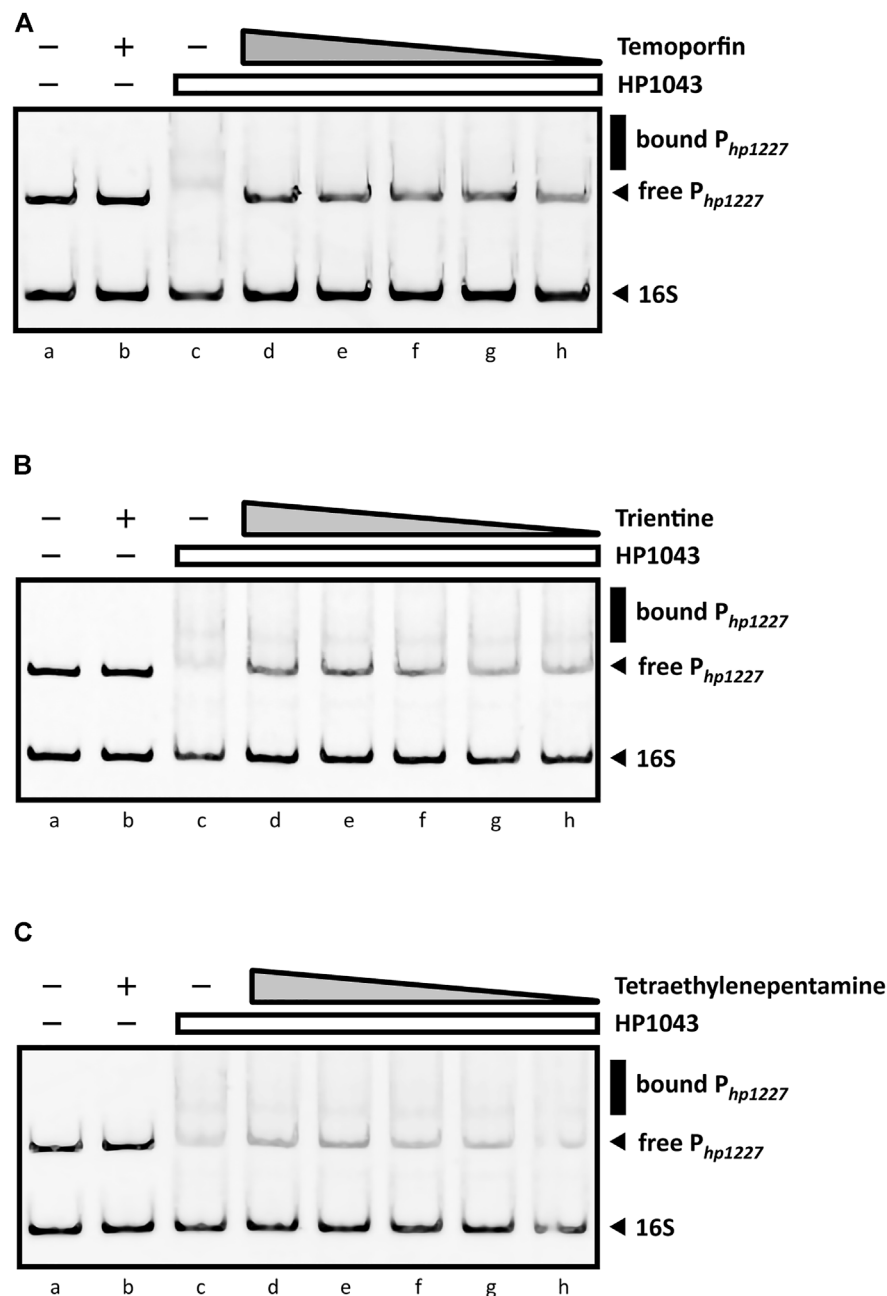


exhibit a less marked but appreciable loss of DNA-binding activity of HP1043. However, a faint band corresponding to the bound promoter DNA was still detected in both EMSA. Regardless, all three ligands were able to restore the electrophoretic mobility of free  $P_{hp1227}$  in a concentration-dependent manner. In contrast, four ligands did not show an appreciable reduction of the DNA–protein complex under the used experimental conditions (Supplementary Figure S3). In conclusion, three candidates (temoporfin, tetraethylenepentamine and trientine) showed a sharp inhibition

pattern on HP1043 binding, thus suggesting that these drugs can bind with a high affinity to HP1043.

## 4 DISCUSSION

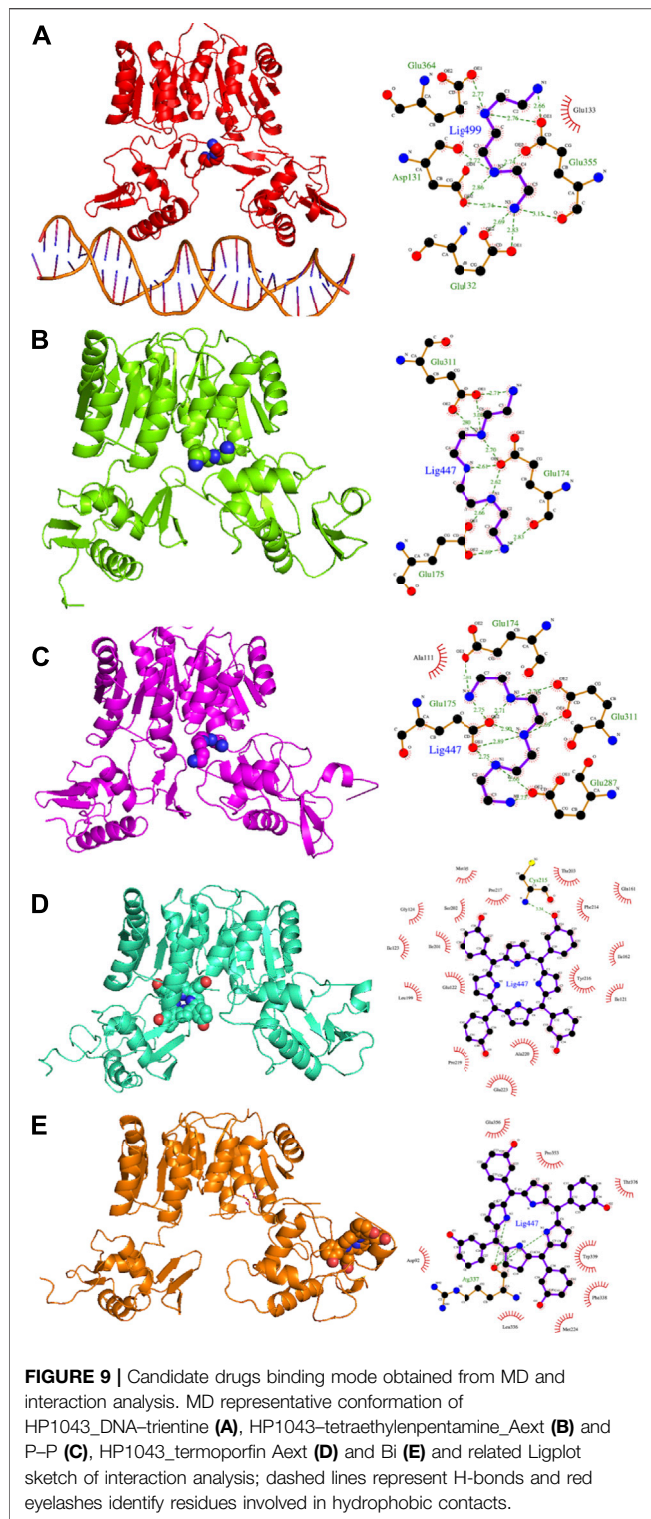
*H. pylori* colonizes the gastric mucosa of about 50% of the human population. It is strongly associated with the inflammation of the upper gastrointestinal tract, and it is related to several diseases



**FIGURE 8 |** EMSAs in the presence of DNA-binding inhibitors. All EMSAs show the same amount of specific ( $P_{hp1227}$ ) and non-specific (16S rRNA gene) DNA probes and the same sample order: lane (a) DNA probes control, protein- and ligand-free; (b) compound control at 1 mM (indicated by a "+") without the protein; (c) DNA-binding control in the presence of HP1043; lane (d) to (h) show samples containing a fixed amount of HP1043 monomeric protein (4  $\mu$ M) with a decreasing concentration of the ligand, respectively, 1, 0.5, 0.2, 0.1, and 0.05 mM; the absence of protein and compound is indicated by a "-"; the compound concentration is depicted as a gray triangle, while a white rectangle is used for the HP1043 protein fixed concentration. **(A)** Addition of temoporfin results in a reduced *in vitro* affinity of HP1043 protein for its target promoter region. Temoporfin restored the free DNA probe mobility and reduced the smear of the specific DNA probe at higher concentrations. For **(B)** trientine and **(C)** tetraethylenepentamine, similar inhibition effects were observed. Both chemicals induced a decreased DNA-binding activity for HP1043. However, a faint up-shifted band is still visible with the tested conditions. Symbols are as described in the legend in **Figure 7**.

including gastric cancer. Currently, the treatment for eradication of *H. pylori* infection mainly consists of triple standard therapy, including a proton pump inhibitor, amoxicillin and clarithromycin (Zagari et al., 2018; Roszczenko-Jasińska et al., 2020), frequently supplemented with bismuth salt, and substituted by tetracycline and metronidazole when

required. The main issue of this therapy is the antibiotic resistance to clarithromycin and metronidazole, affecting the treatment efficacy in about 70% of the cases (Su Young Kim et al., 2015; Hooi et al., 2017; Roszczenko-Jasińska et al., 2020). Known antibiotics target different bacterial enzymes and ribosomal subunits (Hu et al., 2017), but none



target a transcription factor. Recently, several novel molecules with an anti-*H. pylori* effect were proposed (Nishimori et al., 2007; Geng et al., 2009; Makobongo et al., 2014), also targeting the HP1043 transcription factor (González et al., 2019b; 2019a). The discovery of new therapies against *H. pylori* requires the identification and

validation of novel drug targets essential for *in vivo* growth or pathogenicity (González et al., 2018; Roszczenko-Jasińska et al., 2020). Transcription factors belong to the genes essential for the growth and virulence of pathogenic bacteria, as they act on more than one target gene. In *H. pylori*, one of these is HP1043 (Pellicciari et al., 2017). This allows us to consider HP1043 as a stable protein, as it is subjected to an evolutionary pressure that avoids the emergence of new mutants.

The development of new drugs is a time-consuming and expensive process with a high failure risk. Drug repositioning has become a strategy to reduce time and costs, proposing new applications for known drugs. This strategy was recently applied to identify treatments for several diseases such as tumors (Aydin et al., 2022; Persico et al., 2022; Petrosyan et al., 2022), cardiac diseases (Rouhana et al., 2021), and neurodegenerative diseases (Bhat et al., 2020; Agostini et al., 2021).

In the present study, we propose a repositioning of approved drugs for HP1043 by applying the VHTS protocol to select molecules with a high affinity for this protein both in the free form and bound to the DNA. The binding mode of top results was deeply evaluated by molecular dynamics and experimentally tested for inhibition properties through EMSA. Presented results evidenced three promising drugs (binding mode obtained from molecular dynamics and interaction analysis are reported in Figure 9), displaying an appreciable impairment of DNA-binding activity of HP1043, excluding non-specific binding of DNA. Trientine was *in silico* analyzed as bound to the HP1043–DNA complex, and displayed a reduced affinity for DNA molecule, compared to the other tested complexes, just as an increased distance between the B chain DNA-binding domain and the DNA molecule. The other two drugs, namely, tetraethylenepentamine and temoporfin, were simulated and bound to HP1043 in two different conformations. The first one showed high affinity for the transcription factor without evidence on protein flexibility or conformation, while the latter induced the increase of chain distance, inducing conformational changes on the protein structure.

From the pharmacological viewpoint, trientine is a Cu chelating agent used in the treatment of Wilson's disease, and it is orally delivered and poorly absorbed from the gastrointestinal tract. Tetraethylenepentamine is an ethylenamine with metal-chelating properties, while temoporfin is a photosensitizer used in photodynamic therapy of tumor cells; it is intravenously administered and collected in tumor tissues. These molecules belong to different pharmaceutical categories, but all display a similar inhibitory activity on HP1043 DNA binding.

Starting from the presented results, these three compounds can be considered to propose new molecules for *H. pylori* treatment, after having tested directly on bacteria to assess experimental  $K_i$ , as well as assays to determine the minimal inhibitory concentration (MIC) and the minimal bactericidal concentration (MBC) must be performed.

## DATA AVAILABILITY STATEMENT

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



## AUTHOR CONTRIBUTIONS

FA, AZ, and EC performed the experimental analysis. DR designed and supervised the experimental analysis. VS conceived the study and critically revised the manuscript. FC designed and performed the computational analysis. FA, DR, and FC wrote the manuscript.

## FUNDING

This work was supported by funding of the Italian Ministry of University and Research (MUR), grant: PRIN 2020YXFSW5 to VS and FC.

## REFERENCES

- Agostini, F., Masato, A., Bubacco, L., and Bisaglia, M. (2021). Metformin Repurposing for Parkinson Disease Therapy: Opportunities and Challenges. *Ijms* 23, 398. doi:10.3390/IJMS23010398
- Aydin, B., Yildirim, E., Erdogan, O., Arga, K. Y., Yilmaz, B. K., Bozkurt, S. U., et al. (2022). Past, Present, and Future of Therapies for Pituitary Neuroendocrine Tumors: Need for Omics and Drug Repositioning Guidance. *OMICS: A J. Integr. Biol.* 26, 115–129. doi:10.1089/OMI.2021.0221
- Bhat, A., Dalvi, H., Jain, H., Rangaraj, N., Singh, S. B., and Srivastava, S. (2021). Perspective Insights of Repurposing the Pleiotropic Efficacy of Statins in Neurodegenerative Disorders: An Expository Appraisal. *Curr. Res. Pharmacol. Drug Discov.* 2, 100012. doi:10.1016/J.CRPHAR.2020.100012
- Cheatham, T. E., and Case, D. A. (2013). Twenty-five Years of Nucleic Acid Simulations. *Biopolymers* 99, 969–977. doi:10.1002/bip.22331
- Danielli, A., and Scarlato, V. (2010). Regulatory Circuits in *Helicobacter Pylori*: Network Motifs and Regulators Involved in Metal-dependent Responses. *FEMS Microbiol. Rev.* 34, 738–752. doi:10.1111/j.1574-6976.2010.00233.x
- Geng, B., Basarab, G., Comita-Prevoir, J., Gowravaram, M., Hill, P., Kiely, A., et al. (2009). Potent and Selective Inhibitors of *Helicobacter pylori* Glutamate Racemase (MurI): Pyridodiazepine Amines. *Bioorg. Med. Chem. Lett.* 19, 930–936. doi:10.1016/J.BMCL.2008.11.113
- González, A., Casado, J., Chueca, E., Salillas, S., Velázquez-Campoy, A., Espinosa Angarica, V., et al. (2019a2019). Repurposing Dihydropyridines for Treatment of *Helicobacter pylori* Infection. *Pharmaceutics* 11, 681. doi:10.3390/PHARMACEUTICS11120681
- González, A., Fillat, M. F., and Lanás, Á. (2018). Transcriptional Regulators: Valuable Targets for Novel Antibacterial Strategies. *Future Med. Chem.* 10, 541–560. doi:10.4155/fmc-2017-0181
- González, A., Salillas, S., Velázquez-Campoy, A., Espinosa Angarica, V., Fillat, M. F., Sancho, J., et al. (2019b). Identifying Potential Novel Drugs against *Helicobacter pylori* by Targeting the Essential Response Regulator HsrA. *Sci. Rep.* 9, 11294. doi:10.1038/s41598-019-47746-9
- Hooi, J. K. Y., Lai, W. Y., Ng, W. K., Suen, M. M. Y., Underwood, F. E., Tanyingoh, D., et al. (2017). Global Prevalence of *Helicobacter pylori* Infection: Systematic Review and Meta-Analysis. *Gastroenterology* 153, 420–429. doi:10.1053/j.gastro.2017.04.022
- Hu, Y., Zhu, Y., and Lu, N.-H. (2017). Novel and Effective Therapeutic Regimens for *Helicobacter Pylori* in an Era of Increasing Antibiotic Resistance. *Front. Cel. Infect. Microbiol.* 7, 1–20. doi:10.3389/fcimb.2017.00168
- Jenks, P. J., and Edwards, D. I. (2002). Metronidazole Resistance in *Helicobacter pylori*. *Int. J. Antimicrob. Agents* 19, 1–7. doi:10.1016/S0924-8579(01)00468-X
- Kim, S. Y., Choi, J.-W. C., Kim, S. Y., Choi, D. J., and Chung, J. W. (2015). Antibiotic Treatment for *Helicobacter Pylori*: Is the End Coming? *Wjgpt* 6, 183–198. doi:10.4292/wjgpt.v6.i4.183
- Laskowski, R. A., and Swindells, M. B. (2011). LigPlot+: Multiple Ligand-Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* 51, 2778–2786. doi:10.1021/ci200227u
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theor. Comput.* 11, 3696–3713. doi:10.1021/ACS.JCTC.5B00255
- Makobongo, M. O., Gilbreath, J. J., and Merrell, D. S. (2014). Nontraditional Therapies to Treat *Helicobacter pylori* Infection. *J. Microbiol.* 52, 259–272. doi:10.1007/S12275-014-3603-5
- Miller, B. R., Mcgee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012). MMPBSA . Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theor. Comput.* 8, 3314–3321. doi:10.1021/ct300418h
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* 30, 2785–2791. doi:10.1002/jcc.21256
- Nishimori, I., Minakuchi, T., Kohsaki, T., Onishi, S., Takeuchi, H., Vullo, D., et al. (2007). Carbonic Anhydrase Inhibitors: The  $\beta$ -carbonic Anhydrase from *Helicobacter pylori* Is a New Target for Sulfonamide and Sulfamate Inhibitors. *Bioorg. Med. Chem. Lett.* 17, 3585–3594. doi:10.1016/J.BMCL.2007.04.063
- Onufriev, A., Bashford, D., and Case, D. A. (2004). Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* 55, 383–394. doi:10.1002/prot.20033
- Pellicciari, S., Pinatel, E., Vannini, A., Peano, C., Puccio, S., De Bellis, G., et al. (2017). Insight into the Essential Role of the *Helicobacter pylori* HP1043 Orphan Response Regulator: Genome-wide Identification and Characterization of the DNA-Binding Sites. *Sci. Rep.* 7. doi:10.1038/srep41063
- Persico, M., Abbruzzese, C., Matteoni, S., Matarrese, P., Campana, A. M., Villani, V., et al. (2022). Tackling the Behavior of Cancer Cells: Molecular Bases for Repurposing Antipsychotic Drugs in the Treatment of Glioblastoma. *Cells* 11, 263. doi:10.3390/CELLS11020263
- Petrosyan, E., Fares, J., Cordero, A., Rashidi, A., Arrieta, V. A., Kanojia, D., et al. (2022). Repurposing Autophagy Regulators in Brain Tumors. *Intl J. Cancer.* doi:10.1002/IJC.33965
- Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., et al. (2019). Drug Repurposing: Progress, Challenges and Recommendations. *Nat. Rev. Drug Discov.* 18, 41–58. doi:10.1038/nrd.2018.168
- Roncarati, D., Scarlato, V., and Vannini, A. (2022). Targeting of Regulators as a Promising Approach in the Search for Novel Antimicrobial Agents. *Microorganisms* 10, 185. doi:10.3390/microorganisms10010185
- Roszczenko-Jasińska, P., Wojtyś, M. I., and Jagusztyn-Krynica, E. K. (2020). *Helicobacter pylori* Treatment in the post-antibiotics Era-Searching for New Drug Targets. *Appl. Microbiol. Biotechnol.* 104, 9891–9905. doi:10.1007/s00253-020-10945-w
- Rouhana, S., Virsolvy, A., Fares, N., Richard, S., and Thireau, J. (2021). Ranolazine: An Old Drug with Emerging Potential; Lessons from Pre-clinical and Clinical Investigations for Possible Repositioning. *Pharmaceutics* 15, 31. doi:10.3390/PH15010031
- Salomon-Ferrer, R., Case, D. a., and Walker, R. C. (2013). An Overview of the Amber Biomolecular Simulation Package. *Wires Comput. Mol. Sci.* 3, 198–210. doi:10.1002/wcms.1121

## ACKNOWLEDGMENTS

The authors acknowledge Marco Moscatelli of the Institute for Biomedical Technologies (CNR-ITB) for computational support.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.887564/full#supplementary-material>

- Schär, J., Sickmann, A., and Beier, D. (2005). Phosphorylation-independent Activity of Atypical Response Regulators of *Helicobacter pylori*. *J. Bacteriol.* 187, 3100–3109. doi:10.1128/JB.187.9.3100-3109.2005
- Seshasayee, A. S., Bertone, P., Fraser, G. M., and Luscombe, N. M. (2006). Transcriptional Regulatory Networks in Bacteria: from Input Signals to Output Responses. *Curr. Opin. Microbiol.* 9, 511–519. doi:10.1016/j.mib.2006.08.007
- Simonovic, M., and Volz, K. (2001). A Distinct Meta-Active Conformation in the 1.1-Å Resolution Structure of Wild-type ApoCheY. *J. Biol. Chem.* 276, 28637–28640. doi:10.1074/jbc.C100295200
- Sterling, T., and Irwin, J. J. (2015). ZINC 15 - Ligand Discovery for Everyone. *J. Chem. Inf. Model.* 55, 2324–2337. doi:10.1021/acs.jcim.5b00559
- Sugano, K., Tack, J., Kuipers, E. J., Graham, D. Y., El-Omar, E. M., Miura, S., et al. (2015). Kyoto Global Consensus Report on *Helicobacter Pylori* gastritis. *Gut* 64, 1353–1367. doi:10.1136/gutjnl-2015-309252
- Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., et al. (1997). The Complete Genome Sequence of the Gastric Pathogen *Helicobacter pylori*. *Nature* 388, 539–547. doi:10.1038/41483
- Vianna, J. S., Ramis, I. B., Ramos, D. F., Von Groll, A., and Silva, P. E. A. d. (2016). Drug Resistance in *Helicobacter Pylori*. *Arq. Gastroenterol.* 53, 215–223. doi:10.1590/S0004-28032016000400002
- World Health Organization (2017). Prioritization of Pathogens to Guide Discovery, Research and Development of New Antibiotics for Drug-Resistant Bacterial Infections, Including Tuberculosis. *WHO Rep.* 148.
- Zagari, R. M., Rabitti, S., Eusebi, L. H., and Bazzoli, F. (2018). Treatment of *Helicobacter pylori* Infection: A Clinical Practice Update. *Eur. J. Clin. Invest.* 48, e12857. doi:10.1111/ECI.12857
- Zannoni, A., Pellicciari, S., Musiani, F., Chiappori, F., Roncarati, D., and Scarlato, V. (2021). Definition of the Binding Architecture to a Target Promoter of HP1043, the Essential Master Regulator of *Helicobacter pylori*. *Ijms* 22, 7848. doi:10.3390/ijms22157848

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

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

Copyright © 2022 Antoniciello, Roncarati, Zannoni, Chiti, Scarlato and Chiappori. 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.



## OPEN ACCESS

## Edited by:

Marco Nardini,  
University of Milan, Italy

## Reviewed by:

Ramandeep Singh,  
Translational Health Science and  
Technology Institute (THSTI), India  
Anshu Bhardwaj,  
Institute of Microbial Technology  
(CSIR), India

## \*Correspondence:

Sherine E. Thomas  
set44@cam.ac.uk  
Tom L. Blundell  
tlb20@cam.ac.uk

## †Present addresses:

Sherine E. Thomas,  
Department of Pathology, University of  
Cambridge, Cambridge,  
United Kingdom  
William J. McCarthy,  
The Francis Crick Institute, London,  
United Kingdom  
Jamal El Bakali,  
Inserm, CHU Lille, UMR-S 1172-LINC-  
Lille Neuroscience & Cognition,  
University of Lille, Lille, France  
Michal Blaszczyk,  
Department of Medicine, Cambridge  
Institute of Therapeutic Immunology  
and Infectious Disease, University of  
Cambridge, Cambridge,  
United Kingdom

## Specialty section:

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

Received: 21 February 2022

Accepted: 09 May 2022

Published: 30 May 2022

## Citation:

Thomas SE, McCarthy WJ, El Bakali J,  
Brown KP, Kim SY, Blaszczyk M,  
Mendes V, Abell C, Floto RA,  
Coyne AG and Blundell TL (2022)  
Structural Characterization of  
*Mycobacterium abscessus*  
Phosphopantetheine Adenylyl  
Transferase Ligand Interactions:  
Implications for Fragment-Based  
Drug Design.  
Front. Mol. Biosci. 9:880432.  
doi: 10.3389/fmolb.2022.880432

# Structural Characterization of *Mycobacterium abscessus* Phosphopantetheine Adenylyl Transferase Ligand Interactions: Implications for Fragment-Based Drug Design

Sherine E. Thomas<sup>1\*†</sup>, William J. McCarthy<sup>2†</sup>, Jamal El Bakali<sup>2†</sup>, Karen P. Brown<sup>3</sup>,  
So Yeon Kim<sup>1</sup>, Michal Blaszczyk<sup>1†</sup>, Vitor Mendes<sup>1,3</sup>, Chris Abell<sup>2</sup>, R. Andres Floto<sup>3,4</sup>,  
Anthony G. Coyne<sup>2</sup> and Tom L. Blundell<sup>1\*</sup>

<sup>1</sup>Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom, <sup>2</sup>Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, United Kingdom, <sup>3</sup>MRC Laboratory of Molecular Biology, Molecular Immunity Unit, Department of Medicine, University of Cambridge, Cambridge, United Kingdom, <sup>4</sup>Cambridge Centre for Lung Infection, Royal Papworth Hospital, Cambridge, United Kingdom

Anti-microbial resistance is a rising global healthcare concern that needs urgent attention as growing number of infections become difficult to treat with the currently available antibiotics. This is particularly true for mycobacterial infections like tuberculosis and leprosy and those with emerging opportunistic pathogens such as *Mycobacterium abscessus*, where multi-drug resistance leads to increased healthcare cost and mortality. *M. abscessus* is a highly drug-resistant non-tuberculous mycobacterium which causes life-threatening infections in people with chronic lung conditions such as cystic fibrosis. In this study, we explore *M. abscessus* phosphopantetheine adenylyl transferase (PPAT), an enzyme involved in the biosynthesis of Coenzyme A, as a target for the development of new antibiotics. We provide structural insights into substrate and feedback inhibitor binding modes of *M. abscessus* PPAT, thereby setting the basis for further chemical exploration of the enzyme. We then utilize a multi-dimensional fragment screening approach involving biophysical and structural analysis, followed by evaluation of compounds from a previous fragment-based drug discovery campaign against *M. tuberculosis* PPAT ortholog. This allowed the identification of an early-stage lead molecule exhibiting low micro molar affinity against *M. abscessus* PPAT ( $K_d$   $3.2 \pm 0.8 \mu M$ ) and potential new ways to design inhibitors against this enzyme. The resulting crystal structures reveal striking conformational changes and closure of solvent channel of *M. abscessus* PPAT hexamer providing novel strategies of inhibition. The study thus

**Abbreviations:** PPAT, Phosphopantetheine adenylyl transferase; CoA, Coenzyme A; *Mab*, *Mycobacterium abscessus*; *Mtb*, *Mycobacterium tuberculosis*; ATP, Adenosine 5' triphosphate; PhP, 4'-phosphopantetheine; dpCoA, 3'-dephospho CoA; NTM, nontuberculous mycobacteria; AMPCPP, Adenosine 5'-[ $\alpha,\beta$ -methylene]triphosphate.

validates the ligandability of *M. abscessus* PPAT as an antibiotic target and identifies crucial starting points for structure-guided drug discovery against this bacterium.

**Keywords:** *Mycobacterium abscessus*, *Mycobacterium tuberculosis*, Coenzyme A pathway, CoaD/ PPAT, drug discovery, fragment-based, antibiotics

## INTRODUCTION

The emergence of highly drug-resistant non-tuberculous mycobacteria (NTM) in recent years has intensified the challenges to clinically manage chronic lung conditions such as cystic fibrosis (Nessar et al., 2012; Bryant et al., 2021). *Mycobacterium abscessus* is one of the most common NTMs found in cystic fibrosis (CF) related lung infections and being intrinsically resistant to most antibiotics, causes very high rates of lung function decline (Nessar et al., 2012; Stephenson et al., 2017). Although previously thought to be a benign environmental microbe, *M. abscessus* is increasingly observed as a cause of chronic lung and soft-tissues infections, usually in the context of lung diseases like CF and in immunocompromised individuals (Bar-On et al., 2015; Parkins and Floto, 2015). *M. abscessus* is highly challenging to treat due to uniform resistance to standard anti-tuberculosis drugs, in addition to most antibiotics. High virulence and resistance to chemotherapy of this bacterium is attributed to a combination of intrinsic and acquired resistance mechanisms (Hill et al., 2012). Current clinical management of *M. abscessus* involves prolonged therapy using antibiotic combinations such as amikacin, cefoxitin and macrolides which are poorly tolerated, often resulting in therapeutic failure (Yang et al., 2017). This NTM is also associated with life-threatening disseminated infection in lung transplant recipients (Hill et al., 2012). Thus, there is an urgent need for effective and less toxic drugs to treat *M. abscessus* infections.

Coenzyme A (CoA) is an essential cellular co-factor that acts as an important acyl-group carrier in all organisms. CoA plays a key role in mediating numerous biosynthetic, degradative, and metabolic pathways (Geerloff et al., 1999). The biosynthesis of CoA consists of a five-step reaction involving pantothenate (Vitamin B5), cysteine and ATP as starting substrates (Abiko, 1967). Phosphopantetheine adenyltransferase (PPAT or CoaD) catalyses the penultimate step in the biosynthesis of CoA in prokaryotes. The enzyme catalyses the reversible transfer of an adenylyl group from ATP to 4'-phosphopantetheine (PhP) to yield 3'-dephospho-CoA (dpCoA) and pyrophosphate. The product dpCoA in turn is phosphorylated by dephospho-CoA kinase (DPCK/CoaE) to generate the final product of the pathway, Coenzyme A (Robishaw and Neely, 1985; Izard and Geerloff, 1999) (Figure 1B).

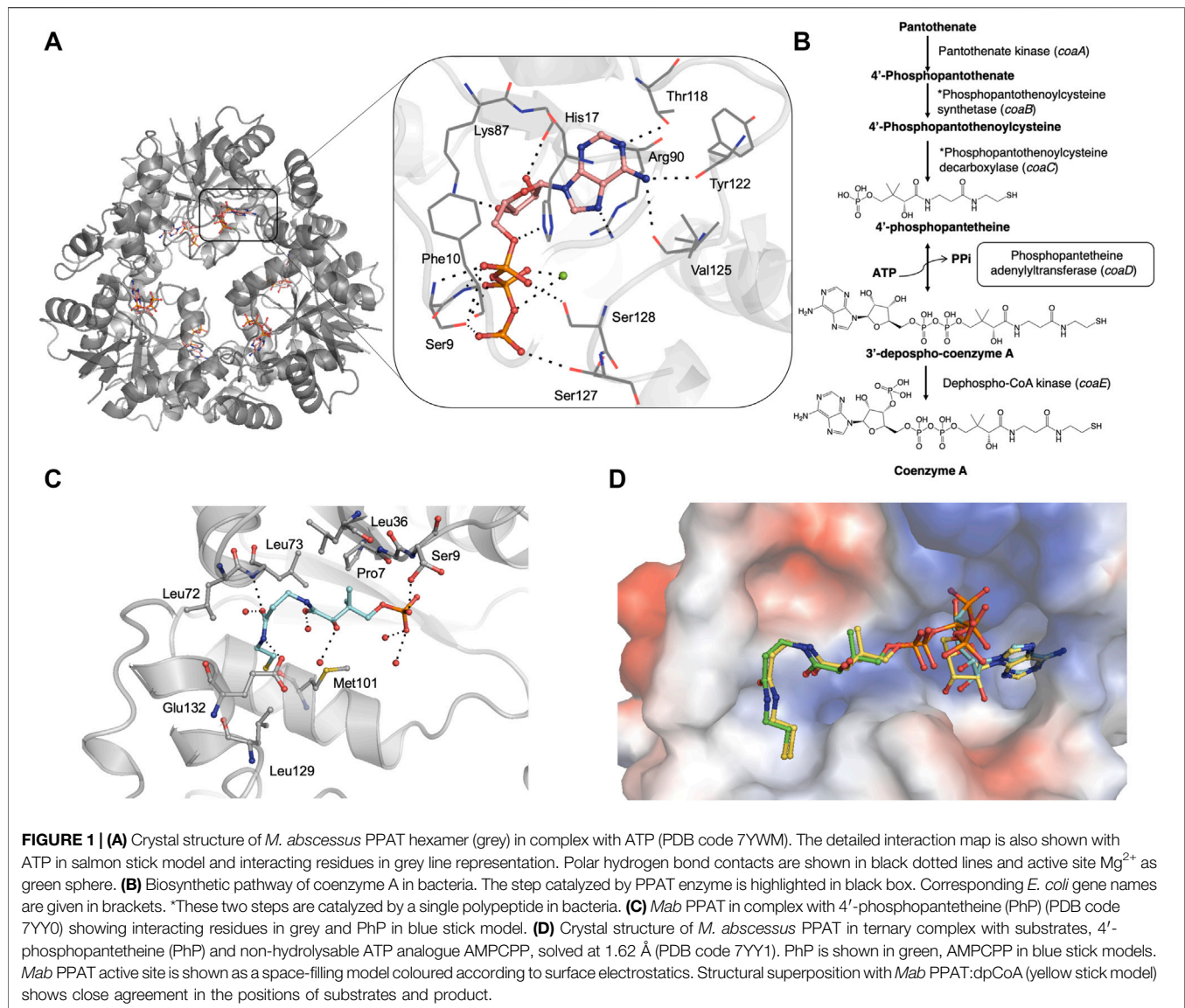
Enzymes in the CoA biosynthetic pathway have been recognised as attractive antibiotic targets (Cole et al., 1998). Previous studies on CoA pathway intermediates in *E. coli* show that pantothenate and 4'-phosphopantetheine (PhP) accumulate in the cell, suggesting an important rate-limiting role of PPAT in the pathway for regulating the cellular content of CoA (Jackowski and Rock, 1984; Gerdes et al., 2002; Wubben and Mesecar, 2011). Further studies on *M.*

*tuberculosis* show the essential role played by the *coaD* gene in mycobacterial growth *in vitro*, confirming the potential of PPAT as an antibiotic target (Ambady et al., 2012) (El Bakali et al., 2020). In higher eukaryotes, the final two steps of CoA biosynthesis are catalysed by a single bi-functional enzyme, CoA synthase, containing a PPAT like domain. The marked structural differences between bacterial and human PPAT domain architectures further facilitate target specific antibiotic discovery without inducing mechanism-based toxicity (Aghajanian and Worrall, 2002; Daugherty et al., 2002; Zhyvoloup et al., 2002).

Previous attempts to develop inhibitors targeting *E. coli* PPAT enzyme, using combinatorial synthesis approaches, were only partially successful as the resulting candidate compounds showed poor inhibitory activity in whole cell-based assays (Zhao et al., 2003). A high throughput screening (HTS) campaign against *E. coli* PPAT screened 750,000 compounds, resulting in the identification of 99 validated hits. Some of these hits were developed into compounds exhibiting low micro molar IC<sub>50</sub> against *E. coli* PPAT (Miller et al., 2010), however further structural biology evidence is needed to establish the ligand binding mode. Yet another HTS of the AstraZeneca compound library and subsequent structure-guided lead optimization, targeting PPAT from Gram-positive bacteria such as *S. pneumoniae* and *S. aureus*, led to the development of compounds that significantly reduced the bacterial burden both *in vitro* and in animal models of infection (De Jonge et al., 2013). However, these compounds were not progressed further into clinical development due to unfavourable toxicity and pharmacokinetic profiles. A recent fragment-based drug discovery campaign (Skepper et al., 2018) resulted in the development of several unique chemical scaffolds targeting the *E. coli* PPAT active site. Although some of these compounds exhibited nanomolar affinity *in vitro*, sufficient cellular potency could not be achieved against *E. coli* and other Gram-negative bacteria to warrant further chemical development of the series.

In this study, we report some of the first structural insights into the substrate and feedback inhibitor binding modes of PPAT from *M. abscessus*. We then explore the chemical space of *M. abscessus* PPAT using a multi-dimensional fragment screening approach involving biophysical and X-ray crystallographic analysis. The study is further extended by incorporating the knowledge gained from a previous fragment-based drug discovery campaign in our research groups, against *M. tuberculosis* PPAT ortholog (El Bakali et al., 2020), to identify potential early-stage lead compound targeting *M. abscessus* PPAT. Results from the study validate the ligandability of *M. abscessus* PPAT as an antibiotic target and facilitate the initiation of structure-guided drug discovery targeting this highly drug-resistant mycobacterium.





## MATERIALS AND METHODS

### PCR and Molecular Cloning

The *coaD* gene (MAB\_3259c) was amplified from *Mycobacterium abscessus* (ATCC 19977) genomic DNA using the following primers (Sigma):

Forward Primer: 5'-ATAGGATCCATGACGGGAGCGGTG  
TGCCC-3' and

Reverse Primer: 5'-ACCAAGCTTCTATTGTGCCTGGCC  
ACGCAGTTTC -3'

The purified PCR products and pET28a SUMO vector were subjected to restriction digestion with BamHI and HindIII restriction endonucleases (ThermoScientific). The ligation of digested insert and vector was done using T4 DNA ligase (New England Biolabs) by incubation at room

temperature for 10 min. The ligation product was transformed into *E. coli* DH5α competent cells by the heat-shock method, plated on LB agar-kanamycin plates and incubated at 37°C. Single colonies were randomly picked on the following day and inoculated in LB media with kanamycin (30 μg/ml) and grown overnight at 37°C. Plasmids from the resulting cultures were isolated, purified and the integrity of the clones was confirmed by sequencing (DNA Sequencing Facility, Department of Biochemistry, Cambridge).

### Expression and Purification of Full-Length *Mab* PPAT

*E. coli* BL21 (DE3) strain containing pET28a N-His SUMO-PPAT plasmid was grown overnight at 37°C in LB-media containing Kanamycin (30 μg/ml). This seed stage culture was

used to inoculate 6 L of 2x YT media with Kanamycin (30 µg/ml) until optical density ( $A_{600\text{nm}}$ ) reached 0.6. The expression of recombinant construct was induced by the addition of Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM and further allowed to grow at 18°C for 18 h. Cells were harvested by centrifugation at 4°C for 20 min at 5,000 g and the pellet was re-suspended in buffer A (50 mM Tris-HCl pH 8.0, 350 mM NaCl, 20 mM Imidazole). 10 µg/ml DNaseI, 5 mM  $\text{MgCl}_2$  and 3 protease inhibitor cocktail tablets (New England Biolabs) were added to the cell suspension. The cells were lysed by sonication (Branson). The lysate clarified by centrifugation at 4°C for 40 min at 25,568 g, was passed through a pre-equilibrated (with buffer A), 10 ml pre-packed Nickel-sepharose column (HiTrap IMAC FF, GE Healthcare). The column was washed with buffer A and the bound protein was eluted using buffer B (50 mM Tris-HCl pH 8.0, 350 mM NaCl and 500 mM Imidazole). Eluates from HiTrap IMAC column were pooled and Ulp1 Sumo protease was added in 1:100 mg (Ulp1: target protein) ratio and subjected to dialysis against 2 L of buffer C (50 mM Tris-HCl pH 8.0, 350 mM NaCl) overnight at 4°C. After overnight dialysis and cleavage of N-His tag, the protein was passed through a pre-equilibrated (buffer A) 5 ml HiTrap IMAC FF Nickel column (GE Healthcare). The flow-through was collected and the column was washed with buffer A and passed buffer B at a linear gradient of 150 ml of 100% buffer B. 4 ml fractions were collected and analyzed on a 15% SDS-PAGE. Flow-through and wash fractions from the above column were pooled and concentrated using a 30 kDa centrifugal concentrator (Amicon) and loaded onto a pre-equilibrated (with buffer D: 50 mM Tris-HCl pH 8.0, 150 mM NaCl) 120 ml Superdex200 16/600 column (GE Healthcare). 2 ml fractions were collected and analyzed on an SDS-PAGE gel. Fractions corresponding to pure PPAT protein were pooled and concentrated to 24 mg/ml, flash frozen in liquid nitrogen and stored at -80°C. The overall protein yield from a 6 L starting culture was approximately 100 mg. Identity of the purified protein was further confirmed by MALDI mass fingerprinting.

## Preliminary Fragment Screening

Thermal shift assays were carried out in a 96-well format with each well containing 25 µl of reaction mixture of 10 µM PPAT protein in buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl), 5 mM compound, 5% DMSO and 5x Sypro orange dye. Appropriate positive (Protein, DMSO and Coenzyme A) and negative (Protein, DMSO only) controls were also included. The measurements were performed in a Biorad-CFX connect thermal cycler using the following program: 25°C for 10 min followed by a linear increment of 0.5°C every 30 s to reach a final temperature of 95°C. The results were analyzed using Microsoft excel.

## Soaking of PPAT Native Crystals

Crystals for this experiment were grown at 19°C in 48-well sitting drop plates (Swiss CDI) in the following condition: 0.2 M Sodium bromide, 20%–24% PEG3350, 0.1 M Bis-Tris propane pH 6.5–7.5. Crystals were picked and allowed to soak in a 4 µl drop containing reservoir solution and 10 mM compound (in

DMSO), which was then equilibrated against 250 µl of the corresponding reservoir solution overnight at 19°C in 24-well hanging drop vapour diffusion set up.

## Co-Crystallization of PPAT Protein

2 mM final concentration of compound in DMSO was added to 20 mg/ml of PPAT protein, mixed and incubated for 2 h on ice. Crystals were grown in the following condition: 0.2 M Sodium bromide, 20%–24% PEG3350, 0.1 M Bis-Tris propane pH 6.5–7.5 or in sparse matrix screens: Wizard 1&2 (Molecular Dimensions), Wizard 3&4 (Molecular Dimensions), JCSG + Suite (Molecular Dimensions). The crystallization drops were set up at a protein to reservoir drop ratio of 0.3 µl: 0.3 µl, in 96-well (MRC2) sitting drop plate, using Mosquito crystallization robot (TTP labtech) and the drops were equilibrated against 70 µl of reservoir at 19°C.

## Data Collection and Processing

X-ray data sets for PPAT apo and ligand-bound crystals were collected on I04, I03, I04-1 or I24 beamlines at the Diamond Light Source in the United Kingdom. The crystals were flash-cooled in cryo-protectant containing precipitant solution and 25% Ethylene glycol. The data sets were collected using the rotation method at wavelength of 0.979 Å, Omega start: 0°, Omega Oscillation: 0.15°, total oscillation: 240°, total images: 2,400, exposure time: 0.05 s. The experimental intensities were processed to 1.5 Å and the diffraction images were processed using AutoPROC (Vonrhein et al., 2011) utilizing XDS (Kabsch, 2010) for indexing, integration, followed by POINTLESS (Evans, 2011), AIMLESS (Evans and Murshudov, 2013) and TRUNCATE (French, 1978) programs from CCP4 Suite (Winn et al., 2011) for data reduction, scaling and calculation of structure factor amplitudes and intensity statistics. All crystals belonged to space group C2 2 21 and consist of three protomers in the asymmetric unit.

## Crystal Structure Solution and Refinement

The *Mycobacterium abscessus* PPAT ligand-bound structures were solved by molecular replacement using PHASER (McCoy et al., 2007) with the atomic coordinates of *Mycobacterium abscessus* PPAT apo-structure (PDB entry: 5O06) as search model. Structure refinement was carried out using REFMAC (Murshudov et al., 2011) and PHENIX (Adams et al., 2010). The models obtained were manually re-built using COOT interactive graphics program (Emsley and Cowtan, 2004) and electron density maps were calculated with 2|Fo| - |Fc| and |Fo| - |Fc| coefficients. Position of ligands in the protein active site and water molecules were located in difference electron density maps.

## Isothermal Titration Calorimetry (ITC)

ITC experiments were done using an ITC200 (MicroCal) instrument. The *Mab* PPAT protein for ITC was dialysed overnight at 4°C in storage buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl) the same buffer was used for preparing the ligand solutions. The protein and ligands/fragments were used at a concentration of 72 µM and 1–5 mM, respectively. ITC recorded for ATP was done in buffer: 50 mM Hepes pH 8.0, 150 mM NaCl, 5 mM  $\text{MgCl}_2$ , at *Mab* PPAT concentration of

200  $\mu$ M and 3 mM ATP. Buffer components and other constituents like DMSO were kept constant in both protein and ligand solutions. Ligands to buffer titrations were subtracted in all cases and experiments were repeated at least twice. Data were analysed using the Origin software (OriginLab, Northampton, MA, United States).

## RESULTS

### *Mab* PPAT Binary Complex With Substrates ATP and 4'-Phosphopantetheine

To characterize the *Mab* PPAT ligand binding propensities, we determined crystal structures of *Mab* PPAT in complex with natural substrates ATP and 4'-phosphopantetheine (PhP). ATP occupies a positively charged pocket in the *Mab* PPAT substrate binding groove with the adenine ring making hydrogen-bond contacts with side chain hydroxyl of Thr118 (Figure 1A). The catalytic Arg90 at the base of the active site makes further H-bond and  $\pi$ -stacking interactions with the indole ring, while the adenine amino group is engaged in two additional hydrogen bonds with the backbone carbonyl groups of Val125 and Tyr122, respectively. The ATP ribose moiety mediates H-bonds to the side chain of Lys-87 and to three water molecules in the active site. The three phosphate groups of ATP are seen engaged in polar and H-bond interactions to the invariant catalytic residues Ser127, Ser128, Ser9, Phe10 and His-17, in addition to  $Mg^{2+}$  ion and several active site water molecules (Figure 1A).

In contrast to ATP, substrate 4'-phosphopantetheine (PhP) adopts a bent conformation in a predominantly hydrophobic part of the active site pocket. Here the 4'-phosphate group of PhP is engaged in a hydrogen-bond contact to the sidechain hydroxyl group of Ser9 and the terminal amide of the PhP  $\beta$ -mercaptoethylamine moiety makes further hydrogen bonds to the backbone nitrogen of Leu73 and side chain carboxylate of Glu132 at the edge of the active site. Major hydrophobic contacts in the region are mediated by Leu36, Pro7 and Leu73 at the top of the substrate site and Met101 and Leu129 at the base of the site. Figure 1C shows a detailed interaction map of the *Mab* PPAT-PhP complex.

### *Mab* PPAT Ternary Complex With PhP and AMPCPP: Insights Into Catalytic Cycle

Previous studies carried out on other bacterial PPAT orthologs have shown that the catalytic mechanism does not involve a covalent participation of residues in the functional site (Izard, 2002; Izard, 2003). Rather, the enzyme acts by providing the most favourable positioning of substrates to reduce the activation energy of the transition state. This allows the 4'-phosphate of 4'-phosphopantetheine (PhP) to undergo nucleophilic attack on the  $\alpha$ -phosphate of ATP in an in-line displacement mechanism. The activation energy barrier of this reaction is decreased by PPAT by way of orienting the ATP  $\beta$  and  $\gamma$ -phosphates, thereby stabilizing the penta-covalent transition state (Izard, 2002; Izard, 2003). Here we determined, for the first time, a ternary complex of PPAT with 4'-phosphopantetheine and non-hydrolysable ATP analogue, AMPCPP. A superposition of the substrate-bound ternary

structure of *Mab* PPAT with that of the reaction product dpCoA bound *Mab* PPAT structure, determined earlier (Thomas et al., 2017), shows close agreement in the positions of substrates and product (Figure 1D). 4'-phosphopantetheine (PhP) and the corresponding moiety in dpCoA overlaps perfectly in the PPAT active site. AMPCPP  $\beta$  and  $\gamma$  phosphates show greater flexibility in the corresponding crystal structure and therefore could be modelled in two different ways oriented away from the active site. The  $\alpha$ -phosphate that undergoes nucleophilic attack by PhP could be identified unambiguously and is seen in close proximity (1.7 Å) to the  $\alpha$ -phosphate of the reaction product dpCoA in the crystal structure, providing further evidence for the proposed catalytic mechanism (Figure 1D).

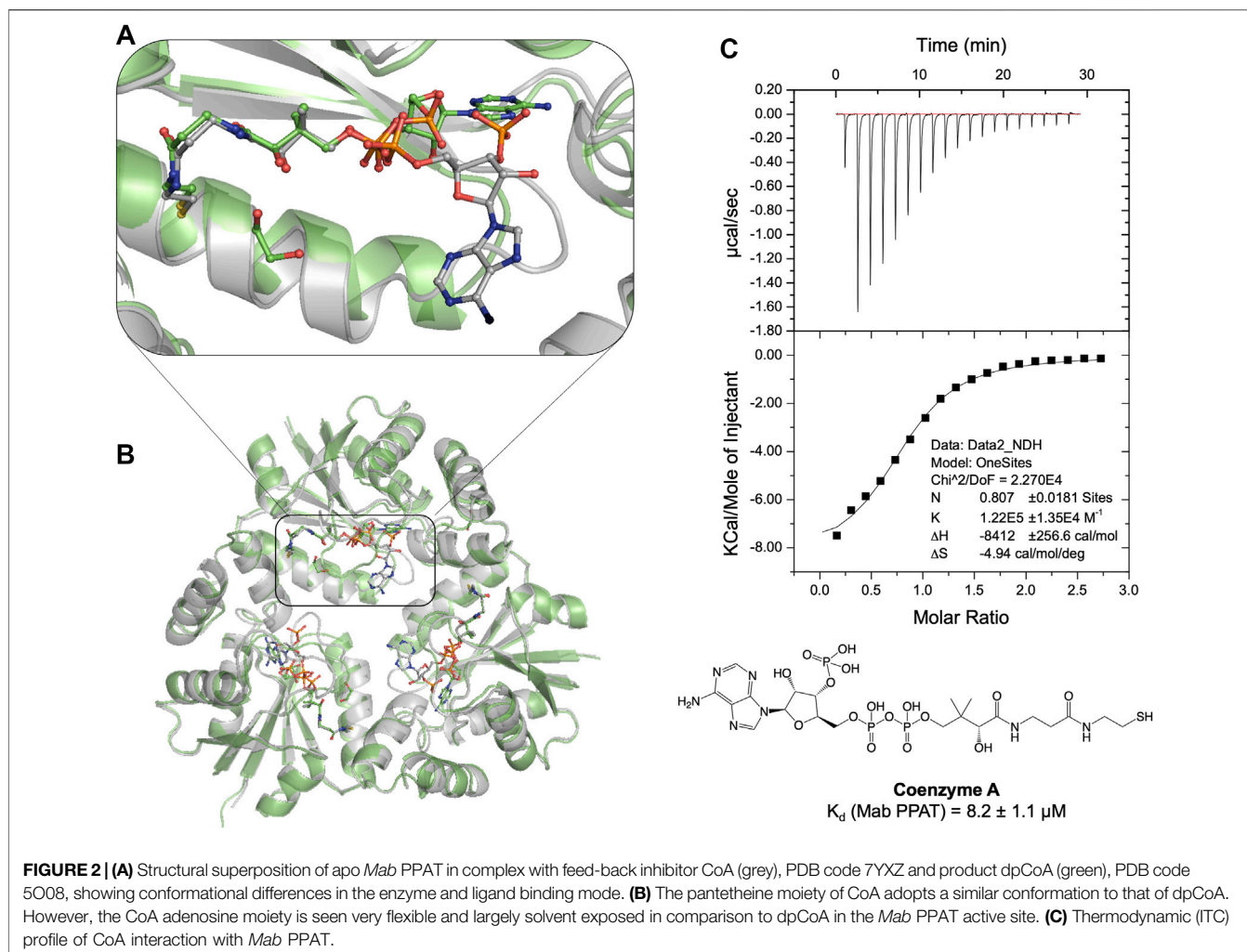
### *Mab* PPAT Complex With Feedback Inhibitor and Comparison With Product Binding

Previous studies on *E. coli* PPAT have shown that Coenzyme A (CoA), the end-product of the biosynthetic pathway, acts as feedback inhibitor of the PPAT enzyme catalyzed reaction (Wubben and Mesecar, 2011). To further understand the binding mode of CoA, we determined the crystal structure of *Mab* PPAT in complex with CoA. CoA and the enzyme product 3'-dephospho CoA (dpCoA) differ only by the presence of a 3-phosphate on the ribose ring. A structural superposition of the *Mab* PPAT CoA structure with that of dpCoA bound form determined, previously within our research group (Thomas et al., 2017), shows that the pantetheine moiety of CoA and that of dpCoA adopts a similar conformation in the *Mab* PPAT active site (Figures 2A,B). On the contrary, the adenosine moieties of CoA and dpCoA do not coincide in the *Mab* PPAT active site. The adenine ring along with the 3'-phosphate group of CoA is found largely exposed to the central solvent filled channel in *Mab* PPAT-CoA structure whereas in *Mab* PPAT-dpCoA, the adenine ring adopts a different position near the catalytic Arg90 (Figures 2A,B). Further, the adenylyl moiety of CoA exhibits high B-factors in comparison to that of the dpCoA complex, suggesting greater movements of this moiety in CoA-bound *Mab* PPAT. This could be attributed to the steric hindrance in the molecule imparted by the additional phosphate group of CoA. These observations further corroborate the earlier reports on PPAT-CoA structures from *E. coli* and *M. tuberculosis* (Morris and Izard, 2004; Timofeev et al., 2012), which show that the CoA adenine ring adopts a highly solvent-exposed conformation and/or exhibiting low electron-density.

### Fragment Screening Explores the Chemical Space of *Mab* PPAT Active Site

Having identified the ligand binding modes and characterized the corresponding interactions, a fragment-based drug discovery campaign was explored to design inhibitors of *Mab* PPAT. The initial fragment screening of an in-house library (consisting of 960 fragments) was performed using fluorescence-based thermal shift assays. The hits obtained, exceeding a thermal shift cut-off of 3 standard deviations from the negative control (the PPAT and DMSO only), were further validated by X-ray crystallography and quantified using isothermal





titration calorimetry (ITC). Preliminary results from this fragment screening effort have been described earlier (Thomas et al., 2017). A complete list of the validated fragment hits is shown in **Table 1** and (**Figure 3A** and **Supplementary Figure S1**).

The fragment hits were shown by X-ray crystallography to occupy one of four sub-pockets in the CoaD active site, arbitrarily numbered **I–IV**, (**Figure 3A**). Sub-pockets **I** and **II** correspond to the ATP adenine and ribose phosphate binding regions of *Mab* PPAT. Fragments binding to this region engage in key interactions seen in the binding of *Mab* PPAT with ATP such as H-bond contacts with Thr118, Try122, Ser127, His17, Ser9 and  $\pi$ -stacking interactions with Arg90, His17 and Phe10, in addition to novel binding interactions mediated by Gly16 and Thr14 (**Figure 3A** and **Supplementary Figure S1**).

Fragment interactions at sub-pockets **III** and **IV** coincide with the binding mode of *Mab* PPAT in complex with PhP. Therefore, crucial contacts at this part of the active site, such as those mediated by Leu72-73, Leu36, Pro7, Ser9 and Glu132, were recapitulated by these fragments. Sub-pocket **III** is seen in an occluded state in apo, and native ligand bound *Mab* PPAT structures. Whereas fragment binding results in an open conformation allowing further chemical exploration of this

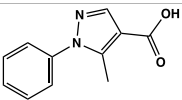
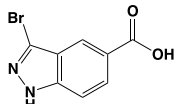
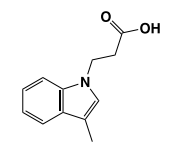
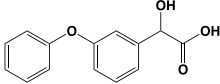
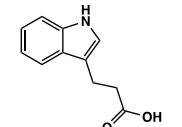
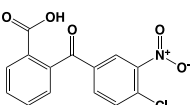
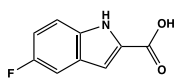
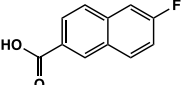
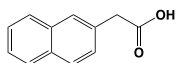
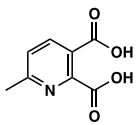
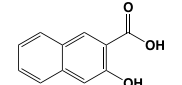
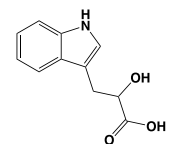
cryptic site. Interestingly, many fragments are seen mediating additional  $\pi$ -stacking contacts such as those involving Phe76 and Tyr136 in this active site region (**Figure 3A** and **Supplementary Figure S1**). Identification of fragment hits were further informed by analysis using our in-house hotspot mapping program (Radoux et al., 2016). As described earlier (Thomas et al., 2017), this analysis performed on *Mab* PPAT apo structure, provided additional insights on how some of these fragment hits can be chemically elaborated into drug-like compounds.

## Evaluation of Fragment-Based Lead Compound Designed to Inhibit *M. tuberculosis* PPAT

While the chemical elaboration and lead optimization of the above fragments are currently underway (McCarthy, W.J., Thomas, S.E., et al, unpublished), we compared the fragment hits identified on *Mab* PPAT with the corresponding fragments studied earlier on PPAT ortholog from *M. tuberculosis* (sequence identity 77%). A multiple sequence alignment of PPAT homologues from seven bacterial species, including *M. tuberculosis*, reveals conservation of key amino acids with *Mab*

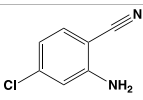
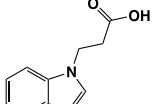
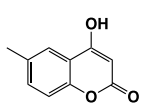
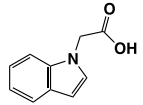


**TABLE 1 |** Summary of fragment hits identified on *M. abscessus* PPAT shown along with corresponding thermal shift values and binding sites. Fragment chemical structures are also shown.

Fragment	PDB code	$\Delta T_m$ (°C)	Site	Chemical structure
1	5O0A	+6	I	
2	5O0B	+12.5	I, II	
3	5O0C	+0.5	II	
4	5O0D	+2.5	III	
5	5O0F	+3.5	III	
6	5O0H	+4.5	III, IV	
7	7YY3	+3.5	III	
8	7YY4	+2.9	III	
9	7YY5	+2.3	III	
10	7YY6	+2.0	I	
11	7YY7	+2.0	III	
12	7YY8	+1.5	III	

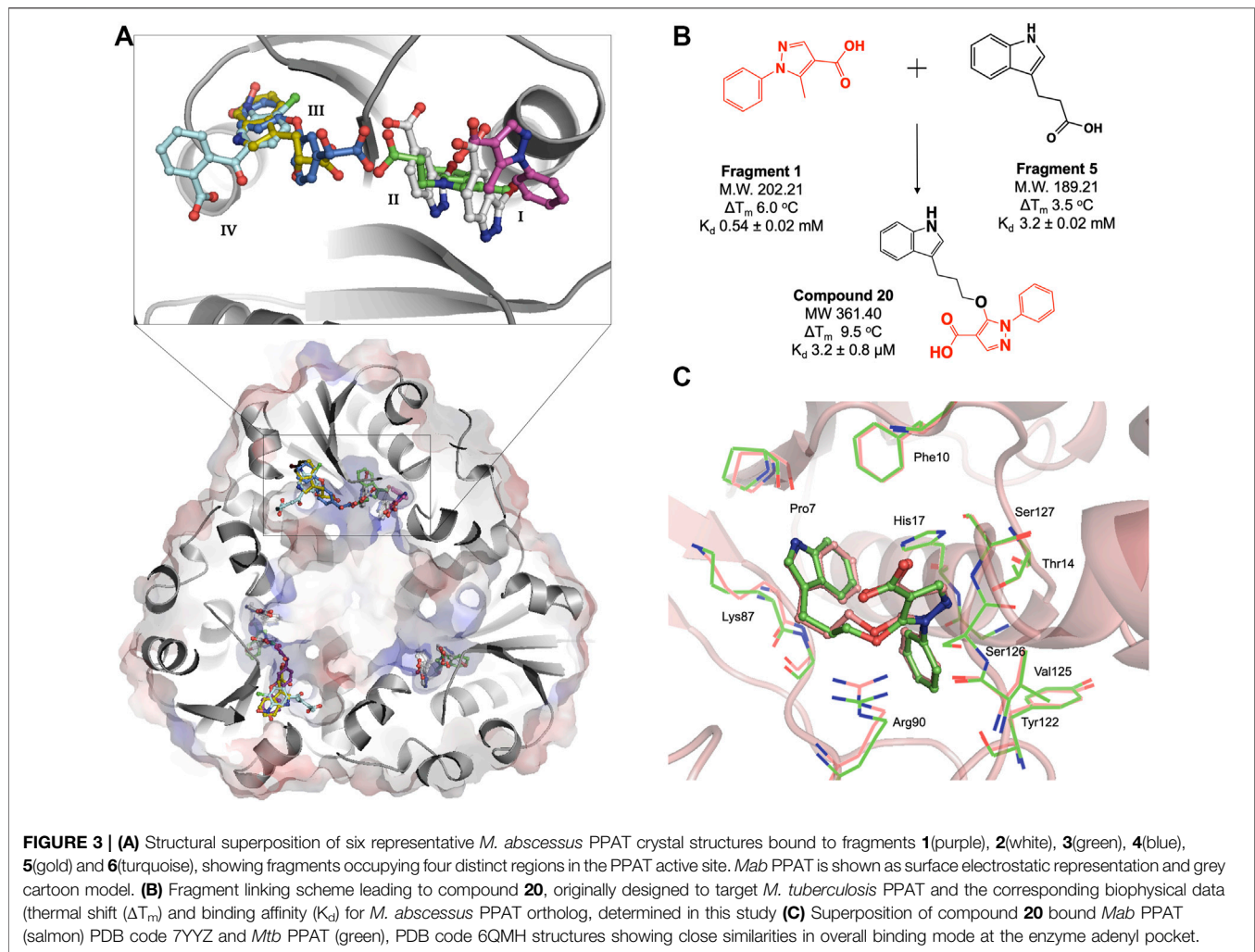
(Continued in next column)

**TABLE 1 |** (Continued) Summary of fragment hits identified on *M. abscessus* PPAT shown along with corresponding thermal shift values and binding sites. Fragment chemical structures are also shown.

Fragment	PDB code	$\Delta T_m$ (°C)	Site	Chemical structure
13	7YY9	+1.5	I	
14	7YYA	+1	II	
15	7YYB	+0.5	III	
16	7YYC	-0.5	II	

PPAT. These include the proposed catalytic residues such as His17, Lys41, Arg90, Ser9 and Ser127 (Geerlof et al., 1999; Izard, 2002), as well as the highly conserved T/HxGH motif, the signature of the nucleotidyltransferase  $\alpha/\beta$  phosphodiesterase superfamily (**Supplementary Figure S2**).

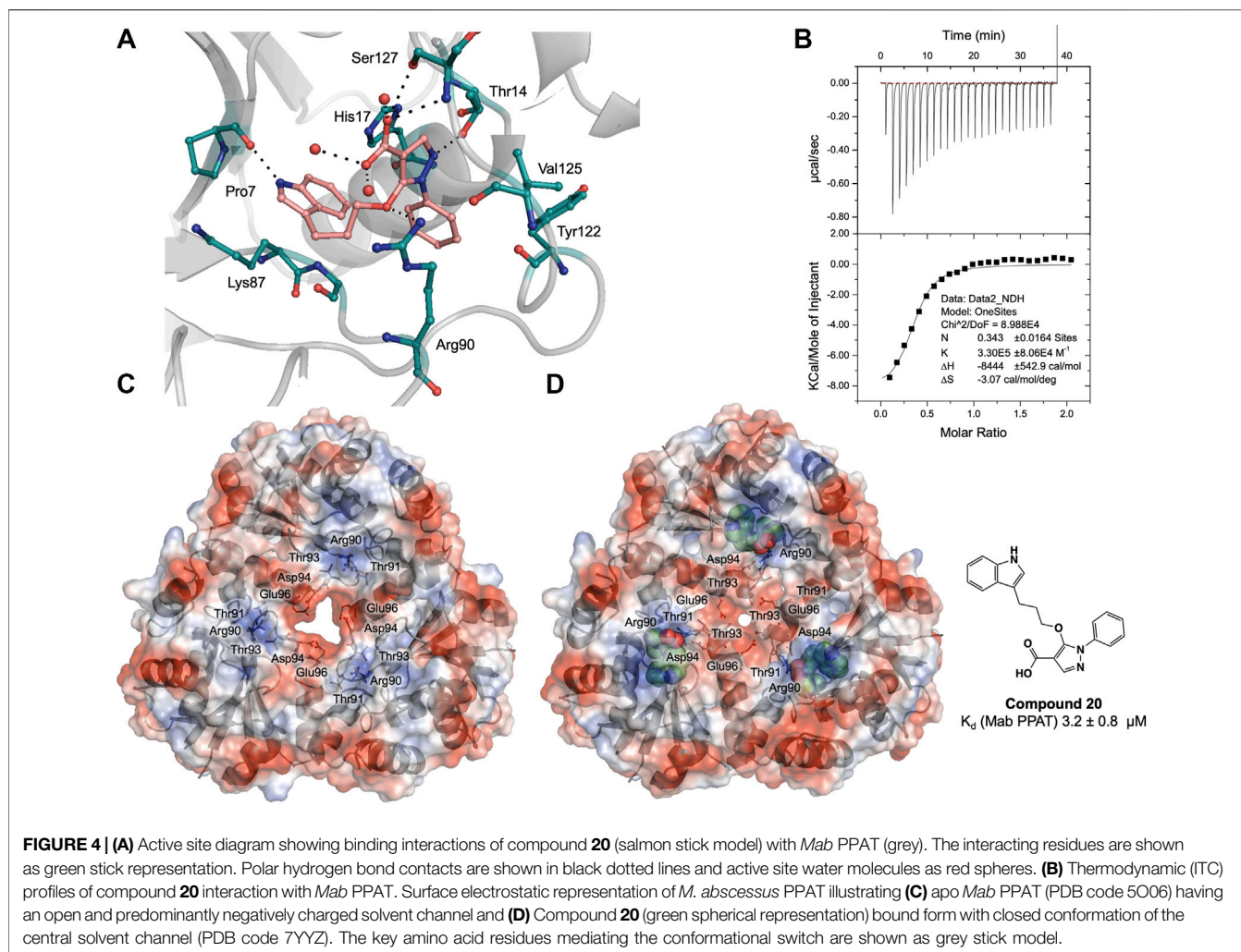
We therefore analysed the interactions of fragment hits, previously studied on *Mtb* PPAT, with respect to the corresponding *Mab* PPAT complexes. The binding modes of four different fragments were verified earlier on *Mtb* PPAT, using biophysical characterization and determination of crystal structures (El Bakali et al., 2020), of which the pyrazole fragment 3 recapitulates the binding mode adopted by the corresponding fragment (Fragment 1) on the sub-pocket I of *Mab* PPAT (**Figure 3A** and **Supplementary Figure S1**). This fragment engages all the key contacts seen in the *Mtb* PPAT structure, such as H-bond to Thr14, Arg90, His17 and Ser127. Similarly, the Benzophenone fragment identified on *Mtb* PPAT (El Bakali et al., 2020) is involved in hydrophobic interactions with Leu73 and Leu36 and polar contacts to Asn105 and Glu132 as seen in the equivalent fragment analogue (Fragment 6) on *Mab* PPAT sub-pockets III and IV (**Figure 3A** and **Supplementary Figure S1**). In contrast, unique binding modes were observed by fragments containing the indole moiety. For instance, in the case of *Mtb* PPAT, the indole fragments (11 and 12) bind solely at sub-pocket II of the active site chiefly mediated by active site His17 and Arg90 (El Bakali et al., 2020). While the corresponding fragments (14 and 5) in *Mab* PPAT occupy sub-pockets II and III respectively, engaging additional interactions through both hydrophobic interactions to Leu73 and Leu36 and polar contacts to Ser9 and Lys87 (**Figure 3A** and **Supplementary Figure S1**).



Having identified similarities and unique features in the fragment binding modes of *Mab* PPAT and *Mtb* PPAT, the question was asked whether the lead compound developed thereof against PPAT from *M. tuberculosis* would inhibit *Mab* PPAT, with comparable affinity. We therefore determined the crystal structure of lead compound **20** in complex with *Mab* PPAT and performed thermodynamic studies of the interaction (**Figures 4A,B**). Compound **20** was developed by linking of the pyrazole fragment **3** and the indole fragment **12** which bind at the *Mtb* PPAT sub-pockets **I** and **II**, respectively (El Bakali et al., 2020). As discussed earlier, although the corresponding pyrazole fragment **1** occupies sub-pocket **I** on *Mab* PPAT, the indole fragment **5** preferentially binds sub-pocket **III**, as do the indole fragments **7** and **12** on *Mab* PPAT. In contrast, three other indole fragments identified on *Mab* PPAT (**2**, **3** and **16**), preferentially bind at sub-pocket **II** mediating interactions with His17, Arg90, Ser127, analogous to fragment **12** on *Mtb* PPAT (**Figure 3A** and **Supplementary Figure S1**).

Interestingly, the crystal structure of *Mab* PPAT in complex with compound **20**, solved in this study, shows similarities in

binding mode and interactions, with that of *Mtb* PPAT at the ATP binding site (**Figures 3C, 4A**). The pyrazole part of the compound recapitulates interactions of the corresponding pyrazole fragment **1**, while the indole moiety, now positioned at sub-pocket **II**, makes H-bond contact with backbone carbonyl of Pro7 and cation- $\pi$  interaction to Lys87. This change in binding preference of the indole moiety is not surprising, as interactions with this region of the active site are recapitulated by other indole fragments that occupy sub-pocket **II** as well as ATP. Furthermore, the overall affinity of the pyrazole moiety (fragment **1**,  $K_d$   $0.54 \pm 0.02$  mM), quantified by ITC on *Mab* PPAT, is higher than indole fragment represented by fragment **5** ( $K_d$   $3.2 \pm 0.02$  mM). The interactions gained by fragment **5** in sub-site **II** such as those mediated by the backbone carbonyl of Pro7 seems to compensate for those lost in sub-site **III**, such as hydrophobic contacts to Leu73 and Leu36, thereby further facilitating the alternate binding mode of the indole moiety of compound **20** (**Figure 4A** and **Supplementary Figure S1**).



**FIGURE 4 | (A)** Active site diagram showing binding interactions of compound **20** (salmon stick model) with *Mab* PPAT (grey). The interacting residues are shown as green stick representation. Polar hydrogen bond contacts are shown in black dotted lines and active site water molecules as red spheres. **(B)** Thermodynamic (ITC) profiles of compound **20** interaction with *Mab* PPAT. Surface electrostatic representation of *M. abscessus* PPAT illustrating **(C)** apo *Mab* PPAT (PDB code 5O06) having an open and predominantly negatively charged solvent channel and **(D)** Compound **20** (green spherical representation) bound form with closed conformation of the central solvent channel (PDB code 7YYZ). The key amino acid residues mediating the conformational switch are shown as grey stick model.

## Compound 20 Results in Conformational Switching and Closure of Solvent Channel

As observed in PPAT from other bacterial species, *Mab* PPAT contains a solvent channel running through the centre of the hexamer along the 3-fold axis. Overlay of crystal structures of apo and compound-**20** bound to *Mab* PPAT shows that compound **20** results in large conformational change at the base of the PPAT active site and ordering of loop IV, connecting helix  $\alpha 4$  and strand  $\beta 4$ , to two complete  $\alpha$ -helical turns. Such a dramatic shift is mediated by a change in orientation of Arg90 side chain which now makes hydrogen-bond and  $\pi$ -stacking contacts with the ligand and is no longer available to interact with Thr-91 and Thr-93 at the edge of the channel. The Arg-90 further makes polar contacts with the Asp-94 of the loop region thereby lowering the negative charge of the solvent channel. These conformational rearrangements of residues 89–96 exposed to the channel lead to closure of the PPAT solvent channel when bound to compound **20** (Figure 4D).

Similar rearrangements can be observed in ATP and dpCoA bound PPAT from *Mab* (Supplementary Figure S3) and other bacteria (Izard, 2003; Timofeev et al., 2012). Whereas in apo,

PhP-bound and CoA-bound PPAT structures, the solvent channel adopts an open conformation and has a predominantly negative charge distribution from the acidic groups like Asp-94 and Glu-96 that are exposed to the channel (Figure 4C and Supplementary Figure S3). It was proposed earlier that these negatively charged residues lead to repulsion between protomers of the hexamer at the trimer interface playing an important role in subunit communication and in transitioning the enzyme from its symmetric to asymmetric forms. Additionally, the channel opening may also facilitate entry and binding of substrate to the active site (Morris and Izard, 2004; Timofeev et al., 2012). Therefore, the compound **20** bound *Mab* PPAT structure determined in this study presents an additional approach to enhance inhibition of the enzyme by blocking the central solvent channel, which may in turn prevent inter-subunit communication and access of substrates to the PPAT active site.

## Binding Affinity and Anti-Microbial Effect of Compound 20

Thermodynamic study of compound **20** binding to *Mab* PPAT shows a very significant improvement in binding affinity ( $K_d$  3.2 ±

0.8  $\mu\text{M}$ ) by over two orders of magnitude in comparison to starting fragments ( $K_d \geq 0.5 \text{ mM}$ ) (**Figure 3B**). An 80-fold improvement in the binding affinity is achieved by introduction of the linked indole ring. The affinity of compound **20** is comparable to that of *Mab* PPAT for feedback inhibitor CoA ( $K_d 8.2 \pm 1.1 \mu\text{M}$ ) (**Figure 2C**) and  $\approx 7$ -fold greater than affinity of *Mab* PPAT for the substrate ATP ( $K_d 20.1 \pm 3.3 \mu\text{M}$ ) (**Supplementary Figure S4**). Notably, the binding affinity of compound **20** on *Mab* PPAT is nearly 5-fold higher than that on *Mtb* PPAT ( $K_d 15 \pm 4 \mu\text{M}$ ) (El Bakali et al., 2020).

However, compound **20** afforded only low cellular activity when profiled on *M. abscessus* ATCC19977 whole-cell cultures using a luciferase-reporter based luminescence assay up to a concentration of 200  $\mu\text{M}$  in comparison to positive control, Amikacin (data not shown). NTMs, such as *M. abscessus*, are unique in their drug susceptibilities and ability to evade antibiotics through several powerful intrinsic resistance mechanisms. These include antibiotic-inactivating enzymes, target-modifying enzymes, drug efflux pumps and several additional genes acquired through horizontal gene transfer (Nessar et al., 2012; Nasiri et al., 2017). Further investigations are now underway to assess the lack of cellular activity of compound **20** and improve the physico-chemical properties of this early-stage lead compound.

## DISCUSSION

In this study, phosphopantetheine adenyltransferase (PPAT/CoaD) enzyme, involved in the coenzyme A biosynthesis pathway, was evaluated in the context of fragment-based discovery of new inhibitors targeting *Mycobacterium abscessus*. The structures of PPAT from *M. abscessus* in binary complex substrates 4'-phosphopantetheine (PhP) and ATP, feedback inhibitor Coenzyme A (CoA) and ternary complex with 4'-phosphopantetheine (PhP) and AMPCPP were solved at 1.5–1.8 Å resolution. These structures and corresponding thermodynamic studies provided detailed insights into the mode of ligand binding and catalytic process of *Mab* PPAT, facilitating the initiation of a structure-guided fragment-based drug discovery against this intrinsically drug-resistant *mycobacterium*.

Screening of a fragment library of 960 molecules by differential scanning fluorimetry and hit validation using X-ray crystallography resulted in the identification of 16 fragment hits. These fragment hits occupy four distinct sub-pockets in *Mab* PPAT, spanning the entire active site groove of the enzyme thus providing several potential starting points for drug discovery. While the chemical elaboration of the above fragments are on-going (McCarthy, W.J., Thomas, S.E., et al, unpublished), we investigated the compounds developed from a previous fragment-based drug discovery effort on *M. tuberculosis* PPAT ortholog (El Bakali et al., 2020), in the context of *M. abscessus*. The compound **20** studied was developed by linking of two fragments that occupy the PPAT adenyl binding regions. This elaborated compound afforded low micromolar affinity against *Mab* PPAT with a dissociation constant of  $3.2 \pm 0.8 \mu\text{M}$  and the resulting crystal structures reveal important conformational changes imparted by the compounds on *Mab*

PPAT. However, these compounds failed to show any significant inhibitory activity, when profiled on *M. abscessus* ATCC19977 whole-cell model. Further investigations are underway to assess whether the lack of cellular activity is related to permeability of the compounds across the thick cell envelope of mycobacteria or due to inactivation of compounds by the bacteria, to improve the physico-chemical properties of compound **20** and aid future drug discovery. The structural insights into substrate, feedback inhibitor and fragment binding modes described in this study can be utilized to further explore the chemical space of mycobacterial PPAT. In addition, the validated fragment hits identified in this study can also act as building blocks for new chemical entities targeting this essential bacterial enzyme. Overall, the study validates the ligandability of *M. abscessus* PPAT as an antibiotic target and identifies crucial starting points to facilitate structure-guided drug discovery against this highly-drug resistant *mycobacterium*.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**. Coordinates and structure factors related to this work have been deposited in the PDB with accession numbers 7YWM, 7YXZ, 7YY0, 7YY1, 7YY2, 7YY3, 7YY4, 7YY5, 7YY6, 7YY7, 7YY8, 7YY9, 7YYA, 7YYB and 7YYC.

## AUTHOR CONTRIBUTIONS

ST and SK performed molecular biology, expression, protein purification, crystallography, and fragment library screening. WM and JE designed, synthesized, and characterized the compounds. ST and WM performed the biophysical studies. KB performed the microbiological experiments on *M. abscessus*. MB provided structural biology support. TB, RF, AC, CA, and VM managed the project. ST wrote the manuscript.

## FUNDING

This work is funded by the Cystic Fibrosis Trust (Registered as a charity in England and Wales (1079049) and in Scotland (SC040196)) Strategic research consortium. ST is funded by the Cystic Fibrosis Trust (RG 70975) and Foundation Botnar (RG91317); VM is funded by the Bill and Melinda Gates Foundation SHORTEL-TB (OPP1158806) The project was funded by grants from the Foundation for the National Institutes of Health with support from the Bill and Melinda Gates Foundation (grant #OPP1024021 to (CA, TB, JE), and the EU FP7 (project HEALTH-F3-2011-260872, More Medicines for Tuberculosis (MM4TB) (CA and TB); TLB is funded by the Wellcome Trust (Wellcome Trust Investigator Award 200814\_Z\_16\_Z: RG83114).



## ACKNOWLEDGMENTS

The authors would like to thank the Diamond Light Source for beam-time (proposals mx9537, mx14043, mx18548) and the staff of beamlines I03, I02, I04, I04-1 and I24 for assistance with data collection.

## REFERENCES

- Abiko, Y. (1967). Investigations on Pantothenic Acid and its Related Compounds: IX. Biochemical Studies (4).<sup>\*</sup> Separation and Substrate Specificity of Pantothenate Kinase and Phosphopantothenoylcysteine Synthetase. *J. Biochem.* 61, 290–299. doi:10.1093/oxfordjournals.jbchem.a128547
- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. D. Biol. Cryst.* 66, 213–221. doi:10.1107/s0907444909052925
- Aghajanian, S., and Worrall, D. M. (2002). Identification and Characterization of the Gene Encoding the Human Phosphopantetheine Adenylyltransferase and Dephospho-CoA Kinase Bifunctional Enzyme (CoA Synthase). *Biochem. J.* 365, 13–18. doi:10.1042/bj20020569
- Ambady, A., Awasthy, D., Yadav, R., Basuthkar, S., Seshadri, K., and Sharma, U. (2012). Evaluation of CoA Biosynthesis Proteins of *Mycobacterium tuberculosis* as Potential Drug Targets. *Tuberculosis* 92, 521–528. doi:10.1016/j.tube.2012.08.001
- Bar-On, O., Mussaffi, H., Mei-Zahav, M., Prais, D., Steuer, G., Stafler, P., et al. (2015). Increasing Nontuberculous Mycobacteria Infection in Cystic Fibrosis. *J. Cyst. Fibros.* 14, 53–62. doi:10.1016/j.jcf.2014.05.008
- Bryant, J. M., Brown, K. P., Burbau, S., Everall, I., Belardinelli, J. M., Rodriguez-Rincon, D., et al. (2021). Stepwise Pathogenic Evolution of *Mycobacterium abscessus*. *Science* 372, eabb8699. doi:10.1126/science.abb8699
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., et al. (1998). Deciphering the Biology of *Mycobacterium tuberculosis* from the Complete Genome Sequence. *Nature* 393, 537–544. doi:10.1038/31159
- Daugherty, M., Polanuyer, B., Farrell, M., Scholle, M., Lykidis, A., de Crécy-Lagard, V., et al. (2002). Complete Reconstitution of the Human Coenzyme A Biosynthetic Pathway via Comparative Genomics. *J. Biol. Chem.* 277, 21431–21439. doi:10.1074/jbc.m201708200
- De Jonge, B. L. M., Walkup, G. K., Lahiri, S. D., Huynh, H., Neckermann, G., Utley, L., et al. (2013). Discovery of Inhibitors of 4'-Phosphopantetheine Adenylyltransferase (PPAT) to Validate PPAT as a Target for Antibacterial Therapy. *Antimicrob. Agents Chemother.* 57, 6005–6015. doi:10.1128/aac.01661-13
- El Bakali, J., Blaszczyk, M., Evans, J. C., Boland, J. A., McCarthy, W. J., Dias, M. V. B., et al. (2020). Chemical Validation of *Mycobacterium tuberculosis* Phosphopantetheine Adenylyltransferase Using Fragment Linking and CRISPR Interference. *bioRxiv*. doi:10.1101/2020.09.04.280388
- Emsley, P., and Cowtan, K. (2004). Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. D. Biol. Cryst.* 60, 2126–2132. doi:10.1107/s0907444904019158
- Evans, P. R. (2011). An Introduction to Data Reduction: Space-Group Determination, Scaling and Intensity Statistics. *Acta Crystallogr. D. Biol. Cryst.* 67, 282–292. doi:10.1107/s090744491003982x
- Evans, P. R., and Murshudov, G. N. (2013). How Good Are My Data and what Is the Resolution? *Acta Crystallogr. D. Biol. Cryst.* 69, 1204–1214. doi:10.1107/s0907444913000061
- French, S., and Wilson, K. (1978). On the Treatment of Negative Intensity Observations. *Acta Cryst. Sect. A* 34, 517–525. doi:10.1107/s0567739478001114
- Geerlof, A., Lewendon, A., and Shaw, W. V. (1999). Purification and Characterization of Phosphopantetheine Adenylyltransferase from *Escherichia coli*. *J. Biol. Chem.* 274, 27105–27111. doi:10.1074/jbc.274.38.27105
- Gerdes, S. Y., Scholle, M. D., D'souza, M., Bernal, A., Baev, M. V., Farrell, M., et al. (2002). From Genetic Footprinting to Antimicrobial Drug Targets: Examples in Cofactor Biosynthetic Pathways. *J. Bacteriol.* 184, 4555–4572. doi:10.1128/jb.184.16.4555-4572.2002
- Hill, U. G., Floto, R. A., and Haworth, C. S. (2012). Non-tuberculous Mycobacteria in Cystic Fibrosis. *J. R. Soc. Med.* 105(Suppl. 2), S14–S18. doi:10.1258/jrsm.2012.12s003
- Izard, T., and Geerlof, A. (1999). The Crystal Structure of a Novel Bacterial Adenylyltransferase Reveals Half of Sites Reactivity. *EMBO J.* 18, 2021–2030. doi:10.1093/emboj/18.8.2021
- Izard, T. (2003). A Novel Adenylate Binding Site Confers Phosphopantetheine Adenylyltransferase Interactions with Coenzyme A. *J. Bacteriol.* 185, 4074–4080. doi:10.1128/jb.185.14.4074-4080.2003
- Izard, T. (2002). The Crystal Structures of Phosphopantetheine Adenylyltransferase with Bound Substrates Reveal the Enzyme's Catalytic Mechanism 1 Edited by K. Nagai. *J. Mol. Biol.* 315, 487–495. doi:10.1006/jmbi.2001.5272
- Jackowski, S., and Rock, C. O. (1984). Metabolism of 4'-phosphopantetheine in *Escherichia coli*. *J. Bacteriol.* 158, 115–120. doi:10.1128/jb.158.1.115-120.1984
- Kabsch, W. (2010). XDS. *Acta Crystallogr. D. Biol. Cryst.* 66, 125–132. doi:10.1107/s0907444909047337
- Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic Software. *J. Appl. Cryst.* 40, 658–674. doi:10.1107/s0021889807021206
- Miller, J. R., Thanabal, V., Melnick, M. M., Lall, M., Donovan, C., Sarver, R. W., et al. (2010). The Use of Biochemical and Biophysical Tools for Triage of High-Throughput Screening Hits - A Case Study with *Escherichia coli* Phosphopantetheine Adenylyltransferase. *Chem. Biol. Drug Des.* 75, 444–454. doi:10.1111/j.1747-0285.2010.00957.x
- Morris, V. K., and Izard, T. (2004). Substrate-induced Asymmetry and Channel Closure Revealed by the Apoenzyme Structure of *Mycobacterium tuberculosis* phosphopantetheine Adenylyltransferase. *Protein Sci.* 13, 2547–2552. doi:10.1110/ps.04816904
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., et al. (2011). REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. D. Biol. Cryst.* 67, 355–367. doi:10.1107/s0907444911001314
- Nasiri, M. J., Haeili, M., Ghazi, M., Goudarzi, H., Pormohammad, A., Imani Fooladi, A. A., et al. (2017). New Insights in to the Intrinsic and Acquired Drug Resistance Mechanisms in Mycobacteria. *Front. Microbiol.* 8, 681. doi:10.3389/fmicb.2017.00681
- Nessar, R., Cambau, E., Reyat, J. M., Murray, A., and Gicquel, B. (2012). *Mycobacterium abscessus*: a New Antibiotic Nightmare. *J. Antimicrob. Chemother.* 67, 810–818. doi:10.1093/jac/ckr578
- Parkins, M. D., and Floto, R. A. (2015). Emerging Bacterial Pathogens and Changing Concepts of Bacterial Pathogenesis in Cystic Fibrosis. *J. Cyst. Fibros.* 14, 293–304. doi:10.1016/j.jcf.2015.03.012
- Radoux, C. J., Olsson, T. S. G., Pitt, W. R., Groom, C. R., and Blundell, T. L. (2016). Identifying Interactions that Determine Fragment Binding at Protein Hotspots. *J. Med. Chem.* 59, 4314–4325. doi:10.1021/acs.jmedchem.5b01980
- Robishaw, J. D., and Neely, J. R. (1985). Coenzyme A Metabolism. *Am J Physiol* 248, E1–E9. doi:10.1152/ajpendo.1985.248.1.e1
- Skepper, C. K., Moreau, R. J., Appleton, B. A., Benton, B. M., Drumm, J. E., 3rd, Feng, B. Y., et al. (2018). Discovery and Optimization of Phosphopantetheine Adenylyltransferase Inhibitors with Gram-Negative Antibacterial Activity. *J. Med. Chem.* 61, 3325–3349. doi:10.1021/acs.jmedchem.7b01861
- Stephenson, A. L., Stanojevic, S., Sykes, J., and Burgel, P.-R. (2017). The Changing Epidemiology and Demography of Cystic Fibrosis. *La Presse Médicale* 46, e87–e95. doi:10.1016/j.lpm.2017.04.012
- Thomas, S. E., Mendes, V., Kim, S. Y., Malhotra, S., Ochoa-Montano, B., Blaszczyk, M., et al. (2017). Structural Biology and the Design of New Therapeutics: From HIV and Cancer to Mycobacterial Infections. *J. Mol. Biol.* 429, 2677–2693. doi:10.1016/j.jmb.2017.06.014
- Timofeev, V., Smirnova, E., Chupova, L., Esipov, R., and Kuranova, I. (2012). X-ray Study of the Conformational Changes in the Molecule of Phosphopantetheine

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.880432/full#supplementary-material>

- Adenylyltransferase from *Mycobacterium Tuberculosis* during the Catalyzed Reaction. *Acta Crystallogr. D. Biol. Cryst.* 68, 1660–1670. doi:10.1107/s0907444912040206
- Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., et al. (2011). Data Processing and Analysis with the autoPROC toolbox. *Acta Crystallogr. D. Biol. Cryst.* 67, 293–302. doi:10.1107/s0907444911007773
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., et al. (2011). Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr. D. Biol. Cryst.* 67, 235–242. doi:10.1107/s0907444910045749
- Wubben, T., and Mesecar, A. D. (2011). Structure of *Mycobacterium Tuberculosis* phosphopantetheine Adenylyltransferase in Complex with the Feedback Inhibitor CoA Reveals Only One Active-Site Conformation. *Acta Cryst. Sect. F.* 67, 541–545. doi:10.1107/s1744309111010761
- Yang, B., Jhun, B. W., Moon, S. M., Lee, H., Park, H. Y., Jeon, K., et al. (2017). Clofazimine-Containing Regimen for the Treatment of *Mycobacterium Abscessus* Lung Disease. *Antimicrob. Agents Chemother.* 61, e02052-16. doi:10.1128/AAC.02052-16
- Zhao, L., Allanson, N. M., Thomson, S. P., Maclean, J. K. F., Barker, J. J., Primrose, W. U., et al. (2003). Inhibitors of Phosphopantetheine Adenylyltransferase. *Eur. J. Med. Chem.* 38, 345–349. doi:10.1016/s0223-5234(03)00047-3
- Zhyvoloup, A., Nemazanyy, I., Babich, A., Panasyuk, G., Pobigailo, N., Vudmaska, M., et al. (2002). Molecular Cloning of CoA Synthase. *J. Biol. Chem.* 277, 22107–22110. doi:10.1074/jbc.c200195200

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

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

Copyright © 2022 Thomas, McCarthy, El Bakali, Brown, Kim, Blaszczyk, Mendes, Abell, Floto, Coyne and Blundell. 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.



# Innovative Approach for a Classic Target: Fragment Screening on Trypanothione Reductase Reveals New Opportunities for Drug Design

Annarita Fiorillo<sup>1,2\*</sup>, Gianni Colotti<sup>1,2</sup>, Cécile Exertier<sup>1</sup>, Anastasia Liuzzi<sup>1</sup>, Francesca Seghetti<sup>3</sup>, Alessandra Salerno<sup>3</sup>, Jessica Caciolla<sup>3</sup> and Andrea Ilari<sup>1</sup>

<sup>1</sup>Institute of Molecular Biology and Pathology, Italian National Research Council, IBPM-CNR, Rome, Italy, <sup>2</sup>Department of Biochemical Sciences, Sapienza University, Rome, Italy, <sup>3</sup>Department of Pharmacy and Biotechnology (FaBIT), Alma Mater Studiorum, University of Bologna, Bologna, Italy

## OPEN ACCESS

### Edited by:

Pietro Roversi,  
Italian National Research Council, Italy

### Reviewed by:

Samuel Kojo Kwofie,  
University of Ghana, Ghana  
Antonio Jimenez-Ruiz,  
University of Alcalá, Spain

### \*Correspondence:

Annarita Fiorillo  
annarita.fiorillo@uniroma1.it

### Specialty section:

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

**Received:** 21 March 2022

**Accepted:** 19 May 2022

**Published:** 04 July 2022

### Citation:

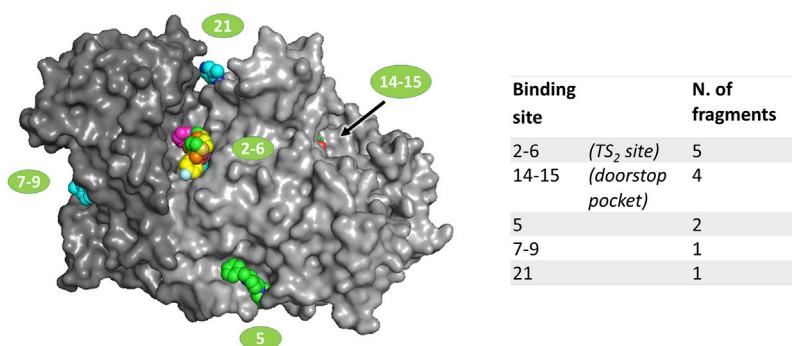
Fiorillo A, Colotti G, Exertier C, Liuzzi A, Seghetti F, Salerno A, Caciolla J and Ilari A (2022) Innovative Approach for a Classic Target: Fragment Screening on Trypanothione Reductase Reveals New Opportunities for Drug Design. *Front. Mol. Biosci.* 9:900882. doi: 10.3389/fmolb.2022.900882

Trypanothione reductase (TR) is a key factor in the redox homeostasis of trypanosomatid parasites, critical for survival in the hostile oxidative environment generated by the host to fight infection. TR is considered an attractive target for the development of new trypanocidal agents as it is essential for parasite survival but has no close homolog in humans. However, the high efficiency and turnover of TR challenging targets since only potent inhibitors, with nanomolar IC<sub>50</sub>, can significantly affect parasite redox state and viability. To aid the design of effective compounds targeting TR, we performed a fragment-based crystal screening at the Diamond Light Source XChem facility using a library optimized for follow-up synthesis steps. The experiment, allowing for testing over 300 compounds, resulted in the identification of 12 new ligands binding five different sites. Interestingly, the screening revealed the existence of an allosteric pocket close to the NADPH binding site, named the “doorstop pocket” since ligands binding at this site interfere with TR activity by hampering the “opening movement” needed to allow cofactor binding. The second remarkable site, known as the Z-site, identified by the screening, is located within the large trypanothione cavity but corresponds to a region not yet exploited for inhibition. The fragments binding to this site are close to each other and have some remarkable features making them ideal for follow-up optimization as a piperazine moiety in three out of five fragments.

**Keywords:** fragment screening, trypanosomatid infection, trypanothione reductase, rational drug discovery, protein crystallography

## 1 INTRODUCTION

Leishmaniasis and trypanosomiasis are vector-borne zoonoses transmitted by parasites from *Leishmania* and *Trypanosoma* spp., belonging to the Trypanosomatidae family. It is estimated that, overall, these parasites affect 20 million people worldwide, causing 100,000 deaths every year (WHO, 2022b; WHO, 2022c; WHO, 2022d). These infections are included in the list of neglected tropical diseases (WHO, 2022a) and are diffused in some of the poorest countries in the world. However, leishmaniasis is endemic in the whole Mediterranean basin, and climate change is



**FIGURE 1** | Location of fragments binding sites on TbTR surface. The two subunits of the dimer are colored in two different grey shades. Ligands are represented as spheres. Sites numbering is reported according to the output generated by PanDDA analysis; double numbers indicate symmetrically equivalent sites on the dimer.

promoting the spreading of the insect vector, expanding the risk of human exposure to regions previously out of the range of the disease.

Currently available treatments are very unsatisfying due to poor efficacy, high toxicity, and increasing resistance. Despite the urgent need, no new drugs reached the market in the last 30 years and the number of compounds in a clinical trial is very low (Field et al., 2017), indicating that more efforts are necessary for the development of new classes of therapeutics.

Target-based drug discovery is a powerful approach to the early stages of drug development. The success of this method strongly relies on the careful selection of the target that must fulfill some requirements in order to assure efficacy and selectivity of the hits (Frearson et al., 2007).

Parasitic redox systems are a valuable source of promising targets since the maintenance of intracellular redox homeostasis is essential for parasite survival inside their hosts; moreover, these systems often depend on unique and nonredundant core redox enzymes (Salinas, 2013). In the case of the Trypanosomatidae, redox homeostasis relies on the dithiol trypanothione (TSH), a peculiar variant of glutathione (GSH) in which two GSH molecules are joined by the polyamine spermidine. TSH is kept in the reduced state by the action of the NADPH/FAD-dependent protein trypanothione reductase, homolog to the human glutathione reductase (GR).

Since its identification, TR has been considered a convincing target for drug discovery because of some attracting characteristics: 1) it is essential for parasite viability; 2) it differs from the homolog GR enough to allow for selectivity; and 3) it is druggable since it has been shown to be efficiently addressed by inhibitors (Battista et al., 2020). Moreover, the high sequence conservation of TRs from different species, reaching 100% for the TSH cavity, supports the possibility to develop a broad-spectrum drug against all trypanosomatid infections (Ilari et al., 2018).

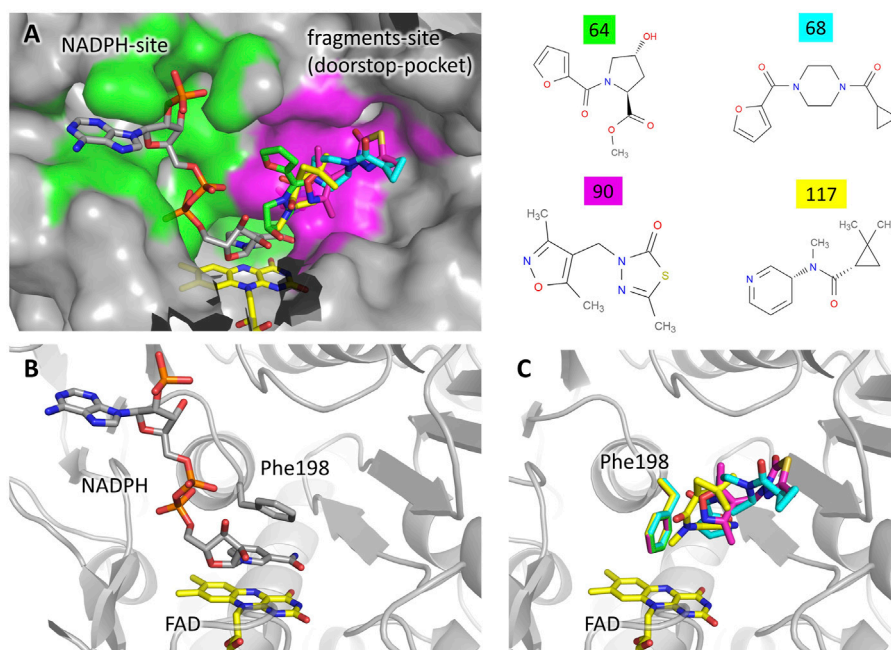
However, TR has a significant limitation that must be taken into account in drug discovery. Indeed, previous studies showed that, in order to significantly affect parasite viability in culture, TR activity must be reduced by at least 90%, meaning that only inhibitors with sub-micromolar IC<sub>50</sub> should be considered

suitable drug candidates (Krieger et al., 2000). The identification of such inhibitors is particularly challenging considering that the substrate cavity is wide and solvent exposed, thus making it difficult to efficiently target it with small drug-like compounds (Patterson et al., 2011).

Considerable effort has been put together to identify new hits targeting TR, rationalize the interaction through structural characterization, and improve known scaffolds through SAR studies or structure-based design (Bernardes et al., 2013; Field et al., 2017; Battista et al., 2020), but despite some remarkable results (Patterson et al., 2011; De Gasparo et al., 2018; Turcano et al., 2018; De Gasparo et al., 2019; Turcano et al., 2020), none of the leads proposed has been yet promoted to preclinical trials, due to sub-optimal activity or toxicity issue. Most characterized inhibitors were found to target the TSH cavity, mainly the so-called mepacrine binding site (MBS) located at the entrance of the TSH cavity. More recently, an alternative structure-based strategy has been successfully explored, aimed to disrupt TR functional dimer by targeting a cavity located at the dimerization interface (Revuelto et al., 2021). However, it is very likely that other sites as well as other modes of inhibition not yet identified can be exploited for TR inactivation.

In this context, we decided to perform the first crystallographic fragment screening on TR, since we believe that it can make an important contribution to TR-targeted drug development. In fact, fragment screening is now well-established as a powerful approach to early drug discovery since it gives the unique opportunity to probe protein surfaces with limited collections of small molecules (150–250 Da) covering a large chemical space (Murray and Rees, 2009). Among the biophysical detection methods available for screening, crystallography has the advantage to identify the hits and define the binding mode in one single step. Therefore, it is possible to obtain a detailed mapping of the interaction hot spots of a target and to unravel unexpected mechanisms of inhibition. These results may be particularly useful to overcome the issues encountered so far with TR as a target. In fact, the detailed scanning of the TSH cavity as well as the identification of previously unexplored allosteric pockets has the potential to disclose new opportunities for lead design.





**FIGURE 2** | Fragments binding to the “doorstop pocket”. **(A)** View of the NADPH binding site (green) and the fragment binding site (magenta). Although soaking has been carried out in absence of NADPH, it is shown in the figure for clarity. The figure was prepared by adding the fragments (color code on the left) to the structure of the complex TbTR-NADPH (pdb: 2wov). **(B)** Conformation of Phe198 upon NADPH binding. **(C)** Conformation of Phe198 in presence of the fragments, hampering NADPH binding.

## 2 EXPERIMENTAL METHODS

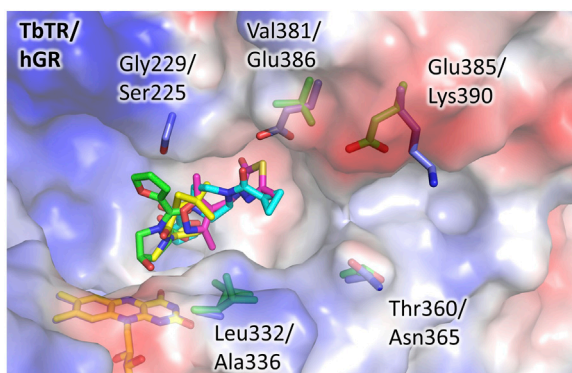
### 2.1 Protein Production and Crystallization

TR from *L. infantum* was produced as previously reported. TR from *T. brucei* (TbTR) was produced as previously reported (Turcano et al., 2020), with minor modifications. Briefly, the gene corresponding to TbTR (Uniprot ID: Q389T8) was cloned in the pET15b vector using the restriction site NdeI and XhoI to add an N-terminal His-tag. The construct TbTR-pET15b was transformed in *E. coli* BL21 (DE3) cells and expressed in LB medium by o/n induction with 0.5 mM IPTG (OD<sub>600</sub> = 0.6) at 37°C. Cells were recovered by centrifugation, lysis was performed by sonication on ice (lysis buffer: 20 mM Tris-HCl pH 8, 300 mM NaCl, 5 mM imidazole, 0.1 mM PMSF, DNase, 5 mM MgCl<sub>2</sub>). The soluble fraction was collected by centrifugation at 40,000 g for 20 min, clarified by filtration through a 0.2 μm syringe filter, and then applied to prepacked 5 ml HisTrap columns (Cytiva). The protein was eluted by a linear gradient against 500 mM imidazole (elution peak around 150 mM imidazole). The His-tag was removed by thrombin cleavage, 1 U of thrombin per mg of TbTR, overnight incubation at 4°C in 50 mM TRIS-HCl pH 8, 100 mM NaCl, 2 mM CaCl<sub>2</sub>. Thrombin was removed by Benzamidine-Sepharose 6B resin (Cytiva) and tag-free TbTR was further purified by reverse immobilized metal affinity chromatography (Protino Ni-TED resin, Macherey-Nagel). Finally, the buffer was exchanged in a centrifugal concentrator with 20 mM HEPES pH 7.5, the protein was concentrated at 12 mg/ml, frozen in liquid nitrogen, and stored at -80°C.

The crystallization condition suitable for screening at the Diamond Light Source XChem facility was identified by optimizing the conditions previously reported (Patterson et al., 2011). Little differences in the concentration of the precipitants led to two different crystal forms: plate-like monoclinic crystals with the tetrameric asymmetric unit (AU), same as reported in the original study, and tridimensional orthorhombic crystals with dimeric AU. The orthorhombic crystals have been considered more favorable since they resulted to be more resistant to mechanical stress and diffract consistently up to 1.6 Å; moreover, the dimeric AU, corresponding to the biological unit of TR, is more convenient than the tetrameric one for the inspection of the events identified by the Pan-Dataset Density Analysis (PanDDA). The selected crystallization condition is composed as follows: 22% 2-Methyl-2,4-pentanediol (MPD), 14% PEG3350, imidazole 40 mM pH 8. 50 mM NaBr must be added to the protein solution to get regular and well-diffracting crystals. Protein and crystallization solution was then shipped in dry ice to the XChem facility where crystallization plates have been set up in SwissCi 3-drops midi plates using 200 nl of protein solution plus 200 nl of crystallization solution.

### 2.2 Fragment Screening Experiment and Structure Solution

A total of 381 fragments from the DSI poised library (<https://enamine.net/compound-libraries/fragment-libraries/dsi-poised-library>) (500 mM stock concentration dissolved in DMSO) were transferred directly to TbTR crystallization drops using an ECHO



**FIGURE 3 |** The doorstep pocket, differences between TbTR and hGR. The electrostatic semitransparent surface of TbTR in complex with the fragment 64 (pdb: 5S9T) is reported as generated by PyMol. Yellow arrows indicate the main sequence substitutions, the corresponding side chains are shown as sticks (green for TbTR, blue for hGR (pdb: 3djj)).

liquid handler (50 mM nominal final concentration) and soaked for 1–2 h before harvesting. Data were collected at Diamond light source beamline I04-1. A total of 371 datasets were collected in the resolution range of 2.9 Å or higher with the majority being in the range of 1.6 Å to 2.2 Å.

Data processing was performed using the automated XChem Explorer pipeline (Krojer et al., 2017). Fragment hits were

identified using the PanDDA algorithm (Pearce et al., 2017) followed by visual inspection. Refinement was performed using REFMAC (Murshudov et al., 2011). A summary of data collection and refinement statistics for all fragment-bound datasets and the reference dataset is shown in **Supplementary Table 1**.

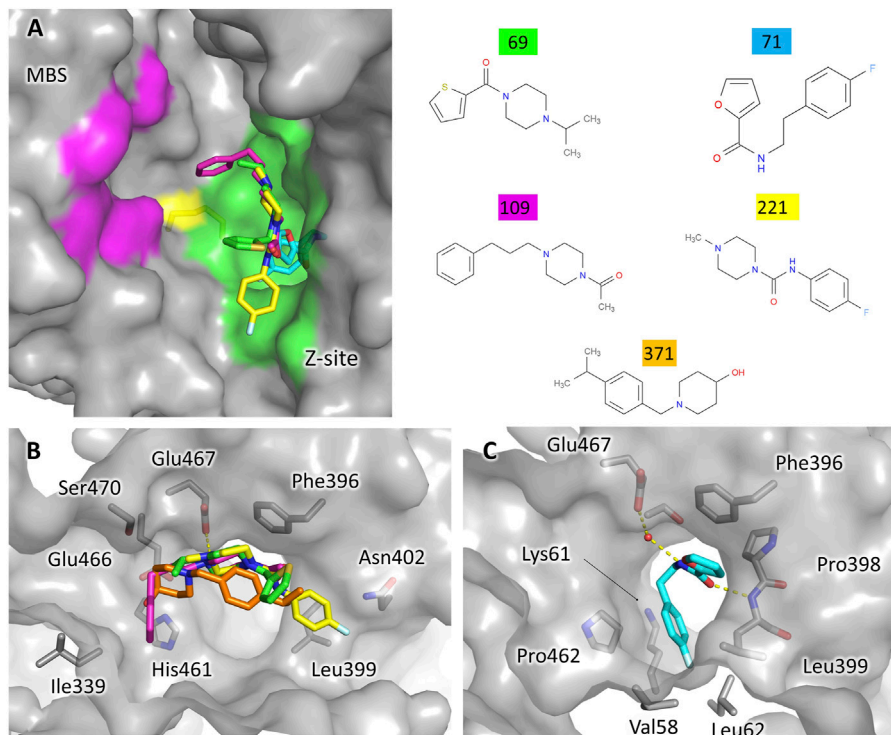
## 2.3 Enzymatic Assay

Enzymatic inhibition assays were performed on LiTR as reported in (Ilari et al., 2018). Briefly, each assay was started by the addition of NADPH 100 μM to the pre-equilibrated reaction solution comprising HEPES 50 mM at pH 7.4, NaCl 40 mM, LiTR 50 nM, oxidized TSH 150 μM, and inhibitor. The decrease of absorbance at 340 nm, indicating the oxidation of NADPH, was measured to follow reaction progression.

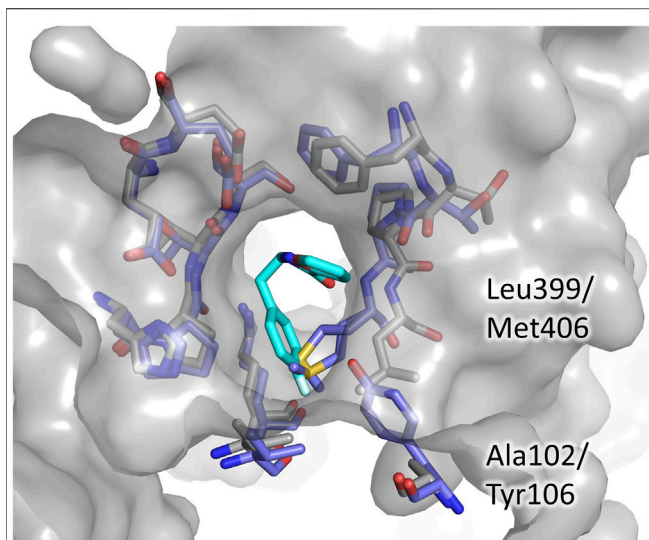
The selected compounds were tested at a fixed concentration (100 μM, triplicates), and the residual activity with respect to control was considered.

## 2.4 Chemistry

All the commercially available reagents and solvents were purchased from Sigma-Aldrich, Alfa Aesar, VWR, and TCI and used without further purification. Reactions were followed by analytical thin-layer chromatography (TLC) on pre-coated TLC plates (layer: 0.20 mm silica gel 60 with a fluorescent indicator UV254, from Sigma-Aldrich). Developed plates were air-dried and analyzed under a UV lamp (UV 254/365 nm). A CEM Discover SP-focused microwave reactor was used for microwave-assisted



**FIGURE 4 |** Fragments binding at the TSH cavity. **(A)** View of the whole TSH cavity. The surfaces corresponding to the Z-site and the MBS are colored green and magenta, respectively. The redox active cysteines are shown as sticks and colored yellow to help locate the fragments with respect to the catalytic site. **(B)** Binding of the fragments 69, 109, 221, and 371 are characterized by an amino heterocycle. **(C)** Binding of the fragment 71.



**FIGURE 5 |** The Z-site, differences between TbTR and hGR. The residues lining the Z-site are shown as sticks, grey for TbTR, blue for hGR, and fragment 71 are colored cyan. Labels are reported only for residues differing in TR and GR.

reactions. Nuclear magnetic resonance (NMR) experiments were run on a Varian VXR 400 (400 MHz for  $^1\text{H}$  and 100 MHz for  $^{13}\text{C}$ ).  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were acquired at 300 K using deuterated chloroform (chloroform- $d$ ) and dimethyl sulfoxide (dimethyl sulfoxide- $d_6$ ) as solvents. Chemical shifts ( $\delta$ ) are reported in parts per million (ppm) relative to tetramethylsilane (TMS) as the internal reference and coupling constants ( $J$ ) are reported in hertz (Hz). The spin multiplicities are reported as s (singlet), br s (broad singlet), d (doublet), t (triplet), q (quartet), and m (multiplet). Mass spectra were recorded on a Waters ZQ4000, XevoG2-XSQTof, Acquity arc-QDA LC-MS apparatus with electrospray ionization (ESI) in positive mode. Compounds were named following IUPAC rules as applied by ChemBioDraw Ultra (version 19.0). The purity of compounds was determined using a Kinetex 5  $\mu\text{M}$  EVO C18 100 Å, LC column 150  $\times$  4.6 mm, and an HPLC JASCO Corporation (Tokyo, Japan) instrument (PU-1585 UV equipped with a 20  $\mu\text{l}$  loop valve). All the tested compounds showed  $\geq 95\%$  purity by analytical HPLC.

#### 2.4.1 1-(4-(3-Phenylpropyl)piperazin-1-yl)ethan-1-one (1)

To a stirred solution of **8** (79 mg; 0.38 mmol, 1.0 eq) and triethylamine (0.07 ml, 0.48 mmol, 1.25 eq) in dichloromethane (4 ml) at  $0^\circ\text{C}$ , acetic anhydride (0.04 ml, 0.42 mmol, 1.1 eq) was added dropwise. The resulting reaction mixture was stirred at room temperature for 3 h. Upon completion, the reaction mixture was diluted with additional dichloromethane (10 ml) and washed with saturated  $\text{NaHCO}_3$  aqueous solution (2  $\times$  10 ml). The organic phase was dried over anhydrous  $\text{Na}_2\text{SO}_4$ , filtered, and concentrated under *vacuum* to afford **1** as a colorless oil (86 mg, 92%).  $^1\text{H}$  NMR (400 MHz, chloroform- $d$ )  $\delta$  7.32–7.24 (m, 3H), 7.23–7.15 (m, 2H), 3.62 (t,  $J$  = 5.2 Hz, 2H), 3.50–3.43 (m, 2H), 2.69–2.61 (m, 2H), 2.46–2.34 (m,

6H), 2.08 (s, 3H), 1.82 (p,  $J$  = 7.6 Hz, 2H).  $^{13}\text{C}$  NMR (101 MHz, chloroform- $d$ )  $\delta$  169.0, 142.0, 128.5 (2C), 128.5 (2C), 126.0, 57.9 (2C), 53.4, 52.9, 46.5, 41.6, 33.7, 21.8.

#### 2.4.2 N-(4-fluorophenyl)-4-methylpiperazine-1-carboxamide (2)

To a solution of phenyl (4-fluorophenyl)carbamate (**13**, 84 mg, 0.36 mmol) in dichloromethane (1.85 ml) was added triethylamine (2.00 eq) and the methylpiperazine (**12**, 0.03 ml, 0.36 mmol). The reaction mixture was heated at  $40^\circ\text{C}$  overnight. Solvent was removed under *vacuum* and the resulting crude purified by silica gel chromatography. Purification by silica gel column chromatography (dichloromethane/methanol, 9.5:0.5), gave **2** as pale-yellow powder (54 mg, 64%).  $^1\text{H}$  NMR (400 MHz, chloroform- $d$ )  $\delta$  7.32–7.27 (m, 2H), 7.01–6.90 (m, 2H), 6.26 (br, 1H), 3.52–3.49 (m, 4H), 2.47–2.43 (m, 4H), 2.34 (s, 3H).  $^{13}\text{C}$  NMR (101 MHz, chloroform- $d$ )  $\delta$  159.1 (d,  $J_{\text{C-F}}$  = 242.0 Hz), 155.2, 135.0 (d,  $J_{\text{Cp-F}}$  = 2.8 Hz), 122.2 (d,  $J_{\text{Cm-F}}$  = 8.0 Hz, 2C), 115.6 (d,  $J_{\text{Co-F}}$  = 22.4 Hz, 2C), 54.8, 46.3, 44.2.

#### 2.4.3 N-(4-fluorophenethyl)furan-2-carboxamide (3)

2-(4-Fluorophenyl)ethan-1-amine (**15**, 0.20 ml, 1.53 mmol) was added dropwise to a microwave vial charged with a magnetic stirring bar and the acyl chloride **14** (0.43 ml, 3.28 mmol) in toluene (3 ml). The reaction was carried out under microwave irradiation at  $100^\circ\text{C}$ , 150 W for 10 min. The solvent was removed under vacuum, and the crude purified by silica gel chromatography. Purification by silica gel column chromatography (dichloromethane/methanol, 9.8:0.2) gave **3** as yellowish oil (345 mg, 94%).  $^1\text{H}$  NMR (400 MHz, chloroform- $d$ )  $\delta$  7.40 (dd,  $J$  = 1.8 and 0.8 Hz, 1H), 7.22–7.15 (m, 2H), 7.10 (dd,  $J$  = 3.6 and 0.8 Hz, 1H), 7.04–6.96 (m, 2H), 6.49 (dd,  $J$  = 3.6 and 1.8 Hz, 1H), 6.36 (s, 1H), 3.67–3.61 (m, 2H), 2.89 (t,  $J$  = 7.2 Hz, 2H).  $^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  161.8 (d,  $J_{\text{C-F}}$  = 244.4 Hz), 158.5, 148.1, 143.9, 137.5, 134.5 (d,  $J_{\text{Cp-F}}$  = 3.2 Hz), 130.3 (d,  $J_{\text{Cm-F}}$  = 7.8 Hz, 2C), 115.6 (d,  $J_{\text{Co-F}}$  = 21.2 Hz, 2C), 114.3, 112.3, 40.5, 35.2, 21.7.

#### 2.4.4 Tert-Butyl Piperazine-1-carboxylate (5)

To a solution of piperazine (**4**, 4.04 g, 48.00 mmol, 6.0 eq) in dichloromethane (24 ml) at  $0^\circ\text{C}$ , was added a solution of  $\text{Boc}_2\text{O}$  (1.75 g, 8.00 mmol, 1.0 eq) in dichloromethane (40 ml) over a period of 2 h. The reaction mixture was stirred at room temperature for 18 h. Upon completion, the resulting suspension was washed with  $\text{H}_2\text{O}$  (5  $\times$  100 ml) to remove the unreacted piperazine. The organic phase was dried over anhydrous  $\text{Na}_2\text{SO}_4$ , filtered, and concentrated under *vacuum* to afford **5** as a colorless solid (947 mg, 64%).  $^1\text{H}$  NMR (400 MHz, chloroform- $d$ )  $\delta$  3.44–3.35 (m, 4H), 2.84–2.75 (m, 4H), 2.16 (br, 1H), 1.45 (s, 9H).

#### 2.4.5 Tert-Butyl 4-(3-Phenylpropyl)piperazine-1-carboxylate (7)

In a 50-ml sealed vessel, a stirred mixture of **5** (555 mg, 2.97 mmol, 1.5 eq) and commercially available (3-bromopropyl)benzene (**6**, 0.30 ml, 1.98 mmol, 1.0 eq) in acetonitrile (6 ml) in the presence of  $\text{K}_2\text{CO}_3$  (411 mg; 2.97 mmol, 1.5 eq) as base and KI (3 mg; 0.02 mmol, 0.01 eq)



**TABLE 1** | Overall data quality for fragment hits. R/R<sub>f</sub>: R factor and free R factor as calculated by the last refinement cycle in Refmac. RSCC: real space correlation coefficient. The reference model is the structure of unbound TbTR based on the best dataset collected.

Sample	Pdb code	R/R <sub>f</sub>	Resolution	Binding site	Occupancy	RSCC
60	5S9S	20.4/24.3	1.8	5	0.88	0.85
64	5S9T	18.9/22.1	1.66	14–15	0.7–0.64	0.93–0.77
68	5S9U	19.3/22.8	1.73	14–15	0.68–0.63	0.85–0.8
69	5S9V	18.7/22.9	1.9	2–6	1–0.96	0.83–0.87
71	5S9W	19.9/24.6	1.96	2–6	0.66–0.72	0.86–0.89
90	5S9X	19.7/23.8	1.84	5–14–15	0.29–0.5–0.63	0.87–0.89–0.85
94	5S9Y	19.8/23.6	1.75	9–7	0.86–0.7	0.9–0.84
109	5S9Z	19.3/22.5	1.73	6–2	0.8–0.92	0.86–0.91
117	5SA0	19.6/25.1	1.97	14–15	0.63–0.72	0.78–0.81
221	5SA1	19.4/23.1	1.84	2	1	0.93
296	5SA2	20/23.8	1.78	21	0.8	0.84
371	5SA3	18.7/21.8	1.74	6	0.76	0.87
Reference model	5SMJ	19.3/21.8	1.65	-	-	-

as a catalyst, was heated to 80°C for 3 h. Upon completion, the hot suspension was filtered, and the residue was washed with acetone several times. The collected filtrates were concentrated under *vacuum* and the resultant crude was purified by silica gel column chromatography (dichloromethane/methanol/ammonia 32% aqueous solution, 9.8:0.2:0.02) to afford **7** as a colorless oil (554 mg, 92%). <sup>1</sup>H NMR (400 MHz, chloroform-*d*) δ 7.31–7.23 (m, 2H), 7.22–7.14 (m, 3H), 3.44 (t, *J* = 5.2 Hz, 4H), 2.64 (t, *J* = 7.6 Hz, 2H), 2.40–2.35 (m, 6H), 1.84 (q, *J* = 7.6 Hz, 2H), 1.45 (s, 9H).

#### 2.4.6 1-(3-Phenylpropyl)piperazine (**8**)

**7** (368 mg, 1.21 mmol, 1.0 eq) was treated with trifluoroacetic acid (1.85 ml, 24.2 mmol, 20.0 eq) in dichloromethane (12 ml) at 0°C. The ice bath was removed, and the resulting mixture was stirred at room temperature for 2 h. Upon reaction completion, the mixture was diluted with additional dichloromethane (10 ml) and washed with saturated NaHCO<sub>3</sub> aqueous solution (2 × 15 ml). The organic phase was dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>, filtered, and concentrated under *vacuum* to afford **8** as a colorless oil (233 mg, 94%). <sup>1</sup>H NMR (400 MHz, chloroform-*d*) δ 7.28–7.20 (m, 2H), 7.19–7.10 (m, 3H), 2.87 (t, *J* = 4.8 Hz, 4H), 2.65–2.56 (m, 2H), 2.46–2.27 (m, 6H), 2.20 (br, 1H), 1.85–1.73 (m, 2H).

#### 2.4.7 Phenyl (4-Fluorophenyl)carbamate (**11**)

To a solution of **9** (0.09 ml, 0.90 mmol, 1.0 eq) and Na<sub>2</sub>CO<sub>3</sub> (57.2 mg, 0.54 mmol, 0.6 eq) in a mixture of ethyl acetate, tetrahydrofuran and H<sub>2</sub>O (1.36, 0.27 and 0.27 ml, respectively) at 0°C, phenyl chloroformate (**10**, 0.12 ml, 0.99 mmol, 1.1 eq) was added dropwise. The reaction mixture was stirred at room temperature overnight, then it was concentrated under a *vacuum* to remove organic solvents. Water was added to the residue and the resulting precipitate was recovered by filtration under *vacuum*, washed with water, and dried to give compound **11** as a grey solid (357 mg, quantitative yield). <sup>1</sup>H NMR (400 MHz, chloroform-*d*) δ 7.44–7.37 (m, 4H), 7.26–7.22 (m, 1H), 7.21–7.17 (m, 2H), 7.08–7.01 (m, 2H), 6.88 (br, 1H).

#### 2.4.8 Furan-2-carbonyl Chloride (**14**)

SOCl<sub>2</sub> (5.26 ml, 44.6 mmol) was added dropwise to a suspension of the furan-2-carboxylic acid (**13**, 500 mg, 4.46 mmol) in toluene (22 ml). The reaction refluxed at 110°C for 2 h before heating was

stopped. Evaporation of the volatiles under *vacuum* gave the desired compound (assumed 100% yield), which was employed in the next synthetic step without further purification.

## 3 RESULTS AND DISCUSSION

### 3.1 Fragment Screening: 12 Hits Identified in 5 Sites

Crystallographic fragment screening performed at the XChem facility allowed us to test 352 small compounds from the DSiP library, a collection specifically designed to ensure rapid and cheap follow-up synthesis (Cox et al., 2016).

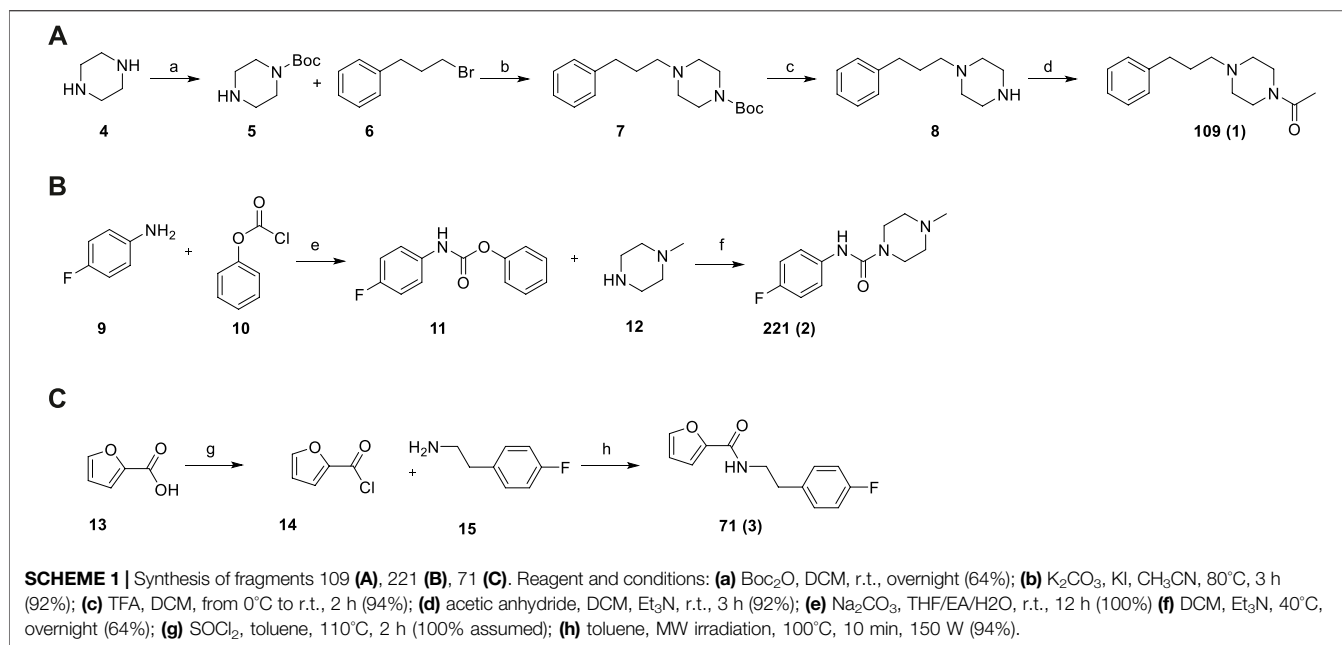
Data analysis, performed by the PanDDA algorithm (Pearce et al., 2017), reported 480 interesting “events” clustered in 33 sites possibly associated to fragment binding. Visual inspection of these putative events led to the identification of 21 true binding events in eight sites, corresponding to 12 fragment hits distributed in 5 independent binding sites (Figure 1), described later in detail. The experimental results have been compared to the prediction made by FTmap (Kozakov et al., 2015), a computational mapping server that identifies binding hot spots (Supplementary Figure S3).

Table 1 gives an overview of data quality for the hits. All datasets have a resolution higher than 2 Å and the modeled fragments show a good correlation with the electron density maps, as proven by the real-space correlation coefficient (RSCC) (Pearce et al., 2017). Indeed, the RSCC is around 0.9 for most hits and always in the range of 0.77–0.93. Refined occupancy is relatively high for all the fragments apart from sample 90, but even in this case, the ligand has been unambiguously located thanks to the event map generated by PanDDA (see Supplementary Material for a view of the event map for each ligand).

### 3.2 Fragments Binding in Proximity of the NADPH-Site: The Doorstop-Pocket

Four different fragments bind to the site 14–15. The binding does not cause any significant structural variation with respect to the unbound protein. A detailed description of the interactions engaged by each fragment is reported in





**Supplementary Figure S1.** The analysis of the unbound TbTR structure made by FTmap predicts a hot spot located at this site, although it does not stand out as a remarkable druggable site (**Supplementary Figure S3**). Site 14–15 is located in close proximity to the NADPH binding cavity (**Figure 2**). As shown in **Figure 2A**, the two binding pockets are almost independent (note that fragment soaking has been performed in the absence of NADPH). Although there is no significant overlap between the fragments and NADPH (**Figure 2A**), the presence of the fragments would hamper the entrance of NADPH into its cavity due to the conformation assumed by Phe198 (**Figures 2B,C**). In fact, Phe198 is located over the FAD and, in the apo conformation, its side chain is perpendicular to the isoalloxazine ring. Upon NADPH binding, the nicotinamide moiety inserts between FAD and Phe198 forming an aromatic sandwich in which the phenyl ring lays parallel to the FAD (**Figure 2B**). Therefore, for the NADPH to bind to TbTR, it is essential that Phe198 is able to switch from its perpendicular position to a parallel one with respect to the FAD. The binding of the fragments in the adjacent pocket blocks Phe198 in the perpendicular conformation, thus the binding of the cofactor is prevented (**Figure 2C**).

A very similar mechanism of inhibition has been observed in 2018 for the protein thioredoxin glutathione reductase from *Schistosoma mansoni* (SmTGR) by Angelucci and coworkers (Silvestri et al., 2018). The authors defined the fragment binding site as the “doorstop pocket”, according to the effect to block the aromatic side chain that acts as a door for the entrance of NADPH. Indeed, the mode of binding of NADPH described above for TR from *Trypanosoma brucei* is common to NAD/FAD reductases in general, only differing by the nature of the “door” residue that may either be a phenylalanine or a tyrosine residue depending on the species.

By sequence comparison with homologous, the authors noted that the residues contributing to the doorstep pocket in SmTGR

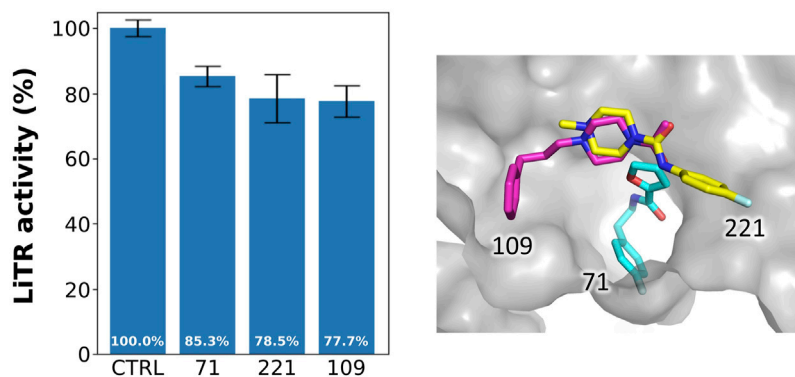
are highly conserved, and they predicted the existence of a similar pocket in TR from leishmania and trypanosoma as well as in human GR and thioredoxin reductase (TrxR).

To evaluate the feasibility to exploit this pocket to develop selective inhibitors of TbTR, we compared its structure to hGR and detected some significant differences (**Figure 3**). We found few substitutions among the residues lining the pocket, namely Gly229 and Leu332 replaced by Ser225 and Ala336 in hGR, that cause a shape variation that is limited but capable to interfere with the binding. However, more relevant substitutions are present not far away from the pocket, and concern residues Val381, Glu385, and Thr360 which are replaced by Glu386, Lys390, and Asn365 in hGR. These substitutions, affecting both shape and charge distribution, are located along a small contiguous cleft (**Figure 3**) that could be targeted by fragment growing. Notably, other human NAD-dependent reductases, such as TRxR, should be taken into account as potential off-targets.

### 3.3 Fragments Binding to the TSH Cavity

Five fragments were found to bind at the site two to six, located in the TSH cavity. The ligands occupy the same region of the cavity, corresponding to the so-called Z-site, a sub-pocket predicted by bioinformatic analysis (Khan et al., 2000) but never confirmed experimentally until now (**Figure 4**).

Surprisingly, no ligands could be modeled at the mepacrine binding site (MBS), although most of the yet structurally characterized inhibitors of TR are known to target this hot-spot (Battista et al., 2020), including the ones mispredicted to bind the Z-site (Saravanamuthu et al., 2004). Indeed, the PanDDA analysis reported some events at the MBS, but the density map was so poorly defined that these putative hits were not further considered. This result is confirmed by FTmap since no hot spots are predicted at the MBS (**Supplementary Figure S3**).



**FIGURE 6 |** Activity of the selected fragments. Residual activity of LiTR after treatment with fixed concentration (100  $\mu$ M) of the selected fragments.

Some relevant features can be detected among the fragments targeting the Z-site. Four fragments contain an amino heterocycle, namely a piperazine ring for hits 69, 109, 221 or a piperidine ring for hits 371 (**Figure 4B**). It is worth noticing that the piperazine assumes the same position in all the structures. In general, the amino heterocycle strongly contributes to binding, in fact, the positively charged amino group forms a salt bridge with Glu467 (N-O distance: 2.7 Å for piperazine, 3.6 Å for piperidine), while the other main interactions are hydrophobic and involve Ile339, Phe396, Leu399 (**Figure 4B**).

The fifth fragment, compound 71, strongly differs in the mode of binding: it explores a small cavity corresponding to the narrow entrance of a wide interfacial cavity that connects the two TSH sites on the dimer (**Figure 4C**). The interaction is mainly hydrophobic: the fluorophenyl group targets a sub-pocket lined mostly by hydrophobic side chains (**Figure 4C**) while the furan makes a stacking interaction with Phe396. Moreover, the amide moiety engages in hydrogen bonds with the Leu399 backbone and a water molecule bridging to Glu367.

The binding of the ligands to the Z-site does not require any conformational rearrangement of the site with respect to the apo structure (ground state model, PDB ID 5SMJ), except a very limited movement of the stretch 395–397 toward the ligand in the case of fragment 71.

Although no known inhibitor binds this area of the cavity, a HEPES molecule was serendipitously found in the TbTR structure solved in a complex with an inhibitor targeting the MBS (De Gasparo et al., 2019). It is interesting to note that the piperazine of HEPES interacts with Glu467 while the sulfonate points toward the subpocket. This finding stresses a role for the piperazine as an anchoring moiety for the Z-site.

Intriguingly, the interfacial cavity, whose entrance appears to be occluded by hit 71, has been exploited to carry out an alternative inhibition strategy of TR aimed at the disruption of the dimeric assembly (Ruiz-Santaquiteria et al., 2017; Revuelto et al., 2021). Moreover, a similar cavity is present in GR and is targeted by some non-competitive or uncompetitive inhibitors such as xanthenes (Savvides and Karplus, 1996), menadione (Karplus et al., 1989), and N-arylisoalloxazines (Schönleben-Janans et al., 1996). Indeed, the analysis made using the FTMap

web server (**Supplementary Figure S3**) highlights hot-spots corresponding to hit 71 and indicates the whole interfacial cavity as a significant druggable site. The fact that no hits have been found within the cavity could be due to poor accessibility, resulting from the combination of the narrow entrance and the structural rigidity caused by crystal packing.

Even in the case of the Z-site, we assessed the potentiality for selective inhibition by comparison to hGR. As shown in **Figure 5**, almost all the residues contributing to fragment binding are conserved apart for Leu399 which is replaced by Met406. This finding is not surprising because, during catalysis, this zone of the cavity hosts the carboxylic ends shared by both TSH and glutathione. However, structure superimposition shows that in hGR the small cavity targeted by fragment 71 is narrower due to the shift of the segment 396–399 (403–406 in hGR). The difference in conformation is justified by the replacement of Ala102 with the bulkier Tyr106: the side chain, therefore, pushes Met406 and results in a shrunk subpocket where the entrance of the ligand would be hampered (**Figure 5**).

This observation indicates that, in order to develop a selective inhibitor targeting the Z-site, it is essential to exploit this small cavity explored by fragment 71.

### 3.4 Other Binding Sites

The remaining binding sites (5, 7–9, 21) are less populated by fragments and are located in regions that seem not particularly significant for the modulation of the activity of the protein. Thus, these sites are likely of little relevance and the corresponding fragments will be only briefly described.

The binding sites of the fragments 60 and 90 are very close, thus they are grouped as site 5 even if they are distinct. In fact, while compound 60 binds at a mainly hydrophobic pocket on the protein surface, compound 90 is completely buried within the dimerization interface (**Supplementary Figure S2**). Note that compound 90 mainly binds to sites 14–15 (the doorstep pocket) while its occupancy in site 5 is low. Site 5 is far away from the catalytic site, and the compound binding does not cause any significant conformational variation, except the shift of the side chains of Leu73 and Phe83 in the case of hit 90 (**Supplementary Figure S2B**), suggesting that the binding of these ligands may not

affect TbTR activity. However, it is worth noticing that fragment 90 in site 5 is adjacent to the interfacial cavity where the xanthene inhibitor of GR is bound (**Supplementary Figure S3**).

Fragments 94 and 296 are placed at sites 7–9 and 21 that are rather peripheral and solvent exposed.

### 3.5 Synthesis of Selected Fragments Targeting the Z-Site

Based on the structure analysis, we decided to focus on the fragments binding to the Z-site. Indeed, these hits offer the opportunity to improve already known inhibitors as well as to design completely new compounds by merging and linking approaches. We selected three representative fragments, i.e., two fragments containing the piperazine ring (109 and 221) and the one targeting the sub-pocket (71).

Given the in-house availability of the building blocks, we decided to synthesize the compounds **109** (**1**), **221** (**2**), and **71** (**3**), over purchasing them from commercial vendors. Fragments **1–3** were thus synthesized by slightly modified reported procedures (**Scheme 1**) (Glennon et al., 1981; McNaughton et al., 2009; Li et al., 2019). In this way, the optimized synthetic routes also give clues about the synthetic feasibility, and the scale-up and provide the basis for analog expansion. All the synthesized compounds showed  $\geq 95\%$  purity by analytical HPLC and were characterized by spectroscopic data ( $^1\text{H}$  and  $^{13}\text{C}$  NMR), which are reported in the *Methods* Section.

### 3.6 Activity Assay of Selected Fragments

We tested the activity of the selected fragments by preliminary inhibition assay on TR from *Leishmania infantum*. Indeed, the high degree of sequence conservation, reaching 100% for the TSH cavity, allows using TR from any species.

As shown in **Figure 6**, reporting the residual activity of LiTR after treatment with a fixed concentration (100  $\mu\text{M}$ ) of the compounds, the fragments have a low but evident effect on protein activity, in line with the general behavior of fragments, whose affinity for the target is usually in the millimolar range due to the small size of these compounds that limits the number of possible interactions.

## 4 CONCLUSION AND PERSPECTIVES

TR is an appealing but challenging target for the development of new trypanocidal drugs. Indeed, due to its high efficiency and turnover, only very potent inhibitors can be considered as leads, but the structural characteristics of the substrate binding cavity, wide and solvent exposed, make this goal difficult to achieve.

In order to bring out new opportunities for inhibition, such as new binding molecular scaffolds as well as uncharacterized secondary sites, we performed the first crystallographic fragment screening ever reported on TR. The experiment resulted in 12 fragments targeting 5 independent sites, two of which are of particular interest.

The screening revealed the existence of an allosteric pocket close to the NADPH binding site, named “doorstop pocket” in

reference to the mechanism of inhibition observed. A very similar site has been recently detected in a homolog protein from the parasite *Schistosoma* and was successfully exploited to develop efficient inhibitors active in the low micromolar range (Silvestri et al., 2018), stressing the potentiality of the new druggable site identified by our study.

The second site, known as the Z-site, is located within the large TSH cavity but corresponds to a region not yet exploited for inhibition. The fragments binding to this site have some remarkable features making them ideal for follow-up optimization. Indeed, the recurrence of a piperazine moiety in three out of five fragments suggests a convenient anchoring point, suitable for fragment merging. Moreover, the capability of one fragment to target a small and specific subpocket provides an opportunity for specific inhibition. Based on these observations, we selected and synthesized three hits particularly suited for future developments and tested their activity on TR resulting to be in line with inhibitors with such a limited size. We are currently working on the design and synthesis of derivatives of the selected fragments based on the merging and linking approach. Moreover, the fact that most structurally characterized inhibitors bind to other regions of the same cavity suggests the possibility to exploit our results to modify known inhibitors to improve their activity and selectivity, possibly overcoming the constitutive limitations of TR as a drug target.

The promising results gathered from this first fragment screening campaign bring new light to TR as a target for a structure-based approach and encourage the expansion of the chemical space explored by testing new fragments collections.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <http://www.wwpdb.org/>, 5S9S, 5S9T, 5S9U, 5S9V, 5S9W, 5S9X, 5S9Y, 5S9Z, 5SA0, 5SA1, 5SA2, 5SA3, and 5SMJ.

## AUTHOR CONTRIBUTIONS

AF and AI designed and conducted the research. AF performed the experiments that yielded the crystallographic data. AF analyzed the crystal structures. FS, AS, and JC synthesized the compounds. CE performed the activity assays. GC and AL were involved in discussions of the project and revising the manuscript. All authors reviewed the results and approved the final version of the manuscript.

## FUNDING

This work was supported by a grant (ID: FIS2019\_03796) from the MIUR (Italian Ministry of Education, University, and Research) to AI. Access to XChem facility was supported by iNEXT-Discovery (PID: 13997).

## ACKNOWLEDGMENTS

The authors thank Alice Douangamath and Daren Faeron (Diamond Light Source) for their kind assistance with the XChem experiment.

## REFERENCES

- Battista, T., Colotti, G., Ilari, A., and Fiorillo, A. (2020). Targeting Trypanothione Reductase, a Key Enzyme in the Redox Trypanosomatid Metabolism, to Develop New Drugs against Leishmaniasis and Trypanosomiasis. *Molecules* 25 (8), 1924. doi:10.3390/molecules25081924
- Bernardes, L. S. C., Zani, C. L., and Carvalho, I. (2013). Trypanosomatidae Diseases: from the Current Therapy to the Efficacious Role of Trypanothione Reductase in Drug Discovery. *Curr. Med. Chem.* 20 (21), 2673–2696. doi:10.2174/0929867311320210005
- Cox, O. B., Krojer, T., Collins, P., Monteiro, O., Talon, R., Bradley, A., et al. (2016). A Poised Fragment Library Enables Rapid Synthetic Expansion Yielding the First Reported Inhibitors of PHIP(2), an Atypical Bromodomain. *Chem. Sci.* 7 (3), 2322–2330. doi:10.1039/c5sc03115j
- De Gasparo, R., Brodbeck-Persch, E., Bryson, S., Hentzen, N. B., Kaiser, M., Pai, E. F., et al. (2018). Biological Evaluation and X-Ray Co-Crystal Structures of Cyclohexylpyrrolidine Ligands for Trypanothione Reductase, an Enzyme from the Redox Metabolism of Trypanosoma. *ChemMedChem* 13 (9), 957–967. doi:10.1002/cmdc.201800067
- De Gasparo, R., Halgas, O., Harangozo, D., Kaiser, M., Pai, E. F., Krauth-Siegel, R. L., et al. (2019). Targeting a Large Active Site: Structure-Based Design of Nanomolar Inhibitors of Trypanosoma Brucei Trypanothione Reductase. *Chemistry* 25 (49), 11416–11421. doi:10.1002/chem.201901664
- Field, M. C., Horn, D., Fairlamb, A. H., Ferguson, M. A. J., Gray, D. W., Read, K. D., et al. (2017). Anti-trypanosomatid Drug Discovery: an Ongoing Challenge and a Continuing Need. *Nat. Rev. Microbiol.* 15 (4), 217–231. doi:10.1038/nrmicro.2016.193
- Frearson, J. A., Wyatt, P. G., Gilbert, I. H., and Fairlamb, A. H. (2007). Target Assessment for Antiparasitic Drug Discovery. *Trends Parasitol.* 23 (12), 589–595. doi:10.1016/j.pt.2007.08.019
- Glennon, R. A., Salley, J. J., Steinsland, O. S., and Nelson, S. (1981). Synthesis and Evaluation of Novel Alkylpiperazines as Potential Dopamine Antagonists. *J. Med. Chem.* 24 (6), 678–683. doi:10.1021/jm00138a007
- Ilari, A., Genovese, I., Fiorillo, F., Battista, T., De Ionna, I., Fiorillo, A., et al. (2018). Toward a Drug against All Kinetoplastids: From LeishBox to Specific and Potent Trypanothione Reductase Inhibitors. *Mol. Pharm.* 15 (8), 3069–3078. doi:10.1021/acs.molpharmaceut.8b00185
- Karplus, P. A., Pai, E. F., and Schulz, G. E. (1989). A Crystallographic Study of the Glutathione Binding Site of Glutathione Reductase at 0.3-nm Resolution. *Eur. J. Biochem.* 178 (3), 693–703. doi:10.1111/j.1432-1033.1989.tb14500.x
- Khan, M. O. F., Austin, S. E., Chan, C., Yin, H., Marks, D., Vaghjiani, S. N., et al. (2000). Use of an Additional Hydrophobic Binding Site, the Z Site, in the Rational Drug Design of a New Class of Stronger Trypanothione Reductase Inhibitor, Quaternary Alkylammonium Phenothiazines. *J. Med. Chem.* 43 (16), 3148–3156. doi:10.1021/jm000156+
- Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., et al. (2015). The FTMap Family of Web Servers for Determining and Characterizing Ligand-Binding Hot Spots of Proteins. *Nat. Protoc.* 10 (5), 733–755. doi:10.1038/nprot.2015.043
- Krieger, S., Schwarz, W., Ariyanayagam, M. R., Fairlamb, A. H., Krauth-Siegel, R. L., and Clayton, C. (2000). Trypanosomes Lacking Trypanothione Reductase Are Avirulent and Show Increased Sensitivity to Oxidative Stress. *Mol. Microbiol.* 35 (3), 542–552. doi:10.1046/j.1365-2958.2000.01721.x
- Krojer, T., Talon, R., Pearce, N., Collins, P., Douangamath, A., Brandao-Neto, J., et al. (2017). The XChemExplorer Graphical Workflow Tool for Routine or Large-Scale Protein-Ligand Structure Determination. *Acta Crystallogr. D. Struct. Biol.* 73 (Pt 3), 267–278. doi:10.1107/S2059798316020234
- Li, Y., Ye, T., Xu, L., Dong, Y., Luo, Y., Wang, C., et al. (2019). Discovery of 4-Piperazinyl-2-Aminopyrimidine Derivatives as Dual Inhibitors of JAK2 and FLT3. *Eur. J. Med. Chem.* 181, 111590. doi:10.1016/j.ejmech.2019.111590
- McNaughton, B. R., Gareiss, P. C., Jacobs, S. E., Fricke, A. F., Scott, G. A., and Miller, B. L. (2009). A Potent Activator of Melanogenesis Identified from Small-Molecule Screening. *ChemMedChem* 4 (10), 1583–1589. doi:10.1002/cmdc.200900194
- Murray, C. W., and Rees, D. C. (2009). The Rise of Fragment-Based Drug Discovery. *Nat. Chem.* 1 (3), 187–192. doi:10.1038/nchem.217
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., et al. (2011). REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. D. Biol. Crystallogr.* 67 (Pt 4), 355–367. doi:10.1107/S0907444911001314
- Patterson, S., Alphey, M. S., Jones, D. C., Shanks, E. J., Street, I. P., Frearson, J. A., et al. (2011). Dihydroquinazolines as a Novel Class of Trypanosoma Brucei Trypanothione Reductase Inhibitors: Discovery, Synthesis, and Characterization of Their Binding Mode by Protein Crystallography. *J. Med. Chem.* 54 (19), 6514–6530. doi:10.1021/jm200312v
- Pearce, N. M., Krojer, T., Bradley, A. R., Collins, P., Nowak, R. P., Talon, R., et al. (2017). A Multi-Crystal Method for Extracting Obscured Crystallographic States from Conventionally Uninterpretable Electron Density. *Nat. Commun.* 8, 15123. doi:10.1038/ncomms15123
- Revuelto, A., de Lucio, H., García-Soriano, J. C., Sánchez-Murcia, P. A., Gago, F., Jiménez-Ruiz, A., et al. (2021). Efficient Dimerization Disruption of Leishmania Infantum Trypanothione Reductase by Triazole-Phenyl-Thiazoles. *J. Med. Chem.* 64 (9), 6137–6160. doi:10.1021/acs.jmedchem.1c00206
- Ruiz-Santaquiteria, M., Sánchez-Murcia, P. A., Toro, M. A., de Lucio, H., Gutiérrez, K. J., de Castro, S., et al. (2017). First Example of Peptides Targeting the Dimer Interface of Leishmania Infantum Trypanothione Reductase with Potent *In Vitro* Antileishmanial Activity. *Eur. J. Med. Chem.* 135, 49–59. doi:10.1016/j.ejmech.2017.04.020
- Salinas, G. (2013). An Update on Redox Biology of Parasites. *Antioxid. Redox Signal.* 19 (7), 661–664. doi:10.1089/ars.2013.5348
- Saravanamuthu, A., Vickers, T. J., Bond, C. S., Peterson, M. R., Hunter, W. N., and Fairlamb, A. H. (2004). Two Interacting Binding Sites for Quinacrine Derivatives in the Active Site of Trypanothione Reductase. *J. Biol. Chem.* 279 (28), 29493–29500. doi:10.1074/jbc.m403187200
- Savvides, S. N., and Karplus, P. A. (1996). Kinetics and Crystallographic Analysis of Human Glutathione Reductase in Complex with a Xanthene Inhibitor. *J. Biol. Chem.* 271 (14), 8101–8107. doi:10.1074/jbc.271.14.8101
- Schönleben-Janasz, A., Kirsch, P., Mittl, P. R., Schirmer, R. H., and Krauth-Siegel, R. L. (1996). Inhibition of Human Glutathione Reductase by 10-arylisoalloxazines: Crystalline, Kinetic, and Electrochemical Studies. *J. Med. Chem.* 39 (7), 1549–1554. doi:10.1021/jm950511+
- Silvestri, I., Lyu, H., Fata, F., Boumis, G., Miele, A. E., Ardini, M., et al. (2018). Fragment-Based Discovery of a Regulatory Site in Thioredoxin Glutathione Reductase Acting as "Doorstop" for NADPH Entry. *ACS Chem. Biol.* 13 (8), 2190–2202. doi:10.1021/acschembio.8b00349
- Turcano, L., Torrente, E., Missineo, A., Andreini, M., Gramiccia, M., Di Muccio, T., et al. (2018). Identification and Binding Mode of a Novel Leishmania Trypanothione Reductase Inhibitor from High Throughput Screening. *PLoS Negl. Trop. Dis.* 12 (11), e0006969. doi:10.1371/journal.pntd.0006969
- Turcano, L. B. T., Torrente De Haro, E., Missineo, A., Alli, C., Paonessa, G., Colotti, G., et al. (2020). Spiro-Containing Derivatives Show Antiparasitic Activity against Trypanosoma Brucei through Inhibition of the Trypanothione Reductase Enzyme. *Negl. Trop. Dis.* 14 (5), e0008339. doi:10.1371/journal.pntd.0008339
- WHO (2022a). WHO Publishes On-Line Key Information about Leishmaniasis. Available online: <https://www.who.int/news-room/fact-sheets/detail/leishmaniasis> (Accessed March, 2022).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.900882/full#supplementary-material>



WHO (2022b). WHO Publishes On-Line Key Information about Chagas Disease. Available online: [https://www.who.int/news-room/fact-sheets/detail/chagas-disease-\(american-trypanosomiasis\)](https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis)) (Accessed March, 2022).

WHO (2022c). WHO Publishes On-Line Key Information about HAT. Available online: [https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-\(sleeping-sickness\)](https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness)) (Accessed March, 2022).

WHO (2022d). Neglected Tropical Diseases. Available online: [https://www.who.int/health-topics/neglected-tropical-diseases#tab=tab\\_1](https://www.who.int/health-topics/neglected-tropical-diseases#tab=tab_1) (Accessed March, 2022).

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

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

*Copyright © 2022 Fiorillo, Colotti, Exertier, Liuzzi, Seghetti, Salerno, Caciolla and Ilari. 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.*



## OPEN ACCESS

EDITED BY  
Gianluca Molla,  
University of Insubria, Italy

REVIEWED BY  
Aiwu Zhou,  
Shanghai Jiao Tong University, China  
Teresita Padilla-Benavides,  
Wesleyan University, United States  
Gabi U. Dachs,  
University of Otago, Christchurch,  
New Zealand

\*CORRESPONDENCE  
Luigi Scietti,  
luigi.scietti@unipv.it  
Federico Forneris,  
federico.forneris@unipv.it

<sup>†</sup>Present address  
Luigi Scietti, Biochemistry and Structural  
Biology Unit, Department of  
Experimental Oncology, IRCCS  
European Institute of Oncology (IEO),  
Milan, Italy.

<sup>‡</sup>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 15 February 2022  
ACCEPTED 11 July 2022  
PUBLISHED 25 August 2022

CITATION  
Scietti L, Moroni E, Mattoteia D, Fumagalli M,  
De Marco M, Negro L, Chiapparino A,  
Serapian SA, De Giorgi F, Faravelli S,  
Colombo G and Forneris F (2022) A Fe<sup>2+</sup>-  
dependent self-inhibited state influences the  
druggability of human collagen lysyl  
hydroxylase (LH/PLOD) enzymes.  
*Front. Mol. Biosci.* 9:876352.  
doi: 10.3389/fmolb.2022.876352

COPYRIGHT  
© 2022 Scietti, Moroni, Mattoteia,  
Fumagalli, De Marco, Negro, Chiapparino,  
Serapian, De Giorgi, Faravelli, Colombo and  
Forneris. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# A Fe<sup>2+</sup>-dependent self-inhibited state influences the druggability of human collagen lysyl hydroxylase (LH/PLOD) enzymes

Luigi Scietti<sup>1\*†‡</sup>, Elisabetta Moroni<sup>2‡</sup>, Daiana Mattoteia<sup>1‡</sup>,  
Marco Fumagalli<sup>1</sup>, Matteo De Marco<sup>1</sup>, Lisa Negro<sup>1</sup>,  
Antonella Chiapparino<sup>1</sup>, Stefano A. Serapian<sup>3</sup>,  
Francesca De Giorgi<sup>1</sup>, Silvia Faravelli<sup>1</sup>, Giorgio Colombo<sup>3</sup> and  
Federico Forneris<sup>1\*</sup>

<sup>1</sup>The Armenise-Harvard Laboratory of Structural Biology, Department of Biology and Biotechnology, University of Pavia, Pavia, Italy, <sup>2</sup>Consiglio Nazionale delle Ricerche, Istituto di Scienze e Tecnologie Chimiche "Giulio Natta" (SCITEC-CNR), Milano, Italy, <sup>3</sup>Department of Chemistry, University of Pavia, Pavia, Italy

Multifunctional human collagen lysyl hydroxylase (LH/PLOD) enzymes catalyze post-translational hydroxylation and subsequent glycosylation of collagens, enabling their maturation and supramolecular organization in the extracellular matrix (ECM). Recently, the overexpression of LH/PLODs in the tumor microenvironment results in abnormal accumulation of these collagen post-translational modifications, which has been correlated with increased metastatic progression of a wide variety of solid tumors. These observations make LH/PLODs excellent candidates for prospective treatment of aggressive cancers. The recent years have witnessed significant research efforts to facilitate drug discovery on LH/PLODs, including molecular structure characterizations and development of reliable high-throughput enzymatic assays. Using a combination of biochemistry and *in silico* studies, we characterized the dual role of Fe<sup>2+</sup> as simultaneous cofactor and inhibitor of lysyl hydroxylase activity and studied the effect of a promiscuous Fe<sup>2+</sup> chelating agent, 2,2'-bipyridil, broadly considered a lysyl hydroxylase inhibitor. We found that at low concentrations, 2,2'-bipyridil unexpectedly enhances the LH enzymatic activity by reducing the inhibitory effect of excess Fe<sup>2+</sup>. Together, our results show a fine balance between Fe<sup>2+</sup>-dependent enzymatic activity and Fe<sup>2+</sup>-induced self-inhibited states, highlighting exquisite differences between LH/PLODs and related Fe<sup>2+</sup>, 2-oxoglutarate dioxygenases and suggesting that conventional structure-based approaches may not be suited for successful inhibitor development. These insights address outstanding questions regarding druggability of LH/PLOD lysyl hydroxylase catalytic site and provide a solid ground for upcoming drug discovery and screening campaigns.

## KEYWORDS

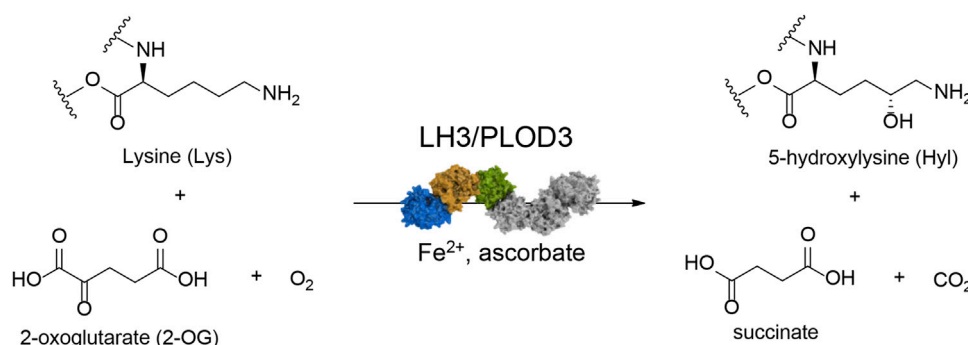
collagen, lysyl hydroxylase (LH), Fe<sup>2+</sup>/2-oxoglutarate-dependent dioxygenases, structure-based drug design, molecular dynamics simulations, cancer metastasis, 2-2'-bipyridil

## Introduction

The supramolecular organization of collagen in the extracellular matrix (ECM) depends on various post-translational modifications (PTMs) that occur during its biosynthesis. Among the different PTMs, lysine (Lys) hydroxylation is key for proper collagen fibril formation, thus defining the overall physiochemical properties of ECM (Yamauchi and Sricholpech, 2012). The collagen lysyl hydroxylase (LH/PLOD) enzyme family comprises the three isoforms LH1/PLOD1, LH2/PLOD2 and LH3/PLOD3 (encoded by the procollagen-lysine, 2-oxoglutarate 5-dioxygenase (*PLOD*) genes) and is the sole enzyme capable of hydroxylating collagen Lys in humans (Sciatti and Forneris, 2020). These enzymes use  $\text{Fe}^{2+}$ , 2-oxoglutarate (2-OG), ascorbate and  $\text{O}_2$  to catalyze the addition of a hydroxyl group in position 5 of collagen Lys, yielding 5-hydroxylysine (Hyl) with the release of succinate and  $\text{CO}_2$  (Figure 1). Unmodified collagen Lys and modified Hyl are both substrates of collagen lysyl oxidases (LOX), which catalyze the oxidative deamination of Lys and Hyl forming highly reactive aldehydes ( $\text{Lys}^{\text{ald}}$  and  $\text{Hyl}^{\text{ald}}$ , respectively) that spontaneously rearrange to form Lys-derived collagen cross-links (LCC) and Hyl-derived collagen cross-links (HLCC) in the ECM. A physiological ratio between LCC and HLCC is essential to establish and maintain a proper ECM functionality. Conversely, excess HLCC in the tumor microenvironment has been linked to biomechanical alterations and increased ECM tension and stiffness (Levental et al., 2009; Pankova et al., 2016). The deposition of ordered thicker collagen fibers, as consequence of the abnormal LCC/HLCC ratio, is a characteristic of severe tissue fibrosis, one of the hallmarks of cancer (Van Der Slot et al., 2004; Chen et al., 2015). Cancer cells take advantage of these “collagen highways” to migrate toward blood vessels that sustain metastatic progression (Provenzano et al., 2006; Du et al., 2017b; Gkretsi

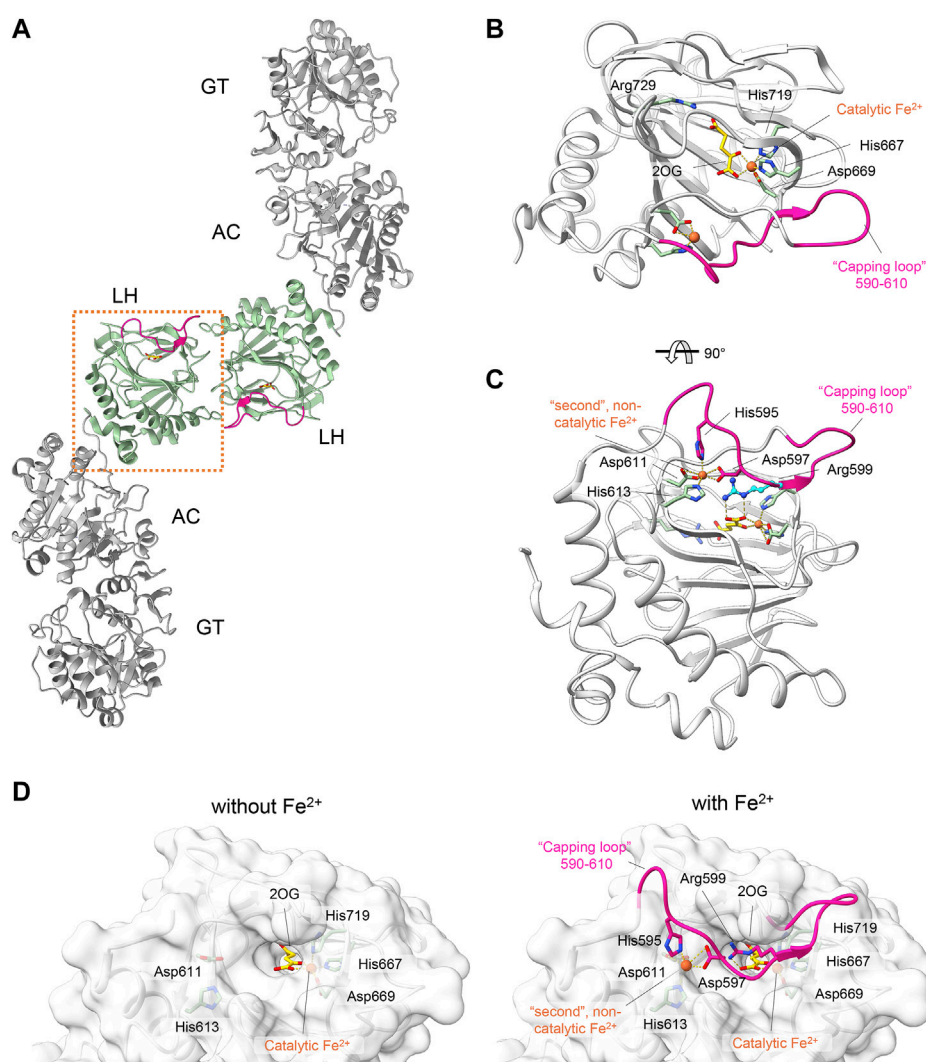
and Stylianopoulos, 2018). Over the last years, multiple studies correlated both hypoxia-dependent and independent overexpression and mislocalization of LH enzymes with increased propensity to metastatization in a wide variety of solid tumors, recognizing these enzymes as markers of adverse prognosis (Chen et al., 2015; Chen et al., 2016; Pankova et al., 2016; Sato et al., 2021).

Initially, the hypoxia-driven overexpression of *PLOD2* was the first identified prognostic factor in several tumors, as hepatocellular carcinoma (Noda et al., 2012; Du et al., 2017a), sarcoma (Eisinger-Mathason et al., 2013), lung and colon cancer (Du et al., 2017a), renal carcinoma (Kurozumi et al., 2016), glioma (Song et al., 2017; Xu et al., 2017), oral squamous cell and endometrial carcinoma (Saito et al., 2019; Wan et al., 2020) bone and breast metastasis (Blanco et al., 2012; Gilkes et al., 2013; Du et al., 2017b) and cervical cancer (Li et al., 2021). Interestingly, the downregulation of LH2/PLOD2 isoform in renal cell carcinoma via tumor suppressing miRNA significantly inhibited cell migration and invasion (Kurozumi et al., 2016), confirming the importance of LH in tumor progression. Later on, the LH1/PLOD1 and the LH3/PLOD3 isoforms were also identified as biomarkers in many different types of solid tumors. High *PLOD1* expression levels were found in gastrointestinal carcinoma (Wang et al., 2018), osteosarcoma (Jiang et al., 2020), glioma (Tian et al., 2021; Wang et al., 2021) and bladder cancer (Yamada et al., 2019). The LH3/PLOD3 isoform was identified to be upregulated in glioma (Tsai et al., 2018; Baek et al., 2019), gastric cancer (Wang et al., 2019) and colorectal cancer (Deng et al., 2021; Shi et al., 2021), acting as a promoter of metastatization in different cancer types (Gong et al., 2021). In agreement with the observations on *PLOD2*, also *PLOD3* knockdown suppressed the malignant phenotype in renal cell carcinoma (Xie et al., 2020). Furthermore, the entire LH/PLOD family was found correlated with metastatization of solid tumors as



**FIGURE 1**

Reaction scheme for collagen lysine hydroxylation by LH3/PLOD3. The enzyme requires  $\text{Fe}^{2+}$  as an essential catalytic cofactor,  $\text{O}_2$ , and 2-oxoglutarate (2-OG) as co-substrates, and ascorbate to maintain the redox state of  $\text{Fe}^{2+}$  to enable catalytic turnover. Upon lysine hydroxylation, 2-OG is converted into succinate with release of one molecule of  $\text{CO}_2$ .

**FIGURE 2**

Structural features of the LH domain and catalytic site in LH3/PLOD3. **(A)** Cartoon representation of the dimeric assembly of multidomain LH3/PLOD3. The labels indicate the domain organization of LH3/PLOD3, showing the positions of the GT, AC, and LH domains along the quaternary structure. The box highlights the position of the LH domain (green) at the dimer interface. **(B)** Details of the LH catalytic site. Amino acids surrounding the  $\text{Fe}^{2+}$  cofactor and 2-OG are shown as sticks. The capping loop is shown in magenta, indicating the substrate entrance site. **(C)** Details of the second  $\text{Fe}^{2+}$  binding site interlocking the capping loop into the self-inhibitory state. Residues interacting with the second  $\text{Fe}^{2+}$  as shown as sticks. Residue Arg599, mimicking the collagen lysine substrate subject to hydroxylation, is shown in blue. **(D)** Side-by-side surface representations of the accessibility to the LH active site in absence (left) and in presence (right) of the capping loop. Residues involved in catalysis and in maintaining the self-inhibitory state of the capping loop are shown as in panels **(B,C)**.

hepatocellular and renal cell carcinomas (Xu et al., 2019; Yang et al., 2020), gliomas (Zhao et al., 2021), gastric (Li et al., 2020), and ovarian (Guo T. et al., 2021) cancers. Taken together, these observations strongly point these enzymes as a hot topic in cancer research: LH/PLODs are not only widely recognized prognostic markers of cancer metastatization with poor outcome, but also very promising druggable targets for anticancer therapy.

The lack of a structural templates of LH/PLODs have hampered for many years targeted drug discovery campaigns.

Only recently, our group and others determined molecular structures suitable as templates for *in silico* drug discovery (Guo et al., 2018; Sciatti et al., 2018). In particular, the crystal structure of the full-length LH3/PLOD3, the first of a human collagen lysyl hydroxylase, provided key insights on catalytic pockets. LH3/PLOD3 is a multifunctional enzyme capable of performing collagen lysine hydroxylation (as the other two isoforms) and glycosylation (Sciatti and Forneris, 2020; De Giorgi et al., 2021). Indeed, LH3/PLOD3 can additionally catalyze the galactosylation and further glucosylation of Hyl to



form  $\alpha$ -(1,2)-glucosyl- $\beta$ -(1,0)-galactosyl-5-hydroxylysine *in vitro*. The identification of specific genes (*COLGALT1/2*) encoding for collagen galactosyltransferases (*GLT25D1/2*) makes the physiological relevance of the galactosyltransferase activity of LH3/PLOD3 under debate. In addition, there are increasing indications that also LH1/PLOD1 and LH2/PLOD2 possess glycosyltransferase activity, although less pronounced (Ewans et al., 2019; Guo H. F. et al., 2021).

LH/PLOD enzymes belong to the  $\text{Fe}^{2+}$ , 2-OG-dependent dioxygenase superfamily (Martinez and Hausinger, 2015), a widespread class of enzymes that catalyzes oxidative reactions such as epimerization, demethylation, and hydroxylation (Hausinger, 2004; Flashman and Schofield, 2007; Loenarz and Schofield, 2008; Loenarz and Schofield, 2011). Despite the broad range of functions carried out, all  $\text{Fe}^{2+}$ , 2-OG-dependent dioxygenases display a common double-stranded  $\beta$ -helix folding (DSBH) topology with highly conserved binding sites and catalytic mechanisms (Costas et al., 2004; Clifton et al., 2006). The catalytic domain responsible for Lys hydroxylation is located at the C-terminus of the LH/PLOD structure and is essential for its unique dimeric quaternary structure, a fundamental prerequisite for collagen lysyl hydroxylase activity (Guo et al., 2018; Sciatti et al., 2018). Within the enzymatic pocket of human LH3/PLOD3, a catalytic  $\text{Fe}^{2+}$  is coordinated by His667, Asp669 and His719 (Figures 2A,B). In a LH/PLOD viral homolog, this  $\text{Fe}^{2+}$  is a fundamental structural element of LH/PLOD enzymes and its chelation from the active site completely disrupt protein folding, dimer formation and catalytic activity (Guo et al., 2018). Pioneering work (Kivirikko and Prockop, 1967; Myllylä et al., 1979; Puistola et al., 1980) highlighted the complexity and binding promiscuity of LH/PLODs towards binding of different metal ions, and the associated impact on the LH enzymatic activity. Structural studies also revealed the presence of a possible second  $\text{Fe}^{2+}$  bound within the LH domain, shaping a unique site never observed in other  $\text{Fe}^{2+}$ , 2-OG-dependent dioxygenases. This second  $\text{Fe}^{2+}$  is coordinated by two Asp and two His residues (in human LH3/PLOD3, His595, Asp597, Asp611 and His613), whose side chain interactions with the metal ion induce a well-defined conformation of a “capping loop”, a stretch comprising residues Gly590-Glu610 that closes the entrance of the catalytic site, mimicking the collagen Lys substrate by positioning Arg599 exactly in front of the 2-OG donor substrate (Sciatti et al., 2018) (Figures 2C,D). The same region was characterized by pronounced flexibility in absence of  $\text{Fe}^{2+}$  (Guo et al., 2018; Sciatti et al., 2018), but did not allow to unambiguously rule out crystallization-induced stabilization of the unique conformation observed in the presence of a second  $\text{Fe}^{2+}$  bound.

The simultaneous presence of features common to all  $\text{Fe}^{2+}$ , 2-OG-dependent dioxygenases and unique, distinguishing elements exclusively present in the LH/

TABLE 1 List of oligonucleotides used to generate LH3/PLOD3 mutants.

Oligonucleotide name	Oligonucleotide sequence
D597A-Fw	CTTCAAGGCTGGCTGGAGGCTAC
D597A-Rv	CCTCATGCCGCCGCGCTGAC
D611A-Fw	CCATCCACATGAAGCAGGTGGGG
D611A-Rv	CCACGGTGGGCACATTCTCGTAG

PLOD family, makes these enzymes ideal targets for structure-based drug discovery campaigns. In this study, we combined MD simulations with structure-guided mutagenesis of LH3/PLOD3 and used biochemical assays to elucidate the role of the capping loop in the accessibility of the active site. Our work sets the grounds for successful drug discovery campaigns on LH/PLOD enzymes to fight cancer metastasis.

## Materials and methods

### Chemicals

All chemicals were purchased from Sigma-Aldrich (Merck) unless specified otherwise.

### Molecular cloning and site-directed mutagenesis

The coding sequence for wild-type human *PLOD3* gene (GenBank accession number BC011674.2) was obtained from Source Bioscience. Oligonucleotides containing in-frame 5'-BamHI and 3'-NotI were designed and used to sub-clone the coding sequence devoid of the N-terminal signal peptide into a pCR8 vector, that was also used as a template for subsequent experiments. Single-point mutations were generated using Phusion Site Directed Mutagenesis (Invitrogen) with the oligonucleotides listed in Table 1. The linear mutagenized plasmids were phosphorylated using T4 polynucleotide kinase (Invitrogen) prior to ligation using T4 DNA ligase (Invitrogen). All plasmids were checked by Sanger sequencing prior to cloning into the pUPE.106.08 expression vector. This expression vector, kindly provided by U-protein Express, BV (U-PE, Netherlands) provides the N-terminal cystatin signal peptide, followed by a N-terminal 6xHis-tag and a recognition site for Tobacco Etch Virus (TEV) protease prior to the in-frame BamHI-NotI restriction cassette, followed by an in-frame stop codon.

## Production of recombinant LH3/PLOD3 expression using transiently-transfected HEK293F cells

Recombinant tagged LH3/PLOD3 were produced using suspension cultures of HEK293F (Invitrogen) cells, maintained and transfected according to (Faravelli et al., 2021). Cells were not authenticated and not tested for *mycoplasma* contamination. Briefly, cells were transfected at cell densities of 1 million/ml using 3 µg of polyethylenimine (PEI; Polysciences) for 1 µg of pUPE.106.08-LH3/PLOD3 plasmid DNA per mL of cells. Cultures were supplemented with 0.6% Primatone RL 4 h after transfection. The cell medium containing secreted LH3/PLOD3 was collected 6 days after transfection by centrifugation at  $1,000 \times g$  for 15 min.

## Purification of recombinant LH3/PLOD3 enzymes

The LH3/PLOD3-containing medium from HEK293F cell cultures was filtered through a syringe 0.8 µm filter (Sartorius). The pH and ionic strength of the filtered medium were adjusted using a 5X concentrated buffer stock to reach a final concentration of 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)/NaOH, 500 mM NaCl, 30 mM imidazole, pH 8.0. Recombinant LH3/PLOD3 was purified using a combination of affinity and size-exclusion chromatography on Äkta systems (GE Healthcare) according to (Sciatti et al., 2018). The filtered supernatant was first loaded onto a 20 ml His-Prep FF column (GE Healthcare) and eluted using 250 mM imidazole. The eluate was then loaded onto a 5 ml HiPrep desalting FF column (GE Healthcare) equilibrated in 25 mM HEPES/NaOH, 500 mM NaCl, pH 8.0. The N-terminal His-tag was cleaved using overnight His-tagged TEV protease digestion at 4°C followed by affinity-based removal of TEV protease and the cleaved His-tag using a 5 ml HisTrap FF (GE Healthcare). Recombinant LH3 was concentrated to 5 mg/ml using 30,000 MWCO Vivaspins Turbo centrifugal filters (Sartorius), then loaded onto a Superdex 200 10/300 GL (preparative scale) or onto a Superdex 200 5/150 GL (analytical scale) columns (GE Healthcare) equilibrated with 25 mM HEPES/NaOH, 200 mM NaCl, pH 8.0. LH3/PLOD3-containing fractions as assessed from SDS-PAGE analysis were pooled, concentrated, and stored at  $-80^{\circ}\text{C}$  until further usage.

## LH assays using LC-MS and analysis of 2,2'-bipyridil effects on enzymatic activity

Synthetic collagen peptides were purchased from China Peptides. Peptides tested were ARGIKGIRGFS and GIKGIKGIKGIK sequences (Sciatti et al., 2018). 5 µM LH3/

PLOD3 was incubated with 50 µM  $\text{FeCl}_2$ , 100 µM 2-OG, 500 µM ascorbate, 1 mM peptide substrate and 0–500 µM 2,2'-bipyridine. Reactions were allowed to proceed for 3 h at  $37^{\circ}\text{C}$ . 10 µl of each sample were supplemented with 38 µl of Milli-Q water and acidified by addition of 2 µl of formic acid (FA) to reach a final volume of 50 µl, then analyzed on an UHPLC-HRMS/MS system (AB Sciex, United States). LC unit (ExionLC AD) consists of a column oven thermostated at  $40^{\circ}\text{C}$ , an autosampler cooled at  $10^{\circ}\text{C}$  and a binary gradient pump system. MS instrument consists of a high resolution QTOF mass spectrometer (AB Sciex X500B) equipped with a Turbo V Ion source and a Twin Sprayer ESI (electrospray ionization) probe, controlled by SCIEX OS 2.1 software. Peptides were separated by reverse phase (RP) HPLC on a Hypersil Gold C18 column ( $150 \times 2.1$  mm, 3 µm particle size, 175 Å pore size, Thermo Fisher Scientific) using a linear gradient (2–50% solvent B in 15 min) in which solvent A consisted of 0.1% aqueous FA and solvent B of acetonitrile (CAN) containing 0.1% FA. Flow rate was 0.2 ml/min. Mass spectra were generated in positive polarity under constant instrumental conditions: ion spray voltage 4,500 V, declustering potential 100 V, curtain gas 30 psi, ion source gas 1 40 psi, ion source gas 2 45 psi, temperature  $350^{\circ}\text{C}$ , collision energy 10 V. Spectra analyses were performed using SCIEX OS 2.1 software. Statistical evaluations based on pair sample comparisons between uncoupled and coupled assay values using Student's *t*-test in Prism 7 (Graphpad software).

## Luminescence-based LH assays

Reaction mixtures (5 µl total volume) were prepared according to (Sciatti et al., 2018) by sequentially adding LH3/PLOD3 at 0.2 mg/ml, 0–1 mM peptide substrate or 4 mg/ml gelatin in water (solubilized through heating denaturation at  $95^{\circ}\text{C}$  for 10 min), 500 µM ascorbate, 100 µM 2-OG, variable concentrations of  $\text{FeCl}_2$  (0–200 µM) and let incubate for 1 h at  $37^{\circ}\text{C}$ . Reactions were stopped by heating samples at  $95^{\circ}\text{C}$  for 2 min prior to transfer into Proxiplate white 384-well plates (Perkin-Elmer), then 5 µl of the Succinate-Glo reagent I (Promega) were added and let incubate 1 h at  $25^{\circ}\text{C}$ , after that 10 µl of the Succinate-Glo reagent II (Promega) were added and let incubate 10 min at  $25^{\circ}\text{C}$ . The plates were then transferred into a GloMax Discovery plate reader (Promega) configured according to manufacturer's instructions for luminescence detection. All experiments were performed in triplicates. Control experiments were performed using identical conditions by selectively removing LH3/PLOD3, 2-OG or peptide substrates. Data were analyzed and plotted using Prism 7 (Graphpad Software). Statistical evaluations based on pair sample comparisons between uncoupled and coupled

assay values using Student's t-test in Prism 7 (Graphpad software).

## Differential scanning fluorimetry assays

DSF assays were performed on LH3/PLOD3 wild-type using a Tycho NT.6 instrument (NanoTemper Technologies). LH3/PLOD3 samples at a concentration of 1 mg/ml in a buffer composed of 25 mM HEPES/NaOH, 200 mM NaCl, pH 8. Binding assays were performed by incubating LH3/PLOD3 with variable FeCl<sub>2</sub> and 2,2'-bipyridil concentrations. Data were analyzed and plotted using GraphPad Prism 7 (Graphpad Software).

## Fe<sup>2+</sup> binding assays

Recombinant LH3/PLOD3 was subject to labeling using the NHS-RED kit (NanoTemper Technologies) according to manufacturer's instructions. Labeled LH3/PLOD3 at a concentration of 50 nM was incubated in a buffer composed of 25 mM TRIS/HCl, 100 mM NaCl, pH 7.5 with variable concentrations of FeCl<sub>2</sub> for 40 min. The samples were then transferred into Dianthus 384-well plates for Temperature-Related Intensity Change (TRIC) (NanoTemper Technologies), and centrifuged at 1,000 g for 2 min. TRIC measurements were performed immediately after centrifugation using a Dianthus NT.23 instrument (NanoTemper Technologies). The samples were first measured for 1 s without heating and for 5 s with the IR-laser turned on. Normalized fluorescence values ( $F_{norm}$ ), described as ratios between fluorescence values after and prior to infrared laser activation were collected and plotted as a function of ligand concentration (Schulte et al., 2021). Determination of binding affinities was carried out using the DI.SCREENING Analysis software (NanoTemper Technologies). Data were then exported and plotted using GraphPad Prism 7 (Graphpad Software).

## Molecular dynamics simulations

Simulations were started from the extrapolation of the dimeric LH domain of human LH3/PLOD3 from its experimental crystal structure (PDB 6FXR) (Sciatti et al., 2018) using COOT (Emsley et al., 2010); two systems were prepared: in one system both Fe<sup>2+</sup> cations were left (LH3<sub>Fe2</sub>), while in the other one only the catalytic Fe<sup>2+</sup> was left in (LH3<sub>Fe1</sub>). Residues were modeled in their standard protonation states at physiological pH, as predicted by PROPKA, version 3.1 (Sondergaard et al., 2011): this resulted in one disulfide bridge (between Cys 563 and 698), histidines 546, 586, 643, 681,

717 being protonated on Nε2, and histidines 595, 613, 667, 711, 719 protonated on Nδ1 (comprising histidines in both Fe<sup>2+</sup> binding sites). Crystallographic waters were taken from the published crystal structure of LH3/PLOD3 (PDB: 6FXR) (Sciatti et al., 2018). Hydrogen atoms were introduced using the *tleap* utility in *AmberTools* (version 19) (Case et al., 2005), as well as -NH<sub>3</sub><sup>+</sup> and -COO<sup>-</sup> caps at the N- and C-termini, respectively. All molecular dynamics simulations (MD) were carried out with the *AMBER* software package (version 18) (Case et al., 2005; Case et al., 2018), using its GPU-accelerated (Salomon-Ferrer et al., 2013) *pmemd.cuda* utility during equilibration and production, and *sander* otherwise; three independent MD replicas (different random seeds) were carried out for LH3<sub>Fe1</sub> and LH3<sub>Fe2</sub> alike. A 8.0 Å cutoff was applied for the calculation of Lennard-Jones and Coulomb interactions between nonbonded atoms; beyond this limit, only Coulomb interactions were computed, using the particle mesh Ewald approach (Darden et al., 1993). Each replica's production stage was 1 μs in length, and conducted with a 2 fs time-step in the *NpT* ensemble (with a temperature of 300 K enforced via Langevin's thermostat (Loncharich et al., 1992); collision frequency 1 ps<sup>-1</sup>, and a 1 atm pressure enforced by Berendsen's barostat (Berendsen et al., 1984). Preproduction stages for each replica consisted in minimization (10 steps of steepest descent + 290 steps of conjugate gradient); heating (20 ps; *NpT*; 25–300 K; increasingly softer harmonic restraints on Ca atoms;  $k = 5.0$  kcal mol<sup>-1</sup> Å, collision frequency 0.75 ps<sup>-1</sup>, with 2 fs time-step); and equilibration (1.0 ns; *NpT*; 300 K collision frequency 1 ps<sup>-1</sup>, with 2 fs time-step). Analyses of MD trajectories were carried out with the *CPPTRAJ* program distributed within the *AmberTools* suite (version 19) (Case et al., 2005) or with code written in-house.

## Distance fluctuation

To characterize the impact of the second Fe<sup>2+</sup> on the internal dynamics of LH domain of LH3, we made use of the previously introduced distance fluctuation (DF) analysis (Morra et al., 2012; Moroni et al., 2018). For each MD trajectory of the two systems, we computed on the combined meta trajectory the matrix of distance fluctuations, in which each element of the matrix corresponds to the DF parameter. DF is defined, for a couple of amino acids *i* and *j*, as the variance of the time-dependent distance  $d_{ij}$  of the C α atoms:

$$DF_{ij} = \langle (d_{ij} - \langle d_{ij} \rangle)^2 \rangle$$

where the brackets indicate the time-average over the trajectory. This parameter is invariant under translations and rotations of the molecules and, unlike the covariance matrix, does not depend on the choice of a particular protein reference structure. The resulting DF matrix can be used to assess the intrinsic flexibility of proteins. This parameter characterizes residues that move in a

highly coordinated fashion, and it is actually able to reflect the presence of specific coordination patterns and quasi-rigid domains motion in the protein of interest. In particular, pairs of amino acids belonging to the same quasi-rigid domain are associated with small distance fluctuations and vice versa.

## Forcefield and parametrization of cofactor 2-OG and Fe<sup>2+</sup> binding sites

Lennard-Jones and intramolecular bonded parameters for the 2-oxoglutaric acid cofactor (2-OG) in the catalytic site of LH3/PLD3 were assigned according to the *generalized Amber forcefield* (GAFF) (Wang et al., 2004), using *AmberTools'* *antechamber* and *parmchk2* utilities (Case et al., 2005), after adding methyl hydrogens using the *reduce* tool (Case et al., 2005). Assignment of charges and (intermolecular) 2-OG-Fe<sup>2+</sup> bonded parameters were performed as discussed below. Parametrization of both Fe<sup>2+</sup> binding sites was carried out using the *MCPB.py* utility (Li and Merz, 2016) in conjunction with density functional theory calculations (DFT) using the *Gaussian09* suite (Frisch et al., 2009). Intermolecular bonded parameters for all residues in both Fe<sup>2+</sup>-binding sites and for the 2-OG cofactor were derived by *MCPB.py* (Li and Merz, 2016) applying the Seminario method (Seminario, 1996) on a DFT-derived Hessian matrix. More specifically, this Hessian matrix was calculated at the B3LYP (Lee et al., 1988; Becke, 1993)/6-31G(d) level of theory after optimizing a “small” model of both binding sites at the same level to a confirmed minimum (no imaginary frequencies). Such “small” models (generated by *MCPB.py* (Li and Merz, 2016)) included: 1) the Fe<sup>2+</sup> cation in that particular binding site; 2) binding site residue sidechains up to their Cβ, with a shorter Cβ-H bond replacing Cα-Cβ; and 3) in the catalytic site, the entire 2-OG cofactor. Residues' Cβ atoms and 2-OG's carboxylate oxygens farthest from Fe<sup>2+</sup> were frozen during optimization and excluded from frequency calculations. Atomic point charges on both Fe<sup>2+</sup> cations on individual binding site residues and on the entire 2-OG cofactor were fitted by *MCPB.py* (Li and Merz, 2016) using the *RESP* method (Bayly et al., 1993) based on the outcome of (single-point) *ESP* charge fitting calculations (Besler et al., 1990) at a higher DFT level, on a “large” version of each Fe<sup>2+</sup> binding site. More specifically, the chosen level of DFT for *ESP* charge fitting (Besler et al., 1990) was B3LYP (Lee et al., 1988; Becke, 1993)/6-31G(d)/def2-SV(P), (Weigend and Ahlrichs, 2005) with the def2-SV(P) basis set specifically applied to Fe<sup>2+</sup>. “Large” binding site models (generated by *MCPB.py* (Li and Merz, 2016)) comprised Fe<sup>2+</sup>, 2-OG, binding site residues in their entirety (i.e., with their backbone) as well as backbones of Glu596 and His668, contiguous to His595/Asp597 and His667/Asp669, respectively. All contiguous backbone fragments in the large models are capped by acetyl and *N*-methyl moieties at their N- and C-termini, respectively. *Gaussian09* (Frisch et al., 2009)

TABLE 2 List of 2-OG analogs tested. For each compound, the chemical formula of the associated free acid is reported. None of the compounds yielded detectable binding to LH3/PLD3 or inhibition of the conversion of 2-OG into succinate in presence or absence of acceptor substrate.

Compound	Formula
Formate	CH <sub>2</sub> O <sub>2</sub>
Oxalate	C <sub>2</sub> H <sub>2</sub> O <sub>4</sub>
Malonate	C <sub>3</sub> H <sub>4</sub> O <sub>4</sub>
Tartronate	C <sub>3</sub> H <sub>4</sub> O <sub>5</sub>
Mesoxalate	C <sub>3</sub> H <sub>2</sub> O <sub>5</sub>
Aminomaltonate	C <sub>3</sub> H <sub>5</sub> NO <sub>4</sub>
Fumarate	C <sub>4</sub> H <sub>4</sub> O <sub>4</sub>
Oxalacetate	C <sub>4</sub> H <sub>4</sub> O <sub>5</sub>
Malate	C <sub>4</sub> H <sub>6</sub> O <sub>5</sub>
Aspartate	C <sub>4</sub> H <sub>7</sub> NO <sub>4</sub>
Tartrate	C <sub>4</sub> H <sub>6</sub> O <sub>6</sub>
Glutamate	C <sub>5</sub> H <sub>9</sub> NO <sub>4</sub>
Glutarate	C <sub>5</sub> H <sub>8</sub> O <sub>4</sub>
Acetonedicarboxylate	C <sub>5</sub> H <sub>6</sub> O <sub>5</sub>
2-hydroxyglutarate	C <sub>5</sub> H <sub>8</sub> O <sub>5</sub>
Adipate	C <sub>6</sub> H <sub>10</sub> O <sub>4</sub>

was programmed to perform *ESP* charge fitting (Besler et al., 1990) over 10 spherical shells around each atom, with 17 grid points per square Bohr. As *per MCPB.py's* default (Li and Merz, 2016), the atomic radius of both Fe<sup>2+</sup> cations was taken to be 1.409 Å. In all DFT calculations, Fe<sup>2+</sup> centers in both the catalytic and noncatalytic binding sites were modeled in their quintet state, after comparative optimizations of each site with Fe<sup>2+</sup> in the triplet and singlet state confirmed—at the B3LYP (Lee et al., 1988; Becke, 1993)/6-31G(d) level of theory—that the quintet state is the most energetically stable in both cases (data not published). All remaining LH3/PLD3 residues—including intra-residue bonded parameters for residues in both Fe<sup>2+</sup>-binding sites—were treated with the *ff14SB* forcefield (Maier et al., 2015), whereas Na<sup>+</sup> counteranions were modeled with parameters by Joung and Cheatham (Joung and Cheatham, 2008): these are compatible with the TIP3P model in use for water (Jorgensen et al., 1983).

## Results

### The LH catalytic site does not accommodate competitive inhibitors

The presence of amino acid networks shared by Fe<sup>2+</sup>, 2-OG-dependent dioxygenase enzymes in their catalytic sites provides a general template for structure-based design of potential inhibitors. The analysis of the residues surrounding the



catalytic  $\text{Fe}^{2+}$  and the 2-OG indeed supported the possibility that 2-OG analogs may act as competitive inhibitors of the co-substrate molecule (Rose et al., 2011). In human LH3/PLOD3, the catalytic  $\text{Fe}^{2+}$  is strongly coordinated by the side chains of His667, Asp669, and His719 of the DSBH fold, and by the 2-OG co-substrate, capped towards the outer solvent by a flexible capping loop defined by residues 590–610 (Figure 2). Using a combination of nano-differential scanning fluorimetry (nanoDSF) and luminescence-based activity assays, we screened a small, focused library of 2-OG analogs (Table 2) searching for compounds capable of inhibiting LH activity. After thorough testing, none of the compounds tested showed binding/folding stabilization in nanoDSF, nor inhibition of enzymatic conversion of 2-OG into succinate. Likewise, a custom-designed library of compounds selected through *in silico* virtual screening of specific candidate binders of the LH3/PLOD3 catalytic site did not provide suitable hints for inhibitors of LH activity using these assays.

## A second $\text{Fe}^{2+}$ binding site on the capping loop modulates accessibility to the LH catalytic site

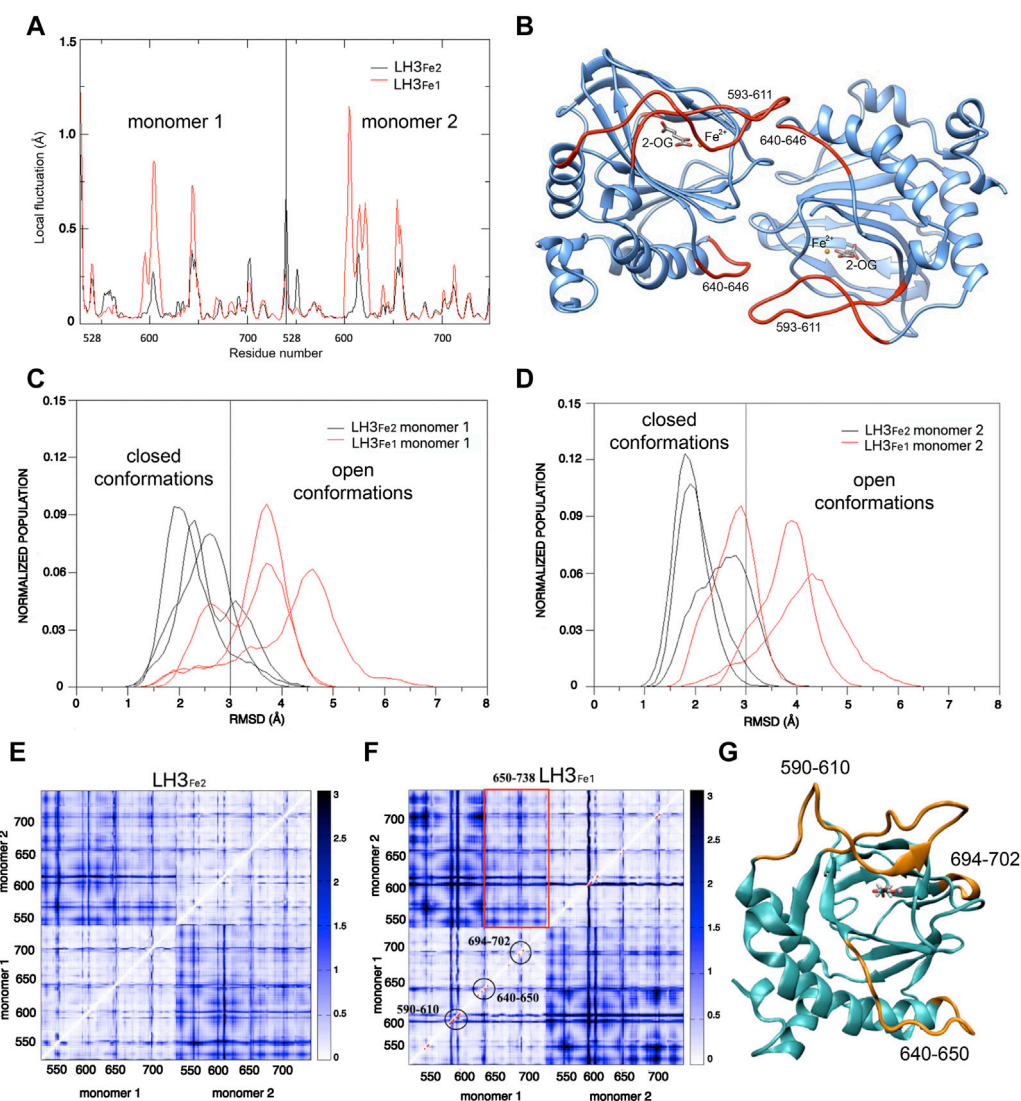
Intrigued by the recalcitrance to inhibition of the LH catalytic site by 2-OG analogs, we focused our attention to the distinguishing features displayed by this domain when compared to homologous  $\text{Fe}^{2+}$ , 2-OG-dependent dioxygenases, and in particular to the capping loop and the stable conformation adopted in the presence of excess  $[\text{Fe}^{2+}]$ . This interlocked state may indeed constitute an obstacle when dealing with inhibition of the LH catalytic site, and the relative positioning of the capping loop is crucial for inhibitor accessibility to the active pocket. We decided to perform a thorough investigation of the flexibility of this loop *in silico* and with site-directed mutagenesis *in vitro* to validate the possible significance of the second  $\text{Fe}^{2+}$  binding site prior to attempting to quantitatively probe the specific metal ion binding to the two distinct LH3/PLOD sites within the LH domain.

To investigate the impact of the second  $\text{Fe}^{2+}$  on the structural conformation of the capping loop 590–610, the C-terminal LH domain involved in LH/PLOD dimerization underwent all-atom MD simulation in explicit solvent, with and without the second  $\text{Fe}^{2+}$  ion. In the following, we refer to the simulated system with the second  $\text{Fe}^{2+}$  as LH3<sub>Fe2</sub>, while the system simulated without it has been named LH3<sub>Fe1</sub>. Three independent replicate simulations were carried out for the two systems, each 1  $\mu\text{s}$  long. In each independent replicate we used identical simulation parameters (see Material and Methods), varying only the initial velocities of atoms via random assignments from a Maxwell distribution consistent with the required temperature. Visual inspection of MD simulations shows that both systems are characterized by minimal atom fluctuations,

suggesting that this fragment of the enzyme is stiff and allows for minimal protein motions away from the starting (crystal) structure, except for the gate loop and some protein region at the dimer interface.

The first step in the analysis of the dynamics of LH3/PLOD3 C-terminus and its potential variation as a function of the presence/absence of the second  $\text{Fe}^{2+}$ , was the identification of protein regions displaying higher levels of local flexibility during dynamics, through the computation of the local fluctuations (LF). This parameter detects the flexibility of a given residue with respect to the neighboring amino acids; the comparison of this calculation for the two systems (i.e., LH3<sub>Fe1</sub> and LH3<sub>Fe2</sub>), allows to extrapolate variations in protein regions more affected by the presence/absence of the second  $\text{Fe}^{2+}$ . LF is calculated for residue  $i$ , as the mean of the variances of the distance  $d_{ij}$  of the  $\text{C}_\alpha$  ( $i$ ) and  $\text{C}_\alpha$  ( $j$ ) atoms, for  $j = i-2, i-1, i+1, i+2$ . In the LH3<sub>Fe1</sub> system residues 593–611 of the gate loop of both monomers displayed higher flexibility as compared to the LH3<sub>Fe2</sub> system (Figures 3A,B). The peak at residues 640–646, constituted by residues of one monomer which contact the gate loop of the other monomer, is more defined in the LH3<sub>Fe1</sub> system (Figures 3A,B). To quantify the observed differences in the dynamics of the enzyme during MD in the two simulated systems, we calculate the root-mean-square deviation (RMSD) distributions of MD trajectories, after optimal rigid body superposition of backbone atoms of each snapshot with the corresponding atoms of the crystal structure, where the gate loop is in the closed conformation.

The RMSD distributions of the loop 590–610 in the two monomers of the single trajectories are reported in Figures 3C,D, showing the different behavior of the two systems and the variations in the single trajectories. A threshold of 3 Å was used to distinguish the closed from the open conformation, based on visual inspection of MD simulations and superposition of snapshots representing the two states. These plots show that the gate loop is stabilized in the closed conformation in the LH3<sub>Fe2</sub>, while the absence of second  $\text{Fe}^{2+}$  impacts the stability of the loop, shifting the population of LH3<sub>Fe1</sub> towards the open conformation, even though both systems can visit the two states. To understand the impact of the second  $\text{Fe}^{2+}$  on the internal dynamics of LH3/PLOD3, we performed the distance fluctuation analysis on the meta trajectories of the two systems, obtained concatenating the three MD replicas of each system. The DF parameter, which is the variance of the inter-residue distance  $d_{ij}$ , was calculated for any pair of residues during the trajectory (see Material and Methods). The resulting DF matrices calculated for LH3<sub>Fe1</sub> and LH3<sub>Fe2</sub> shown in Figures 3E,F, report on the fluctuation of the inter-residue distance in the corresponding residue pairs and matrix regions, describing the intrinsic flexibility and coordination of the LH domains. Relatively low DF values identify protein regions that move together, in coordination. The comparison of these matrices can be used to evaluate possible changes of the internal

**FIGURE 3**

Computational analysis of the dynamic behavior of LH capping loop in modulating accessibility to the LH catalytic site. **(A)** Mean values of the local fluctuations for each residue of the dimerization LH domains of LH3, calculated over the three replicate simulations for LH3<sub>Fe1</sub> (red profile) and LH3<sub>Fe2</sub> (black profile) systems. **(B)** Structure of the dimerization LH domains of LH3: residues corresponding to the highest peaks in the local fluctuations profile of LH3<sub>Fe1</sub> (593–611 of the gate loop and residues 640–646) are colored in red. **(C,D)** RMSD distributions (loop 590–610) of the LH3<sub>Fe2</sub> (black lines) and LH3<sub>Fe1</sub> systems, for both LH domains. **(E,F)** Distance Fluctuations analysis. Color-coded DF matrices resulting from the DF analysis for the LH3<sub>Fe2</sub> **(E)** and LH3<sub>Fe1</sub> **(F)** systems. Pairs of amino acids that move in a coordinated way, that is with a quasi-rigid motion, are characterized by small DF values (white dots) and vice versa (blue dots). Red dots in the LH3<sub>Fe1</sub> matrix correspond to residues which show significant difference in DF values compared to the LH3<sub>Fe2</sub> matrix, according to the F-test. Red rectangle encloses residues 650–738 of one monomer, corresponding to the beta-sheet core of the LH domain, that move in a coordinated way with the other monomer. **(G)** Structure of the LH domain of LH3: residues which show significant differences in protein motion for the two systems are colored in orange.

dynamics and coordination due to the presence/absence of the second Fe<sup>2+</sup>.

Overall, the matrices for both systems turned out to be similar, with largely overlapping patterns of small and large fluctuations of inter-residue distances. In both systems, the internal dynamics of the two domains is characterized by small atomic fluctuations, confirming that protein structure is

stiff. Indeed, the highest DF parameters in both systems correspond to protein regions that do not adopt well-ordered 3D structures, which are intrinsically more flexible. Generally, the inter-domain motion is characterized by higher fluctuations than the internal domain motion. In particular, in the LH3<sub>Fe1</sub> system, residues 650–738 of one monomer, corresponding to the beta-sheet core, move in a more

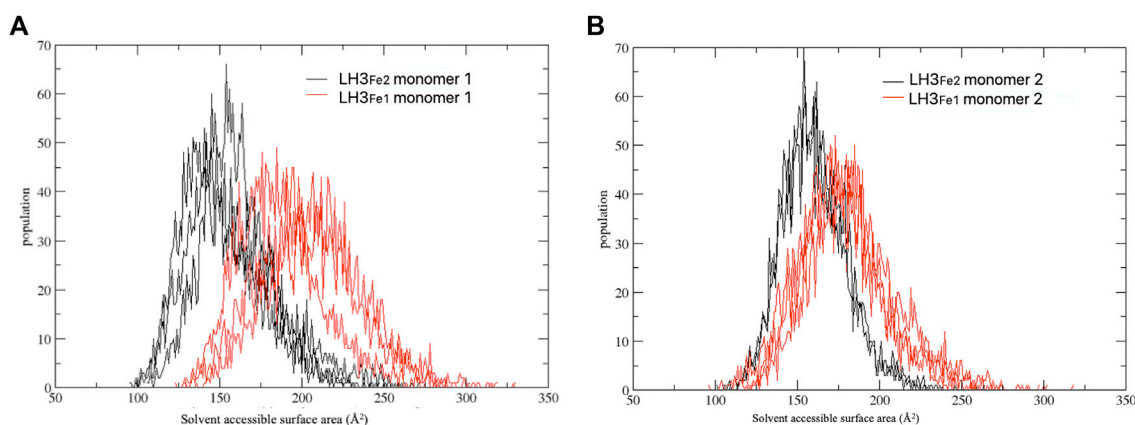


FIGURE 4

Distribution of the solvent accessible surface area calculated for the three replicate simulations. The simulations have been carried out separately for each computed system (LH3<sub>Fe1</sub> in red, LH3<sub>Fe2</sub> in black) and monomer (A,B).

coordinated way with the other monomer than the rest of the protein (Figure 3F).

The significance of the differences observed between the two DF matrices was evaluated with a statistical analysis based on F-test. We used LH3<sub>Fe2</sub> as a reference state for comparing the two matrices. Red dots in the LH3<sub>Fe1</sub> system (Figure 3F) correspond to residues which show significant difference in DF values in respect to the corresponding DF values in LH3<sub>Fe2</sub>, according to this test. This analysis highlights that relevant differences in protein motion for the two systems concern residues of the gate loop 590-610 and residues 640-650 forming the loop that connects the alpha helix 618-639 with the beta-sheet core of the protein. This loop partially forms the dimerization interface of the two domains, as well as residues 694-702, which displays higher DF values in the LH3<sub>Fe1</sub> system (Figures 3F,G).

To further investigate the role of the relative positioning of the capping loop for the accessibility to the active pocket depending on the second Fe<sup>2+</sup> ion, we computed the part of the van der Waals surface of residues forming the catalytic site (protein residues within 6 Å of 2-OG and Fe<sup>2+</sup>) that are accessible to solvent, that is the solvent accessible surface area, across each frame of MD trajectories, using the VMD program (Humphrey et al., 1996). This analysis gives a measure of the capability of the loop in regulating the access of the substrate to the catalytic site, depending on the presence of the second Fe<sup>2+</sup>. Figure 4 shows the distribution of the solvent accessible surface area calculated for the three replicas, for each system and monomer. In both monomers the distributions of the LH3<sub>Fe1</sub> system are shifted towards slightly higher values compared to the LH3<sub>Fe2</sub>, suggesting that the presence of the Fe<sup>2+</sup> has an impact on the access to the active site by modulating the opening/closure of the loop, even though both systems can visit the two distinct states, in agreement with the RMSD analysis discussed above.

To inspect the effect of the second Fe<sup>2+</sup> on the dynamics of the capping loop at a finer level, we evaluated the persistence over the simulation time of the native and new interactions established by this loop with the rest of the protein. The results of this analysis show that most of the native contacts are maintained during the simulation time in both systems (Supplementary Figure S1A,B), while in LH3<sub>Fe1</sub> some residues of the loop tend to weaken some of these interactions compared to LH3<sub>Fe2</sub>. As for new stable contacts, we observed that in LH3<sub>Fe1</sub> the loop can't form new interactions with the rest of the protein due to its higher mobility, while in the LH3<sub>Fe2</sub> the loop establishes new steady interactions with residues 643-648 of the other monomer (Supplementary Figure S1C).

As mentioned above, in the crystal structure of the LH domain obtained in the presence of excess [Fe<sup>2+</sup>], residue Arg599 belonging to the loop 590-610 forms a salt bridge with the 2-OG co-substrate, yielding a conformation that mimics the collagen lysine substrate (Sciatti et al., 2018) (Figures 2C,D). To check the impact of the presence/absence of the second Fe<sup>2+</sup> on the stabilization of this bond, we tracked the hydrogen bonds between Arg599 and 2-OG co-substrate over the course of the trajectories. This analysis showed that in most of the conformations visited during the dynamics, Arg599 forms a salt-bridge with 2-OG in both systems, with a slight prevalence in LH3<sub>Fe2</sub> (Supplementary Figure S1D-F). Taken together, the results from the *in silico* analysis corroborate the observation that the second Fe<sup>2+</sup> constrains the LH domain in a tightly interlocked conformation with reduced conformational motility in proximity to the enzyme's catalytic site.

To better examine the impact of the second Fe<sup>2+</sup>-binding site on enzymatic activity and substrate accessibility, we generated the Asp597Ala and Asp611Ala variants of human LH3/PLOD3 using site-directed mutagenesis. In particular, we

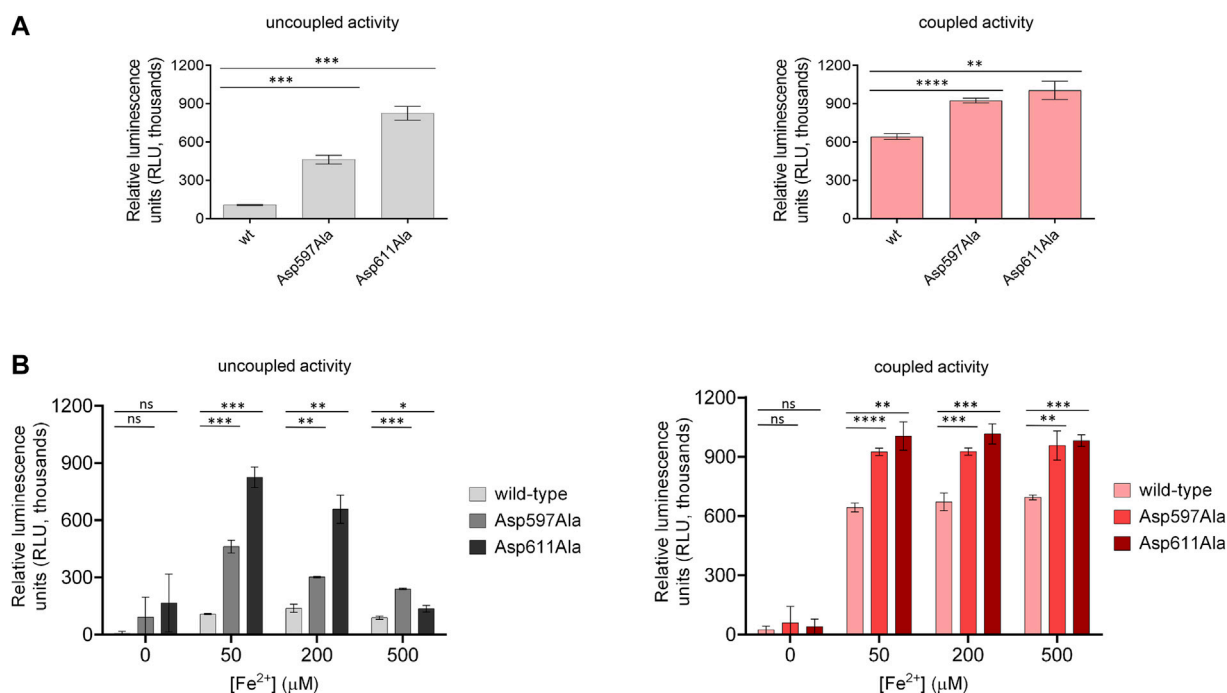


FIGURE 5

Comparison of enzymatic activity of LH3/PLOD3 wild-type and mutants affecting the second  $Fe^{2+}$  binding site. (A) Luminescence-based uncoupled (i.e., with no acceptor substrate, left) and coupled (i.e., using gelatin as acceptor substrate, right) enzymatic activity assays detecting succinate formation. Error bars represent standard deviations from average of triplicate independent experiments. Statistical evaluations based on pair sample comparisons between luminescence values for wild-type LH3 and mutants using Student's *t*-test. ns, non-significant; \*, *p*-value < 0.05; \*\*, *p*-value < 0.01; \*\*\*, *p*-value < 0.001; \*\*\*\*, *p*-value < 0.0001. (B) Evaluation of the uncoupled (left) and coupled (right) enzymatic activities [as in panel (A)] in presence of increasing concentrations of  $Fe^{2+}$ . Error bars and statistical analyses as in panel (A).

focused on Asp597 as a critical capping loop residue involved in the coordination of the second  $Fe^{2+}$ , whereas Asp611 could represent its counterpart as central to the  $Fe^{2+}$  coordination platform on the surface of the LH domain (Figures 2C,D). Both mutants could be expressed and purified to homogeneity, and showed yields, folding stability and oligomerization states comparable to wild-type LH3/PLOD3 (Supplementary Figure S2). We probed their  $Fe^{2+}$  binding affinity using Temperature-Related Intensity Change (TRIC) and found that both mutants showed 4-times lower affinity compared to wild-type LH3/PLOD3 (Supplementary Figure S3). We were not surprised of the weak binding observed: the catalytic  $Fe^{2+}$  is known to have structural roles within the LH domain and cannot be removed to probe its binding (Guo et al., 2018; Sciatti et al., 2018), therefore we assumed that the binding data collected exclusively refer to possible additional  $Fe^{2+}$  binding sites. When tested for their enzymatic activity in the presence of acceptor substrates such as gelatin (i.e., coupled activity), the mutants showed slightly higher (i.e., 1.5X) activity than their wild-type counterpart (Figure 5A, right). Nevertheless, we consistently observed a much higher (i.e., 3-5X) degree of 2-OG conversion into succinate without the need of an acceptor

substrate (i.e., uncoupled activity, Figure 5A, left). Prompted by this observation, we decided to investigate the modulatory effect of  $[Fe^{2+}]$  on both uncoupled and coupled enzymatic activities of these mutants. We found that the uncoupled activities of both LH3/PLOD3 Asp597Ala and Asp611Ala mutants were less inhibited by high  $[Fe^{2+}]$  compared to wild-type LH3/PLOD3 (Figure 5B, left), whereas the modulation of the coupled enzymatic activities was almost unaffected by the presence of the point mutations (Figure 5B, right). Based on the combined results obtained from the MD simulations and mutagenesis, we reasoned that binding of a second metal ion on the surface of the LH domain induces a conformationally stable, self-inhibited state which in turn reduces the ability of LH/PLOD3 enzymes to process 2-OG into succinate via uncoupled activity.

## $Fe^{2+}$ chelating agents produce unexpected effects on LH3/PLOD3 enzymatic activity

Given the unique presence of two distinct  $Fe^{2+}$  binding sites in close proximity, but with opposite effects on enzymatic



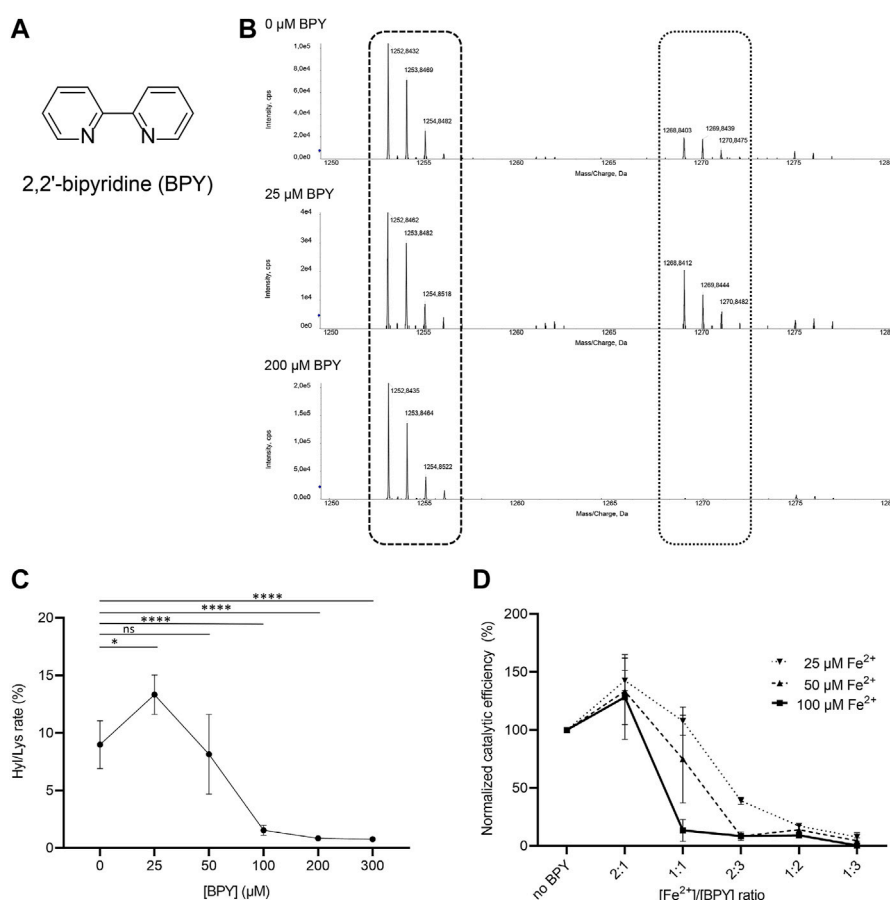


FIGURE 6

Evaluation of the effect of 2,2'-bipyridil (BPY) on LH3/PLOD3 enzymatic activity. **(A)** Chemical structure of BPY. **(B)** Detection of Hyl formation using direct mass spectrometry (MS) assays, and evaluation of the effect of BPY on Lys-to-Hyl conversion. The scheme shows the comparison between three MS spectra, showing the peaks consistently detected for unmodified Lys in the synthetic peptides (dashed box) and the presence of additional peaks (dotted box) indicating Hyl formation. **(C)** Results of MS analysis of LH3/PLOD3 enzymatic activity as a function of BPY concentration. The plot shows the presence of an unexpected increase of enzymatic activity at low BPY concentrations, followed by a drop consistent with inhibition due to sequestration of catalytic Fe<sup>2+</sup>. Error bars represent standard deviations from average of triplicate independent experiments. Statistical evaluations based on pair sample comparisons between data points collected in absence or in presence of [Fe<sup>2+</sup>] using Student's *t*-test. ns, non-significant; \*, *p*-value < 0.05; \*\*, *p*-value < 0.01; \*\*\*, *p*-value < 0.001; \*\*\*\*, *p*-value < 0.0001. **(D)** Evaluation of LH3/PLOD3 enzymatic activity as a function of varying [BPY]/[Fe<sup>2+</sup>] ratios. The analysis shows that the enzymatic activity increase, already observed at low BPY concentration, depends on the residual Fe<sup>2+</sup> concentration available for catalysis. Error bars represent standard deviations from average of triplicate independent experiments. Statistical analyses conducted by comparing data points collected at the same [Fe<sup>2+</sup>]/[BPY] ratio using one-way ANOVA analysis highlight non-significant differences for experiments performed using different concentrations of Fe<sup>2+</sup> in the assay. Additional statistical evaluations for each data point are provided in [Supplementary Figure S3](#).

activity, we wondered whether small-molecule inhibitors acting through metal ion chelation could efficiently modulate substrate processing in LH/PLOD enzymes. To date, despite large scale screenings identified potential hits ([Devkota et al., 2019](#)), specific inhibitors of the lysyl hydroxylase activity are missing. The only known inhibitor of LH/PLOD enzymes is the 2,2'-bipyridine (BPY), a non-specific inhibitor that act as chelating agent. BPY chemical structure is characterized by two pyridyl rings, heterocyclic chemical moieties containing nitrogens ([Figure 6A](#)). Although the mechanism of action of BPY on LH/PLOD enzymes is not well characterized, being a metal

chelator, it likely acts on the Fe<sup>2+</sup> ion present in the lysyl hydroxylase domain ([Ikeda et al., 1994](#); [Rose et al., 2011](#); [Vasta and Raines, 2015](#); [Jover et al., 2018](#)). We therefore decided to explore the actual impact of BPY inhibition on human LH3/PLOD3 enzymatic activity. Firstly, we attempted to evaluate concentration-dependent inhibition using luminescence-based assays, but we realized that usage of BPY was not compatible with the assay setup, likely due to inhibition of the Mg<sup>2+</sup>-dependent luciferase reaction by the reagent. We therefore focused on a strategy to directly investigate hydroxylysine formation on synthetic collagen peptides using

mass spectrometry (MS) by evaluating the relative ratios between non-hydroxylated and hydroxylated lysine side chains (Figure 6B). Using this method, we could perform accurate quantitation of the dose-dependent effects of BPY on LH3/PLOD3 enzymatic activity.

In absence of BPY, collagen peptide substrate processing was consistent with previously reported data (Scietti et al., 2018). Given a BPY:Fe<sup>2+</sup> chelation stoichiometry of 3:1, and a half-maximal BPY concentration required to form complexes with 20  $\mu$ M Fe<sup>2+</sup> (Fe<sub>20</sub>-EC<sub>50</sub>) experimentally determined in about 40  $\mu$ M (Vasta and Raines, 2015), we would have expected an initial reduction of LH activity at a [Fe<sup>2+</sup>]/[BPY] ratio of 2:1. Surprisingly, assays performed with concentrations of BPY up to 25  $\mu$ M unexpectedly revealed increasing Hyl/Lys ratios, suggesting enhancement of enzymatic activity, whereas only higher BPY concentrations caused the expected dramatic decay in substrate processing (Figures 6B,C). Considering that standard assays are carried out by supplementing a fixed concentration of Fe<sup>2+</sup> (i.e., 50  $\mu$ M), we wondered whether the activity boost observed at relatively low BPY concentrations could be associated to Fe<sup>2+</sup> sequestration from the non-catalytic site on capping loop, whereas inhibition could be caused by chelation of the catalytic Fe<sup>2+</sup>.

We therefore carried out additional assays, in which we simultaneously varied [Fe<sup>2+</sup>] and [BPY]. Our results (Figure 6D, Supplementary Figure S4) showed that the LH3/PLOD3 catalytic activity always reaches its maximum at [Fe<sup>2+</sup>]/[BPY] ratio of 2:1, supporting a fine balance between sequestration of excess inhibitory metal ions bound to the capping loop and inhibition through chelation of Fe<sup>2+</sup> in the catalytic site. These results suggest that usage of BPY as a lead compound for the development of inhibitors of LH/PLOD enzymatic activity demands particular care, as unexpected concentration-dependent opposite effects may be produced through release of the self-inhibitory, Fe<sup>2+</sup>-dependent capping loop.

## Discussion

Human LH/PLOD enzymes are becoming a hot topic in cancer research, due to their involvement in fibrotic conversion of collagens in the tumor microenvironment and the association to higher risk of metastasis. As cancer metastatization has been clearly correlated to excess Lys-Hyl enzymatic conversion, development of highly specific LH/PLOD inhibitors is desirable. With the release of the molecular structure of full-length human LH3/PLOD3, the challenge of securing detailed atomic structures for at least one human LH/PLOD isoform has been overcome (Scietti et al., 2018), however as of yet no LH/PLOD inhibitors are available. In this work, we carried out a small molecule screening aiming at finding hits to be used for the development of LH/PLOD inhibitors. To achieve this goal, we

used structural and mechanistic insights from related Fe<sup>2+</sup>, 2-OG-dependent dioxygenases, and found that the challenge of developing specific LH/PLOD inhibitors may present additional obstacles for which extra care is needed.

Despite the high resemblance of the LH3/PLOD3 catalytic pocket with structurally-related Fe<sup>2+</sup>, 2-OG dioxygenases that could be inhibited by 2-OG analogs (Rose et al., 2011), our initial campaign focused on development of small-molecule inhibitors based on their ability to compete with 2-OG in the LH catalytic site were not successful. Likewise, dedicated libraries of compounds developed *in silico* based on high-resolution structural templates of the enzyme cavity did not provide any useful leads towards LH/PLOD inhibition. Using MD simulations, we could interpret the systematic recalcitrance to inhibition with limited accessibility to the catalytic site, caused by a very stable, self-inhibited state generated by specific conformations adopted by the capping loop in the presence of excess [Fe<sup>2+</sup>]. Previous structural studies demonstrated that the capping loop folding was strictly dependent on the iron coordination by four residues, three of which laying on the surface of the LH domain (i.e., His595, Asp611, His613), and one (Asp597) being part of the capping loop. By mutagenizing either Asp611 or Asp597 to alanine, we could observe a decrease in the binding affinity for non-catalytic Fe<sup>2+</sup> and a much-increased uncoupled conversion of 2-OG into succinate which can be explained with improved accessibility to the catalytic site, supporting the physiological role of the Fe<sup>2+</sup>-dependent self-inhibited conformation adopted by the capping loop.

Taken together, these results highlight the extremely delicate balance of Fe<sup>2+</sup> concentration needed in LH/PLOD enzymes to enable their function: while too little [Fe<sup>2+</sup>] hampers catalysis due to lack of an essential component in the catalytic site, even a little excess can instead interlock the LH domain into a self-inhibited state. This behavior is unique for LH/PLODs and different from related Fe<sup>2+</sup>, 2-OG dioxygenases, and is likely responsible for the differential responses observed during treatment with Fe<sup>2+</sup> chelators. Indeed, when testing BPY as candidate inhibitor of LH activity, we found that at low concentrations this compound is capable of enhancing the enzymatic activity rather than blocking it. We interpreted this boost in enzymatic activity with sequestration of the Fe<sup>2+</sup> bound on the LH domain surface trapping the capping loop, and we could demonstrate that such effect depends on the ratio between [Fe<sup>2+</sup>] and [BPY].

Collectively, the results obtained provide interesting perspectives regarding the mechanisms of substrate processing by LH/PLOD enzymes, as well as guidance for future inhibitor design. The strong stability of the Fe<sup>2+</sup>-induced conformation adopted by the capping loop, together with the self-inhibited state obtained through substrate mimicry by LH3/PLOD3 Arg599 implies that physiological substrate processing may depend on the release of the second Fe<sup>2+</sup> interlock through long-range interactions on the surface of the LH domain. Given the extended conformation of collagen polypeptide substrates, this is a likely option and demands further investigation, in

particular considering the possible roles that metal ions bound to collagen substrates may have upon/during post-translational modification of lysines, but also after the processed collagen has detached from the LH/PLOD enzymes. As for inhibitor design, the presence of this metal ion-dependent self-inhibited state represents an additional challenge, which may require the development of synergistic strategies acting simultaneously on the release of the capping loop and on the competition with either 2-OG in the catalytic site, or the Lys side chain subjected to hydroxylation.

## Data availability statement

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

## Author contributions

LS and FF carried out preliminary structural analyses, designed and supervised the research work. LS, DM, and AC produced wild-type LH3, with support from SF. DM, MDM, LN, and FDG performed enzymatic activity assays and analyzed results. DM performed TRIC studies to assess  $\text{Fe}^{2+}$  binding. AC carried out mutagenesis experiments and purified the LH3/PLOD3 mutants, with support from SF. MF carried out mass spectrometry experiments and analyzed data. SAS carried out the calculations for the parametrization of  $\text{Fe}^{2+}$  binding sites. EM, SS and GC carried out molecular simulations and analysis. LS, EM, DM and FF wrote the manuscript, with contributions from all authors.

## Funding

This project has received funding from the Italian Association for Cancer Research (AIRC, “My First AIRC Grant” id. 20075 to FF), by the Mizutani Foundation for Glycoscience (grant id. 200039 to FF), by Fondazione Giovanni Armenise-Harvard (CDA2013 to FF), and by the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Program (2018–2022, to the Department of Biology and

Biotechnology “L. Spallanzani,” University of Pavia). AC carried out research while recipient of a Marie Curie Individual Fellowship (MSCA-IF) from the European Union’s Horizon 2020 research and innovation program (grant agreement COTETHERS—n. 745934). The TRIC instrumentation used for this research was acquired through funding by Regione Lombardia, regional law n° 9/2020, resolution n° 3776/2020. None of the funding sources had roles in study design, collection, analysis and interpretation of data, in the writing of the report and in the decision to submit this article for publication.

## Acknowledgments

We thank the PASS-BioMed Facility (Centro Grandi Strumenti) of the University of Pavia for provision of TRIC instrumentation and Sristi Raj Rai for help with enzymatic assays. SS wishes to acknowledge Prof. Ulf Ryde (Lund University, Sweden) for his kind advice on using the def2-SV(P) basis set for  $\text{Fe}^{2+}$ .

## Conflict of interest

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

## Publisher’s note

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

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.876352/full#supplementary-material>

## References

- Back, J. H., Yun, H. S., Kwon, G. T., Lee, J., Kim, J. Y., Jo, Y., et al. (2019). PLOD3 suppression exerts an anti-tumor effect on human lung cancer cells by modulating the PKC-delta signaling pathway. *Cell Death Dis.* 10, 156. doi:10.1038/s41419-019-1405-8
- Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* 97, 10269–10280. doi:10.1021/j100142a004
- Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi:10.1063/1.464913
- Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. V., 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
- Besler, B. H., Merz, K. M., Jr., and Kollman, P. A. (1990). Atomic charges derived from semiempirical methods. *J. Comput. Chem.* 11, 431–439. doi:10.1002/jcc.540110404

- Blanco, M. A., Leroy, G., Khan, Z., Aleckovic, M., Zee, B. M., Garcia, B. A., et al. (2012). Global secretome analysis identifies novel mediators of bone metastasis. *Cell Res.* 22, 1339–1355. doi:10.1038/cr.2012.89
- Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., Cruzeiro, V. W. D., et al. (2018). *Amber 2018*. San Francisco: University of California.
- Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., et al. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688. doi:10.1002/jcc.20290
- Chen, Y. L., Guo, H. F., Terajima, M., Banerjee, P., Liu, X., Yu, J., et al. (2016). Lysyl hydroxylase 2 is secreted by tumor cells and can modify collagen in the extracellular space. *J. Biol. Chem.* 291, 25799–25808. doi:10.1074/jbc.M116.759803
- Chen, Y., Terajima, M., Yang, Y., Sun, L., Ahn, Y. H., Pankova, D., et al. (2015). Lysyl hydroxylase 2 induces a collagen cross-link switch in tumor stroma. *J. Clin. Invest.* 125, 1147–1162. doi:10.1172/JCI74725
- Clifton, I. J., McDonough, M. A., Ehrismann, D., Kershaw, N. J., Granatino, N., and Schofield, C. J. (2006). Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J. Inorg. Biochem.* 100, 644–669. doi:10.1016/j.jinorgbio.2006.01.024
- Costas, M., Mehn, M. P., Jensen, M. P., and Que, L., Jr. (2004). Dioxygen activation at mononuclear nonheme iron active sites: Enzymes, models, and intermediates. *Chem. Rev.* 104, 939–986. doi:10.1021/cr020628n
- 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
- De Giorgi, F., Fumagalli, M., Sciatti, L., and Forneris, F. (2021). Collagen hydroxyllysine glycosylation: Non-conventional substrates for atypical glycosyltransferase enzymes. *Biochem. Soc. Trans.* 49 (2), 855–866. doi:10.1042/BST20200767
- Deng, X., Pan, Y., Yang, M., Liu, Y., and Li, J. (2021). PLOD3 is associated with immune cell infiltration and genomic instability in colon adenocarcinoma. *Biomed. Res. Int.* 2021, 4714526. doi:10.1155/2021/4714526
- Devkota, A. K., Veloria, J. R., Guo, H. F., Kurie, J. M., Cho, E. J., and Dalby, K. N. (2019). Development of a high-throughput lysyl hydroxylase (LH) assay and identification of small-molecule inhibitors against LH2. *SLAS Discov.* 24, 484–491. doi:10.1177/2472555218817057
- Du, H., Chen, Y., Hou, X., Huang, Y., Wei, X., Yu, X., et al. (2017a). PLOD2 regulated by transcription factor FOXA1 promotes metastasis in NSCLC. *Cell Death Dis.* 8, e3143. doi:10.1038/cddis.2017.553
- Du, H., Pang, M., Hou, X., Yuan, S., and Sun, L. (2017b). PLOD2 in cancer research. *Biomed. Pharmacother.* 90, 670–676. doi:10.1016/j.biopha.2017.04.023
- Eisinger-Mathason, T. S., Zhang, M., Qiu, Q., Skuli, N., Nakazawa, M. S., Karakasheva, T., et al. (2013). Hypoxia-dependent modification of collagen networks promotes sarcoma metastasis. *Cancer Discov.* 3, 1190–1205. doi:10.1158/2159-8290.CD-13-0118
- Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of coot. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 486–501. doi:10.1107/S0907444910007493
- Ewans, L. J., Colley, A., Gaston-Massuet, C., Gualtieri, A., Cowley, M. J., McCabe, M. J., et al. (2019). Pathogenic variants in PLOD3 result in a Stickler syndrome-like connective tissue disorder with vascular complications. *J. Med. Genet.* 56, 629–638. doi:10.1136/jmedgenet-2019-106019
- Faravelli, S., Campioni, M., Palamini, M., Canciani, A., Chiapparino, A., and Forneris, F. (2021). Optimized recombinant production of secreted proteins using human embryonic kidney (HEK293) cells grown in suspension. *Bio. Protoc.* 11, e3998. doi:10.21769/BioProtoc.3998
- Flashman, E., and Schofield, C. J. (2007). The most versatile of all reactive intermediates? *Nat. Chem. Biol.* 3, 86–87. doi:10.1038/nchembio0207-86
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian 09 revision A.2*. Wallingford, CT: Gaussian, Inc.
- Gilkes, D. M., Bajpai, S., Wong, C. C., Chaturvedi, P., Hubbi, M. E., Wirtz, D., et al. (2013). Procollagen lysyl hydroxylase 2 is essential for hypoxia-induced breast cancer metastasis. *Mol. Cancer Res.* 11, 456–466. doi:10.1158/1541-7786.MCR-12-0629
- Gkretsi, V., and Stylianopoulos, T. (2018). Cell adhesion and matrix stiffness: Coordinating cancer cell invasion and metastasis. *Front. Oncol.* 8, 145. doi:10.3389/fonc.2018.00145
- Gong, S., Duan, Y., Wu, C., Osterhoff, G., Schopow, N., and Kallendrusch, S. (2021). A human pan-cancer system Analysis of procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3 (PLOD3). *Int. J. Mol. Sci.* 22, 9903. doi:10.3390/ijms22189903
- Guo, H. F., Bota-Rabassadas, N., Terajima, M., Leticia Rodriguez, B., Gibbons, D. L., Chen, Y., et al. (2021). A collagen glucosyltransferase drives lung adenocarcinoma progression in mice. *Commun. Biol.* 4, 482. doi:10.1038/s42003-021-01982-w
- Guo, H. F., Tsai, C. L., Terajima, M., Tan, X., Banerjee, P., Miller, M. D., et al. (2018). Pro-metastatic collagen lysyl hydroxylase dimer assemblies stabilized by Fe(2+)-binding. *Nat. Commun.* 9, 512. doi:10.1038/s41467-018-02859-z
- Guo, T., Gu, C., Li, B., and Xu, C. (2021). PLODs are overexpressed in ovarian cancer and are associated with gap junctions via connexin 43. *Lab. Invest.* 101, 564–569. doi:10.1038/s41374-021-00533-5
- Hausinger, R. P. (2004). FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* 39, 21–68. doi:10.1080/10409230490440541
- 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
- Ikeda, H., Ogata, I., and Fujiwara, K. (1994). Evidence that impaired intracellular collagen synthesis reduces proliferation in cultured rat hepatocytes. *Biochem. Biophys. Res. Commun.* 200, 1701–1707. doi:10.1006/bbrc.1994.1648
- Jiang, H., Guo, W., Yuan, S., and Song, L. (2020). PLOD1 is a prognostic biomarker and mediator of proliferation and invasion in osteosarcoma. *Biomed. Res. Int.* 2020, 3418398. doi:10.1155/2020/3418398
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi:10.1063/1.445869
- Joung, I. S., and Cheatham, T. E. (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* 112, 9020–9041. doi:10.1021/jp8001614
- Jover, E., Silvente, A., Marin, F., Martinez-Gonzalez, J., Orriols, M., Martinez, C. M., et al. (2018). Inhibition of enzymes involved in collagen cross-linking reduces vascular smooth muscle cell calcification. *FASEB J.* 32, 4459–4469. doi:10.1096/fj.201700653R
- Kivirikko, K. I., and Prockop, D. J. (1967). Enzymatic hydroxylation of proline and lysine in procollagen. *Proc. Natl. Acad. Sci. U. S. A.* 57, 782–789. doi:10.1073/pnas.57.3.782
- Kurozumi, A., Kato, M., Goto, Y., Matsushita, R., Nishikawa, R., Okato, A., et al. (2016). Regulation of the collagen cross-linking enzymes LOXL2 and PLOD2 by tumor-suppressive microRNA-26a/b in renal cell carcinoma. *Int. J. Oncol.* 48, 1837–1846. doi:10.3892/ijo.2016.3440
- Lee, C., Yang, W., and Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter* 37, 785–789. doi:10.1103/physrevb.37.785
- Levental, K. R., Yu, H., Kass, L., Lakins, J. N., Egeblad, M., Erler, J. T., et al. (2009). Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* 139, 891–906. doi:10.1016/j.cell.2009.10.027
- Li, G., Wang, X., and Liu, G. (2021). Corrigendum to “PLOD2 is a potent prognostic marker and associates with immune infiltration in cervical cancer”. *Biomed. Res. Int.* 2021, 9762405. doi:10.1155/2021/9762405
- Li, P., and Merz, K. M., Jr. (2016). MCPB.py: A Python based metal center parameter builder. *J. Chem. Inf. Model.* 56, 599–604. doi:10.1021/acs.jcim.5b00674
- Li, S. S., Lian, Y. F., Huang, Y. L., Huang, Y. H., and Xiao, J. (2020). Overexpressing PLOD family genes predict poor prognosis in gastric cancer. *J. Cancer* 11, 121–131. doi:10.7150/jca.35763
- Loenarz, C., and Schofield, C. J. (2008). Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* 4, 152–156. doi:10.1038/nchembio0308-152
- Loenarz, C., and Schofield, C. J. (2011). Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends biochem. Sci.* 36, 7–18. doi:10.1016/j.tibs.2010.07.002
- Loncharich, R. J., Brooks, B. R., and Pastor, R. W. (1992). Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanine-N'-methylamide. *Biopolymers* 32, 523–535. doi:10.1002/bip.360320508
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi:10.1021/acs.jctc.5b00255
- Martinez, S., and Hausinger, R. P. (2015). Catalytic mechanisms of Fe(II)- and 2-Oxoglutarate-dependent oxygenases. *J. Biol. Chem.* 290, 20702–20711. doi:10.1074/jbc.R115.648691
- Moroni, E., Agard, D. A., and Colombo, G. (2018). The structural asymmetry of mitochondrial Hsp90 (Trap1) determines fine tuning of functional dynamics. *J. Chem. Theory Comput.* 14, 1033–1044. doi:10.1021/acs.jctc.7b00766



- Morra, G., Potestio, R., Micheletti, C., and Colombo, G. (2012). Corresponding functional dynamics across the Hsp90 chaperone family: Insights from a multiscale analysis of MD simulations. *PLoS Comput. Biol.* 8, e1002433. doi:10.1371/journal.pcbi.1002433
- Myllylä, R., Schubotz, L. M., Weser, U., and Kivirikko, K. I. (1979). Involvement of superoxide in the prolyl and lysyl hydroxylase reactions. *Biochem. Biophys. Res. Commun.* 89, 98–102. doi:10.1016/0006-291x(79)90948-3
- Noda, T., Yamamoto, H., Takemasa, I., Yamada, D., Uemura, M., Wada, H., et al. (2012). PLOD2 induced under hypoxia is a novel prognostic factor for hepatocellular carcinoma after curative resection. *Liver Int.* 32, 110–118. doi:10.1111/j.1478-3231.2011.02619.x
- Pankova, D., Chen, Y., Terajima, M., Schliekelman, M. J., Baird, B. N., Fahrenholtz, M., et al. (2016). Cancer-associated fibroblasts induce a collagen cross-link switch in tumor stroma. *Mol. Cancer Res.* 14, 287–295. doi:10.1158/1541-7786.MCR-15-0307
- Provenzano, P. P., Eliceiri, K. W., Campbell, J. M., Inman, D. R., White, J. G., and Keely, P. J. (2006). Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC Med.* 4, 38. doi:10.1186/1741-7015-4-38
- Puistola, U., Turpeenniemi-Hujanen, T. M., Myllylä, R., and Kivirikko, K. I. (1980). Studies on the lysyl hydroxylase reaction. II. Inhibition kinetics and the reaction mechanism. *Biochim. Biophys. Acta* 611, 51–60. doi:10.1016/0005-2744(80)90041-8
- Rose, N. R., McDonough, M. A., King, O. N., Kawamura, A., and Schofield, C. J. (2011). Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* 40, 4364–4397. doi:10.1039/c0cs00203h
- Salomon-Ferrer, R., Gotz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* 9, 3878–3888. doi:10.1021/ct400314y
- Saito, T., Uzawa, K., Terajima, M., Shiiba, M., Amelio, A. L., Tanzawa, H., et al. (2019). Aberrant collagen cross-linking in human oral squamous cell carcinoma. *J. Dent. Res.* 98, 517–525. doi:10.1177/0022034519828710
- Sato, K., Parag-Sharma, K., Terajima, M., Musicant, A. M., Murphy, R. M., Ramsey, M. R., et al. (2021). Lysyl hydroxylase 2-induced collagen cross-link switching promotes metastasis in head and neck squamous cell carcinomas. *Neoplasia* 23, 594–606. doi:10.1016/j.neo.2021.05.014
- Schulte, C., Khayenko, V., Nordblom, N. F., Toppel, F., Peck, V., Gupta, A. J., et al. (2021). High-throughput determination of protein affinities using unmodified peptide libraries in nanomolar scale. *iScience* 24, 101898. doi:10.1016/j.isci.2020.101898
- Sciatti, L., Chiapparino, A., De Giorgi, F., Fumagalli, M., Khorai, L., Nergadze, S., et al. (2018). Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. *Nat. Commun.* 9, 3163. doi:10.1038/s41467-018-05631-5
- Sciatti, L., and Forneris, F. (2020). “Full-length human collagen lysyl hydroxylases,” in *Encyclopedia of inorganic and bioinorganic chemistry*. Editor R. A. Scott (Hoboken, NJ: Wiley), 1–12.
- Seminario, J. M. (1996). Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem.* 60, 1271–1277. doi:10.1002/(sici)1097-461x(1996)60:7<1271::aid-qua8>3.0.co;2-w
- Shi, J., Bao, M., Wang, W., Wu, X., Li, Y., Zhao, C., et al. (2021). Integrated profiling identifies PLOD3 as a potential prognostic and immunotherapy relevant biomarker in colorectal cancer. *Front. Immunol.* 12, 722807. doi:10.3389/fimmu.2021.722807
- Sondergaard, C. R., Olsson, M. H., Rostkowski, M., and Jensen, J. H. (2011). Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* 7, 2284–2295. doi:10.1021/ct200133y
- Song, Y., Zheng, S., Wang, J., Long, H., Fang, L., Wang, G., et al. (2017). Hypoxia-induced PLOD2 promotes proliferation, migration and invasion via PI3K/Akt signaling in glioma. *Oncotarget* 8, 41947–41962. doi:10.18632/oncotarget.16710
- Tian, L., Zhou, H., Wang, G., Wang, W. Y., Li, Y., and Xue, X. (2021). The relationship between PLOD1 expression level and glioma prognosis investigated using public databases. *PeerJ* 9, e11422. doi:10.7717/peerj.11422
- Tsai, C. K., Huang, L. C., Tsai, W. C., Huang, S. M., Lee, J. T., and Hueng, D. Y. (2018). Overexpression of PLOD3 promotes tumor progression and poor prognosis in gliomas. *Oncotarget* 9, 15705–15720. doi:10.18632/oncotarget.24594
- Van Der Slot, A. J., Zuurmond, A. M., Van Den Bogaert, A. J., Ulrich, M. M., Middelkoop, E., Boers, W., et al. (2004). Increased formation of pyridinoline cross-links due to higher telopeptide lysyl hydroxylase levels is a general fibrotic phenomenon. *Matrix Biol.* 23, 251–257. doi:10.1016/j.matbio.2004.06.001
- Vasta, J. D., and Raines, R. T. (2015). Selective inhibition of prolyl 4-hydroxylases by bipyrinedicarboxylates. *Bioorg. Med. Chem.* 23, 3081–3090. doi:10.1016/j.bmc.2015.05.003
- Wan, J., Qin, J., Cao, Q., Hu, P., Zhong, C., and Tu, C. (2020). Hypoxia-induced PLOD2 regulates invasion and epithelial-mesenchymal transition in endometrial carcinoma cells. *Genes Genomics* 42, 317–324. doi:10.1007/s13258-019-00901-y
- Wang, B., Xu, L., Ge, Y., Cai, X., Li, Q., Yu, Z., et al. (2019). PLOD3 is upregulated in gastric cancer and correlated with clinicopathologic characteristics. *Clin. Lab.* 65, 2901. doi:10.7754/Clin.Lab.2018.180541
- Wang, D., Zhang, S., and Chen, F. (2018). High expression of PLOD1 drives tumorigenesis and affects clinical outcome in gastrointestinal carcinoma. *Genet. Test. Mol. Biomarkers* 22, 366–373. doi:10.1089/gtmb.2018.0009
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi:10.1002/jcc.20035
- Wang, Z., Shi, Y., Ying, C., Jiang, Y., and Hu, J. (2021). Hypoxia-induced PLOD1 overexpression contributes to the malignant phenotype of glioblastoma via NF-κB signaling. *Oncogene* 40, 1458–1475. doi:10.1038/s41388-020-01635-y
- Weigend, F., and Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 7, 3297–3305. doi:10.1039/b508541a
- Xie, D., Li, J., Wei, S., Qi, P., Ji, H., Su, J., et al. (2020). Knockdown of PLOD3 suppresses the malignant progression of renal cell carcinoma via reducing TWIST1 expression. *Mol. Cell. Probes* 53, 101608. doi:10.1016/j.mcp.2020.101608
- Xu, W. H., Xu, Y., Wang, J., Tian, X., Wu, J., Wan, F. N., et al. (2019). Procollagen-lysine, 2-oxoglutarate 5-dioxygenases 1, 2, and 3 are potential prognostic indicators in patients with clear cell renal cell carcinoma. *Aging (Albany NY)* 11, 6503–6521. doi:10.18632/aging.102206
- Xu, Y., Zhang, L., Wei, Y., Zhang, X., Xu, R., Han, M., et al. (2017). Procollagen-lysine 2-oxoglutarate 5-dioxygenase 2 promotes hypoxia-induced glioma migration and invasion. *Oncotarget* 8, 23401–23413. doi:10.18632/oncotarget.15581
- Yamada, Y., Kato, M., Arai, T., Sanada, H., Uchida, A., Misono, S., et al. (2019). Aberrantly expressed PLOD1 promotes cancer aggressiveness in bladder cancer: A potential prognostic marker and therapeutic target. *Mol. Oncol.* 13, 1898–1912. doi:10.1002/1878-0261.12532
- Yamauchi, M., and Sricholpech, M. (2012). Lysine post-translational modifications of collagen. *Essays Biochem.* 52, 113–133. doi:10.1042/bse0520113
- Yang, B., Zhao, Y., Wang, L., Zhao, Y., Wei, L., Chen, D., et al. (2020). Identification of PLOD family genes as novel prognostic biomarkers for hepatocellular carcinoma. *Front. Oncol.* 10, 1695. doi:10.3389/fonc.2020.01695
- Zhao, Y., Zhang, X., and Yao, J. (2021). Comprehensive analysis of PLOD family members in low-grade gliomas using bioinformatics methods. *PLoS One* 16, e0246097. doi:10.1371/journal.pone.0246097



## OPEN ACCESS

## EDITED BY

Anastassios C. Papageorgiou,  
University of Turku, Finland

## REVIEWED BY

Manfred S. Weiss,  
Helmholtz Association of German  
Research Centers (HZ), Germany  
Max Nanao,  
European Synchrotron Radiation  
Facility, France

## \*CORRESPONDENCE

Nicole Zitzmann,  
nicole.zitzmann@bioch.ox.ac.uk  
Pietro Roversi,  
pietro.roversi@ibba.cnr.it

<sup>†</sup>These authors have contributed equally  
to this work and share first authorship

## SPECIALTY SECTION

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

RECEIVED 02 June 2022

ACCEPTED 11 November 2022

PUBLISHED 14 December 2022


## CITATION

Caputo AT, Ibba R, Le Cornu JD,  
Darlot B, Hensen M, Lipp CB,  
Marcianò G, Vasiljević S, Zitzmann N and  
Roversi P (2022), Crystal polymorphism  
in fragment-based lead discovery of  
ligands of the catalytic domain of UGGT,  
the glycoprotein folding quality  
control checkpoint.  
*Front. Mol. Biosci.* 9:960248.  
doi: 10.3389/fmolb.2022.960248

## COPYRIGHT

© 2022 Caputo, Ibba, Le Cornu, Darlot,  
Hensen, Lipp, Marcianò, Vasiljević,  
Zitzmann and Roversi. 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.

# Crystal polymorphism in fragment-based lead discovery of ligands of the catalytic domain of UGGT, the glycoprotein folding quality control checkpoint

Alessandro T. Caputo<sup>1,2†</sup>, Roberta Ibba<sup>1,3†</sup>, James D. Le Cornu<sup>1,4</sup>,  
Benoit Darlot<sup>1</sup>, Mario Hensen<sup>1</sup>, Colette B. Lipp<sup>1</sup>,  
Gabriele Marcianò <sup>5</sup>, Snežana Vasiljević<sup>1</sup>, Nicole Zitzmann<sup>1\*</sup>  
and Pietro Roversi<sup>6,7\*</sup>

<sup>1</sup>Biochemistry Department, Oxford Glycobiology Institute, University of Oxford, Oxford, United Kingdom, <sup>2</sup>Commonwealth Scientific and Industrial Research Organisation, Clayton, VIC, Australia, <sup>3</sup>Department of Medicine, Surgery and Pharmacy, University of Sassari, Sassari, Italy, <sup>4</sup>Wellcome Trust Centre for Cell Biology, University of Edinburgh, Scotland, United Kingdom, <sup>5</sup>Biochemistry Department, University of Oxford, Oxford, United Kingdom, <sup>6</sup>IBBA-CNR Unit of Milano, Institute of Agricultural Biology and Biotechnology, Milano, Italy, <sup>7</sup>Department of Molecular and Cell Biology, Leicester Institute of Structural and Chemical Biology, University of Leicester, Leicester, United Kingdom

None of the current data processing pipelines for X-ray crystallography fragment-based lead discovery (FBLD) consults all the information available when deciding on the lattice and symmetry (i.e., the polymorph) of each soaked crystal. Often, X-ray crystallography FBLD pipelines either choose the polymorph based on cell volume and point-group symmetry of the X-ray diffraction data or leave polymorph attribution to manual intervention on the part of the user. Thus, when the FBLD crystals belong to more than one crystal polymorph, the discovery pipeline can be plagued by space group ambiguity, especially if the polymorphs at hand are variations of the same lattice and, therefore, difficult to tell apart from their morphology and/or their apparent crystal lattices and point groups. In the course of a fragment-based lead discovery effort aimed at finding ligands of the catalytic domain of UDP-glucose glycoprotein glucosyltransferase (UGGT), we encountered a mixture of trigonal crystals and pseudotrigonal triclinic crystals—with the two lattices closely related. In order to resolve that polymorphism ambiguity, we have written and described here a series of Unix shell scripts called *CoALLA* (crystal polymorph and ligand likelihood-based assignment). The *CoALLA* scripts are written in Unix shell and use *autoPROC* for data processing, *CCP4-Dimple/REFMAC5* and *BUSTER* for refinement, and *RHOFIT* for ligand docking. The choice of the polymorph is effected by carrying out (in each of the known polymorphs) the tasks of diffraction data indexing, integration, scaling, and structural refinement. The most likely polymorph is then chosen as the one with the best structure refinement  $R_{\text{free}}$  statistic. The *CoALLA* scripts further

implement a likelihood-based ligand assignment strategy, starting with macromolecular refinement and automated water addition, followed by removal of the water molecules that appear to be fitting ligand density, and a final round of refinement after random perturbation of the refined macromolecular model, in order to obtain unbiased difference density maps for automated ligand placement. We illustrate the use of *CoALLA* to discriminate between H3 and P1 crystals used for an FBLD effort to find fragments binding to the catalytic domain of *Chaetomium thermophilum* UGGT.

## KEYWORDS

UGGT, crystal polymorphism, structure-based lead discovery, structure determination pipeline, [(morpholin-4yl)methyl]quinolin-8-ol

## 1 Introduction

There is an urgent medical need to develop novel antiviral drugs, as exemplified by the recent Ebola, Zika, and SARS-CoV-2 outbreaks (Pardi and Weissman, 2020). In particular, an FDA-approved host-targeting broad-spectrum antiviral could revolutionize the treatment of existing and newly emerging viruses (Dwek et al., 2022). Recent works elucidating the structures of key endoplasmic reticulum (ER) enzymes assisting the folding of viral glycoproteins have opened new avenues for identifying novel antivirals (Caputo et al., 2016; Roversi et al., 2017; Warfield et al., 2020).

Host-targeting broad-spectrum antivirals are a possibility because many viruses hijack the same host enzymes during their life cycle (Oksenych and Kainov (2022)). For example, the envelope glycoproteins of many viruses are exquisitely dependent on calnexin-mediated folding, a process enabled by the action of the ER endoplasmic reticulum enzyme UDP-Glc glycoprotein glucosyltransferase (UGGT) and the ER alpha glucosidases I and II (D'Alessio et al., 2010). ER alpha glucosidases I and II usher client proteins in and out of the calnexin cycle, which is part of the glycoprotein folding quality control in the ER, whilst UGGT plays a key role in retaining misfolded glycoproteins in the ER for a “second chance” at folding correctly (Hammond et al., 1994).

One of the more advanced strategies for host-targeting broad-spectrum antiviral drug development is focusing on iminosugars as active site inhibitors of the ER alpha glucosidases (Alonzi et al., 2017; Tyrrell et al., 2017). However, as carbohydrate mimics, iminosugar inhibitors of the ER alpha glucosidases have some undesired off-target effects, as they also inhibit certain other carbohydrate processing enzymes within the human host (Sayce et al., 2016). A new class of molecules inhibiting host glycoprotein folding enzymes that viruses depend upon would have great potential for antiviral therapy (Karade et al., 2021). To complement a programme of development of new allosteric ER alpha glucosidase inhibitors with fewer off-target effects, we endeavoured to investigate the potential of UGGT, the major calnexin cycle misfold sensor (Trombetta et al., 1989),

as a novel antiviral target (Tax et al., 2019). Currently, no UGGT inhibitors are known other than the product UDP (Trombetta and Helenius, 1999) and the UDP-glucose analogue UDP-2-deoxy-2-fluoro-D-glucose (U2F), neither of which is selective for UGGT.

One effective strategy to broaden the knowledge of the chemical space of a given protein target is fragment-based lead discovery (FBLD), a sequence of experiments enabling the extraction of ligands of a chosen target macromolecule from a chemical library (Ciulli et al., 2006; Murray and Blundell, 2010; Chen et al., 2015; Müller et al., 2022). Whenever crystals of the target macromolecule reliably diffract to near-atomic resolution, single-crystal X-ray crystallography is one of the main techniques successfully used for FBLD (Ciulli et al., 2006; Murray and Blundell, 2010; Radoux et al., 2016; Müller et al., 2022).

We recently determined crystal structures of the ER glycoprotein folding quality control checkpoint enzyme, the UDP-Glc glycoprotein glucosyltransferase of *Chaetomium thermophilum* (CtUGGT) (Roversi et al., 2017). As well as a potential drug target against viruses (Dalziel et al., 2014), UGGT could be a target for compounds rescuing slightly misfolded and yet active glycoprotein mutants in certain instances of congenital rare disease (Amara et al., 1992) and against some cancers (Tax et al., 2019). We set out to find small-molecule ligands for this target *via* X-ray crystallography FBLD.

We could not crystallise the GT24 catalytic domain of human UGGT—so we used crystals of the GT24 catalytic domain of CtUGGT (hereinafter CtUGGT<sub>GT24</sub>) instead. The sequence of this fungal UGGT has about 70% similarity and 60% identity to the ones of the same domain of the two human UGGT isoforms (UGGT1 and UGGT2, see Figure 1) so that any ligands found with the crystals of CtUGGT would likely bind the human enzymes too—paving the way to a medicinal chemistry program towards modulators of human UGGT activity. In order to avoid fragments that would bind to the UDP-Glc pocket (and would then likely have some off-target affinity for a number of human glucosyltransferases using the same co-factor (Albesa-Jové and Guerin, 2016), the crystals used for the FBLD



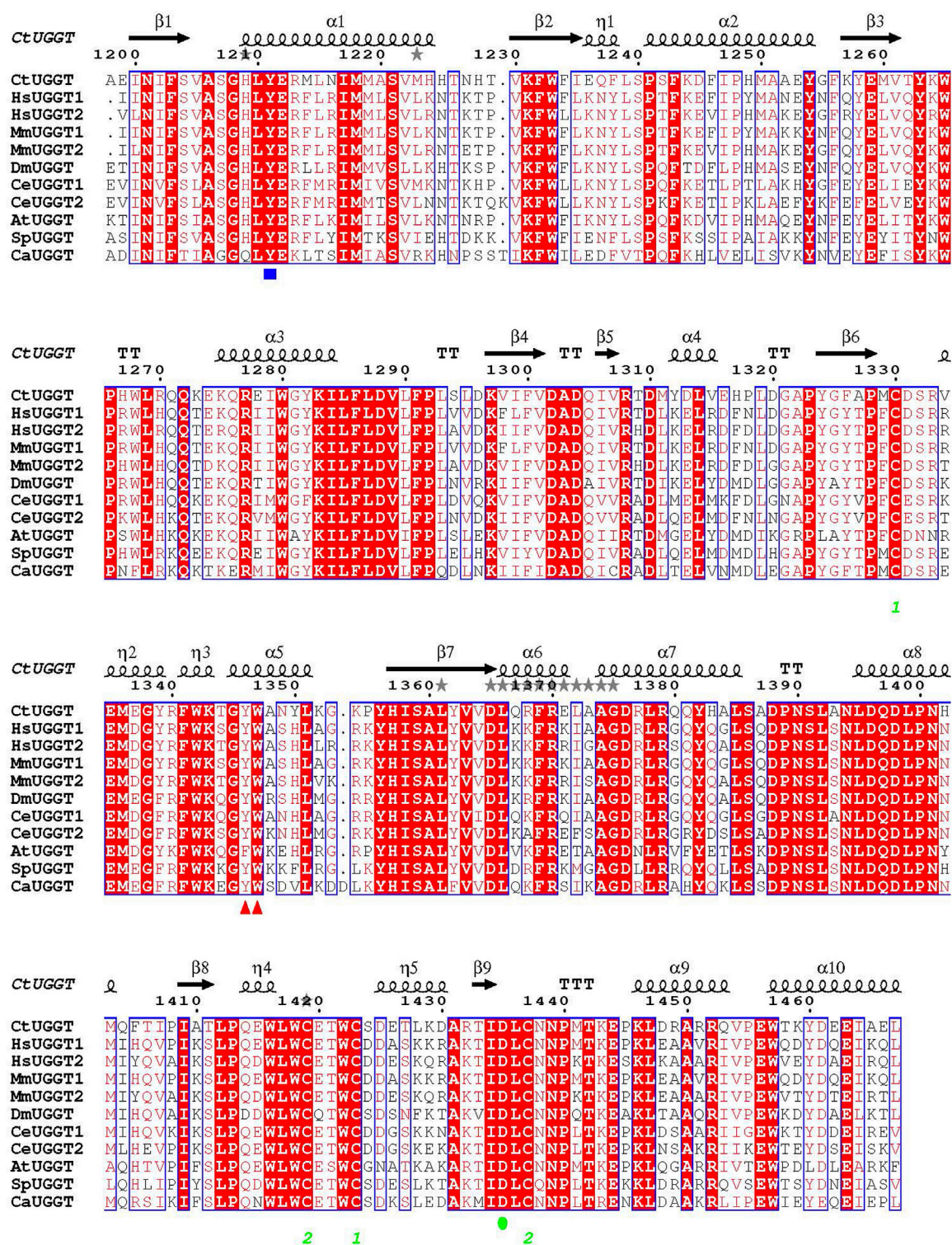


FIGURE 1

Sequence alignment of GT24 domains of a few eukaryotic UGGTs. Ct: *Chaetomium thermophilum*; Hs: *Homo sapiens*; Mm: *Mus musculus*; At: *Arabidopsis thaliana*; Ce: *Caenorhabditis elegans*; Dm: *Drosophila melanogaster*; Sp: *Schizosaccharomyces pombe*; Ca: *Candida albicans*. The CtUGGT D1302 and D1304 residues coordinating the catalytic  $\text{Ca}^{2+}$  ion are completely conserved across these sequences. Red triangles mark the CtUGGT  $^{1346}\text{WY}^{1347}$  clamp. A blue square marks the position of CtUGGT Y1211 (coordinating the UDP-Glc uracyl ring). A green circle marks the position of CtUGGT D1435 (coordinating the  $\text{Ca}^{2+}$  ion).



study were grown in the presence of  $\text{Ca}^{2+}$  and UDP-Glc (hereinafter  $\text{CtUGGT}_{\text{GT24}}^{\text{UDP-Glc}}$ ).

The  $\text{CtUGGT}_{\text{GT24}}^{\text{UDP-Glc}}$  crystals turned out to belong to two different polymorphs. Crystal polymorphism refers to the growth of crystals of the same compound belonging to different crystal forms. Each of the crystal forms for the same compound is referred to as one of its crystal polymorphs. All polymorphs contain the same molecule, but each polymorph has its own distinct crystal lattice and/or space group symmetry and/or asymmetric unit contents (Buerger, 1936a; Buerger, 1936b). A number of different polymorphs can grow from a few related crystallisation conditions—or even from crystallisation solutions that are nominally the same but experience slight stochastic variations in variables such as temperature, rate of evaporation, and impurities (Jurnak, 1985; Carter et al., 1994; Zabara et al., 2011; Yekwa et al., 2017).

The occurrence of crystal polymorphism during FBLD efforts is not uncommon. For example, in a recent crystallographic screening project, 364 diffraction datasets were collected each from a crystal individually soaked with one compound from a library; of these, 16 crystals belonged to the orthorhombic  $P2_12_12_1$  space group instead of the common monoclinic  $P2_1$  form; the two unit cells were also closely related (Schiebel et al., 2016). It is of note that systematic exploration of crystal polymorphism prior to FBLD can be of great advantage: the best diffracting polymorph can be selected, and various lattices expose different potential drug-binding sites (Vera et al., 2013). Unfortunately, implementation of polymorph assignment in automated FBLD data processing pipelines still leaves some to be desired.

For example, the XCE FBLD X-ray data processing pipeline used at Diamond Light Source (Krojer et al., 2017; Collins et al., 2018; Douangamath et al., 2021; Pearce et al., 2022) decides on the crystal polymorph using a comparison of point group symmetry and cell (ignoring the information encoded by the known polymorph atomic models), or it leaves polymorph attribution to manual intervention on the part of the user. The FBLD efforts at the BESSY and MAXIV synchrotrons use *FragMAXapp* for data processing (Lima et al., 2021); at EMBL Grenoble, the *CRIMS* suite is a large-scale, automated fragment screening pipeline enabling evaluation of libraries of over 1,000 fragments (Cornaci et al., 2021); the IspyB system used at some synchrotrons allows for data integration and storage in alternative lattices in parallel, specifically to address the possibility of multiple polymorphs (Monaco et al., 2013); the FBLD efforts at the Swiss Light Source (SLS) rely on the *FFCS* processing pipeline (Kaminski et al., 2022). Regrettably, none of these FBLD data processing systems has a mechanism in place for automated polymorph assignment.

Overall, current implementations of the FBLD discovery process can be plagued by space group and cell ambiguity, especially if the polymorphs at hand are variations of the same lattice and, therefore, difficult to tell apart from their

morphology and/or their apparent crystal lattices and point groups.

In order to expedite the analysis of the  $\text{CtUGGT}_{\text{GT24}}^{\text{UDP-Glc}}$  FBLD X-ray diffraction datasets, we have written and described here a series of Unix shell scripts called *CoALLA* (crystal polymorph and ligand likelihood-based Assignment). The *CoALLA* scripts are written in Unix shell and use *autoPROC* (Vonnrhein et al., 2011) for data processing, *CCP4-Dimple/REFMAC5* (Murshudov et al., 2011; Winn et al., 2011; Keegan et al., 2015) and *BUSTER* (Blanc et al., 2004; Bricogne et al., 2017) for refinement, and *RHOFIT* (Vonnrhein et al., 2011) for ligand placement.

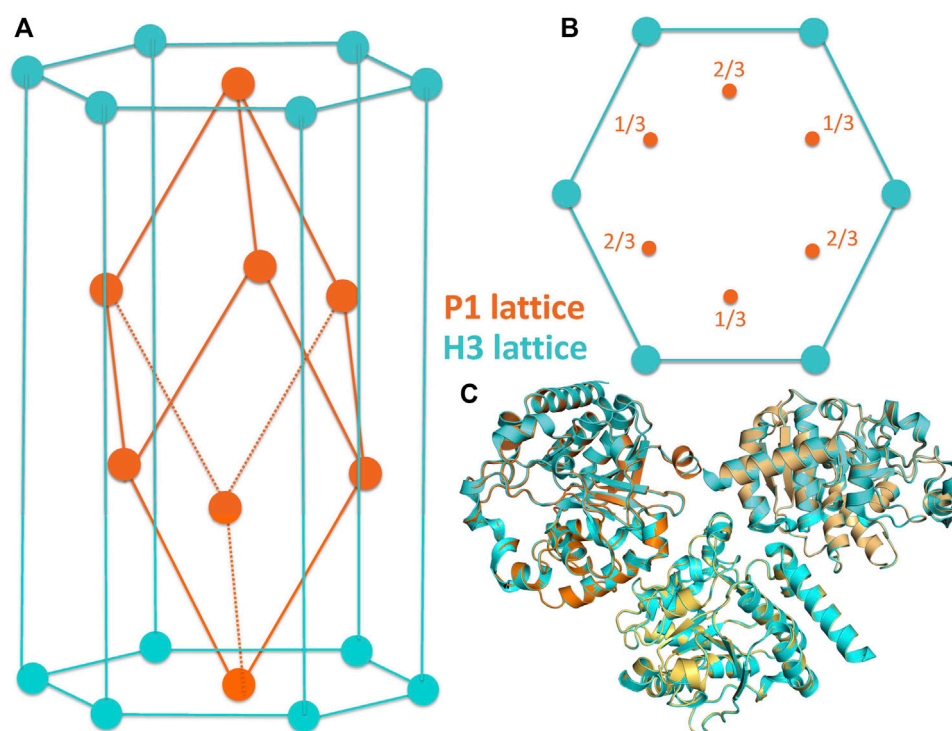
Unique to *CoALLA* is the implementation of the choice of the polymorph, which is effected by carrying out diffraction data indexing, integration, scaling, and structural refinement in each of the possible polymorphs. The most likely polymorph is then chosen based on the best structure refinement statistics.

## 2 $\text{CtUGGT}_{\text{GT24}}^{\text{UDP-Glc}}$ crystal structure

The crystals of  $\text{CtUGGT}_{\text{GT24}}^{\text{UDP-Glc}}$  used for the FBLD effort belonged to the space group H3 (with cell edges  $a = b = 118.8 \text{ \AA}$  and  $c = 68.8 \text{ \AA}$  (cyan cell in Figures 2A,B, PDB ID 6FSN)) with one molecule per asymmetric unit. The crystals likely capture a conformation of UDP-Glc following initial binding to the protein: the co-factor's ribose ring points towards the solvent (Figure 3A). The uracyl ring O4 atom accepts a hydrogen bond from the main chain NH of S1207, and its N<sub>3</sub> atom donates one hydrogen bond to the main chain O of the same residue (top of Figure 3A). Half of the coordination sphere of the  $\text{Ca}^{2+}$  ion in the  $\text{CtUGGT}_{\text{GT24}}$  active site is occupied by the side chains from D1302 and D1304 belonging to the UGGT conserved  $^{1302}\text{DAD}^{1304}$  motif and the side chain of the conserved D1435; the remaining three sites are occupied by an O atom from the  $\beta$  phosphate of UDP-Glc, the O2' atom of the Glc ring, and a water molecule (Figure 3A).

### 2.1 Fragment-soaked $\text{CtUGGT}_{\text{GT24}}^{\text{UDP-Glc}}$ crystal index in related P1 or H3 lattices

To discover  $\text{CtUGGT}_{\text{GT24}}$  ligands by FBLD, 768  $\text{CtUGGT}_{\text{GT24}}^{\text{UDP-Glc}}$  crystals were soaked with as many compounds, and X-ray diffraction datasets were collected from each soaked crystal (see Table 1). During the initial analysis of these X-ray diffraction datasets, we discovered that, upon soaking, crystal symmetry was sometimes lowered to P1, with three molecules in the asymmetric unit of a pseudo-rhombohedral primitive cell (orange cell in Figures 2A,B, PDB ID 7ZLU) of dimensions  $a = 66.3 \text{ \AA}$ ,  $b = 72.7 \text{ \AA}$ ,  $c = 72.7 \text{ \AA}$ ,  $\alpha = 111.2^\circ$ ,  $\beta = 107.7^\circ$ , and  $\gamma = 107.7^\circ$ . This cell is a distortion of the

**FIGURE 2**

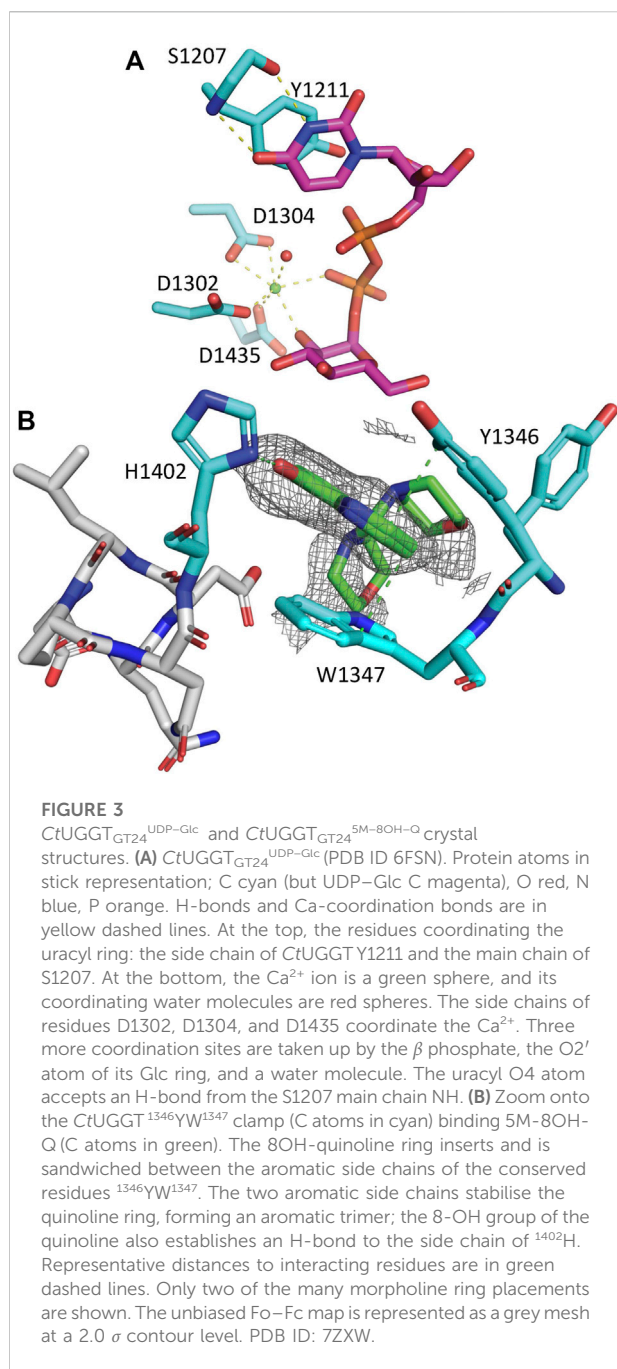
(A,B) Related crystal symmetries and lattices of the H3 (cyan) and P1 (orange) crystal forms of the *CtUGGT<sub>GT24</sub>* crystals. (C) The molecule in the asymmetric unit of the H3 crystal is shown, together with two of its symmetry mates (cyan), in cartoon representation. This portion of the H3 lattice has been overlaid onto the asymmetric unit of the P1 crystals (three chains, painted orange, light orange, and yellow-orange; also in cartoon representation) by superposing the "A" of the H3 crystal to the "A" chain in the P1 crystal.

R3 primitive rhombohedral cell, which for the *CtUGGT<sub>GT24</sub><sup>UDP-Glc</sup>* H3 crystals used for the soaks, would have lattice parameters  $a = b = c = 72.32 \text{ \AA}$  and  $\alpha = \beta = \gamma = 110.43^\circ$ . The three molecules in the P1 crystal asymmetric unit are related by an NCS threefold axis equivalent to the H3 crystallographic threefold axis. The two packings are difficult to distinguish (see Figure 2C), and often a soaked *CtUGGT<sub>GT24</sub><sup>UDP-Glc</sup>* X-ray diffraction dataset can index in both the H3 and P1 lattices, depending, for example, on the program used for indexing or the parameters of the automated indexing algorithm.

The *CtUGGT<sub>GT24</sub>* FBLD datasets were initially processed through the XCE workflow tool (Krojer et al., 2017). Space group assignment in the XCE workflow relies on automated data processing taking place during automated data collection on the Diamond I04-1 beamline (Douangamath et al., 2021), but the decision about the possible space group and cell of a certain dataset is made without consulting available models. In essence, the initial XCE automated decision regarding the space group is based on X-ray diffraction scaling statistics and the cell and point group of the dataset. For example, XCE reads the PDB header of each declared polymorph and determines the point group and

unit cell volume. Then, once processing each fragment-soaked dataset, if the MTZ point group is identical to the one of a reference polymorph and the unit cell volume is similar (within 12%, but this can be tuned), it assumes that they belong to the same crystal form. XCE also allows for the detection of unexpected crystal forms for a subset of the collected crystals: in the presence of different polymorphs, discrepancies between the reference files provided and some of the datasets arise—either in terms of the space groups and/or the unit cells' volumes or because of high  $R_{\text{free}}$  values after initial refinement. In the presence of crystal polymorphism, a certain degree of manual curation is, therefore, needed to run XCE successfully.

When faced with processing hundreds of datasets belonging to different and yet related crystal forms, we reasoned that from a Bayesian statistical standpoint (Bricogne, 1997), the best way of deciding on the correct crystal polymorph is the one that consults all the available data. In this conceptual framework, the likelihood of a certain dataset belonging to one of the known polymorphs can be best evaluated by refining, in turn, each known polymorphic structure against the dataset processed in that symmetry/cell and then choosing the one with the best refinement  $R_{\text{free}}$  statistic.



### 3 Results on the CtUGGT<sub>GT24</sub> FBLD effort

More than 1,200 CtUGGT<sub>GT24</sub> crystals were grown, and 768 of them were soaked with as many compounds of the DSi-Poised Library (Diamond-SGC-iNext, ex DSPL, <https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening/Fragment-Libraries/DSi-Poised-Library.html>).

Each compound was at a concentration of 500 mM in deuterated DMSO; 40 nl of compound stock solution was

dispensed by the Echo robot onto each 200-nl drop, making the final compound concentration 83 mM in 20% DMSO.

Automated data collection was carried out on 692 soaked crystals, but only 493 scaled datasets were obtained, spanning the resolution range of 1.72–11.4 Å. The majority of datasets diffracted to 3.5 Å resolution or better, see Figure 5A.

Only for 439 of these 493 datasets, a structure could be successfully refined in either H3 or P1 or both. Table 1 reports the overall counts, while Figures 5A,B report a few statistics for the 439 diffraction datasets that refined better either in H3 (390 datasets) or P1 (49 datasets).

On average, H3 crystals diffract better than P1 ones:  $\langle {}^{\text{H3}}\text{Resolution} \rangle = 2.53 \text{ Å}$  and  $\langle {}^{\text{P1}}\text{Resolution} \rangle = 3.62 \text{ Å}$  (Figure 5A). Indeed, the scaling statistics are on average better in H3 than in P1:  $\langle {}^{\text{H3}}R_{\text{meas}} \rangle = 0.171$  and  $\langle {}^{\text{P1}}R_{\text{meas}} \rangle = 0.341$  (Figure 5B).

None of the 39 P1 crystals revealed any bound fragments. The P1 2.05 Å structure of a co-crystal of the CtUGGT<sub>GT24</sub> domain with the UDP-Glc analogue (and UGGT inhibitor) UDP-2-deoxy-2-fluoro-D-glucose (U2F) was obtained after the FBLD effort and is available at the Protein Databank (<https://www.rcsb.org>) as PDB ID 7ZLU.

Three of the H3-soaked crystals turned out to contain density for a bound CtUGGT<sub>GT24</sub> ligand.

#### 3.1 Fragment x0441

A CtUGGT<sub>GT24</sub><sup>UDP-Glc</sup> crystal soaked in (1-(1-ethyl-1H-pyrazol-5-yl)-N-methylmethanamine (SMILES string CCN1N=CC=C1CNC, “fragment x0441”) shows density for the ligand at a crystal contact, but the binding site is not particularly conserved: CtUGGT Y1350, H1402, Q1381, and M1403 (which in HsUGGT1 corresponds to H1406, N1458, Q1437, and M1459) (Figures 1, 4). The pose is rather ambiguous at this resolution (2.44 Å). This site, in the context of the full-length UGGT molecule, faces the UGGT central saddle and is proximal to the putative binding site of the first GlcNAc of the UGGT client’s glycan (close to CtUGGT H1402, which is HsUGGT1 N1508).

#### 3.2 Fragment x0763

The CtUGGT<sub>GT24</sub><sup>UDP-Glc</sup> crystal soaked in (1,3-dimethyl-N-(propan-2-yl)-1H-pyrazole-5-carboxamide (SMILES string CC(C)NC(=O)C=1C=C(C)N(C)N1, “fragment x0763”) also showed residual electron density for the ligand (in two orientations), this time in a conserved pocket: CtUGGT T1442, M1441, R1333, R1452, and E1444 (which in HsUGGT1 corresponds to T1498, M1497, R1389, R1508, and E1500) (Figures 1, 4). The pocket is not far from the putative binding site for the C-branch of the client glycoprotein glycan (CtUGGT M1336, Y1339, and M1441 or HsUGGT1 M1392, Y1395, and M1507), but the interactions of the fragment with the

TABLE 1 Summary of the FBLD effort on CtUGGT<sub>GT24</sub>.

Crystallisation experiments (sitting drops)~ 1,800	
Soak experiments	768
Soaked crystals mounted/cryoprotected	692
Diffraction experiments	732
Processed diffraction datasets (H3-only, P1-only, or both)	493 (266, 25, 202)
Successfully refined (lower R <sub>free</sub> in H3 and lower R <sub>free</sub> in P1)	439 (390, 49)
Fragments bound	3

More than one dataset was collected from some crystals. A total of 96 crystals were soaked and mounted/cryoprotected but were not irradiated due to a mistake in loading the sample changer.

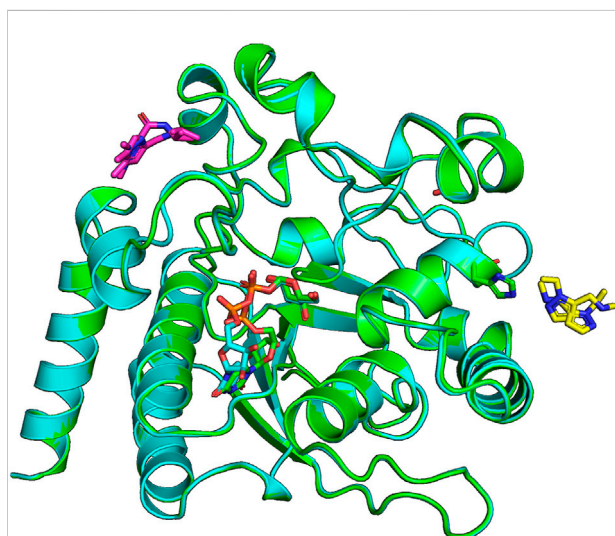


FIGURE 4

Ligands x0441 and x0763 bind to surface pockets of the CtUGGT<sub>GT24</sub> domain. The crystal structures of CtUGGT<sub>GT24</sub><sup>UDP-Glc</sup> soaked in compounds x0441 and x0763 are painted in green and cyan cartoon representation, respectively. The molecule of UDP-Glc in the catalytic pocket of each structure is represented in sticks (its C atoms also in green and cyan for the x0441 and x0763 soaked crystal structure, respectively; N atoms in blue, O atoms in red, and P atoms in orange). Two partially overlapping poses of compound x0673 are represented in sticks with magenta C atoms. Two partially overlapping poses of compound x0441 are represented in sticks with yellow C atoms, next to CtUGGT<sub>GT24</sub> residue H1042, also represented in sticks.

GT domain are weak—perhaps because the binding site is at a crystal contact.

### 3.3 Fragment x0248: CtUGGT<sub>GT24</sub><sup>5M-8OH-Q</sup> crystal structure

The best hit of the FBLD effort was 5-[(morpholin-4-yl)methyl]quinolin-8-ol (SMILES string C1COCCN1CC2=C3C=CC=NC3=C(C=C2)O, “fragment

x0248,” henceforth 5M-8OH-Q), a crystal soaked with which diffracted to 2.25 Å (PDB ID 7ZXW). Diffraction data from this crystal process in H3 with R<sub>meas</sub><sup>H3</sup> = 0.323 and in P1 with R<sub>meas</sub><sup>P1</sup> = 0.206. The CCP4-Dimple/REFMAC5 R<sub>free</sub>s are R<sub>free</sub><sup>H3</sup> = 0.227 and R<sub>free</sub><sup>P1</sup> = 0.2537.

The compound binds to a conserved patch on the surface of the CtUGGT<sub>GT24</sub> domain, about 15 Å away from the UDP-Glc binding site (Figure 3B). The morpholine ring is partially disordered in the crystal, but one of its ring placements is 4.2 Å from the conserved <sup>1396</sup>DQD<sup>1398</sup> motif coordinating the Glc ring of UDP-Glc; the ligand also causes a displacement of the side chain of CtUGGT<sub>GT24</sub><sup>1346</sup>Y. Through this displacement, the 8OH-quinoline ring inserts and is sandwiched between the aromatic side chains of the conserved residues <sup>1346</sup>YW<sup>1347</sup>—which we propose to call the “YW clamp.” The two aromatic side chains stabilise the quinoline ring, forming an aromatic trimer (Lanzarotti et al., 2020); the 8-OH group of the quinoline also establishes an H-bond to the side chain of <sup>1402</sup>H (Figure 3B).

Tables 2, 3 report the crystallographic data and refinement statistics for the CtUGGT<sub>GT24</sub><sup>UDP-Glc</sup> and CtUGGT<sub>GT24</sub><sup>5M-8OH-Q</sup> crystal structures.

## 4 Discussion

In the course of an X-ray crystallography FBLD effort aimed at the discovery of molecules binding the catalytic domain of *Chaetomium thermophilum* UGGT (CtUGGT<sub>GT24</sub>), we encountered crystal polymorphism: the majority of CtUGGT<sub>GT24</sub> crystals belonged to space group H3, but some crystals lowered their symmetry to an equivalent P1 pseudo-rhombohedral lattice. The space group and lattice ambiguity was tricky to resolve by the automated FBLD data processing algorithm—which essentially did not commit to a specific polymorph and left the polymorph choice to the user. However, how is such a choice best implemented in an automated and reliable fashion?

First, if the coordinates of enough indexing spots are gathered, the reciprocal cell parameters can be rather



TABLE 2 X-ray data collection parameters and data processing statistics for *CtUGGT<sub>GT24</sub>* crystal structures.

Structure	<i>CtUGGT<sub>GT24</sub></i> <sup>UDP-Glc</sup>	<i>CtUGGT<sub>GT24</sub></i> <sup>5M-8OH-Q</sup>
PDB ID	6FSN	7ZXW
Beamline	I03@DLS	I04-1@DLS
Wavelength $\lambda$ (mm, Å)	0.97960	0.91587
Transmission %	100	100
Number of images	1,800	1,000
Oscillation range (°)	0.1	0.18
Exposure time (s)	0.02	0.06
Space group (Z)	H3 (6)	H3 (6)
Cell edges: a, b, and c (Å)	a = b = 118.83 and c = 68.75	a = b = 118.01, c = 68.37
Cell angles $\alpha$ , $\beta$ , $\gamma$ (°)	$\alpha = \beta = 90$ , $\gamma = 120$	$\alpha = \beta = 90$ , $\gamma = 120$
Resolution range (Å)	57.17–1.27 (1.34–1.27)	59.00–2.53 (2.37–2.25)
R <sub>merge</sub>	0.07 (1.40)	0.29 (1.95)
R <sub>meas</sub>	0.08 (1.56)	0.32 (2.16)
Observations	486,484 (68,419)	86,970 (12,914)
Unique observations	94,855 (13,700)	16,926 (2,464)
Average I/ $\sigma$ (I)	10.5 (1.1)	6.1 (1.2)
Completeness %	99.0 (97.7)	99.9 (99.9)
Multiplicity	5.1 (5.0)	5.1 (5.2)
CC <sub>1/2</sub>	0.99 (0.50)	0.98 (0.37)

Each structure was determined using a single crystal. Values in parentheses refer to the highest resolution shell.

TABLE 3 Refinement statistics for *CtUGGT<sub>GT24</sub>* crystal structures. All structures contain a Ca<sup>+</sup> ion coming from the protein solution.

Structure	<i>CtUGGT<sub>GT24</sub></i> <sup>UDP-Glc</sup>	<i>CtUGGT<sub>GT24</sub></i> <sup>5M-8OH-Q</sup>
PDB ID	6FSN	7ZXW
Wavelength $\lambda$ (Å)	0.97960	0.91587
Space group (Z)	H3 (6)	H3 (6)
Resolution range (Å)	57.17–1.27 (1.34–1.27)	59.01–2.25 (2.26–2.25)
R <sub>work</sub> R <sub>free</sub>	0.211 (0.227)	0.202 (0.240)
Protein atoms (< B factor >, Å <sup>2</sup> )	2,426 (25.5)	2,380 (34.5)
Water molecules (< B factor >, Å <sup>2</sup> )	252 (34.69)	194 (40.5)
Ligands (< B factor >, Å <sup>2</sup> )	Ca <sup>2+</sup> , UDP-Glc (20.44)	Ca <sup>2+</sup> , UDP-Glc, 5M-8OH-Q (59.2)
rmsd <sub>bonds</sub> (Å), rmsd <sub>angles</sub> (°)	0.01, 1.06	0.008, 0.95

Each structure was determined using a single crystal. Values in parentheses refer to the highest resolution shell.

precise. Errors in the reciprocal lattice choice can then be estimated by the differences between reciprocal cell parameters that would differ in the two reciprocal lattices being compared. For example, for the *CtUGGT<sub>GT24</sub>* crystals described in this study, *b*<sup>\*</sup> and *c*<sup>\*</sup> would be identical in R3 but differ in P1.

In cases where alternative lattices are pseudo-equivalent and symmetry/cell changes are such that the volume of the asymmetric unit increases/decreases, a discriminating criterion between polymorphs can be based on indexing quality or average

intensity of classes of reflections that are systematically extinct in one polymorph and allowed in another one. For example, if indexing and integration of a lattice-centred dataset are carried out in a primitive lattice, the data will appear pseudo-lattice-centred, with more reflections in the primitive than in the lattice-centred space group, and the question that will help discriminate the space groups is then how strong are the additional reflections only allowed in the primitive cell polymorph compared to those that are present in both polymorphs. The values of the fractional mean intensity of the additional reflections, perhaps as a function

of resolution, would help discriminate between a truly centred lattice and a pseudo-centred one.

In the general case in which the polymorphs have equivalent lattices but different symmetries, or unrelated lattices altogether, a polymorph choice based on indexing quality (or average intensity of classes of reflections) may not be so straightforward.

A second class of statistics that may be conceivably used for polymorph discrimination are scaling statistics, but each commonly evaluated scaling statistic risks opening a different can of worms if chosen as the basis for polymorph discrimination. For example,  $R_{\text{merge}}$  has been shown to privilege lower-symmetry space groups, and  $R_{\text{meas}}$  was introduced to discriminate between space groups with different scaling multiplicities (Diederichs and Karplus, 1997; Karplus and Diederichs, 2015). Unfortunately, a robust estimation of  $R_{\text{meas}}$  requires data multiplicity—and so does the estimation of  $CC_{1/2}$  (Karplus and Diederichs, 2015): both scaling statistics are strongly dependent on random (*i.e.*, photon-counting) errors and may not discriminate very well between low-signal but correct symmetry and high-signal but wrong symmetry. Data multiplicity will be high enough for reliable polymorph choice (in the presence of polymorphs of low symmetry) only if more than half a reciprocal sphere of data is collected and/or a data collection strategy is followed—both of which are uncommon practices in most FBLD efforts. For example, in P1, two (or more) datasets of  $180^\circ$  each, with a large kappa offset, would be needed. Data of this kind may not be always available for low-symmetry space group FBLD datasets, thus ruling out the implementation of strategies based on statistics like  $R_{\text{meas}}$  or  $CC_{1/2}$  or those where the error model comes into play, like average  $I/\sigma(I)$  and its sister statistic, asymptotic  $I/\sigma(I)_{\text{asympt}}$  (ISa) (Diederichs, 2010). Supplementary Figure S1 illustrates scatter plots of outer shell  $I/\sigma(I)$ , outer shell  $CC_{1/2}$ ,  $I/\sigma(I)_{\text{asympt}}$  (ISa), and  $R_{\text{meas}}$  separately for *CtUGGT*<sub>GT24</sub> datasets that had lower  $R_{\text{free}}$  in H3 or P1. None of these data processing-based statistics would lead to polymorph choice in overall agreement with the  $R_{\text{free}}$  one.

We decided to test a polymorph choice strategy that would consult the known H3 and P1 atomic structures, and for each dataset use both of them individually in refinement against the diffraction data (processed in that symmetry), in order to resolve the H3 vs. P1 decision, which would then have the advantage of being based on all the prior information available (Jaynes, 1968).

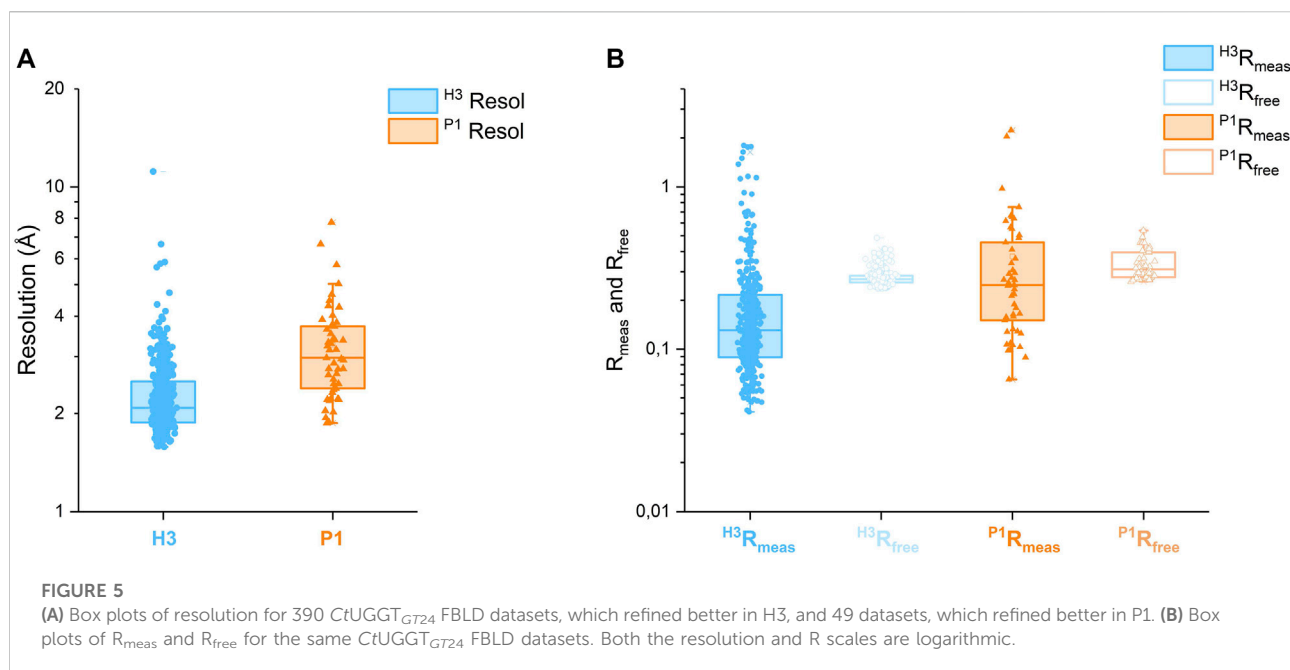
Polymorph choice based on refinement statistics of course has its own risks/drawbacks. First, reliable refinements in all possible polymorphs depend on the availability of a structural model of good quality for each of them. Second, when the lattices of the polymorphs at hand are related, to properly compare  $R_{\text{free}}$  values, the free set of reflections should be chosen consistently: for example, the free set of a lower-symmetry polymorph should be a symmetry expansion of the pseudo-equivalent higher-symmetry one. Finally, the standard error of an R value is (roughly) inversely related to the square root of the number

of reflections. So how big does a difference of  $R_{\text{free}}$  values have to be so that it is statistically significant? A statistical test would be needed to judge if the improvement on  $R_{\text{free}}$  warrants the choice of the lower-symmetry space group over the higher-symmetry one.

More generally, we are aware that, strictly speaking, the use we make of  $R_{\text{free}}$  is not what the statistic was invented for (Brunger, 1992), that is, discrimination between alternative models in view of one set of diffraction structure factor amplitudes. In the case of polymorph choice, the single set of data to account for are the diffraction images, while the alternative models to be tested against the diffraction data comprise lattice parameters (including the orientation matrix) and space group symmetry as well as the usual atomic model parameters.

The best approach to the problem would, therefore, require a single integrated piece of software that refines two classes of parameters against unprocessed X-ray diffraction images: the first class of parameters are the ones traditionally refined during X-ray data processing, and the second class are the parameters traditionally refined during macromolecular refinement. To be most useful for the purpose of polymorph choice, such a program would likely need to deal at least with some of the correlations between parameters belonging to either class (Roversi and Tronrud, 2021). Each refinement of the atomic structures of one of the possible polymorphs directly against the diffraction images would compute a  $R_{\text{free}}$ , a free likelihood or another model comparison metric, (Babcock et al., 2018) and presumably enable the choice of the best polymorph as the one that uses only as few parameters as are needed to fit the signal but not the noise in the data, and no more.

Rather than a statistically solid (and likely time-consuming!) solution to the problem of polymorph choice in FBLD, we have aimed here at the implementation and testing of a simple protocol that would nevertheless choose polymorphs by consulting the information in the available atomic models. The scripts implementing the automated polymorph choice processed each of the 493 *CtUGGT*<sub>GT24</sub> datasets both in H3 and in P1, enabling 439 structural refinements to take place in either or both space groups. For the 156 datasets that were refined in both H3 and P1, the polymorph choice was then based on the symmetry giving the lower  $R_{\text{free}}$ . For each dataset, in each polymorph, our scripts inherit the free set of reflections from the reference dataset in that polymorph so that the choice of free sets for a pair of related-lattice polymorphs can be made consistently once and for all before running the refinements on which the polymorph choice is based. Unsurprisingly, analysis of the distributions of  $R_{\text{meas}}$  and  $R_{\text{free}}$  over these datasets reveals that in both polymorphs, the dispersion of the latter statistic is sharper than the one of the former (see Figure 5B), supporting the choice of the polymorph based on  $R_{\text{free}}$  rather than  $R_{\text{meas}}$ . The best hit was a *CtUGGT*<sub>GT24</sub> 2.2 Å H3 crystal soaked in 5M-8OH-Q. The molecule is bound to a conserved



patch on the surface of the protein. The compound is now the starting point for a medicinal chemistry programme that will develop more potent and selective UGGT inhibitors.

As to the generality/feasibility of our strategy, it is true that its computational requirements scale linearly with the number of possible polymorphs. For a given dataset and  $n$  possible polymorphs, all our approach requires is full data integration, scaling, and refinement in each polymorph. Synchrotron high-throughput automated data processing pipelines run such a series of steps many times per dataset already: for example, at some synchrotrons, the *IspyB* system (Monaco et al., 2013) enables data reduction and storage of each dataset at least three times: with *xia2* and *xia2.multiplex* (Winter et al., 2013; Gildea et al., 2022), *DIALS* (Winter et al., 2022), and *autoPROC* (Vonnrhein et al., 2011), run with a variety of defaults and making use of programs such as *XDS* (Kabsch, 2010) and/or toolkits such as *CCTBX* (Grosse-Kunstleve et al., 2002). Each of these data processing strands is then completed by a structural refinement. One of the reasons for this “redundancy” is indeed related to the possibility that different data processing algorithms may choose different polymorphs, thus alerting the user to the possibility of alternative symmetries/lattices. As we do not know of examples of fragment-based lead discovery efforts encountering more than  $n = 3$  polymorphs, one possible solution to make a strategy equivalent to the one we suggest and yet retain generality and feasibility would be commitment to one of the currently implemented data processing pipelines and its repetition for each of  $n$  possible polymorphs. In most cases, if  $n$  is not large, such an approach would not add much extra time, and it would still systematically sample all possible polymorphs.

Whichever the best solution, implementation of automated polymorph assignment will be an important step towards the realisation of the full potential of crystal polymorphism in FBLD (Vera et al., 2013). It is our hope that with minimal tweaks, existing pipelines for FBLD data processing can be modified to implement ideas similar to the ones we have described here, when discriminating between polymorphs which are related and, therefore, difficult to distinguish on the basis of cell parameters, diffraction data apparent point group, and scaling statistics alone.

## 5 Materials and methods

### 5.1 CtUGGT<sub>GT24</sub> cloning, protein expression, and purification

The C-terminally 6xHis-tagged CtUGGT<sub>GT24</sub> construct corresponding to CtUGGT residues 1,187–1,473 was successfully amplified by PCR starting from the CtUGGT-pHLSec vector (Roversi et al., 2017) using primers CtUGGT<sub>GT24</sub>\_Fwd: ggttgctagctgaaccggtGAGGCAACCAAGTCCGTG and CtUGGT<sub>GT24</sub>\_Rev: gatggtggtgctgtgtaccTTCCC TCACTCTCCTCGC.

The amplified insert was identified by agarose gel electrophoresis (about 900 bp) and purified from the gel.

Following purification of the PCR products, the CtUGGT<sub>GT24</sub> insert was assembled into the AgeI/KpnI linearised pHLSec vector (Aricescu et al., 2006) via ligation-independent cloning (*aka* Gibson assembly). After transformation and plating, *E. coli* colonies containing the

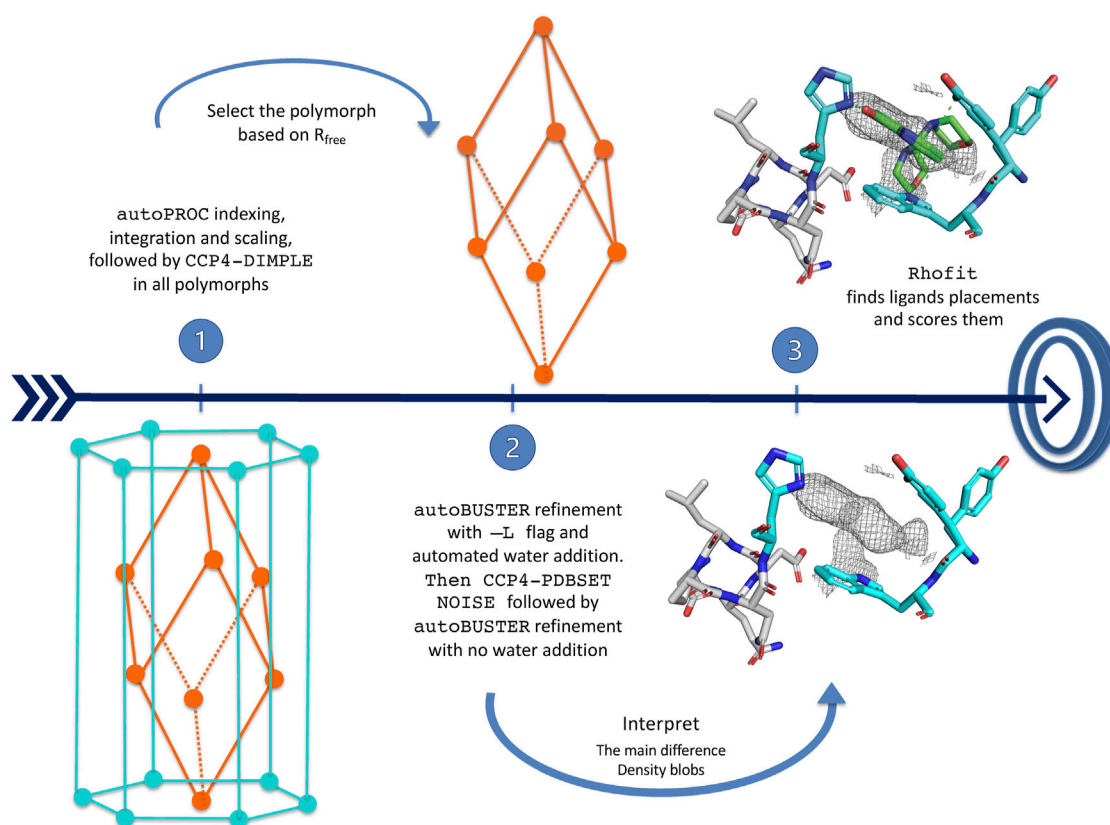


FIGURE 6  
CoALLA flow diagram.

desired construct were identified by colony PCR through identification by agarose gel electrophoresis of the correct size. *CtUGGT*-pHLsec DNA plasmid purification from the correctly identified colonies was carried out *via* DNA miniprep and the resulting plasmid DNA sent for sequencing for confirmation of the desired DNA construct.

The maxiprepmed *CtUGGT*-pHLsec DNA plasmid was transfected into HEK293F cells following the protocol used for *CtUGGT* (Roversi et al., 2017). Purification was achieved by IMAC on an Äkta FPLC system, followed by gel filtration chromatography, after which the proteins were identified by SDS-PAGE. The final buffer was 20 mM HEPES pH 7.4, 50 mM NaCl, and 1 mM  $\text{CaCl}_2$ .

## 5.2 *CtUGGT*<sub>GT24</sub> co-crystallization with UDP-Glc

All crystals were grown at 18°C in sitting drops by the vapour diffusion method, set up with a mosquito liquid handling robot (TTP Labtech). Crystallisation drops had an initial volume of 200 nL. The volume ratio of protein to precipitant was either 1:1 or 2:1.

Crystals of *CtUGGT*<sub>GT24</sub><sup>UDP-Glc</sup> grew in 1 week in a 1:1 mixture of *CtUGGT* at 6 mg/ml, 2 mM  $\text{CaCl}_2$ , and 5 mM UDP-Glc and a number of Morpheus screen conditions (Gorrec, 2009; Gorrec, 2015).

The best crystals came from a crystal grown in Morpheus screen condition 2-9 containing 0.12 M ethylene glycol, 0.1 M buffer system 3 pH 8.5, and 30% v/v precipitant mix 1 (Gorrec, 2009; Gorrec, 2015).

## 5.3 *CtUGGT*<sub>GT24</sub><sup>UDP-Glc</sup> crystal growth for FBLD

A 96-well deep well block was prepared with 500  $\mu\text{L}$  in each well: 257.5  $\mu\text{L}$  of Morpheus precipitant mix 1 (40% v/v PEG 500 MME; 20% w/v PEG 20000); 25.8  $\mu\text{L}$  of 1 M Bicine (buffer system 3 acid component); 24.3  $\mu\text{L}$  of 1 M Tris (buffer system 3 basic component); 192.4  $\mu\text{L}$  of MilliQ Water. The Hydra robot at the Research Complex in Harwell was first used to transfer 25  $\mu\text{L}$  of the crystallisation solution from the deep well block to each mother liquor well of six MRC 3-well crystallisation plates. Vapour diffusion experiments were set up at the Research



Complex in Harwell in six MRC 3-well crystallisation plates—using a mosquito robot equipped with an anti-evaporation cover with 60% controlled humidity in order to avoid drying up of the crystallisation drops during deposition. A total of  $6 \times 96 \times 3 = 1728$  drops were set up. The CtUGGT<sub>GT24</sub> protein at a concentration of 6.5 mg/ml, in the presence of 1 mM CaCl<sub>2</sub> and 5 mM UDP-Glc, was mixed in protein: mother liquor proportions 1.35:1 (drops a and c) and 2:1 (drop d) in drops of total volume 200 nl, and the crystals were left to grow at 18°C. In less than a week, about two-thirds of the experiments yielded crystals.

## 5.4 CtUGGT<sub>GT24</sub><sup>UDP-Glc</sup> crystal soaking for FBLD

Prior to the fragment-based lead discovery effort, 50 of the CtUGGT<sub>GT24</sub><sup>UDP-Glc</sup> crystals were soaked in 0, 5, 10, 15, 20, and 30% DMSO and diffraction tested; no significant deterioration of the diffraction quality was observed. The space group prior to soaking is H3, one molecule per asymmetric unit, with cell edges  $a = b = 119 \text{ \AA}$ , and  $c = 69 \text{ \AA}$ .

All CtUGGT<sub>GT24</sub><sup>UDP-Glc</sup> crystal drops were imaged, and the best crystals were marked for soaking with the Echo robot at the Xchem facility attached to beamline I04-1 at the Diamond Light Source (Collins et al., 2017; Doungamath et al., 2021). 768 compounds of the DSI-Poised Library (Diamond-SGC-iNext, ex DSPL <https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening/Fragment-Libraries/DSi-Poised-Library.html>) and compound stock solutions at a concentration of 500 mM in deuterated DMSO were each soaked into a crystal drop: 40 nl of the compound was dispensed by the Echo robot (Collins et al., 2017) onto each 200-nl drop, making the final compound concentration 83 mM in 20% DMSO. The soaked crystals were left incubating for a variable time between 2 and 4 h. 692 crystals were fished and cryocooled with the aid of the SGC Shifter Robot (Wright et al., 2021).

## 5.5 CtUGGT<sub>GT24</sub><sup>UDP-Glc</sup> soaked crystal X-ray diffraction for FBLD

Automated data collection was carried out on 596 soaked crystals. Automated loop centring failed about 6% of the time, and about 50 crystals were re-measured with optical centring. The symmetry is sometimes lowered by soaking, the crystal can in this case index in space group P1 with three molecules in the asymmetric unit of a pseudo-rhombohedral cell of dimensions  $a = 66.3 \text{ \AA}$   $b = 72.7 \text{ \AA}$   $c = 72.7 \text{ \AA}$ ,  $\alpha = 111.2^\circ$ ,  $\beta = 107.7^\circ$ , and  $\gamma = 107.7^\circ$ . This cell is related to the one of the rhombohedral setting of the H3 crystals,  $a = b = c = 72.32 \text{ \AA}$  and  $\alpha = \beta = \gamma = 110.43^\circ$ .

## 5.6 CtUGGT<sub>GT24</sub><sup>UDP-Glc</sup> X-ray data processing, model refinement, and ligand fitting for FBLD

Data processing, model refinement, and ligand fitting were carried out with the purpose-written shell script pipeline *CoALLA*. In order to decide on the correct symmetry, each dataset was indexed and integrated both in H3 and in P1 using the *autoPROC* suite of programmes; for each dataset, refinement of the protein model was carried out in both the H3 and P1 form in *autoBUSTER*, with the space group giving rise to the lower Rfree being chosen as the correct one for the calculation of the Fo-Fc map. The SMILES string for each compound was fed together with the refined model and phases in order to attempt docking the ligand using *rhofit*. The best hits were listed by ranking the *rhofit* score and/or CC and the hits inspected in *Coot*.

### 5.6.1 Data processing and polymorph assignment

In order to run *CoALLA*, all possible polymorphs must be known; moreover, two reference files must be available for each polymorph: one X-ray diffraction dataset with experimental structure factor amplitudes [in mtz format (Winn et al., 2011)] and its corresponding structure coordinates (in PDB format).

The *CoALLA* pipeline initially processes the diffraction data in each and every one of the polymorphs listed in input, using the data processing suite *autoPROC* Vonnrhein et al. (2011).

The resolution of the data processed in each polymorph is initially chosen by *autoPROC* using the *autoPROC* command flag -M HighResCutOnCChalf, which by default sets the maximum resolution so that  $CC_{1/2}$  in the outer shell is no lower than 0.30.

At the indexing stage, in order to ensure consistent indexing across all FBLD datasets belonging to the same polymorph, the *autoPROC* command line flag -ref <reference dataset>—is used, enforcing the same indexing choice as the reference diffraction dataset for the polymorph being tested. If the reference cell dimensions differ significantly from any of the autoindexed ones, *autoPROC* may be unable to refine a reference-based indexing solution that fits the data: the reference dataset indexing choice is then enforced only after *autoPROC* indexing/integration by a run of *CCP4-pointless* (Evans, 2011) with the keyword TOLERANCE 10<sup>1</sup>. At this stage, the reference

1. Note that in *Phenix*-based FBLD pipelines, models are mapped to the same frame of reference as an isomorphous structure. The *CoALLA* algorithm choice of enforcing indexing to the polymorph's reference dataset has the advantage that simple rigid-body refinement of the reference structure against the data processed in the chosen polymorph is enough to commit to consistent origin choices for all crystals belonging to a certain polymorph, without the need for any molecular replacement steps and/or origin reconciliation.

dataset  $R_{\text{free}}$  flags are also inherited for each polymorph (using CCP4-CAD) and are kept for all subsequent calculations for the dataset in that polymorph.

The *PanDDA* pipeline (Pearce et al., 2017a; Pearce et al., 2017b; Pearce et al., 2017c) in use at Diamond Light Source beamline I04-1 Douangamath et al. (2021)—where the data were collected—uses CCP4-Dimple/REFMAC5 (Murshudov et al., 2011; Winn et al., 2011; Keegan et al., 2015) as its refinement engine (see Figure 6). At the initial stage of deciding on the correct polymorph, it seemed natural to exploit the fast refinement capabilities of CCP4-Dimple. After data processing in all polymorphs, the hypothesis that the data belong to a certain polymorph is tested in *CoALLA* by running structural refinement in CCP4-Dimple (Winn et al., 2011; Keegan et al., 2015) against the data processed in each polymorph in turn.

In each polymorph, the *CoALLA* scripts run two CCP4-Dimple refinements of the reference model against the dataset as processed in that polymorph: a rigid body refinement first, followed by a full atomic one. In order to be able to compare the  $R_{\text{free}}$ s of models refined in different polymorphs, all CCP4-Dimple refinements are run in *CoALLA* up to the best common maximum resolution limit across all polymorphs, which the scripts obtain for each polymorph by harvesting the resolution limit from the relevant *autoPROC* output file (see Figure 6).

Once a CCP4-Dimple refinement of the structure has run for each polymorph, the polymorph with the best CCP4-Dimple refinement  $R_{\text{free}}$  is chosen as the most likely one to continue the analysis. It is at this stage that *CoALLA* improves on automated polymorph choices based on data processing only: the available information about the structures of the polymorphs is added to the one present in the diffraction data to single out the most likely polymorph.

### 5.6.2 BUSTER structure refinements

The structures are then refined in *BUSTER* starting from the CCP4-Dimple output file, with automated NCS restraints and external restraints to the reference structure (Smart et al., 2012), command line keywords—target <ref.pdb>—autoncs).

The first *BUSTER* refinement is run with the -L command line flag: this turns on automated water updating. Waters initially also fill volumes where potential ligands are located, but the water atoms around residual difference density are then removed at the last *BUSTER* refinement cycle. This procedure enables water-building without deteriorating the quality of the difference density in putative ligand regions.

At this stage, CCP4-PDBSET is run on the *BUSTER*-refined model (with keyword NOISE) in order to introduce positional noise in the coordinates resulting from the refinement previously described—including any waters that survived the pruning at the end of the previous *BUSTER* refinement with automated water addition. The positional noise

introduced by CCP4-PDBSET at this stage ought to wipe out any “memory” of the waters that may have been refined (and then were removed) in putative ligand pockets. A second *BUSTER* refinement (without water addition) is then run starting from the CCP4-PDBSET shaken model. The 2Fo-Fc and Fo-Fc density maps necessary for ligand identification and docking are computed at the end of this *BUSTER* refinement.

### 5.6.3 RHOFIT ligand placement

Automatic interpretation of crystal residual difference density fully automates the placement of ligands, providing an unbiased alternative to manual ligand placement, especially for data of low to medium resolution (Langer et al., 2008; Carolan and Lamzin, 2014; Echols et al., 2014; Wlodek et al., 2006). Algorithms can take into account protein–ligand interactions as well as fit to the map (Mooij et al., 2006).

The *RHOFIT* program (Smart et al., 2014) is used within the *CoALLA* pipeline to find the best placements in all the *BUSTER* refined maps.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <http://www.wwpdb.org/>, 6fsn <http://www.wwpdb.org/>, 7zwx.

## Author contributions

PR, ATC, and NZ conceived the study. PR, CBL, and SV cloned the *CtUGGT<sub>GT24</sub>* construct. PR, JDLC, RI, and SV prepared the *CtUGGT<sub>GT24</sub>* protein. PR, RI, and JDLC grew the *CtUGGT<sub>GT24</sub>* crystals. PR, RI, ATC, BD, MH, and NZ carried out fragment-based lead discovery effort. ATC and PR wrote the *CoALLA* scripts. RI, ATC, BD, MH, GM, and PR analysed data. All authors contributed to the writing of the manuscript.

## Funding

The work was funded by the Glycobiology Endowment and by two University of Oxford Confidence in Concept Schemes, grant references MRC-MC\_PC\_15029 and MRC-MC\_PC\_16056 (to NZ). PR was the recipient of an LISC Wellcome Trust ISSF award, grant reference 204801/Z/16/Z and a Wellcome Trust Seed Award in Science, grant reference 214090/Z/18/Z. RI was the recipient of a Sardinian

Regional Government PhD scholarship. AC was funded by a Wellcome Trust 4-Year Studentship reference 097300/Z/11/Z. NZ is a Fellow of Merton College, Oxford.

## Acknowledgments

Tobias Krojer, Anthony Bradley, Patrick Collins, Alice Douangamath, Jose Brandao-Neto, Frank Von Delft, and all the staff of the I04-1 beamline at the Diamond Light Source helped with sample preparation and data collection. Conor Wild gave insightful comments on our choice of  $R_{\text{free}}$  as a model comparison statistic. Constantina Foutinou helped with the preparation of the manuscript. Tobias Krojer, José A. Marquez, and Justyna A. Wojdyla kindly commented on the availability of automated polymorph choice algorithms in current implementations of FBLD data processing pipelines. Kay Diederichs kindly discussed the suitability of various statistics to compare scaling of a dataset in different symmetries. The authors thank the reviewers for helpful suggestions on the first submitted version of the manuscript.

## References

- Albesa-Jové, D., and Guerin, M. E. (2016). The conformational plasticity of glycosyltransferases. *Curr. Opin. Struct. Biol.* 40, 23–32. Carbohydrate-protein interactions and glycosylation • Biophysical and molecular biological methods. doi:10.1016/j.sbi.2016.07.007
- Alonzi, D. S., Scott, K. A., Dwek, R. A., and Zitzmann, N. (2017). Iminosugar antivirals: The therapeutic sweet spot. *Biochem. Soc. Trans.* 45, 571–582. doi:10.1042/BST20160182
- Amara, J. F., Cheng, S. H., and Smith, A. E. (1992). Intracellular protein trafficking defects in human disease. *Trends Cell. Biol.* 2, 145–149. doi:10.1016/0962-8924(92)90101-r
- Aricescu, A. R., Lu, W., and Jones, E. Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr. D. Biol. Crystallogr.* 62, 1243–1250. doi:10.1107/S0907444906029799
- Babcock, N. S., Keedy, D. A., Fraser, J. S., and Sivak, D. A. (2018). Model selection for biological crystallography. *bioRxiv*. doi:10.1101/448795
- Blanc, E., Roversi, P., Vonnrhein, C., Flensburg, C., Lea, S. M., and Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D. Biol. Crystallogr.* 60, 2210–2221. doi:10.1107/S0907444904016427
- Bricogne, G., Blanc, E. M. B., Flensburg, C., Keller, P., Paciorek, W., et al. (2017). *Transition metals in catalysis: The functional relationship*. BUSTER 2.10.3.
- Bricogne, G. (1997). “The bayesian statistical viewpoint on structure determination: Basic concepts and examples,” in *Macromolecular crystallography*. Editors C. W. Carter Jr., and R. M. Sweet (San Diego, CA: Academic Press), 361–423. vol. 276 of *Methods in Enzymology*.
- Brunger, A. T. (1992). Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475. doi:10.1038/355472a0
- Buerger, M. J. (1936a). The general rôle of composition in polymorphism. *Proc. Natl. Acad. Sci. U. S. A.* 22, 685–689. doi:10.1073/pnas.22.12.685
- Buerger, M. J. (1936b). The kinetic basis of crystal polymorphism. *Proc. Natl. Acad. Sci. U. S. A.* 22, 682–685. doi:10.1073/pnas.22.12.682
- Caputo, A. T., Alonzi, D. S., Marti, L., Reca, I.-B., Kiappes, J. L., Struwe, W. B., et al. (2016). Structures of mammalian ER  $\alpha$ -glucosidase II capture the binding modes of broad-spectrum iminosugar antivirals. *Proc. Natl. Acad. Sci. U. S. A.* 113, E4630–E4638. doi:10.1073/pnas.1604463113
- Carolan, C. G., and Lamzin, V. S. (2014). Automated identification of crystallographic ligands using sparse-density representations. *Acta Crystallogr. D. Biol. Crystallogr.* 70, 1844–1853. doi:10.1107/S1399004714008578
- Carter, C. W., Doublé, S., and Coleman, D. E. (1994). Quantitative analysis of crystal growth. Tryptophanyl-tRNA synthetase crystal polymorphism and its relationship to catalysis. *J. Mol. Biol.* 238, 346–365. doi:10.1006/jmbi.1994.1297
- Chen, X., Qin, S., Chen, S., Li, J., Li, L., Wang, Z., et al. (2015). A ligand-observed mass spectrometry approach integrated into the fragment based lead discovery pipeline. *Sci. Rep.* 5, 8361. doi:10.1038/srep08361
- Ciulli, A., Williams, G., Smith, A. G., Blundell, T. L., and Abell, C. (2006). Probing hot spots at protein-ligand binding sites: A fragment-based approach using biophysical methods. *J. Med. Chem.* 49, 4992–5000. doi:10.1021/jm060490r
- Collins, P. M., Douangamath, A., Talon, R., Dias, A., Brandao-Neto, J., Krojer, T., et al. (2018). “Chapter eleven - achieving a good crystal system for crystallographic x-ray fragment screening,” in *Modern approaches in drug discovery*. Editor C. A. Lesburg (Academic Press), 251–264. vol. 610 of *Methods in Enzymology*. doi:10.1016/bs.mie.2018.09.027
- Collins, P. M., Ng, J. T., Talon, R., Nekrošius, K., Krojer, T., Douangamath, A., et al. (2017). Gentle, fast and effective crystal soaking by acoustic dispensing. *Acta Crystallogr. D. Struct. Biol.* 73, 246–255. doi:10.1107/S205979831700331X
- Cornaciu, I., Bourgeois, R., Hoffmann, G., Dupeux, F., Humm, A.-S., Mariaule, V., et al. (2021). The automated crystallography pipelines at the EMBL HTX facility in Grenoble. *J. Vis. Exp.* (172), e62491. doi:10.3791/62491
- D'Alessio, C., Caramelo, J. J., and Parodi, A. J. (2010). UDP-Glc:glycoprotein glucosyltransferase-glucosidase II, the ying-yang of the ER quality control. *Semin. Cell. Dev. Biol.* 21, 491–499. doi:10.1016/j.semcdb.2009.12.014
- Dalziel, M., Crispin, M., Scanlan, C. N., Zitzmann, N., and Dwek, R. A. (2014). Emerging principles for the therapeutic exploitation of glycosylation. *Science* 343, 1235681. doi:10.1126/science.1235681
- Diederichs, K., and Karplus, P. A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* 4, 269–275. doi:10.1038/nsb0497-269
- Diederichs, K. (2010). Quantifying instrument errors in macromolecular X-ray data sets. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 733–740. doi:10.1107/S0907444910014836

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.960248/full#supplementary-material>. The CoALLA shell script will be available from the corresponding authors upon request.

- Douangamath, A., Powell, A., Fearon, D., Collins, P. M., Talon, R., Krojer, T., et al. (2021). Achieving efficient fragment screening at XChem facility at Diamond light Source. *J. Vis. Exp.* (171), e62414 doi:10.3791/62414
- Dwek, R. A., Bell, J. I., Feldmann, M., and Zitzmann, N. (2022). Host-targeting oral antiviral drugs to prevent pandemics. *Lancet* 399, 1381–1382. doi:10.1016/S0140-6736(22)00454-8
- Echols, N., Moriarty, N. W., Klei, H. E., Afonine, P. V., Bunkóczi, G., Headd, J. J., et al. (2014). Automating crystallographic structure solution and refinement of protein–ligand complexes. *Acta Crystallogr. D. Biol. Crystallogr.* 70, 144–154. doi:10.1107/S139900471302748X
- Evans, P. R. (2011). An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallogr. D. Biol. Crystallogr.* 67, 282–292. doi:10.1107/S090744491003982X
- Gildea, R. J., Beilstein-Edmands, J., Axford, D., Horrell, S., Aller, P., Sandy, J., et al. (2022). xia2.multiplex: a multi-crystal data-analysis pipeline. *Acta Crystallogr. D. Struct. Biol.* 78, 752–769. doi:10.1107/S2059798322004399
- Gorrec, F. (2015). The MORPHEUS II protein crystallization screen. *Acta Crystallogr. F. Struct. Biol. Commun.* 71, 831–837. doi:10.1107/S2053230X1500967X
- Gorrec, F. (2009). The MORPHEUS protein crystallization screen. *J. Appl. Crystallogr.* 42, 1035–1042. doi:10.1107/S0021889809042022
- Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W., and Adams, P. D. (2002). The computational crystallography toolbox: Crystallographic algorithms in a reusable software framework. *J. Appl. Crystallogr.* 35, 126–136. doi:10.1107/S0021889801017824
- Hammond, C., Braakman, I., and Helenius, A. (1994). Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc. Natl. Acad. Sci. U. S. A.* 91, 913–917. doi:10.1073/pnas.91.3.913
- Jaynes, E. T. (1968). Prior probabilities. *IEEE Trans. Syst. Sci. Cyber.* 4, 227–241. doi:10.1109/TSSC.1968.300117
- Jurnak, F. (1985). Induction of elongation factor Tu-GDP crystal polymorphism by polyethylene glycol contaminants. *J. Mol. Biol.* 185, 215–217. doi:10.1016/0022-2836(85)90194-9
- Kabsch, W. (2010). “XDS,” in *Acta crystallographica. Section D, Biological crystallography*, 66 (Chester, England: International Union of Crystallography), 125–132. doi:10.1107/S0907444909047337
- Kaminski, J. W., Vera, L., Stegmann, D. P., Vering, J., Eris, D., Smith, K. M. L., et al. (2022). Fast fragment- and compound-screening pipeline at the Swiss Light Source. *Acta Crystallogr. D. Struct. Biol.* 78, 328–336. doi:10.1107/S2059798322000705
- Karade, S. S., Hill, M. L., Kiappes, J. L., Manne, R., Aakula, B., Zitzmann, N., et al. (2021). N-substituted valiolamine derivatives as potent inhibitors of endoplasmic reticulum  $\alpha$ -glucosidases I and II with antiviral activity. *J. Med. Chem.* 64, 18010–18024. doi:10.1021/acs.jmedchem.1c01377
- Karplus, P. A., and Diederichs, K. (2015). Assessing and maximizing data quality in macromolecular crystallography. *Curr. Opin. Struct. Biol.* 34, 60–68. Carbohydrate-protein interactions • Biophysical and molecular biological methods. doi:10.1016/j.sbi.2015.07.003
- Keegan, R., Wojdyr, M., Winter, G., and Ashton, A. (2015). Dimple: A difference map pipeline for the rapid screening of crystals on the beamline. *Acta Crystallogr. A Found. Adv.* 71, s18. doi:10.1107/S2053273315099702
- Krojer, T., Talon, R., Pearce, N., Collins, P., Douangamath, A., Brandao-Neto, J., et al. (2017). The XChemExplorer graphical workflow tool for routine or large-scale protein–ligand structure determination. *Acta Crystallogr. D. Struct. Biol.* 73, 267–278. doi:10.1107/S2059798316020234
- Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* 3, 1171–1179. doi:10.1038/nprot.2008.91
- Lanzarotti, E., Defelipe, L. A., Marti, M. A., and Turjanski, A. G. (2020). Aromatic clusters in protein–protein and protein–drug complexes. *J. Cheminform.* 12, 30. doi:10.1186/s13321-020-00437-4
- Lima, G. M. A., Jagudin, E., Talibov, V. O., Benz, L. S., Marullo, C., Barthel, T., et al. (2021). FragMAXapp: Crystallographic fragment-screening data-analysis and project-management system. *Acta Crystallogr. D. Struct. Biol.* 77, 799–808. doi:10.1107/S2059798321003818
- Monaco, S., Gordon, E., Bowler, M. W., Delagenière, S., Guijarro, M., Spruce, D., et al. (2013). Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF. *J. Appl. Crystallogr.* 46, 804–810. doi:10.1107/S0021889813006195
- Mooij, W., Hartshorn, M., Tickle, I., Sharff, A., Verdonk, M., and Jhoti, H. (2006). Automated protein–ligand crystallography for structure-based drug design. *ChemMedChem* 1, 827–838. doi:10.1002/cmdc.200600074
- Müller, J., Klein, R., Tarkhanova, O., Gryniukova, A., Borysko, P., Merkl, S., et al. (2022). Magnet for the needle in haystack: “crystal structure first” fragment hits unlock active chemical matter using targeted exploration of vast chemical spaces. *J. Med. Chem.* Epub ahead of print. doi:10.1021/acs.jmedchem.2c00813
- Murray, C. W., and Blundell, T. L. (2010). Structural biology in fragment-based drug design. *Curr. Opin. Struct. Biol.* 20, 497–507. doi:10.1016/j.sbi.2010.04.003
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., et al. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.* 67, 355–367. doi:10.1107/S0907444911001314
- Oksenych, V., and Kainov, D. E. (2022). Broad-spectrum antivirals and antiviral drug combinations. *Viruses* 14, 301. doi:10.3390/v14020301
- Pardi, N., and Weissman, D. (2020). Development of vaccines and antivirals for combating viral pandemics. *Nat. Biomed. Eng.* 4, 1128–1133. doi:10.1038/s41551-020-00658-w
- Pearce, N. M., Bradley, A. R., Krojer, T., Marsden, B. D., Deane, C. M., and von Delft, F. (2017a). Partial-occupancy binders identified by the Pan-Dataset Density Analysis method offer new chemical opportunities and reveal cryptic binding sites. *Struct. Dyn.* 4, 032104. doi:10.1063/1.4974176
- Pearce, N. M., Krojer, T., Bradley, A. R., Collins, P., Nowak, R. P., Talon, R., et al. (2017b). A multi-crystal method for extracting obscured crystallographic densities from conventionally uninterpretable electron density. *Nat. Commun.* 8, 15123. doi:10.1038/ncomms15123
- Pearce, N. M., Krojer, T., and von Delft, F. (2017c). Proper modelling of ligand binding requires an ensemble of bound and unbound states. *Acta Crystallogr. D. Struct. Biol.* 73, 256–266. doi:10.1107/S2059798317003412
- Pearce, N. M., Skyner, R., and Krojer, T. (2022). Experiences from developing software for large X-ray crystallography-driven protein–ligand studies. *Front. Mol. Biosci.* 9, 861491. doi:10.3389/fmolb.2022.861491
- Radoux, C. J., Olsson, T. S. G., Pitt, W. R., Groom, C. R., and Blundell, T. L. (2016). Identifying interactions that determine fragment binding at protein hotspots. *J. Med. Chem.* 59, 4314–4325. doi:10.1021/acs.jmedchem.5b01980
- Roversi, P., Marti, L., Caputo, A. T., Alonzi, D. S., Hill, J. C., Dent, K. C., et al. (2017). Interdomain conformational flexibility underpins the activity of UGGT, the eukaryotic glycoprotein secretion checkpoint. *Proc. Natl. Acad. Sci. U. S. A.* 114, 8544–8549. doi:10.1073/pnas.1703682114
- Roversi, P., and Tronrud, D. E. (2021). Ten things I ‘hate’ about refinement. *Acta Crystallogr. D. Struct. Biol.* 77, 1497–1515. doi:10.1107/S2059798321011700
- Sayce, A. C., Alonzi, D. S., Killingbeck, S. S., Tyrrell, B. E., Hill, M. L., Caputo, A. T., et al. (2016). Iminosugars inhibit dengue virus production via inhibition of ER  $\alpha$ -glucosidases-not glycolipid processing enzymes. *PLoS Negl. Trop. Dis.* 10, e0004524. doi:10.1371/journal.pntd.0004524
- Schiebel, J., Krimmer, S. G., Röwer, K., Knörlein, A., Wang, X., Park, A. Y., et al. (2016). High-throughput crystallography: Reliable and efficient identification of fragment hits. *Structure* 24, 1398–1409. doi:10.1016/j.str.2016.06.010
- Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., et al. (2012). Exploiting structure similarity in refinement: Automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D. Biol. Crystallogr.* 68, 368–380. doi:10.1107/S0907444911056058
- Smart, O., Womack, T., Sharff, A., Flensburg, C., Keller, P., Paciorek, W., et al. (2014). *RHOFIT, version 1.2.4*. Cambridge, United Kingdom: Global Phasing Ltd.
- Tax, G., Lia, A., Santino, A., and Roversi, P. (2019). Modulation of erqc and erad: A broad-spectrum spanner in the works of cancer cells? *J. Oncol.* 2019, 8384913. doi:10.1155/2019/8384913
- Trombetta, E. S., and Helenius, A. (1999). Glycoprotein reglucosylation and nucleotide sugar utilization in the secretory pathway: Identification of a nucleoside diphosphatase in the endoplasmic reticulum. *EMBO J.* 18, 3282–3292. doi:10.1093/emboj/18.12.3282
- Trombetta, S. E., Bosch, M., and Parodi, A. J. (1989). Glucosylation of glycoproteins by mammalian, plant, fungal, and trypanosomatid protozoa microsomal membranes. *Biochemistry* 28, 8108–8116. doi:10.1021/bi00446a022
- Tyrrell, B. E., Sayce, A. C., Warfield, K. L., Miller, J. L., and Zitzmann, N. (2017). Iminosugars: Promising therapeutics for influenza infection. *Crit. Rev. Microbiol.* 43, 521–545. doi:10.1080/1040841X.2016.1242868



- Vera, L., Antoni, C., Devel, L., Czarny, B., Cassar-Lajeunesse, E., Rossello, A., et al. (2013). Screening using polymorphs for the crystallization of protein–ligand complexes. *Cryst. Growth & Des.* 13, 1878–1888. doi:10.1021/cg301537n
- Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., et al. (2011). Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D. Biol. Crystallogr.* 67, 293–302. doi:10.1107/S0907444911007773
- Warfield, K. L., Alonzi, D. S., Hill, J. C., Caputo, A. T., Roversi, P., Kiappes, J. L., et al. (2020). Targeting endoplasmic reticulum  $\alpha$ -glucosidase I with a single-dose iminosugar treatment protects against lethal influenza and dengue virus infections. *J. Med. Chem.* 63, 4205–4214. doi:10.1021/acs.jmedchem.0c00067
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* 67, 235–242. doi:10.1107/S0907444910045749
- Winter, G., Beilsten-Edmands, J., Devenish, N., Gerstel, M., Gildea, R. J., McDonagh, D., et al. (2022). DIALS as a toolkit. *Protein Sci.* 31, 232–250. doi:10.1002/pro.4224
- Winter, G., Lobley, C. M. C., and Prince, S. M. (2013). Decision making in xia2. *Acta Crystallogr. D. Biol. Crystallogr.* 69, 1260–1273. doi:10.1107/S0907444913015308
- Wloddek, S., Skillman, A. G., and Nicholls, A. (2006). Automated ligand placement and refinement with a combined force field and shape potential. *Acta Crystallogr. D. Biol. Crystallogr.* 62, 741–749. doi:10.1107/S0907444906016076
- Wright, N. D., Collins, P., Koekemoer, L., Krojer, T., Talon, R., Nelson, E., et al. (2021). The low-cost Shifter microscope stage transforms the speed and robustness of protein crystal harvesting. *Acta Crystallogr. D. Struct. Biol.* 77, 62–74. doi:10.1107/S2059798320014114
- Yekwa, E., Khourieh, J., Canard, B., Papageorgiou, N., and Ferron, F. (2017). Activity inhibition and crystal polymorphism induced by active-site metal swapping. *Acta Crystallogr. D. Struct. Biol.* 73, 641–649. doi:10.1107/S205979831700866X
- Zabara, A., Amar-Yuli, I., and Mezzenga, R. (2011). Tuning in-meso-crystallized lysozyme polymorphism by lyotropic liquid crystal symmetry. *Langmuir* 27, 6418–6425. doi:10.1021/la200710p

# Frontiers in Molecular Biosciences

Explores biological processes in living organisms  
on a molecular scale

Focuses on the molecular mechanisms  
underpinning and regulating biological processes  
in organisms across all branches of life.

## Discover the latest Research Topics

[See more](#) →

### Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne, Switzerland  
[frontiersin.org](https://frontiersin.org)

### Contact us

+41 (0)21 510 17 00  
[frontiersin.org/about/contact](https://frontiersin.org/about/contact)



### Frontiers in Molecular Biosciences

