

**Crystallographic fragment screening -
improvement of workflow, tools and procedures,
and application for the development of enzyme
and protein-protein interaction modulators**

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by
Tatjana Barthel

2023

1st reviewer: Manfred S. Weiss
Macromolecular Crystallography
Helmholtz-Zentrum für Materialien und Energie

2nd reviewer: Markus C. Wahl
Structural Biochemistry
Freie Universität Berlin

Date of defense: 03.07.2023

Acknowledgement

I want to thank first and foremost Markus Wahl and Manfred Weiss for the opportunity to perform my doctoral thesis project. They are both inspirational men, experts in their fields and taught me how to be an independent scientist. Especially, I want to highlight Manfred Weiss' exceptional determination and support to allow me to involve myself also in other aspects as planning workshops or taking up user support.

I am grateful for the prolonged support and faith in my abilities, from Jan Wollenhaupt. He pushed me to become the best scientist I could be and grow in my scientific but also transferable skills. He always helped me out when I needed help and I appreciate it greatly.

Furthermore, I want to thank the Macromolecular Crystallography Group at HZB for the amazing talks, discussions and overall help. I would have not been able to progress through my project as quickly and effectively as I did without their support. Especially, I want to thank Laila Benz for her hard work in the MPro project.

I want to thank another person from HZB in a different department. Dirk Wallacher helped me with 3D-printing, laser-cutting of materials and his great expertise in sample environment needs. It was a pleasure collaborating with him.

I am highly grateful for the possibility to perform some experiments at the Free University Structural Biology Group led by Markus Wahl. They helped me with any question I had, and I enjoyed spending time with the group. I especially want to highlight Nicole Holton for her welcoming atmosphere and helpful nature.

I want to further thank Gustavo Lima, who made it possible for me to handle the about 1000 dataset-large campaign and helped me with the analysis. I appreciate the time he put in and enjoyed our collaboration.

In addition, I am grateful to have been able to work with Maria Kehr. She assisted with a few experiments in the AR project and was an amazingly helpful.

Furthermore, I want to thank Alexander Metz for his continuous help with the understanding of computational approaches and his insights for KNIME-related questions. It was a pleasure working together with him.

I want to thank especially my family and friends for the amazing support. My parents Christina Vogt and Stefan Barthel helped shape me into the person I am today. Together with my sister Jasmin Barthel, my husband Zakaria Kerboua Laouar and my best friend Mareike Blank, they have helped me through all tough times with their unconditional love and celebrate each success with me no matter how small it may seem. They are my pillars and make such ambitious work possible. I love them all and am grateful to have them in my life.

Selbstständigkeitserklärung:

Hierdurch versichere ich, dass ich meine Dissertation selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet habe. Geistiges Eigentum anderer Autoren wurde als entsprechend gekennzeichnet. Ebenso versichere ich, dass ich an keiner anderen Stelle ein Prüfungsverfahren beantragt bzw. die Dissertation in dieser oder anderer Form an keiner anderen Fakultät als Dissertation vorgelegt habe.

Table of contents

1. Introduction	5
1.1 Drug and Tool Compound Discovery.....	5
1.1.1 Screening and Libraries	7
1.1.1.1 Fragment Library Design	11
1.1.1.2 Assessing a Protein's Druggability	13
1.1.1.3 Choosing the Screening Method	13
1.1.2 Crystallographic Fragment Screening.....	14
1.1.2.1 Protein Crystallization.....	14
1.1.2.2 The Fragment Library Choice.....	15
1.1.2.3 Crystal Soaking	15
1.1.2.4 Crystal Harvesting	17
1.1.2.5 Data Collection	19
1.1.2.6 Data Analysis.....	19
1.1.3 Structure-based drug/tool compound design	22
1.1.3.1 Preparation of Virtual Docking Input.....	22
1.1.3.1.1 Definition of Binding Site	23
1.1.3.1.2 Building a Library of Possible Follow-up Compounds.....	23
1.1.3.2 Docking and Scoring	24
1.1.3.2.1 Template-based Docking	24
1.1.3.3 Re-scoring of Docking Poses	25
1.1.3.4 Follow-up Selection	25
1.2 Enzymes as targets	25
1.2.1 Aspartic Proteases - Endothiapepsin	26
1.2.2 Cysteine proteases – SARS-CoV-2 MPro.....	27
1.3 Protein-Protein Interactions as targets	28
1.3.1 Pre-mRNA Splicing – the spliceosome	29
1.3.2 The interaction of Aar2 and Prp8	31

2. Aims.....	34
3. Results.....	35
3.1 Support for crystal manipulation and harvesting with the EasyAccess Frame	35
3.2 FragMAXapp - Automation of crystallographic fragment screening data analysis and management via a user-friendly web application.....	40
3.3 The F2X-Universal Library and the F2X-Entry Screen enable efficient CFS	47
3.4 Crystallographic fragment screening workflow at HZB	55
3.5 Advancing fragment hits via SAR by catalogue – the Frag4Lead workflow.....	61
3.6 Inherent binding pose validation and discovery of putative novel PPI sites by CFS68	
4. Discussion	74
4.1 Reaching high hit rates through efficient libraries, suitable crystal systems and multiplex analysis.....	74
4.1.1 Advantages of the F2X-Universal Library and F2X-Entry Screen.....	74
4.1.2 Crystal Packing Defines CFS Success	76
4.1.3 Key Developments for Efficient Data Analysis	77
4.2 Speeding up and simplifying CFS experiments and data management.....	79
4.2.1 Protein Crystallization	79
4.2.2 Crystal Handling.....	80
4.2.3 Data Collection.....	81
4.2.4 Data Management.....	81
4.3 What to do with your Fragment Hits?	83
4.3.1 Prioritizing Fragment Hits.....	83
4.3.1.1 Prioritization Based on Target Sites in the AR campaign.....	84
4.3.1.2 Prioritization Based on Inherent Binding Mode Validation.....	86
4.3.2 Optimization of Fragment Hits into Lead Compounds	87
4.3.2.1 Automated Growing Approach via Frag4Lead	87
4.3.2.2 Facilitated Fragment Merging and Linking via Similarity Search.....	89
4.3.3 Discovery of Putative Protein Functions	90
4.3.4 Methodological Investigations of CFS Results.....	90
5. Conclusion	92

6. Appendix – further projects	93
6.1 Cryptic pocket identified at a spliceosomal PPI interface	93
6.2 First optimization step of Prp3-Prp8 ^{RNaseH} PPI inhibitor	100
6.3 Crystal packing matters in crystallographic fragment screening.....	104
References	109
Publications	128

List of publications

- I. **Barthel, T.**; Huschmann, F. U.; Wallacher, D.; Feiler, C. G.; Klebe, G.; Weiss, M. S.; Wollenhaupt, J., Facilitated Crystal Handling Using a Simple Device for Evaporation Reduction in Microtiter Plates. *J. Appl. Crystallogr.* **2021**, *54* (1), 376–382. <https://doi.org/10.1107/s1600576720016477>

Contribution:

I performed the experiments and analyzed the results. The assembly of the different iterations of the device's prototypes after 3D-printing was done by me. Additionally, I made most of the visualization and wrote the original draft.

- II. Lima, G. M. A.; Jagudin, E.; Talibov, V. O.; Benz, L. S.; Marullo, C.; **Barthel, T.**; Wollenhaupt, J.; Weiss, M. S.; Mueller, U., FragMAXapp: Crystallographic Fragment-Screening Data-Analysis and Project-Management System. *Acta Crystallogr. Sect. D Struct. Biol.* **2021**, *77* (6), 799–808. <https://doi.org/10.1107/s2059798321003818>.

Contribution:

During development of the FragMAXapp I analyzed my CFS data with the first version of it. This beta-testing included for me to report back any issues or missing features. I also reviewed the manuscript.

- III. Wollenhaupt, J.; Metz, A.; **Barthel, T.**; Lima, G. M. A.; Heine, A.; Mueller, U.; Klebe, G.; Weiss, M. S., F2X-Universal and F2X-Entry: Structurally Diverse Compound Libraries for Crystallographic Fragment Screening. *Structure* **2020**, *28* (6), 694-706.e5. <https://doi.org/10.1016/j.str.2020.04.019>.

Contribution:

I took part in the preparation of the ready-to-use plates of the F2X-Entry and F2X-Universal library via acoustic dispensing. The Aar2 and Prp8^{RNAseH} protein production and the AR complex assembly was performed by me. I also performed the crystallization, soaking, and harvesting for the EP and AR-campaigns with DMSO and the EP-campaign w/o DMSO. I performed 1/2 of the data collections at MAX IV and BESSY II. The data analysis was performed for the most part by me. I prepared many figures for the manuscript and reviewed it.

- IV.** Metz, A.; Wollenhaupt, J.; Glöckner, S.; Messini, N.; Huber, S.; **Barthel, T.**; Merabet, A.; Gerber, H. D.; Heine, A.; Klebe, G.; Weiss, M. S., Frag4Lead: Growing Crystallographic Fragment Hits by Catalog Using Fragment-Guided Template Docking. *Acta Crystallogr. Sect. D Struct. Biol.* **2021**, *77* (9), 1168–1182. <https://doi.org/10.1107/S2059798321008196>.

Contribution:

I refined half of the structures deposited for this work and contributed to writing the manuscript.

- V.** Wollenhaupt, J.; **Barthel, T.**; Lima, G. M. A.; Metz, A.; Wallacher, D.; Jagudin, E.; Huschmann, F. U.; Hauß, T.; Feiler, C. G.; Gerlach, M.; Hellmig, M.; Förster, R.; Steffien, M.; Heine, A.; Klebe, G.; Mueller, U.; Weiss, M. S., Workflow and Tools for Crystallographic Fragment Screening at the Helmholtz-Zentrum Berlin. *J. Vis. Exp.* **2021**, 2021 (169), 1–19. <https://doi.org/10.3791/62208>.

Contribution:

I helped writing the filming script for the video. I prepared the solutions and crystals and performed the presentation of the workflow in the instructional video. Some figures were produced by me and I helped to write the manuscript.

- VI.** **Barthel, T.**; Wollenhaupt, J.; Lima, G. M. A.; Wahl, M. C.; Weiss, M. S., Large-Scale Crystallographic Fragment Screening Expedites Compound Optimization and Identifies Putative Protein-Protein Interaction Sites. *J. Med. Chem.* **2022**, *2022*, 14630–14641. <https://doi.org/10.1021/acs.jmedchem.2c01165>

Contribution:

I performed the protein production, crystallization, soaking and harvesting for the crystallographic screening. For the remote data collections at MAX IV, most of them were performed by me. I analyzed the data and modelled the fragments into the respective electron density. The manuscript was mainly written by me and I produced the figures.

Zusammenfassung

Eine der großen gesellschaftlichen Herausforderungen unserer Zeit ist die Bekämpfung von Krankheiten, die die Lebenserwartung verkürzen und zu hohen wirtschaftlichen Verlusten führen. Sowohl das Verständnis als auch die Behandlung dieser Krankheiten erfordern Forschungsaktivitäten auf allen Ebenen. Ein Aspekt davon ist die Entdeckung und Entwicklung von Werkzeugmolekülen und Arzneimitteln. Werkzeugmoleküle unterstützen die Erforschung von Krankheiten und die Entwicklung von Medikamenten. Seit etwa 20 Jahren versucht man die Entdeckung neuer chemischer Verbindungen durch das Screening kleiner organischer Moleküle mittels Hochdurchsatzverfahren. In den letzten Jahren hat sich die Röntgenkristallographie als das vielversprechendste Verfahren zur Durchführung eines solchen Screenings herausgestellt. Kristallographisches Fragment-Screening (CFS) generiert Bindungsinformationen sowie 3D-Strukturinformationen des Zielproteins im Komplex mit dem gebundenen Fragment. Dieses Promotionsprojekt fokussiert sich hauptsächlich auf die Optimierung des kristallographischen Fragment-Screening-Workflows. Untersucht wurden die Voraussetzungen für erfolgreichere Screening-Kampagnen in Bezug auf das untersuchte Kristallsystem, die Fragmentbibliotheken, den Umgang mit den kristallinen Proben sowie den Umgang mit den mit einer Screening-Kampagne verbundenen Daten. Der verbesserte CFS-Workflow wurde als detailliertes Protokoll und begleitendes Video präsentiert, um auch zukünftige CFS-Nutzer auf optimierte und zugängliche Weise zu schulen. Zusammen machen diese Verbesserungen CFS-Kampagnen zu einer Methode mit höherem Durchsatz, die das Screening größerer Fragmentbibliotheken ermöglicht und eine höhere Anzahl von Kampagnen pro Jahr gewähren. Die Protein-Targets während des gesamten Projekts waren zwei Enzyme und ein spleißosomaler Protein-Protein-Komplex. Die Enzyme umfassten die Aspartat-Protease Endothiapepsin und die SARS-Cov-2-Hauptprotease. Der Protein-Protein-Komplex war die RNaseH-ähnliche Domäne von Prp8, einem wichtigen Strukturprotein im Spleißosom, zusammen mit seinem nuklearen Shuttle-Faktor Aar2. Durch die Durchführung der CFS-Kampagnen gegen krankheitsrelevante Targets könnten die resultierenden Fragment-Hits direkt zur Entwicklung von Werkzeugmolekülen oder Medikamenten verwendet werden. Die ersten Schritte der Optimierung von Fragment-Hits in Binder mit höherer Affinität wurden ebenfalls auf Verbesserungen hin untersucht. Zusammenfassend wurden unzählige neue Ansatzpunkte für die Entwicklung von Wirkstoffen und Arzneimitteln identifiziert.

Abstract

One of the great societal challenges of today is the fight against diseases which reduce life expectancy and lead to high economic losses. Both the understanding and the addressing of these diseases need research activities at all levels. One aspect of this is the discovery and development of tool compounds and drugs. Tool compounds support disease research and the development of drugs. For about 20 years, the discovery of new compounds has been attempted by screening small organic molecules by high-throughput methods. More recently, X-ray crystallography has emerged as the most promising method to conduct such screening. Crystallographic fragment-screening (CFS) generates binding information as well as 3D-structural information of the target protein in complex with the bound fragment. This doctoral research project is focused primarily on the optimization of the crystallographic fragment screening workflow. Investigated were the requirements for more successful screening campaigns with respect to the crystal system studied, the fragment libraries, the handling of the crystalline samples, as well as the handling of the data associated with a screening campaign. The improved CFS workflow was presented as a detailed protocol and as an accompanying video to train future CFS users in a streamlined and accessible way. Together, these improvements make CFS campaigns a more high-throughput method, offering the ability to screen larger fragment libraries and allowing higher numbers of campaigns performed per year. The protein targets throughout the project were two enzymes and a spliceosomal protein-protein complex. The enzymes comprised the aspartic protease Endothiapepsin and the SARS-Cov-2 main protease. The protein-protein complex was the RNaseH-like domain of Prp8, a vital structural protein in the spliceosome, together with its nuclear shuttling factor Aar2. By performing the CFS campaigns against disease-relevant targets, the resulting fragment hits could be used directly to develop tool compounds or drugs. The first steps of optimization of fragment hits into higher affinity binders were also investigated for improvements. In summary, a plethora of novel starting points for tool compound and drug development was identified.

1. Introduction

The cell is crowded with proteins, nucleic acids and small molecules.¹ These different molecules interact in an organized way to regulate cellular processes. Therefore, to understand the inner workings of a cell, scientists must find ways to examine molecular interactions. This is especially important for society, as such detailed understanding helps us to pinpoint causes of diseases and the possible ways to intervene. Investigations of molecular interactions of a cell often require so-called tool compounds, that manipulate the process of interest and support controlled measurements.² In cases where the mechanism is well studied, drugs can be developed to tackle faulty cellular processes. One method used in drug or tool compound discovery is crystallographic fragment screening (CFS). It has the advantage of providing 3D-structures of the protein-ligand complex directly after screening. It has developed throughout the last years into a popular approach.³ Hence, improvements of the used methodology are sorely needed. CFS campaigns can be performed at the Helmholtz-Zentrum Berlin (HZB), which includes a synchrotron facility (BESSY II).⁴ The macromolecular crystallography (MX) group supports users at HZB in their CFS endeavors. To improve the CFS workflow at HZB, I performed several CFS campaigns and incorporated the resulting knowledge into the CFS workflow. Additionally, screening against the chosen protein targets results in viable starting points for drug and tool compound development. Thus, the experiments offer information on two different viewpoints. In the following work, the methodology of drug and tool compound discovery is explained in more detail leading to the several steps for possible improvements. Additionally, the chosen protein targets are introduced along with their respective relevance.

1.1 Drug and Tool Compound Discovery

For thousands of years humans have utilized drugs to address diseases.^{5,6} It had been common to prepare extracts from plants like teas or ointments. These mixtures included several different compounds without, in most cases, having knowledge of the biologically active ones. Nowadays, the route to reach a Food and Drug Administration-approved (FDA-approved) drug includes detailed scientific experiments with a clear understanding of the interaction between drug and drug target.⁷ Drugs can be different kinds of chemicals, for example small synthetic molecules, synthetic peptides, or antibodies.⁸⁻¹⁰ Among these, small molecules have been the most popular in the pharmaceutical industry.¹¹ The mode of action of a drug is to interact with a disease-relevant protein or molecule and modulate its activity to tackle a specific disease. The most popular targets so far are enzymes, receptors, or ion channels.¹² However, the modulation of protein function has additional applications. In the academic field, beyond the primary purpose of drug discovery it can

be helpful to develop or find molecules that affect cellular processes in order to understand the system in more detail and often this information is used to understand diseases better.² Such molecules can be termed chemical probes or tool compounds. They can be used as a tool to investigate cellular mechanisms.¹³ Both, discovering drugs and discovering tool compounds, can be split into similar steps (**Figure 1**). Often a tool compound can be developed to support drug discovery efforts. Thus, a tool compound is developed and applied during the screening or activity assay of a drug discovery.

Tool Compound Development

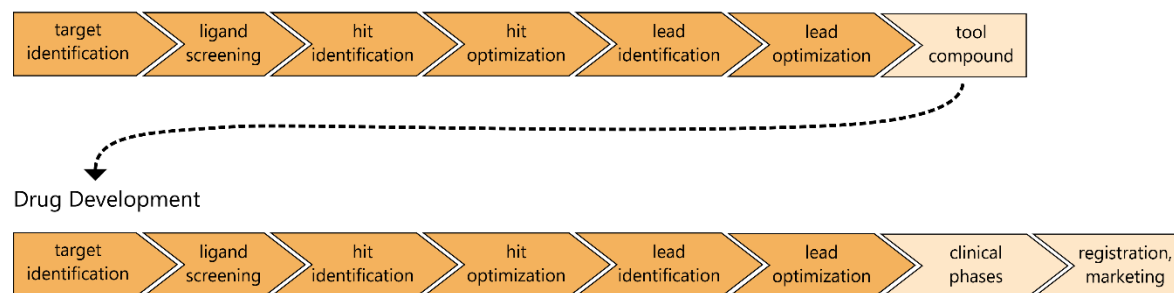


Figure 1: The tool compound development and drug development processes are shown. The individual steps are presented, with similar steps between both processes colored in orange and specific steps for each process are colored in light brown. As a tool compound can be used during drug development, both processes can appear in succession.

First, a target must be identified through identifying key actors in disease-relevant processes using basic research. Second, an assay is established, either biochemical or biophysical, and ligand screening is performed. Third, the initial binders are identified, which are called hits. These are starting points for further developments. Fourth, the hits are optimized into higher affinity binders. Fifth, lead structures are identified from the optimized hits, which are molecules that have been chosen to improve activity against the target and chemical properties like solubility. Sixth, the lead structures are then optimized into molecules that have an effect at low concentrations. After these mentioned steps the optimized final molecule provides a well-measured activity against the target, and the chemical properties allow it to be used in assays. In this way, a tool compound is established. A tool compound does not have to be administered to humans; therefore, additional clinical tests are unnecessary. However, in the case of drug discovery, after the lead optimization step, the clinical phases are entered. The possible drug must be tested for safety and health benefits and be registered before entering the market.

The first steps of ligand screening and hit identification have a great impact on the following optimization process. Additionally, the characteristics of the screening library and the cut-off criteria chosen to distinguish between a binder and non-binder will limit the available

starting points. Thus, it is vital to choose the best strategy fitting to the target and available infrastructure from the beginning.

1.1.1 Screening and Libraries

There are two main branches of screening approaches: high-throughput screening (HTS)¹⁴ and fragment screening (FS)^{15,16} (**Table 1**). In HTS, drug-like compounds are screened. This refers to compounds that are usually following the rule-of-5, meaning they are around 500 Da (~30 heavy atoms), have up to five hydrogen bond donors, up to ten hydrogen bond acceptors and a clogP up to five.¹⁷ The clogP represents the solubility of a compound as a partition coefficient for octanol/water.¹⁸ This rule-of-5 intends to guide scientists to an orally available molecule.¹⁹ However, this rule is simply a guidance. Several drugs do not strictly fall under the rule-of-5, and thus it should not be considered too stringently.²⁰ The benefit of screening such compounds is the commonly high affinity of found hits (nanomolar to low micromolar), due to the presence of several pharmacophores that can engage in interactions. However, they usually result only in hit rates between 0.01 – 0.2%.²¹ The low hit rates derive from the large size of these compounds, which might include several functional groups that can either interact favorably or unfavorably with the target. Even one unfavorable interaction might be enough to abolish binding. Thus, the screening libraries contain usually between tens of thousands to millions of compounds. Screening such large numbers of compounds is only possible with biochemical assays performed in minimal volume, often using plate-based assays that utilize fluorescent signals. These assays have the advantage of a fast signal read-out and can often be combined with the increasing or decreasing activity of the target, which presents the biological activity of the compound. Due to the size of the compounds, the typical steps of optimization include exchanging functional groups and scaffolds to improve binding.

In FS, smaller compounds called fragments are utilized. They loosely follow the rule-of-3, with the molecular weight up to 300 Da, up to three hydrogen bond donors, up to three hydrogen acceptors and a clogP up to three.²² The expected binding affinities are between high micromolar to millimolar, but the hit rates can span between 1 to 40%. This is due to the small size of fragments, and consequently lower binding specificity. FS libraries often comprise between hundreds to thousands of fragments. Therefore, medium-throughput assays can be applied. Sensitive biophysical methods are needed to detect the weak binders, like nuclear magnetic resonance (NMR), X-ray crystallography or surface plasmon resonance (SPR). The small size of fragments allows for new optimization strategies called growing, merging and linking.²³

Table 1: Comparison of the HTS and FS approach. Different aspects of HTS and FS are compared, that define the individual approaches.

	HTS	FS
Library size	10.000 - 1.000.000	100 - 5.000
Screening compound MW	500 Da	300 Da
Expected binding affinity	nM - μ M	μ M - mM
Expected hit rates	0.01 - 0.2 %	1 - 40 %
Assays	Biochemical	Biophysical
Optimization strategies	Exchange of functional groups and scaffolds	Growing, linking, merging

HTS is the more traditional screening approach used in pharmaceutical industry, however a shift towards FS is becoming tangible. The growing popularity and success of FS is reflected by the rising number of FDA-approved fragment-based drugs (**Table 2**). Since 2011, 6 drugs made it to the market for various cancer types³ and many more compounds are in clinical trials.²⁴ Additionally, one of the FDA-approved drugs targets the G-protein KRAS^{G12C}, which had been thought to be undruggable until last year.^{25,26} The general success of the FS approach can be attributed to the following reasons. First, the library size can be limited to a few thousand compounds while maintaining a similar or even higher diversity in chemical and structural features compared to HTS. This is due to the chemical space spanned by smaller and larger molecules, respectively. The smaller the molecules, the smaller their chemical space. In the case of larger, drug-like compounds it has been estimated that 10^{63} molecules make up their chemical space. In the case of molecules that are 17 heavy atoms in size, it is about 10^9 molecules.²⁷ Therefore, to investigate the same percentage of the chemical spaces, in case of fragments, a smaller number of molecules must be screened, compared to the drug-like compounds. Second, different optimization strategies can be applied. In case of HTS, the size of the compounds cannot be increased further during optimization. Drug-like molecules are often around 500 Da in molecular weight, as shown by the rule-of-5. Growing the molecule further often results in unfavorable chemical properties for drugs. Thus, it is only feasible to exchange groups which limits the possibilities for new structural features. In FS the compounds are smaller, which allows the addition of groups and provides higher flexibility in synthesis of follow-up molecules. Additionally, fragments are better starting points as they bind more efficiently. That means that fragments present a limited availability for interactions and thus bind with great energetic gain in relation to their size, which is also referred to as ligand efficiency.²⁸

Furthermore, due to the higher hit rates achieved by FS, the possibilities to merge compounds bound in a similar position or link them when they are close to each other broadens the available optimization paths.^{29,30}

Table 2: List of current fragment-based drugs that are either FDA-approved or in clinical trials.³¹ The name or code is given for each drug with the respective company developing it and the target. The table was adapted from the Practical Fragments Blog by Dan Erlanson, published on 21st November 2023.

	Drug	Company	Target
Approved	Asciminib	Novartis	BCR-ABL1
	Erdafitinib	Astex/J&J	FGFR1-4
	Pexidartinib	Plexikon	CSF1R, KIT
	Sotorasib	Amgen	KRAS ^{G12C}
	Vemurafenib	Plexikon	B-RAF ^{V600E}
	Venetoclax	AbbVie/Genentech	Selective BCL-2
Phase 3	Capivasertib	AstraZeneca/Astex/CR-UK	AKT
	Lanabecestat	Astex/AstraZeneca/Lilly	BACE1
	Navitoclax (ABT-263)	Abbott	BCL-2/BCL _{XL}
	Pelabresib (CP-0610)	Constellation	BET
	Verubecestat	Merck	BACE1
Phase 2	ASTX029	Astex	ERK1,2
	ASTX660	Astex	XIAP/cIAP1
	AT7519	Astex	CDK1,2,4,5,9
	AT9283	Astex	Aurora, JAK2
	AUY-922	Vernalis/Novartis	HSP90
	AZD5991	AstraZeneca	MCL1
	DG-051	deCODE	LTA4H
	eFT508	eFFECTOR	MNK1/2
	Indeglitazar	Plexikon	pan-PPAR agonist
	LY2886721	Lilly	BACE1
	LY3202626	Lilly	BACE1
	LY3372689	Lilly	OGA
	LY517717	Lilly/Protherics	FXa
	LYS006	Novartis	LTA4H
	MAK683	Novartis	PRC2 EED
	Onalespib	Astex	HSP90
	PF-06650833	Pfizer	IRAK4
	PF-06835919	Pfizer	KHK
	PLX51107	Plexikon	BET
	S64315	Vernalis/Servier/Novartis	MCL1
VK-2019	Cullinan Oncology/Wistar	EBNA1	

	Drug	Company	Target
Phase 1	AG-270	Agios/Servier	MAT2A
	ABBV-744	Abbott	BD2-selective BET
	ABT-518	Abbott	MMP-2 & 9
	ABT-737	Abbott	BCL-2/BCL _{xL}
	AT13148	Astex	AKT, p70S6K, ROCK
	AZD3839	AstraZeneca	BACE1
	AZD5099	AstraZeneca	Bacterial topoisomerase II
	BI 1823911	Boehringer Ingelheim	KRAS ^{G12C}
	BI 691751	Boehringer Ingelheim	LTA4H
	CFTX-1554	Confo Therapeutics	AT ₂ receptor
	ETC-206	D3	MNK1/2
	GDC-0994	Genentech/Array	ERK2
	HTL0014242	Sosei Heptares	mGlu5 NAM
	HTL0018318	Sosei Heptares	M1-receptor partial agonist
	HTL9936	Sosei Heptares	M1-receptor partial agonist
	IC-776	Lilly/ICOS	LFA-1
	LP-261	Locus	Tubulin
	LY2811376	Lilly	BACE1
	Mivebresib	AbbVie	BRD2-4
	MRTX1719	Mirati	PRMT5 MTA
	Navoximod	New Link/Genentech	IDO1
	PLX5568	Plexikon	RAF
	SGX-393	SGX	BCR-ABL
	SGX-523	SGX	MET
	SNS-314	Sunesis	Aurora
	TAK-020	Takeda	BTK

1.1.1.1 Fragment Library Design

The mentioned benefits of FS depend on a good fragment library. A good fragment library should enable scientists to find a high number of novel starting points for their target that can potentially be optimized into lead compounds. This goal can be achieved by maximizing the diversity of the fragments. There are different characteristics that can define diversity. It mainly refers to the chemical and structural features of a molecule. However, in the last years, shape diversity has gained popularity. There are different ways to compare fragments based on their similarity. Such comparisons can be used to pick the most dissimilar fragments and reach high diversity. One of the first steps can be a 2D

comparison via the Tanimoto coefficient.³² For that, a 2D fingerprint is generated for each molecule. These fingerprints are a binary string where each value represents a specific feature of a molecule. If that feature is present in the fragment, it is given a 1, if it is not present it is given a 0. Thus, each fragment is given a binary string that can be compared. An additional layer of comparison can be added via newer tools like the ROCS 3D-shape and pharmacophore comparison.³³ The 3D-shape comparison calculates a Gaussian-based volume for the heavy atoms of a molecule. This volume represents the shape of the molecule. The volumes of different molecules are overlaid, and the similarity is calculated. The pharmacophore comparison is based on a list of known pharmacophores and each pharmacophore is assigned different colors. The combination of colors per molecule is compared to other molecules and again a similarity is calculated. Both comparisons can be combined. Fragment libraries can then be compared based on their diversity via different metrics like principal moment of inertia (PMI) plots³⁴ that visualize how the fragments spread across a triangle with each corner representing rod-like, disc-like and sphere-like compounds. Carbery *et al.* argue that chemical or structural diversity differs from functional diversity, which should be the focus for diverse libraries.³⁵ They propose that a library built to be functionally diverse could decrease the number of fragments screened while gaining a lot of information about the proteins binding site.

Other considerations should be included in fragment library design, as for example chemical stability, solubility or the synthetic tractability. The typical bottleneck in academic FS campaigns is the follow-up step after hit identification. This refers to the first optimization step to find binders with greater binding affinity. Depending on the fragment chosen for optimization, several superstructures of the fragment might be commercially available and thus straight-forward to follow up, or medicinal chemists might be needed to assess the probability of synthesizing larger compounds. Involving medicinal chemists can result in higher costs during optimization and longer times to establish novel syntheses. This is especially a problem for academic groups, which do not have medicinal chemists. This problem can be circumvented partly if the library has been built to consider synthetic tractability. For example, only fragments are chosen that include reactive handles.³⁶ Thus, if one of the promising exit vectors of the fragment hit has a reactive handle, additional functional groups can be added with relatively low synthetic effort. However, this approach might limit the achievable diversity of the library. Moreover, the reactive handle itself will also often engage in interactions, thus be buried in the protein binding pocket and not offer a usable exit vector for further follow-up.

Considering the different aspects of library design, each fragment library will present a different set of fragments, based on their criteria. They might be specialized based on their

3D-shape or focused on a specific protein target. Thus, it is vital to understand the design of the library and base the decision which library to apply on the planned screening campaign.

1.1.1.2 Assessing a Protein's Druggability

Fragments have another application besides identifying promising starting points. Specifically designed fragment libraries like FragLites³⁷ or MiniFragments³⁸ comprise very small fragments below 150 Da and are used to probe a protein's surface for its druggability. The druggability of a target protein is meant to predict the success of drug discovery for the target.³⁹ Proteins that have a low druggability, meaning few surface areas that chemicals bind to, are not as promising as proteins with high druggability. The experimental probing approach requires X-ray crystallography to visualize the binding position of each molecule.⁴⁰ Thus, FS has two main objectives. On the one side it allows to find starting points for drug or tool compound development and on the other side it provides the possibility to probe a proteins surface for druggable sites.

1.1.1.3 Choosing the Screening Method

The screening assay applied during FS will provide a defined set of information for each fragment hit. This will also have an influence on the choice of the best fragment hit to continue with optimization. There are different ways to perform the screening step. It is often suggested to perform orthogonal assays to gather as much information as possible about the fragment hits and validate their binding.²³ This can be a combination of any of the typical FS methods for example NMR, X-ray crystallography, SPR, and thermal shift assay (TSA). A recent paper stated to complement this approach by combining a structural assay, a binding assay, and a biological assay.⁴¹ The goal of this approach is simple; to understand the binding and activity of each fragment hit as best as possible. However, there are two drawbacks. First, establishing two or three separate assays for the same target is very time-consuming and challenging especially in the academic setting. Second, another paper published a few years back had shown a very small overlap between orthogonal FSs against the same target.⁴² Six biophysical methods had been applied to screen fragments against Endothiapepsin (**Section 1.2.1**) and the analysis of each screening showed different fragment hits. The number of detected hits and the identity differed greatly, with X-ray crystallography finding the largest number of unique hits. Not one fragment was detected by all six biophysical methods. It is questionable which assay should be trusted most. One could limit oneself to the fragment hits that are detected by at least two assays, however, this might lead to a loss of promising starting points. The authors conclude that X-ray crystallography should be the method of choice as a first screening assay. It finds the most hits, probably due to the possibility to apply very high

concentrations of fragments and thus detecting very low affinity binders. (Further explanation as to why high concentrations are needed for detection will follow in **Section 1.1.2.3.**) The listed advantages fit well with the benefit of a crystallographic fragment screening (CFS) of providing 3D structural information about the bound hit from the start.

1.1.2 Crystallographic Fragment Screening

X-ray crystallography is an integral part in the drug discovery process. It is typically applied to perform structure-based drug design, as it is the only method to elucidate the binding site and binding pose of the protein-ligand complex. Furthermore, it is able to detect very low affinity binders, such as fragments. The amount of information given by CFS makes it more and more popular in academia and industry. Additionally, it has been developed into a medium-throughput method in the last decades.⁴³ The general workflow is seen in **Figure 2**, with 5 steps:

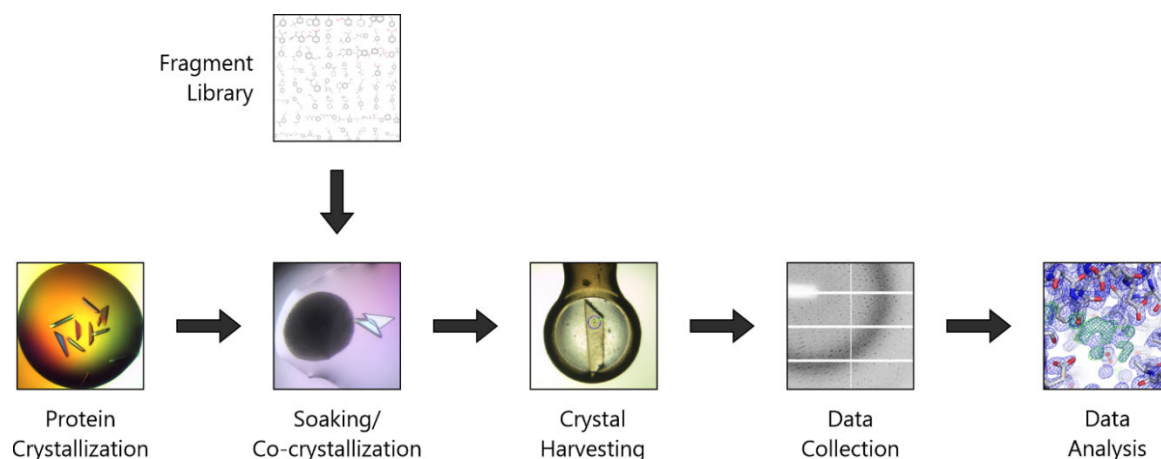


Figure 2: A schematic overview of the CFS workflow. The individual steps are shown with respective exemplary pictures.

1.1.2.1 Protein Crystallization

First, the protein needs to be crystallized. This is eased by robot-assisted setup of crystallization screens. In case crystals can be grown that diffract ideally better than 2.0 Å resolution, the crystallization can be optimized for CFS. Most importantly the crystals need to diffract reliably to high resolution. This means over ten crystals the diffraction should stay in a range of +/- 0.2 Å resolution. Otherwise, there is no guarantee that the crystals used later for soaking diffract well enough for subsequent analysis. Additionally, the space group and solvent content of the crystals should be considered. A greater solvent content translates into larger solvent channels. This eases diffusion for the fragments into the crystal and allows them to reach each protein molecule. Furthermore, the space group is defined by the packing of the crystal. Especially if a certain binding site is targeted, this part of the protein must not be involved in crystal contacts. The target site should be open

towards the solvent to be accessible for fragments. Therefore, if the packing is very tight or obstructs the binding site, effort should be taken to crystallize the protein in a different crystal form.

1.1.2.2 The Fragment Library Choice

Second, a fragment library is chosen. The library can be made directly for CFS or for a specific target or a general screening library. The library determines the chemical space that can be explored and the possible resulting starting points. The presentation of the library has an influence on the possibilities of the next step of soaking or co-crystallization. Thus, for example the method of soaking is dependent on the chosen fragment library.

1.1.2.3 Crystal Soaking

Third, the protein crystals are soaked with the fragments or co-crystallized in their presence. In co-crystallization, the presence of the fragment can induce a change in the crystallization condition, which can be dramatic enough to abolish crystal growth. The soaking approach might also damage protein crystals due to the addition of organic solvents in the soaking condition. However, the soaking experiment may not substantially deteriorate the diffraction quality. Thus, soaking usually results in a higher number of diffracting crystals and is therefore the preferred approach. There are two approaches currently used for soaking. For one approach the fragments are dissolved in an organic solvent, often DMSO, and via an acoustic dispenser, small droplets of this solution are added to the crystallization drop.⁴⁴ In this way the crystal is not touched, but the crystal must withstand as much DMSO as possible so that a sufficient fragment concentration can be reached in the crystallization drop. The other approach for soaking is to dry the fragments onto typical crystallization plates, add a soaking solution to dissolve the fragment again and transfer a crystal into this drop.⁴⁵ In this way the crystal must be transferred, but it is possible to soak with less or even no organic solvent, while keeping the fragment concentration high. The fragment concentration is key to reaching high occupancy of the fragment in the protein crystal. This is based on the thermodynamic equilibrium equations for protein-ligand complexes (**Figure 3**).

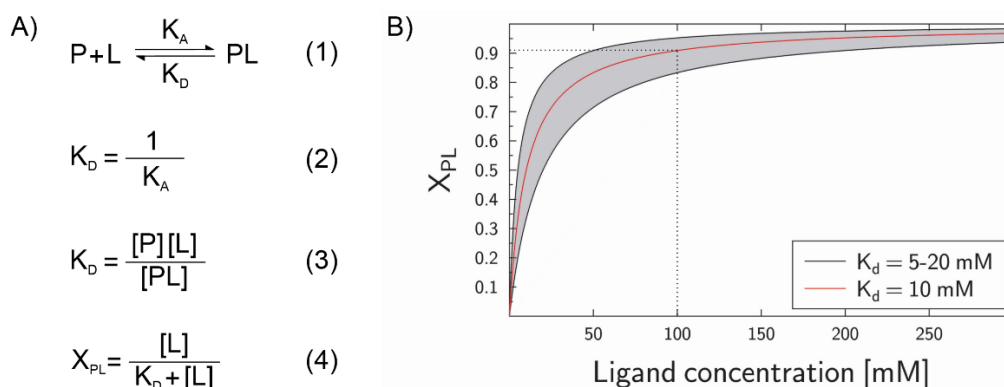


Figure 3: The protein-ligand equilibrium explained as adapted from Manfred S. Weiss (personal communication). A) Important equations for the protein-ligand equilibrium. 1) The relationship between the free protein (P) and ligand (L), the protein-ligand complex (PL) and the association constant K_A and dissociation constant K_D . 2) The relationship between K_D and K_A . 3) The equation to calculate K_D values from protein, ligand, and protein-ligand concentrations. 4) The equation to calculate the fraction of protein-ligand complex compared to the overall protein (X_{PL}) from the ligand concentration and K_D . B) Calculated curves based on equation (4). The dashed line highlights an appropriate ligand concentration to achieve a high fraction of protein-ligand complex.

The equilibrium is described as a reversible reaction between the free protein and free ligand, and the protein-ligand complex (**Equation 1**). This equilibrium is defined through the association constant K_A and the inverse dissociation constant K_D (**Equation 1** and **Equation 2**). In case the K_D is low, the equilibrium is pushed towards a higher concentration of the protein-ligand complex and in case of a high K_D , the equilibrium is pushed towards the free protein and free ligand. Considering fragments, the K_D is usually high which in turn means the concentration of the protein-ligand complex is low. To push the equilibrium towards higher protein-ligand complex concentrations, either the ligand concentration or the protein concentration must be increased. In case of a protein crystal, the protein concentration is fixed, thus, the ligand concentration is increased to achieve this goal. In principle this means that especially for low affinity binders (high K_D) such as fragments, high fragment concentrations are necessary to reach high protein-ligand complex concentrations. The fragment concentration can also be calculated to assess which concentrations are beneficial for fragment screening (**Equation 4** and **Figure 3B**). If we assume a K_D of 10 mM a 10-fold higher fragment concentration (100 mM) would result in about 90% protein-ligand complex formation compared to the overall protein amount. Here only one binding site is considered for simplification. A high concentration of protein-ligand complex translates into a high occupancy of the fragment in the protein crystal and makes it easier or even possible to identify the bound fragment. (**Figure 4A**).⁴² Therefore, high fragment concentrations during soaking are necessary. Furthermore, in comparison soaking times are not as vital for detection of fragments. Schiebel *et al.* compared the influence of soaking time with soaking concentration and found that at concentrations of 90 mM fragment even a soaking time of 2 min can be sufficient (**Figure 4B**). It is noteworthy that soaking times can differ depending on the crystal size

and crystal packing. In general, it is convenient to soak overnight if it is tolerated by the protein crystals. Even if such long soaking times are not feasible, shorter times do not prevent detection of fragments.

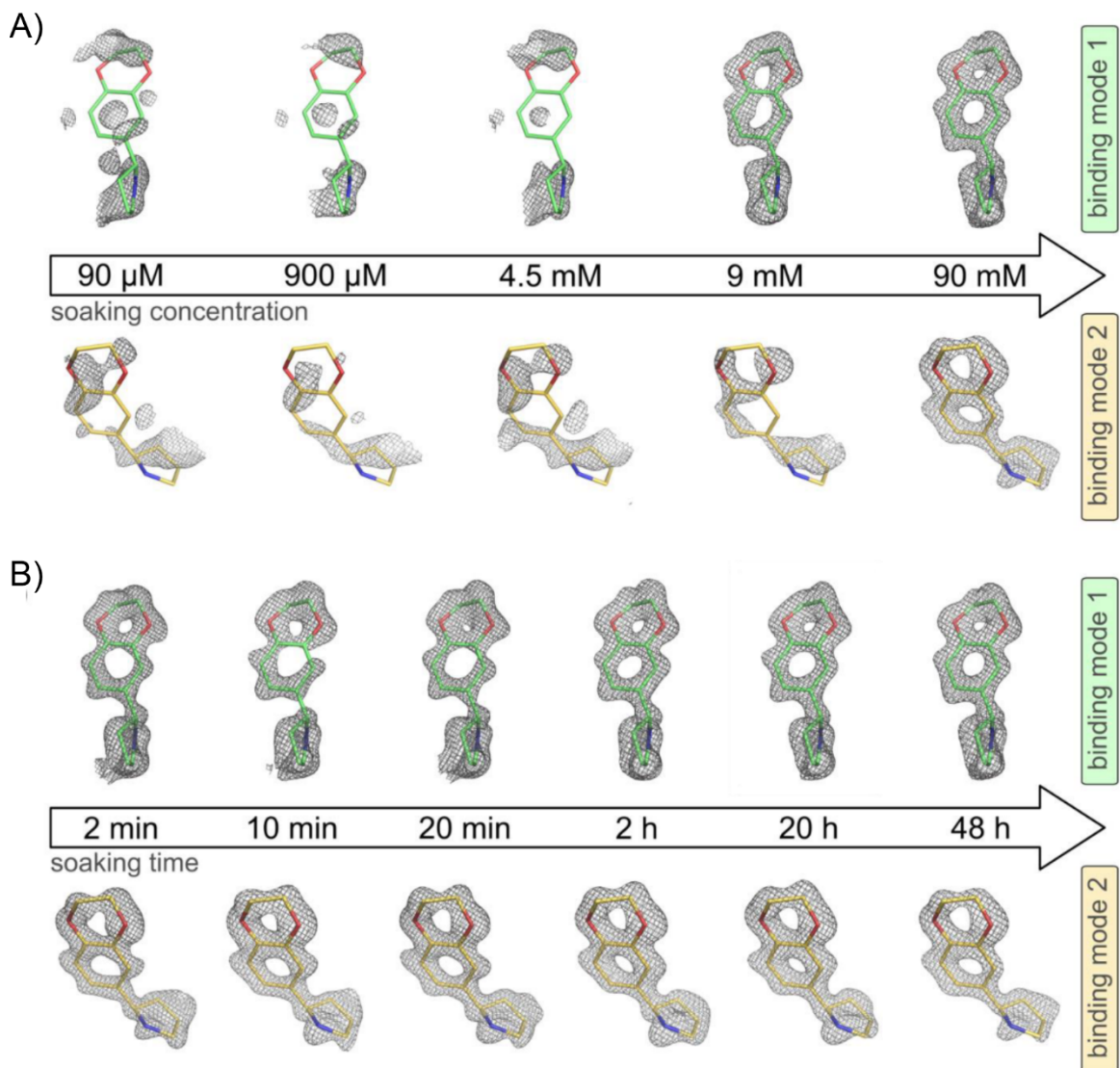


Figure 4: A comparison is shown between soaking time and soaking fragment concentration. The figure was adapted from Schiebel *et al.*⁴² The fragment structure is shown as sticks, colored in green for the first binding site and in yellow for the second binding site. The respective electron density $mF_o - DF_c$ is shown at 3σ . A) The influence of the soaking concentration on detectability of the fragment is shown. Each experiment was performed for 48 h. B) The influence of soaking times on the detectability of the fragment is presented. Each experiment was performed with 90 mM fragment concentration.

1.1.2.4 Crystal Harvesting

Fourth, the protein crystals are harvested (exemplary visual presentation of harvested CFS samples: **Figure 5**). This is usually done manually by cutting off the crystallization foil that protects the crystals, mount the crystals into a sample holder called a loop, and plunge-freeze it in liquid nitrogen. This is repeated for all the samples. This procedure is still one of the most time-consuming and labor-intensive steps. In a medium- or high-throughput approach, the cumbersome foil cutting must be repeated hundreds of times. To simplify

and speed up this step of crystal harvesting, several different devices have been developed.⁴⁶⁻⁴⁹ For example, at ESRF a fully automated harvesting machine, called the CrystalDirect harvester, had been built which allows manual crystal harvesting to be completely avoided.^{48,50} However, this machine is expensive and requires the crystallization to be performed with specific crystallization plates. So far it has only been implemented at ESRF. Another example is the robot-assisted device, the Crystal Shifter,⁴⁹ developed at Diamond Light Source, which negates the foil cutting. It is employed now at several synchrotron sites (MAXIV, DESY, SLS and Melbourne Synchrotron - personal communication). The developed solutions to automate crystal harvesting or assist crystallographers are usually rather expensive and need specific environments. Therefore, there is still no inclusive solution for smaller academic laboratories.

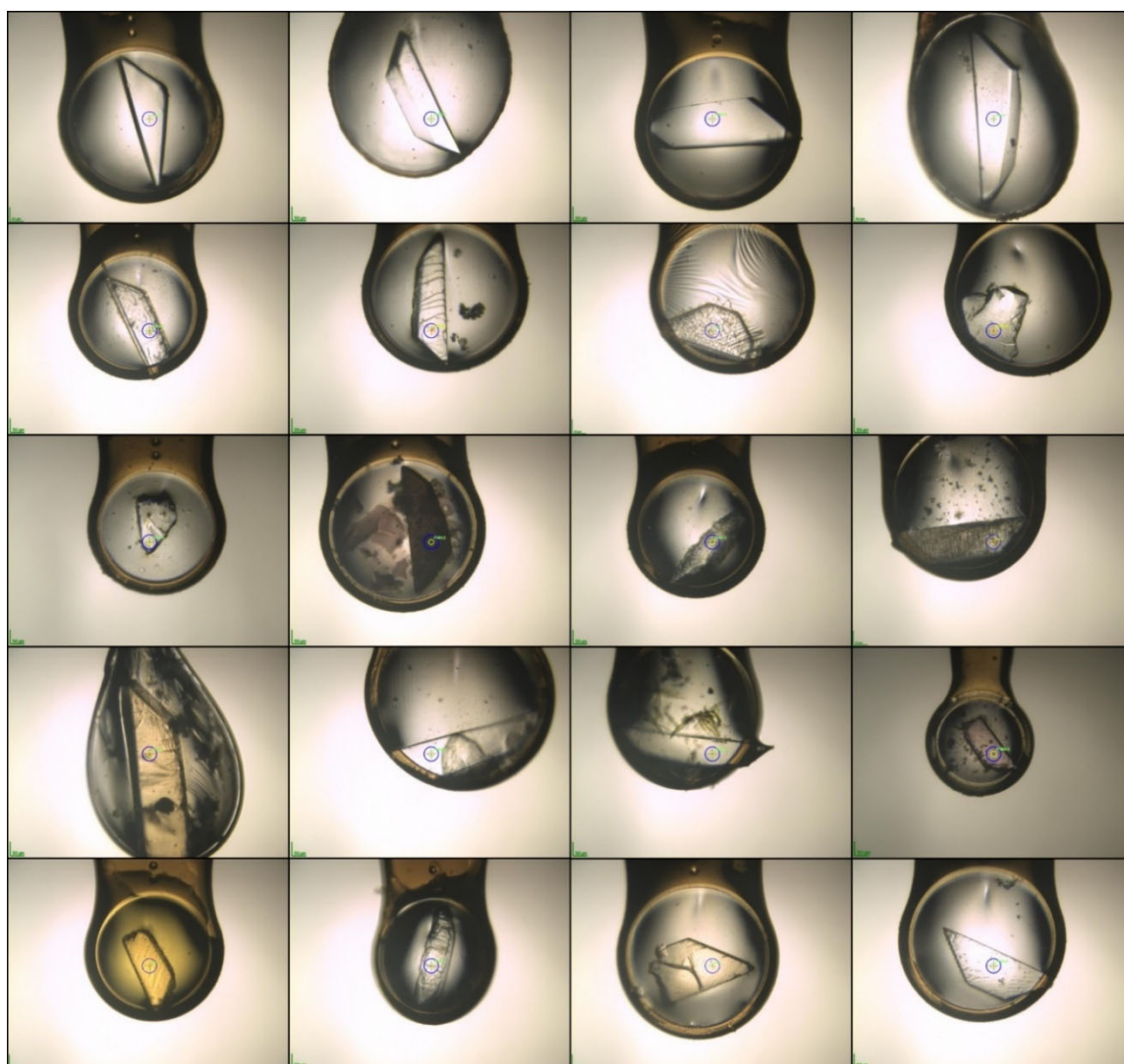


Figure 5: Possible visual appearances of CFS samples. Depending on the fragment contained in the soaking drop, the crystal's appearance might change.

1.1.2.5 Data Collection

Fifth, the data is collected from each crystal. Bright synchrotron sources, fast read-out detectors, robot-assisted mounting and novel software results in a high-throughput set up for data collection.⁵¹⁻⁵⁵ This allows screening of hundreds of crystals per day with minimal intervention by the scientist. Furthermore, two synchrotron facilities (ESRF and Diamond Light Source) are operating fully-automated beamlines.^{56,57} There are even recent developments to increase the number of collected datasets at Diamond Light Source and DESY. The new beamline K04 at Diamond Light Source will be built as an automated screening beamline, which is planned to allow the measurements of 1,000 to 2,000 crystals per day.⁵⁸ At DESY the beamline HiPhax starts its first operation as a high-throughput pharmaceutical X-ray screening endstation which will allow collection of over 1000 datasets per 24 hours.^{59,60} Thus, the throughput in CFS becomes more reliant on other steps of the CFS workflow.

1.1.2.6 Data Analysis

Sixth, the collected data is analyzed. Each dataset must be processed, refined and examined for binders. This process has been automated as best as possible via automatic processing software, automatic refinement software and fragment-screening-specific software. The automatic refinement software especially saves time throughout the campaign, as this is a time-consuming labor-intensive step. Three pipelines have been written for this purpose so far, BUSTER,⁶¹ dimple⁶² and fspipeline.⁶³ Another important fragment-focused software helps scientists to identify fragment hits: the Pan-Dataset Density Analysis (PanDDA).⁶⁴ PanDDA allows identification of even very weakly bound fragments that would not be visible in conventional electron density maps (**Figure 6**).

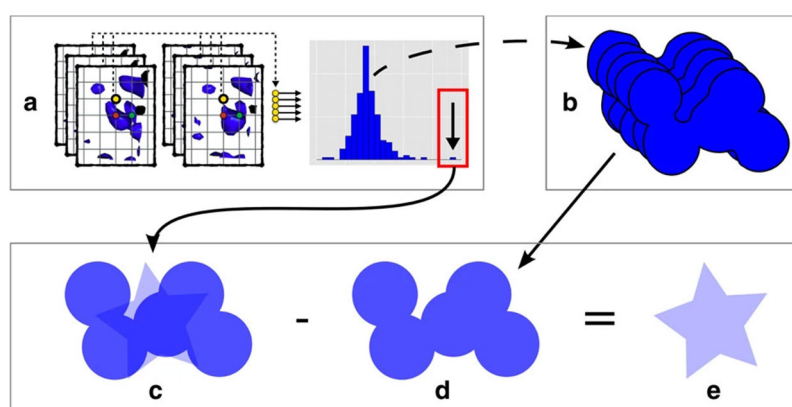


Figure 6: Schematic view of the PanDDA approach adapted from Pearce *et al.*⁶⁴ A) The comparison of each dataset to each other is shown at one specific coordinate in the map. Outliers hint towards a larger change like a bound fragment. B) The datasets from the main distribution for this voxel representing the ground state are averaged. C) The outlier consists of the superposition of the ground state and fragment bound state. D) The average ground state can be subtracted. E) The fragment bound state is extracted.

To explain this in more detail, one must understand the complexity of one single protein crystal. A protein crystal is built up of individual protein molecules. Only a certain percentage of the proteins inside a crystal will bind the fragment, depending on the affinity

of the fragment. In the case of a bound fragment, a protein crystal comprises of a major state (the unbound state) and a minor state (the fragment-bound state). The data collected from a protein crystal averages these states which makes it hard to distinguish the minor and major state by conventional electron density maps. In order to separate the states, PanDDA locally aligns the $2mF_o-DF_c$ electron density maps of each dataset collected during a CFS campaign in real space. These electron density maps are then compared against each other at each voxel and a statistical Z-score is calculated. Based on that, a spatial Z-map is assembled. Whenever a cluster of high Z-scores is identified, the dataset is highlighted as interesting. Datasets without such clusters are then used to calculate an average map. This can be referred to as a background of the fragment-bound state. This background is subtracted from the $2mF_o-DF_c$ electron density maps stepwise. Depending on the ratio of unbound and fragment-bound state, a right amount of the background needs to be subtracted, which is referred to as the Background Density Correction factor (BDC). The final calculation is shown in **Figure 7**. The BDC is determined algorithmically as the value at which the background-corrected map is least correlated with the ground state map. That BDC value is then multiplied with the ground state map. This is then subtracted from the dataset map to derive the event map (**Figure 7A**).

A) [event-map] = [dataset map] - BDC x [ground state map]

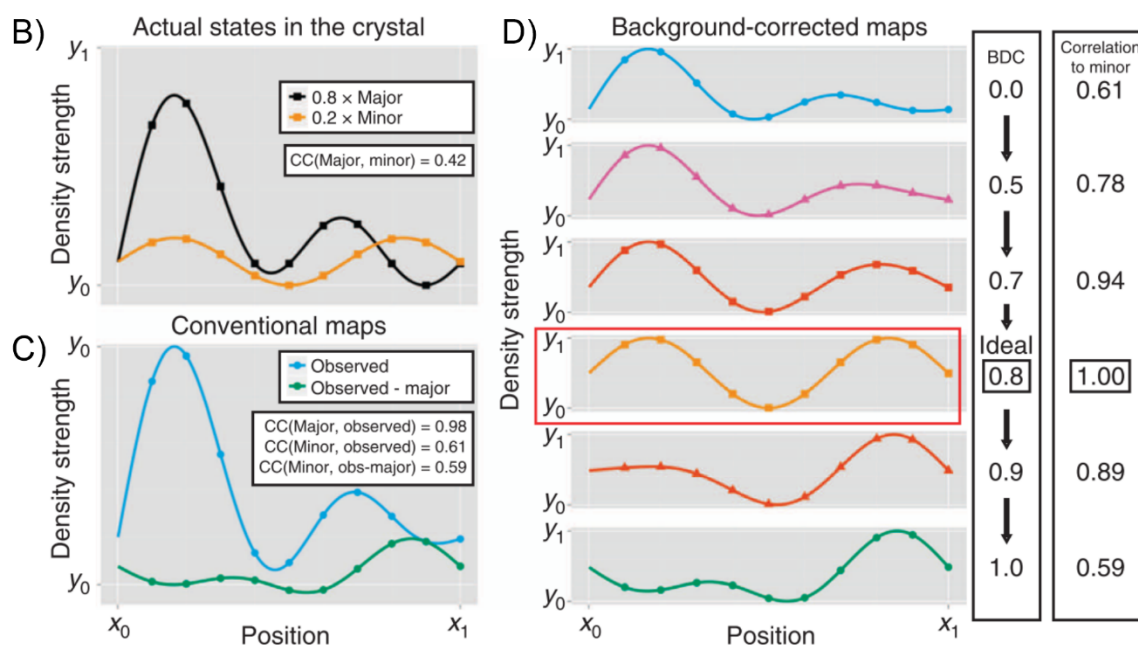
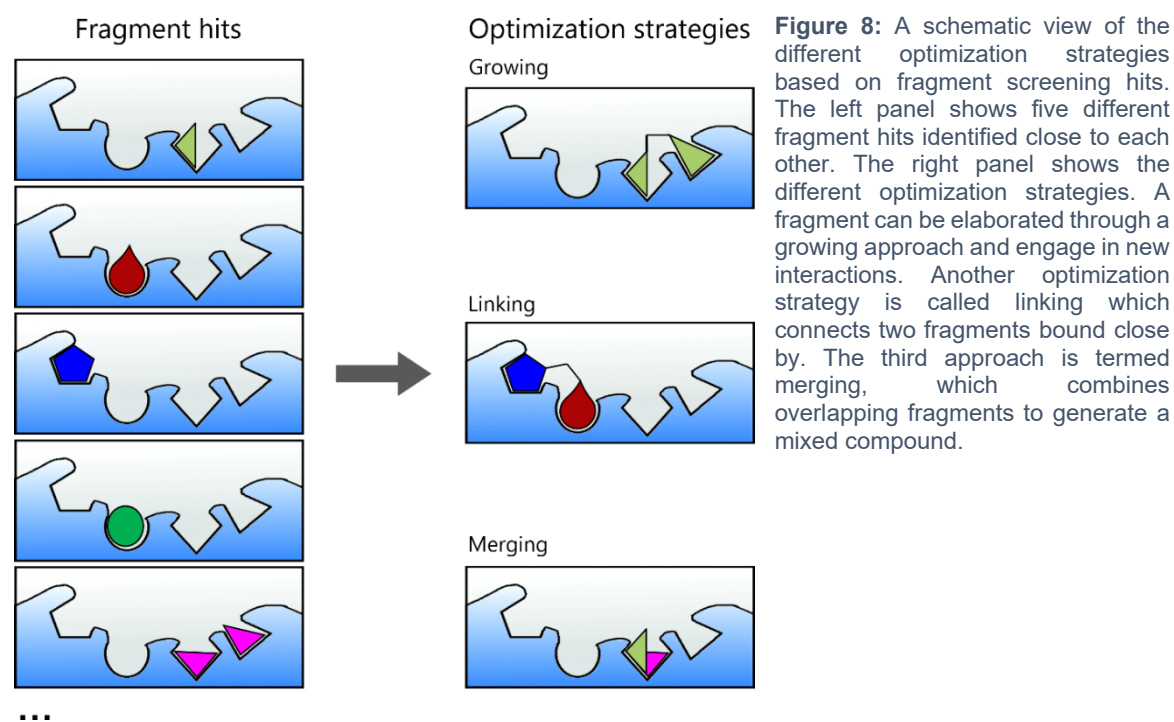


Figure 7: Equation and schematic explanation of the event map calculation and BDC level adapted from Pearce *et al.*⁶⁴ A) The equation to derive the event map is shown. B) An example is presented for a protein crystal consisting of 80% major state (black) and 20% minor state (orange). C) The observed $2mF_o-DF_c$ map (blue) and F_o-F_c map (green) is shown for the given example. Both do not correlate well with the minor state. D) Background-corrected maps are shown for different BDC levels in different colors. The orange curve at BDC 0.8 correlates well to the minor state.

The resulting map after the subtraction is named the “event map”, which represents the fragment-bound state, and enhances the features at possible binding events. In this way PanDDA allows identification of weak binders. It is possible that several binding events occur for one fragment hit, meaning the fragment binds to more than one surface region of the protein. Therefore, the number of binding events can differ from the number of unique fragment hits, and both should be considered in reporting CFS results.



The resulting hit structures are refined manually after this workflow and can be used for structure-based drug/tool compound design. There are three different approaches that hits can be optimized; namely growing, merging, and linking (**Figure 8**).^{23,65} The growing approach is the most popular path so far, as it only requires one fragment hit to continue. The structure is investigated in detail to identify viable exit vectors, i.e. growing directions. Some vectors might be blocked by the protein surface, or show indispensable interaction features, or point away from the protein towards the solvent. A viable exit vector has the potential to add more groups to the fragment, leading to additional favorable interactions. The merging and linking approaches require at least two fragment hits. In the case of merging, the fragments need to have a part of their structure to overlap. This will be used to combine favorable functional groups that are already known to be able to engage in interactions at that position. The linking approach requires two fragments that are located close to each other. Usually, a distance of about 3 Å is favored as the fragments are close but still leave room to add different structural features or linkers.⁶⁶ The two fragments are then connected via a linker. This addition is proposed to have an additive effect on the binding energy.²⁹ This has been shown to be not as simple, mainly depending on the linker,

as it can change the possible geometry and flexibility of the compounds.⁶⁷ These optimization strategies require the rational optimization approach as structural information about the fragments binding pose and position is necessary.

1.1.3 Structure-based drug/tool compound design

The rational structure-based approach to drug discovery takes advantage of 3D-structural information to improve the hit or lead compound. Structural information is an integral part of the drug discovery pipeline.⁶ During optimization, hundreds to millions of possible follow-up molecules could be worth considering for further binding experiments. However, testing and synthesizing is usually restricted to a limited number, possibly a few dozen or hundreds of such compounds. This restriction is necessary because of feasibility and time constraints. To focus the experimental work, computer-aided drug discovery (CADD) approaches can be employed to ease and speed-up the optimization step.⁶⁸⁻⁷⁰ CADD workflows typically include the preparation of the computational input, docking and scoring of candidate molecules, re-scoring of the docking poses, and post-processing of the results like filtering docking poses and calculating visual aids for selection (**Figure 9**).

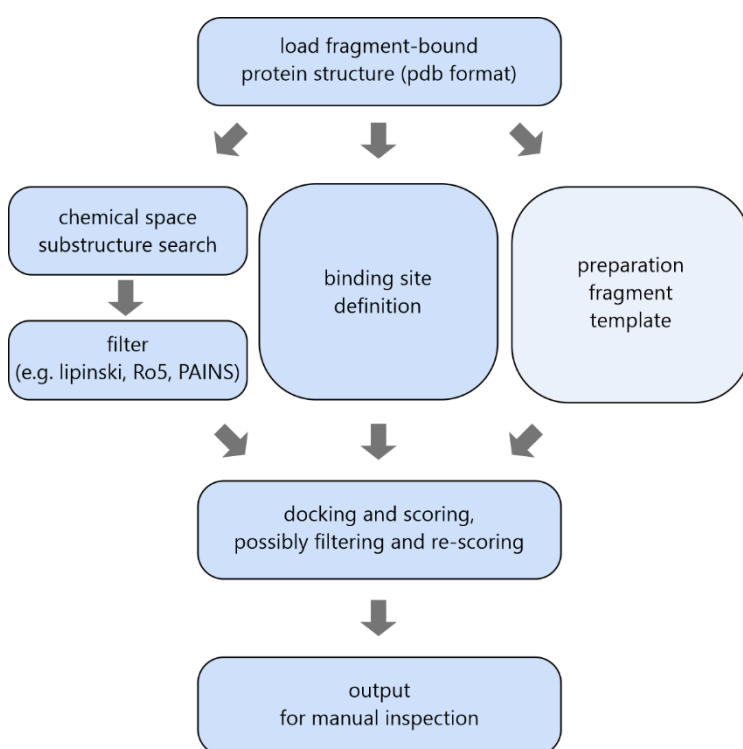


Figure 9: Schematic overview of the CADD workflow presented here. Each step is shown in a blue box, the preparation of the fragment template is highlighted in a lighter blue, as this depends on the docking mode (template-based docking).

1.1.3.1 Preparation of Virtual Docking Input

Before the actual docking step, one needs to define or specify the binding site and the virtual library of molecules to be docked. Additional steps might be needed in case of ambiguous X-ray structures, though such cases are not focused on in this work.⁷¹

1.1.3.1.1 *Definition of Binding Site*

Defining a binding site means to select a number of amino acids of the protein that make up a ligand binding site. This selection of a part of the protein molecule is then considered for energy calculations during docking calculations. The definition of the binding site is necessary to minimize computation time, as each individual atom is considered for calculations.⁷² If the complete protein would be given, more calculations are necessary compared to a reduced input like a binding site. The binding site does not necessarily consist of amino acids alone. Based on crystal structures, water molecules or co-factors might be known and might be relevant for ligand binding.^{73,74} In such cases these additional molecules are included in binding site definition. The binding site definition can be performed via various programs. For example, the commercial software SeeSAR from BioSolveIT incorporated the LeadIT program to assist scientists with the binding site definition.⁷⁵ The software can automatically select relevant amino acids based on a protein-ligand complex structure, where the ligand binds in the target site. In case no such complex structure is available, SeeSAR performs a binding site search and offers suggestions for possible binding sites. The automatically defined binding site can be manually adjusted. Utilizing a software like SeeSAR allows scientist to perform the step of binding site definition in a rather straight-forward way. Parallel to the binding site definition another input can be generated: a virtual library with compounds used in docking.

1.1.3.1.2 *Building a Library of Possible Follow-up Compounds*

The second necessary input for a CADD workflow is a virtual library of to-be-docked compounds. Such compounds can also be referred to as follow-up compounds, as they follow up on an initial starting point. The virtual library can be built by hand, e.g., based on ideas for merging or linking fragments. Alternatively, the initial fragment hit may be used in a substructure search in commercially available chemical spaces to identify larger molecules containing the same pharmacophore.⁷⁶ Another way to create a library of compounds would be to perform a similarity search in commercially available chemical spaces to find analogs of the initial fragment.⁷⁷ The virtual library can then be filtered before docking to make sure undesired molecule are removed. One prominent example of such undesired molecules are potentially toxic molecules.⁷⁸ Removing such compounds at the beginning avoids difficulty later on when the project is moved to in vivo studies. Another example are Pan-assay interference compounds (PAINS).⁷⁹ PAINS are molecules that are often false-positives in high-throughput assays. The extraction of PAINS gives a certain assurance to the scientist that compounds purchased for study are not known to interfere with typical assay setups. By filtering the virtual library to exclude toxic compounds and PAINS, the library is decreased in size, speeding up the docking step. After the creation

of both inputs, the binding site definition and the virtual library, the actual docking procedure is applied.

1.1.3.2 Docking and Scoring

The next steps of the CADD workflow explained here, are docking and scoring, which are both performed by docking programs. The previously prepared receptor and virtual library for to-be-docked compounds are required for the docking program. Docking is the computational calculation of possible receptor-ligand complexes.⁷² To reduce calculation time the receptor is usually kept fixed. This means no rotational or translational movement of the amino acids, water molecules or other ligands included in the binding site definition are considered. Solely the ligand of interest is scrutinized. Each movement of the ligand defines one docking pose. After docking, the energy released or needed for interactions between the ligand and the receptor are estimated by the program. These estimations are referred to as scoring. Scoring functions describe physical interaction energies in mathematical terms.⁸⁰ Each docking program usually utilizes a different scoring function.⁸¹ The differences are mainly based on the implementation or prioritization of different interaction types. Therefore, depending on which docking program is used, a different score might result for the same docking pose. A docking score value can be the first filter in a CADD workflow to remove undesired docking poses. Docking approaches are supposed to distinguish between binders and non-binders based on the docking score value. Thus, thousand to millions of compounds can be filtered to enrich binders.

1.1.3.2.1 Template-based Docking

In case of CFS, the resulting fragment might be used for a special version of docking in a CADD workflow. This docking version is referred to as template-based docking. A docking program offering such template-based docking is FlexX⁸² in SeeSAR. Template-based docking means that the information of the fragment binding pose is utilized for follow-up compound positioning during docking. The 3D-structure of the fragment bound to the protein is known and can be used as a template to place and orient the docking poses in a certain direction. Each individual follow-up compound from the library is superimposed to the template fragment, matching the common substructure as best as possible. In this way a biased approach of docking is performed. It can be assumed that follow-up compounds will retain the original binding mode for most parts. Thus, follow-up compounds that do not fit well to the receptor based on the fragment template are filtered out. The number of successfully docked follow-up compounds can be energy-minimized further through re-scoring.

1.1.3.3 Re-scoring of Docking Poses

A so-called re-scoring can be employed after the docking program to obtain score values from a different perspective, possibly less biased than the score that was optimized in the original docking procedure.⁸³ A possible bias results from the link between docking and scoring. The docking program will optimize the follow-up compound pose to the inherent scoring function. In this way an artificially high docking score might be calculated because docking also optimizes for the specific imperfections inherent to any individual scoring function. A possibility for re-scoring is the HYdrogen bond and DEhydration scoring function (HYDE)⁸⁴ from SeeSAR. This scoring function especially considers, as the name states, hydrogen bonds of the follow-up compound-receptor complex and dehydration energies upon binding. The HYDE program also has a further advantage. Based on these binding energy estimations for the receptor and the docked follow-up compound, the program allows visualization of favorable and unfavorable interactions of the follow-up compounds as colored spheres. Thus, during the next step of visual examination of the docking poses, the docking poses can be assessed for promising follow-up compounds.

1.1.3.4 Follow-up Selection

The final docking poses are examined visually in a graphical user interface (e.g. the GUI of SeeSAR) or via additional programs able to display the ligand-protein complexes.^{85,86} During visual inspection, various information of protein-ligand interactions is employed to judge the poses and choose promising molecules for experimental testing. Such protein-ligand interaction knowledge includes, the interaction angles, interaction distances, and the complementarity of the binding pocket shape and the molecule shape.⁸⁷ Based on the HYDE re-scoring, an estimation of the binding affinity is made, which additionally helps the scientist to decide on a limited number of follow-up compounds to be purchased.

The computational approach of docking allows to test millions of molecules and reduce them to hundreds of molecules that a scientist can investigate. Thus, CADD workflows are useful for *in silico* high-throughput screening of possible follow-up compounds. However, computational calculations or predictions of binding poses are not as reliable as experimental results and thus it has been shown in many cases that predicted poses differ from the experimental poses or molecules do not bind at all.^{36,88} Experimental validation of the calculated docking poses is required to confirm actual binding and the respective binding pose.

1.2 Enzymes as targets

One of the most popular drug target classes are enzymes.^{12,89} They have mostly a well-defined active site, often a well-understood mechanism, and a clear biological activity. Information about their substrate, product and natural inhibitors can already be used as

starting points for drug development.⁹⁰ Differences in substrate and inhibitor provide especially valuable information for understanding which features are vital for binding. One well-studied enzyme class are the proteases. Proteases can be categorized based on their catalytic center; namely aspartic proteases, serine proteases, cysteine proteases, metalloproteases and the less abundant glutamic proteases and threonine proteases.^{91,92} Each protease category has its own specific targets in drug discovery. Two of them will be addressed in this work.

1.2.1 Aspartic Proteases - Endothiapepsin

Aspartic proteases play key roles in diseases like hypertension, AIDS, amyloid diseases, and malaria and therefore are interesting targets in drug development.⁹³ In academia often a model protein is investigated. They can support research as they are easy to produce and easy to handle compared to for example human proteins. The results gained from model proteins can be translated to other proteins in their family or be used to improve or validate experimental setups.

Endothiapepsin (EP) is such a model protein for aspartate proteases.⁹⁴ The protein comprises 330 amino acids and acts as a monomer, with two domains forming the active site. The active site is made up primarily of two aspartic acids (Asp32 and Asp215) connected via a water molecule. The water molecule is positioned in a way that it can nucleophilically attack the peptide bond of a bound peptide and thus hydrolyze it.⁹⁵ The bound peptide is recognized by several other amino acids, which are referred to as specificity pockets (S). Thus, the binding site is made up of the catalytic center, and the pockets S1 to S6 and S1' to S3' (**Figure 10**).⁹⁶

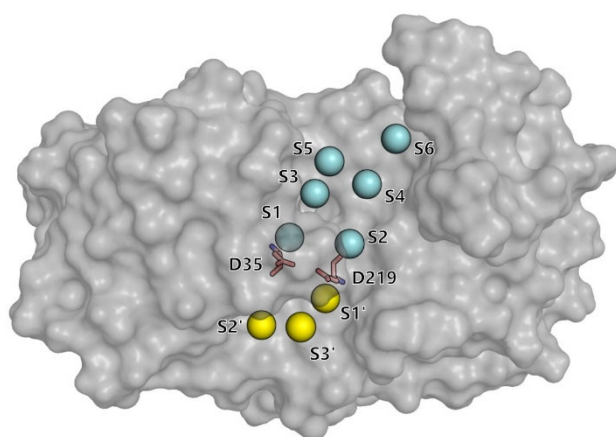


Figure 10: Overview of the active site of EP. The protein is shown as a grey surface. The active site aspartates are shown as pink sticks. The specificity pockets are shown as spheres. The depicted structure is PDB ID 4Y5L. The figure has been prepared using PyMOL.⁸⁵

EP is a suitable model protein as it can be produced easily in high amounts, is stable at room temperature, and forms highly diffracting crystals.⁹³ The EP crystals are robust and thus especially suitable for crystallographic methods, like CFS. Therefore, EP had been screened with several fragment libraries. This allows comparison of newly developed

fragment libraries with established ones to assess their performance. The protein-ligand interactions of EP found through CFS campaigns are also investigated in much detail.^{36,45,96} For example, it has been shown that positively charged fragments are common binders, while other fragments are found less abundantly. Based on the manifold information available for EP, it was chosen as a model protein for method development during the doctoral thesis work.

1.2.2 Cysteine proteases – SARS-CoV-2 MPro

Cysteine proteases are another typical target protein class. Due to their high activity, they are usually produced in an inactive form, as are also some other proteases. A prodomain acts as the inhibitor of the protease. The activation is achieved through different mechanisms, as for example an accessory molecule or a drop in pH.⁹¹ Cysteine proteases are involved in various diseases, for example, in osteoporosis, rheumatoid arthritis, or in hemoglobin hydrolysis in the case of parasitic cysteine proteases.⁹⁷ Additionally, cysteine proteases are vital for viruses, such as severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2), to act inside the host cell.⁹⁸ This cysteine protease is named main protease (MPro) and has been the first and foremost drug target since the pandemic started in 2020. The protease is similar to its related cysteine proteases in SARS-CoV-1 and middle east respiratory syndrome (MERS) that have been intensely studied as a reaction to their respective pandemics.⁹⁹ MPro acts as a dimer after cleaving the prodomain of itself. The active site is similar to the aspartic protease, built up by the catalytic center and the peptide pockets. Here, the catalytic dyad is comprised of the catalytically active Cys145 and the His41.¹⁰⁰ The specificity pockets are categorized into S1 to S5 and S1' (**Figure 11**).¹⁰¹

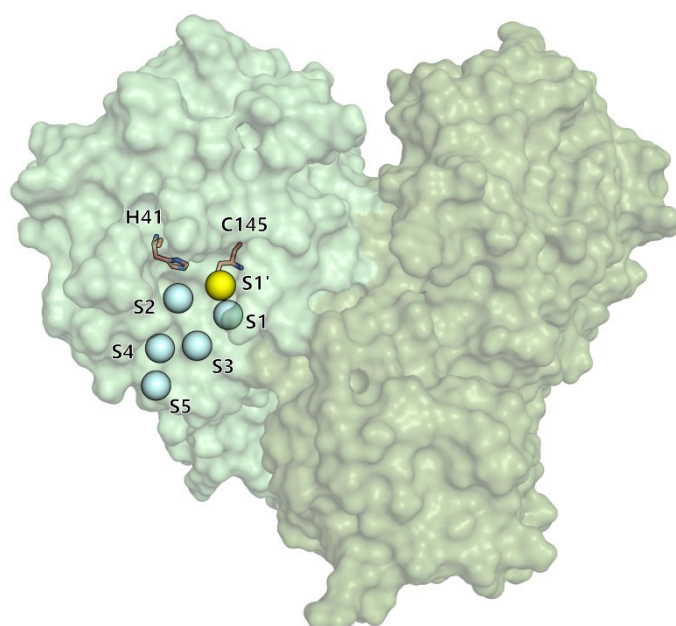


Figure 11: Overview of the active site of MPro. The protein is shown as a green surface, each monomer in a different shade. The active site histidine and cysteine are shown as pink sticks. The specificity pockets are shown as spheres. The depicted structure is an apo structure collected at BESSY II BL14.1, not deposited in the PDB. The figure has been prepared using PyMOL.⁸⁵

The global focus on MPro as a drug target resulted in various ligand screenings to find starting points for drug development (for example ^{102–104}). One highly prominent fragment screening campaign, called Covid Moonshot, aimed to develop a drug against MPro in an open science way.¹⁰⁵ This means that each step of the project was made publicly available, and every scientist was able to contribute to the development. However, even though several studies have been performed in the last years and projects like Covid Moonshot have progressed quite far, more and better inhibitors are required. Society needs to be prepared for possible virus variants that bypass so far developed drugs and society especially needs to be prepared for future pandemics. Therefore, MPro is chosen as a presently highly relevant target for this work.

1.3 Protein-Protein Interactions as targets

Enzymes, receptors, and ion channels play key roles in cellular functions, which is why they are common targets in drug discovery.¹² However, such proteins often have highly conserved binding sites across species or across their isoforms. For example, kinases are promising targets for several cancer types, but their active sites can differ in only one amino acid.¹⁰⁶ Furthermore, the pool of possible drug targets is highly limited when focusing on these proteins. It has been estimated that only 1.5% of the human genome has been targeted by the pharmaceutical industry with small molecules.^{107,108} The human protein interactome was estimated to encompass about 130.000 to 650.000 binary PPIs.^{109,110}

In recent years, there has been increased interest in targeting protein-protein interactions (PPIs). PPIs are necessary for many cellular processes like signal transduction, protein degradation, and gene expression. For example, the transcription factor HNF1 β is only transferred from the cytosol into the nucleus upon binding of Importin α .¹¹¹ Another example of a vital PPI is the activation of the histone deacetylase enzyme HDAC3, which is activated by the structural rearrangement of the SMRT-DAD domain upon binding of HDAC3.¹¹² There are also so-called hub proteins, that engage in several PPIs with different proteins. For example, Prp8, also known as the heart of the spliceosome, is responsible for arranging various proteins throughout the splicing process.¹¹³ Some PPIs induce a disorder-to-order conformational change, which acts as a regulation mechanism, for example the molecular regulator proteins 14-3-3 that bind mainly intrinsically disordered proteins.¹¹⁴ Therefore, by including PPIs in drug discovery efforts, a novel field of drug targets opens up.

PPIs are still an underexplored target due to the challenging nature of PPI sites. While enzymes have a confined active site surface area of about 300 – 1000 Å², PPI sites are often larger with about 1500–3000 Å².¹⁰ The amino acids responsible for the PPI are spread across the binding sites in so-called hot spots. To modulate an interaction the

molecule will have to hit several hot spots. Additionally, active sites often present hydrophobic surfaces, while PPI sites are more hydrophilic and hot spots often include amino acids like Trp, Arg, Tyr, Asp, or His.^{10,115} Disordered proteins are more difficult to target as they do not always present their active binding site. Though as two proteins are involved in PPIs, the 14-3-3 protein for example can be targeted instead of its binder.¹¹⁶ Additionally, PPI sites are highly unique compared to active sites of enzymes. Thus, adverse binding can be avoided, and side effects are minimized at later drug development stages. PPIs present different challenges as targets compared to more typical targets like enzymes, but it has been shown to work successfully. The second fragment-based FDA-approved drug Venetoclax targets the BCL-2 protein involved in chronic lymphocytic leukemia.¹¹⁷ Thus, demonstrating the possibility to target PPIs.

1.3.1 Pre-mRNA Splicing – the spliceosome

There are several steps during gene expression inside a cell. In simple terms the cell must transcribe the DNA to RNA and then translate RNA to proteins. However, this is only applicable for prokaryotes. In eukaryotes the gene expression is more complex. Eukaryotic cells have higher compartmentalization and a higher number of protein variants. Additional steps like pre-mRNA splicing and post-translational modifications are present. Pre-mRNA splicing is needed as eukaryotic genes are made of introns and exons. Introns are parts of the gene that are excised from the pre-mRNA to yield mature mRNA. The splicing process is a cycle, as most of the small nuclear ribonucleoproteins (snRNPs) involved are recycled for the next splicing event. A large multiprotein-RNA machine is responsible for splicing the pre-mRNA, the spliceosome. It is made up of 5 small noncoding RNAs (snRNAs) and hundreds more of proteins. The snRNAs are part of larger protein-RNA complexes: the snRNPs. SnRNPs are defined by their snRNA and are called U1, U2, U4, U5 and U6. The snRNPs catalyze the excision of the introns by two transesterification steps, while the additional proteins in the spliceosome regulate the process and recognize the pre-mRNA. The spliceosome is a highly dynamic complex, involving major compositional and conformational rearrangements throughout the splicing cycle. This highly complex process has been studied for decades mainly focusing on the system in yeast. Yeasts are less complex than as for example human cells that include different splicing levels depending on the cell type. An additional process called alternative splicing, which adds another level of complexity is not considered here even though it was shown recently that yeasts also perform alternative splicing.¹¹⁸ In simplified terms the splicing cycle can be described as follows (**Figure 12**)^{119,120}:

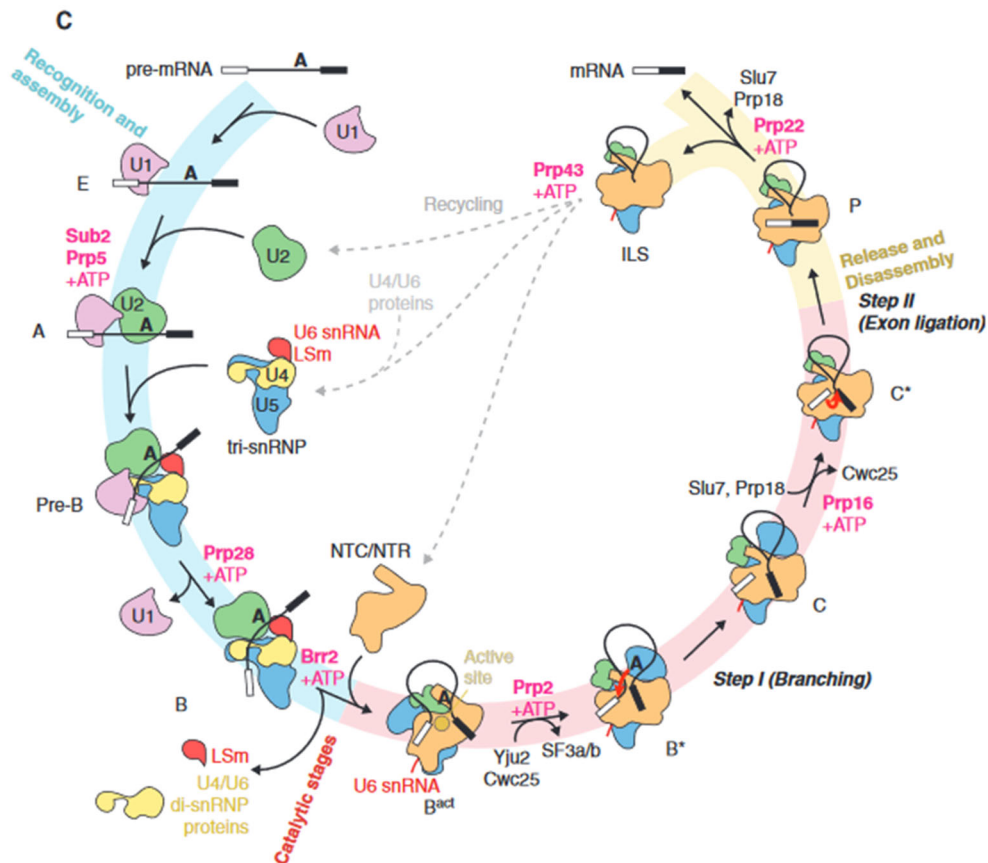


Figure 12: Simplified schematic view of the splicing cycle in yeast adapted from Plaschka *et al.*¹²⁰ The cycle is split into three parts; the recognition and assembly step, the catalytic stages and the release and disassembly step. The pre-mRNA is shown as a line for the intron and a white and black box for the exons. The spliceosome is displayed as blobs for each snRNP and the NTC/NTR. Important helicases are written in pink. Other important proteins are written in black entering and exiting the spliceosome at different steps. Each spliceosomal complex is identified with a letter. The recycling step is shown with a grey dotted line.

The first step is the recognition of the pre-mRNA. The U1 snRNP recognizes the intron via its 5' splice site (SS). This complex is termed complex E. Next, the U2 snRNP binds to the 3' SS and the branch point site and thus changes the complex to the pre-spliceosome, also complex A. To build up the pre-B complex the U4/U6-U5 tri-snRNP enters. Afterwards the pre-catalytic complex (B complex) is formed by binding of Prp28 and the removal of the U1 snRNP. Another important helicase Brr2 binds to the spliceosome and unwinds the U4/U6 snRNA complex. Through the unwinding the U4 snRNPs exit the spliceosome, which results in the activated complex B^{act}. Another important protein joins the spliceosome, Prp2. Only upon binding of Prp2 the spliceosome is catalytically activated (complex B^{*}). Now the first transesterification reaction can be performed on the 5' SS. Upon that the intron is folded towards the branchpoint and is connected there as the lariat structure. In this conformation the spliceosome is considered as complex C. The complex P (post-spliceosomal complex) is reached after the second transesterification step, while the U2, U5 and U6 snRNPs are still bound to the mature mRNA and the

intron-lariat. The last few steps involve the regulated disassembly of the individual components, which enter again a new cycle of splicing on a new pre-mRNA substrate.

Although studied for decades, the understanding of the splice cycle is still not complete. For example, due to the recent developments in the cryo-EM field,^{121,122} it was possible to visualize the different spliceosomal complexes and discover new distinct steps in the splicing cycle.^{120,123,124} More work is needed to fully understand all players involved and their possible connections to various diseases. It has been shown that spinal muscular atrophy, or retinitis pigmentosa, or neurological diseases like Parkinson result from aberrant pre-mRNA splicing.^{125,126} Some splicing modulators have been developed in the recent years. Prominent examples are some of the first splicing modulators identified, including spliceostatin A,^{127,128} pladienolide B¹²⁹ and herboxidiene.¹³⁰ All three small molecules target the SF3B protein.¹³¹ SF3B is part of the U2 snRNP and recognizes the branch point.¹³² The modulators were derived from natural products and the target was unknown at first. Through the efforts to identify the target of these splicing modulators the structure and function of SF3B had been uncovered.^{133–135} Based on this knowledge a synthetic small molecule, H3B-8800, had been designed, and is still undergoing clinical trials.¹³⁶ It had also been designated by the FDA in 2017 as an orphan drug to target leukemia.¹³⁷ The example shows that detailed mechanistic knowledge is a key prerequisite for directed and specific drug development. To support further investigation of the splicing process, tool compounds need to be developed that modulate certain proteins, RNA-protein interactions (RPIs), or PPIs and thus foster detailed functional understanding of the spliceosome.

1.3.2 The interaction of Aar2 and Prp8

One key protein in the spliceosome was named the heart of the spliceosome: Prp8. It is a hub-protein that is highly conserved and about 230 – 280 kDa in molecular weight, depending on the exact species.¹¹³ It is part of the U5 snRNP and thus enters the spliceosome for the formation of the B complex. It organizes various proteins throughout the splicing process to regulate the different steps. Therefore, it is involved in many PPIs and RPIs. Several domains have been annotated to Prp8, for example the Jab1/MPN domain, the RNaseH-like domain, the reverse transcriptase-like domain and the endonuclease domain (**Figure 13**).¹³⁸ The RNaseH-like domain has a similar sequence and fold to the RNaseH however it does not show any activity, i.e. is a pseudo-enzymatic domain.

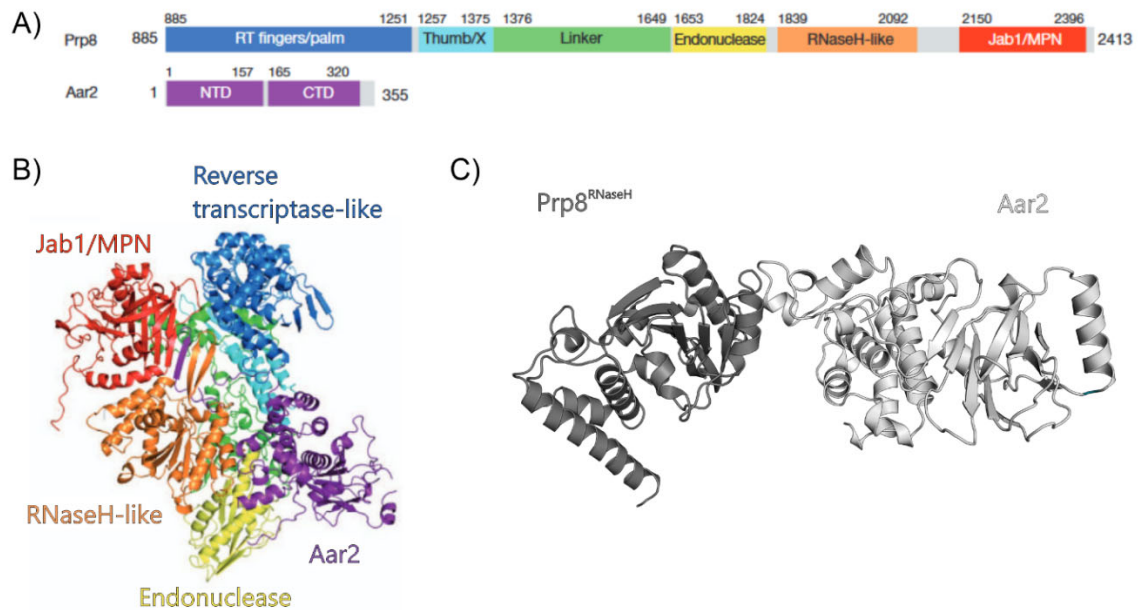


Figure 13: Overview of Prp8 domains and the interaction with Aar2 adapted from Galej *et al.*¹³⁸. A) The domain architecture is shown of Prp8⁸⁸⁵⁻²⁴¹³ and Aar2 with different color per domain. B) Complex structure of Prp8⁸⁸⁵⁻²⁴¹³ and Aar2 with each domain colored according to A. (PDB ID 4I43) C) AR complex structure as worked with during this doctoral research project (collected at BESSY II BL14.1, not deposited in the PDB).

Aar2 is one of the proteins that interact with Prp8, although not in the context of the spliceosome itself. Prp8 is present in the cytosol in a pre-U5 snRNP complex and thus must enter the nucleus.¹³⁹ This is orchestrated by Aar2. It binds to Prp8 and transfers the pre-U5 snRNP into the nucleus. Aar2 binds to the RNaseH-like domain of Prp8 with a high affinity in the low nanomolar range.¹³⁹ The pre-U5 snRNP is transported into the nucleus and premature binding of other proteins or RNA to Prp8 is prevented by Aar2. After two sequential phosphorylations of Aar2 via an unknown kinase, Aar2 is removed from this complex. The freed-up binding site is taken by the main helicase Brr2. Interestingly, Aar2 was required to gain a crystal structure of a near full-length Prp8 (**Figure 13**).¹³⁸

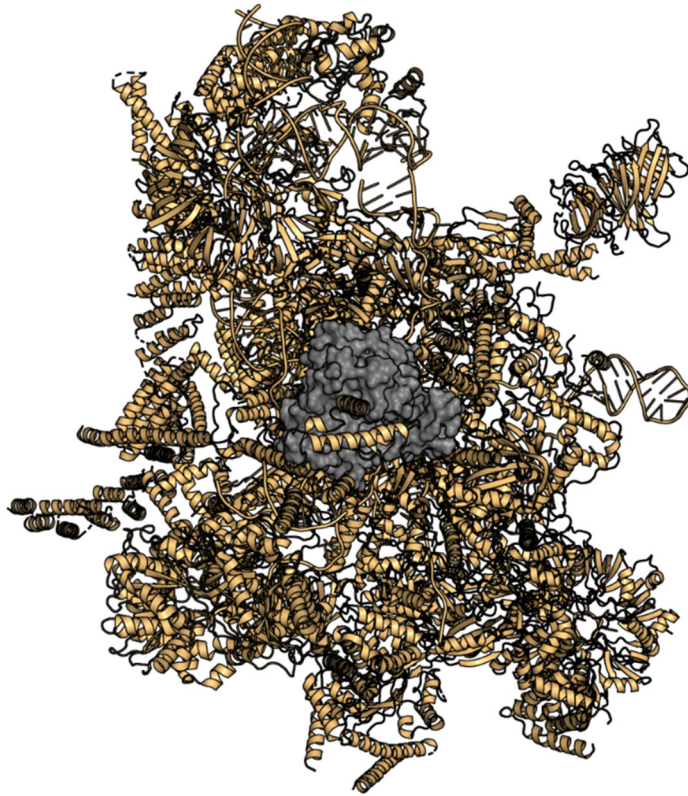


Figure 14: Presentation of an exemplary cryo-EM structure of the P-complex. The Prp8^{RNaseH} domain is shown in surface view in grey and the remaining proteins and RNA is shown in yellow cartoon. The depicted structure is PDB ID 6BK8. The figure has been prepared using PyMOL.⁸⁵

Further information about Aar2 is currently not known. There have been studies showing additional functions in humans, but so far not in yeast.¹⁴⁰ Therefore, it could be highly interesting to find out more about this protein, possibly via a probing of its surface with CFS. Additionally, the interaction between Prp8^{RNaseH} and Aar2 could be enhanced via a tool compound. It could be examined if this stalls the splicing process completely, or this modulation could be circumvented by an alternative splicing process. Furthermore, by screening the protein-protein complex, the campaign acts as two campaigns at once. Fragments bound to Prp8^{RNaseH} can also be developed into inhibitors for known interaction sites with other proteins. As Prp8^{RNaseH} is located in the middle of the spliceosome it engages in various additional PPIs, that can be targeted (**Figure 14**). One CFS screen will result in multiple possible pathways after analysis.

2. Aims

This project combines two main objectives: the improvement of the current CFS workflow and the development of small molecule modulators. The two objectives are related with each other. The CFS campaigns can only be performed in a time-efficient manner due to developments that support the user. Additionally, the challenges that occur while developing small molecule modulators point towards the limitations of the current workflow. These limitations will be improved throughout the project and tested again. The two main objectives can be split into several underlying goals which were tackled in this work:

- 1) *Improvements of the crystal handling step to speed up crystal soaking and harvesting.*
- 2) *Simplifying data management with a computational interface to track and handle CFS data.*
- 3) *Establishing an efficient CFS workflow, which new CFS users can follow easily.*
- 4) *Performing the CFS campaign of the F2X-Universal Library against the AR target.*
- 5) *The development of a computational workflow to help with follow-up design.*
- 6) *The design of small molecule spliceosome modulators.*
- 7) *The investigation of the influence of space groups on hit rates.*

3. Results

3.1 Support for crystal manipulation and harvesting with the EasyAccess Frame



Figure taken from Barthel *et al.*

Publication:

Barthel, T.; Huschmann, F. U.; Wallacher, D.; Feiler, C. G.; Klebe, G.; Weiss, M. S.; Wollenhaupt, J.

Facilitated Crystal Handling Using a Simple Device for Evaporation Reduction in Microtiter Plates.

J. Appl. Crystallogr. **2021**, 54 (1), 376–382.

<https://doi.org/10.1107/s160057672001647>

Many steps of the CFS workflow have been successfully automated already, especially regarding the protein crystallization, the data collection and data analysis.^{51,52,55,63,141–146} However, one main bottleneck remains: the crystal handling. Harvesting hundreds to thousands of crystals is time-consuming and labor intense. To tackle this, different solutions were developed, which are either only available for a specific facility or quite expensive.^{46–49} Therefore, we arrived at a different approach to tackle the bottleneck of crystal handling: a device that is affordable even for smaller academic groups, and that can be transported easily between laboratories. The original inspiration was taken from sliding puzzles by Franziska Huschmann, a postdoctoral fellow at HZB at that point. Based on her initial developments we continued to improve the device further until commercialization. We named this device the EasyAccess Frame (**Figure 15**).

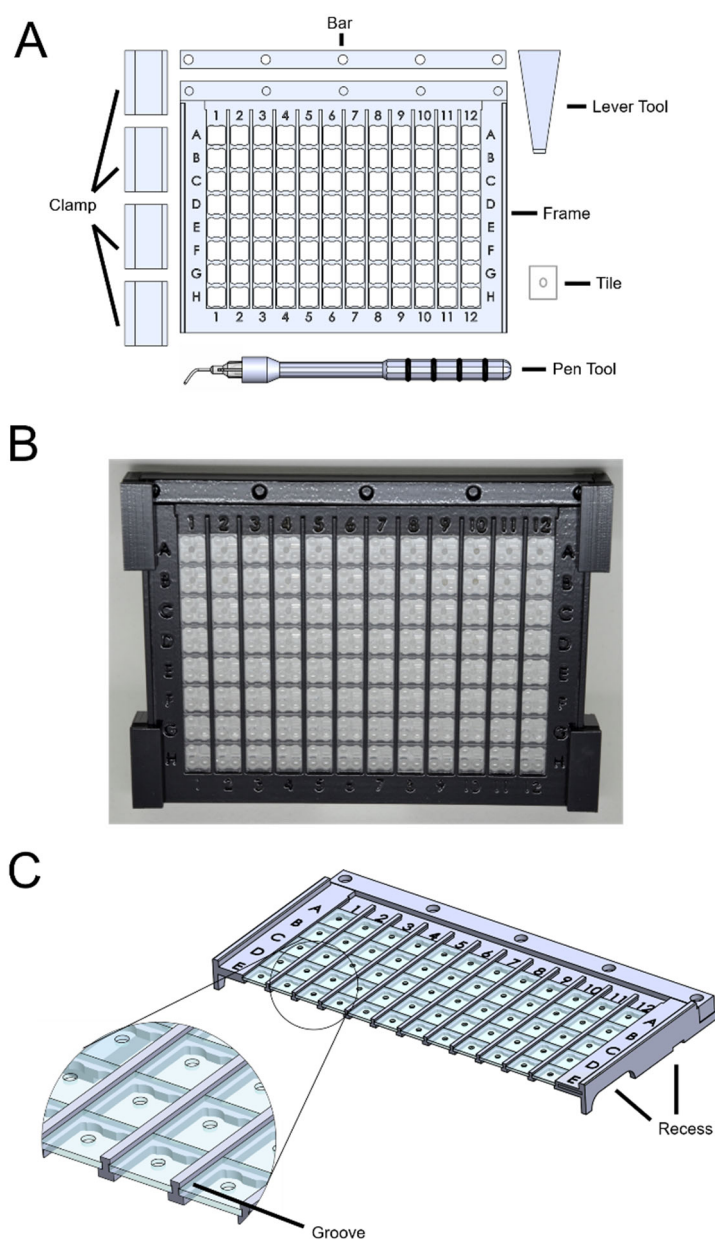


Figure 15: The EasyAccess Frame set. A) A schematic top-view illustration of the EasyAccess Frame and all parts included in the set. Shown are the frame, the clamps, the bar, the lever tool, one exemplary acrylic glass tile with a depression in the middle of the tile, and the pen tool. B) A photo taken of an EasyAccess Frame on top of a crystallization plate with the inserted acrylic glass tiles, the bar and four clamps. C) A schematic illustration of the EasyAccess Frame from a side angle showing the recesses on the side of the frame. The frame is shown cut in half with a zoom-in to indicate the grooves. Figure adapted from Barthel *et al.*¹⁴⁷

The device is a lid that is placed on top of a typical crystallization plate. In this case it has been optimized for MRC 3-lens 96-well low-profile plates, a widely used format in crystallization and fragment screening facilities. However, it could be optimized for other crystallization plates if needed. It is composed of a 3D-printed frame and 96 transparent acrylic glass tiles which can be moved up and down in provided grooves of the EasyAccess Frame. A 3D-printed bar prevents the accidental removal of these tiles. Additionally, further 3D-printed tools have been developed that ease the handling of the EasyAccess Frame. Four clamps are provided which fix the crystallization plate to the frame. Further included is a lever tool to remove the crystallization plate safely and easily from the frame. The frame provides special recesses on the sides which can be accessed via the lever tool or a thumb (**Figure 15C**). A pen tool is provided with a metal top piece which fits the depression of the acrylic glass tiles. This way the tiles can be moved up and down. Alternatively, they can also be moved with the tip of a finger. Owing to the minimal cost of the materials used, the easy production and by avoiding robotics, the EasyAccess Frame is kept affordable for academic groups and small businesses. Due to the nature of the frame the working angle is reduced, calculated as a decrease of 2.4° (**Figure 16**). This minimal restriction still allows for successful crystal manipulation and harvest.

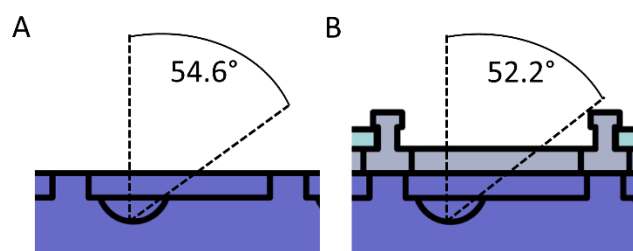


Figure 16: Schematic side view of one of the crystallization plate wells (in purple). The depression shown on the left side represents a lens for the placement of a crystallization drop. The working angle measured from the deepest point of the lens to the edge of the lens is shown in dashed lines. A) The working angle of 54.6° without the EasyAccess Frame is shown. B) The working angle of 52.2° with the EasyAccess Frame (grey) is shown. Figure adapted from Barthel *et al.*¹⁴⁷

The performance of the device was tested with a commercially available crystallization screen (Core Suite II) that provides an array of 96 different crystallization conditions, including alcohol-containing solutions, PEG-based solution, and high-salt solutions. The experiment was performed for 6 hours, which mimics a typical working session in a CFS campaign with about 300 crystals. The drops were set via a pipetting robot and directly afterwards the EasyAccess Frame was placed on top. Photos were taken with the microscope at 3 time points, a starting photo, after 30 min and 360 min. The drops were usually not perfectly circular, but slightly elliptical. Therefore, the diameter was measured in two directions, at the major and the minor axes, for each solution and each time point.

The mean was calculated from the two diameters. Based on these values the solutions were sorted into three categories: good (<15% reduction), medium (15-30% reduction), and bad (>30% reduction) (**Figure 17**). Additionally, in case a drop diameter stayed below 30% reduction but showed signs of phase separation or crystallization of the contained chemicals, it was categorized as bad (**Figure 17C**). Overall, 59 solutions were categorized as good, 18 solutions as medium, 17 solutions as bad, and 2 solutions were not included in the analysis due to low surface tension.

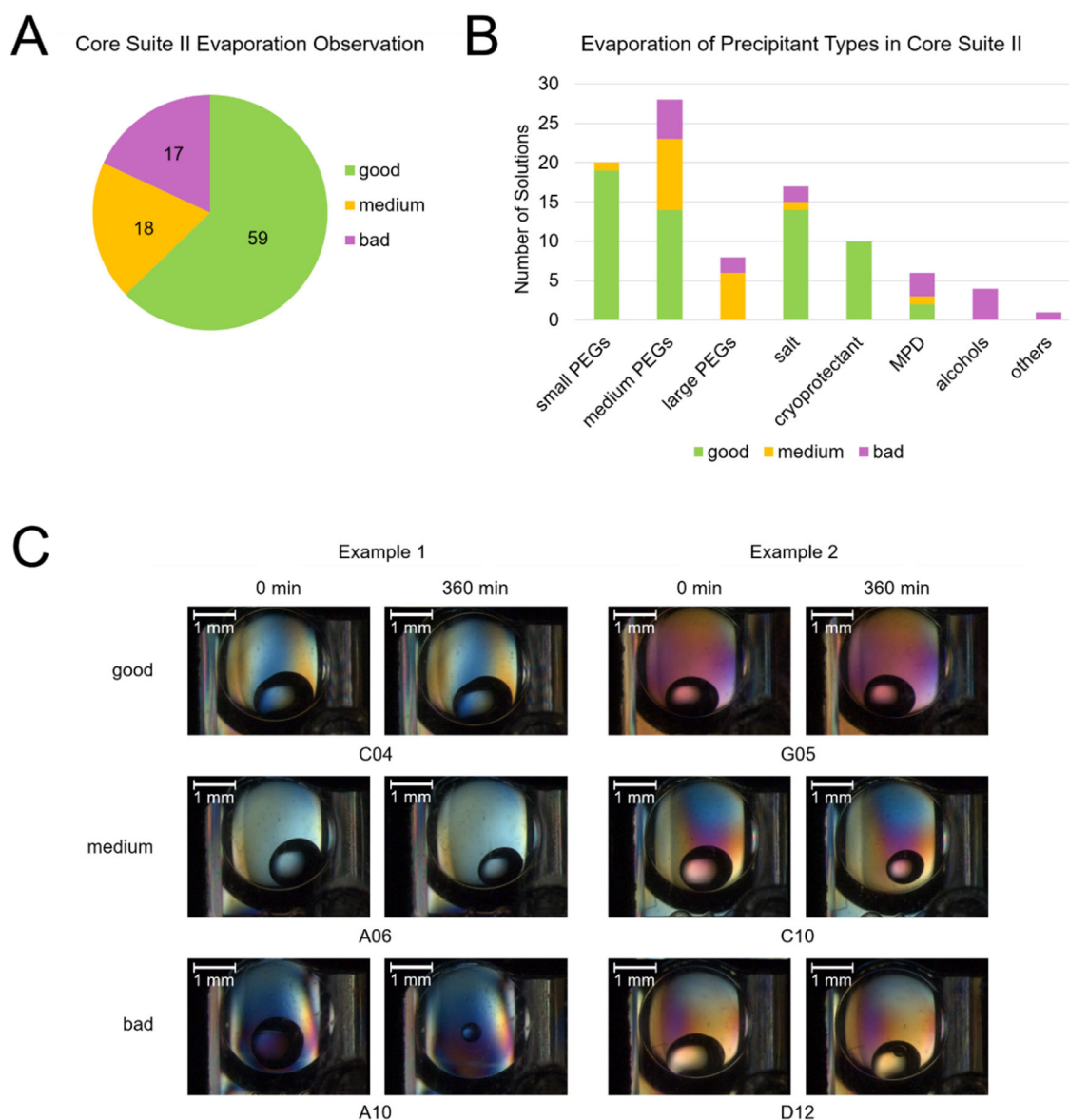
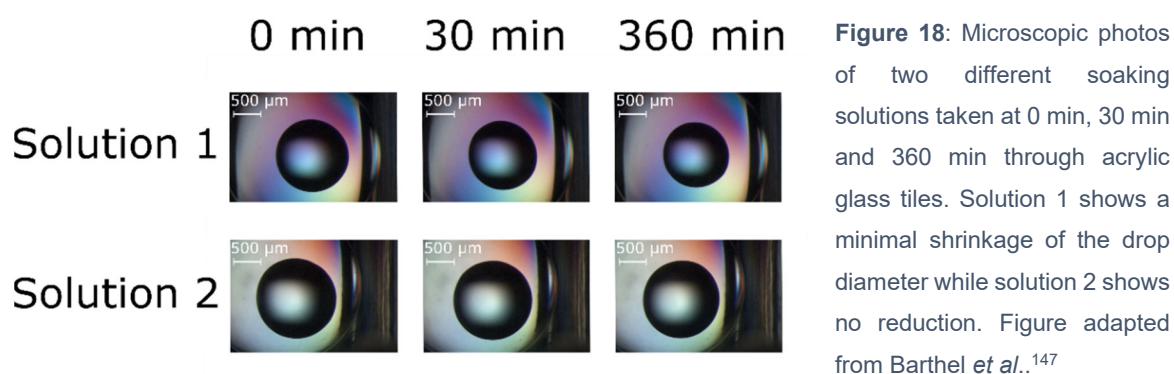


Figure 17: The results of the 96 different solutions tested for evaporation reduction with the EasyAccess Frame. A) A pie chart showing the number of solutions categorized into good, medium, and bad. B) The categorization of the solutions into precipitant categories with the respective number of good, medium, and bad solutions. C) Exemplary microscopic photos of crystallization drops taken at 0 min and 360 min taken through acrylic glass tiles. Two examples are shown for each category. Figure adapted from Barthel *et al.*¹⁴⁷

These results show that more than 50% of the solutions evaluated worked well for up to 6 hours. The problematic solutions often included alcohols and larger PEGs (**Figure 17B**) though for most chemicals it is not as clear. It is not possible to conclusively categorize certain chemicals as bad or good. Additionally, no correlation could be identified between the drop diameter reduction over time and the relative humidity of the drop considering the main crystallization condition compound.^{148–150} Therefore, it cannot be predicted how fast a crystallization drop will evaporate under the EasyAccess Frame based on its main component's concentration. Additionally, two further conditions, i.e. drop compositions, were tested, which have been used at the soaking step in previous CFS campaigns. Soaking conditions differ often from crystallization conditions as they must be cryoprotected and often include organic solvents like DMSO. Both solutions showed a minimal drop diameter reduction (<15%) after 360 min (**Figure 18**). Taken together this result and the fact, that solutions of the 96 conditions including typical cryoprotectants were categorized exclusively as good, there is a high possibility to work with soaking solutions for several hours. However, no prediction can be made for a specific solution. In each case, the working solution should be tested by the experimenter beforehand to assure safe working for up to 6 hours.



To summarize, we developed an affordable evaporation protection device, which speeds up crystal handling and restricts the working space in a minimal way. The frame ensures safe crystal handling and harvesting. It has by now also been commercialized. The EasyAccess Frame is small and lightweight, which allows for easy transportation to different laboratories. Thus, the device supports users conducting CFS campaigns without the requirement for the experimenter to be on site at the synchrotron.

3.2 FragMAXapp - Automation of crystallographic fragment screening data analysis and management via a user-friendly web application

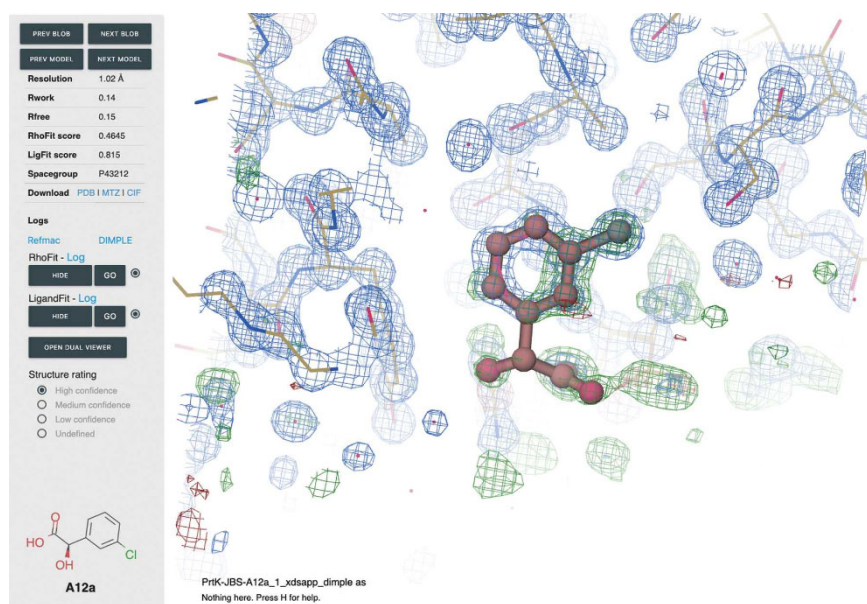


Figure taken from Lima *et al.*

Publication:

Lima, G. M. A.; Jagudin, E.; Talibov, V. O.; Benz, L. S.; Marullo, C.; **Barthel, T.**;
Wollenhaupt, J.; Weiss, M. S.; Mueller, U.

FragMAXapp: Crystallographic Fragment-Screening Data-Analysis and Project-
Management System

Acta Crystallogr. Sect. D Struct. Biol. **2021**, 77 (6), 799–808.

<https://doi.org/10.1107/s2059798321003818>.

Nowadays, hundreds of datasets can be collected at modern synchrotron MX-beamlines in a typical eight-hour shift. This is based in the efforts undertaken by synchrotron facilities to increase photon flux, speed up sample exchange via robotic mounting and incorporating sensitive and fast-readout detectors.^{4,51,53,55,151–154} The hundreds or thousands of datasets produced in typical screening campaigns cannot be analyzed individually in reasonable time. This results in a higher demand for subsequent automation for data analysis. Therefore, several solutions have been developed tailored to specific screening facilities at synchrotrons. One of them is the XChemExplorer platform at Diamond Light Source for CFS campaigns, which allows users to process, refine, and identify hits via a computational pipeline.¹⁴⁴ At the EMBL another pipeline called CRIMS has been build up for such purpose to support the users at their site.¹⁵⁵ However, transferring such software to other facilities is often accompanied by challenges, and thus the software has been not implemented at other synchrotron facilities. Here, we developed together with the FragMAX platform at MAX IV¹⁴⁵ a data analysis and CFS project management application. Its specialties are the option to perform multiplex analysis, it is a browser-based application, and it has a modular architecture, which allows for easier implementation at other sites.

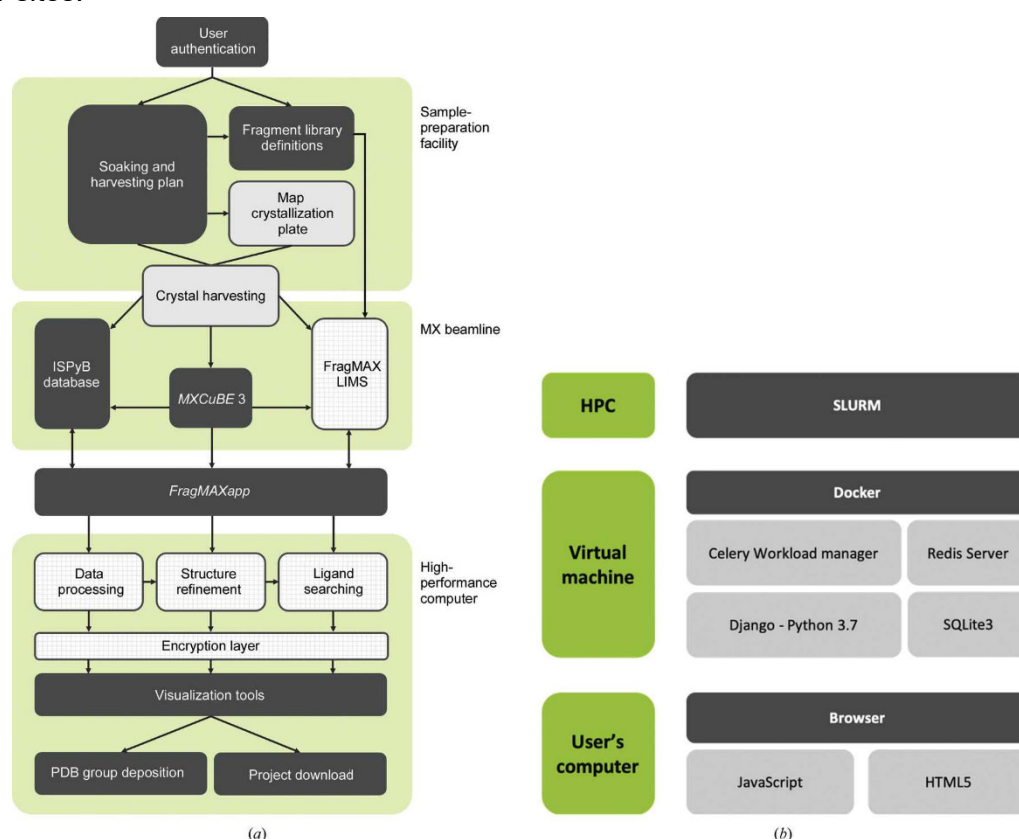


Figure 19: A schematic overview of the FragMAXApp. A) The different steps encompassed by the FragMAXApp. The dark grey boxes are steps performed via the HTML interface, light grey boxes are steps performed in the laboratory and the checkered boxes are steps performed in the background. The lines indicate the connection between individual steps. B) The computational environments applied in the FragMAXApp are depicted. Figure adapted from Lima *et al.*¹⁵⁶

The FragMAXapp has been developed in Python3/Django with JavaScript for the web interface. The individual steps that are included in the pipeline are the design of the experiment, the definition of the chosen fragment library, the sample definition via ISPyB,¹⁵⁷ the data collection via MXCuBE 3,^{158,159} the data analysis (processing, refinement, hit identification), the visualization of the data, the submission to the PDB and the download of the complete project (**Figure 19**). The data analysis pipeline is the heart of FragMAXapp. It includes various crystallographic software the user can choose from for data analysis. For processing it offers fastdp,¹⁶⁰ EDNAproc,¹⁶¹ autoPROC,¹⁶² DIALS,¹⁶³ XDS,¹⁶⁴ and XDSAPP.¹⁶⁵ For automatic refinement, it provides the options BUSTER,^{61,166} DIMPLE,⁶² and fspipeline.⁶³ Additionally, it includes software for hit identification, namely Phenix LigandFit,^{167,168} Rhofit,¹⁶⁹ and PanDDA.⁶⁴ For each hit identification program, the ligands must be available as a pdb or cif-file. Therefore, the structural information about each individual fragment is given by the scientist as SMILES and these are converted via AceDRG,¹⁷⁰ GRADE,¹⁷¹ or Phenix eLBOW¹⁷² into the respective file types. Thus, the web application offers a plethora of different analysis software, which results in the possibility to create 54 possible different combinations. Each combination might provide new information, leading to additionally identified hits.¹⁷³ All combinations can be run in parallel and the FragMAXapp suggests helpful options for each program, for example a reasonable resolution cutoff. The option to analyze all datasets at once or analyze a specific dataset are provided. The success or failure of data analysis can be monitored via the data collection tab of the web application (**Figure 20**). The user can follow the analysis of each dataset, the corresponding photographs of the sample mounted at the goniometer taken before the data collection and the 2D structure of the soaked fragment. In this way the user is given the possibility to handle their data flexibly and gain analysis information in one overview.

Data Collection Overview

● Successful
● Not started
● Failed



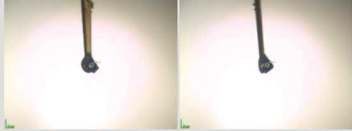

Info/Process	Protein	Sample name	Run	Proc Status	Ref Status	Lig Status	Resolution [Å]	Frames	Crystal Picture	Ligand
INFO	PrtK	A10a	1	<ul style="list-style-type: none"> ● autoPROC ● XIA2/DIALS ● EDNA_proc ● fastdp ● XDSAPP ● XIA2/XDS ● Pipedream 	<ul style="list-style-type: none"> ● BUSTER ● DIMPLE ● fspipeline ● Pipedream 	<ul style="list-style-type: none"> ● LigandFit ● RhoFit ● Pipedream 	1.24	3600		
INFO	PrtK	A11a	1	<ul style="list-style-type: none"> ● autoPROC ● XIA2/DIALS ● EDNA_proc ● fastdp ● XDSAPP ● XIA2/XDS ● Pipedream 	<ul style="list-style-type: none"> ● BUSTER ● DIMPLE ● fspipeline ● Pipedream 	<ul style="list-style-type: none"> ● LigandFit ● RhoFit ● Pipedream 	1.24	3600		

Figure 20: Exemplary screenshot of the FragMAXapp Data Collection Overview tab. For each individual dataset the data collection information is presented with the sample name. The processing, refinement, and ligand status are shown for each analysis program with corresponding circles indicating the status of the run. Additionally, the data collection parameters for detector distance (resolution) and number of frames are shown. Furthermore, two photos of the sample at the goniometer are depicted with 2D structures of the soaked fragment. Figure adapted from Lima *et al.*¹⁵⁶

Each analysis software has their own log files, offering information about the progress and errors of the analysis. A first overview is given by the traffic light display in the processing status column (**Figure 20**), which gives hints about failed processes. For more detailed information, log and error files can be accessed via FragMAXapp. In case special parameters are needed for successful analysis, the processing tab provides additional input options. The user has the possibility to add any changes to the software command as it would be possible with the standalone program. Additionally, the resulting electron density with the fitted ligand is visualized via UglyMol.¹⁷⁴ It allows for similar functions as Coot,¹⁷⁵ which is a familiar software for every crystallographer. Thus, the scientist can manually check the resulting electron density and fitted ligand and use this information to re-analyze datasets (**Figure 21**). All the analyzed data can be exported after analysis and downloaded. Thus, the data can be handled solely with FragMAXapp or be extracted for other analyses.

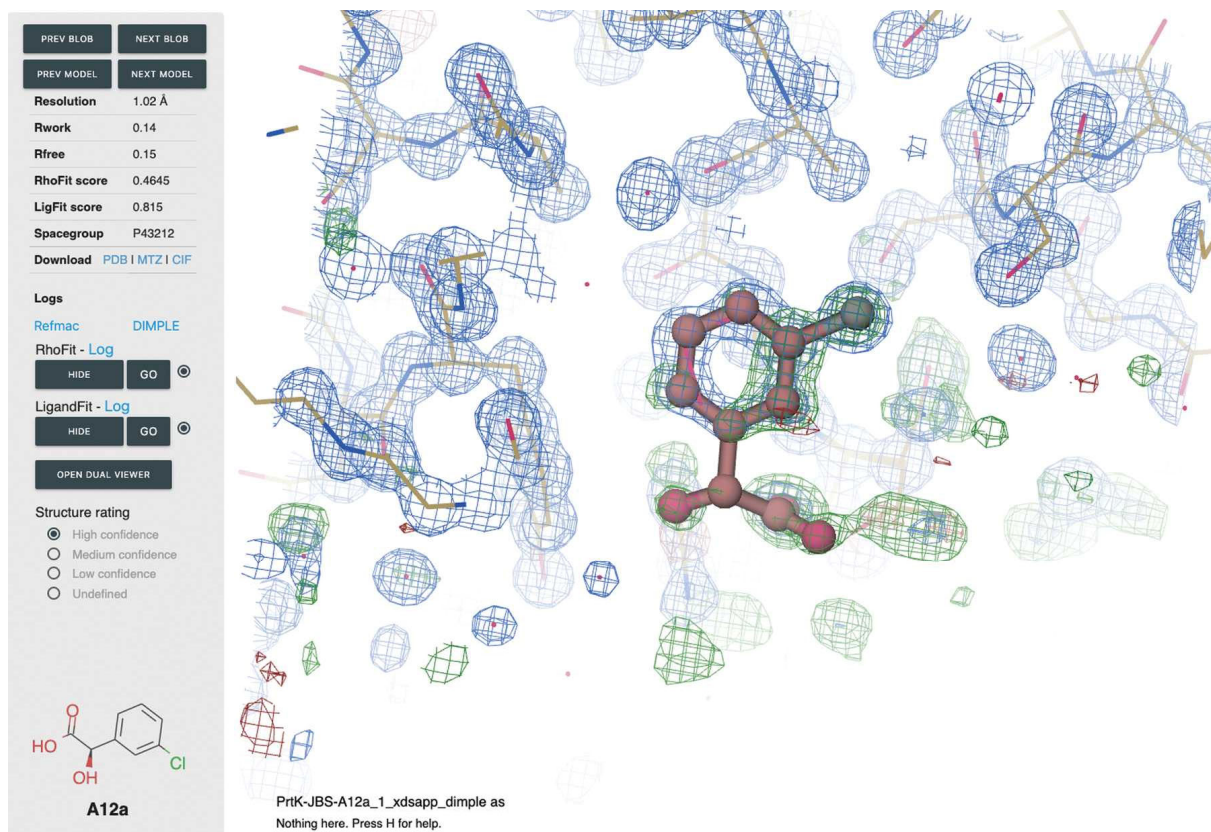
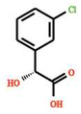
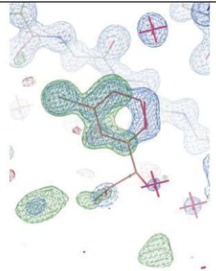
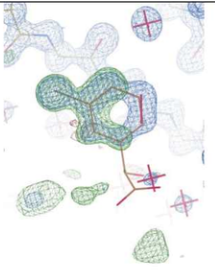
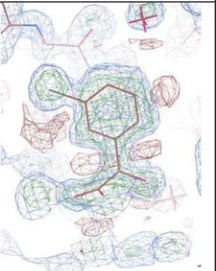
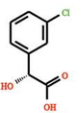
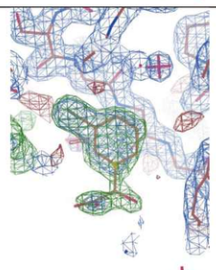
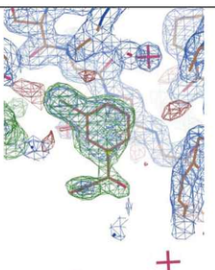
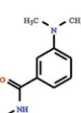
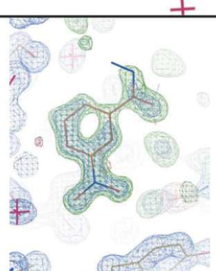
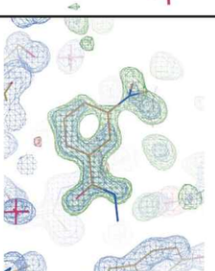
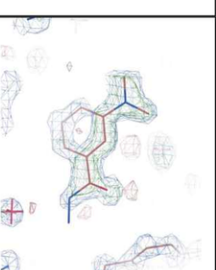
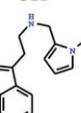
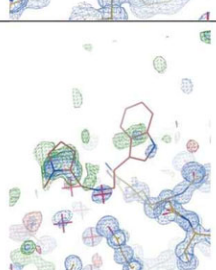
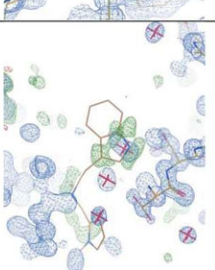
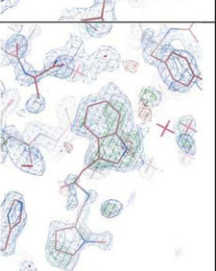

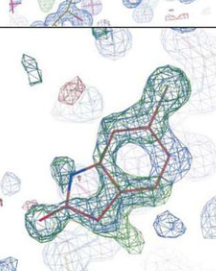
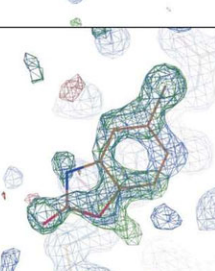
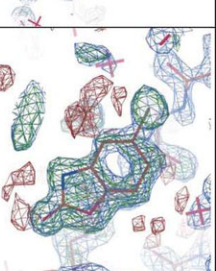
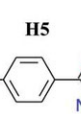
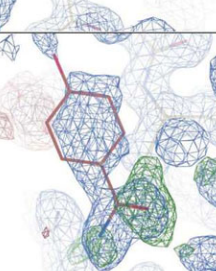
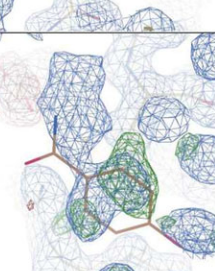
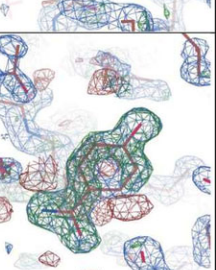


Figure 21: An exemplary UglyMol view of a dataset. A possible binding event is shown with a fragment structure in red. On the left side information regarding the dataset are shown with the possibility to rank the confidence in the presence of the fragment. Figure adapted from Lima *et al.*¹⁵⁶

The FragMAXapp has been applied to several different CFS campaigns,^{173,176–178} including an in-house campaign with proteinase K. The proteinase K campaign is presented here as an example for handling data with FragMAXapp in a real-life scenario. The Frag Xtal Screen library was chosen,⁴⁵ which comprises of 96 fragments. An additional 40 apo crystals were included in the campaign. The analysis of this campaign required two weeks to be ready for submission to the PDB, which is a rather short time for 136 datasets. The provided hit identification programs Ligandfit, RhoFit and PanDDA allowed identification of 18 binding events. The identified binding events differed between the applied programs (**Figure 22**). Therefore, the combination of several programs can be beneficial. This campaign shows the successful application of FragMAXapp to CFS data and the high throughput it enables.

Data set information	LigandFit	Rhofit	PanDDA
A12  <i>autoPROC fspipeline</i> 5roq			
A12 RT  <i>xia2/DIALS fspipeline</i> 5rop			No solution
B5  <i>EDNAproc fspipeline</i> 5rol			
C11*  <i>autoPROC fspipeline</i> 5rof			
F6  <i>xia2/DIALS fspipeline</i> 5row			
H5  <i>EDNAproc fspipeline</i> 5rpm			

* The binding sites identified by *RhoFit*, *LigandFit* and *PanDDA* are not the same.

Figure 22: A comparison between the automatic hit identification programs. In case of *LigandFit* and *Rhofit* the fragment has been built in automatically, in the case of *PanDDA* the fragment was modelled manually. Figure adapted from Lima *et al.*¹⁵⁶

The FragMAXapp covers all steps of the CFS campaign from experiment design to PDB submission. This simplifies the work for each user, as they have an intuitive web application that helps them with each step. The FragMAXapp has a special advantage compared to the other developed pipelines. It is built with a modular architecture that grants the possibility to transfer this system easily to other facilities. The heart of FragMAXapp has been successfully implemented at BESSYII and SIRIUS,¹⁷⁶ which resulted in published CFS campaigns. The FragMAXapp supports users especially by speeding up and simplifying data analysis.

3.3 The F2X-Universal Library and the F2X-Entry Screen enable efficient CFS

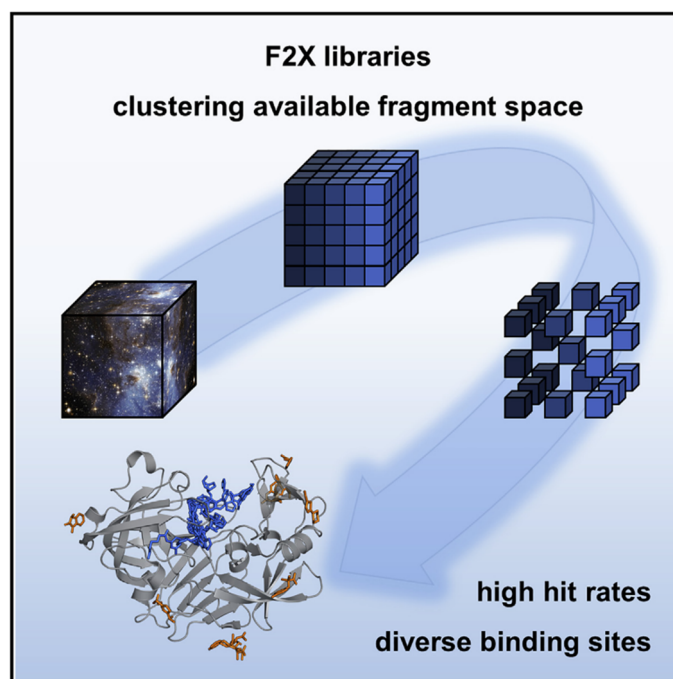


Figure taken from Wollenhaupt *et al.*

Publication:

Wollenhaupt, J.; Metz, A.; **Barthel, T.**; Lima, G. M. A.; Heine, A.; Mueller, U.; Klebe, G.; Weiss, M. S.

F2X-Universal and F2X-Entry: Structurally Diverse Compound Libraries for Crystallographic Fragment Screening.

Structure **2020**, 28 [6], 694-706.e5.

<https://doi.org/10.1016/j.str.2020.04.019>.

The success of a CFS campaign does not rely solely on the throughput provided by machine and software developments. One major determinant for success is the fragment library. It provides the chemical and structural diversity that can be covered by the campaign and can also be tailored to specific protein targets.¹⁷⁹ Several fragment libraries have been designed ranging from a few hundred to thousands of fragments. There are facility-specific libraries and commercially available fragment libraries that can offer unique fragments.^{180,181} Additionally, CFS offers the unique opportunity to screen with very high concentrations of fragments in the higher millimolar to lower molar range. However, to achieve such high concentrations in CFS a specialized presentation of the library is needed. The F2X-Universal Library represents such a specialized library for CFS. The compounds are dried onto typical crystallization plates, removing the solvent of these fragments, typically DMSO. Depending on the solubility of the fragment, high concentrations can be achieved by dissolving the fragment in a small volume of 0.4 μ l soaking solution. This setup has been specifically designed for CFS campaigns and provides users a library for any target.

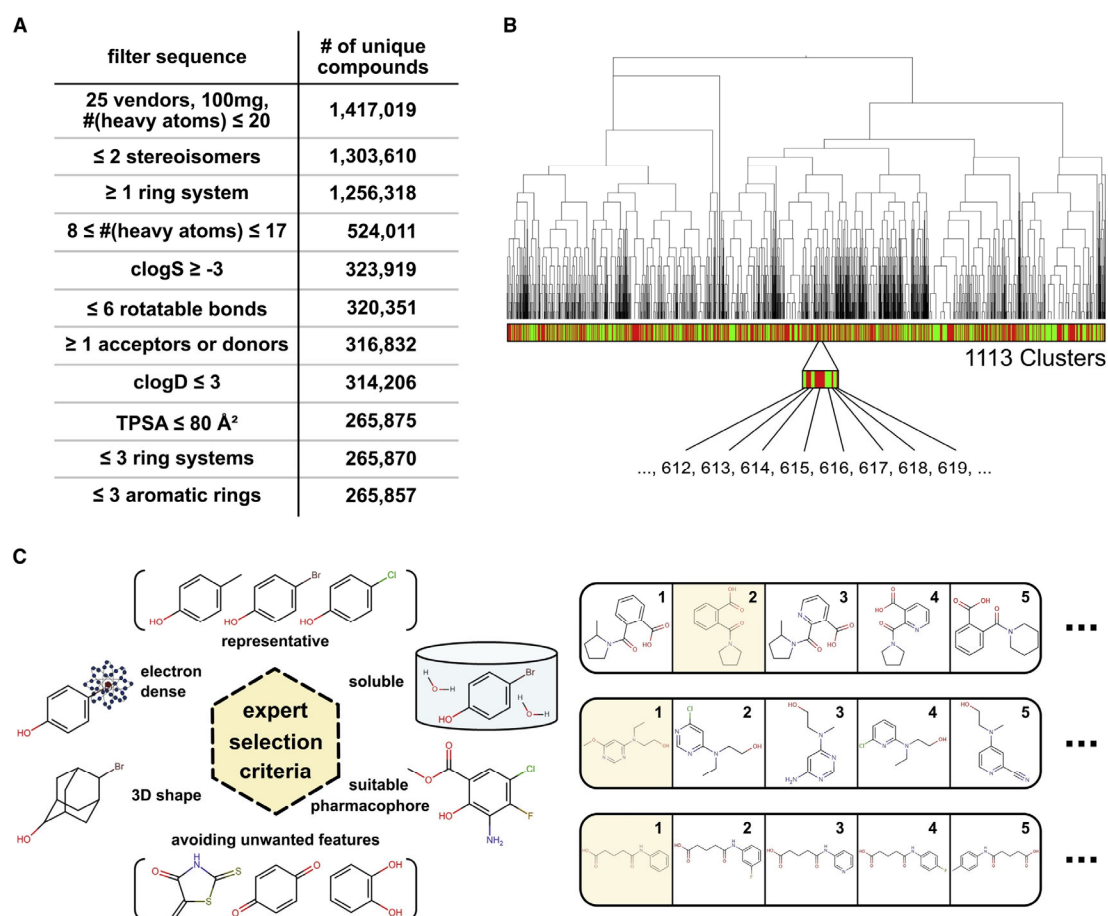


Figure 23: The design of the F2X-Universal Library. A) The standard medicinal chemistry filters applied to remove unwanted chemical properties are listed. B) A schematic view of the cluster tree cut at 1113 clusters. C) An exemplary presentation of the criteria considered at the expert selection step. Figure adapted from Wollenhaupt & Metz *et al.*¹⁷³

The F2X-Universal Library comprises 1103 fragments that have been chosen from the large pool of commercially available compounds. The first filters comprised the heavy atom cut-off of 20 and relaxed rule-of-three criteria.²² The further selection included standard medicinal chemistry filters, clustering based on 3D-shape and pharmacophore similarity, and manual expert selection of representatives (**Figure 23**). Out of the 1103 chosen compounds, 96 fragments were picked to construct a sub-selection named F2X-Entry Screen (**Figure 24A**). The F2X-Entry Screen shows a similar distribution of chemical properties to the F2X-Universal Library (**Figure 24B**).

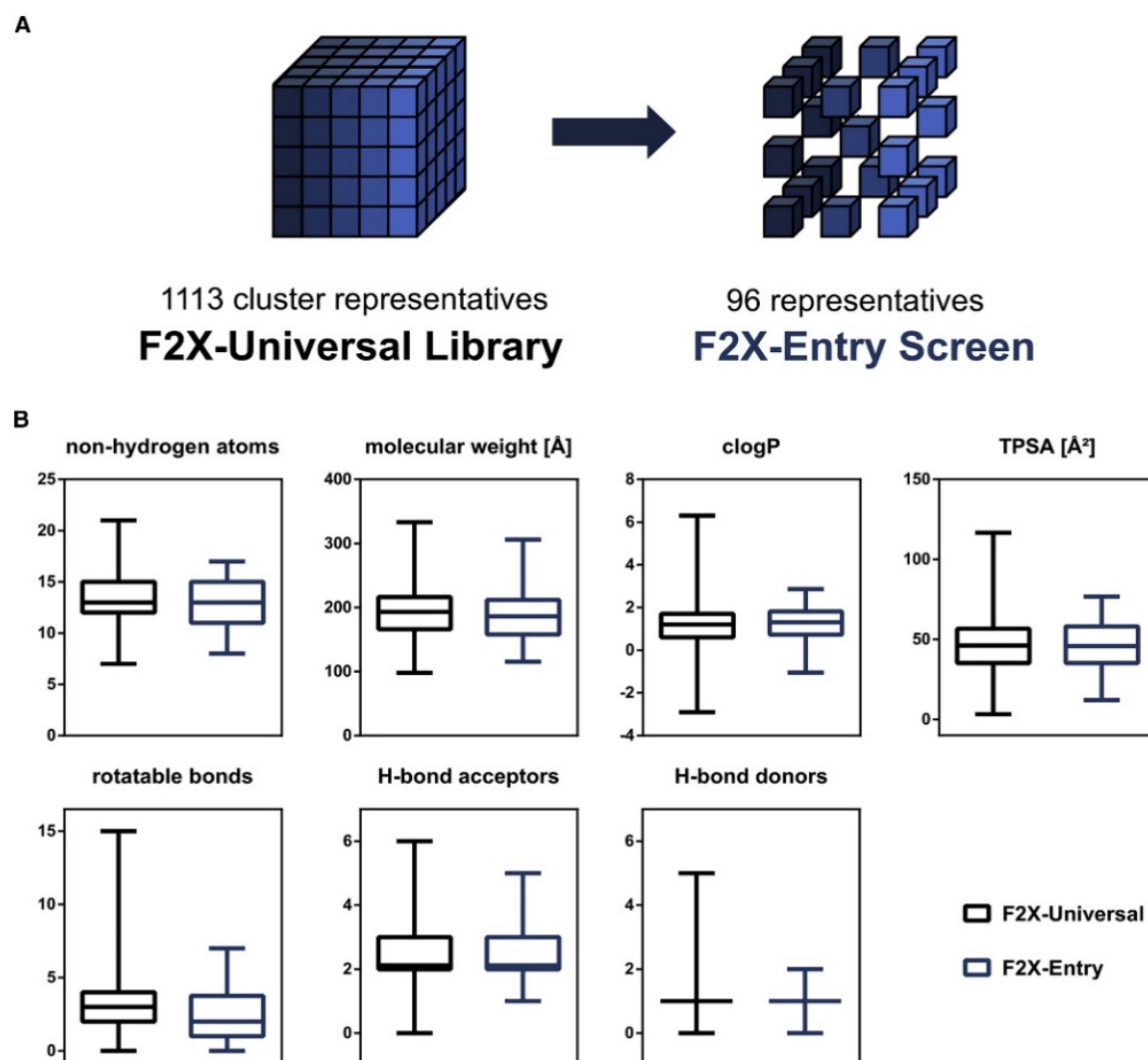


Figure 24: The chemical properties of the F2X-Universal Library and the F2X-Entry Screen. A) The relation between the F2X-Universal Library and the F2X-Entry Screen is depicted in a schematic. B) Seven standard chemical properties important for fragment libraries are shown. The distribution of the fragments contained in the F2X-Universal Library (black) and its sub-selection, the F2X-Entry Screen (blue), are shown as boxplots. Boxplots were prepared via Microsoft Office Excel. Figure adapted from Wollenhaupt & Metz *et al.*¹⁷³

The F2X-Entry Screen has been validated via two separate CFS campaigns. One campaign was performed against a model protein, the aspartic protease Endothiapepsin (EP). The other was performed against a spliceosomal protein-protein complex of Prp8^{RNaseH} and Aar2 (AR). Both crystal systems allowed collection of high-resolution and high-quality data based on the high-resolution cut-off, the ISa value¹⁸² and completeness (**Figure 25**). In both cases duplicates of the fragment soaks were prepared and at least 30 apo crystals (i.e. controls handled like the fragment soaks without the presence of a fragment). The analysis was performed via the before-mentioned FragMAXapp including the PanDDA hit identification program.

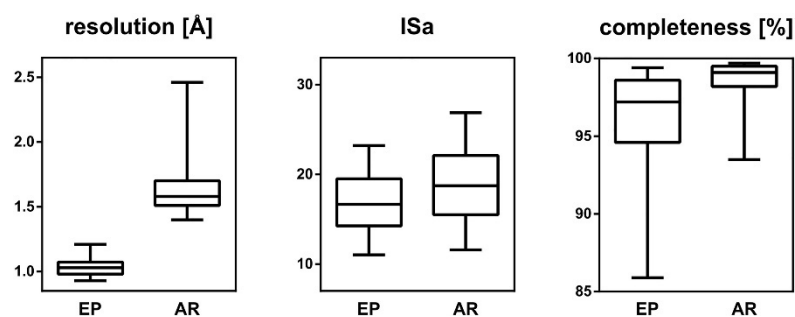


Figure 25: Data quality indicators of the collected crystallographic data. The distribution of the data across the resolution range, ISa values and completeness are given for both CFS campaigns as boxplots. Boxplots were prepared via Microsoft Office Excel. Figure adapted from Wollenhaupt & Metz *et al.*¹⁷³

In case of the EP campaign, 29 hits (37 binding events) were identified through various combinations of analysis software. This translates into a hit rate of 30%. The majority of these hits, 22 fragments (27 binding events), bound in the active site (**Figure 26**). Out of these 27 binding events, 8 bound in direct or water-mediated contact to the catalytic dyad, while the remaining fragments cover the different selectivity pockets. This is comparable to three other CFS campaigns performed with EP.^{45,63,96,183} Additional binding events were found in remote locations by 7 fragments. This includes three novel binding events that have not been observed before.

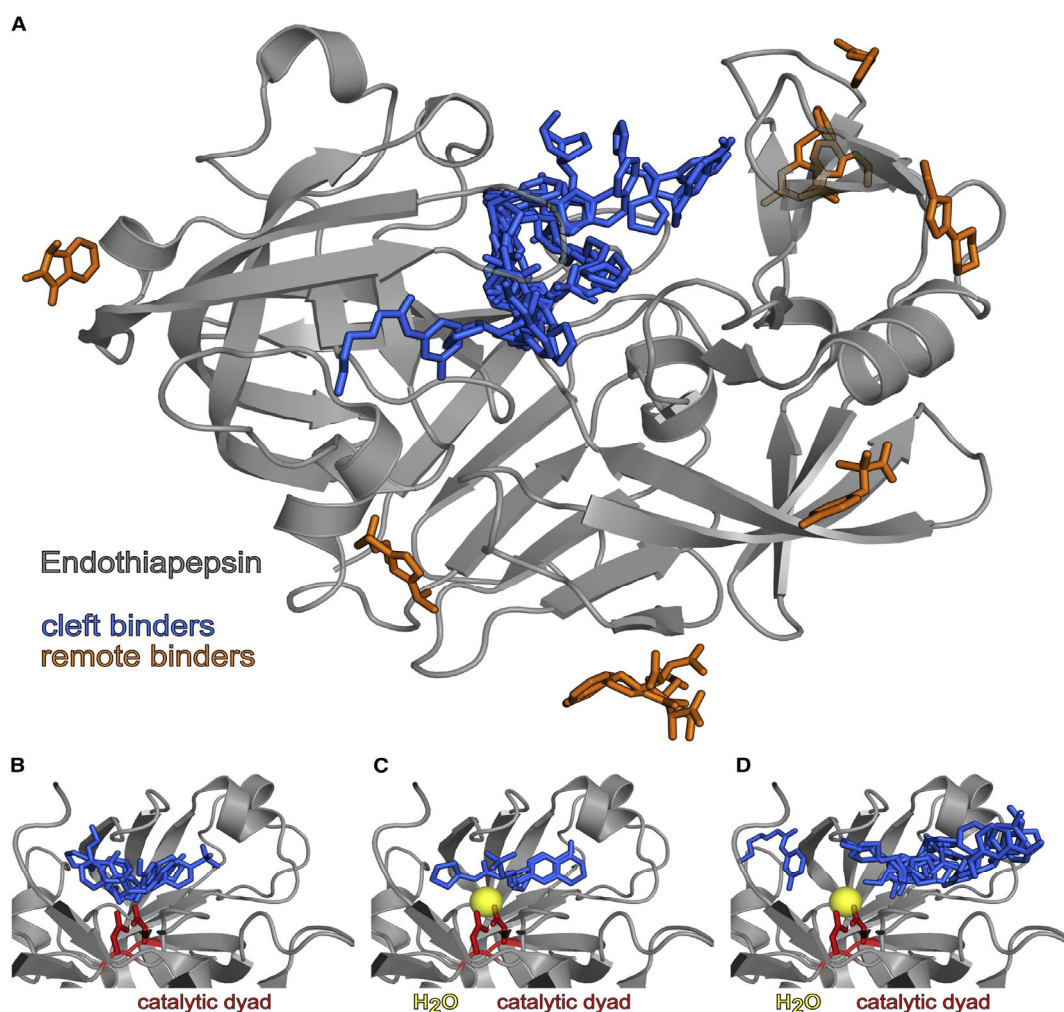


Figure 26: An illustration of the identified EP fragment hits. EP is shown as cartoon in light grey (PDB group deposition G_1002147), while the fragments are shown as sticks with blue colored fragments as the cleft binders and orange fragments for remote binders. The catalytic dyad amino acids are shown as sticks in red and the bound water molecule is shown as a yellow sphere. A) An overview of the fragment hits covering EP. B) The identified hits bound directly to the catalytic dyad. C) The identified hits bound via a water-mediated interaction to the catalytic dyad. D) Active site binders that do not engage in interactions with the catalytic dyad. The figure has been prepared via PyMOL.⁸⁵ Figure adapted from Wollenhaupt & Metz *et al.*¹⁷³

The CFS campaign with AR resulted in a similarly successful outcome. A hit rate of 21% was achieved with 20 fragment hits (23 binding events) (**Figure 27**). One of the identified hits bound directly at the interface of the two proteins. The other binding events are spread equally across the two protein surfaces. Out of the 20 fragments, 10 fragments (10 binding events) bound to Aar2, and 11 fragments (12 binding events) were observed on Prp8^{RNaseH}. Some binding sites of the identified fragment hits overlap with known PPI interfaces. Thus, different target sites can be chosen for further PPI modulator development.

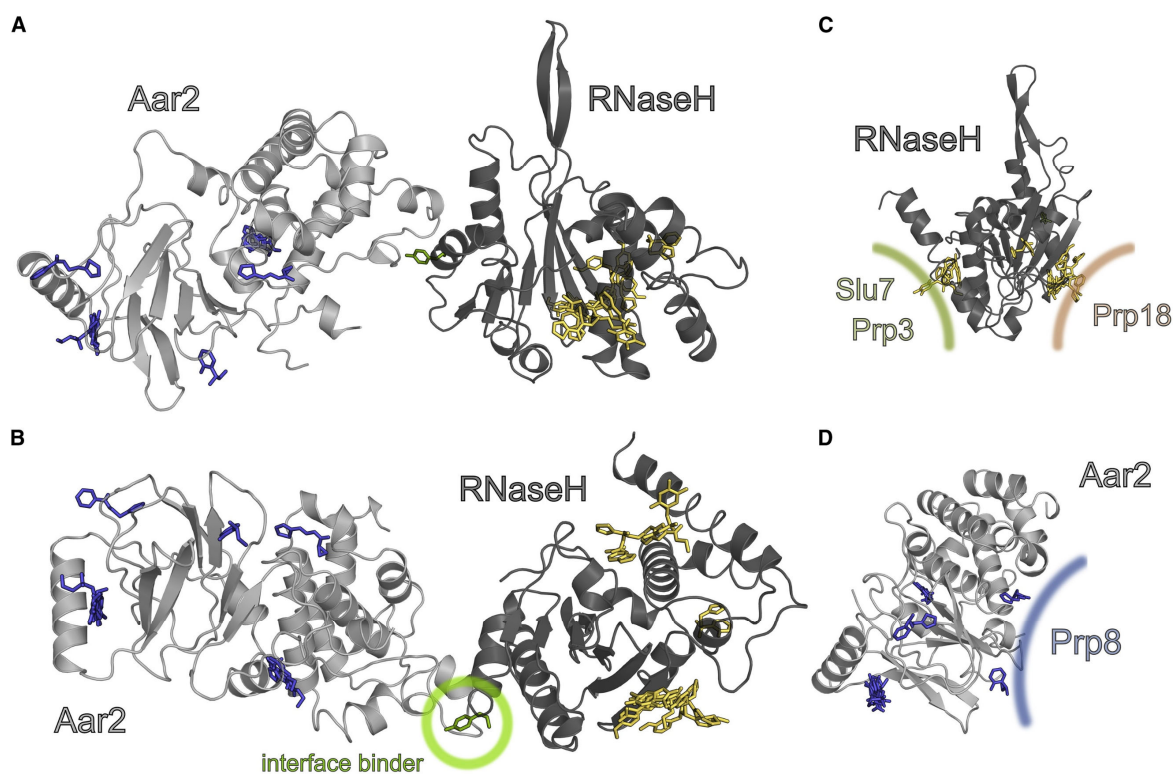


Figure 27: An illustration of the AR fragment hits. Both proteins are shown in cartoon, Aar2 in light grey and Prp8^{RNaseH} in dark grey (PDB group deposition G_1002115). The fragments are shown as sticks with yellow fragments bound to Prp8^{RNaseH}, blue fragments bound to Aar2, and a green fragment bound to the interface. A) The side view of the identified fragments bound to AR. B) A bottom view of the complex with the bound fragments. C) The Prp8^{RNaseH} domain with its bound fragments. Known interaction sites with other spliceosomal proteins are indicated via the colored lines. (PPIs based on PDB 6BK8 for Slu7;¹⁸⁴ PDB 3JCM for Prp3;¹⁸⁵ PDB 5WSG for Prp18¹⁸⁶. D) Aar2 with its bound fragments. Another interaction with Prp8 is indicated via a line (PDB 4I43).¹³⁸ The figure has been prepared via PyMOL.⁸⁵ Figure adapted from Wollenhaupt & Metz *et al.*¹⁷³

Both campaigns were performed with DMSO present in the soaking solution. This is usually a necessity for fragment soaking, as fragments are often dissolved in organic solvents such as DMSO. Therefore, to allow for high concentrations, a certain amount of DMSO containing the fragment needs to be present. However, protein crystals can disintegrate upon contact with DMSO, or their diffraction properties might deteriorate. The F2X-Universal Library and thus the F2X-Entry Screen as well, offer the possibility of soaking without DMSO. Due to the drying of the fragments onto a crystallization plate the DMSO is evaporated. The soaking solution can be prepared without DMSO. Soaking without DMSO has been tested with both targets. In both cases about 70% of the originally found fragment hits could be found without DMSO (**Figure 28**). However, it is important to note that new binding events could be identified without DMSO. Therefore, CFS campaigns without DMSO appear feasible and might result in different binding events and thus additional information for drug development.

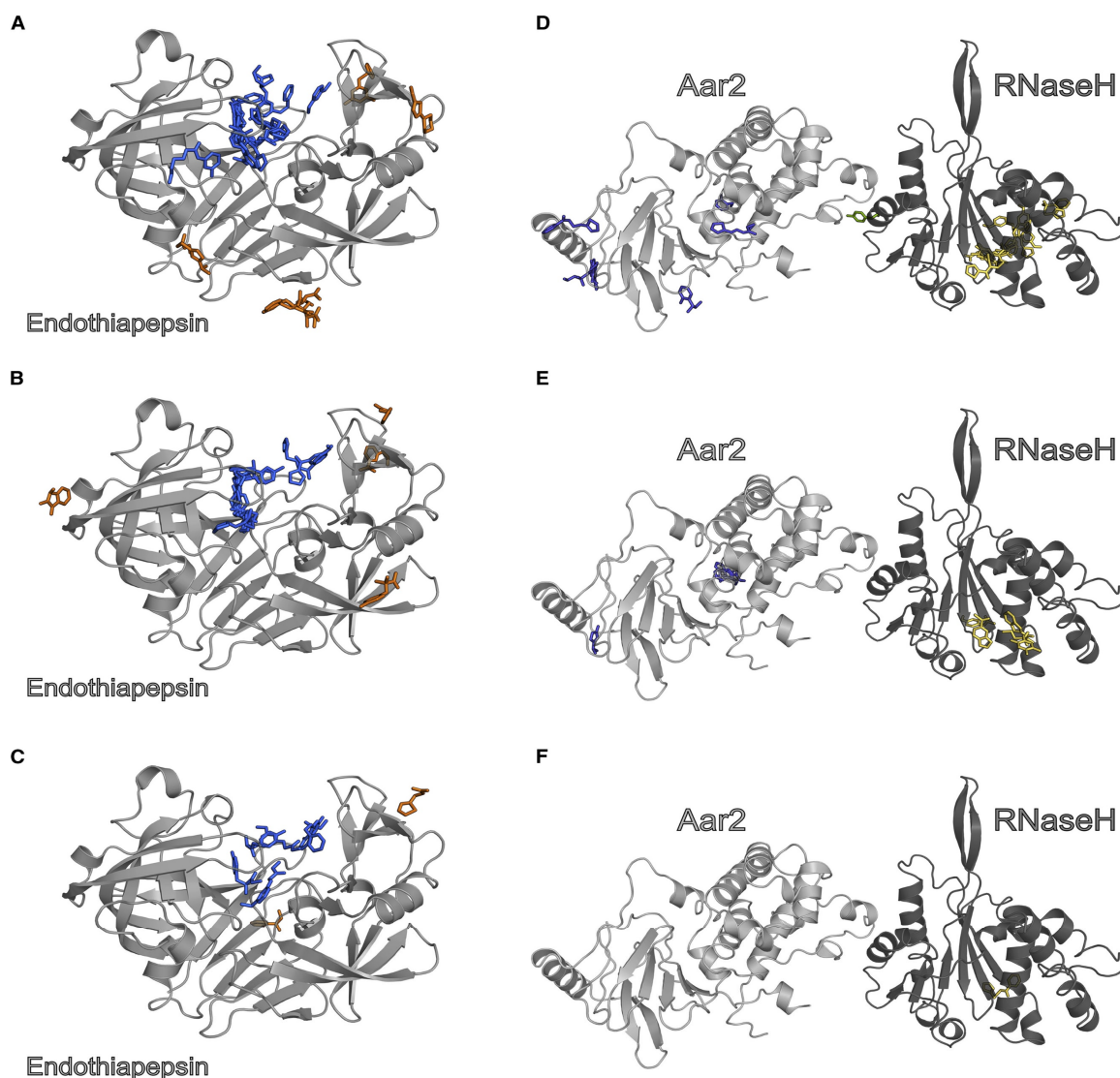


Figure 28: The comparison between soaking with DMSO and soaking without DMSO. On the left side EP is shown in light grey cartoon (PDB group deposition for with DMSO soaking G_1002147; PDB group deposition for DMSO-free soaking G_1002120) and on the right side Aar2 is shown in light grey cartoon and Prp8^{RNaseH} in dark grey cartoon (PDB group deposition for with DMSO soaking G_1002115; PDB group deposition for DMSO-free soaking G_1002119). Fragments are shown as sticks. A) All binding events of EP are shown that appear independent of DMSO presence. B) Binding events of EP that were only detected in presence of DMSO are depicted. C) Binding events of EP that were only identified without DMSO are shown. D) All binding events of AR that were detected regardless of DMSO presence are shown. E) Binding events that only appeared in presence of DMSO are depicted. F) Binding events that were identified only without DMSO are shown. The figure has been prepared via PyMOL.⁸⁵ Figure adapted from Wollenhaupt & Metz *et al.*¹⁷³

The F2X-Universal Library depicts a novel, structurally highly diverse fragment library, with a sub-selection that allows for screening of only 96 fragments. This subset named F2X-Entry Screen represents a similar distribution of chemical properties as the F2X-Universal Library and has been shown in this work to perform successfully. In two exemplary campaigns with a model protein and a more challenging PPI target, high hit rates of 20 – 30% were achieved. The identified binding events differ in their binding sites

allowing for various optimization pathways. Additionally, the specialized presentation of the library provides the possibility to omit organic solvents from the soaking condition. Thus, for example DMSO-sensitive targets can still be screened. Overall, the F2X-Entry Screen and F2X-Universal Library present valid library options for CFS campaigns.

3.4 Crystallographic fragment screening workflow at HZB



Screenshot taken from video from Wollenhaupt *et al.*

Publication:

Wollenhaupt, J.; **Barthel, T.**; Lima, G. M. A.; Metz, A.; Wallacher, D.; Jagudin, E.; Huschmann, F. U.; Hauß, T.; Feiler, C. G.; Gerlach, M.; Hellmig, M.; Förster, R.; Steffien, M.; Heine, A.; Klebe, G.; Mueller, U.; Weiss, M. S.

Workflow and Tools for Crystallographic Fragment Screening at the Helmholtz-Zentrum Berlin.

J. Vis. Exp., **2021** (169), 1–19.

<https://doi.org/10.3791/62208>.

Based on the previously presented developments, an optimized CFS workflow was established on site. This CFS workflow allows the user to perform a high-quality CFS campaign with minimal time and effort. However, transferring the knowledge and additional prerequisites needed for a CFS campaign to the experimenter turned out to be challenging. A simple checklist did not sufficiently prepare users for the experiment. Thus, a detailed protocol and a video presenting each step of the protocol were produced. In this way a user unfamiliar with this method is taught how to perform CFS campaigns and is presented with an example of a successful campaign.

Before the protocol starts, and thus before a campaign can start, the requirements for such an experiment are given. The crystal system should be examined for a favorable crystal lattice, reliable growth of crystals, the diffraction quality of those crystals and volatile chemicals in the crystallization condition. Additionally, the soaking condition is optimized, ideally to include DMSO and a higher concentration of a buffer to avoid pH-shifts, and to be cryoprotected. Assessing such points beforehand is vital for collecting high-quality data and thus for the success of a CFS campaign.

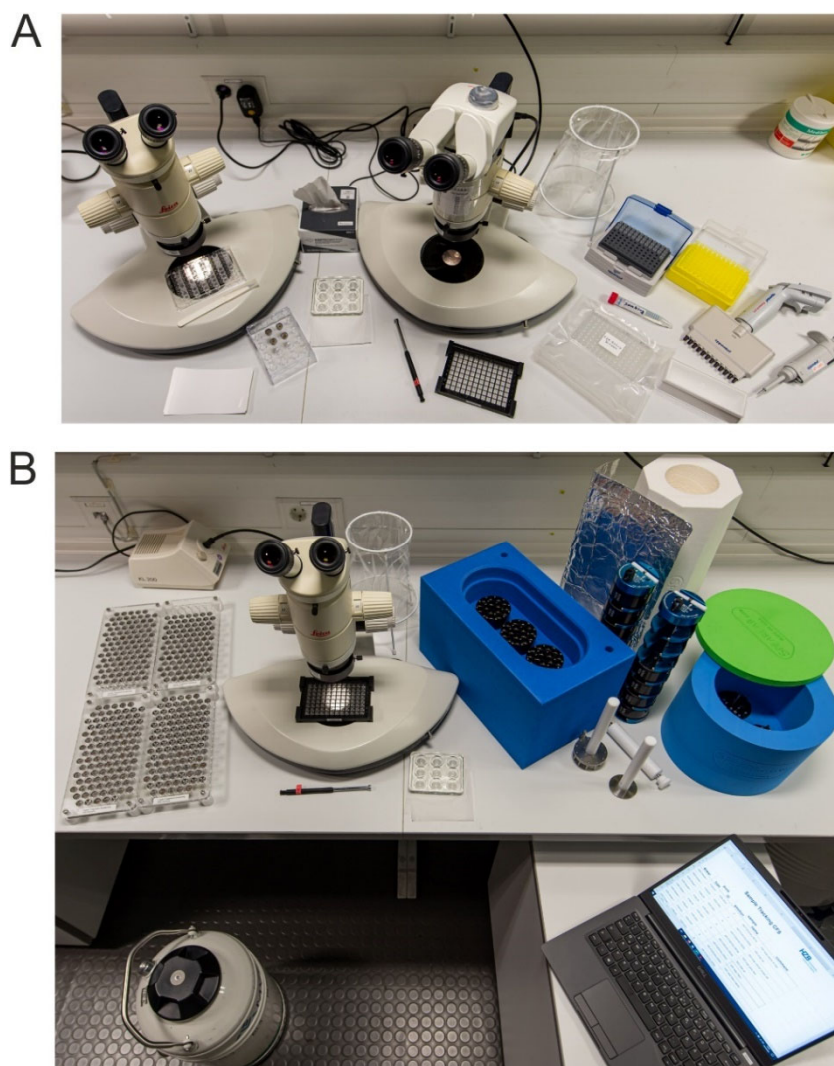


Figure 29: An exemplary experimental setup for CFS crystal soaking and harvesting. A) The setup for crystal soaking includes two microscopes to ease the transfer of crystals from one plate to the other. B) The setup for crystal harvesting includes a liquid nitrogen dewar and a laptop for note taking. Figure adapted from Wollenhaupt *et al.*¹⁸⁷

The first step of the CFS workflow at HZB refers to crystal soaking. The experimental setup is shown in **Figure 29**. The soaking plate, containing the frozen dried-on fragments needs to be prepared (**Section 3.3**). The plate is thawed, removed from its vacuum bag and the sealing foil is removed. The reservoir wells are filled with the soaking solution and the EasyAccess Frame is placed on the plate. Afterwards, a drop is pipetted onto the first lens with a dried-on fragment and two crystals are transferred to that drop. This is repeated for each fragment. Then, the EasyAccess Frame is removed, the plate is sealed with sealing foil and placed into an incubator overnight. Additionally, apo-crystals can be prepared for analysis purposes in a similar way. The same soaking solution is used, but no fragment is present while soaking. In this way hundreds of crystals can be soaked in a few hours.

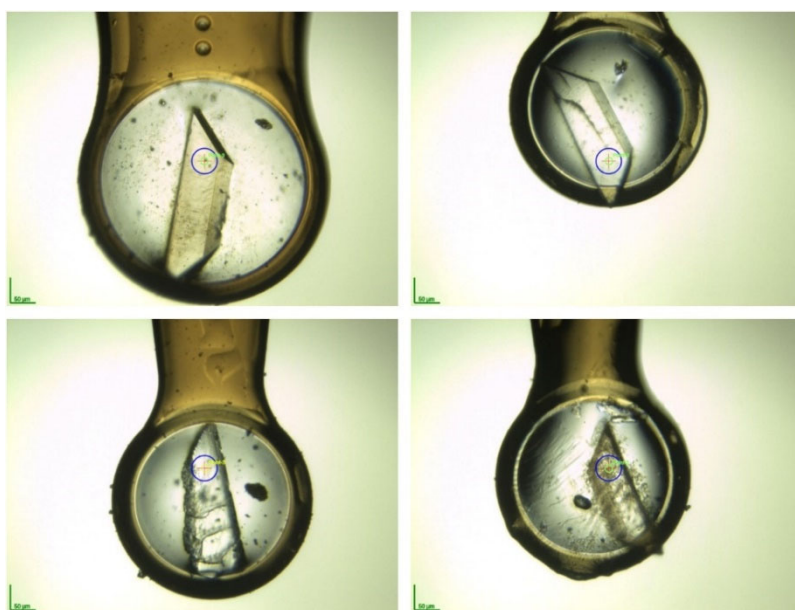
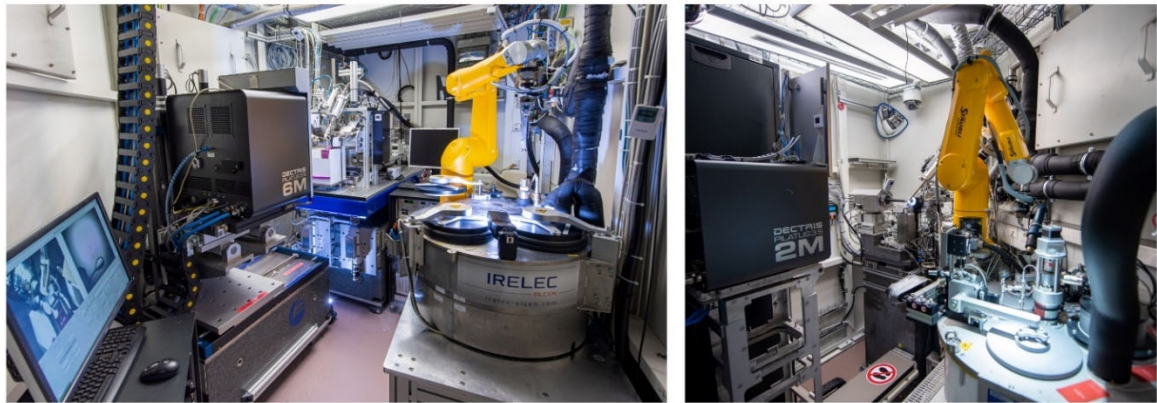


Figure 30: Depictions of possible protein crystal morphology after fragment soaking. Four harvested protein crystals are shown: one slightly colored, two cracked crystals and one with additional particles surrounding it. Figure adapted from Wollenhaupt *et al.*¹⁸⁷

The second step is crystal harvesting. The pucks are cooled down with liquid nitrogen and a laptop or sheet of paper is prepared to note down the position of each harvested crystal and comments throughout the process. The plate is taken from the incubator, the sealing foil is removed, and the EasyAccess Frame is again placed on top. From each soaking drop the crystals are harvested regardless of their appearance (**Figure 30**). In case a cryoprotection step is necessary, a cryo-protected solution is added to the second dried-on fragment lens. Each crystal is first transferred to the cryoprotection drop and then frozen in liquid nitrogen. If apo-crystals have been prepared, the same procedure is applied to these crystals. The filled pucks are stored in a storage dewar until the measurement.

A



B

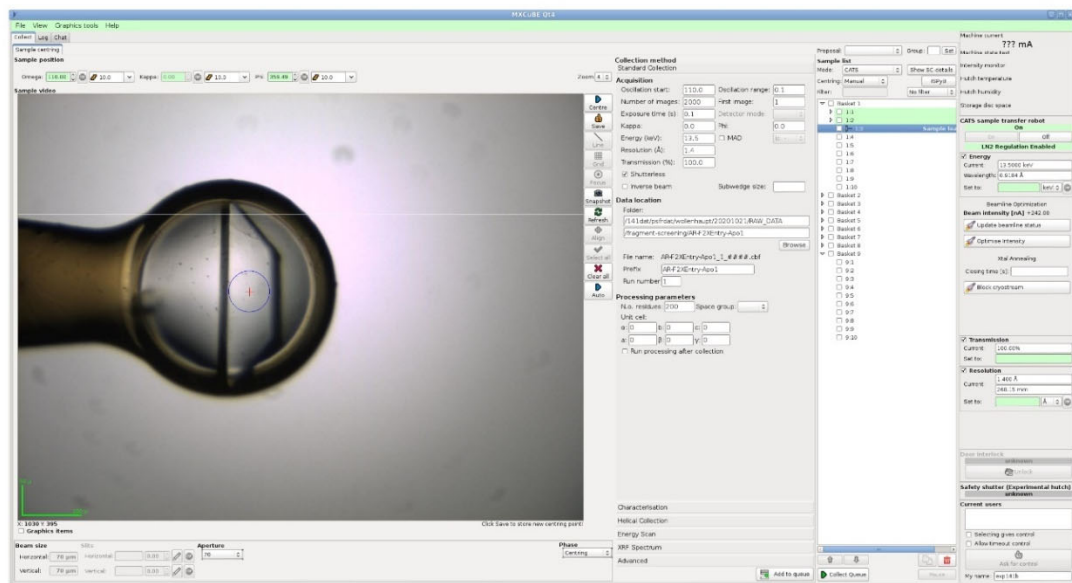


Figure 31: The beamline and data collection software are depicted. A) The beamlines BL14.1 and BL14.2 at BESSY II are shown in photographs taken inside the hutch.⁴ B) A screenshot of MXCuBE¹⁵⁸ is shown, which is used to control the machines at the respective beamline. Figure adapted from Wollenhaupt *et al.*¹⁸⁷

The third step, and the final experimental stage, is the data collection (**Figure 31**). In case of CFS campaigns, the focus is turned towards speed. Therefore, test images are avoided, and the detector distance is kept at a predefined distance, based on the higher end of previous measured apo-crystals of the same project. Ideally an optimized collection strategy, including rotation increments and total rotation has been established beforehand with crystals undergoing a mock treatment in soaking solution. This will allow collection of a full dataset with minimal time.

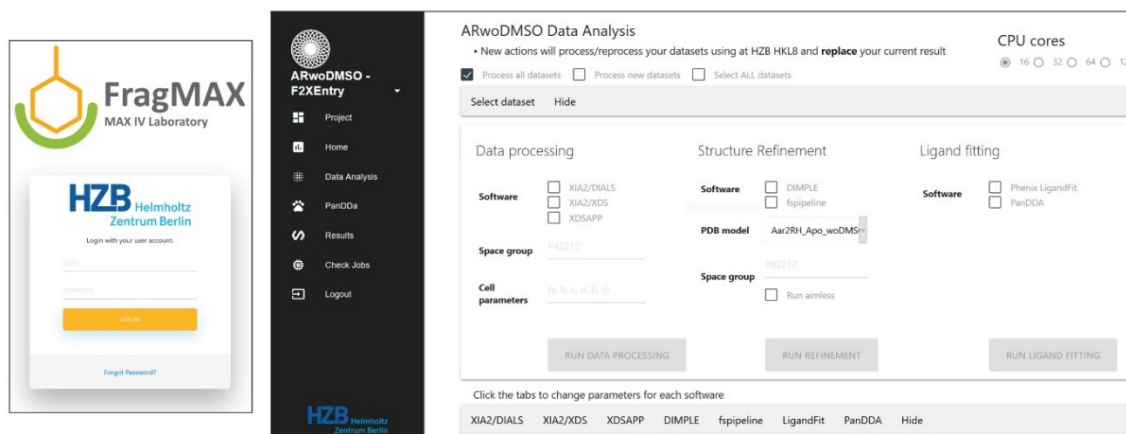


Figure 32: Screenshots of the FragMAXapp interface are depicted. On the left side the login screen is shown, and the right side presents the data analysis tab of the web application. Figure adapted from Wollenhaupt *et al.*¹⁸⁷

The last steps include auto-processing, auto-refinement and semi-automated hit identification. All of this is performed via FragMAXapp,¹⁵⁶ implemented at HZB (**Figure 32**). The collected data are loaded into the web application with the respective information about the fragment library. Processing of each dataset can be performed by several programs, which the user can choose from. After processing all datasets that allowed for it, they can be auto-refined again by several pipelines. At HZB we recommend choosing the combination of XDSAPP¹⁶⁵ and fspipeline⁶³ first and in case of a low hit rate it can be helpful to try all possible combinations of processing programs and refinement pipelines. The auto-refined datasets can be analyzed via PanDDA for hit identification. Here, the option is given to use known apo datasets for ground state characterization. The visual inspection of the PanDDA⁶⁴ results is the only step performed outside of FragMAXapp, using the pandda.inspect GUI in coot. After that, data are loaded back into the web application. In case fragment-containing datasets have been used for ground state characterization, it can be helpful to perform a second, improved run of PanDDA excluding the known fragment-bound datasets. Thus, the initial fragment hits are detected in a semi-automated way and the user can enter the first optimization phase.

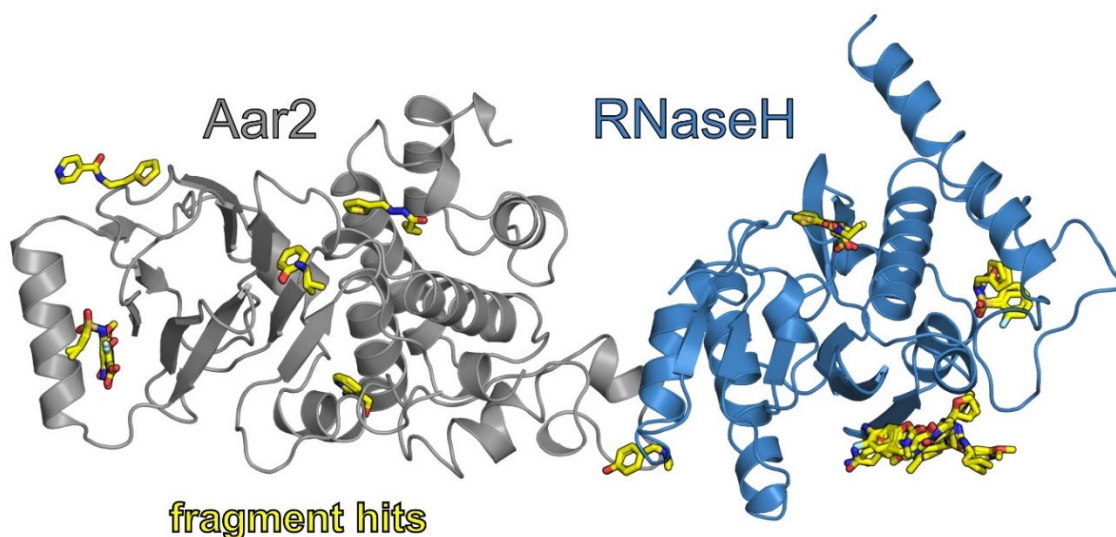


Figure 33: Overview of the identified fragment hits from the CFS campaign of the F2X-Entry Screen against the AR protein-protein complex (PDB group deposition G_1002119). The proteins are presented in cartoon, Aar2 in grey and Prp8^{RNaseH} in blue. The fragments are shown as sticks in yellow. The figure has been prepared via PyMOL.⁸⁵ Figure adapted from Wollenhaupt *et al.*¹⁸⁷

This workflow had been successfully applied in several campaigns, including a campaign with the protein-protein complex AR without DMSO in its soaking solution. This campaign demonstrated the possibility of performing a CFS campaign with the F2X-Entry Screen without DMSO. An important feature especially for protein crystals that do not tolerate organic solvents. However, in this campaign only the identified fragment hits from the previous DMSO-containing campaign were soaked again without DMSO. The possibility of new binders has not been explored. The identified fragment hits are shown in **Figure 33**. Even without DMSO during soaking, the campaign resulted in 15 binders.

In conclusion, the CFS workflow applied at HZB has been presented in a detailed step-by-step explanation. An 8-minute-long video constitutes the heart of the publication. As of March 2023, the video has already been viewed almost 4000 times. The description of the workflow allows for users to understand the requirements for a CFS campaign and reproduce the optimized steps and perform high-quality CFS experiments.

3.5 Advancing fragment hits via SAR by catalogue – the Frag4Lead workflow

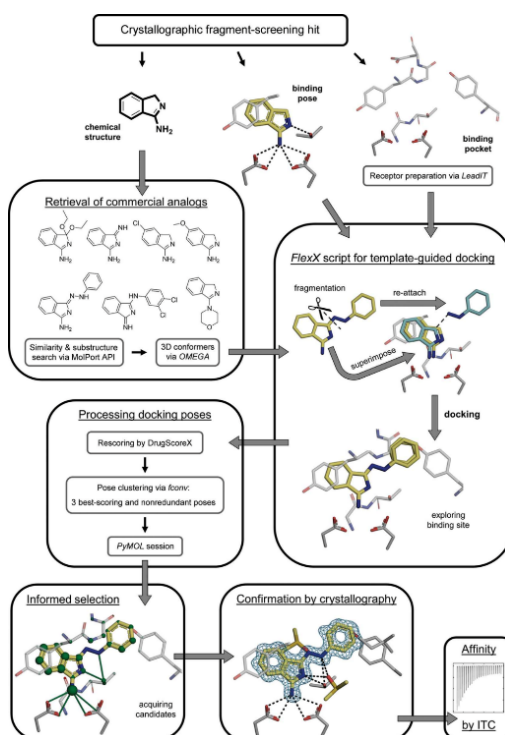


Figure taken from Metz & Wollenhaupt *et al.*

Publication:

Metz, A.; Wollenhaupt, J.; Glöckner, S.; Messini, N.; Huber, S.; **Barthel, T.**; Merabet, A.; Gerber, H. D.; Heine, A.; Klebe, G.; Weiss, M. S.

Frag4Lead: Growing Crystallographic Fragment Hits by Catalog Using Fragment-Guided Template Docking.

Acta Crystallogr. Sect. D Struct. Biol. **2021**, 77 (9), 1168–1182.

<https://doi.org/10.1107/S2059798321008196>

A CFS campaign results in the identification of promising fragment hits. However, the development of tool compounds or drugs only starts with this screening step. Afterwards, the fragment hits are optimized into higher affinity binders. Such optimization often includes collaborations with medicinal chemists to synthesize larger compounds based on the fragment hit. It can be challenging to find collaboration partners and the specifically tailored syntheses are often time-consuming and expensive. Another approach at the beginning of optimization can be SAR by catalogue.^{188–190} Commercially available chemical spaces are searched for promising compounds, that either present an exchanged substituent on the fragment or that grow along a certain exit vector. This method allows the finding of higher affinity binders without the involvement of expensive chemical synthesis. To support users at BESSY II, a computational workflow has been built which provides such a SAR by catalogue approach for users.

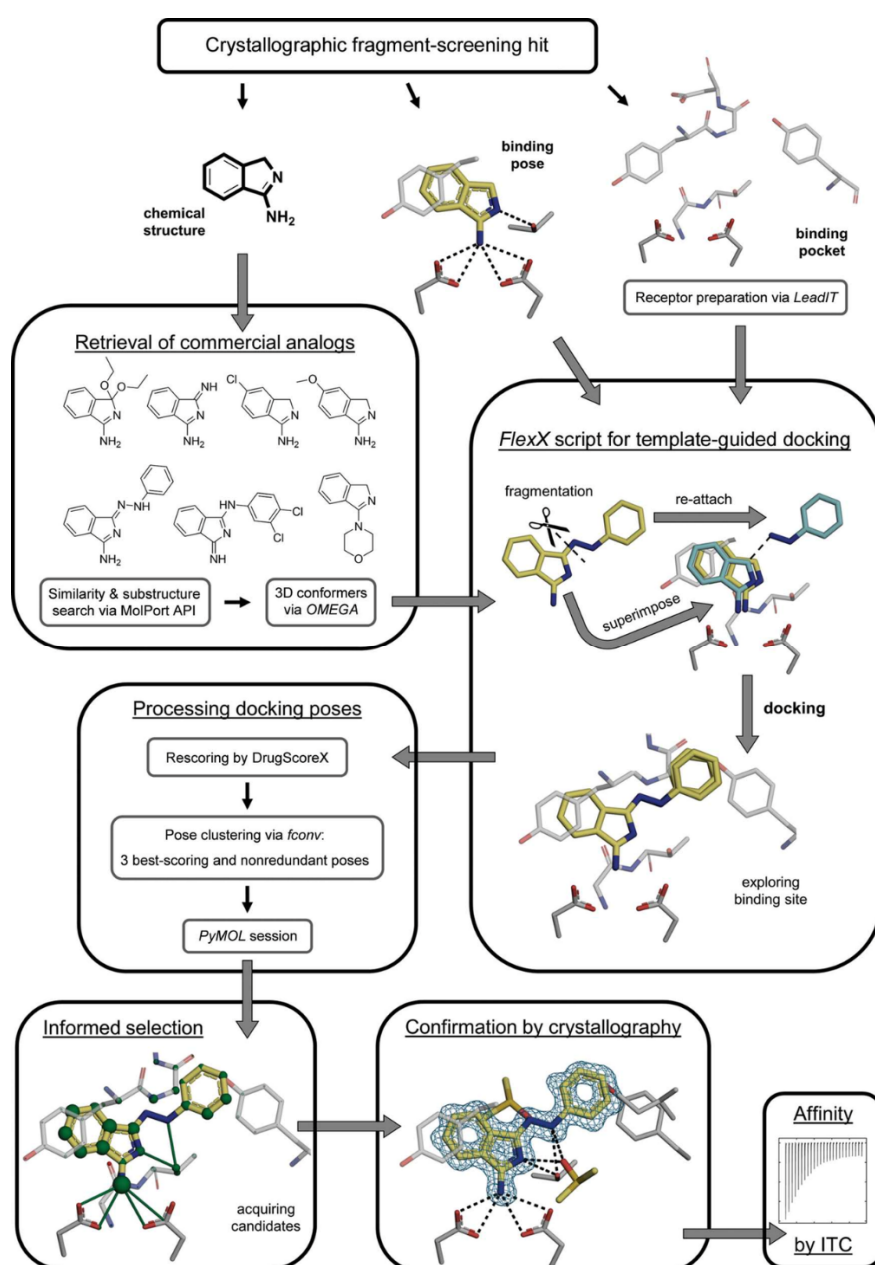

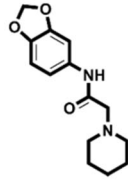
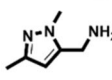
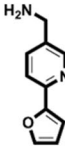
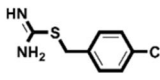


Figure 34: Schematic overview of the Frag4Lead workflow. The different steps are presented through boxes and an example fragment hit is used to illustrate the workflow. Figure adapted from Metz & Wollenhaupt *et al.*⁸⁸

The workflow was termed Frag4Lead and is implemented via KNIME¹⁹¹ (**Figure 34**). The Frag4Lead workflow starts with the CFS hit given as a fragment-bound protein structure pdb-file. The fragment structure is extracted from the complex structure and used to search for analogs. Similarity and substructure searches can be performed to search the commercial vendor MolPort.¹⁹² The found analogs are then prepared as 3D structures with the OMEGA application.¹⁹³ Based on the complex structure, the binding site is defined via LeadIT. The defined binding site, the commercial analogs and the binding pose information is combined during template-based docking via FlexX.⁸² The docked poses are re-scored by the DrugScoreX software⁸³ and filtered to exclude docking poses with bad scores. Afterwards, fconv¹⁹⁴ is used to cluster the poses to only keep the three highest ranked but still most dissimilar poses per compound. The resulting poses are inspected manually via a PyMOL session.⁸⁵ Favorable interactions are visualized with green spheres and unfavorable interactions with red spheres, offering in this way an expert selection of follow-up compounds. Finally, the follow-up compounds are tested again for binding via crystallography and biophysical affinity determination, e.g. via isothermal titration calorimetry (ITC), and successful binders can be used for a second round of optimization.

Table 3: The five initial fragment hits are presented with their 2D structure, K_D value, LE value, number of identified follow-up compounds and number of successfully docked follow-up compounds. The respective PDB code is given for each fragment-bound structure. Table adapted from Metz & Wollenhaupt *et al.*⁸⁸

Fragment	PDB code	Chemical structure	K_d (μM)	LE (kcal mol^{-1} per atom)	No. of identified follow-up candidates	No. of successfully docked follow-up candidates
F005	4y3e		1700	0.38	556	67
F041	4y3z		900	0.22	1013	88
F058	4y56		8800	0.31	10022	>1000†
F066	5dq4		400	0.35	615	395
F290	4y35		100	0.45	267	32

† Only the 1000 highest-ranking fragment analogs were considered.

To present the successful application of this workflow, previously known crystallographic fragment hits on the model protein EP were used as starting points.^{63,96,183} Five different starting points were picked in a way to avoid cherry-picking and, instead, imitating a typical CFS campaign on a more challenging target with only a few promising hits as the outcome (**Table 3**). Each fragment hit progressed through the computational workflow with a varied number of successfully docked analogs. These were then checked visually to narrow the selection down to around 25-30 final follow-up compounds, which is a realistic purchase in the context of a typical academic project.

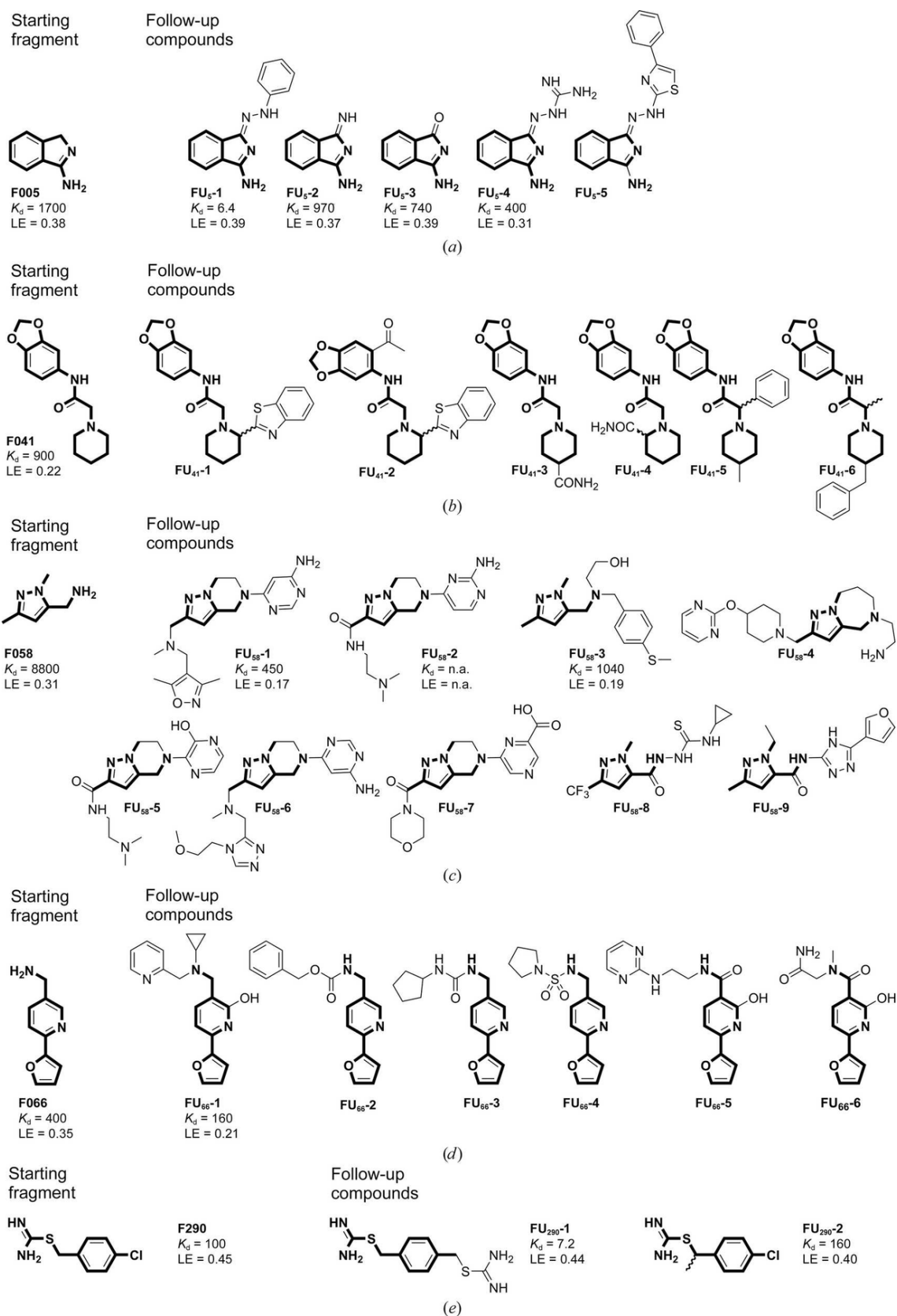


Figure 35: Overview of the purchased follow-up compounds. The starting fragment structure is shown on the left side and the substructure is highlighted in bold lines throughout the follow-up compound structures. The code of each follow-up compound is written underneath or next to the respective 2D structure. Additionally, the K_D derived from ITC measurements and the calculated LE values are given for the bound follow-up compounds. Figure adapted from Metz & Wollenhaupt *et al.*⁸⁸

In total 28 compounds were purchased (**Figure 35**). The follow-up compounds included similar analogs with marginal changes for hit validation and some with larger changes applying the growing optimization strategy. Each compound was tested in crystallographic experiments by soaking EP crystals in solutions containing an excess of compound before flash-freezing. After subsequent data collection and automated data treatment using the aforementioned tools, ten of these compounds could be visualized in the electron density. Nine follow-up compounds were also investigated via ITC. One could not be used for ITC measurements as not enough material was available. ITC delivers information about the affinity (K_D) and the LE, which are typical indicators monitored during optimization. Most of the follow-up compounds showed an improved binding affinity. There was one fragment that did not result in a higher affinity binder (FU₂₉₀₋₂). The highest binding affinity reached in this study was measured from FU₅₋₁ with 6.4 μ M, i.e. a 266-fold improvement of the compared K_D . Taken together, 33 % of the purchased follow-ups were identified as binders and delivered new information for further iterative compound development.

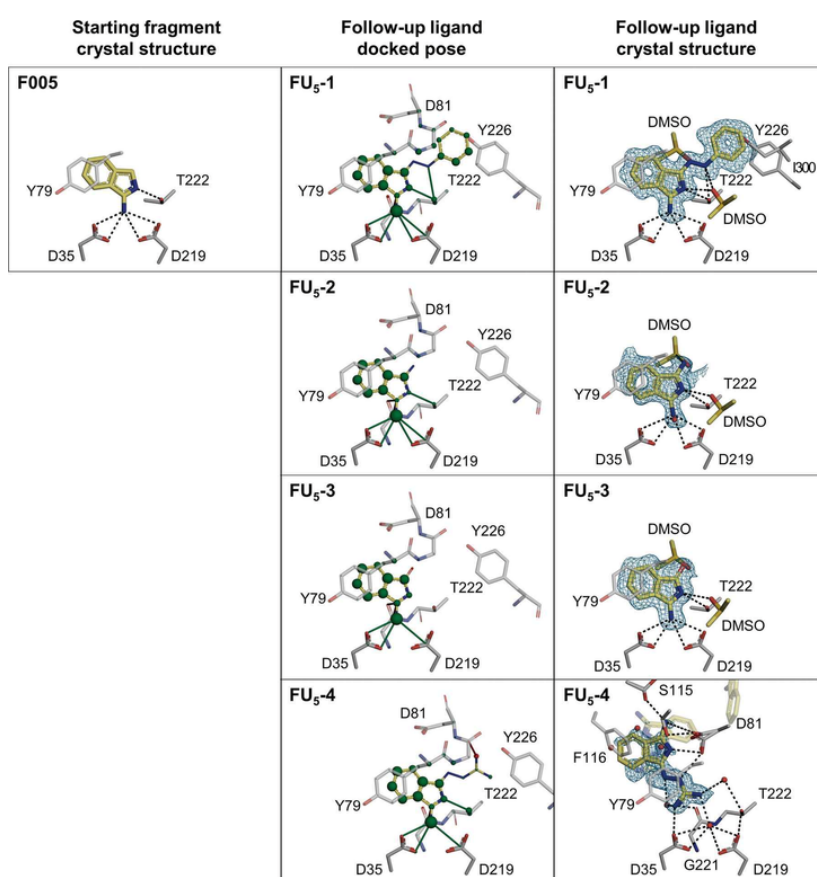
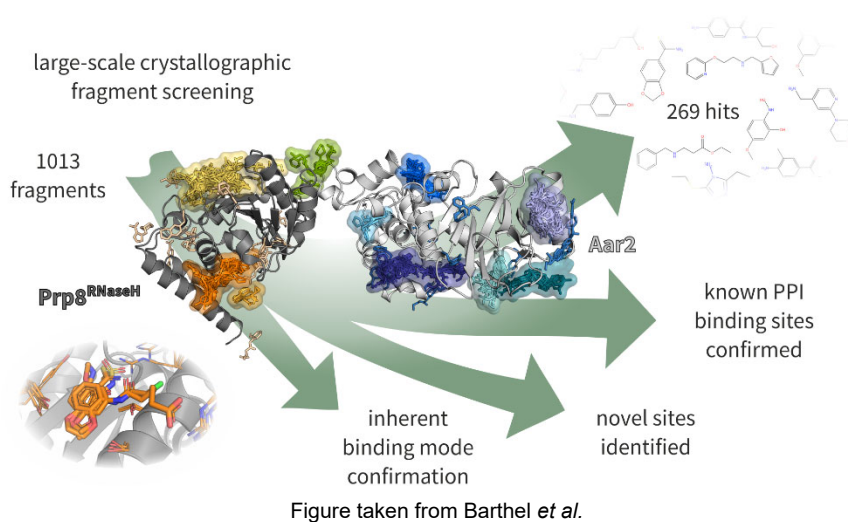


Figure 36: Comparison of the docking pose and crystal structure of the F005 follow-up compounds. The compounds are presented as yellow sticks and the protein binding pocket is presented as grey sticks. Left the crystal structure is shown of the initial fragment hit. In the middle column the docked poses are shown with green spheres to indicate favorable binding energies for each atom and green lines to illustrate favorable contact distances. Red lines indicate unfavorable contact distances. On the right side the crystal structure is shown with the respective polder OMIT mF_o-DF_c electron density maps¹⁹⁵ contoured at $\sigma = 3$. Water molecules are shown as red spheres and polar interactions are shown as dashed lines. Figure adapted from Metz & Wollenhaupt *et al.*⁸⁸

Out of the ten found follow-up compounds, five retained their expected binding pose. One example is given by fragment F005 (**Figure 36**). Four of the five purchased follow-up compounds were identified as binders. Two of these, FU₅-2 and FU₅-3, add one additional heavy atom to the starting fragment and bind as expected. The remaining two, FU₅-1 and FU₅-4, present larger additions to the original fragment hit. FU₅-1 shows the expected binding pose, while FU₅-4 was flipped and binds to the catalytic dyad with its aminoguanidine moiety instead of the isoindole moiety. Similar binding of a guanidinium group to the catalytic dyad had been observed in another screening.¹⁸³ The presented example shows that such changes in the binding mode upon experimental validation occur, but also that maintained binding poses with higher affinity are possible.

Overall, the application of the Frag4Lead workflow resulted in the successful identification of ten of the purchased follow-up compounds bound in the active site. Most of them bound with higher affinity than the original fragment, and half of the compounds retained the expected binding pose. This workflow allowed for fast optimization at the beginning of the development of lead structures, without involving custom synthesis. Therefore, the Frag4Lead workflow provides a platform for users to optimize their fragment hits in an easy, fast, and low-cost fashion.

3.6 Inherent binding pose validation and discovery of putative novel PPI sites by CFS



Publication:

Barthel, T.; Wollenhaupt, J.; Lima, G. M. A.; Wahl, M. C.; Weiss, M. S.

Large-Scale Crystallographic Fragment Screening Expedites Compound Optimization and Identifies Putative Protein-Protein Interaction Sites.

J. Med. Chem. **2022**, 2022, 14630–14641.

<https://doi.org/10.1021/acs.jmedchem.2c01165>

Based on the aforementioned developments to support CFS campaigns at HZB, a large-scale CFS was performed against the AR target. AR is part of the pre-mature U5 snRNP and the dissociation of the complex is necessary for formation of the mature U5 snRNP and subsequent splicing. Screening about 1000 fragments against such a target presents the opportunity to investigate two hypotheses: I) A large number of fragment hits will be identified that may potentially repeat binding modes. This bypasses the typical step of binding mode confirmation performed after hit identification. II) Bound fragments will cluster at biologically relevant PPI sites. Thus, fragment clusters hint towards unknown functional binding sites of the target protein. These results offer new insights into the protein target and speed up compound optimization.

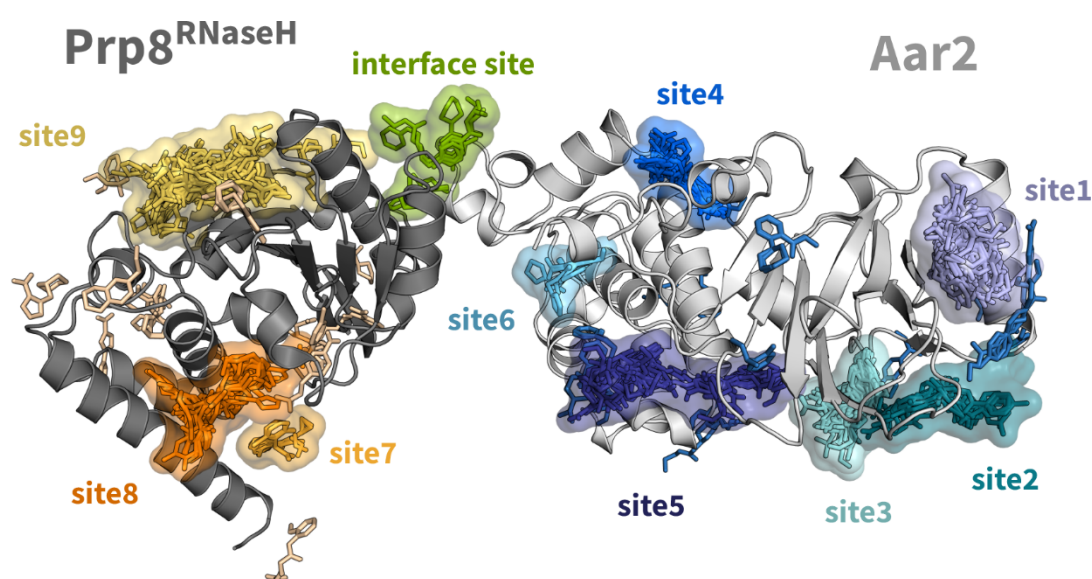


Figure 37: Overview of all identified fragment hits bound to the AR complex. The proteins are shown in cartoon presentation, Prp8^{RNaseH} in dark grey and Aar2 in light grey (PDB group deposition G_1002241). The fragments are depicted as sticks colored differently regarding the protein they bind to and their fragment cluster. Fragments bound to Aar2 are shown in blue colors. Fragments bound to Prp8^{RNaseH} are shown in yellow colors. The interface binders are shown in green. Binding sites are indicated via the combined surface view of the bound fragments shown as clouds. The figure has been prepared via PyMOL.⁸⁵ Figure adapted from Barthel *et al.*¹⁷⁸

The CFS campaign resulted in 249 identified fragment hits that cover both protein surfaces. The analysis was based on the FragMAXapp with an additional clustering step via cluster4x.¹⁹⁶ The program cluster4x offers the possibility for the experimenter to group datasets based on their similarity in amplitudes and C_α position. By analyzing such clusters via PanDDA the ground state model can be improved. This approach allowed identification of 22 additional fragment hits and shows the validity of clustering large datasets for certain analysis steps. The identified fragments cluster at different binding sites, referred to as site 1 – 9 and the interface site. Taken together with the previously found 20 hits of the

F2X-Entry Screen, a total of 269 fragment hits resulted from the complete F2X-Universal Library (**Figure 37**). This translates into a hit rate of 27%. Due to the screening of the complete F2X-Universal Library additional binding sites were identified compared to solely screening the F2X-Entry Screen (**Figure 38**). Furthermore, binding sites that originally presented only one binder are covered by several fragments of the F2X-Universal Library.

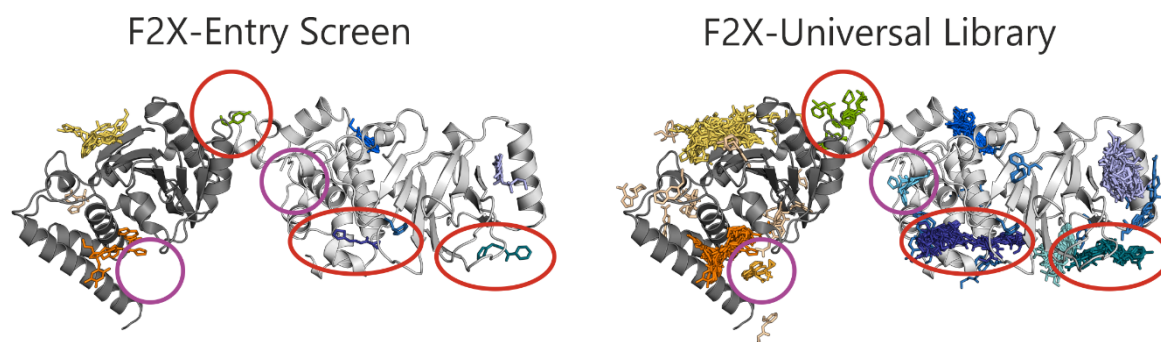


Figure 38: Comparison of F2X-Entry Screen hits (PDB group deposition G_1002115) and F2X-Universal Library hits (PDB group deposition G_1002241). The proteins are depicted as cartoons, Prp8^{RNaseH} in dark grey, Aar2 in light grey. Fragments are shown as sticks, colored depending on their binding site. Purple circles point towards newfound binding sites by the F2X-Universal Library. Red circles show binding sites that presented one fragment from the F2X-Entry Screen and is now highly populated by several fragments from the F2X-Universal Library. The figure has been prepared via PyMOL.⁸⁵ Figure adapted from Barthel *et al.*¹⁷⁸

The chemical properties of the binders have been compared with the overall distribution of these properties throughout the F2X-Universal Library (**Figure 39**). This is important as strong trends towards certain properties could point out limitations of this library or preferences of the AR target. Overall, the distribution of the fragments present in the F2X-Universal Library and the bound fragments is very similar. Therefore, the library offers a fitting range of diverse fragments for this target.

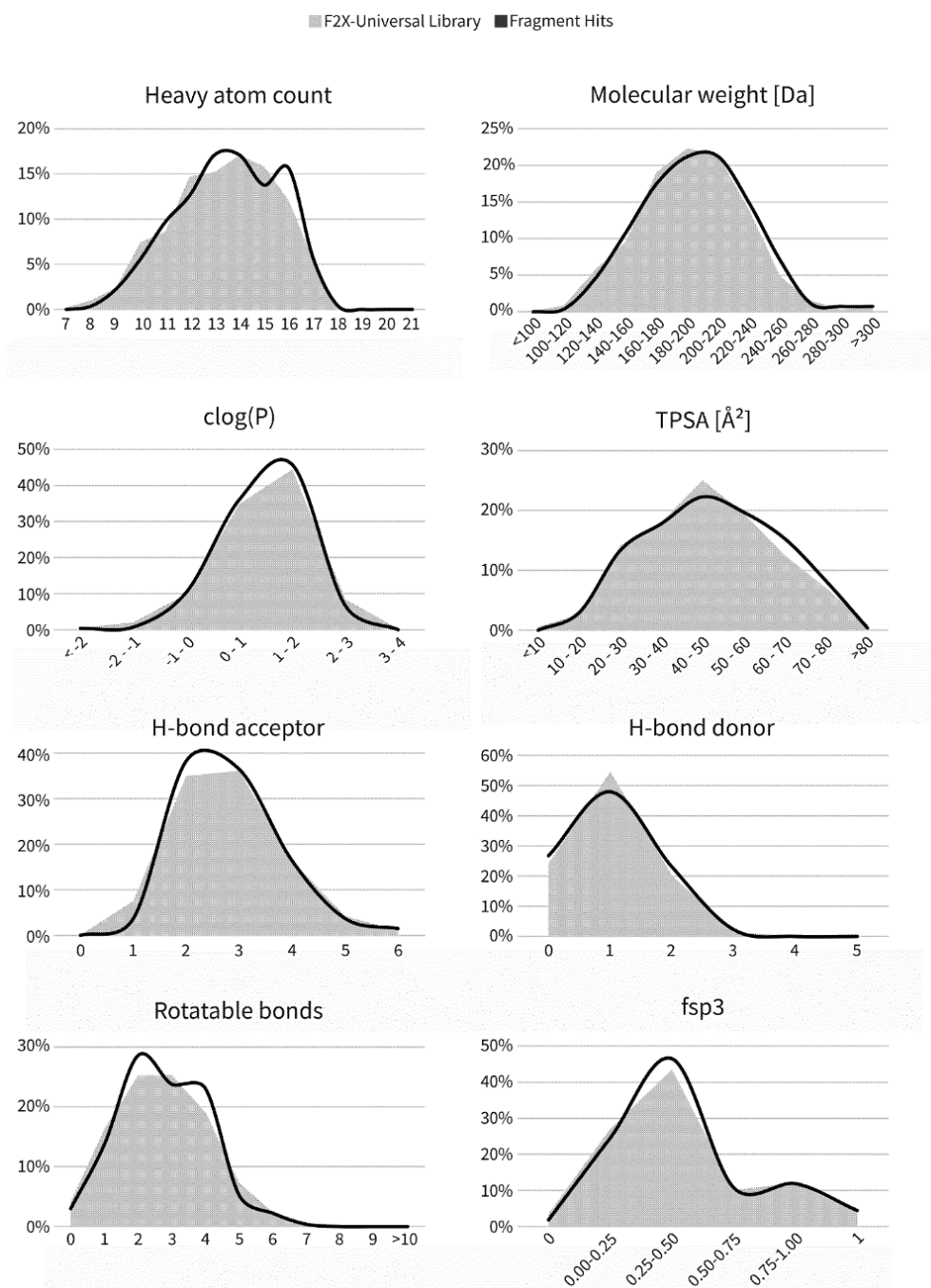


Figure 39: Comparison of chemical properties between the complete F2X-Universal Library fragments (grey area) and the found binders (black line). The curves have been prepared via Microsoft Office Excel. Figure adapted from Barthel *et al.*¹⁷⁸

The screening of the complete F2X-Universal Library presented additional binders at sites that were only occupied by one or two fragments before. A closer look at such sites, for example the interface site or site 8, display the expected binding mode confirmation. At the interface site a fragment had been found to displace Arg186, which is positioned outside its usual binding pocket (**Figure 40**). This replacement of the Arg186 side chain by a ring structure has been repeated by five additional fragments. The presented protein-fragment interactions are similar between the different fragments. A similar observation has been made for site 8. This shows how a large-scale CFS can speed up

compound optimization by avoiding additional binding mode confirmation subsequent to the screening analysis.

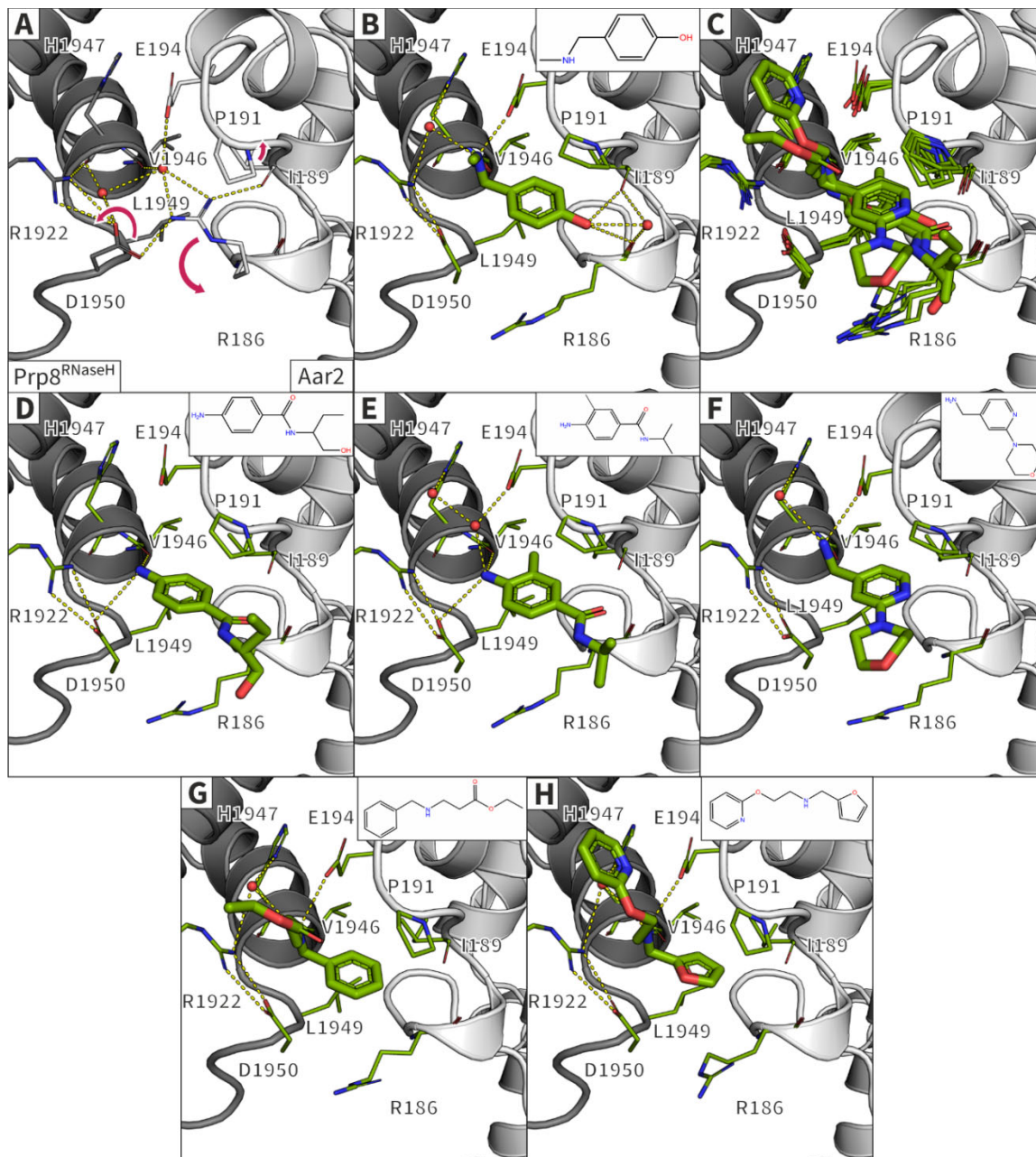


Figure 40: Illustration of binding mode confirmation at the AR interface site. The proteins are shown as cartoon, Prp8^{RNaseH} in dark grey, Aar2 in light grey (inhouse refined apo structure). Important amino acid side chains of the proteins are shown as sticks. Water molecules engaged in interactions with the protein and/or fragment are shown as red spheres. Hydrogen bonds are shown as yellow dashed lines. The fragments are shown as green sticks. For each fragment a 2D structure is given in the right upper corner. A) The unbound state of the protein complex is depicted with red arrows pointing out the movement of the amino acid chains upon binding of the fragments. B) The F2X-Entry Screen hit fragment shown with its binding mode (PDB ID 5QYG). C) An overlay of all identified fragments bound similarly to the interface site. D-H) Additional F2X-Universal Library fragment hits bound in a similar fashion as the F2X-Entry Screen fragment hit (PDB IDs 7FNH, 7FN1, 7FN2, 7FMK, 7FML). The figure has been prepared via PyMOL.⁸⁵ Figure adapted from Barthel *et al.*¹⁷⁸

To address the second hypothesis, the identified binding sites of AR were compared to known PPIs throughout the splicing cycle (**Figure 41**). It became evident that several biologically functional sites overlap with the fragment clusters. Three known PPI sites were corroborated, another Prp8-Aar2 PPI, the interaction site of Prp18 with Prp8^{RNaseH}, and the interaction site of Prp3 and Slu7 with Prp8^{RNaseH}. Due to the well-fitting overlap of the known PPI sites and the fragment clusters, it can be hypothesized that the other fragment clusters hint at unknown biologically relevant PPI sites.

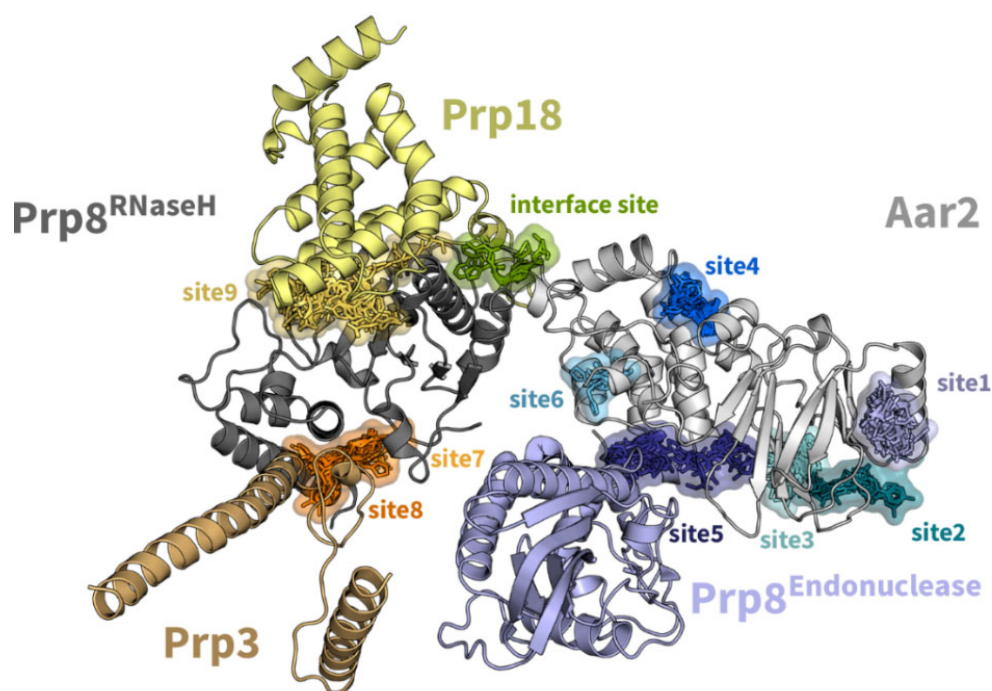


Figure 41: Overlapping fragment clusters with known PPI sites. Proteins are shown in cartoon, Prp8^{RNaseH} in dark grey, Aar2 in light grey (inhouse refined apo structure), Prp18 in yellow (PDB ID 6BK8),¹⁹⁷ Prp3 in brown (PDB ID 3JCM),¹⁸⁵ Prp8^{Endonuclease} in violet (PDB ID 4I43).¹³⁸ Identified fragments are shown as sticks in different colors depending on their binding site. Binding sites are depicted as respectively colored clouds. The figure has been prepared via PyMOL.⁸⁵ Figure adapted from Barthel *et al.*¹⁷⁸

The large-scale CFS campaign performed with the F2X-Universal Library against the spliceosomal protein-protein complex AR resulted in information-rich results. Hundreds of starting points were identified for a multitude of subsequent compound optimization pathways to develop PPI enhancers or inhibitors. Additionally, binding modes were confirmed by the overlapping of several fragments at one binding site engaging in similar interactions. Furthermore, fragment clusters point towards potentially biological relevant PPIs. Thus, large-scale CFS campaigns offer more than the number of starting points for drug or tool compound developments.

4. Discussion

In the last decade, CFS has evolved to an attractive method in drug and tool compound discovery. Simultaneously, the number of performed CFS campaigns has risen and is rising still. Prior to that, the data collection time per crystal was in the range of several minutes, which prevented high-throughput screening of crystals and thus the establishment of X-ray crystallography as an efficient screening method. However, through the last years synchrotron facilities have found ways to increase their photon flux and provide robot-assisted crystal mounting. As the data collection step ceased to be an issue in CFS campaigns, it became apparent that several other steps of the workflow require further improvements to support experimenters. Sample and data handling need to be simplified and sped up. Efficient fragment libraries are needed, that are available for users at the synchrotron site. Hit rates should be increased to broaden the pool of starting points for optimization. Additionally, ways to utilize found fragment hits in a cost-effective way need to be developed, especially to support academic groups with less resources than pharmaceutical companies. The step of optimization requires possibilities to derive a higher affinity binder without time-consuming and costly syntheses. Furthermore, fragment binding might lead to conclusions beyond tool compound or drug development. Fragments could for example uncover unknown protein functions. The presented work here addressed the mentioned points for improvement and resulted in several ways to support CFS campaigns.

4.1 Reaching high hit rates through efficient libraries, suitable crystal systems and multiplex analysis

The success of a CFS campaign is usually measured by the number of resulting fragment hits. They should be numerous, bind to the target site and offer diverse starting points. To support users in performing successful campaigns, efficient fragment libraries are needed. Furthermore, the fragments should be able to reach the target site, and each bound fragment hit should be clearly identified in the electron density, which relies on the analysis of the measured data. All points were addressed during this work. A CFS-specialized fragment library was developed where I helped to prepare and validate it. The influence of crystal packing on hit rates was investigated by me on a SARS-CoV-2 target. Additionally, I tested different analysis strategies to identify ways to find as many binders as possible.

4.1.1 Advantages of the F2X-Universal Library and F2X-Entry Screen

The F2X-Universal Library and F2X-Entry Screen comprise pharmacophorically diversified, yet also chemically diverse fragments. The representation of the commercially available fragment space was assured through clustering by the combination of 3D-shape

and pharmacophore similarity. The inclusion of pharmacophore information during clustering is a unique feature of the F2X libraries. The libraries are meant to be used for a wide range of targets and, therefore, no specific focus had been applied, like exclusively incorporating pharmacophores known to bind kinases or bias towards certain charges or polarities. However, most fragment libraries are built to reach high diversity in some aspect.^{198,199} This common parameter is usually not the unique selling point of a library, but a certain specialization is required that elevates one fragment library further compared to others. Therefore, a specialized method of preparing the fragments for crystallographic experiments was developed in cooperation with the Drug Design Group in Marburg already in the context of the HZB fragment library and the later commercialized Frag Xtal Screen.²⁰⁰ The fragments are dried onto a typical crystallization plate, as had been done in the case of the commercialized Frag Xtal Screen based on the work of Franziska Huschmann.⁴⁵ The F2X libraries offer the special opportunity to soak with high concentration without the need for organic solvents in the soaking solution, as volume and composition of the drop added to the fragment can be chosen by the experimenter. The fragments of the F2X libraries are dissolved in whichever soaking solution fits the crystal system. Organic solvents are often applied to support dissolving more hydrophobic fragments, though this can sometimes already sufficiently be achieved by PEGs, which are often part of crystallization conditions. In this way the experimenter has more freedom in choosing the best soaking condition. This will lead to a minimum of crystals deteriorating based on the soaking condition, resulting in more diffraction data, and ultimately more hits in the campaign. Additionally, as the fragment concentration is not bound to the added organic solvent content, concentrations up to 100 mM can be reached by adjusting the drop size applied to dissolve the fragments. The fragment presentation of the F2X libraries distinguishes this fragment library from available CFS libraries from larger companies like Enamine or Life Chemicals. Additionally, most systems offered at other facilities, for example Diamond Light Source, present their fragment library in a different way. They make use of an acoustic dispenser to combine droplets of fragment stocks to the protein crystallization drop.⁵⁷ These stocks are usually fragments dissolved in organic solvents, such as DMSO. This has the advantage of less manual stress to the protein crystal, as it does not have to be transferred to the soaking drop. In the case of protein crystals that are especially sensitive to manual stress, this way of soaking is favorable. However, it requires addition of, for example, DMSO to the crystallization drop, which is known to potentially deteriorate protein crystals. Furthermore, the tolerated DMSO amount predefines the highest concentration of fragment reachable inside the soaking drop. The importance of high soaking concentrations has been discussed earlier in this work (**Section 1.1.2.3**) and illustrated by Schiebel *et al.* in their Figure S1 (**Figure 4**), where soaking time and fragment

concentration were compared.⁴² Thus, to identify weakly binding fragments, high concentrations are vital. The advantages of the composition and presentation of the F2X libraries has manifested in typically high hit rates so far, ranging between 5-42%, but typically in the range of 10-25%, based on the presented campaigns in this work and the user campaigns at HZB (data not shown).

Due to its convenient size of 96 fragments, the sub-selection F2X-Entry Screen offers an easy entry into CFS for new targets. It allows users to test their target protein with its crystal system in a CFS setting, but also present the user with viable starting points already after this first campaign, compared to other small libraries as for example the MiniFragments³⁸, which provide a very useful probing and hot spot identification of the target surface, but are too small to start a fragment optimization. However, as shown in **Section 3.6, Figure 38**, the limited size of the F2X-Entry Screen might not provide the possibility to find many fragment clusters. Nevertheless, by screening the complete F2X-Universal Library, several binding sites of AR, that only displayed one fragment bound in the F2X-Entry Screen campaign, show large fragment clusters in the F2X-Universal Library campaign. Additionally, new binding sites were identified. These clusters provide inherent binding mode validation (further discussed in **Section 4.3.1.2**), enable merging and linking approaches (further discussed in **Section 4.3.2.2**), and hint towards putative protein functions (further discussed in **Section 4.3.3**). This additional information gained from one single CFS campaign demonstrates the benefit of screening larger fragment libraries. Therefore, the F2X-Entry Screen makes CFS campaigns more accessible to novice CFS users that would like to test their system first and get initial starting points. For more elaborate investigations of the target protein, screening with the F2X-Universal Library is particularly beneficial. Especially considering that the CFS workflow is nowadays optimized to perform screenings of 1000 fragments in a short amount of time, it requires a reasonable effort to screen the complete F2X-Universal Library. Taken together with the specialized setup of the libraries with the dried-on fragments, the F2X libraries present an advantageous way to perform CFS campaigns.

4.1.2 Crystal Packing Defines CFS Success

A high hit rate in CFS relies on the possibility for the fragments to access the protein's surface. Because of crystal contacts between individual protein molecules, certain surface areas are blocked by the neighboring protein molecule. This can include the target site of a protein, like the active site or a known PPI site. If the target site is hardly accessible because of the neighboring protein molecule, it is likely that the number of fragment hits at that site is small. Thus, the crystal packing has a direct influence on the hit rate. This situation had been examined for an exemplary protein target in detail in **Section 6.3**. The

study confirmed the dependency of crystal packing and hit rate through comparison of two different space groups screened for the same target, the SARS-CoV-2 main protease. By choosing the space group with better access to the active site, seven more fragment hits could be found. Thus, the hit rate was almost tripled. To further corroborate this, more crystal systems that allow for different packings should be investigated in the same way. A similar comparison could be performed by utilizing the analyzed CFS campaigns presented during this work. Prp8^{RNaseH} also crystallizes alone, and the crystals diffract to suitable resolution for CFS, and thus can be screened instead of the AR complex crystals. Already, during my student assistant work in the Wahl group, I screened the Frag Xtal Screen⁴⁵ against Prp8^{RNaseH}, resulting in only one hit, which could not be further reproduced (data not shown). These crystals however have a low solvent content, and the protein surface is largely covered in crystal contacts, which likely explains the poor outcome of this campaign. However, this CFS campaign cannot be directly compared to the AR campaign, as different fragment libraries were applied. For example, both libraries had been screened against EP.^{45,173} The Frag Xtal Screen against EP campaign resulted in a 9.4% hit rate, while the F2X-Entry Screen against EP campaign resulted in a 30% hit rate. Therefore, the shown lack of bound fragments cannot solely be ascribed to the crystal packing. Repeating the CFS campaign of Prp8^{RNaseH} with the F2X-Entry Screen, it would be possible to compare the hit rates of the single crystal system and the AR complex crystal system. The connection between crystal packing and hit rate could also be tested with a broader variety of target classes. For that, the PDB could be searched for protein structures that were solved from different crystal packings and apply CFS to them. This could be especially interesting in case of protein targets that have been screened before but resulted in a few fragment hits. The potentially additional found fragment hits through this study can directly be used for drug or tool compound development.

4.1.3 Key Developments for Efficient Data Analysis

The achieved hit rates in CFS rely further on the possibility to identify each weak binding fragment. The identification depends on the analysis of the data, including processing and refinement. In X-ray crystallography several different programs have been written. Each individual program presents distinct ways of handling the collected data, which results in slightly different electron density maps. The most common programs are included in FragMAXapp. This allows users to either choose their favorite programs or to perform data analysis with 54 different combinations of software. Nowadays, high performing computers allow analysis of the data in a few hours, which makes it possible for FragMAXapp to provide multiple analyses without prolonging the process substantially. Other data management tools do not present the same complexity, but usually focus on software developed in-house. For instance, XChemExplorer mentions the possibility to add other

analysis software but at Diamond Light Source, xia2 and duple are used exclusively. This can have a profound influence on the hit rate of a CFS campaign. In case of the F2X-Entry Screen against EP campaign, the full plethora of analysis programs in FragMAXapp were applied. Three additional fragments were identified because of the different combinations. The rather small number of additional fragments might occur as EP is a model protein and usually shows a high number of strong hits. Nonetheless, each individual fragment hit presents additional information about possible binding interactions for optimization. At HZB other CFS campaigns have been performed that utilized the multiplex analysis approach and confirmed the resulting increase in hit rates with real-life cases. For example, the F2X-Entry Screen against Nsp1 of SARS-CoV-2 campaign resulted in 6 fragment hits based on one combination. An additional 5 fragments were found by applying several combinations (Frank Lennartz, personal communication, data not shown). Thus, the FragMAXapp multiplex analysis supports users in finding as many bound fragments as possible. Especially in campaigns, which might result in small hit rates below 10% after one combination, it is recommended to use the full power of FragMAXapp analysis.

The push towards larger screening campaigns of several hundreds to thousands of fragments,^{102,103,201,202} as in the case of the F2X-Universal Library against AR campaign, brought up another way to improve the analysis of CFS data. Inherent polymorphism between individual protein crystals tend to manifest more clearly with larger sample size. Such polymorphism is especially tangible in PanDDA. This software compares the individual datasets with each other and marks datasets that show differences to the remaining ones. The datasets that show not such clear difference are chosen to build an averaged ground state model. If all datasets are presented for ground-state model building at the PanDDA step, the ground state model will represent an average of the full complexity in the CFS data set. However, if the data is separated before this analysis step, a more uniform ground state model could be built and help to find weak binding fragments. Such uniform ground state model could be reached through removal of outliers, which might result from faulty processing or refinement. Outliers would obscure the averaged ground state model. In case of large-scale campaigns, the data can be separated into several clusters. Each cluster is then used to build a more refined average ground state model. PanDDA itself performs an internal pre-clustering, solely based on the unit cell parameters. Differences in C_{α} positions or amplitudes between datasets are not considered. Such detailed clustering had been tested in the publication presented in **Section 3.6** by applying the program cluster4x.¹⁹⁶ Cluster4x allows users to look at their data through a correlation matrix, an SVD plot and the C_{α} positions of the protein structures. Mainly clustering is performed through separating the data based on the SVD plot. Several axes are calculated

which can be picked by the user to find a plot that visualizes the separate groups best. Datasets that are similar to each other are close to each other. By clustering the final 837 analyzed datasets into 11 clusters, it was possible to identify 22 additional fragment hits and 53 additional binding events. Compared to 249 fragment hits found without clustering, it is not a substantial increase. However, these additional fragment hits and binding events provide further information for optimization into follow-up compounds as shown in **Section 3.6** and **Section 6.1**. Four of the six presented fragments had been identified solely through the clustering analysis. Thus, validating the reproducibility of the binding mode was possible because of the extra clustering analysis step. An additional trial was performed in the F2X-Entry Screen against MPro (orthorhombic) campaign presented in **Section 6.3**. The campaign included less datasets than in the F2X-Universal campaign, but through clustering the data into three clusters, 6 further fragment hits were identified. Especially in this case, where a few fragment hits are found by conventional analysis methods, the benefit of clustering is confirmed. The presented data agrees with the original findings of the cluster4x publication. Ginn reported a 65% increase in the hit rate of the published CFS data targeting PTP1B.¹⁹⁶ Thus, it was shown that more isomorphous ground state models allowed to increase the number of found fragment hits. Additionally, future incorporation of cluster4x directly in FragMAXapp might allow to push the number of fragment hits even further.

4.2 Speeding up and simplifying CFS experiments and data management

Screening hundreds to thousands of crystals brings along the new challenges of handling such numbers of crystals and the resulting datasets. To tackle these bottlenecks, I took part in developing an evaporation protection device to help with crystal handling. Additionally, we published a protocol and accompanying video that teach novice users how to perform a CFS campaign. Furthermore, I beta-tested a web-based software application to keep track of collected data, their meta-information, and their analysis.

4.2.1 Protein Crystallization

The preparation of a CFS campaign is vital for its success. The experimenter needs to know beforehand the requirements for the experiment. Additionally, it is advantageous if the experimenter already knows the individual steps performed during a CFS campaign, to better understand the requirements. For a crystallographer untrained in CFS, such a campaign includes many unknowns. To prepare users at HZB and increase the chance for a successful CFS campaign, the CFS workflow we developed on site was explained in form of a detailed protocol, an accompanying video, and exemplary data of a CFS campaign performed at HZB. Similar presentations for other synchrotron sites have been

published, as such workflows differ between facilities.^{57,145,203} However, the main steps are common for all CFS workflows (**Figure 2**). At other synchrotron sites besides HZB, it is mostly required to crystallize the target protein on site of the synchrotron. The crystallization also must be optimized for special crystallization plates to be compatible with other machinery later in the workflow. However, to transfer a crystallization system to another lab often results in failed crystallization attempts or prolonged optimization, leading to the loss of valuable time or, in some cases might even lead to termination of a project. Therefore, the fragment screening platform at HZB allows users to prepare their samples simply at their home lab as shown in the publication. All necessary tools for sample preparation can be sent by the MX-group to the user. Thus, users do not need to invest more time for crystallization optimization on site at the synchrotron.

4.2.2 Crystal Handling

The soaking procedure following the crystallization is often done by addition of a fragment stock solution to the crystallization drop. This can be performed either via acoustic dispensing (Diamond Light Source, SLS) or pipetting robots (MAXIV). It has the advantage of imparting minimal mechanical stress to the crystal. The harvesting is either robot-assisted as with the Crystal Shifter at Diamond Light Source, MAXIV, SLS and ANSTO or the CrystalDirect system at ESRF. The crystal shifter⁴⁹ developed at diamond light source was developed to support scientists in crystal handling. It includes a motorized stage for the crystallization plate and a software to trigger the movements. The scientist prepares a csv-sheet to program the movement of the plate underneath the device and then the scientist only triggers the next move. This possibility increases the crystal handling speed substantially. Nevertheless, the crystal shifter is set up at one microscope and the plate moves exclusively translationally, which restricts the experimenter's movement. The CrystalDirect harvester^{48,50} developed at EMBL, which was implemented at the MASSIF-1 beamline in collaboration between ESRF and EMBL,²⁰⁴ is a fully automated option for crystal harvesting. The CrystalDirect harvester is a robotic system connected to a software, which allows the experimenter to plan which crystal should be harvested by the system. Afterwards, the crystals are harvested without further manual intervention and can be mounted directly at the beamline for measurement. Both options for crystal harvesting are rather expensive due to their robotic parts and therefore are mainly located at facilities like synchrotrons. At HZB, soaking is performed with dried-on fragments that are dissolved in a soaking solution. This approach demands the transfer of crystals from their crystallization drop to the soaking drop, which presents more mechanical stress than the options at other synchrotron sites. However, the soaking experiment does not need machines like acoustic dispensers or pipetting robots. Therefore, it is not necessary for the experiment to be performed at the synchrotron site. To also speed up the harvesting step at HZB, the

evaporation-protection device, the EasyAccess Frame, was developed and commercialized afterwards. It speeds up crystal handling by avoiding the cutting open and re-sealing of individual wells of crystallization plates. In this way, about 70 crystals per hour can be harvested by an experienced scientist. The device has a low cost as we refrained from using robotic parts or costly metal parts. This results in a mainly manually handled device, which provides certain advantages. For example, the EasyAccess Frame, can be moved between microscopes and labs and allows the experimenter to rotate the plate in any way to support more comfortable and easier crystal manipulation. The EasyAccess Frame has been in use at HZB for about 5 years and through this time it has supported users and scientists in the macromolecular crystallography group at HZB. The collected feedback throughout these years from the BESSY II MX-users has been positive. The harvesting process remains manual labor compared to other systems but provides a good alternative product to the existing offers. During the last years, the EasyAccess Frame allowed users to prepare large numbers of high-quality crystals for CFS campaigns in a matter of hours. Going further, speeding up the harvesting process with the frame even more could be imagined, by the addition of voice-tracked sample notes during harvesting. A tablet could be placed close by, and the user gives their notes orally to a voice-to-text software. The software could write the notes into a tracking sheet used later directly for sample tracking in ISPyB.²⁰⁵

4.2.3 Data Collection

To support the following data collection at Diamond Light Source, a specific fragment screening beamline had been built with high automation to a point that no user intervention is necessary during the measurements.¹⁴⁶ Also, at ESRF a fully automated beamline had been built, which is now also used for CFS campaigns.⁵⁶ They have the advantage that the beamlines can be used without much supervision by the scientist and thus frees up time for other experiments. Especially in cases of CFS campaigns the data collection includes many repetitive steps which can be taken over by machines. However, platforms with high automation usually lack the possibility for more individual applications. At HZB data collection is performed by the scientist manually, though with a degree of automation. It is planned in future endeavors to automate the data collection further at HZB, by integration of ISPyB and automatic sample centering.^{54,206–208} However, this part of the CFS workflow was not focused on during this work.

4.2.4 Data Management

Supporting the preparation of hundreds of samples in a short amount of time brings along another challenge after data collection, namely the data management. Each collected dataset must be tracked from the measurement to the final protein structure file. Additional

information, for example the fragment structure needs to relate to the respective datasets. To provide such an overview and the possibility to handle all data in one place, FragMAXapp was developed. Software developments largely rely on real-life case scenarios to identify bugs, missing features and how intuitive the interface is to users. Therefore, the CFS campaigns presented in this doctoral thesis project applied FragMAXapp during the software's early development stages. The published version of FragMAXapp presented here has been further improved based on the large-scale CFS campaign of the F2X-Universal Library against AR. A part of the FragMAXapp that still required additional work, included the preparation of the final data for PDB submission. The CFS campaign provided data collected throughout several beamtimes and the campaign required 269 structures to be prepared for PDB deposition. Thus, necessary scripts and additions to FragMAXapp were implemented based on this test case. Furthermore, FragMAXapp was implemented at other synchrotron sites in the last years, due to its modular architecture. A running version is currently present at BESSY II and SIRIUS. The successful transfer of the application shows the feasibility of such installations. Other analysis pipelines developed at Diamond Light Source or EMBL are also available for download, however their architecture cannot be transferred as easily to different platforms. At Diamond Light Source the XChemExplorer software has been developed.¹⁴⁴ It can be downloaded as part of the CCP4 software package²⁰⁹ and comprises the management of all data sets and their analysis. The PDB submission has not been implemented at the point of publication of the software but was mentioned as a future plan. The software does not include planning of sample preparation, as does FragMAXapp. At Diamond Light Source a different software takes care of such organization. Thus, FragMAXapp simplifies data management for users even further by including crystallization strategies and soaking plans. The management of crystallization and soaking experiments are also included in the CRIMS platform developed by EMBL and implemented at ESRF and DESY.¹⁵⁵ It is an overarching software that allows the user to plan their crystallography experiments from crystallization, via crystal harvesting with the CrystalDirect harvester, to the final structure. However, this software does not include the preparation of the final structures for PDB submission. The lack of such preparation is not an essential factor for single crystal structures, but in the context of a CFS campaign it is rather time-consuming. Therefore, it is a necessary addition to such data management software in order to speed up CFS workflows and make such experiments attractive for users working in a fast-paced publication environment. FragMAXapp addresses exactly these points, connecting the planning of all necessary experiments with data analysis and final PDB submission preparation.

All in all, the tools developed and the detailed CFS workflow presented allow users at HZB to perform their CFS campaign in an organized, time-efficient, and easily understandable way. This becomes increasingly important due to the advancements in high-throughput crystallography, which allows to increase the number of screened fragments. For example, the new beamlines planned at Diamond Light Source (K04)⁵⁸ and DESY (HiPhax)^{59,60}, which will increase throughput resulting in thousands of crystals measured a day. Such developments will without question lead to even larger CFS campaigns, which require developments as have been presented here.

4.3 What to do with your Fragment Hits?

After a CFS campaign, users are often left with the question what to do with the identified fragments. The main purpose of fragment hits is the further optimization into tool compounds or drugs. To support users in their further endeavors, the computational Frag4Lead workflow was developed in an HZB-Uni Marburg collaboration, which allows for fast and cost-effective optimization via catalog compounds. The Frag4Lead workflow was transferred into a KNIME-workflow that could be set up at HZB and then continuously improved also by me. Thus, the main purpose of fragment hits is addressed, however there is more information in the result of a CFS that can be extracted. To support users to get as much as possible out of their data, the F2X-Universal Library against AR campaign was analyzed by me with different aspects in mind to pinpoint additional information gained through large-scale CFS campaigns. The analysis resulted in the identification of additional benefits of such a CFS campaign for fragment prioritization and conclusions about biological functions of the screened target.

4.3.1 Prioritizing Fragment Hits

In the case of screening large fragment libraries of about 1000 fragments like the F2X-Universal Library, a plethora of starting points might be generated. As discussed before, these can be useful for merging and linking approaches. However, the user must select promising starting points out of the multiple fragment hits. Two steps of prioritization are discussed here, which were also considered during optimization for spliceosomal PPI modulator development based on the AR campaign. Depending on the results of a CFS campaign several binding sites may have been identified through clustering of fragments at specific surface areas. Therefore, each binding site must be examined for its potential biological function and thus, its prioritization during hit optimization. In the case of enzymes such as Endothiapepsin or MPro most fragments will bind to the active site which are likely to also be the target site for tool compound or drug development. There, the prioritization is often rather straight-forward. In cases like AR, where no explicit active site is known, other considerations are gathered to prioritize certain fragment clusters.

4.3.1.1 Prioritization Based on Target Sites in the AR campaign

Ten distinct fragment clusters were identified in the AR campaign (**Section 3.6, Figure 40**). In the following work, the meaning of these binding sites in the context of tool compound development will be explored.

A prominent binding site that is distinguishable at first sight from all other binding sites is the AR interface site. The interaction between Prp8^{RNaseH} and Aar2 is vital for the transport of the pre-U5 snRNP into the nucleus and regulates protein and RNA binding to Prp8 in the U5 snRNP context. Without the interaction of these proteins the pre-U5 snRNP would be unable to enter the spliceosome to form the pre-B complex. Therefore, inhibition of this interaction would lead to a halt of the splicing process at the stage of complex A. In contrast enhancing this particular interaction would result in the shuttling of the pre-U5 snRNP into the nucleus but prevention of the dissociation of Aar2 from the U5 snRNP. While, the splicing cycle would be halted at the same overall stage, it would nevertheless be interesting to elucidate if the pre-U5 snRNP can still be incorporated into the U4/U6.U5 tri-snRNP. On the other hand it is possible that exclusively single, vital components like the helicase Brr2 are not able to bind at the B complex stage anymore, or even that the pre-U5 snRNP stays detached from the spliceosome.¹³⁹ In both interventions, inhibition and enhancement, the importance of the interaction between Prp8^{RNaseH} and Aar2 is examined specifically compared to a total knockout of the proteins, which might lead to further biological effects unrelated from the incorporation of the U5-snRNP into the spliceosome. Targeting the AR interface site allows investigation of the early steps of the splicing process and provide further understanding of Aar2's role during splicing.

One further PPI site is known on the surface of Aar2 with Prp8. In this case Aar2 binds to the endonuclease domain of Prp8. As this is a second interaction between Aar2 and Prp8, it might lead to similar results as described previously, and thus would be just as interesting as the AR interface site. One argument against targeting this binding site would be the proximity of a crystal contact. The starting points for further developments must be chosen carefully and possibly validated in other binding assays to assure binding of the fragment in solution.

In the case of Prp8^{RNaseH} two other PPI sites are known, one is completely solvent exposed, and the other is also close to a crystal contact site. The binding site that is completely exposed to solvent overlaps with the PPI of Prp8^{RNaseH} and Prp3. Both proteins come into contact upon assembly of the tri-snRNP.²¹⁰ Prp3 is essential for the assembly and stability of the U4/U6 snRNP and binds directly to the U4/U6 duplex.¹⁸⁵ Prp3 mutants result in a halted assembly of the pre-B complex, due to defects in U4/U6 snRNP or tri-snRNP assembly.^{210,211} By developing inhibitors of the Prp8^{RNaseH} and Prp3 interaction, the binding

of Prp3 to the U4/U6 duplex would still be possible. Therefore, the assembly of the U4/U6 snRNP is not disturbed, but possibly the assembly of the tri-snRNP. A more focused examination of the importance of this specific PPI could be examined. Again, as with the AR interface site, the initial steps of spliceosome assembly would be investigated in more detail.

To target later stages of the splicing cycle, the second known PPI site of Prp8^{RNaseH} covered by a fragment cluster can be targeted: the interaction with Prp18. Prp18 is incorporated into the spliceosome at the stage of the C* complex responsible for the 2nd transesterification step.¹²⁰ It was shown that Prp18 is not vital for pre-mRNA splicing and thus likely does not halt the complete splicing cycle.²¹² However, it is still involved in the stabilization of the interaction between U5 snRNP and the exon.²¹³ Thus, an inhibitor for this PPI could simply elucidate the necessity of the Prp8^{RNaseH}-Prp18 interaction instead of restricting Prp18 function. As mentioned before, this binding site is in proximity to a crystal contact and therefore chosen starting points may have to undergo validation in solution. In all cases at known PPI sites a hypothesis can be postulated for the effect a tool compound might have and the resulting applications of it.

Out of the ten identified binding sites, six binding sites have no known overlapping PPI site or other known biological functions. The meaning of these binding sites is not clear after the CFS campaign and thus are not assigned as high a priority for immediate tool compound development as the binding sites with known functions. However, especially in the case of Aar2 which has not been investigated in as much detail as Prp8, the found fragment clusters could point to a variety of potential functions. Approaches to assess their function could include computational methods²¹⁴⁻²¹⁷ or directed mutations at these sites. Implications of fragments clustering at unknown binding sites are discussed further in **Section 4.3.3**. Again, some binding sites are close to crystal contacts which could hinder optimization of starting points. One might argue that binding sites at crystal contacts might not have any biological function at all. This has been discussed by Geerds *et al.*²¹⁸ where they showed that a specific crystal contact site overlaps with a PPI site. Thus, binding sites at crystal contacts should not be entirely disregarded, but rather first critically examined.

Taken together the information given for the different binding sites, priorities can be set to decide which target site should be examined first. As four sites have known biological functions, these should be prioritized during the development of spliceosomal tool compounds. Between the four binding sites targeting the AR interface site and Prp3 site has a high probability of resulting in a halted splicing cycle, thus imparting a more drastic change in splicing compared to targeting the Prp18 site. Additionally, the Prp18 site and Prp8^{Endonuclease} site are close to crystal contacts which will require further validation of bound

fragment hits before optimization. Therefore, during this project the AR interface site and the Prp3 site have been prioritized (**Section 6.1 and Section 6.2**).

4.3.1.2 Prioritization Based on Inherent Binding Mode Validation

A method of prioritization of individual fragment hits of a chosen target site was shown by Martin *et al.* through performing three independent screening assays.⁴¹ This is time-consuming and often not feasible for academic groups. Another way to prioritize fragment hits inside the same binding site is mentioned by Metz & Wollenhaupt *et al.*, who suggest validating binding modes of fragments and continuing with robust binding modes.⁸⁸ Such binding modes usually include strong interactions, that are not disturbed by slight changes to the fragment. Therefore, fragment hits with robust binding modes are favorable in optimization efforts. However, the additional step of validating binding modes includes purchasing analogs of the fragment hits and experimentally verifying their binding. Therefore, it requires additional costs and time. However, based on the F2X-Universal Library against AR campaign, an inherent binding mode validation emerged. Several fragments despite pharmacophore diversity repeat similar binding modes. This information can be useful to understand the binding mode in more detail and utilize this information in the first step of optimization. Such redundancy was criticized by Carbery *et al.*³⁵ They claim that the overlap of several fragments that are structurally diverse does not add more information and thus fragment libraries should focus on functionally diverse fragments that do not overlap. The idea to incorporate functionally diverse fragments is already included in the F2X libraries through the pharmacophore diversity. Thus, the identified overlapping fragments are not solely structurally diverse. However, it is noteworthy that the identification of repeating binding modes and novel binding modes are both important to consider in CFS. Here, the inherent binding mode validation can be used to prioritize certain starting points and circumvent the need to purchase additional analogs.

For example, in the case of the AR interface binding site, screening only the 96 compounds of the F2X-Entry Screen resulted in one single fragment hit, with a very particular binding mode including significant side chain rearrangements (**Section 6.1**). On its own this could be a singleton and would usually not be prioritized for further development unless the binding mode is validated. Especially if the binding mode includes a cryptic pocket, where it might be unclear which moieties are needed to trigger binding. Through screening the complete F2X-Universal Library it was possible to identify further fragments repeating the binding mode (**Section 3.6, Figure 38**) and to avoid purchasing analogs. The interface binders clearly show two viable exit vectors and a ring structure as a core, highlighting interactions necessary for binding (**Section 3.6., Figure 40**). This information was then

applied for optimization (**Section 6.1**). Even though until now no successful development of a higher affinity binder was yet achieved, the validation of this binding mode raised attention to the interface site, nonetheless. The performed ITC experiments (**Section 6.1**) also demonstrate that the fragments at the interface site are promising starting points.

In the case of the Prp3 binding site (**Section 6.2**), the repetition of the binding mode allowed for merging ideas which resulted in a follow-up compound binding as expected. The electron density of the follow-up compound is clearer than the electron density for the initial fragment hit. This observation is not because of better resolution, as the P06C11-soaked crystal diffracted to about 1.3 Å resolution and the follow-up-soaked crystal diffracted to about 1.5 Å resolution. The better-defined electron density of the follow-up compound might hint towards higher occupancy and in turn reflect an increase in binding affinity. To further corroborate this, one could establish a binding assay such as ITC or perform additional crystallographic experiments. Different follow-up compound concentrations could be tested in crystallographic experiments to find out which concentration is needed to visualize the follow-up compound in the electron density. Considering the thermodynamic explanation of the importance of the fragment concentration explained in **Section 1.1.2.3**, it was presented that the ligand concentration and K_D can be set in mathematical relation. Thus, if less follow-up compound is needed to push the equilibrium towards the protein-ligand complex, the K_D will be lower. An estimation of the K_D could be given.

The given examples point out the benefit of inherent binding mode validation and utilizing this information during hit optimization. The fragment hits that do show repeating binding modes could be prioritized in the beginning of optimization. Thus, the follow-up development is sped up by performing large-scale CFS campaigns.

4.3.2 Optimization of Fragment Hits into Lead Compounds

The optimization of fragment hits often involves custom synthesis or collaborations with medicinal chemists. Such solutions are usually expensive and cumbersome, especially for academic groups. As most users at HZB are academic groups, they often do not have the resources for extensive chemical synthesis in the beginning of a tool compound development or drug development. Therefore, the Frag4Lead workflow was developed as an answer to how users could perform the first step of compound optimization in a fast and cost-effective way.

4.3.2.1 Automated Growing Approach via Frag4Lead

The successful application of the Frag4Lead workflow had been presented with the model protein EP. Based on five starting points, 28 follow-ups were purchased, which is a feasible

number for academic groups. Out of these 28 follow-ups, 10 follow-ups were identified in the electron density, resulting in a success rate of 30 % successful binders. For the best of the follow-up compounds, a 266-fold increase in binding affinity was measured, while the follow-up also maintained the binding pose of the initial fragment.⁸⁸ The approach to start with commercially available molecules for follow-up compounds makes the first optimization step more accessible. A slightly different approach is to choose a specialized library, which is focused on fast follow-up development. The DSi-Poised Library developed at Diamond Light Source, providing so-called poised fragments as starting points.²¹⁹ Poised fragments were characterized as fragments that can be synthesized via well-known chemical reactions. Based on such reactions, analogs of each fragment can be synthesized in an easy way. The library was developed together with the Structural Genomic Consortium, iNEXT and Enamine. Enamine provides these low-cost analogs of each fragment included in the DSi-Poised Library. In this way users can easily find analogs of their fragment hits to first validate the binding pose. Subsequently, the bound analogs are used to build up a follow-up library. Another platform utilizing well-known chemical reactions was presented by Bentley *et al.* called rapid elaboration of fragments into leads using X-ray crystallography (REFiL_X).²²⁰ Here, the fragment hit might result from any fragment library, but analogs with different exit vectors are tested to find feasible starting points for their approach. After choosing promising starting points, a library of compounds tailored to the individual fragment hit is prepared and parallel synthesis is applied. For this approach the focus is put on well-known, high-yield reactions, which allows for microscale synthesis for example in 96-well plates. The crude reaction mixtures are used for crystallographic experiments and in case of a binding event, the reaction product is purified and tested again via X-ray crystallography. This platform allows for fast synthesis and testing of follow-up compounds. Both platforms limit their follow-up chemical space to molecules easily synthesized via certain well-known reactions. Thus, not the full repertoire of commercially available compounds is used in these approaches. In contrast the Frag4Lead workflow is not directly dependent on the chemical reaction possibly performed with the fragment hit. The Frag4Lead workflow can search a plethora of catalogs or chemical spaces. Thus, even natural compounds or compounds resulting from multiple-step reactions are considered. Additionally, the Frag4Lead workflow includes necessary steps like virtual template-based docking and filtering the resulting docking poses in a reasonable way. Such work is still required for example with the DSi-posed library, as a selection of analogs and follow-ups is needed. Therefore, the Frag4Lead workflow presents an alternative approach, for optimization by growing, while assisting the user in compound selection.

4.3.2.2 Facilitated Fragment Merging and Linking via Similarity Search

The Frag4Lead workflow supports fragment growing as an optimization strategy, which had been the most popular strategy these last years. Growing is advantageous because in case of growing, basically one fragment hit is enough for optimization. Additionally, with the help of Frag4Lead, growing is rather straight-forward based on the approach to search for commercially available compounds that are superstructures of the fragment hit or through other approaches as the elaboration of fragments via well-known chemical reactions. However, in recent years the paradigm shifts further towards merging and linking, as an increasing number of CFS campaigns are performed with up to 1000 screened fragments.^{102,103,201,202} Such screenings usually result in hundreds of fragment hits. Some of these fragments will have overlapping structures (as shown in **Section 3.6**) or bind close enough to each other for linking. Thus, merging and linking strategies become more feasible. However, both strategies require so far rather manual work in finding appropriate starting points for this and promising follow-ups. In case of merging often the scientist must identify starting points that overlap in a meaningful way and create the new compound themselves. In the case of linking, the scientist must find two starting points that are oriented in a way which allows the placement of a linker and are at the same time far enough separated to leave space for a linker. First ideas to support users in their endeavors have been presented by the development of Fragmenstein.²²¹ This program allows users to present several fragment hit poses that might work for merging or linking and the program will combine these fragments. However, the synthetic feasibility is not considered and thus the resulting follow-up compounds might neither be commercially available nor simple to synthesize. Despite the current lack of automation in these optimization strategies, the merging approach had been taken on in **Section 6.2** to generate promising follow-up compounds. Even though the Frag4Lead workflow does not help with merging directly, it was still applied in an adapted way in this project. Often the merged compound might not be commercially available and thus results in the difficulty of finding a medicinal chemist to synthesize the desired compound. To avoid this, the desired compound can be used in a similarity search. Found analogs can also undergo template-based docking, similar to Frag4Lead in order to prioritize candidates. In the future a combination of programs like Fragmenstein and the Frag4Lead workflow could be imagined, streamlining the process even further. The complete CFS dataset could be entered into the workflow with automatic selection of merging and linking fragment candidates through the overlap of similar structures and close proximity. These chosen fragment hits would be combined as done by Fragmenstein and directly feed back into the template-based docking. The resulting merged or linked follow-up compounds would be investigated by the scientist.

4.3.3 Discovery of Putative Protein Functions

The F2X-Universal Library against AR campaign pointed out another interesting use of fragment hits. The identified fragment hits cluster in certain areas of the protein surface. Three of the presented fragment clusters coincide with known PPI sites in context of AR within the spliceosome, based on cryo-EM structures. Therefore, a connection between the fragment clusters and PPI sites could be hypothesized for the remaining fragment clusters. In this way CFS results allow users to understand their target protein better. Putative PPI sites could be investigated for their biological function and interaction partners could be determined. Such studies could be performed with tool compounds developed based on the identified fragment hits. Going one step further, other functional binding sites could also be found through CFS campaigns. For example, the possibility to identify allosteric sites of proteins through CFS had been discussed before by Ludlow *et al.* and Krojer *et al.*, though assessing the actual function of such putative allosteric sites remained challenging.^{222,223} Another study had been performed by Shumilin *et al.* focusing on metabolite screening to identify natural ligands of a protein and thus understanding the protein's function.²²⁴ The chosen metabolites were separated into cocktails based on their structural similarity. Each cocktail included smaller ligands (e.g. uric acid) and larger ligands (e.g. FMN). They were able to elucidate the function of two proteins from different protein families, where function had been unknown at that point. A similar approach is currently examined by a master student Paula Fröling at the Macromolecular Crystallography Group at HZB. It is noteworthy to highlight the use of larger ligands like FMN by Shumilin *et al.* instead of cutting such ligands into smaller fragments for screening. The second approach had been examined by Barelier *et al.*²²⁵ and discussed by Dan Erlanson.^{226,227} The promise of finding fragments that directly elucidate the natural substrate seems not as straight-forward. However, the publications by Shumilin *et al.*, Ludlow *et al.* and Krojer *et al.* are in line with a broader hypothesis: fragment hits from CFS campaigns bind to biologically relevant binding sites of proteins. CFS campaigns thus, offer the possibility of gaining additional information about the target protein which were previously unknown. Additionally, the possibility of fragments binding to biologically relevant binding sites means that each CFS campaign potentially leads to several target sites for compound optimization.

4.3.4 Methodological Investigations of CFS Results

From a methodological point of view large-scale CFS campaigns offer even further insight into protein-ligand interactions. For example, they provide the opportunity to investigate in detail which kinds of moieties bind to PPI sites. This might differ compared to active sites, also shown by fragment libraries especially designed for PPI sites (Enamine, Life Chemicals). Such a viewpoint had not been presented in the published articles. However,

screening the F2X-Universal Library against AR campaign especially provides a large dataset with much more information than can be discussed within the limitations of this work. One could compare all CFS campaigns performed at HZB with the F2X-Entry Screen and extract fragments binding mainly to PPI sites or known active sites. Thereby, a tendency for certain bound fragments might come to light. In case specific moieties can be attributed to specific binding sites, one could imagine pinpointing the function of an identified binding site simply by screening a library built from these results. For example, some fragments would be known to bind almost exclusively to PPI sites and others almost exclusively to certain active sites. The binding sites identified through a CFS campaign could be analyzed for the bound fragments and already based on these the biological function would be known. This approach would speed up the assessment of functional sites on proteins.

5. Conclusion

Within this thesis project, almost each step of the CFS workflow was improved to make screening more efficient, concise, and simple. Several CFS campaigns by users were supported through these developments. The performed CFS campaigns did show the large hit rates achievable by screening the F2X libraries and utilizing FragMAXapp for data analysis. Besides this, they also showed that additional information can be gained from fragment hits. The putative PPI sites of Aar2 can be further investigated and help to find out more about this protein. It was possible to start the development of spliceosomal PPI modulators and through applying the CFS workflow at HZB the offered support for CFS users was improved in several aspects of the workflow. The next steps of spliceosomal modulator design will continue to derive a functional tool compound. Additionally, further investigations to automate the CFS workflow as much as possible are planned, as well as keeping the follow-up workflow Frag4Lead up to date.

By improving the CFS workflow, users are supported in performing such campaigns with minimal effort and limited time. Thus, tool compounds needed to investigate cellular processes could be developed in shorter times. As drug development also utilizes the same CFS workflow during the screening step, drugs can equally be developed faster. Time is especially critical for drug development, as it has economic value and a direct link to the health of individual people. Therefore, more time-efficient structure-based drug discovery is required for a functional and healthy society.

6. Appendix – further projects

The presented publications document an improvement of the CFS workflow applied at HZB and the first step of developing several PPI modulators for the AR target. However, the project is not concluded by the presentation of the found fragment hits. Additional projects were started that did not reach the point of publication yet. The identified starting points were investigated in more detail, taking the first steps towards an optimized binder. The optimized CFS workflow was further applied to highly relevant targets such as the SARS-CoV-2 cysteine protease MPro. This target could be crystallized in two different space groups, which were both screened, and revealed the importance of crystal packing in CFS campaigns. These projects are presented here with the initial results and future plans to support these findings.

6.1 Cryptic pocket identified at a spliceosomal PPI interface

The large-scale CFS campaign of the F2X-Universal Library against the AR target revealed hundreds of starting points for fragment-based ligand design. A few fragments bound directly to the interface of AR. Most of these fragments bind at the same position, repeating a similar binding mode (**Section 3.6**). Binding of the fragments results in a movement of the Arg186 of Aar2, exposing the side chain into the solvent and opening up a novel binding pocket. Such pockets are referred to as cryptic pockets.²²⁸ This pocket has been examined further to understand the interplay of the displacement of the arginine side chain and the binding of the fragment in more detail.

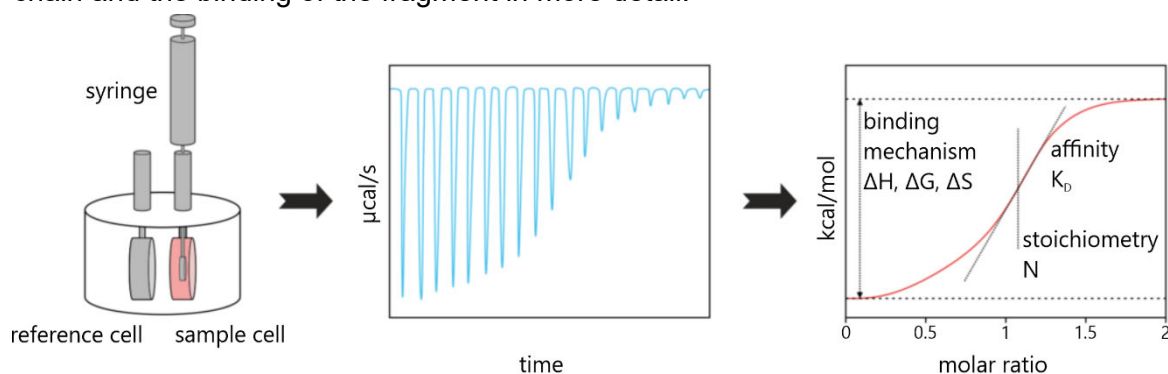


Figure 42: Schematic view of the ITC assay adapted from <https://2bind.com/itc/>.²²⁹ On the left side the setup is shown which shows the syringe, the reference cell and sample cell. The exemplary resulting raw data is shown in the middle. Based on this data a sigmoidal curve can be calculated which allows to extract the binding affinity, the stoichiometry, the enthalpy and solve the Gibbs free energy equation for the entropic term.

An already established ITC assay for the AR complex was applied (student assistant project at AG Wahl). ITC is a prominent method to investigate protein-protein or protein-ligand interactions. It is a unique technique as it reveals the thermodynamics of that interaction. The setup can be seen in **Figure 42**. A reference cell is filled with water while the sample cell is filled with an aqueous solution including the protein of interest. Via

a rotating syringe the second component (either the second protein of interest or a ligand) is added in a stepwise manner to the sample cell. During each titration step in case binding occurs, the temperature of the sample cell will slightly change. Either an exothermic or endothermic binding reaction happens, which will lead to a differing temperature of the sample cell to the reference cell. The machine will heat or cool down the sample cell with as much energy as needed to eliminate the temperature difference. The energy needed will decrease after a certain amount of ligand or protein is added, resulting in a sigmoidal curve. Based on the data, the binding affinity and stoichiometry can be calculated. Additionally, the enthalpy is extracted and, together with the affinity, can be used to solve the Gibbs free energy equation for the entropic term. Thus, by applying ITC it is possible to learn more about the binding mode and the possible influence the displacement of Arg186 has on binding.

total injection	16
cell temperature	25°C
reference power	5 µcal/s
initial delay	120 s
stirring speed	750 rpm
first injection volume	0.5 µl
first injection duration	1 s
injection volume	2.5 µl
injection duration	5 s
injection spacing	120 s
filter period	5 s
feedback mode	high
cell protein concentration	14 µM
syringe protein concentration	140 µM
DMSO concentration	4 % (v/v)
fragment concentration	40 mM
buffer	20 mM Tris pH 7.5 @ RT, 150 mM NaCl

Table 4: The ITC assay parameter for AR binding examination. The protein sample composition is listed as well as the machine parameter for measurement.

The assay was performed in the following way (**Table 4**): The Aar2 protein was present in the sample cell. The Prp8^{RNaseH} protein was present in the syringe. The fragment was present in both sample solutions at a concentration of 40mM. To investigate the influence of Arg186, an alanine mutant was used in the experiment.¹³⁹ Based on the publication of Weber *et al.*, it is known that Arg186 is not vital for the AR interaction. Thus, the influence of P01F08 on the binding affinity of AR could be examined in presence and in absence of Arg186.

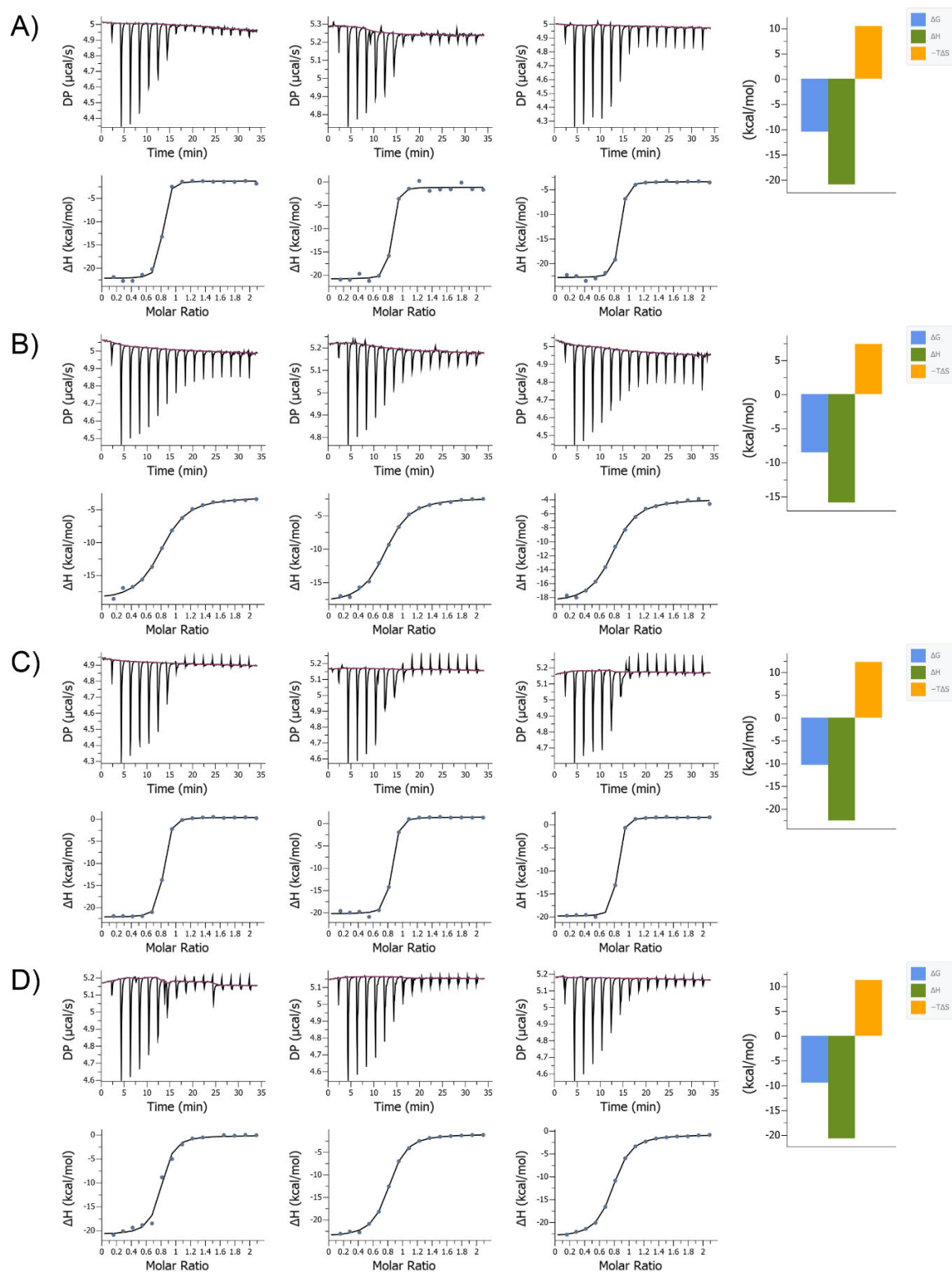


Figure 43: Raw data of all triplicate measurements are depicted. Additionally, for the first triplicate the thermodynamic signature plot is shown. A) Raw data and an exemplary signature plot are given for the wildtype AR binding measurements. B) Raw data and an exemplary signature plot are given for the binding of the Aar2 mutant R186A and Prp8^{RNaseH}. C) The raw data and one signature plot are shown for wildtype AR binding with the addition of fragment P01F08. D) Raw data with one signature plot are depicted for binding of the Aar2 mutant R186A and Prp8^{RNaseH} in presence of P01F08.

The raw data and the resulting curve are shown in **Figure 43**. Each sample was measured in triplicates. The resulting values were averaged, and a standard deviation was calculated (**Table 5**). The average binding affinity for the wildtype protein complex is 17.6 nM. In the absence of R186, the binding affinity is decreased to 52.8 nM. The addition of P01F08 slightly decreases the binding affinity of the wildtype complex to 21.4 nM, however considering the standard deviation of 3.3 nM, only a trend towards lower affinity can be seen. In case of the AR complex with the Aar2 R186A mutant, P01F08 restores the binding affinity almost to the wildtype binding affinity (24.3 nM). Thus, the displacement of R186 by P01F08 takes about the same energy as is released upon binding of P01F08. This is reflected again in case of the R186A mutant complex. Binding of P01F08 can almost restore the binding energy in the absence of R186.

	average K_D	standard deviation
wildtype complex	17.6 nM	2.7 nM
Aar2 R186A complex	52.8 nM	1.9 nM
wildtype complex plus P01F08	21.4 nM	3.3 nM
Aar2 R186A complex plus P01F08	24.3 nM	1.0 nM

Table 5: Calculated binding affinities based on the ITC measurements. For each sample type triplicates were measured. The resulting three values for each sample type were averaged and the standard deviation was calculated.

The first results presented here reveal the underlying nature of the binding mode shown by P01F08. The binding mode can be reproduced and shows a measurable influence on the binding affinity of the AR complex. However, due to the energy needed to displace R186, this starting point needs to be adapted further to the binding site for more favorable interactions. Additional room temperature crystallography experiments were performed with the AR complex and the P01F08 fragment (data not shown). The collected data showed no difference to the cryogenic temperature data. Thus, the binding of P01F08 is a reproducible binding mode with measurable influence on the AR interaction. Based on this promising starting point, the Frag4Lead workflow was employed to find higher affinity binders. To broaden the chemical space that can be searched for follow-up compounds, two further fragments binding in a similar way were chosen, namely P06G06 and P07C05 (**Figure 44**).

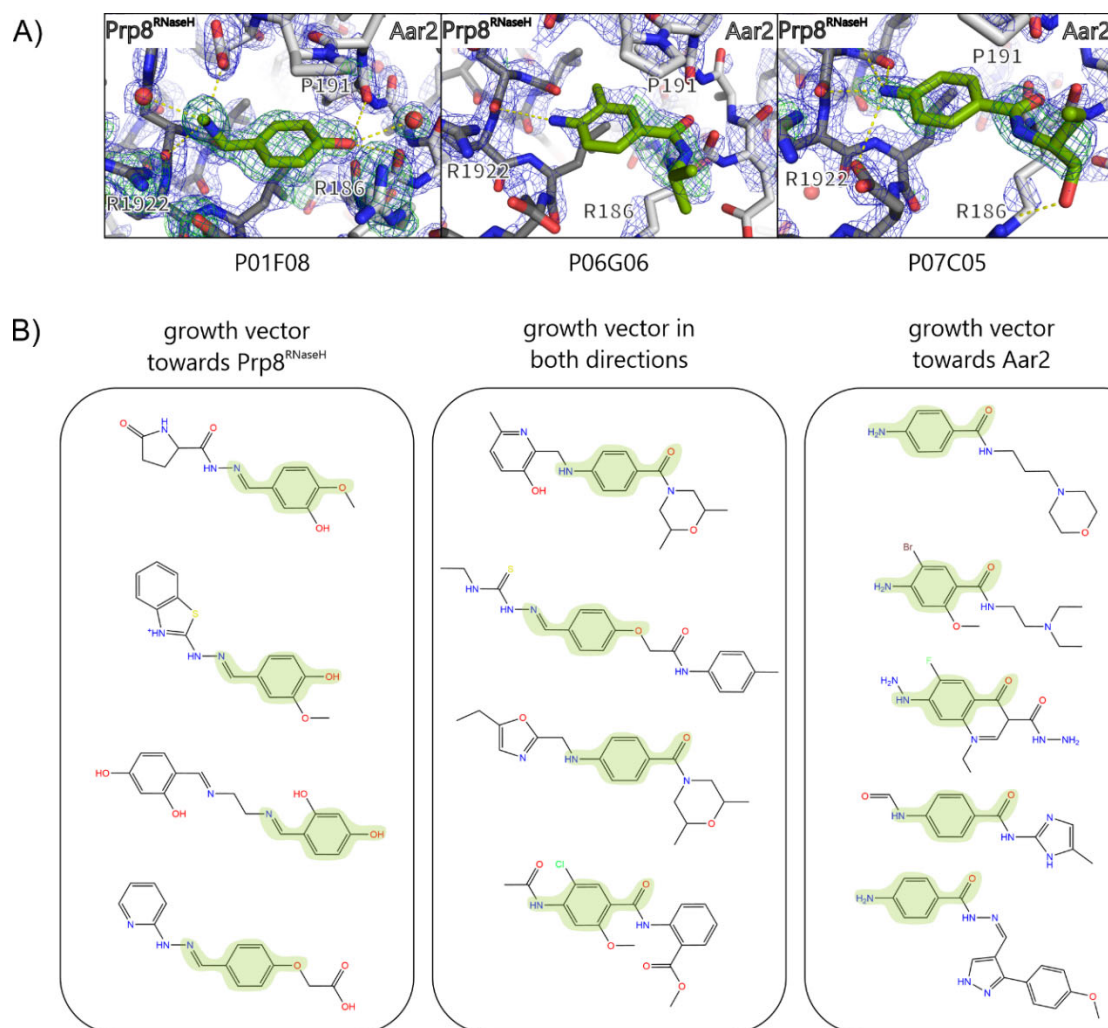


Figure 44: The first growing strategy for an AR PPI enhancer. A) The fragment-bound protein structure is shown in stick format for the three fragments chosen for the growing strategy. Prp8^{RNaseH} is shown in dark grey, Aar2 in light grey (inhouse refined apo structure) and the fragments in green (PDB ID 5QYG, 7FN1, 7FNH). Water molecules are shown as red spheres. The event map ($\sigma=2$) and Z-map ($Z=3$) are shown for each fragment. The figure has been prepared via PyMOL.⁸⁵ B) The purchased follow-up compounds are depicted in 2D presentation. The compounds are categorized based on the direction they grow along the proteins. The green cloud highlights the original fragment structure used for substructure search.

Different runs of the Frag4Lead workflow were performed to identify the best way to run this application. These tests included searching different commercial chemical spaces (for example MolPort, Chemspace) and performing docking with and without water molecules present in the binding site. In the end, 13 molecules were ordered as promising follow-up compounds (**Figure 44B**). These compounds were tested via crystallographic approaches by soaking and co-crystallization. Additionally, all compounds were tested with the established ITC assay. In all cases no binding at the interface was observed. This led to further examination of the binding mode to potentially find new ideas for compound optimization.

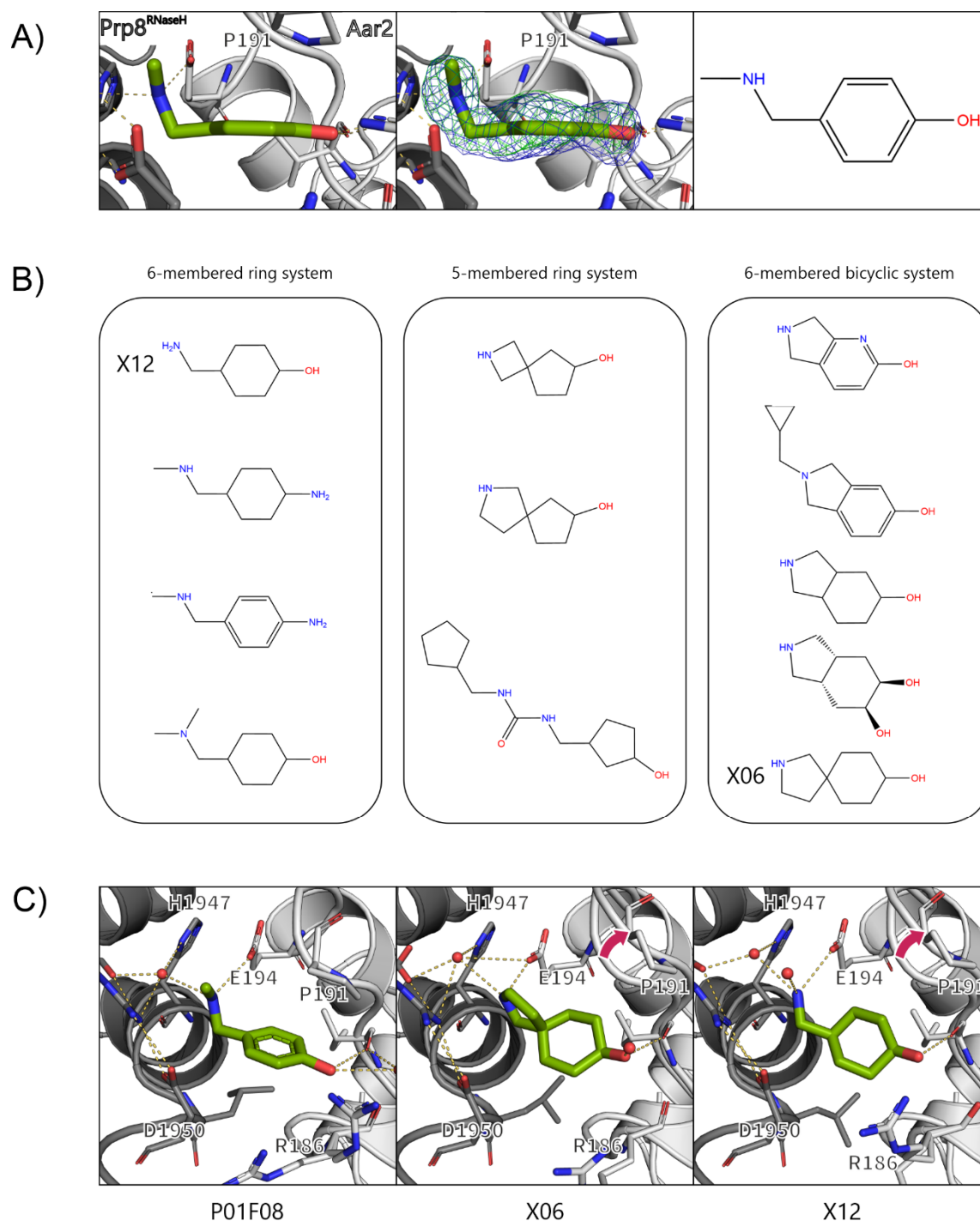


Figure 45: Overview of tested analogs to avoid possibly strained P01F08 binding. A) From left to right: The P01F08 fragment (green sticks) is shown from the side binding to the AR interface (Prp8^{RNaseH} in dark grey, Aar2 in light grey, PDB ID 5QYG). The same view is shown with the respective event map ($\sigma=2$) and Z-map ($Z=3$) resulting from PanDDA. The figures have been prepared via PyMOL.⁸⁵ The 2D structure is shown of the P01F08 fragment. B) The twelve purchased analogs are shown in 2D representation. They are categorized based on their ring system. The two bound analogs are marked with their ID X06 and X12. C) The 3D crystallographic structures are shown of all binders (P01F08 PDB ID 5QYG, the other structures are not yet submitted to the PDB). The fragments are shown as green sticks, the proteins in grey, the water molecules as red spheres and the possible hydrogen bonds as dashed yellow lines. The pink arrow shows a larger movement of the X06- and X12-bound structure compared to the P01F08-bound structure. The figures have been prepared via PyMOL.⁸⁵

A closer look at the binding mode of the clearest hit P01F08 shows an intrinsic strain, which might result in unfavorable binding energies (**Figure 45A**). To find more possible suitable starting points, a core replacement was performed. Several fragment-sized molecules were ordered that kept the atoms necessary for H-bond formation and exchanged the fragment's ring structure with differently sized aromatic and aliphatic rings (**Figure 45B**). These molecules were again tested in soaking and co-crystallization experiments. Two of the 12 molecules were found as binders, X06 and X12 (**Figure 45C**). They were used as starting points for the Frag4Lead workflows, but no promising follow-up compounds were identified. Therefore, no optimized binder could be developed at this stage.

Taken together, the binding mode of P01F08 at the interface of AR has been examined in detail and first efforts were performed to find higher affinity binders. The so far unsuccessful optimization of P01F08, P06G06 and P07C05 reveals the challenging nature of this binding pocket. Future plans could include molecular dynamics simulations of the interface to understand the inherent flexibility of R186 in more detail and simulate the binding of P01F08. Additionally, it could be possible to involve medicinal chemists to test further smaller changes to the P01F08 structure to pinpoint favorable, tolerated or unfavorable changes. In this way it could be possible to develop an AR PPI enhancer.

6.2 First optimization step of Prp3-Prp8^{RNaseH} PPI inhibitor

The CFS campaign against AR resulted in various possible target sites for PPI inhibitor development. One of these sites is the binding site of Prp3 to Prp8^{RNaseH}. Their interaction is necessary throughout the splicing cycle and if inhibited it could lead to a termination.²¹⁰ However, a complete knockout of Prp3 could possibly result in another phenotype than inhibiting solely one PPI. Therefore, developing a PPI modulator for the interaction of Prp3 to Prp8^{RNaseH} would be more direct and specific.

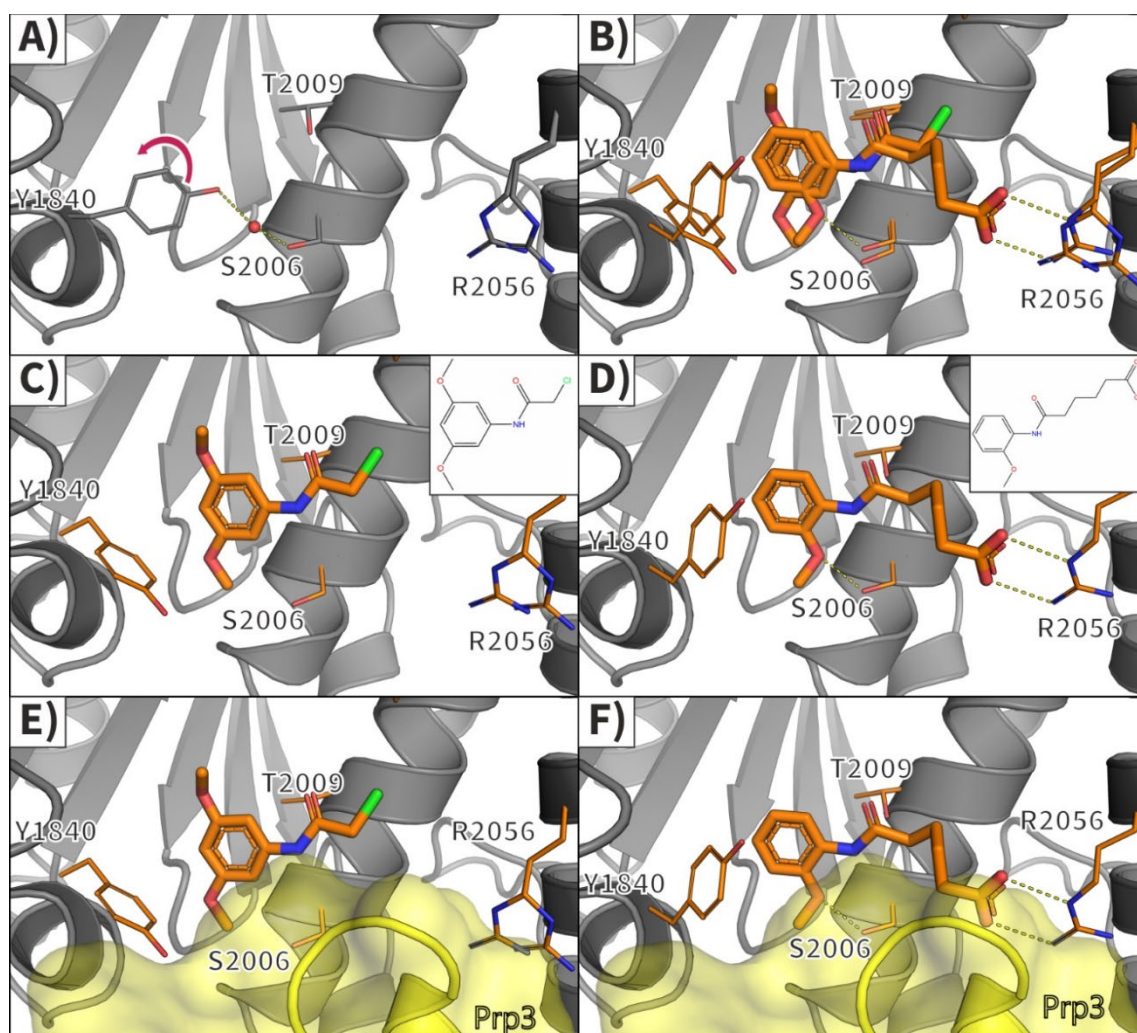


Figure 46: Two exemplary fragments bound at the Prp3 binding site. Fragments are always shown as orange sticks, the Prp8^{RNaseH} protein is shown as dark grey cartoon (inhouse refined apo structure) and the Prp3 protein is shown as yellow cartoon and transparent surface (PDB ID 3JCM).¹⁸⁵ Water molecules are shown as red spheres and hydrogen bonds as yellow dashed lines. A) The apo structure is displayed with an arrow next to Y1840 to indicate the movement upon fragment binding. B) An overlay of the two exemplary fragments is shown, presenting overlapping structural features. C) P03C07 bound to Prp8^{RNaseH} (PDB ID 5STU) is shown with the respective 2D representation in the upper right corner. C) P06C11 bound to Prp8^{RNaseH} (PDB ID 7FMI) is shown with the respective 2D representation in the upper right corner. E) The overlap of Prp3 with P03C07 is displayed to indicate possible clashes. F) The overlap of Prp3 with P06C11 is displayed to indicate possible clashes. The figures have been prepared via PyMOL.⁸⁵

Some fragment hits bound to the binding site of Prp3 have been discussed in Barthel *et al.* A movement of Tyr1840 had been observed in this binding mode. Therefore, another cryptic site was identified by performing a crystallographic fragment screening. The new pocket opens up a larger hydrophobic area, which was shown to bind different aromatic ring systems with ether moieties. This pocket is close to the binding site of Prp3, but most fragment hits found lack an elongated structure that could disturb the PPI. Only P03C07 and P06C11 reach across the protein surface, closer towards the space Prp3 would occupy upon binding (**Figure 46**). Therefore, larger compounds should grow further across the binding site.

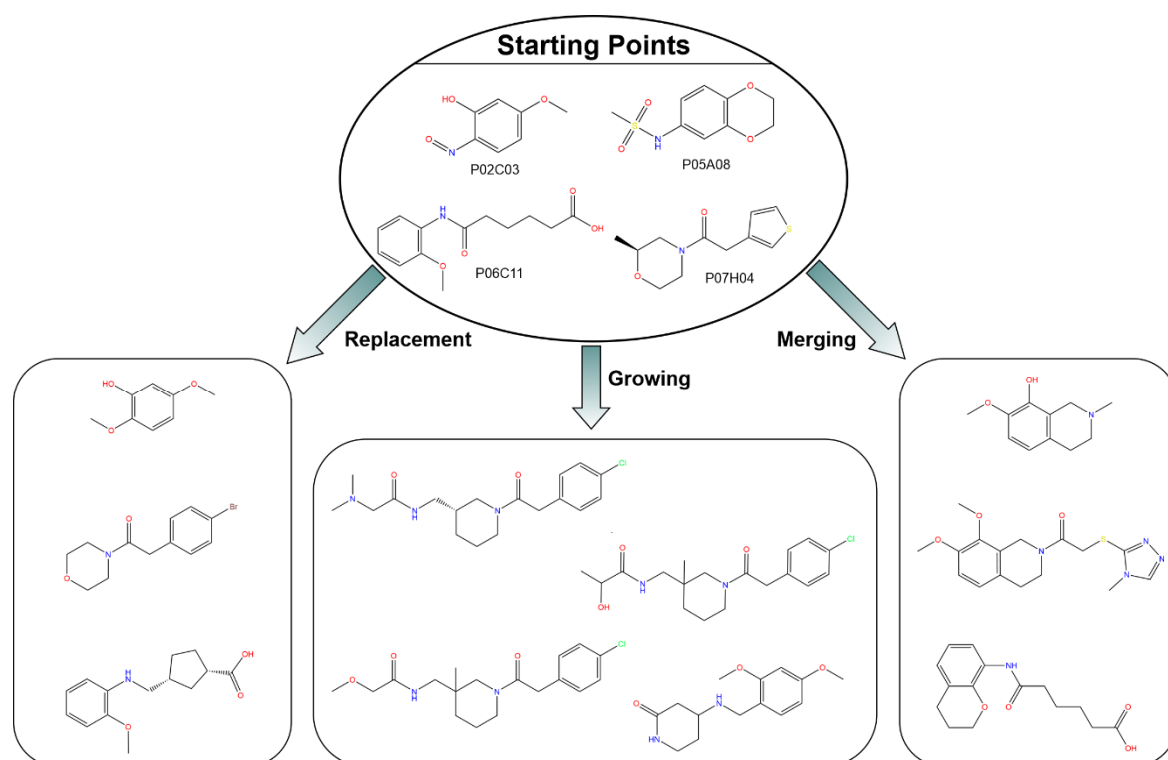


Figure 47: Overview of optimization strategies applied for purchased compounds. Four fragments were chosen as starting points. One strategy applied was to replace certain groups for better fit towards the binding site. The typical growing approach was used partly for the original starting point or a replacement compound. Due to the overlapping fragment binding poses merging ideas were tested. Overall, ten compounds were purchased for the Prp3 binding site.

For this optimization path several starting points were chosen (P01C10, P02C03, P05A08, P06C11) (**Figure 47**). The first chosen optimization approach was the growing method. The updated Frag4Lead workflow mentioned in **Section 6.1** was applied. Further changes were made to the workflow based on the experiences during the development of an AR PPI enhancer. The number of chemical spaces subjected to a substructure search were increased, including now non-commercial spaces. This approach will broaden the searchable chemical space, as different chemical spaces were shown to have a rather

small overlap in offered compounds.²³⁰ To be able to search large or ultra-large chemical spaces in a reasonable time, further programs are needed, like SpaceMACS.⁷⁶ SpaceMACS allows searching such spaces without full enumeration. The largest commercial chemical space that was searched was the Enamine REAL space.²³¹ Additionally, the MolPort space¹⁹² and the Freedom space were searched.²³² Furthermore, the Knowledge Space was included as a non-commercial chemical space. The resulting superstructures were processed via the Frag4Lead workflow. The output was manually examined via SeeSAR.⁷⁵

Through this rather straight-forward growing approach, only a small number of viable follow-ups were identified. To expand the list of possible follow-ups, several further approaches were applied. Merging ideas were explored based on the overlapping structures of the identified fragments. Furthermore, a core replacement was performed for fragment P06C11. This fragment presents a rather flexible elongated aliphatic structure, which could be favorably stabilized through fitting ring structures. The core replacement was performed via SeeSAR's inspirator mode. The resulting chemicals from both strategies were searched via a similarity search in the previously mentioned commercial spaces to find available options. Next to these more typical optimization strategies other ideas were tested. The identified fragments were used in a similarity search with Enamine building blocks.²³³ This identified new starting points, which were used for substructure searches in the REAL space. Additionally, manual alterations of the fragments were performed. Thus, unfavorable atoms were replaced with more favorable ones, as for example replacing an oxygen atom with a carbon atom, which points towards a hydrophobic area. The most promising compounds were purchased (**Figure 47**) and tested via soaking and co-crystallization.

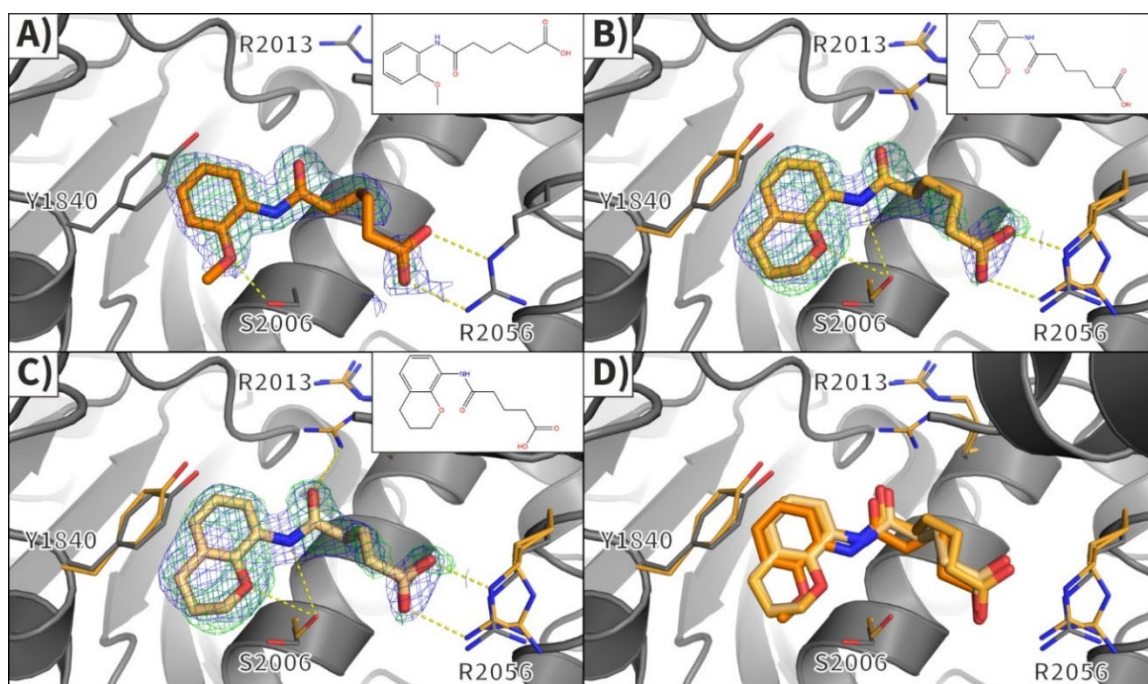


Figure 48: Crystal structure of a bound follow-up compound. The protein Prp8^{RNaseH} is shown as grey cartoon. The compounds are shown in different orange colors. Important side chains are shown as sticks, either in grey for the protein structure of the starting point P06C11, or in orange in case of the follow-up compound protein structure. The 2mF_o-DF_c map is shown in blue ($\sigma=1$) and the mF_o-DF_c map is shown in green ($\sigma=3$). Hydrogen bonds are shown as yellow dashed lines and the 2D structure of each compound is shown in the upper right corner of the respective panel. A) The starting point P06C11 (PDB ID 7FMI) is shown with its respective electron density and hydrogen bonds to the protein. B) The purchased follow-up compound is presented modelled into the electron density after auto-refinement with its respective hydrogen bonds. C) The purchased follow-up compound with a shortened arm (one less C-atom) is presented modelled into the electron density after auto-refinement with its respective hydrogen bonds. D) An overlay of each of the structures is shown to highlight the repeated binding mode.

One of the compounds was found to bind as expected in the Prp3 binding pocket (**Figure 48**). It is a merging follow-up with only marginal changes to the original structure. However, based already on docking and scoring calculations, it was expected to substantially improve binding compared to the original fragments. The fragment P06C11 binds with a maximum estimated affinity (by the HYDE scoring function in SeeSAR) of 1 μ M to the protein. The bound follow-up compound binds with a maximum estimated affinity of 7 nM. In the crystal structure of P06C11, the flexible part of the molecule is not resolved well. The follow-up compound shows much clearer electron density, hinting towards a possible increase in binding affinity. Interestingly, close examination of the electron density showed a better fit for a shortened compound shown in **Figure 48C**. Upon request, the company Enamine provided data of their quality control, which agrees with the longer compound that had been purchased (data not shown). Additionally, impurities identified in the compound batch did not show signs of the proposed shortened construct. Thus, further investigations are needed to fully understand the follow-up compounds' binding.

To demonstrate the compounds' potential inhibitory function, the interaction between Prp3 and Prp8^{RNaseH} must be measured. Although both proteins clearly interact according to cryo-EM,²³⁴ so far it was not possible to show binding of the binary interaction via analytical SEC. Additional TSA and ITC experiments were performed with a Prp3-peptide. In the TSA experiment a shift of 1 °C could be measured, which could hint towards weak binding of the peptide to the protein (data not shown). However, the ITC experiments did not show any binding (data not shown). To further investigate these ambiguous results, protein-observed NMR could be applied. If binding could be confirmed, a titration experiment could be performed to measure the binding affinity. Based on the knowledge of the PPI binding affinity with and without follow-up compound it can be decided if further optimization towards higher binding affinity is necessary or the focus should be put on other properties such as solubility.

Overall, the identified hits offered a plethora of optimization strategies to identify follow-up compounds. Through exploring different optimization pathways, it became clear which parts of the established computational workflow Frag4Lead could be improved. Additionally, the merging ideas resulted in a higher affinity binder that reaches across the PPI and could potentially inhibit Prp3 binding. Thus, an optimized binder was designed, which in the future could be developed further into a spliceosomal PPI modulator.

6.3 Crystal packing matters in crystallographic fragment screening

CFS campaigns require a reliable crystal system that allows for high quality data and resolution better than 2.0 Å. Different methodological CFS publications inform scientists about optimizing their crystal system.^{57,187,235} However, one more very important factor has been come across the several CFS campaigns performed at BESSY II. Crystal packing might have an influence on the success of such a campaign. The target site could be occluded by a crystal contact, or the solvent channels might be too small for efficient molecule diffusion through the crystal. Thus, the crystal packing has a direct influence on the hit rate. To investigate this hypothesis further, the viral cysteine protease MPro from SARS-CoV-2 was crystallized in two different space groups and subjected to CFS. The influence of the different space groups of Mpro was also discussed in Costanzi *et al.*,²³⁶ though no direct comparison in screening endeavors had been shown. Therefore, both systems were used in CFS campaigns with the F2X-Entry Screen. In this way the importance of the space group is examined based on the hit rate and associated number of interesting starting points.

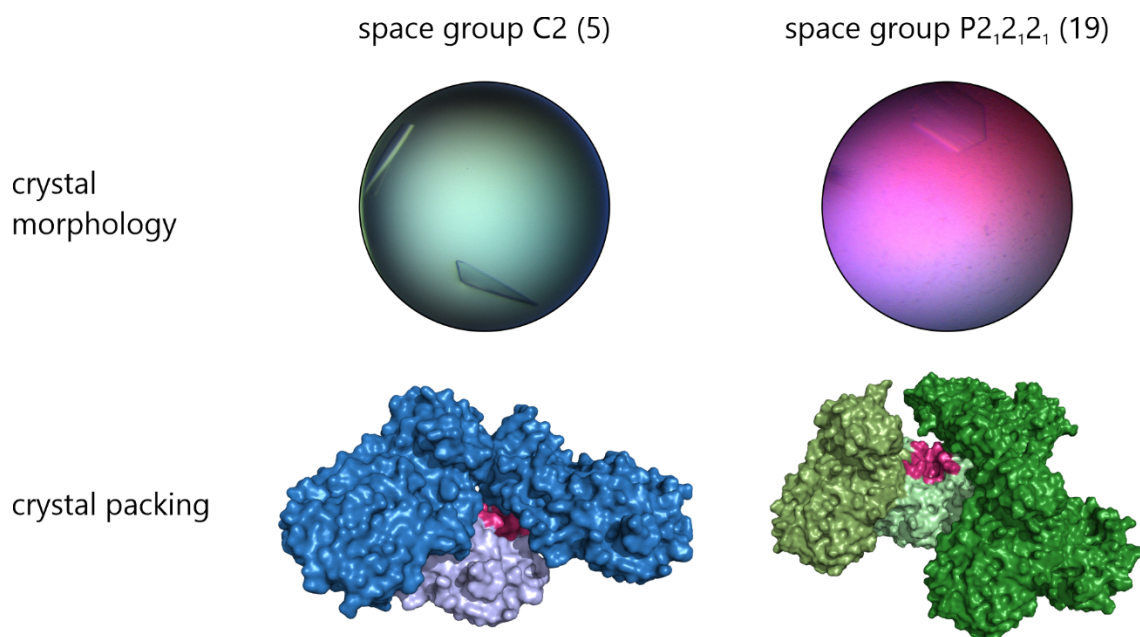


Figure 49: Comparison of MPro crystallized in two different space groups. An exemplary photo is shown of the crystal morphology for each space group. Both crystal systems result in plate-like crystals. The crystal packing is compared, with blue for space group C2 (PDB ID 6Y2E) and green for space group P2₁2₁2₁ (PDB ID 7NUK). The protein molecules of the asymmetric unit are shown in lighter colors. In case of space group P2₁2₁2₁ the monomers of the dimer are distinguished through two different light greens. Symmetry related molecules that bind close to the active site are shown in darker colors. All molecules are presented in surface view, with the active site colored in pink. The protein structure model figures have been prepared via PyMOL.⁸⁵

Several crystallization conditions were published in the PDB for two different space groups of MPro: space group number 5 (C2, monoclinic crystals) and space group number 19 (P2₁2₁2₁, orthorhombic crystals) (**Figure 49**). Both crystal systems result in plate-like crystals. To obtain orthorhombic crystals, several conditions published until May 2022 were reproduced at our site (by Laila Benz and me). Each of the tested conditions resulted in monoclinic crystals. The next step to achieve crystallization of orthorhombic crystals was to get access to orthorhombic crystal seeds. Such seeds were provided to us by the PSI and orthorhombic crystals were obtained in the same crystallization condition as for the monoclinic crystals when adding the orthorhombic crystal seeds. In case of the monoclinic crystals, the asymmetric unit includes one protein molecule, and the active site is slightly occluded by a crystal contact. The orthorhombic crystals on the other hand, have two protein molecules in the asymmetric unit and presents a highly accessible active site. This already points towards the orthorhombic crystal system being a more favorable space group for CFS than the monoclinic crystal system.

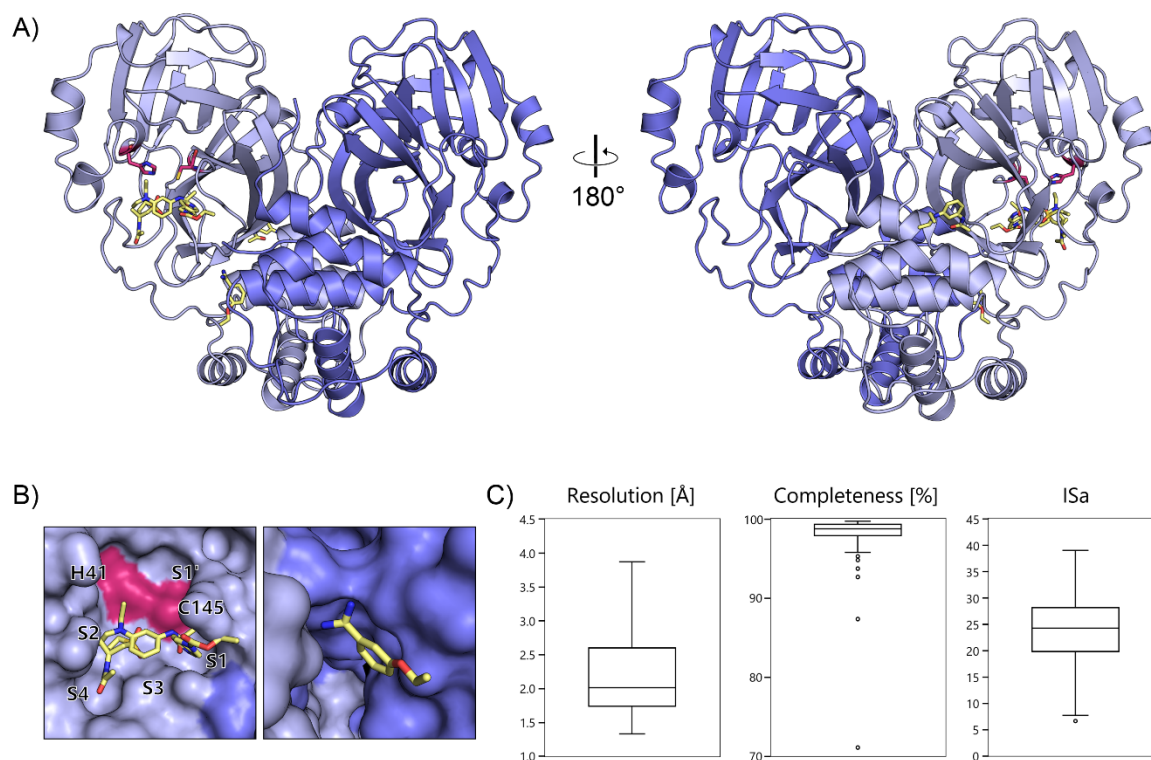


Figure 50: CFS results shown for the monoclinic crystal campaign. A) overview of all found fragment hits. The protein is shown as cartoon in blue (inhouse apo structure), both monomers of the dimer are distinguished by different blue colors. The active site His41 and Cys145 are highlighted in pink as sticks and the fragments are represented as yellow sticks. B) A zoom in to the active site binders is shown and a zoom in to the dimer binder. The figures have been prepared via PyMOL.⁸⁵ C) Boxplots are shown for typical quality indicators for CFS campaigns, the resolution, the completeness and the ISa value. Each boxplot shows the 25th and 75th percentile as a box, with the median as a line in the middle. The whiskers define the maximum and minimum. Outliers are defined as a value 1.5-times higher or lower than the sum of interquartile and 25th or 75th percentile. The boxplots were made via Microsoft Office Excel.

The CFS campaign of the F2X-Entry Screen against the MPro monoclinic system resulted in a high-quality data set (**Figure 50**). The mean resolution throughout the campaign is 2.19 Å, the mean completeness is 98.2, and the mean ISa value is 23.8. In total four fragment hits were identified, one binding at the dimer interface (P01G03) and three binding in the active site (P01B07, P01B08, P01D08). This translates into a hit rate of 4%. The active site binders cover the S1, S2, S3 and S4 pockets usually with one binder per pocket. The fragment hits are diverse and show minimal overlap, usually only by one atom. Thus, analogs or hits from other screening campaigns need to be looked at to understand the reproducibility of these binding modes. For example, the fragments found in the Covid Moonshot project could be overlaid with these identified fragments.^{105,237} The dimer interface binder is binding more closely to one of the two protein molecules and engages in interactions with both protein molecules. Therefore, an enhancer or inhibitor could be developed. As the protein is active as a dimer, a dimerization inhibitor could be more promising for future drug development. Each of the fragment hits could be used as a starting point for optimization. However, their binding modes need to be validated.

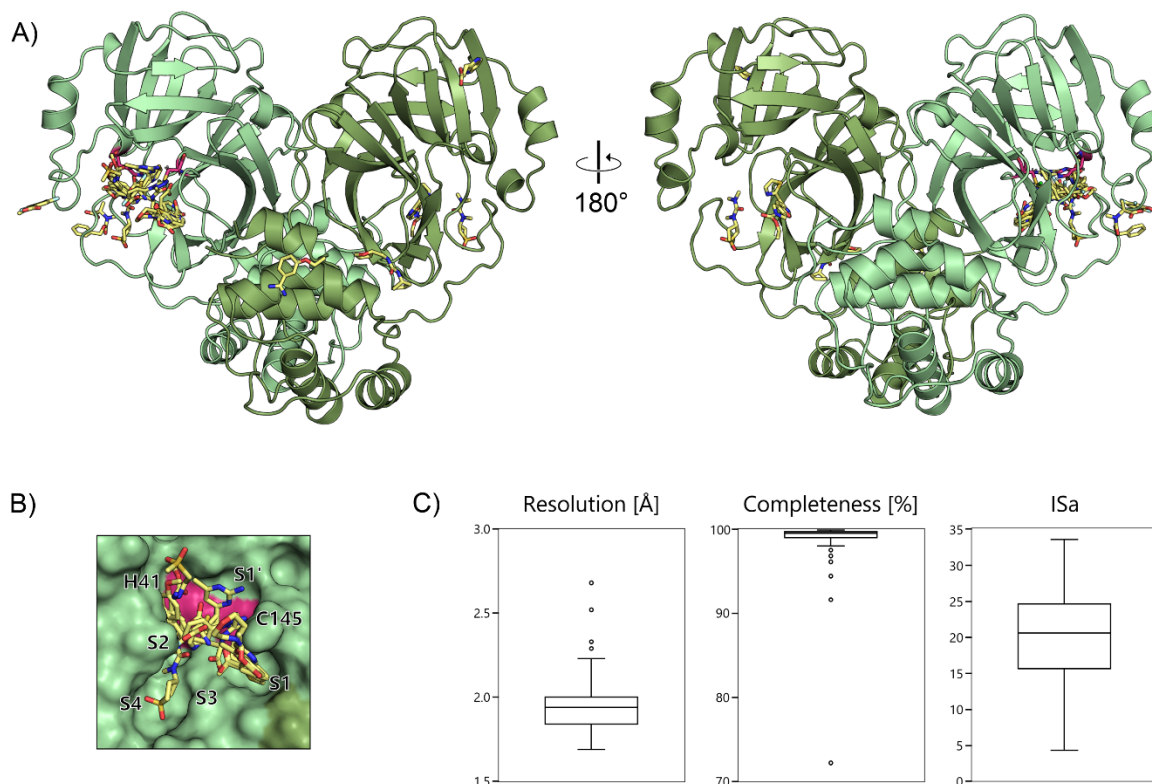


Figure 51: CFS results shown for the orthorhombic crystal campaign. A) Overview of all found fragment hits. The protein is shown as cartoon in green (inhouse apo structure), both monomers of the dimer are distinguished by different green colors. The active site His41 and Cys145 are highlighted in pink as sticks and the fragments are represented as yellow sticks. B) A zoom in to the active site binders is shown. The figures have been prepared via PyMOL.⁸⁵ C) Boxplots are shown for typical quality indicators for CFS campaigns, the resolution, the completeness and the ISa value. Each boxplot shows the 25th and 75th percentile as a box, with the median as a line in the middle. The whiskers define the maximum and minimum. Outliers are defined as a value 1.5-times lower or higher than the sum of interquartile and 25th or 75th percentile respectively. The boxplots were made via Microsoft Office Excel.

The campaign of the F2X-Entry Screen against the MPro orthorhombic system resulted in a similarly high-quality data (**Figure 51**). The mean resolution is 1.94 Å, the mean completeness is 98.9 and the mean ISa value is 20.5. From this campaign 11 fragments could be identified with 2 binding remote from the active site, 7 directly at the active site and 2 presenting two binding events, one at the active site and one remotely. This translates into a 14% hit rate, which is a typical hit rate observed so far for this fragment library. From previous experience we noticed that clustering CFS data via cluster4x, can result in more identified fragment hits.^{178,196} Therefore, the data was clustered into 3 clusters and each one had been analyzed with PanDDA. In this way 6 additional fragments could be found. Hence, the final hit rate is 18 %, with 3 fragments bound remotely (P01G03, P01H03, plus P01H11), 12 fragments bound to the active site (P01B05, P01B08, P01C10, P01D04, P01D08, P01F04, P01G09, plus P01C07, P01E11, P01G04, P01G10, P01H05) and 2 presenting two binding events (P01C02, P01D11), one at the active site and one remotely. No binders were found at the dimer interface. The active site binders explore the active site further than the monoclinic crystal campaign binders. The fragments cover the S1', S1, S2, S3 and S4 pockets with usually several binders per

pocket. Some of the fragments also overlap with their moieties. Thus, some merging opportunities are available to cover the active site. In case of the remote binders, it is not known so far if they could be interesting in further optimization efforts. Therefore, the focus is solely on the active site binders for now.

Taken together, the presented work gives a first answer to the stated hypothesis that crystal packing influences the hit rate of a CFS campaign. The campaign with the monoclinic crystal system resulted in less hits in the active site probably due to the crystal contacts close by. The mean resolution of both campaigns differs slightly, with the monoclinic campaign presenting a lower mean resolution. This should be considered too, as lower resolutions can lead to more difficulty identifying fragments. Still, due to the substantial difference in crystal packing and slight difference in resolution, the two campaigns demonstrate the influence of the crystal packing on CFS results. Therefore, users are advised in early stages of CFS projects to explore different space groups/crystal packing of their target protein. Considering this recommendation, the chance of performing highly successful CFS campaigns increases.

References

- (1) Pielak, G. J. A Model of Intracellular Organization. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (17), 5901–5902. <https://doi.org/10.1073/PNAS.0501812102/>.
- (2) Butler, K. V.; Macdonald, I. A.; Hathaway, N. A.; Jin, J. Report and Application of a Tool Compound Data Set. *J. Chem. Inf. Model.* **2017**, *57* (11), 2699–26706. <https://doi.org/10.1021/ACS.JCIM.7B00343/>.
- (3) De Esch, I. J. P.; Erlanson, D. A.; Jahnke, W.; Johnson, C. N.; Walsh, L. Fragment-to-Lead Medicinal Chemistry Publications in 2020. *J. Med. Chem.* **2022**, *65* (1), 84–99. <https://doi.org/10.1021/ACS.JMEDCHEM.1C01803/>.
- (4) Mueller, U.; Förster, R.; Hellmig, M.; Huschmann, F. U.; Kastner, A.; Malecki, P.; Pühringer, S.; Röwer, M.; Sparta, K.; Steffien, M.; Ühlein, M.; Wilk, P.; Weiss, M. S. The Macromolecular Crystallography Beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current Status and Perspectives. *Eur. Phys. J. Plus* **2015**, *1307* **2015**, *130* (7), 1–10. <https://doi.org/10.1140/EPJP/I2015-15141-2>.
- (5) Ng, R. Appendix 1: History of Drug Discovery and Development. In *Drugs: From Discovery to Approval*; John Wiley & Sons, Ltd, 2009; pp 391–397. <https://doi.org/10.1002/9780470403587.app1>.
- (6) Klebe, G. *Wirkstoffdesign: Entwurf Und Wirkung von Arzneistoffen*, 2nd ed.; Springer Spektrum, 2009.
- (7) Hughes, J. P.; Rees, S. S.; Kalindjian, S. B.; Philpott, K. L. Principles of Early Drug Discovery. *Br. J. Pharmacol.* **2011**, *162* (6), 1239–1249. <https://doi.org/10.1111/J.1476-5381.2010.01127.X>.
- (8) Tiz, B.; Bagnoli, D.; Rosati, L.; Marini, O.; Santi, F.; De Matos, A. M.; Benedetto Tiz, D.; Bagnoli, L.; Rosati, O.; Marini, F.; Santi, C.; Sancineto, L. FDA-Approved Small Molecules in 2022: Clinical Uses and Their Synthesis. *Pharm.* **2022**, *Vol. 14*, Page 2538 **2022**, *14* (11), 2538. <https://doi.org/10.3390/PHARMACEUTICS14112538>.
- (9) Wang, X.; Ni, D.; Liu, Y.; Lu, S. Rational Design of Peptide-Based Inhibitors Disrupting Protein-Protein Interactions. *Front. Chem.* **2021**, *9*, 294. <https://doi.org/10.3389/FCHEM.2021.682675/>.
- (10) Lu, H.; Zhou, Q.; He, J.; Jiang, Z.; Peng, C.; Tong, R.; Shi, J. Recent Advances in the Development of Protein–Protein Interactions Modulators: Mechanisms and Clinical Trials. *Signal Transduct. Target. Ther.* **2020**, *5* (1), 1–23. <https://doi.org/10.1038/s41392-020-00315-3>.
- (11) Beck, H.; Härter, M.; Haß, B.; Schmeck, C.; Baerfacker, L. Small Molecules and Their Impact in Drug Discovery: A Perspective on the Occasion of the 125th Anniversary of the Bayer Chemical Research Laboratory. *Drug Discov. Today* **2022**, *27* (6), 1560–1574. <https://doi.org/10.1016/J.DRUDIS.2022.02.015>.
- (12) Bull, S. C.; Doig, A. J. Properties of Protein Drug Target Classes. *PLoS One* **2015**, *10* (3). <https://doi.org/10.1371/JOURNAL.PONE.0117955>.
- (13) Antolin, A. A.; Sanfelice, D.; Crisp, A.; Villasclaras Fernandez, E.; Mica, I. L.; Chen, Y.; Collins, I.; Edwards, A.; Müller, S.; Al-Lazikani, B.; Workman, P. The Chemical Probes Portal: An Expert Review-Based Public Resource to Empower Chemical Probe Assessment, Selection and Use. *Nucleic Acids Res.* **2023**, *51* (D1), D1492. <https://doi.org/10.1093/NAR/GKAC909>.
- (14) Entzeroth, M.; Flotow, H.; Condron, P. Overview of High-Throughput Screening. *Curr. Protoc. Pharmacol.* **2009**, *44* (1), 9.4.1-9.4.27.

<https://doi.org/10.1002/0471141755.PH0904S44>.

- (15) Erlanson, D. A. Introduction to Fragment-Based Drug Discovery. *Top. Curr. Chem.* **2012**, *317*, 1–32. https://doi.org/10.1007/128_2011_180.
- (16) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science*. **1996**, *274* (5292), 1531–1534. <https://doi.org/10.1126/science.274.5292.1531>.
- (17) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **2001**, *46* (1–3), 3–26. [https://doi.org/10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0).
- (18) Collander, R. The Partition of Organic Compounds Between Higher Alcohols and Water. *Acta Schemica Scand.* **1951**, *5*, 774–780.
- (19) Benet, L. Z.; Hosey, C. M.; Ursu, O.; Oprea, T. I. BDDCS, the Rule of 5 and Drugability. *Adv. Drug Deliv. Rev.* **2016**, *101*, 89. <https://doi.org/10.1016/J.ADDR.2016.05.007>.
- (20) Shultz, M. D. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs. *J. Med. Chem.* **2019**, *62* (4), 1701–1714. <https://doi.org/10.1021/ACS.JMEDCHEM.8B00686/>.
- (21) Zhu, T.; Cao, S.; Su, P. C.; Patel, R.; Shah, D.; Chokshi, H. B.; Szukala, R.; Johnson, M. E.; Hevener, K. E. Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based Upon a Critical Literature Analysis. *J. Med. Chem.* **2013**, *56* (17), 6560. <https://doi.org/10.1021/JM301916B>.
- (22) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A ‘Rule of Three’ for Fragment-Based Lead Discovery? *Drug Discov. Today* **2003**, *8* (19), 876–877. [https://doi.org/10.1016/S1359-6446\(03\)02831-9](https://doi.org/10.1016/S1359-6446(03)02831-9).
- (23) Kirsch, P.; Hartman, A. M.; Hirsch, A. K. H.; Empting, M. Concepts and Core Principles of Fragment-Based Drug Design. *Molecules* **2019**, *24* (23). <https://doi.org/10.3390/MOLECULES24234309>.
- (24) Erlanson, D. A. *Practical Fragments: Fragments in the clinic: 2020 edition*. <https://practicalfragments.blogspot.com/2020/03/fragments-in-clinic-2020-edition.html> (accessed 2020-10-14).
- (25) *FDA Approves First Targeted Therapy for Lung Cancer Mutation Previously Considered Resistant to Drug Therapy | FDA*. <https://www.fda.gov/news-events/press-announcements/fda-approves-first-targeted-therapy-lung-cancer-mutation-previously-considered-resistant-drug> (accessed 2023-01-06).
- (26) Nakajima, E. C.; Drezner, N.; Li, X.; Mishra-Kalyani, P. S.; Liu, Y.; Zhao, H.; Bi, Y.; Liu, J.; Rahman, A.; Wearne, E.; Ojofeitimi, I.; Hotaki, L. T.; Spillman, D.; Pazdur, R.; Beaver, J. A.; Singh, H. FDA Approval Summary: Sotorasib for KRAS G12C Mutated Metastatic NSCLC. *Clin. Cancer Res.* **2022**, *28* (8), 1482. <https://doi.org/10.1158/1078-0432.CCR-21-3074>.
- (27) Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. Twenty Years on: The Impact of Fragments on Drug Discovery. *Nat. Rev. Drug Discov.* **2016**, *15* (9), 605–619. <https://doi.org/10.1038/nrd.2016.109>.
- (28) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand Efficiency: A Useful Metric for Lead Selection. *Drug Discov. Today* **2004**, *9* (10), 430–431. [https://doi.org/10.1016/S1359-6446\(04\)03069-7](https://doi.org/10.1016/S1359-6446(04)03069-7).
- (29) Ichihara, O.; Barker, J.; Law, R. J.; Whittaker, M. Compound Design by Fragment-

- Linking. *Mol. Inform.* **2011**, *30* (4), 298–306. <https://doi.org/10.1002/MINF.201000174>.
- (30) Hoffer, L.; Muller, C.; Roche, P.; Morelli, X. Chemistry-Driven Hit-to-Lead Optimization Guided by Structure-Based Approaches. *Mol. Inform.* **2018**, *37* (9–10), 1800059. <https://doi.org/10.1002/minf.201800059>.
- (31) *Practical Fragments: Fragments in the clinic: 2022 edition*. <http://practicalfragments.blogspot.com/2022/11/fragments-in-clinic-2022-edition.html> (accessed 2023-03-09).
- (32) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (6), 983–996. <https://doi.org/10.1021/C19800211>.
- (33) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50* (1), 74–82. <https://doi.org/10.1021/JM0603365/>.
- (34) Aldeghi, M.; Malhotra, S.; Selwood, D. L.; Chan, A. W. E. Two- and Three-Dimensional Rings in Drugs. *Chem. Biol. Drug Des.* **2014**, *83* (4), 450–461. <https://doi.org/10.1111/CBDD.12260>.
- (35) Carbery, A.; Skyner, R.; Delft, F. von; Deane, C. M. Fragment Libraries Designed to Be Functionally Diverse Recover Protein Binding Information More Efficiently Than Standard Structurally Diverse Libraries. *J. Med. Chem.* **2022**. <https://doi.org/10.1021/ACS.JMEDCHEM.2C01004>.
- (36) Mondal, M.; Radeva, N.; Fanlo-Virgós, H.; Otto, S.; Klebe, G.; Hirsch, A. K. H. Fragment Linking and Optimization of Inhibitors of the Aspartic Protease Endothiapepsin: Fragment-Based Drug Design Facilitated by Dynamic Combinatorial Chemistry. *Angew. Chemie Int. Ed.* **2016**, *55* (32), 9422–9426. <https://doi.org/10.1002/ANIE.201603074>.
- (37) Wood, D. J.; Lopez-Fernandez, J. D.; Knight, L. E.; Al-Khawaldeh, I.; Gai, C.; Lin, S.; Martin, M. P.; Miller, D. C.; Cano, C.; Endicott, J. A.; Hardcastle, I. R.; Noble, M. E. M.; Waring, M. J. FragLites - Minimal, Halogenated Fragments Displaying Pharmacophore Doublets. An Efficient Approach to Druggability Assessment and Hit Generation. *J. Med. Chem.* **2019**, *62* (7), 3741–3752. <https://doi.org/10.1021/ACS.JMEDCHEM.9B00304/>.
- (38) O'Reilly, M.; Cleasby, A.; Davies, T. G.; Hall, R. J.; Ludlow, R. F.; Murray, C. W.; Tisi, D.; Jhoti, H. Crystallographic Screening Using Ultra-Low-Molecular-Weight Ligands to Guide Drug Design. *Drug Discov. Today* **2019**. <https://doi.org/10.1016/j.drudis.2019.03.009>.
- (39) Owens, J. Determining Druggability. *Nat. Rev. Drug Discov.* **2007**, *6* (3), 187–187. <https://doi.org/10.1038/nrd2275>.
- (40) Thomas, S. E.; Collins, P.; James, R. H.; Mendes, V.; Charoensutthivarakul, S.; Radoux, C.; Abell, C.; Coyne, A. G.; Floto, R. A.; Von Delft, F.; Blundell, T. L. Structure-Guided Fragment-Based Drug Discovery at the Synchrotron: Screening Binding Sites and Correlations with Hotspot Mapping. *Philos. Trans. R. Soc. A* **2019**, *377* (2147). <https://doi.org/10.1098/RSTA.2018.0422>.
- (41) Martin, M. P.; Noble, M. E. M. Exiting the Tunnel of Uncertainty: Crystal Soak to Validated Hit. *Acta Crystallogr. Sect. D, Struct. Biol.* **2022**, *78* (Pt 11), 1294. <https://doi.org/10.1107/S2059798322009986>.
- (42) Schiebel, J.; Radeva, N.; Krimmer, S. G.; Wang, X.; Stieler, M.; Ehrmann, F. R.; Fu, K.; Metz, A.; Huschmann, F. U.; Weiss, M. S.; Mueller, U.; Heine, A.; Klebe, G. Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study. *ACS Chem. Biol.* **2016**, *11* (6), 1693–1701. <https://doi.org/10.1021/acschembio.5b01034>.

- (43) Blundell, T. L.; Patel, S. High-Throughput X-Ray Crystallography for Drug Discovery. *Curr. Opin. Pharmacol.* **2004**, *4* (5), 490–496. <https://doi.org/10.1016/j.coph.2004.04.007>.
- (44) Collins, P. M.; Ng, J. T.; Talon, R.; Nekrosiute, K.; Krojer, T.; Douangamath, A.; Brandao-Neto, J.; Wright, N.; Pearce, N. M.; Von Delft, F. Gentle, Fast and Effective Crystal Soaking by Acoustic Dispensing. *Acta Crystallogr. Sect. D Struct. Biol.* **2017**, *73* (3), 246–255. <https://doi.org/10.1107/S205979831700331X/>.
- (45) Huschmann, F. U.; Linnik, J.; Sparta, K.; Ühlein, M.; Wang, X.; Metz, A.; Schiebel, J.; Heine, A.; Klebe, G.; Weiss, M. S.; Mueller, U. Structures of Endothiapepsin-Fragment Complexes from Crystallographic Fragment Screening Using a Novel, Diverse and Affordable 96-Compound Fragment Library. *Acta Crystallogr. Sect. F Structural Biol. Commun.* **2016**, *72* (5), 346–355. <https://doi.org/10.1107/S2053230X16004623/>.
- (46) Deller, M. C.; Rupp, B. Approaches to Automated Protein Crystal Harvesting. *Acta Crystallogr. Sect. F* **2014**, *70* (2), 133–155. <https://doi.org/10.1107/S2053230X14000387>.
- (47) Cuttitta, C. M.; Ericson, D. L.; Scalia, A.; Roessler, C. G.; Teplitsky, E.; Joshi, K.; Campos, O.; Agarwal, R.; Allaire, M.; Orville, A. M.; Sweet, R. M.; Soares, A. S. Acoustic Transfer of Protein Crystals from Agarose Pedestals to Micromeshes for High-Throughput Screening. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2015**, *71* (1), 94–103. <https://doi.org/10.1107/S1399004714013728/>.
- (48) Cipriani, F.; Röwer, M.; Landret, C.; Zander, U.; Felisaz, F.; Márquez, J. A. CrystalDirect: A New Method for Automated Crystal Harvesting Based on Laser-Induced Photoablation of Thin Films. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68* (10), 1393–1399. <https://doi.org/10.1107/S0907444912031459/>.
- (49) Wright, N. D.; Collins, P.; Talon, R.; Nelson, E.; Koekemoer, L.; Ye, M.; Nowak, R.; Newman, J.; Tsing, J.; Mitrovich, N.; Wright, N. D.; Collins, P.; Talon, R.; Nelson, E.; Koekemoer, L.; Ye, M.; Nowak, R.; Newman, J.; Ng, J. T.; Mitrovich, N.; Wiggers, H.; von Delft, F. The Low-Cost, Semi-Automated Shifter Microscope Stage Transforms Speed and Robustness of Manual Protein Crystal Harvesting. *bioRxiv* **2019**, 2019.12.20.875674. <https://doi.org/10.1101/2019.12.20.875674>.
- (50) Márquez, J. A.; Cipriani, F. CrystalDirect™: A Novel Approach for Automated Crystal Harvesting Based on Photoablation of Thin Films. *Methods Mol. Biol.* **2014**, *1091*, 197–203. https://doi.org/10.1007/978-1-62703-691-7_14.
- (51) Förster, A.; Brandstetter, S.; Schulze-Briese, C. Transforming X-Ray Detection with Hybrid Photon Counting Detectors. *Philos. Trans. R. Soc. A* **2019**, *377* (2147), 20180241. <https://doi.org/10.1098/rsta.2018.0241>.
- (52) Leonarski, F.; Redford, S.; Mozzanica, A.; Lopez-Cuenca, C.; Panepucci, E.; Nass, K.; Ozerov, D.; Vera, L.; Oliéric, V.; Buntschu, D.; Schneider, R.; Tinti, G.; Froejdh, E.; Diederichs, K.; Bunk, O.; Schmitt, B.; Wang, M. Fast and Accurate Data Collection for Macromolecular Crystallography Using the JUNGFRÄU Detector. *Nat. Methods* **2018**, *15* (10), 799–804. <https://doi.org/10.1038/s41592-018-0143-7>.
- (53) Martiel, I.; Buntschu, D.; Meier, N.; Gobbo, A.; Panepucci, E.; Schneider, R.; Heimgartner, P.; Müller, D.; Bühlmann, K.; Birri, M.; Kaminski, J. W.; Leuenberger, J.; Oliéric, V.; Glettig, W.; Wang, M. The TELL Automatic Sample Changer for Macromolecular Crystallography. *J. Synchrotron Radiat.* **2020**, *27* (3), 860–863. <https://doi.org/10.1107/S1600577520002416/>.
- (54) Song, J.; Mathew, D.; Jacob, S. A.; Corbett, L.; Moorhead, P.; Soltis, S. M. Synchrotron Radiation Diffraction-Based Automated Crystal Centering. *J. Synchrotron Rad* **2007**, *14*, 191–195. <https://doi.org/10.1107/S0909049507004803>.

- (55) Owen, R. L.; Juanhuix, J.; Fuchs, M. Current Advances in Synchrotron Radiation Instrumentation for MX Experiments. *Arch. Biochem. Biophys.* **2016**, *602*, 21–31. <https://doi.org/10.1016/j.abb.2016.03.021>.
- (56) Bowler, M. W.; Nurizzo, D.; Barrett, R.; Beteva, A.; Bodin, M.; Caserotto, H.; Delagenière, S.; Dobias, F.; Flot, D.; Giraud, T.; Guichard, N.; Guijarro, M.; Lentini, M.; Leonard, G. A.; McSweeney, S.; Oskarsson, M.; Schmidt, W.; Snigirev, A.; Von Stetten, D.; Surr, J.; Svensson, O.; Theveneau, P.; Mueller-Dieckmann, C. MASSIF-1: A Beamline Dedicated to the Fully Automatic Characterization and Data Collection from Crystals of Biological Macromolecules. *J. Synchrotron Radiat.* **2015**, *22* (Pt 6), 1540–1547. <https://doi.org/10.1107/S1600577515016604/>.
- (57) Douangamath, A.; Powell, A.; Fearon, D.; Collins, P. M.; Talon, R.; Krojer, T.; Skyner, R.; Brandao-Neto, J.; Dunnett, L.; Dias, A.; Aimon, A.; Pearce, N. M.; Wild, C.; Gorrie-Stone, T.; von Delft, F. Achieving Efficient Fragment Screening at XChem Facility at Diamond Light Source. *J. Vis. Exp.* **2021**, No. 171. <https://doi.org/10.3791/62414>.
- (58) *K04 Ultra-high throughput beamline for MX and XChem - - Diamond Light Source.* <https://www.diamond.ac.uk/Diamond-II/New-Beamline-Proposals/first-proposals/K04-Ultra-high-throughput-beamline-for-MX-and-XChem.html> (accessed 2023-03-09).
- (59) Meents, A.; Gumprecht, L.; Fischer, P.; Meyer, J.; Günther, S.; Reinke, P.; Ewert, W.; Lieske, J.; Falke, S.; Pakendorf, T.; Hennicke, V.; Ehrh, C.; Francoual, S.; Betzel, C.; Weiss, M. S.; Rarey, M.; Chapman, H. N. HiPhaX-A Fully Automated High-Throughput Pharmaceutical X-Ray Screening Endstation at the PETRA-III Synchrotron in Hamburg.
- (60) *HiPhax has started operations.* https://bmx.cfel.de/news/2022/hiphax_has_started_operations/ (accessed 2023-03-16).
- (61) Bricogne, G.; Blanc, E.; Brandl, M.; Flensburg, C.; Keller, P.; Paciorek, W.; Roversi, P.; Sharff, A.; Smart, O.; Vonrhein, C.; Womack, T. BUSTER. **2011**, v.2.11.2.
- (62) Wojdyr, M.; Keegan, R.; Winter, G.; Ashton, A.; IUCr. DIMPLE - a Pipeline for the Rapid Generation of Difference Maps from Protein Crystals with Putatively Bound Ligands. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69* (a1), 299–299. <https://doi.org/10.1107/S0108767313097419>.
- (63) Schiebel, J.; Krimmer, S. G.; Röwer, K.; Knörlein, A.; Wang, X.; Park, A. Y.; Stieler, M.; Ehrmann, F. R.; Fu, K.; Radeva, N.; Krug, M.; Huschmann, F. U.; Glöckner, S.; Weiss, M. S.; Mueller, U.; Klebe, G.; Heine, A. High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure* **2016**, *24* (8), 1398–1409. <https://doi.org/10.1016/j.str.2016.06.010>.
- (64) Pearce, N. M.; Krojer, T.; Bradley, A. R.; Collins, P.; Nowak, R. P.; Talon, R.; Marsden, B. D.; Kelm, S.; Shi, J.; Deane, C. M.; von Delft, F. A Multi-Crystal Method for Extracting Obscured Crystallographic States from Conventionally Uninterpretable Electron Density. *Nat. Commun.* **2017**, *8*, 15123. <https://doi.org/10.1038/ncomms15123>.
- (65) de Souza Neto, L. R.; Moreira-Filho, J. T.; Neves, B. J.; Maidana, R. L. B. R.; Guimarães, A. C. R.; Furnham, N.; Andrade, C. H.; Silva, F. P. In Silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery. *Frontiers in Chemistry*. Frontiers Media S.A. February 18, 2020. <https://doi.org/10.3389/fchem.2020.00093>.
- (66) Bancet, A.; Raingeval, C.; Lomberget, T.; Le Borgne, M.; Guichou, J.-F. O.; Krimm, I. Fragment Linking Strategies for Structure-Based Drug Design. *J. Med. Chem* **2020**, *63*, 11420–11435. <https://doi.org/10.1021/acs.jmedchem.0c00242>.
- (67) Yu, H. S.; Modugula, K.; Ichihara, O.; Kramschuster, K.; Keng, S.; Abel, R.; Wang, L. General Theory of Fragment Linking in Molecular Design: Why Fragment Linking Rarely

- Succeeds and How to Improve Outcomes. *J. Chem. Theory Comput.* **2020**, *17*, 462. <https://doi.org/10.1021/acs.jctc.0c01004>.
- (68) Patra, J.; Singh, D.; Jain, S.; Mahindroo, N. Application of Docking for Lead Optimization. *Mol. Docking Comput. Drug Des. Fundam. Tech. Resour. Appl.* **2021**, 271–294. <https://doi.org/10.1016/B978-0-12-822312-3.00012-6>.
- (69) Beroza, P.; Crawford, J. J.; Ganichkin, O.; Gendele, L.; Harris, S. F.; Klein, R.; Miu, A.; Steinbacher, S.; Klingler, F. M.; Lemmen, C. Chemical Space Docking Enables Large-Scale Structure-Based Virtual Screening to Discover ROCK1 Kinase Inhibitors. *Nat. Commun.* **2022**, *13* (1), 1–10. <https://doi.org/10.1038/s41467-022-33981-8>.
- (70) de Ruyck, J.; Brysbaert, G.; Blossey, R.; Lensink, M. F. Molecular Docking as a Popular Tool in Drug Design, an in Silico Travel. *Adv. Appl. Bioinforma. Chem.* **2016**, *9* (1), 1–11. <https://doi.org/10.2147/AABC.S105289>.
- (71) Bender, B. J.; Gahbauer, S.; Lutten, A.; Lyu, J.; Webb, C. M.; Stein, R. M.; Fink, E. A.; Balus, T. E.; Carlsson, J.; Irwin, J. J.; Shoichet, B. K. A Practical Guide to Large-Scale Docking. *Nat. Protoc.* **2021**, *16* (10), 4799–4832. <https://doi.org/10.1038/s41596-021-00597-z>.
- (72) Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided. Drug Des.* **2011**, *7* (2), 146. <https://doi.org/10.2174/157340911795677602>.
- (73) Schiebel, J.; Gaspari, R.; Wulsdorf, T.; Ngo, K.; Sohn, C.; Schrader, T. E.; Cavalli, A.; Ostermann, A.; Heine, A.; Klebe, G. Intriguing Role of Water in Protein-Ligand Binding Studied by Neutron Crystallography on Trypsin Complexes. *Nat. Commun.* **2018**, *9* (1), 1–15. <https://doi.org/10.1038/s41467-018-05769-2>.
- (74) Roberts, B. C.; Mancera, R. L. Ligand - Protein Docking with Water Molecules. *J. Chem. Inf. Model.* **2008**, *48* (2), 397–408. <https://doi.org/10.1021/CI700285E/>.
- (75) BioSolveIT. SeeSAR SeeSAR version 11.2.2; BioSolveIT GmbH, Sankt Augustin, Germany, 2022, www.biosolveit.de/SeeSAR. <https://www.biosolveit.de/SeeSAR/> (accessed 2020-05-05).
- (76) Schmidt, R.; Klein, R.; Rarey, M. Maximum Common Substructure Searching in Combinatorial Make-on-Demand Compound Spaces. *J. Chem. Inf. Model.* **2021**, *2022*. <https://doi.org/10.1021/ACS.JCIM.1C00640/>.
- (77) Yuriev, E.; Ramsland, P. A. Latest Developments in Molecular Docking: 2010–2011 in Review. *J. Mol. Recognit.* **2013**, *26* (5), 215–239. <https://doi.org/10.1002/JMR.2266>.
- (78) Schmidt, U.; Struck, S.; Gruening, B.; Hossbach, J.; Jaeger, I. S.; Parol, R.; Lindequist, U.; Teuscher, E.; Preissner, R. SuperToxic: A Comprehensive Database of Toxic Compounds. *Nucleic Acids Res.* **2009**, *37*, D295. <https://doi.org/10.1093/NAR/GKN850>.
- (79) Baell, J. B.; Nissink, J. W. M. Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017 - Utility and Limitations. *ACS Chem. Biol.* **2018**, *13* (1), 36–44. <https://doi.org/10.1021/ACSCHEMBIO.7B00903/>.
- (80) Jain, A. N. Scoring Functions for Protein-Ligand Docking. *Curr. Protein Pept. Sci.* **2006**, *7* (5), 407–420. <https://doi.org/10.2174/138920306778559395>.
- (81) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49* (20), 5912–5931. <https://doi.org/10.1021/JM050362N/>.

- (82) Rarey, M.; Wefing, S.; Lengauer, T. Placement of Medium-Sized Molecular Fragments into Active Sites of Proteins. *J. Comput. Aided. Mol. Des.* **1996**, *10*, 41–54.
- (83) Neudert, G.; Klebe, G. DSX: A Knowledge-Based Scoring Function for the Assessment of Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2011**, *51* (10), 2731–2745. <https://doi.org/10.1021/CI200274Q/>.
- (84) Schneider, N.; Lange, G.; Hindle, S.; Klein, R.; Rarey, M. A Consistent Description of HYdrogen Bond and DEhydration Energies in Protein-Ligand Complexes: Methods behind the HYDE Scoring Function. *J. Comput. Aided. Mol. Des.* **2013**, *27* (1), 15–29. <https://doi.org/10.1007/s10822-012-9626-2>.
- (85) Schrödinger, L. *The PyMOL Molecular Graphics System, Version 2.4*; 2020.
- (86) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. <https://doi.org/10.1002/JCC.20084>.
- (87) Fischer, A.; Smiesko, M.; Sellner, M.; Lill, M. A. Decision Making in Structure-Based Drug Discovery: Visual Inspection of Docking Results. *J. Med. Chem.* **2020**. <https://doi.org/10.1021/acs.jmedchem.0c02227>.
- (88) Metz, A.; Wollenhaupt, J.; Glöckner, S.; Messini, N.; Huber, S.; Barthel, T.; Merabet, A.; Gerber, H. D.; Heine, A.; Klebe, G.; Weiss, M. S. Frag4Lead: Growing Crystallographic Fragment Hits by Catalog Using Fragment-Guided Template Docking. *Acta Crystallogr. Sect. D Struct. Biol.* **2021**, *77* (9), 1168–1182. <https://doi.org/10.1107/S2059798321008196/>.
- (89) Hopkins, A. L.; Groom, C. R. The Druggable Genome. *Nat. Rev. Drug Discov.* **2002**, *1* (9), 727–730. <https://doi.org/10.1038/nrd892>.
- (90) Kumaran, D.; Rawat, R.; Ahmed, S. A.; Swaminathan, S. Substrate Binding Mode and Its Implication on Drug Design for Botulinum Neurotoxin A. *PLoS Pathog.* **2008**, *4* (9), 1000165. <https://doi.org/10.1371/JOURNAL.PPAT.1000165>.
- (91) Verma, S.; Dixit, R.; Pandey, K. C. Cysteine Proteases: Modes of Activation and Future Prospects as Pharmacological Targets. *Front. Pharmacol.* **2016**, *7* (APR). <https://doi.org/10.3389/FPHAR.2016.00107>.
- (92) López-Otín, C.; Bond, J. S. Proteases: Multifunctional Enzymes in Life and Disease. *J. Biol. Chem.* **2008**, *283* (45), 30433. <https://doi.org/10.1074/JBC.R800035200>.
- (93) Hartman, A. M.; Mondal, M.; Radeva, N.; Klebe, G.; Hirsch, A. K. H. Structure-Based Optimization of Inhibitors of the Aspartic Protease Endothiapepsin. *Int. J. Mol. Sci.* **2015**, *16* (8), 19184. <https://doi.org/10.3390/IJMS160819184>.
- (94) Blundell, T.; Jenkins, J.; Pearl, L.; Sewell, T.; Pedersen, V. The High Resolution Structure of Endothiapepsin. *Aspartic Proteinases Their Inhib.* **1985**, 151–162. <https://doi.org/10.1515/9783111649788-019>.
- (95) Coates, L.; Tuan, H. F.; Tomanicek, S.; Kovalevsky, A.; Mustyakimov, M.; Erskine, P.; Cooper, J. The Catalytic Mechanism of an Aspartic Proteinase Explored with Neutron and X-Ray Diffraction. *J. Am. Chem. Soc.* **2008**, *130* (23), 7235–7237. <https://doi.org/10.1021/JA801269X/>.
- (96) Radeva, N.; Krimmer, S. G.; Stieler, M.; Fu, K.; Wang, X.; Ehrmann, F. R.; Metz, A.; Huschmann, F. U.; Weiss, M. S.; Mueller, U.; Schiebel, J.; Heine, A.; Klebe, G. Experimental Active-Site Mapping by Fragments: Hot Spots Remote from the Catalytic Center of Endothiapepsin. *J. Med. Chem.* **2016**, *59* (16), 7561–7575. <https://doi.org/10.1021/ACS.JMEDCHEM.6B00645/>.

- (97) Vasiljeva, O.; Reinheckel, T.; Peters, C.; Turk, D.; Turk, V.; Turk, B. Emerging Roles of Cysteine Cathepsins in Disease and Their Potential as Drug Targets. *Curr. Pharm. Des.* **2007**, *13* (4), 387–403. <https://doi.org/10.2174/138161207780162962>.
- (98) Hu, Q.; Xiong, Y.; Zhu, G. H.; Zhang, Y. N.; Zhang, Y. W.; Huang, P.; Ge, G. B. The SARS-CoV-2 Main Protease (Mpro): Structure, Function, and Emerging Therapies for COVID-19. *MedComm* **2022**, *3* (3), e151. <https://doi.org/10.1002/MCO2.151>.
- (99) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α -Ketoamide Inhibitors. *Science (80-.)*. **2020**, *368* (6489), 409–412. <https://doi.org/10.1126/science.abb3405>.
- (100) Kneller, D. W.; Phillips, G.; O'Neill, H. M.; Jedrzejczak, R.; Stols, L.; Langan, P.; Joachimiak, A.; Coates, L.; Kovalevsky, A. Structural Plasticity of SARS-CoV-2 3CL Mpro Active Site Cavity Revealed by Room Temperature X-Ray Crystallography. *Nat. Commun.* **2020**, *11* (1), 1–6. <https://doi.org/10.1038/s41467-020-16954-7>.
- (101) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H.; Shanghai, in; Lab, Z. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nat.* **2020**, *582* (7811), 289–293. <https://doi.org/10.1038/s41586-020-2223-y>.
- (102) Günther, S.; Reinke, P. Y. A.; Fernández-García, Y.; Lieske, J.; Lane, T. J.; Ginn, H. M.; Koua, F. H. M.; Ehart, C.; Ewert, W.; Oberthuer, D.; Yefanov, O.; Meier, S.; Lorenzen, K.; Krichel, B.; Kopicki, J. D.; Gelisio, L.; Brehm, W.; Dunkel, I.; Seychell, B.; Gieseler, H.; Norton-Baker, B.; Escudero-Pérez, B.; Domaracky, M.; Saouane, S.; Tolstikova, A.; White, T. A.; Hänle, A.; Groessler, M.; Fleckenstein, H.; Trost, F.; Galchenkova, M.; Gevorkov, Y.; Li, C.; Awel, S.; Peck, A.; Barthelmess, M.; Schlünzen, F.; Xavier, P. L.; Werner, N.; Andaleeb, H.; Ullah, N.; Falke, S.; Srinivasan, V.; França, B. A.; Schwinzer, M.; Brognaro, H.; Rogers, C.; Melo, D.; Zaitseva-Doyle, J. J.; Knoska, J.; Penã-Murillo, G. E.; Mashhour, A. R.; Hennicke, V.; Fischer, P.; Hakanpaä, J.; Meyer, J.; Gribbon, P.; Ellinger, B.; Kuzikov, M.; Wolf, M.; Beccari, A. R.; Bourenkov, G.; Stetten, D. Von; Pompidor, G.; Bento, I.; Panneerselvam, S.; Karpics, I.; Schneider, T. R.; Garcia-Alai, M. M.; Niebling, S.; Günther, C.; Schmidt, C.; Schubert, R.; Han, H.; Boger, J.; Monteiro, D. C. F.; Zhang, L.; Sun, X.; Pletzer-Zelgert, J.; Wollenhaupt, J.; Feiler, C. G.; Weiss, M. S.; Schulz, E. C.; Mehrabi, P.; Karnıcar, K.; Usenik, A.; Loboda, J.; Tidow, H.; Chari, A.; Hilgenfeld, R.; Uetrech, C.; Cox, R.; Zaliani, A.; Beck, T.; Rarey, M.; Günther, S.; Turk, D.; Hinrichs, W.; Chapman, H. N.; Pearson, A. R.; Betzel, C.; Meents, A. X-Ray Screening Identifies Active Site and Allosteric Inhibitors of SARS-CoV-2 Main Protease. *Science* **2021**, *372* (6542), 642–646. <https://doi.org/10.1126/SCIENCE.ABF7945>.
- (103) Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.; Owen, C. D.; Resnick, E.; Strain-Damerell, C.; Aimon, A.; Ábrányi-Balogh, P.; Brandão-Neto, J.; Carbery, A.; Davison, G.; Dias, A.; Downes, T. D.; Dunnett, L.; Fairhead, M.; Firth, J. D.; Jones, S. P.; Keeley, A.; Keserü, G. M.; Klein, H. F.; Martin, M. P.; Noble, M. E. M.; O'Brien, P.; Powell, A.; Reddi, R. N.; Skyner, R.; Snee, M.; Waring, M. J.; Wild, C.; London, N.; von Delft, F.; Walsh, M. A. Crystallographic and Electrophilic Fragment Screening of the SARS-CoV-2 Main Protease. *Nat. Commun.* **2020**, *11* (1), 1–11. <https://doi.org/10.1038/s41467-020-18709-w>.
- (104) Luttens, A.; Gullberg, H.; Abdurakhmanov, E.; Vo, D. D.; Akaberi, D.; Talibov, V. O.; Nekhotiaeva, N.; Vangeel, L.; De Jonghe, S.; Jochmans, D.; Krambrich, J.; Tas, A.; Lundgren, B.; Gravenfors, Y.; Craig, A. J.; Atilaw, Y.; Sandström, A.; Moodie, L. W. K.; Lundkvist, Å.; van Hemert, M. J.; Neyts, J.; Lennerstrand, J.; Kihlberg, J.; Sandberg, K.; Danielson, U. H.; Carlsson, J. Ultralarge Virtual Screening Identifies SARS-CoV-2 Main

- Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. *J. Am. Chem. Soc.* **2022**, *144* (7), 2905–2920. <https://doi.org/10.1021/JACS.1C08402/>.
- (105) Consortium, T. C. M.; Chodera, J.; Lee, A.; London, N.; Delft, F. von. COVID Moonshot: Open Science Discovery of SARS-CoV-2 Main Protease Inhibitors by Combining Crowdsourcing, High-Throughput Experiments, Computational Simulations, and Machine Learning. **2020**. <https://doi.org/10.26434/CHEMRXIV.13158218.V1>.
- (106) Cohen, P.; Cross, D.; Jänne, P. A. Kinase Drug Discovery 20 Years after Imatinib: Progress and Future Directions. *Nat. Rev. Drug Discov.* **2021**, *20* (7), 551–569. <https://doi.org/10.1038/s41573-021-00195-4>.
- (107) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How Many Drug Targets Are There? *Nat. Rev. Drug Discov.* **2006**, *5* (12), 993–996. <https://doi.org/10.1038/nrd2199>.
- (108) Laraia, L.; McKenzie, G.; Spring, D. R.; Venkitaraman, A. R.; Huggins, D. J. Overcoming Chemical, Biological, and Computational Challenges in the Development of Inhibitors Targeting Protein-Protein Interactions. *Chem. Biol.* **2015**, *22* (6), 689–703. <https://doi.org/10.1016/J.CHEMBIOL.2015.04.019>.
- (109) Nero, T. L.; Morton, C. J.; Holien, J. K.; Wielens, J.; Parker, M. W. Oncogenic Protein Interfaces: Small Molecules, Big Challenges. *Nat. Rev. Cancer* **2014**, *14* (4), 248–262. <https://doi.org/10.1038/nrc3690>.
- (110) Venkatesan, K.; Rual, J. F.; Vazquez, A.; Stelzl, U.; Lemmens, I.; Hirozane-Kishikawa, T.; Hao, T.; Zenkner, M.; Xin, X.; Goh, K. II; Yildirim, M. A.; Simonis, N.; Heinzmann, K.; Gebreab, F.; Sahalie, J. M.; Cevik, S.; Simon, C.; de Smet, A. S.; Dann, E.; Smolyar, A.; Vinayagam, A.; Yu, H.; Szeto, D.; Borick, H.; Dricot, A.; Klitgord, N.; Murray, R. R.; Lin, C.; Lalowski, M.; Timm, J.; Rau, K.; Boone, C.; Braun, P.; Cusick, M. E.; Roth, F. P.; Hill, D. E.; Tavernier, J.; Wanker, E. E.; Barabási, A. L.; Vidal, M. An Empirical Framework for Binary Interactome Mapping. *Nat. Methods* **2008**, *6* (1), 83–90. <https://doi.org/10.1038/nmeth.1280>.
- (111) Wiedmann, M. M.; Tan, Y. S.; Wu, Y.; Aibara, S.; Xu, W.; Sore, H. F.; Verma, C. S.; Itzhaki, L.; Stewart, M.; Brenton, J. D.; Spring, D. R. Development of Cell-Permeable, Non-Helical Constrained Peptides to Target a Key Protein–Protein Interaction in Ovarian Cancer. *Angew. Chem. Int. Ed. Engl.* **2017**, *56* (2), 524–529. <https://doi.org/10.1002/anie.201609427>.
- (112) Watson, P. J.; Fairall, L.; Santos, G. M.; Schwabe, J. W. R. Structure of HDAC3 Bound to Co-Repressor and Inositol Tetraphosphate. *Nature* **2012**, *481* (7381), 335–340. <https://doi.org/10.1038/nature10728>.
- (113) Grainger, R. J.; Beggs, J. D. Prp8 Protein: At the Heart of the Spliceosome. *RNA* **2005**, *11* (5), 533–557. <https://doi.org/10.1261/rna.2220705>.
- (114) Bustos, D. M. The Role of Protein Disorder in the 14-3-3 Interaction Network. *Mol. Biosyst.* **2011**, *8* (1), 178–184. <https://doi.org/10.1039/C1MB05216K>.
- (115) Ni, D.; Lu, S.; Zhang, J. Emerging Roles of Allosteric Modulators in the Regulation of Protein-Protein Interactions (PPIs): A New Paradigm for PPI Drug Discovery. *Med. Res. Rev.* **2019**, *39* (6), 2314–2342. <https://doi.org/10.1002/MED.21585>.
- (116) Stevers, L. M.; Sijbesma, E.; Botta, M.; Mackintosh, C.; Obsil, T.; Landrieu, I.; Cau, Y.; Wilson, A. J.; Karawajczyk, A.; Eickhoff, J.; Davis, J.; Hann, M.; O'Mahony, G.; Doveston, R. G.; Brunsveld, L.; Ottmann, C. Modulators of 14-3-3 Protein-Protein Interactions. *J. Med. Chem.* **2018**, *61* (9), 3755–3778. <https://doi.org/10.1021/acs.jmedchem.7b00574>.

- (117) Juárez-Salcedo, L. M.; Desai, V.; Dalia, S. Venetoclax: Evidence to Date and Clinical Potential. *Drugs Context* **2019**, *8*. <https://doi.org/10.7573/DIC.212574>.
- (118) Schreiber, K.; Csaba, G.; Haslbeck, M.; Zimmer, R. Alternative Splicing in next Generation Sequencing Data of *Saccharomyces Cerevisiae*. *PLoS One* **2015**, *10* (10), 1–18. <https://doi.org/10.1371/journal.pone.0140487>.
- (119) Wahl, M. C.; Will, C. L.; Lührmann, R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **2009**, *136* (4), 701–718. <https://doi.org/10.1016/j.cell.2009.02.009>.
- (120) Plaschka, C.; Newman, A. J.; Nagai, K. Structural Basis of Nuclear Pre-mRNA Splicing: Lessons from Yeast. *Cold Spring Harb. Perspect. Biol.* **2019**, *11* (5). <https://doi.org/10.1101/CSHPERSPECT.A032391>.
- (121) Callaway, E. Revolutionary Cryo-EM Is Taking over Structural Biology. *Nature* **2020**, *578* (7794), 201. <https://doi.org/10.1038/D41586-020-00341-9>.
- (122) Benjin, X.; Ling, L. Developments, Applications, and Prospects of Cryo-electron Microscopy. *Protein Sci.* **2020**, *29* (4), 872. <https://doi.org/10.1002/PRO.3805>.
- (123) Kastner, B.; Will, C. L.; Stark, H.; Lührmann, R. Structural Insights into Nuclear Pre-mRNA Splicing in Higher Eukaryotes. *Cold Spring Harb. Perspect. Biol.* **2019**, *11* (11), a032417. <https://doi.org/10.1101/CSHPERSPECT.A032417>.
- (124) Fica, S. M. Cryo-EM Snapshots of the Human Spliceosome Reveal Structural Adaptions for Splicing Regulation. *Curr. Opin. Struct. Biol.* **2020**, *65*, 139–148. <https://doi.org/10.1016/J.SBI.2020.06.018>.
- (125) Scotti, M. M.; Swanson, M. S. RNA Mis-Splicing in Disease. *Nat. Rev. Genet.* **2015**, *17* (1), 19–32. <https://doi.org/10.1038/nrg.2015.3>.
- (126) Li, D.; McIntosh, C. S.; Mastaglia, F. L.; Wilton, S. D.; Aung-Htut, M. T. Neurodegenerative Diseases: A Hotbed for Splicing Defects and the Potential Therapies. *Transl. Neurodegener.* **2021**, *10* (1), 1–18. <https://doi.org/10.1186/S40035-021-00240-7>.
- (127) Nakajima, H.; Sato, B.; Fujita, T.; Takase, S.; Terano, H.; Okuhara, M. New Antitumor Substances, FR901463, FR901464 and FR901465 I. Taxonomy, Fermentation, Isolation, Physico-Chemical Properties and Biological Activities. *J. Antibiot. (Tokyo)*. **1996**, *49* (12), 1196–1203. <https://doi.org/10.7164/ANTIBIOTICS.49.1196>.
- (128) Kaida, D.; Motoyoshi, H.; Tashiro, E.; Nojima, T.; Hagiwara, M.; Ishigami, K.; Watanabe, H.; Kitahara, T.; Yoshida, T.; Nakajima, H.; Tani, T.; Horinouchi, S.; Yoshida, M. Spliceostatin A Targets SF3b and Inhibits Both Splicing and Nuclear Retention of Pre-mRNA. *Nat. Chem. Biol.* **2007**, *3* (9), 576–583. <https://doi.org/10.1038/nchembio.2007.18>.
- (129) Mizui, Y.; Sakai, T.; Iwata, M.; Uenaka, T.; Okamoto, K.; Shimizu, H.; Yamori, T.; Yoshimatsu, K.; Asada, M. Pladienolides, New Substances from Culture of *Streptomyces Platensis* Mer-11107 III. *In Vitro* and *In Vivo* Antitumor Activities. *J. Antibiot. (Tokyo)*. **2004**, *57* (3), 188–196. <https://doi.org/10.7164/ANTIBIOTICS.57.188>.
- (130) Miller-Wideman, M.; Makkar, N.; Tran, M.; Isaac, B.; Biest, N.; Stonard, R. HERBOXIDIENE, A NEW HERBICIDAL SUBSTANCE FROM *Streptomyces Chromofuscus* A7847 TAXONOMY, FERMENTATION, ISOLATION, PHYSICO-CHEMICAL AND BIOLOGICAL PROPERTIES. *J. Antibiot. (Tokyo)*. **1992**, *45* (6), 914–921. <https://doi.org/10.7164/ANTIBIOTICS.45.914>.
- (131) Webb, T. R.; Joyner, A. S.; Potter, P. M. The Development and Application of Small Molecule Modulators of SF3b as Therapeutic Agents for Cancer. *Drug Discov. Today* **2013**, *18* (1–2), 43. <https://doi.org/10.1016/J.DRUDIS.2012.07.013>.

- (132) Kumar, D.; Kashyap, M. K.; La Clair, J. J.; Villa, R.; Spaanderman, I.; Chien, S.; Rassenti, L. Z.; Kipps, T. J.; Burkart, M. D.; Castro, J. E. Selectivity in Small Molecule Splicing Modulation. *ACS Chem. Biol.* **2016**, *11* (10), 2716–2723. <https://doi.org/10.1021/acscchembio.6b00399>.
- (133) Larsen, N. A.; Larsen, N. A. The SF3b Complex Is an Integral Component of the Spliceosome and Targeted by Natural Product-Based Inhibitors. **2021**. https://doi.org/10.1007/978-3-030-58971-4_12.
- (134) Effenberger, K. A.; Urabe, V. K.; Jurica, M. S. Modulating Splicing with Small Molecular Inhibitors of the Spliceosome HHS Public Access. *Wiley Interdiscip Rev RNA* **2017**, *8* (2). <https://doi.org/10.1002/wrna.1381>.
- (135) Lopez, A. G.; Allu, S. R.; Mendez, P.; Reddy, G. C.; Maul-newby, H. M.; Ghosh, A. K.; Jurica, M. S. Herboxidiene Features That Mediate Conformation-Dependent SF3B1 Interactions to Inhibit Splicing. *J. Med. Chem.* **2021**. <https://doi.org/10.1021/acscchembio.0c00965>.
- (136) Seiler, M.; Yoshimi, A.; Darman, R.; Chan, B.; Keaney, G.; Thomas, M.; Agrawal, A. A.; Caleb, B.; Csibi, A.; Sean, E.; Fekkes, P.; Karr, C.; Klimek, V.; Lai, G.; Lee, L.; Kumar, P.; Lee, S. C. W.; Liu, X.; MacKenzie, C.; Meeske, C.; Mizui, Y.; Padron, E.; Park, E.; Pazolli, E.; Peng, S.; Prajapati, S.; Taylor, J.; Teng, T.; Wang, J.; Warmuth, M.; Yao, H.; Yu, L.; Zhu, P.; Abdel-Wahab, O.; Smith, P. G.; Buonamici, S. H3B-8800, an Orally Available Small-Molecule Splicing Modulator, Induces Lethality in Spliceosome-Mutant Cancers. *Nat. Med.* **2018**, *24* (4), 497–504. <https://doi.org/10.1038/nm.4493>.
- (137) FDA grants orphan drug status to H3 Biomedicine's cancer treatment H3B-8800 - Pharmaceutical Technology. <https://www.pharmaceutical-technology.com/news/newsfda-grants-orphan-drug-status-to-h3-biomedicines-cancer-treatment-h3b-8800-5900828/> (accessed 2023-03-09).
- (138) Galej, W. P.; Oubridge, C.; Newman, A. J.; Nagai, K. Crystal Structure of Prp8 Reveals Active Site Cavity of the Spliceosome. *Nature* **2013**, *493* (7434), 638–643. <https://doi.org/10.1038/nature11843>.
- (139) Weber, G.; Cristão, V. F.; Santos, K. F.; Jovin, S. M.; Heroven, A. C.; Holton, N.; Lührmann, R.; Beggs, J. D.; Wahl, M. C. Structural Basis for Dual Roles of Aar2p in U5 SnRNP Assembly. *Genes Dev.* **2013**, *27* (5), 525–540. <https://doi.org/10.1101/gad.213207.113>.
- (140) Preussner, M.; Santos, K. F.; Alles, J.; Heroven, C.; Heyd, F.; Wahl, M. C.; Weber, G. Structural and Functional Investigation of the Human SnRNP Assembly Factor AAR2 in Complex with the RNase H-like Domain of PRPF8. *Acta Crystallogr. Sect. D, Struct. Biol.* **2022**, *78* (Pt 11), 1373. <https://doi.org/10.1107/S2059798322009755>.
- (141) Shaw Stewart, P.; Mueller-Dieckmann, J. Automation in Biological Crystallization. *Acta Crystallogr. Sect. F, Struct. Biol. Commun.* **2014**, *70* (Pt 6), 686. <https://doi.org/10.1107/S2053230X14011601>.
- (142) Bowler, M. W.; Svensson, O.; Nurizzo, D. Fully Automatic Macromolecular Crystallography: The Impact of MASSIF-1 on the Optimum Acquisition and Quality of Data. *Crystallogr. Rev.* **2016**, *22* (4), 233–249. <https://doi.org/10.1080/0889311X.2016.1155050>.
- (143) Powell, H. R. X-Ray Data Processing. *Biosci. Rep.* **2017**, *37* (5). <https://doi.org/10.1042/BSR20170227>.
- (144) Krojer, T.; Talon, R.; Pearce, N.; Collins, P.; Douangamath, A.; Brandao-Neto, J.; Dias, A.; Marsden, B.; Von Delft, F. The XChemExplorer Graphical Workflow Tool for Routine or Large-Scale Protein-Ligand Structure Determination. *Acta Crystallogr. Sect. D Struct.*

Biol. **2017**, 73 (Pt 3), 267–278. <https://doi.org/10.1107/S2059798316020234/>.

- (145) Lima, G. M. A.; Talibov, V. O.; Jagudin, E.; Sele, C.; Nyblom, M.; Knecht, W.; Logan, D. T.; Sjögren, T.; Mueller, U. FragMAX: The Fragment-Screening Platform at the MAX IV Laboratory. *Acta Crystallogr. Sect. D* **2020**, 76 (8), 771–777. <https://doi.org/10.1107/S205979832000889X>.
- (146) Douangamath, A.; Brandao-Neto, J.; Williams, M.; Fearn, R.; Krojer, T.; Aller, P.; Henderson, K.; Longhi, E.; Allianelli, L.; von Delft, F. I04-1, a Future “fragment Screening” Beamline Facility at Diamond Light Source.
- (147) Barthel, T.; Huschmann, F. U.; Wallacher, D.; Feiler, C. G.; Klebe, G.; Weiss, M. S.; Wollenhaupt, J. Facilitated Crystal Handling Using a Simple Device for Evaporation Reduction in Microtiter Plates. *J. Appl. Crystallogr.* **2021**, 54 (1), 376–382. <https://doi.org/10.1107/s1600576720016477>.
- (148) Wheeler, M. J.; Russi, S.; Bowler, M. G.; Bowler, M. W. Measurement of the Equilibrium Relative Humidity for Common Precipitant Concentrations: Facilitating Controlled Dehydration Experiments. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2012**, 68 (1), 111–114. <https://doi.org/10.1107/S1744309111054029>.
- (149) Bowler, M. W.; Mueller, U.; Weiss, M. S.; Sanchez-Weatherby, J.; Sorensen, T. L. M.; Thunnissen, M. M. G. M.; Ursby, T.; Gobbo, A.; Russi, S.; Bowler, M. G.; Brockhauser, S.; Svensson, O.; Cipriani, F. Automation and Experience of Controlled Crystal Dehydration: Results from the European Synchrotron Hc1 Collaboration. *Cryst. Growth Des.* **2015**, 15 (3), 1043–1054. <https://doi.org/10.1021/cg500890r>.
- (150) Bowler, M. G.; Bowler, D. R.; Bowler, M. W. Raoult’s Law Revisited: Accurately Predicting Equilibrium Relative Humidity Points for Humidity Control Experiments. *J. Appl. Cryst* **2017**, 50, 631–638. <https://doi.org/10.1107/S1600576717003636>.
- (151) Ursby, T.; Åhnberg, K.; Appio, R.; Aurelius, O.; Barczyk, A.; Bartalesi, A.; Bjelčić, M.; Bolmsten, F.; Cerenius, Y.; Doak, R. B.; Eguiraun, M.; Eriksson, T.; Friel, R. J.; Gorgisyan, I.; Gross, A.; Haghighat, V.; Hennies, F.; Jagudin, E.; Norsk Jensen, B.; Jeppsson, T.; Kloos, M.; Lidon-Simon, J.; de Lima, G. M. A.; Lizatovic, R.; Lundin, M.; Milan-Otero, A.; Milas, M.; Nan, J.; Nardella, A.; Rosborg, A.; Shilova, A.; Shoeman, R. L.; Siewert, F.; Sondhauss, P.; Talibov, V. O.; Tarawneh, H.; Thånell, J.; Thunnissen, M.; Unge, J.; Ward, C.; Gonzalez, A.; Mueller, U.; IUCr. BioMAX – the First Macromolecular Crystallography Beamline at MAX IV Laboratory. *J. Synchrotron Rad.* **2020**, 27 (5), 1415–1429. <https://doi.org/10.1107/S1600577520008723>.
- (152) Fuchs, M. R.; Bhogadi, D. K.; Jakoncic, J.; Myers, S.; Sweet, R. M.; Berman, L. E.; Skinner, J.; Idir, M.; Chubar, O.; McSweeney, S.; Schneider, D. K. NSLS-II Biomedical Beamlines for Micro-Crystallography, FMX, and for Highly Automated Crystallography, AMX: New Opportunities for Advanced Data Collection. *AIP Conf. Proc.* **2016**, 1741 (1), 030006. <https://doi.org/10.1063/1.4952829>.
- (153) Von Stetten, D.; Carpentier, P.; Flot, D.; Beteva, A.; Caserotto, H.; Dobias, F.; Guijarro, M.; Giraud, T.; Lentini, M.; McSweeney, S.; Royant, A.; Petitdemange, S.; Sinoir, J.; Surr, J.; Svensson, O.; Theveneau, P.; Leonard, G. A.; Mueller-Dieckmann, C. ID30A-3 (MASSIF-3) – a Beamline for Macromolecular Crystallography at the ESRF with a Small Intense Beam. *J. Synchrotron Radiat.* **2020**, 27 (3), 844–851. <https://doi.org/10.1107/S1600577520004002>.
- (154) McCarthy, A. A.; Barrett, R.; Beteva, A.; Caserotto, H.; Dobias, F.; Felisaz, F.; Giraud, T.; Guijarro, M.; Janocha, R.; Khadrouche, A.; Lentini, M.; Leonard, G. A.; Lopez Marrero, M.; Malbet-Monaco, S.; McSweeney, S.; Nurizzo, D.; Papp, G.; Rossi, C.; Sinoir, J.; Sorez, C.; Surr, J.; Svensson, O.; Zander, U.; Cipriani, F.; Theveneau, P.; Mueller-Dieckmann, C. ID30B – a Versatile Beamline for Macromolecular

- Crystallography Experiments at the ESRF. *J. Synchrotron Radiat.* **2018**, *25* (4), 1249–1260. <https://doi.org/10.1107/S1600577518007166/>.
- (155) CRIMS | Home. <https://htxlab.embl.fr/#/> (accessed 2023-03-10).
- (156) Lima, G. M. A.; Jagudin, E.; Talibov, V. O.; Benz, L. S.; Marullo, C.; Barthel, T.; Wollenhaupt, J.; Weiss, M. S.; Mueller, U. FragMAXapp: Crystallographic Fragment-Screening Data-Analysis and Project-Management System. *Acta Crystallogr. Sect. D Struct. Biol.* **2021**, *77* (6), 799–808. <https://doi.org/10.1107/s2059798321003818>.
- (157) Beteva, A.; Cipriani, F.; Cusack, S.; Delageniere, S.; Gabadinho, J.; Gordon, E. J.; Guijarro, M.; Hall, D. R.; Larsen, S.; Launer, L.; Lavault, C. B.; Leonard, G. A.; Mairs, T.; McCarthy, A.; McCarthy, J.; Meyer, J.; Mitchell, E.; Monaco, S.; Nurizzo, D.; Pernot, P.; Pieritz, R.; Ravelli, R. G. B.; Rey, V.; Shepard, W.; Spruce, D.; Stuart, D. I.; Svensson, O.; Theveneau, P.; Thibault, X.; Turkenburg, J.; Walsh, M.; McSweeney, S. M. High-Throughput Sample Handling and Data Collection at Synchrotrons: Embedding the ESRF into the High-Throughput Gene-to-Structure Pipeline. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, *62* (10), 1162–1169. <https://doi.org/10.1107/S0907444906032859>.
- (158) Mueller, U.; Thunnissen, M.; Nan, J.; Eguiraun, M.; Bolmsten, F.; Milàn-Otero, A.; Guijarro, M.; Oscarsson, M.; de Sanctis, D.; Leonard, G. MXCuBE3: A New Era of MX-Beamline Control Begins. *Synchrotron Radiat. News* **2017**, *30* (1), 22–27. <https://doi.org/10.1080/08940886.2017.1267564>.
- (159) Oscarsson, M.; Beteva, A.; Flot, D.; Gordon, E.; Guijarro, M.; Leonard, G.; McSweeney, S.; Monaco, S.; Mueller-Dieckmann, C.; Nanao, M.; Nurizzo, D.; Popov, A. N.; Von Stetten, D.; Svensson, O.; Rey-Bakaikoa, V.; Chado, I.; Chavas, L. M. G.; Gadea, L.; Gourhant, P.; Isabet, T.; Legrand, P.; Savko, M.; Sirigu, S.; Shepard, W.; Thompson, A.; Mueller, U.; Nan, J.; Eguiraun, M.; Bolmsten, F.; Nardella, A.; Milàn-Otero, A.; Thunnissen, M.; Hellmig, M.; Kastner, A.; Schmuckermaier, L.; Gerlach, M.; Feiler, C.; Weiss, M. S.; Bowler, M. W.; Gobbo, A.; Papp, G.; Sinoir, J.; McCarthy, A. A.; Karpics, I.; Nikolova, M.; Bourenkov, G.; Schneider, T.; Andreu, J.; Cuní, G.; Juanhuix, J.; Boer, R.; Fogh, R.; Keller, P.; Flensburg, C.; Paciorek, W.; Vonrhein, C.; Bricogne, G.; De Sanctis, D. MXCuBE2: The Dawn of MXCuBE Collaboration. *J. Synchrotron Radiat.* **2019**, *26* (2), 393–405. <https://doi.org/10.1107/S1600577519001267>.
- (160) Winter, G.; McAuley, K. E. Automated Data Collection for Macromolecular Crystallography. *Methods* **2011**, *55* (1), 81–93. <https://doi.org/10.1016/J.YMETH.2011.06.010>.
- (161) Incardona, M. F.; Bourenkov, G. P.; Levik, K.; Pieritz, R. A.; Popov, A. N.; Svensson, O. EDNA: A Framework for Plugin-Based Applications Applied to X-Ray Experiment Online Data Analysis. *J. Synchrotron Radiat.* **2009**, *16* (Pt 6), 872–879. <https://doi.org/10.1107/S0909049509036681>.
- (162) Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.; Bricogne, G. Data Processing and Analysis with the AutoPROC Toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 293–302. <https://doi.org/10.1107/S0907444911007773>.
- (163) Winter, G.; Waterman, D. G.; Parkhurst, J. M.; Brewster, A. S.; Gildea, R. J.; Gerstel, M.; Fuentes-Montero, L.; Vollmar, M.; Michels-Clark, T.; Young, I. D.; Sauter, N. K.; Evans, G. DIALS: Implementation and Evaluation of a New Integration Package. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2018**, *74* (2), 85–97. <https://doi.org/10.1107/S2059798317017235>.
- (164) Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (Pt 2), 125–132. <https://doi.org/10.1107/S0907444909047337>.

- (165) Sparta, K. M.; Krug, M.; Heinemann, U.; Mueller, U.; Weiss, M. S. XDSAPP2.0. *J. Appl. Crystallogr.* **2016**, *49* (3), 1085–1092. <https://doi.org/10.1107/S1600576716004416/>.
- (166) Smart, O. S.; Womack, T. O.; Flensburg, C.; Keller, P.; Paciorek, W.; Sharff, A.; Vonrhein, C.; Bricogne, G. Exploiting Structure Similarity in Refinement: Automated NCS and Target-Structure Restraints in BUSTER. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68* (4), 368–380. <https://doi.org/10.1107/S0907444911056058>.
- (167) Terwilliger, T. C.; Klei, H.; Adams, P. D.; Moriarty, N. W.; Cohn, J. D. Automated Ligand Fitting by Core-Fragment Fitting and Extension into Density. *Acta Crystallogr. Sect. D Struct. Biol.* **2006**, *62* (8), 915–922. <https://doi.org/10.1107/S0907444906017161>.
- (168) Terwilliger, T. C.; Adams, P. D.; Moriarty, N. W.; Cohn, J. D. Ligand Identification Using Electron-Density Map Correlations. *Acta Crystallogr. Sect. D Struct. Biol.* **2006**, *63* (1), 101–107. <https://doi.org/10.1107/S0907444906046233>.
- (169) Smart, O. S.; Womack, T. O.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G. Rhofit. *Glob. Phasing Ltd, Cambridge, UK* **2014**.
- (170) Long, F.; Nicholls, R. A.; Emsley, P.; Gražulis, S.; Merkys, A.; Vaitkus, A.; Murshudov, G. N. AceDRG: A Stereochemical Description Generator for Ligands. *Acta Crystallogr. Sect. D Struct. Biol.* **2017**, *73* (2), 112–122. <https://doi.org/10.1107/S2059798317000067/>.
- (171) Smart, O. S.; Womack, T. O.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G. Grade. *Glob. Phasing Ltd, Cambridge, UK* **2011**.
- (172) Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D. Electronic Ligand Builder and Optimization Workbench (ELBOW): A Tool for Ligand Coordinate and Restraint Generation. *Acta Crystallogr. Sect. D Struct. Biol.* **2009**, *65* (10), 1074–1080. <https://doi.org/10.1107/S0907444909029436>.
- (173) Wollenhaupt, J.; Metz, A.; Barthel, T.; Lima, G. M. A.; Heine, A.; Mueller, U.; Klebe, G.; Weiss, M. S. F2X-Universal and F2X-Entry: Structurally Diverse Compound Libraries for Crystallographic Fragment Screening. *Structure* **2020**, *28* (6), 694-706.e5. <https://doi.org/10.1016/j.str.2020.04.019>.
- (174) Wojdyr, M. UglyMol. **2016**. <https://doi.org/https://uglymol.github.io/>.
- (175) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (4), 486–501. <https://doi.org/10.1107/S0907444910007493>.
- (176) Noske, G. D.; Nakamura, A. M.; Gawriljuk, V. O.; Fernandes, R. S.; Lima, G. M. A.; Rosa, H. V.D.; Pereira, H. D.; Zeri, A. C. M.; Nascimento, A. F. Z.; Freire, M. C. L. C.; Fearon, D.; Douangamath, A.; von Delft, F.; Oliva, G.; Godoy, A. S. A Crystallographic Snapshot of SARS-CoV-2 Main Protease Maturation Process. *J. Mol. Biol.* **2021**, *433* (18), 167118. <https://doi.org/10.1016/J.JMB.2021.167118>.
- (177) Kozielski, F.; Sele, C.; Talibov, V. O.; Lou, J.; Dong, D.; Wang, Q.; Shi, X.; Nyblom, M.; Rogstam, A.; Krojer, T.; Fisher, Z.; Knecht, W. Identification of Fragments Binding to SARS-CoV-2 Nsp10 Reveals Ligand-Binding Sites in Conserved Interfaces between Nsp10 and Nsp14/Nsp16. *RSC Chem. Biol.* **2022**, *3* (1), 44. <https://doi.org/10.1039/D1CB00135C>.
- (178) Barthel, T.; Wollenhaupt, J.; Lima, G. M. A.; Wahl, M. C.; Weiss, M. S. Large-Scale Crystallographic Fragment Screening Expedites Compound Optimization and Identifies Putative Protein-Protein Interaction Sites. *J. Med. Chem.* **2022**, *2022*, 14630–14641. <https://doi.org/10.1021/ACS.JMEDCHEM.2C01165/>.
- (179) Zartler, E.; Swain, C.; Pearce, S. Library Design the Evolution of Fragment-Based Lead

Discovery. *Drug Discov. World Winter 2012/13* **2012**.

- (180) Reville Imbernon, J.; Jacquemard, C.; Bret, G.; Marcou, G.; Kellenberger, E. Comprehensive Analysis of Commercial Fragment Libraries. *RSC Med. Chem.* **2022**, *13* (3), 300–310. <https://doi.org/10.1039/D1MD00363A>.
- (181) Reville Imbernon, J.; Chiesa, L.; Kellenberger, E. Mining the Protein Data Bank to Inspire Fragment Library Design. *Front. Chem.* **2023**, *11* (February), 1–11. <https://doi.org/10.3389/fchem.2023.1089714>.
- (182) Diederichs, K. *ISa Website*. <https://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/ISa> (accessed 2019-09-18).
- (183) Radeva, N.; Schiebel, J.; Wang, X.; Krimmer, S. G.; Fu, K.; Stieler, M.; Ehrmann, F. R.; Metz, A.; Rickmeyer, T.; Betz, M.; Winkvist, J.; Park, A. Y.; Huschmann, F. U.; Weiss, M. S.; Mueller, U.; Heine, A.; Klebe, G. Active Site Mapping of an Aspartic Protease by Multiple Fragment Crystal Structures: Versatile Warheads to Address a Catalytic Dyad. *J. Med. Chem.* **2016**, *59* (21), 9743–9759. <https://doi.org/10.1021/acs.jmedchem.6b01195>.
- (184) Liu, S.; Li, X.; Zhang, L.; Jiang, J.; Hill, R. C.; Cui, Y.; Hansen, K. C.; Zhou, Z. H.; Zhao, R. Structure of the Yeast Spliceosomal Post-Catalytic P Complex. *Science* **2017**, *358* (6368), 1278. <https://doi.org/10.1126/SCIENCE.AAR3462>.
- (185) Wan, R.; Yan, C.; Bai, R.; Wang, L.; Huang, M.; Wong, C. C. L.; Shi, Y. The 3.8 Å Structure of the U4/U6.U5 Tri-SnRNP: Insights into Spliceosome Assembly and Catalysis. *Science* (80-.). **2016**, *351* (6272), 466–475. <https://doi.org/10.1126/science.aad6466>.
- (186) Yan, C.; Wan, R.; Bai, R.; Huang, G.; Shi, Y. Structure of a Yeast Step II Catalytically Activated Spliceosome. *Science* (80-.). **2017**, *355* (6321), 13. https://doi.org/10.1126/SCIENCE.AAK9979/SUPPL_FILE/YAN-SM.PDF.
- (187) Wollenhaupt, J.; Barthel, T.; Lima, G. M. A.; Metz, A.; Wallacher, D.; Jagudin, E.; Huschmann, F. U.; Hauß, T.; Feiler, C. G.; Gerlach, M.; Hellmig, M.; Förster, R.; Steffien, M.; Heine, A.; Klebe, G.; Mueller, U.; Weiss, M. S. Workflow and Tools for Crystallographic Fragment Screening at the Helmholtz-Zentrum Berlin. *J. Vis. Exp.* **2021**, *2021* (169), 1–19. <https://doi.org/10.3791/62208>.
- (188) Hall, R. J.; Murray, C. W.; Verdonk, M. L. The Fragment Network: A Chemistry Recommendation Engine Built Using a Graph Database. **2017**. <https://doi.org/10.1021/acs.jmedchem.7b00809>.
- (189) Erlanson, D. A.; Davis, B. J.; Jahnke, W. Fragment-Based Drug Discovery: Advancing Fragments in the Absence of Crystal Structures. *Cell Chem. Biol.* **2019**, *26* (1), 9–15. <https://doi.org/10.1016/J.CHEMBIOL.2018.10.001>.
- (190) Schulz, M. N.; Landström, J.; Bright, K.; Hubbard, R. E. Design of a Fragment Library That Maximally Represents Available Chemical Space. *J. Comput. Aided. Mol. Des.* **2011**, *25* (7), 611–620. <https://doi.org/10.1007/S10822-011-9461-XI>.
- (191) Berthold, M. R.; Cebon, N.; Dill, F.; Gabriel, T. R.; Kotter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. KNIME-The Konstanz Information Miner Version 2.0 and Beyond. <https://doi.org/10.1145/1656274.1656280>.
- (192) *Compound Sourcing, Selling and Purchasing Platform | Molport*. <https://www.molport.com/shop/index> (accessed 2023-03-10).
- (193) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem.*

- Inf. Model.* **2010**, *50* (4), 572–584. <https://doi.org/10.1021/CI100031X>.
- (194) Neudert, G.; Klebe, G. Fconv: Format Conversion, Manipulation and Feature Computation of Molecular Data. *Bioinformatics* **2011**, *27* (7), 1021–1022. <https://doi.org/10.1093/bioinformatics/btr055>.
- (195) Liebschner, D.; Afonine, P. V.; Moriarty, N. W.; Poon, B. K.; Sobolev, O. V.; Terwilliger, T. C.; Adams, P. D. Polder Maps: Improving OMIT Maps by Excluding Bulk Solvent. *Acta Crystallogr. Sect. D Struct. Biol.* **2017**, *73* (2), 148–157. <https://doi.org/10.1107/S2059798316018210/>.
- (196) Ginn, H. M. Pre-Clustering Data Sets Using Cluster 4 x Improves the Signal-to-Noise Ratio of High-Throughput Crystallography Drug-Screening Analysis. *Acta Crystallogr. Sect. D Struct. Biol.* **2020**, *76* (11), 1134–1144. <https://doi.org/10.1107/S2059798320012619>.
- (197) Bai, R.; Yan, C.; Wan, R.; Lei, J.; Correspondence, Y. S.; Shi, Y. Structure of the Post-Catalytic Spliceosome from *Saccharomyces Cerevisiae* Data Resources 5YLZ Article Structure of the Post-Catalytic Spliceosome from *Saccharomyces Cerevisiae*. *Cell* **2017**, *171*, 1589–1594.e8. <https://doi.org/10.1016/j.cell.2017.10.038>.
- (198) Bower, J.; Pugliese, A.; Drysdale, M. Strategies for Fragment Library Design. **2016**, 99–118. <https://doi.org/10.1002/9783527683604.CH05>.
- (199) Doak, B. C.; Morton, C. J.; Simpson, J. S.; Scanlon, M. J. Assembly of Fragment Screening Libraries. *Appl. Biophys. Drug Discov.* **2017**, 263–283. <https://doi.org/10.1002/9781119099512.ch14>.
- (200) GmbH, J. B. Frag Xtal Screen, Fragment Screen. *Frag Xtal Screen, Fragment Screen*.
- (201) Schuller, M.; Correy, G. J.; Gahbauer, S.; Fearon, D.; Wu, T.; Díaz, R. E.; Young, I. D.; Martins, L. C.; Smith, D. H.; Schulze-Gahmen, U.; Owens, T. W.; Deshpande, I.; Merz, G. E.; Thwin, A. C.; Biel, J. T.; Peters, J. K.; Moritz, M.; Herrera, N.; Kratochvil, H. T.; Aimon, A.; Bennett, J. M.; Neto, J. B.; Cohen, A. E.; Dias, A.; Douangamath, A.; Dunnett, L.; Fedorov, O.; Ferla, M. P.; Fuchs, M. R.; Gorrie-Stone, T. J.; Holton, J. M.; Johnson, M. G.; Krojer, T.; Meigs, G.; Powell, A. J.; Rack, J. G. M.; Rangel, V. L.; Russi, S.; Skyner, R. E.; Smith, C. A.; Soares, A. S.; Wierman, J. L.; Zhu, K.; O'Brien, P.; Jura, N.; Ashworth, A.; Irwin, J. J.; Thompson, M. C.; Gestwicki, J. E.; Von Delft, F.; Shoichet, B. K.; Fraser, J. S.; Ahel, I. Fragment Binding to the Nsp3 Macrodomein of SARS-CoV-2 Identified through Crystallographic Screening and Computational Docking. *Sci. Adv.* **2021**, *7* (16), 8711–8725. <https://doi.org/10.1126/SCIADV.ABF8711>.
- (202) Godoy, A. S.; Nakamura, A. M.; Douangamath, A.; Song, Y.; Noske, G. D.; Gawriljuk, V. O.; Fernandes, R. S.; Pereira, H. D. M.; Oliveira, K. I. Z.; Fearon, D.; Dias, A.; Krojer, T.; Fairhead, M.; Powell, A.; Dunnett, L.; Brandao-Neto, J.; Skyner, R.; Chalk, R.; Delft, F. von; Bajusz, D.; Bege, M.; Borbás, A.; Keserú, G. M.; Oliva, G. Allosteric Regulation and Crystallographic Fragment Screening of SARS-CoV-2 NSP15 Endoribonuclease. *bioRxiv* **2022**, 2022.09.26.509485. <https://doi.org/10.1101/2022.09.26.509485>.
- (203) Cornaciu, I.; Bourgeas, R.; Hoffmann, G.; Dupeux, F.; Humm, A.-S.; Mariaule, V.; Pica, A.; Clavel, D.; Seroul, G.; Murphy, P.; Márquez, J. A. The Automated Crystallography Pipelines at the EMBL HTX Facility in Grenoble. *J. Vis. Exp.* **2021**, No. 172. <https://doi.org/10.3791/62491-V>.
- (204) *Two in one: combining MASSIF-1 and CrystalDirect | EMBL*. <https://www.embl.org/news/lab-matters/two-in-one-combining-massif-1-and-crystaldirect/> (accessed 2023-03-10).
- (205) Delagenière, S.; Brechereau, P.; Launer, L.; Ashton, A. W.; Leal, R.; Veyrier, S.; Gabadinho, J.; Gordon, E. J.; Jones, S. D.; Levik, K. E.; McSweeney, S. M.; Monaco,

- S.; Nanao, M.; Spruce, D.; Svensson, O.; Walsh, M. A.; Leonard, G. A. Data and Text Mining Advance Access Publication ISPyB: An Information Management System for Synchrotron Macromolecular Crystallography. *2011*, 27 (22), 3186–3192. <https://doi.org/10.1093/bioinformatics/btr535>.
- (206) Strutz, T.; Pothineni, S. B.; Lamzin, V. S. Automated Detection and Centring of Cryocooled Protein Crystals. *Acta Crystallogr. D. Biol. Crystallogr.* **2006**, 62 (Pt 11), 1358–1368. <https://doi.org/10.1107/S0907444906031672>.
- (207) Jain, A.; Stojanoff, V. Are You Centered? An Automatic Crystal-Centering Method for High-Throughput Macromolecular Crystallography. *J. Synchrotron Radiat.* **2007**, 14 (4), 355–360. <https://doi.org/10.1107/S0909049507018110>.
- (208) Ito, S.; Ueno, G.; Yamamoto, M. DeepCentering: Fully Automated Crystal Centering Using Deep Learning for Macromolecular Crystallography. *J. Synchrotron Radiat.* **2019**, 26 (4), 1361–1366. <https://doi.org/10.1107/S160057751900434X/>.
- (209) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, 67 (4), 235–242. <https://doi.org/10.1107/S0907444910045749>.
- (210) Anthony, J. G.; Weidenhammer, E. M.; Woolford, J. L. The Yeast Prp3 Protein Is a U4/U6 SnRNP Protein Necessary for Integrity of the U4/U6 SnRNP and the U4/U6.U5 Tri-SnRNP. *RNA* **1997**, 3 (10), 1143.
- (211) Liu, S.; Mozaffari-Jovin, S.; Wollenhaupt, J.; Santos, K. F.; Theuser, M.; Dunin-Horkawicz, S.; Fabrizio, P.; Bujnicki, J. M.; Lührmann, R.; Wahl, M. C. A Composite Double-/Single-Stranded Rna-Binding Region in Protein Prp3 Supports Tri-SnRNP Stability and Splicing. *Elife* **2015**, 4 (JULY 2015). <https://doi.org/10.7554/eLife.07320>.
- (212) James, S.-A.; Turner, W.; Schwer, B. *How Slu7 and Prp18 Cooperate in the Second Step of Yeast Pre-mRNA Splicing*; 2002.
- (213) Aronova, A.; Bacíková, D.; Crotti, L. B.; Horowitz, D. S.; Schwer, B. Functional Interactions between Prp8, Prp18, Slu7, and U5 SnRNA during the Second Step of Pre-mRNA Splicing. *RNA* **2007**, 13 (9), 1437–1444. <https://doi.org/10.1261/rna.572807>.
- (214) Bhadra, A.; Yeturu, K. Site2Vec: A Reference Frame Invariant Algorithm for Vector Embedding of Protein–Ligand Binding Sites. *Mach. Learn. Sci. Technol.* **2020**, 2 (1), 015005. <https://doi.org/10.1088/2632-2153/ABAD88>.
- (215) Ehrt, C.; Brinkjost, T.; Koch, O. A Benchmark Driven Guide to Binding Site Comparison: An Exhaustive Evaluation Using Tailor-Made Data Sets (ProSPECCTs). *PLOS Comput. Biol.* **2018**, 14 (11), e1006483. <https://doi.org/10.1371/JOURNAL.PCBI.1006483>.
- (216) Simonovsky, M.; Meyers, J. Deeply Tough: Learning Structural Comparison of Protein Binding Sites. *J. Chem. Inf. Model.* **2020**, 60 (4), 2356–2366. https://doi.org/10.1021/ACS.JCIM.9B00554/ASSET/IMAGES/LARGE/C19B00554_0007.JPEG.
- (217) Govindaraj, R. G.; Brylinski, M. Comparative Assessment of Strategies to Identify Similar Ligand-Binding Pockets in Proteins. *BMC Bioinformatics* **2018**, 19 (1), 1–17. <https://doi.org/10.1186/S12859-018-2109-2/FIGURES/12>.
- (218) Geerds, C.; Bley Müller, W. M.; Meyer, T.; Widmann, C.; Niemann, H. H. A Recurring Packing Contact in Crystals of InlB Pinpoints Functional Binding Sites in the Internalin Domain and the B Repeat. *Acta Crystallogr. Sect. D Struct. Biol.* **2022**, 78 (3), 310–320.

<https://doi.org/10.1107/S2059798322000432>.

- (219) Cox, O. B.; Krojer, T.; Collins, P.; Monteiro, O.; Talon, R.; Bradley, A.; Fedorov, O.; Amin, J.; Marsden, B. D.; Spencer, J.; Von Delft, F.; Brennan, P. E. A Poised Fragment Library Enables Rapid Synthetic Expansion Yielding the First Reported Inhibitors of PHIP(2), an Atypical Bromodomain. *Chem. Sci.* **2016**, *7* (3), 2322–2330. <https://doi.org/10.1039/C5SC03115J>.
- (220) Bentley, M. R.; Ilyichova, O. V.; Wang, G.; Williams, M. L.; Sharma, G.; Alwan, W. S.; Whitehouse, R. L.; Mohanty, B.; Scammells, P. J.; Heras, B.; Martin, J. L.; Totsika, M.; Capuano, B.; Doak, B.; Scanlon, M. J. Rapid Elaboration of Fragments into Leads by X-Ray Crystallographic Screening of Parallel Chemical Libraries (REFIL_x). *J. Med. Chem.* **2020**, *acs.jmedchem.0c00111*. <https://doi.org/10.1021/acs.jmedchem.0c00111>.
- (221) Ferla, M. *Fragmenstein*. <https://github.com/matteoferla/Fragmenstein>.
- (222) Ludlow, R. F.; Verdonk, M. L.; Saini, H. K.; Tickle, I. J.; Jhoti, H. Detection of Secondary Binding Sites in Proteins Using Fragment Screening. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (52), 15910–15915. <https://doi.org/10.1073/PNAS.1518946112/>.
- (223) Krojer, T.; Fraser, J. S.; von Delft, F. Discovery of Allosteric Binding Sites by Crystallographic Fragment Screening. *Curr. Opin. Struct. Biol.* **2020**, *65*, 209–216. <https://doi.org/10.1016/J.SBI.2020.08.004>.
- (224) Shumilin, I. A.; Cymborowski, M.; Chertihin, O.; Jha, K. N.; Herr, J. C.; Lesley, S. A.; Joachimiak, A.; Minor, W. Identification of Unknown Protein Function Using Metabolite Cocktail Screening. *Structure* **2012**, *20* (10), 1715. <https://doi.org/10.1016/J.STR.2012.07.016>.
- (225) Barelier, S.; Cummings, J. A.; Rauwerdink, A. M.; Hitchcock, D. S.; Farelli, J. D.; Almo, S. C.; Raushel, F. M.; Allen, K. N.; Shoichet, B. K. Substrate Deconstruction and the Nonadditivity of Enzyme Recognition. *J. Am. Chem. Soc.* **2014**, *136* (20), 7374–7382. https://doi.org/10.1021/JA501354Q/SUPPL_FILE/JA501354Q_SI_001.PDF.
- (226) *Practical Fragments: Biofragments: extracting signal from noise, and the limits of three-dimensionality*. <https://practicalfragments.blogspot.com/2014/05/biofragments-extracting-signal-from.html> (accessed 2023-03-18).
- (227) *Practical Fragments: Fragmenting natural products – sometimes PAINfully*. <https://practicalfragments.blogspot.com/2013/02/fragmenting-natural-products-sometimes.html> (accessed 2023-03-18).
- (228) Oleinikovas, V.; Saladino, G.; Cossins, B. P.; Gervasio, F. L. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *J. Am. Chem. Soc.* **2016**, *138* (43), 14257–14263. <https://doi.org/10.1021/JACS.6B05425/>.
- (229) *ITC • Isothermal Titration Calorimetry • Thermodynamics, Affinity, Stoichiometry*. <https://2bind.com/itc/> (accessed 2023-03-13).
- (230) Lessel, U.; Lemmen, C. Comparison of Large Chemical Spaces. *ACS Med. Chem. Lett.* **2019**, *10* (10), 1504–1510. <https://doi.org/10.1021/ACSMEDCHEMLETT.9B00331/>.
- (231) *REAL Space - Enamine*. <https://enamine.net/compound-collections/real-compounds/real-space-navigator> (accessed 2023-03-10).
- (232) *Freedom Space by Chemspace • BioSolveIT*. <https://www.biosolveit.de/2022/10/13/freedom-space-by-chemspace/> (accessed 2023-03-10).
- (233) *Building Blocks - Enamine*. <https://enamine.net/building-blocks> (accessed 2023-03-22).
- (234) Bai, R.; Wan, R.; Yan, C.; Lei, J.; Shi, Y. Structures of the Fully Assembled

Saccharomyces Cerevisiae Spliceosome before Activation. *Science*. **2018**, 360 (6396), 1423–1429. <https://doi.org/10.1126/SCIENCE.AAU0325>.

- (235) Collins, P. M.; Douangamath, A.; Talon, R.; Dias, A.; Brandao-Neto, J.; Krojer, T.; von Delft, F. Achieving a Good Crystal System for Crystallographic X-Ray Fragment Screening. *Methods Enzymol.* **2018**, 610, 251–264. <https://doi.org/10.1016/BS.MIE.2018.09.027>.
- (236) Costanzi, E.; Kuzikov, M.; Esposito, F.; Albani, S.; Demitri, N.; Giabbai, B.; Camasta, M.; Tramontano, E.; Rossetti, G.; Zaliani, A.; Storici, P. Structural and Biochemical Analysis of the Dual Inhibition of MG-132 against SARS-CoV-2 Main Protease (Mpro/3CLpro) and Human Cathepsin-L. *Int. J. Mol. Sci.* **2021**, 22 (21). <https://doi.org/10.3390/IJMS222111779>.
- (237) Consortium, T. C. M.; Chodera, J.; Lee, A.; London, N.; Delft, F. von. Open Science Discovery of Oral Non-Covalent SARS-CoV-2 Main Protease Inhibitors. **2021**. <https://doi.org/10.26434/CHEMRXIV-2021-585KS-V2>.

Publications

- I. **Barthel, T.**; Huschmann, F. U.; Wallacher, D.; Feiler, C. G.; Klebe, G.; Weiss, M. S.; Wollenhaupt, J., Facilitated Crystal Handling Using a Simple Device for Evaporation Reduction in Microtiter Plates. *J. Appl. Crystallogr.* **2021**, *54* (1), 376–382. <https://doi.org/10.1107/s1600576720016477>
CC BY license: <https://creativecommons.org/licenses/by/4.0/>
- II. Lima, G. M. A.; Jagudin, E.; Talibov, V. O.; Benz, L. S.; Marullo, C.; **Barthel, T.**; Wollenhaupt, J.; Weiss, M. S.; Mueller, U., FragMAXapp: Crystallographic Fragment-Screening Data-Analysis and Project-Management System. *Acta Crystallogr. Sect. D Struct. Biol.* **2021**, *77* (6), 799–808. <https://doi.org/10.1107/s2059798321003818>.
CC BY license: <https://creativecommons.org/licenses/by/4.0/>
- III. Wollenhaupt, J.; Metz, A.; **Barthel, T.**; Lima, G. M. A.; Heine, A.; Mueller, U.; Klebe, G.; Weiss, M. S., F2X-Universal and F2X-Entry: Structurally Diverse Compound Libraries for Crystallographic Fragment Screening. *Structure* **2020**, *28* (6), 694-706.e5. <https://doi.org/10.1016/j.str.2020.04.019>.
- IV. Metz, A.; Wollenhaupt, J.; Glöckner, S.; Messini, N.; Huber, S.; **Barthel, T.**; Merabet, A.; Gerber, H. D.; Heine, A.; Klebe, G.; Weiss, M. S., Frag4Lead: Growing Crystallographic Fragment Hits by Catalog Using Fragment-Guided Template Docking. *Acta Crystallogr. Sect. D Struct. Biol.* **2021**, *77* (9), 1168–1182. <https://doi.org/10.1107/S2059798321008196>.
CC BY license: <https://creativecommons.org/licenses/by/4.0/>
- V. Wollenhaupt, J.; **Barthel, T.**; Lima, G. M. A.; Metz, A.; Wallacher, D.; Jagudin, E.; Huschmann, F. U.; Hauß, T.; Feiler, C. G.; Gerlach, M.; Hellmig, M.; Förster, R.; Steffien, M.; Heine, A.; Klebe, G.; Mueller, U.; Weiss, M. S., Workflow and Tools for Crystallographic Fragment Screening at the Helmholtz-Zentrum Berlin. *J. Vis. Exp.* **2021**, 2021 (169), 1–19. <https://doi.org/10.3791/62208>.
CC BY-NC-ND 3.0 license: <https://creativecommons.org/licenses/by-nc-nd/3.0/>
- VI. **Barthel, T.**; Wollenhaupt, J.; Lima, G. M. A.; Wahl, M. C.; Weiss, M. S., Large-Scale Crystallographic Fragment Screening Expedites Compound Optimization and Identifies Putative Protein-Protein Interaction Sites. *J. Med. Chem.* **2022**, *2022*, 14630–14641. <https://doi.org/10.1021/acs.jmedchem.2c01165>



Facilitated crystal handling using a simple device for evaporation reduction in microtiter plates

Tatjana Barthel,^{a,b} Franziska U. Huschmann,^{a,c} Dirk Wallacher,^d Christian G. Feiler,^a Gerhard Klebe,^c Manfred S. Weiss^{a*} and Jan Wollenhaupt^{a*}

^aHelmholtz-Zentrum Berlin, Macromolecular Crystallography, Albert-Einstein-Straße 15, 12489 Berlin, Germany, ^bFreie Universität Berlin, Institute for Chemistry and Biochemistry, Structural Biochemistry Group, Takustraße 5, 14195 Berlin, Germany, ^cPhilipps-Universität Marburg, Institute of Pharmaceutical Chemistry, Drug Design Group, Marbacher Weg 6, 35032 Marburg, Germany, and ^dHelmholtz-Zentrum Berlin, Department Sample Environment, Hahn-Meitner-Platz 1, 14109 Berlin, Germany. *Correspondence e-mail: msweiss@helmholtz-berlin.de, jan.wollenhaupt@helmholtz-berlin.de

Received 22 October 2020

Accepted 20 December 2020

Edited by J. M. García-Ruiz, Instituto Andaluz de Ciencias de la Tierra, Granada, Spain

Keywords: evaporation reduction; crystal handling; crystal harvesting; crystallographic fragment screening.

Supporting information: this article has supporting information at journals.iucr.org/j

In the past two decades, most of the steps in a macromolecular crystallography experiment have undergone tremendous development with respect to speed, feasibility and increase of throughput. The part of the experimental workflow that is still a bottleneck, despite significant efforts, involves the manipulation and harvesting of the crystals for the diffraction experiment. Here, a novel low-cost device is presented that functions as a cover for 96-well crystallization plates. This device enables access to the individual experiments one at a time by its movable parts, while minimizing evaporation of all other experiments of the plate. In initial tests, drops of many typically used crystallization cocktails could be successfully protected for up to 6 h. Therefore, the manipulation and harvesting of crystals is straightforward for the experimenter, enabling significantly higher throughput. This is useful for many macromolecular crystallography experiments, especially multi-crystal screening campaigns.

1. Introduction

In the past two decades, X-ray crystallography, in particular macromolecular X-ray crystallography, has experienced an enormous boost with respect to automation and throughput. This is, in part, due to newer generation synchrotron facilities and increased sensitivity and fast readout of modern detectors (Leonarski *et al.*, 2018; Förster *et al.*, 2019; Owen *et al.*, 2016). Moreover, attempts to automate the entire process have led to advances in crystallization screening technologies, robot-assisted sample mounting at synchrotron beamlines, semi- or completely automated data collection, and more or less completely automated data processing and refinement procedures of the collected data (Shaw Stewart & Mueller-Dieckmann, 2014; Douangamath *et al.*, 2014; Bowler *et al.*, 2016; Powell, 2017). Efforts have also been directed towards the automation of crystal handling and crystal harvesting, tackling this bottleneck of high-throughput crystallography (Deller & Rupp, 2014; Cuttitta *et al.*, 2015; Wright *et al.*, 2021; Cipriani *et al.*, 2012). Most of these devices were built specifically for certain beamlines or laboratories, or have been commercialized, but at rather high costs due to the materials, mechanics and motors involved. Thus, they are only being used by a limited number of facilities and laboratories.

Without such devices at hand, crystal handling is still mainly manual work and time consuming. The crystals grown need to be individually harvested and flash-cooled in liquid nitrogen before data collection. Prior to cooling, crystals are often further manipulated in the crystallization plate environment.



OPEN ACCESS

Such manipulations, *i.e.* post-crystallization treatments where the crystals remain inside the drop, can involve heavy-metal derivatization for phasing experiments, dehydration to improve diffraction quality or soaking experiments with ligands (Heras & Martin, 2005; Rould, 2007). In particular for drug discovery projects the crystallographic screening of small-molecule compounds called fragments has recently been established as a high-throughput technique that requires the harvesting and preparation of a very large number of samples (Schiebel *et al.*, 2016; Krojer *et al.*, 2017; Lima *et al.*, 2020).

The entire process of harvesting and preparation of a crystalline sample for a diffraction experiment usually involves the cutting open of the foil that seals the crystallization plate to allow access to an individual well, the manipulation of crystals of that particular well according to the purpose of the experiment, and then the re-sealing of the respective well for incubation. At a later stage, the well is reopened for harvesting of the samples. Taken together, this procedure is cumbersome, time consuming and limits the number of crystals that can be handled by the experimenter in a given time. Since most of the steps required for sample manipulation usually involve partial evaporation of liquids composing the crystallization drop, any attempt to optimize the process will have to account for evaporation. In this context, two recent developments need to be mentioned: The first is a plate lid with apertures (Zipper *et al.*, 2014). This was mainly developed for reducing evaporation during the drop-setting part of the experiment. The second development is the Crystal Shifter (Wright *et al.*, 2021), a motorized XY microscope stage developed to speed up crystal handling and simultaneously reduce evaporation.

Here we present a small and affordable device which reduces evaporation during crystal manipulation steps while at the same time allowing for easy crystal handling. It thus facilitates rapid crystal treatment and harvesting. It is applied as a temporary lid placed on top of a crystallization plate and can be used in combination with any typical laboratory microscope. It is currently customized for 3-lens 96-well MRC low-profile crystallization plates, but the design is easily adaptable and can be modified to fit any 96-well crystallization plate following the standard SBS footprint.

2. Design and assembly

The design of the device is inspired by a sliding puzzle: *i.e.* individual vertically movable square tiles enable access to the individual wells on the crystallization plate. Concurrently, the tiles protect the other wells, which have to remain covered and thus protected from evaporation, ideally for several hours.

The device consists of a frame, a bar, sliding clamps, a lever tool made from 3D-printed plastic and 96 acrylic glass tiles (Fig. 1). The set is completed by a two-piece pen tool, the handle of which is 3D-printed as well. The top part is designed to accommodate a bent cannula with a rounded tip. All 3D-printed plastic parts are made from thermoplastic (Vero-BlackPlus, stratasys) using an Objet30 Pro printer. The tiles are made from 0.8 mm-thick acrylic glass plates and excized

using a Universal Laser Systems M20 laser. The laser is also used for engraving the tiles in the middle, *i.e.* adding little dips of 1.63 mm diameter. These serve as optional handles when the pen tool is used. After production and cleaning of the parts, the tiles are inserted into the grooves of the frame so that they can be moved vertically. The bar prevents the tiles from accidentally falling off or being slid off the frame. Thus, the tiles are kept inside the frame by the grooves and the bar.

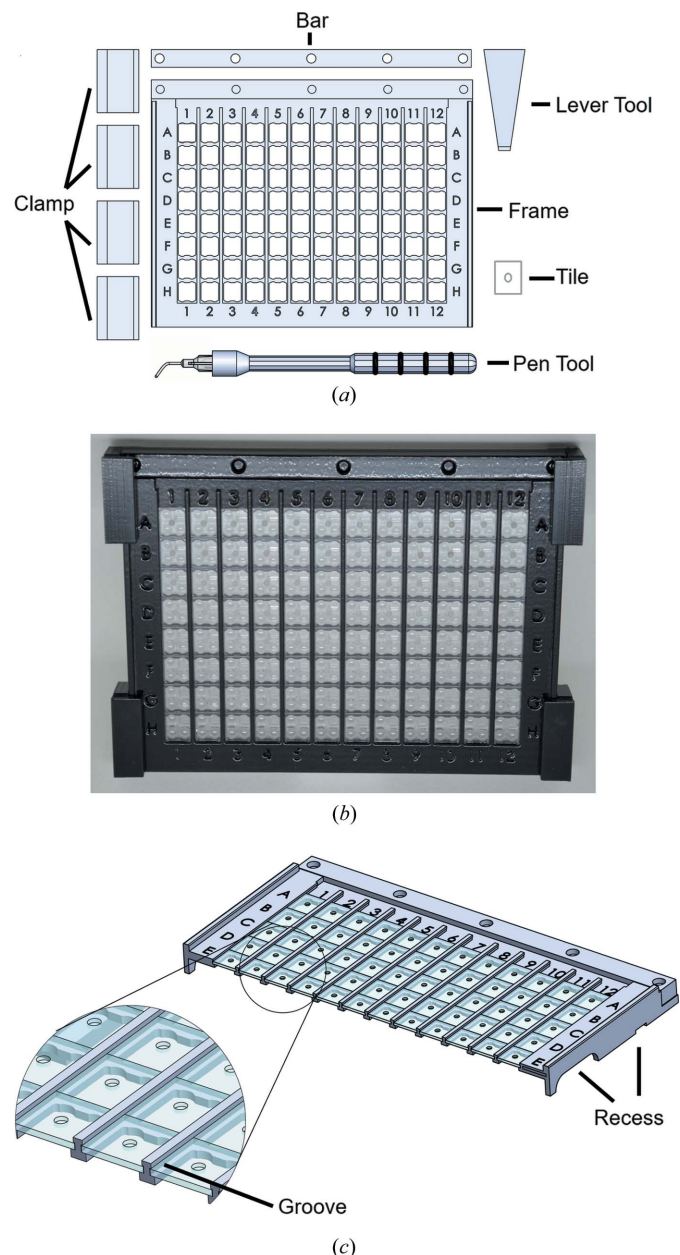


Figure 1 Parts and assembly of the device. (a) Schematic overview of all parts of the frame as well as an acrylic glass tile and the pen and lever tools. (b) Photograph of the fully assembled frame. It was placed on a 3-lens 96-well MRC low-profile crystallization plate and secured by four clamps. (c) The device, seen from another side, as a schematic cross section showing the recesses available on the side. The larger recess can be used via the thumb to hold the plate while lifting the frame; the smaller recesses fit the lever tool to loosen the frame before completely lifting it up. In the close-up view (circle) the groove with the acrylic glass tiles inserted is visible.

The tiles can be moved either with the tip of a finger or, alternatively, using the pen tool to avoid touching them by hand. The production of the frame, including the lasering of the acrylic glass tiles, the printing and cleaning of all 3D-printed parts, and the insertion of the 96 acrylic glass tiles into the frame, takes approximately 10 h per set.

The frame edges are labeled with the corresponding well numbers and letters, helping the experimenter to keep track of each position. In its current design, the frame only fits onto 3-lens 96-well MRC low-profile crystallization plates. The frame is placed onto a prepared crystallization plate after removal of the respective sealing foil. It can be fixed onto the plate without the need for adhesives or grease using the provided clamps. For the safe removal of the frame from the plate, recesses are available on the left and right sides. The large recesses are provided to secure the plate with the thumb. The small recesses can be used in combination with the lever tool, which enables the frame to be lifted off the crystallization plate by a simple turning movement. The removal of the frame from the plate is not adversely affected by any potentially remaining minute amounts of adhesive from the crystallization foil removed earlier. The entire handling process of setting up the frame on the crystallization plate, possible ways of tile movement and removal of the frame from the crystallization plate are visualized in Video S1 of the supporting information.¹

3. Application

By placing the frame on top of the crystallization plate, each well is sealed individually, and the wells can be accessed one at a time. The transparent acrylic glass tiles allow the observation of each experiment in its sealed compartment. The engraved depressions on the acrylic glass tiles do not impair the view of the drops. Due to the height of the frame, the working angle at which it is possible to reach a crystallization drop inside a 3-lens 96-well MRC low-profile crystallization plate decreases by less than three degrees (Fig. 2). Such angles are still comfortable for handling/harvesting crystals under a typical laboratory microscope. The angle was estimated by considering the usual position of protein crystals on the crystallization plate lens. It was assumed that manipulation and harvesting would be performed from the right side, as this gives the maximum space possible with and without the assembled device. If it is difficult to manipulate or harvest a protein crystal from this specific side, it is of course possible to move/rotate the assembly of the plate with the frame to a position that allows for easy handling.

4. Evaporation reduction

The aim of this device is to work with most solutions, except those containing a volatile material as the main component.

¹ Before placing the frame on top of the crystallization plate certain steps need to be performed that are not included in the video. These steps depend on the individual experiment. In the case where foil seals the plate, the foil needs to be removed beforehand. After removing the frame from the plate, the latter can be sealed again by a foil, e.g. for further incubation steps, or the plate can be discarded.

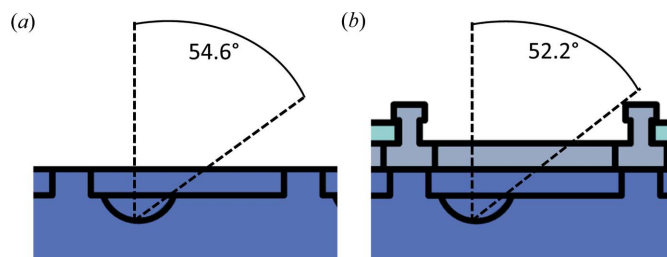


Figure 2

Schematic view of usable space with the device. A side view of one well of a 3-lens 96-well MRC low-profile crystallization plate (purple) is shown without or with the frame (gray). The half circle inside the plate represents one of the lenses of the crystallization plate. Usually the protein crystals will be located towards the bottom of the lens, and thus the angle is measured from this point. As the user will usually choose to work from the side which is most comfortable and provides the largest space for manipulating and harvesting, the angles from the right side are compared. (a) Plate without the frame. The angle shown depicts the possible working angle without the frame of 54.6°. (b) With the frame (gray) placed on top of a crystallization plate, the angle slightly decreases to 52.2°. Using the frame, a reduction of only 2.4° in angular movement range has to be accepted by the experimenter.

Reliable evaporation reduction can be achieved for 1 to 6 h, *i.e.* a typical working session, after the evaporation equilibrium has been reached. If an experiment requires longer incubation times, the frame should be removed and the crystallization plate re-sealed with a standard foil used in macromolecular crystallography. In order to estimate the performance of the device, evaporation experiments were conducted. The working time was assessed while experiments were protected by the acrylic glass tiles. The frame was tested by observing the evaporation of different solutions. Soaking solutions from already performed crystallographic fragment screening (CFS) campaigns and solutions from a typical crystallization screen consisting of 96 different crystallization solutions (JSCG Core Suite II, Quiagen) were tested to obtain an overview of the usability range.

4.1. Evaporation reduction – CFS soaking solutions

In the case of the experiments concerning the evaporation of soaking solutions a 3-lens 96-well MRC low-profile crystallization plate was used with a 40 µl reservoir and 0.4 µl drops in the tested wells. The reservoir was pipetted before the plate was covered by the frame. The drops were set manually after covering the plate with the frame and accessing the individual wells by appropriate tile movements. Each solution was tested in three individual experiments while care was taken to distribute the used wells across the plate (corner versus edge versus center). Subsequently, the drops were observed visually over time at a fixed magnification and focus. The plate was placed on a motorized XY table underneath the microscope lens, allowing plate movements under the microscope in a defined and reproducible manner. As a metric for the extent of evaporation, the diameter of the drops was used. Photographs were taken through the transparent tiles while the experiment was sealed and left undisturbed over the entire course.

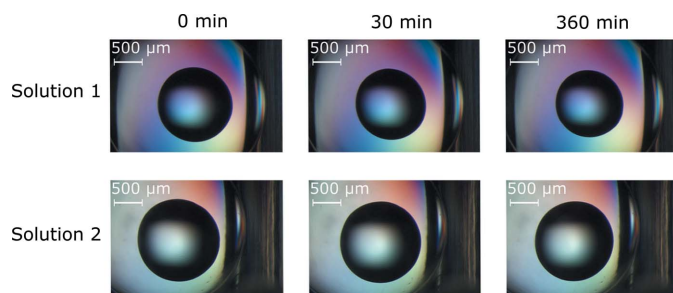


Figure 3

Evaporation test of the device versus exemplary CFS soaking solutions. Depicted are two different soaking solutions used in CFS campaigns at BESSY II, HZB. Both solutions (Solution 1 and Solution 2) already contain DMSO and a cryoprotectant. Immediately after placing the frame on top of the plate and finishing setting all drops, an image at 0 min was taken of each drop as a starting point. The evaporation observations were performed over 6 h at several intermediate time points. The images were taken with the same microscope and settings each time. Depicted here are the 30 and 360 min time points for both conditions tested. Little to no change is observed even after 360 min, thus showing high evaporation protection of the solutions by the frame.

The CFS soaking solutions usually contain the same components of the crystallization conditions and, in addition, a certain amount of dimethyl sulfoxide (DMSO) and cryoprotectant. In a CFS campaign, either crystals are transferred into the soaking solution or DMSO and cryoprotectants are added to the crystallization drop. The crystals will usually be soaked for 30 min to about 24 h, depending on the circumstances. For our purposes, the crystals needed to be transferred into a soaking drop and soaked for about 16 h overnight. The overnight soaking was performed in a foil-sealed plate. Handling or harvesting of 96 crystals usually takes between 1 and 2 h depending on the experimenters' experience and the quality/robustness of the crystals. This means the solutions should optimally stay almost unchanged within that time window to allow for transfer and harvesting of 96 crystals in one session using the frame. Two CFS soaking solutions from previous campaigns (Wollenhaupt *et al.*, 2020) were tested as examples (Fig. 3). Solution 1 contains 17% (*w/v*) PEG 4000, 180 mM Tris pH 8.5, 180 mM Li₂SO₄, 5% (*v/v*) 1,2-propanediol and 8% (*v/v*) DMSO; and solution 2 contains 19.8% (*w/v*) PEG 4000, 68 mM sodium acetate pH 4.6, 68 mM ammonium acetate, 19.3% (*v/v*) glycerol and 9% (*v/v*) DMSO. The drops show minimal evaporation over the course of 6 h, which is a reasonable time frame to carry out crystal soaking and harvesting of about 300 crystals. The results were independent of the position of the wells used on the plate. Several CFS campaigns at BESSY II, Helmholtz-Zentrum Berlin (HZB), have already been successfully performed using the device. Users typically reported comfortable usage and significant speed-up of crystal handling.

4.2. Evaporation reduction – 96-well crystallization solution screen

In protein crystallization experiments, a plethora of crystallization conditions are screened in order to derive diffracting protein crystals. Therefore, numerous crystal-

lization screens have been assembled by combining solutions that potentially facilitate crystal formation (Jancarik & Kim, 1991). The individual solutions typically consist of a combination of precipitant and buffer, sometimes with salts and additives. Considering all possible combinations, an endless number of possible crystallization conditions can be created. Therefore, it is not feasible to perform a comprehensive, all-encompassing experimental evaporation test of all possible combinations. For a first overview, we tested a typical commercial 96-well screen, the JSCG Core Suite II (Quiagen), which includes many common classes of solutions used in macromolecular crystallography (Table S1 of the supporting information). A 3-lens 96-well MRC low-profile crystallization plate was used with 40 µl reservoir and 0.4 µl drops. The reservoir and the sitting drops were pipetted using a 96-syringe head pipetting robot (Gryphon, Art Robbins Instruments) (*i.e.* the drops were pipetted in parallel) before the plate was covered with the frame. The drop size was recorded as described in Section 4.1, leaving the experiments closed.

A typical observation during the first 30 min was that the freshly set drop of each solution decreased slightly in size, probably owing to reaching the evaporation equilibrium inside the sitting-drop experiment. The final decrease of the drop diameter after 360 min was used to assess evaporation. Usually, the drops are not perfectly circular from the top view, but often elliptic. Therefore, the mean between the major and the minor axes of the observed ellipse was taken. Three independent repetitions of the experiment were performed, and the relative drop diameter reductions of those experiments were averaged for each respective solution. Observations of changes in drop sizes were classified into three groups: 'good' – a reduction of less than 15% of the drop diameter; 'medium' – a reduction of about 15–30% of the drop diameter; and 'bad' – a reduction of more than 30% of the drop diameter or drops that show phase separation or crystallization of individual solution components (Fig. 4). In two cases, the conditions B08 and B10 (Table S1), no useful diameter could be measured due to the small surface tension of the solution, which created a large flat drop. These drops were therefore not included in the analysis. From the 94 solutions of the screen used for the analysis, 59 of the solutions were found to be good, 18 were medium and 17 were bad (Table S1). This shows that the device works well for many solutions of this screen. In the experiment, certain tendencies could be seen for the precipitant compound. With a focus on PEGs, which are rather abundant in the screen, it was observed that solutions including smaller PEGs (PEG 200 to PEG 600) are more often categorized as good than larger PEGs (PEG 8000 to PEG 20000). In the case of PEG 6000, additional salt can help to keep the drop size stable. For PEG 1000 to PEG 3350 salt addition seems to have no apparent effect. A salt abundantly used as a precipitant is ammonium sulfate. The categorization is usually good as long as the concentration is above 0.4 M. There is only one exception, where 0.8 M ammonium sulfate is mixed with 0.1 M citric acid pH 4.0 (category medium). Typical alcohols in the screen are ethanol and iso-propanol and solutions containing these alcohols as their main

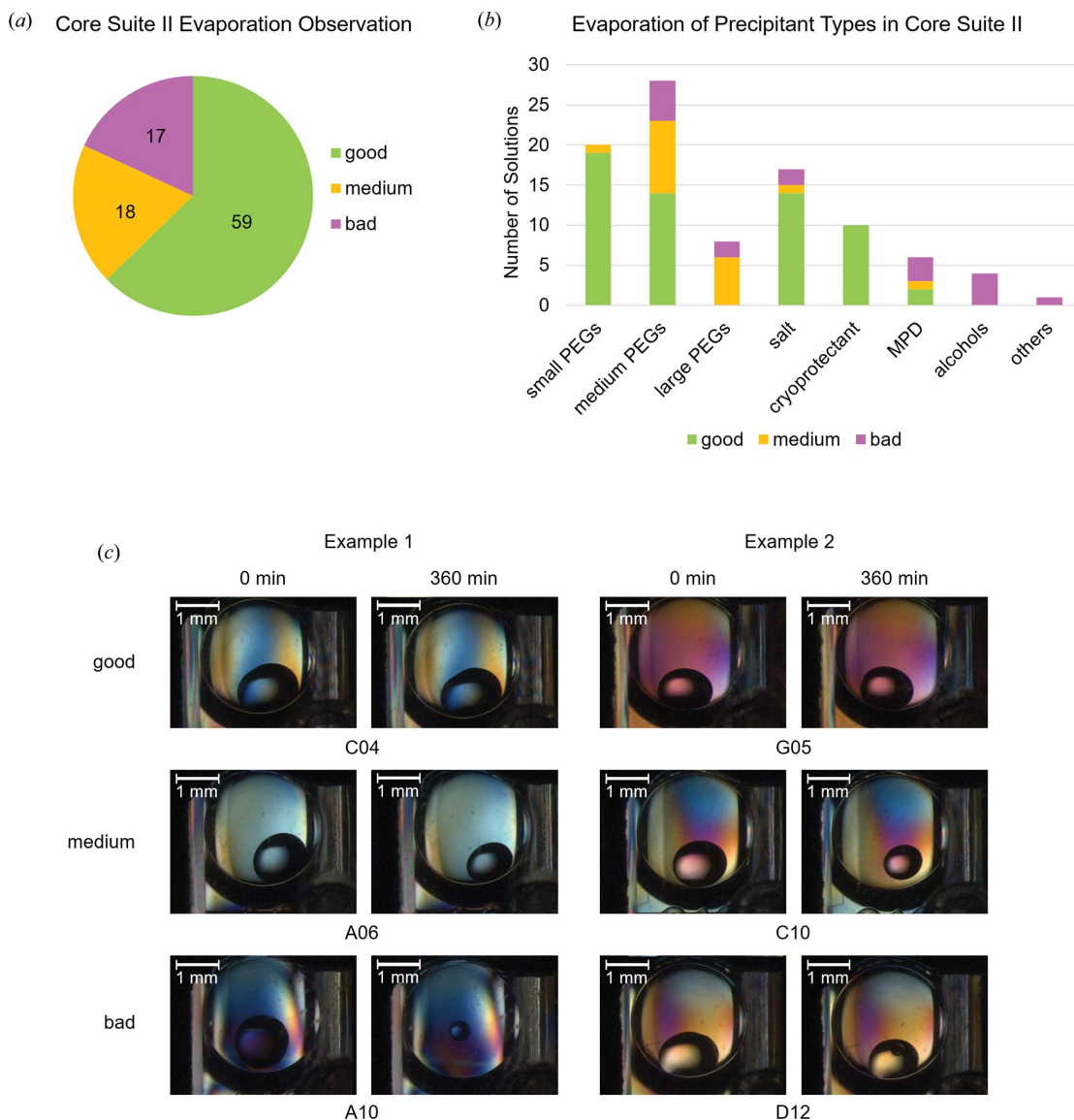


Figure 4 Evaporation test of the device versus the JSCG Core Suite II. The evaporation observations were performed for 6 h following the method described in Fig. 3. (a) Pie chart illustrating the number of solutions categorized as good (green), medium (yellow) and bad (purple). (b) 94 solutions (excluding B08 and B10 due to their low surface tension) were reduced to nine groups that represent different types of precipitants. They show slight tendencies for certain precipitant types. (c) Two examples are shown for each category, each with the 0 and 360 min time points. These time points were used to assess the overall reduction of drop size or, in the case of D12, to detect the appearance of phase separation.

component fell into the bad category. Typical cryoprotectants used in the screen solutions are glycerol, ethylene glycol and 1,2-propanediol. Those conditions were categorized as good. The evaporation results of some precipitants cannot be this easily generalized. In the case of PEG 4000, the solution was usually mixed with glycerol, which might be the actual ingredient turning these solutions into good ones. A similar consideration can be made for 2-methyl-2,4-pentanediol (MPD), where the glycerol-containing solutions were classified as good and the MPD solutions without glycerol as bad.

5. Discussion

All in all, the experiments show that in most cases of the JSCG Core Suite II solutions the user can reliably work for at least

1 h, which is often sufficient to carry out manipulation and harvesting of up to 100 crystals. However, the observations also indicate how unpredictable the evaporation properties seem to be, judging from their components alone. We tried to name certain trends but given the relatively small variety of solutions tested (compared with the vast number of possible combinations in crystallography) no general trends can be applied. Intuitively, the evaporation properties should be connected to the vapor pressure and relative humidity of the individual solutions. There have been several approaches to predict the relative humidity of crystallization conditions on the basis of their main component, to perform controlled dehydration experiments of single crystals outside the crystallization drop (Wheeler *et al.*, 2012; Bowler *et al.*, 2015, 2017).

Online applets for convenient use of the calculations described in the above publications are publicly available (http://go.esrf.eu/UsersAndScience/Experiments/MX/How_to_use_our_beamlines/forms). However, relative humidity values predicted with these formulas based on the main components of the JSCG Core II Suite solutions did not show an obvious relationship by casual comparison with our experimental results of evaporation in crystallization drops. Therefore, factors other than the relative humidity are likely to additionally influence the evaporation properties of crystallization drops in microtiter plate environments. Nevertheless, the device demonstrated excellent performance for more than 50% of the 96 crystallization solutions and the tested soaking solutions from different CFS campaigns. Thus, it could be argued that, in most cases, the device is likely to perform well. However, it is preferable that each solution intended for use for extended time periods should be tested beforehand. This is especially worthwhile for experiments with enhanced throughput like CFS where the composition of the soaking solution is invariant by design.

Unlike the other approach to evaporation reduction specifically made for the setup of crystallization plates mentioned earlier (Zipper *et al.*, 2014), the frame provides sufficient angular space to conveniently manipulate crystals inside the sitting drops. Furthermore, it enables the user to leave open only the well which is being worked on. All other wells remain covered during the entire experiment. In contrast to other devices specifically designed to increase speed for crystal harvesting, like the semi-automated Crystal Shifter (Wright *et al.*, 2021), the frame does not come with integrated automation or digital sample tracking. Still, it has several other advantages. Firstly, it has a much lower production cost as it is simple, made of plastic/acrylic material, and lacks any metal parts, electronics or motors. Secondly, the frame is small and light; thus, it can be transported easily. Although the Crystal Shifter remains the best solution in terms of speed, the frame is more versatile and more widely applicable.

6. Summary

A novel evaporation-protecting device, which we conveniently term the EasyAccess Frame, which minimizes evaporation to facilitate crystal handling and harvesting, is presented. It is small and light and therefore well suited for transportation. The current version of the EasyAccess Frame is designed to fit the 3-lens 96-well MRC low-profile crystallization plate; however, the design can be adapted to any other 96-well crystallization plate in SBS format in the future.

Widely applied experiments in macromolecular crystallography are, for instance, heavy-atom derivatization for phasing and in-drop dehydration experiments to improve diffraction properties of crystals. All these applications can be conveniently performed using the novel tool presented here. It is also possible to customize the frame to a certain extent for different applications. For example, in the case of photo-sensitive protein crystals, the transparent colorless acrylic glass tiles could be exchanged for colored ones, protecting the

crystals from a particular wavelength of light, while still allowing the user to observe the enclosed experiment. For left-handed users, a frame could be designed that fits onto the plate rotated 180° horizontally. All in all, the EasyAccess Frame reduces evaporation in microtiter plates while easing the access to the individual experiments on the plates. In this way it speeds up manual manipulation and harvesting of protein crystals, thereby benefitting X-ray crystallography experiments in general, but especially enhanced-throughput screening experiments. It does not need any special equipment, is re-usable for many different experiments and thus is ideal for every laboratory.

7. Disclosures

A patent application regarding the reported frame device has been filed by Helmholtz-Zentrum Berlin with the German Patent and Trademark Office (registration No. DE 10 2018 111 478.8). Additionally, an international patent application via the PCT route, using the priority of the German patent, has been filed. Furthermore, the handle of the pen tool is protected by a registered design (registration No. DE 40 2018 100 962-0001).

Acknowledgements

We thank all the users of the EasyAccess Frame who provided valuable feedback at several development stages. We are also thankful to the department of Construction at Helmholtz-Zentrum Berlin for their help with 3D printing and access to the 3D-printer facilities. Open access funding enabled and organized by Projekt DEAL.

Funding information

We are thankful for support by the German Federal Ministry of Education and Science (BMBF), via the projects Frag2Xtal and Frag4Lead (grant Nos. 05K13M1; 05K16M1). We are additionally grateful for support via iNEXT-Discovery (project No. 871037), funded by the Horizon 2020 programme of the European Commission.

References

- Bowler, M. G., Bowler, D. R. & Bowler, M. W. (2017). *J. Appl. Cryst.* **50**, 631–638.
- Bowler, M. W., Mueller, U., Weiss, M. S., Sanchez-Weatherby, J., Sorensen, T. L. M., Thunnissen, M. M. G. M., Ursby, T., Gobbo, A., Russi, S., Bowler, M. G., Brockhauser, S., Svensson, O. & Cipriani, F. (2015). *Cryst. Growth Des.* **15**, 1043–1054.
- Bowler, M. W., Svensson, O. & Nurizzo, D. (2016). *Crystallogr. Rev.* **22**, 233–249.
- Cipriani, F., Röwer, M., Landret, C., Zander, U., Felisaz, F. & Márquez, J. A. (2012). *Acta Cryst.* **D68**, 1393–1399.
- Cuttitta, C. M., Ericson, D. L., Scalia, A., Roessler, C. G., Teplitsky, E., Joshi, K., Campos, O., Agarwal, R., Allaire, M., Orville, A. M., Sweet, R. M. & Soares, A. S. (2015). *Acta Cryst.* **D71**, 94–103.
- Deller, M. C. & Rupp, B. (2014). *Acta Cryst.* **F70**, 133–155.

- Douangamath, A., Brandao-Neto, J., Williams, M., Fearn, R., Krojer, T., Aller, P., Henderson, K., Longhi, E., Allianelli, L. & Von Delft, F. (2014). *Acta Cryst.* **A70**, C790.
- Förster, A., Brandstetter, S. & Schulze-Briese, C. (2019). *Philos. Trans. R. Soc. A.* **377**, 20180241.
- Heras, B. & Martin, J. L. (2005). *Acta Cryst.* **D61**, 1173–1180.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Krojer, T., Talon, R., Pearce, N., Collins, P., Douangamath, A., Brandao-Neto, J., Dias, A., Marsden, B. & von Delft, F. (2017). *Acta Cryst.* **D73**, 267–278.
- Leonarski, F., Redford, S., Mozzanica, A., Lopez-Cuenca, C., Panepucci, E., Nass, K., Ozerov, D., Vera, L., Olieric, V., Buntschu, D., Schneider, R., Tinti, G., Froejdh, E., Diederichs, K., Bunk, O., Schmitt, B. & Wang, M. (2018). *Nat. Methods*, **15**, 799–804.
- Lima, G. M. A., Talibov, V. O., Jagudin, E., Sele, C., Nyblom, M., Knecht, W., Logan, D. T., Sjögren, T. & Mueller, U. (2020). *Acta Cryst.* **D76**, 771–777.
- Owen, R. L., Juanhuix, J. & Fuchs, M. (2016). *Arch. Biochem. Biophys.* **602**, 21–31.
- Powell, H. R. (2017). *Biosci. Rep.* **37**, BSR20170227.
- Rould, M. A. (2007). *Methods Mol. Biol.* **364**, 159–182.
- Schiebel, J., Krimmer, S. G., Röwer, K., Knörlein, A., Wang, X., Park, A. Y., Stieler, M., Ehrmann, F. R., Fu, K., Radeva, N., Krug, M., Huschmann, F. U., Glöckner, S., Weiss, M. S., Mueller, U., Klebe, G. & Heine, A. (2016). *Structure*, **24**, 1398–1409.
- Shaw Stewart, P. & Mueller-Dieckmann, J. (2014). *Acta Cryst.* **F70**, 686–696.
- Wheeler, M. J., Russi, S., Bowler, M. G. & Bowler, M. W. (2012). *Acta Cryst.* **F68**, 111–114.
- Wollenhaupt, J., Metz, A., Barthel, T., Lima, G. M. A., Heine, A., Mueller, U., Klebe, G. & Weiss, M. S. (2020). *Structure*, **28**, 694–706.e5.
- Wright, N. D., Collins, P., Koekemoer, L., Krojer, T., Talon, R., Nelson, E., Ye, M., Nowak, R., Newman, J., Ng, J. T., Mitrovich, N., Wiggers, H. & von Delft, F. (2021). *Acta Cryst.* **D77**, 62–74.
- Zipper, L. E., Aristide, X., Bishop, D. P., Joshi, I., Kharzeev, J., Patel, K. B., Santiago, B. M., Joshi, K., Dorsinvil, K., Sweet, R. M. & Soares, A. S. (2014). *Acta Cryst.* **F70**, 1707–1713.



JOURNAL OF
APPLIED
CRYSTALLOGRAPHY

Volume 54 (2021)

Supporting information for article:

Facilitated crystal handling using a simple device for evaporation reduction in microtiter plates

Tatjana Barthel, Franziska U. Huschmann, Dirk Wallacher, Christian G. Feiler, Gerhard Klebe, Manfred S. Weiss and Jan Wollenhaupt

Well	Main component	Component 2	Component 3	Buffer	Precipitant category	Mean diameter reduction (n=3) [%]	Data range (n=3) [%]	Comment	Category
A01	20%(w/v) PEG 8000	0.2 M Sodium chloride		0.1 M CAPS pH 10.5	large PEG	18	16 - 19		medium
A02	1.26 M Ammonium sulfate	0.2 M Sodium chloride		0.1 M CHES pH 9.5	salt	12	10 - 14		good
A03	1.0 M Sodium citrate			0.1 M CHES pH 9.5	salt	1	-2 - 5		good
A04	10%(w/v) PEG 8000	0.2 M Sodium chloride		0.1 M CHES pH 9.5	large PEG	23	15 - 30		medium
A05	10%(w/v) PEG 20000	2%(v/v) 1,4-Dioxane		0.1 M Bicine pH 9.0	large PEG	28	26 - 29		medium
A06	20%(w/v) PEG 550 MME	0.1 M Sodium chloride		0.1 M Bicine pH 9.0	small PEG	16	14 - 18		medium
A07	10%(w/v) PEG 6000	1.0 M Lithium chloride		0.1 M Bicine pH 9.0	medium PEG	15	13 - 17		medium
A08	20%(w/v) PEG 300	5%(w/v) PEG 8000	10%(v/v) Glycerol	0.1M Tris pH 8.5	small PEG	5	3 - 7		good
A09	20%(w/v) PEG 2000 MME	0.01 M Nickel chloride		0.1 M Tris pH 8.5	medium PEG	10	9 - 10		good
A10	20%(v/v) Ethanol			0.1 M Tris pH 8.5	alcohol	63	61 - 64		bad
A11	2.0 M Ammonium dihydrogen phosphate			0.1 M Tris-HCl pH 8.5	salt	11	9 - 14		good
A12	8%(w/v) PEG 8000			0.1 M Tris-HCl pH 8.5	large PEG	13	12 - 15	phase separation	bad
B01	2.0 M Ammonium sulfate			0.1 M Tris-HCl pH 8.5	salt	9	1 - 14		good
B02	40%(v/v) PEG 400	0.2 M Lithium sulfate		0.1M Tris pH 8.5	small PEG	4	2 - 7		good
B03	10%(w/v) PEG 8000	0.2 M Calcium acetate		0.1 M Imidazole pH 8.0	large PEG	27	23 - 33		medium
B04	35%(v/v) MPD	0.2 M Magnesium chloride		0.1 M Imidazole pH 8.0	MPD	13	6 - 19		medium
B05	20%(w/v) PEG 6000	1.0 M Lithium chloride		0.1 M Tris pH 8.5	medium PEG	8	8 - 10		good
B06	20%(w/v) PEG 6000			0.1 M Tris pH 8.5	medium PEG	20	19 - 22		medium
B07	20%(w/v) PEG 3350	0.2 M Lithium Acetate			medium PEG	18	14 - 20		medium
B08	40%(v/v) MPD	0.2 M Magnesium chloride		0.1M Imidazole pH 8.0	MPD	-	-	not possible to measure	-
B09	15%(v/v) Ethanol	0.2 M Magnesium chloride		0.1 M HEPES pH 7.5	alcohol	36	34 - 38		bad
B10	70%(v/v) MPD			0.1 M HEPES pH 7.5	MPD	-	-	not possible to measure	-
B11	17%(w/v) PEG 4000	15%(v/v) Glycerol	8.5%(v/v) Isopropanol	0.085 M Sodium HEPES pH 7.5	medium PEG	5	3 - 7		good
B12	25%(v/v) Glycerol	0.6 M sodium dihydrogen phosphate/0.6 M potassium dihydrogen phosphate		0.075 M Sodium HEPES pH 7.5	cryoprotectant	2	1 - 3		good
C01	27%(v/v) PEG 400	10%(v/v) Glycerol	0.18 M Magnesium chloride	0.09 M Sodium HEPES pH 7.5	small PEG	4	1 - 7		good
C02	2.0 M Ammonium sulfate	2%(v/v) PEG 400		0.1 M Sodium HEPES pH 7.5	salt	5	4 - 7		good
C03	30%(v/v) PEG 400	0.2 M Magnesium chloride		0.1 M Sodium HEPES pH 7.5	small PEG	5	2 - 9		good
C04	50%(v/v) PEG 200	0.2 M Sodium chloride		0.1M Na/K phosphate pH 6.2	small PEG	0	0 - 1		good
C05	20%(w/v) PEG 3350	0.2 M Sodium fluoride			medium PEG	16	15 - 18		medium
C06	2.0 M Ammonium sulfate	0.2 M Lithium sulfate		0.1 M Tris pH 7.0	salt	1	0 - 2		good
C07	40%(v/v) PEG 300	0.2 M Calcium acetate		0.1M Sodium cacodylate pH 6.5	small PEG	1	-1 - 2		good
C08	20%(w/v) PEG 1000			0.1 M Tris pH 7.0	medium PEG	16	12 - 18		medium
C09	10%(w/v) PEG 6000	1.0 M Lithium chloride		0.1 M HEPES pH 7.0	medium PEG	12	11 - 13		good
C10	10%(w/v) PEG 6000			0.1 M HEPES pH 6.5	medium PEG	26	25 - 27		medium
C11	40%(v/v) PEG 400	0.2 M Sodium chloride		0.1M Na/K phosphate pH 6.2	small PEG	1	1 - 1		good
C12	50%(v/v) PEG 200			0.1M Sodium citrate pH 5.5	small PEG	1	1 - 1		good
D01	25%(v/v) 1,2-Propanediol	10%(v/v) Glycerol		0.1M Na/K phosphate pH 6.2	cryoprotectant	4	3 - 6		good
D02	20%(w/v) PEG 3350	0.2 M Sodium nitrate			medium PEG	19	16 - 22		medium
D03	50%(v/v) PEG 200	0.05 M Lithium sulfate		0.1M Tris pH 7.0	small PEG	1	0 - 1		good
D04	20%(w/v) PEG 3350	0.2 M Potassium sulfate			medium PEG	23	21 - 27	crystallization	bad
D05	0.2 M Magnesium formate				salt	48	46 - 51		bad
D06	40%(v/v) PEG 600			0.1M Sodium citrate pH 5.5	small PEG	3	2 - 3		good
D07	20%(w/v) PEG 1000	0.2 M Magnesium chloride		0.1 M Sodium cacodylate pH 6.5	medium PEG	13	11 - 14		good
D08	10%(w/v) PEG 3000	0.2 M Magnesium chloride		0.1 M Sodium cacodylate pH 6.5	medium PEG	23	19 - 27		medium
D09	30%(v/v) PEG 400	0.2 M Lithium sulfate		0.1 M Sodium cacodylate pH 6.5	small PEG	3	2 - 5		good
D10	2.0 M Ammonium sulfate	0.2 M Sodium chloride		0.1 M Sodium cacodylate pH 6.5	salt	1	1 - 1		good
D11	12%(w/v) PEG 20000			0.1 M MES pH 6.5	large PEG	28	28 - 29		medium
D12	20%(w/v) PEG 3350	0.2 M Lithium sulfate			medium PEG	11	11 - 12	phase separation	bad

Well	Main component	Component 2	Component 3	Buffer	Precipitant category	Mean diameter reduction (n=3) [%]	Data range (n=3) [%]	Comment	Category
E01	20%(w/v) PEG 1000	0.2 M Sodium chloride		0.1 M Na/K phosphate pH 6.2	medium PEG	19	16 - 24	phase separation	bad
E02	10%(v/v) MPD			0.1 M MES pH 5.0	MPD	41	40 - 42		bad
E03	20%(w/v) PEG 6000	1.0 M Lithium chloride		0.1 M MES pH 6.0	medium PEG	8	5 - 11		good
E04	10%(w/v) PEG 6000	1.0 M Lithium chloride		0.1 M MES pH 6.0	medium PEG	10	8 - 11		good
E05	5%(w/v) PEG 6000			0.1 M MES pH 5.0	medium PEG	42	41 - 44		bad
E06	25%(v/v) 1,2-Propanediol	10%(v/v) Glycerol	0.2 M Zinc acetate	0.1M Imidazole pH 8.0	cryoprotectant	3	2 - 4		good
E07	40%(v/v) PEG 600	0.2 M Zinc acetate		0.1M Imidazole pH 8.0	small PEG	3	2 - 4		good
E08	30%(v/v) PEG 600	10%(v/v) Glycerol	0.5 M Ammonium sulfate	0.1M Tris pH 7.0	small PEG	1	1 - 1		good
E09	1.0 M Lithium sulfate	0.5 M Ammonium sulfate		0.1 M Sodium citrate pH 5.6	salt	13	12 - 13		good
E10	30%(w/v) PEG 4000	0.2 M Ammonium acetate		0.1 M Sodium citrate pH 5.6	medium PEG	6	4 - 8	phase separation	bad
E11	24%(w/v) PEG 1500	20%(v/v) Glycerol			medium PEG	2	1 - 3		good
E12	40%(v/v) PEG 300	0.2 M Sodium chloride		0.1M Sodium acetate pH 4.5	small PEG	2	2 - 3		good
F01	35%(v/v) MPD	10%(v/v) Glycerol		0.1M Sodium acetate pH 4.5	MPD	9	7 - 10		good
F02	40%(v/v) PEG 300			0.1M Phosphate-citrate pH 4.2	small PEG	3	2 - 4		good
F03	50%(v/v) Ethylene glycol	5%(w/v) PEG 1000		0.1M Sodium acetate pH 4.5	cryoprotectant	1	0 - 2		good
F04	30%(v/v) PEG 200	0.1 M Sodium chloride		0.1M Sodium acetate pH 4.5	small PEG	5	3 - 8		good
F05	40%(v/v) 1,2-Propanediol			0.1M Sodium acetate pH 4.5	cryoprotectant	3	3 - 4		good
F06	40%(v/v) Ethylene glycol			0.1M Sodium acetate pH 4.5	cryoprotectant	2	1 - 2		good
F07	10%(v/v) MPD			0.1 M Sodium acetate pH 5.0	MPD	31	27 - 33		bad
F08	2.4 M Ammonium sulfate			0.1 M Citric acid pH 4.0	salt	4	3 - 5		good
F09	1.6 M Ammonium sulfate			0.1 M Citric acid pH 4.0	salt	9	8 - 9		good
F10	0.8 M Ammonium sulfate			0.1 M Citric acid pH 4.0	salt	23	17 - 33		medium
F11	20%(w/v) PEG 6000	1.0 M Lithium chloride		0.1 M Citric acid pH 5.0	medium PEG	7	6 - 8		good
F12	25%(v/v) 1,2-Propanediol	5%(w/v) PEG 3000	10%(v/v) Glycerol	0.1M Phosphate-citrate pH 4.2	cryoprotectant	1	1 - 1		good
G01	2.0 M Ammonium sulfate	5%(v/v) Isopropanol			salt	12	8 - 16		good
G02	2.0 M Ammonium sulfate				salt	10	9 - 13		good
G03	40%(v/v) PEG 400	0.2 M Magnesium chloride		0.1M MES pH 5.5	small PEG	3	1 - 5		good
G04	1.0 M Hexanediol	0.01 M Cobalt chloride		0.1 M Sodium acetate pH 4.6	alcohol	36	33 - 38		bad
G05	1.6 M Ammonium sulfate	20%(v/v) Glycerol		0.08 M Sodium acetate pH 4.6	salt	1	0 - 1		good
G06	30%(v/v) Glycerol	5.6%(w/v) PEG 4000		0.07 M Sodium acetate pH 4.6	cryoprotectant	4	2 - 8		good
G07	30%(v/v) Glycerol	14%(v/v) Isopropanol	0.14 M Calcium chloride	0.07 M Sodium acetate pH 4.6	cryoprotectant	7	5 - 8		good
G08	20%(w/v) PEG 4000	20%(v/v) Glycerol	0.16 M Ammonium sulfate	0.08 M Sodium acetate pH 4.6	medium PEG	2	1 - 3		good
G09	27%(v/v) MPD	10%(v/v) Glycerol	0.018 M Calcium chloride	0.09 M Sodium acetate pH 4.6	MPD	7	4 - 9		good
G10	2.0 M Ammonium sulfate			0.1 M Sodium acetate pH 4.6	salt	8	7 - 9		good
G11	10%(w/v) PEG 3000	0.2 M Zinc acetate		0.1 M Sodium acetate pH 4.5	medium PEG	29	25 - 34		medium
G12	20%(v/v) PEG 300	10% Glycerol	0.2 M Ammonium sulfate	0.1M Phosphate-citrate pH 4.2	small PEG	3	2 - 5		good
H01	30%(v/v) PEG 400	0.2 M Calcium acetate		0.1 M Sodium acetate pH 4.5	small PEG	9	6 - 11		good
H02	30%(w/v) PEG 8000	0.2 M Lithium sulfate		0.1 M Sodium acetate pH 4.5	large PEG	11	5 - 17	phase separation	bad
H03	25%(v/v) Ethylene glycol				cryoprotectant	15	13 - 17		good
H04	10%(v/v) Isopropanol	0.2 M Lithium sulfate		0.1 M Phosphate-citrate pH 4.2	alcohol	40	35 - 44		bad
H05	20%(w/v) PEG 8000	0.2 M Sodium chloride		0.1 M Phosphate-citrate pH 4.2	large PEG	15	12 - 17		medium
H06	10%(w/v) PEG 1000	10%(w/v) PEG 8000			medium PEG	14	11 - 16		good
H07	25.5%(w/v) PEG 4000	15%(v/v) Glycerol	0.17 M Ammonium sulfate		medium PEG	2	2 - 2		good
H08	30%(w/v) PEG 1500				medium PEG	8	6 - 10		good
H09	0.4 M Ammonium dihydrogen phosphate				salt	45	44 - 47		bad
H10	35%(v/v) 1,4-Dioxane				others	81	78 - 87		bad
H11	10%(v/v) MPD			0.1 M Citric acid pH 2.5	MPD	58	28 - 74		bad
H12	20%(w/v) PEG 6000			0.1 M Citric acid pH 2.5	medium PEG	11	8 - 16		good

FragMAXapp: crystallographic fragment-screening data-analysis and project-management system

Gustavo M. A. Lima,^{a,‡} Elmir Jagudin,^a Vladimir O. Talibov,^a Laila S. Benz,^b Costantino Marullo,^a Tatjana Barthel,^c Jan Wollenhaupt,^c Manfred S. Weiss^c and Uwe Mueller^{c*}

Received 25 September 2020

Accepted 8 April 2021

^aBioMAX, MAX IV Laboratory, Fotongatan 2, 224 84 Lund, Sweden, ^bInstitut für Chemie und Biochemie, Freie Universität Berlin, Thielallee 63, 14195 Berlin, Germany, and ^cMacromolecular Crystallography, Helmholtz-Zentrum Berlin, Albert-Einstein-Strasse 15, 12489 Berlin, Germany. *Correspondence e-mail: uwe.mueller@helmholtz-berlin.de

Edited by M. Schiltz, Fonds National de la Recherche, Luxembourg

‡ Current address: Astex Pharmaceuticals, 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, United Kingdom.

Keywords: fragment screening; high-throughput data analysis; protein crystallography; drug discovery; fragment-based lead discovery.

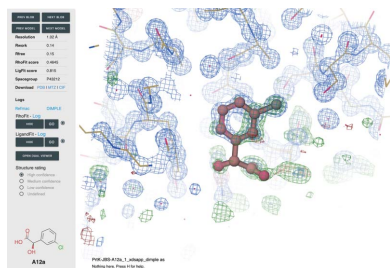
Supporting information: this article has supporting information at journals.iucr.org/d

Crystallographic fragment screening (CFS) has become one of the major techniques for screening compounds in the early stages of drug-discovery projects. Following the advances in automation and throughput at modern macromolecular crystallography beamlines, the bottleneck for CFS has shifted from collecting data to organizing and handling the analysis of such projects. The complexity that emerges from the use of multiple methods for processing and refinement and to search for ligands requires an equally sophisticated solution to summarize the output, allowing researchers to focus on the scientific questions instead of on software technicalities. *FragMAXapp* is the fragment-screening project-management tool designed to handle CFS projects at MAX IV Laboratory. It benefits from the powerful computing infrastructure of large-scale facilities and, as a web application, it is accessible from everywhere.

1. Introduction

Fragment-based lead discovery utilizes a toolbox of biophysical methods, with X-ray crystallography-based fragment screening (CFS) being the main screening technique to obtain 3D structural information on protein–ligand complexes. In its methodological development, CFS co-evolves together with the scientific instrumentation and, more importantly, with the scientific infrastructure that supports contemporary macromolecular X-ray crystallography (MX) (Davies & Tickle, 2011; Grimes *et al.*, 2018). In its current state, CFS relies on experimental setups and data-analysis methodologies that deviate from a classical crystallographic workflow (Krojer *et al.*, 2017; Davies & Tickle, 2011; Wollenhaupt *et al.*, 2021).

Advances in MX data acquisition are setting new standards in structural biology (Grimes *et al.*, 2018). With the advent of MX beamlines at fourth-generation photon sources at synchrotron-radiation facilities such as BioMAX at MAX IV Laboratory (Ursby *et al.*, 2020), Manacá at the SIRIUS Light Source (Nascimento, 2020), AMX and FMX at NSLS II (Fuchs *et al.*, 2016) and MASSIF-1, MASSIF-2 and MASSIF-3 at ESRF–EBS (von Stetten *et al.*, 2020; McCarthy *et al.*, 2018; Svensson *et al.*, 2018), the average data-collection time has been reduced by a factor of ten to less than 40 s for a 360° ω scan. With improved beamline sample capacity allied with reliable and fast robotic sample handling (Martiel *et al.*, 2020; Ursby *et al.*, 2020; Murakami *et al.*, 2020), automated crystal centring (Ito *et al.*, 2019; Di Castro *et al.*, 2008; Schurmann *et al.*, 2019; Svensson *et al.*, 2020) and the possibility of



unattended data collections (Mueller *et al.*, 2017; Cipriani *et al.*, 2012; Nurizzo *et al.*, 2016; Sanchez-Weatherby *et al.*, 2019), contemporary MX experiments generate a massive volume of data within a short time frame, for example more than 100 data sets during a single 8 h session. Finally, advances in crystal-harvesting including new tools (Barthel *et al.*, 2021) as well as new methods such as acoustic crystal harvesting (Samara *et al.*, 2018; Yin *et al.*, 2014), robot-aided harvesting (Wright *et al.*, 2021) and fully automated systems (Cipriani *et al.*, 2012) are increasing the number of crystals harvested by an order of magnitude. Thus, collecting data for CFS at such facilities is a relatively robust procedure and the main bottleneck in MX-based fragment-screening experiments is shifting from sample preparation and data acquisition to data analysis, the building of multiple models and the interpretation of the results.

Currently, automated methods for data processing and structure determination are the best way to deal with the enormous amount of data created in screening experiments. Although those methods are becoming very efficient, one major limitation is still computing power. For example, using a reasonably performing desktop computer, indexing and integrating a single high-resolution data set collected using a high-resolution detector such as an EIGER 16M can take up to 1 h 30 min. Automated and parallel approaches for data analyses benefit from exploring the solutions from several pipelines and cherry-picking the best solution, but the problem of limited computing power is aggravated. To facilitate parallel data processing for hundreds of data sets, data-collection facilities are investing in high-performance computing (HPC). Facilities such as MAX IV provide user access to their computing infrastructure, with the as-yet underestimated advantages of local availability of the data. CFS can benefit greatly from HPC-mediated data processing and analysis, as it improves the quality of the screening results and may possibly increase the sensitivity of the method.

The massive amount of raw experimental data, its processing and the refinement of all potential structures of protein–ligand complexes, the exploration of bound ligands and the corresponding ligand-binding sites, and meta analysis of the data must be treated integrally. This concept was pioneered and first implemented by Astex Therapeutics Ltd (Cambridge, UK; currently part of Astex Pharmaceuticals) within their Pyramid Platform (Davies *et al.*, 2006; Davies & Tickle, 2011). A similar logic was applied by XChem, a fragment-screening facility at Diamond Light Source, Didcot, UK that has operated since 2015. XChem users manage CFS data with a stand-alone graphical application called *XChem Explorer* (Krojer *et al.*, 2017), which is also available within the *CCP4* software suite (Winn *et al.*, 2011). The application guides the user through most stages of the screening experiment, including data processing, analysis, refinement of the models and their deposition in the Protein Data Bank (Berman *et al.*, 2000). Another implementation of the concept was made by EMBL with the web-based *Crystallographic Information Management System (CRIMS)*. This application records all experiments performed with a given target and allows automated data

analysis using *Pipedream* (Vonnrhein *et al.*, 2011; Smart *et al.*, 2014; Bricogne *et al.*, 2017; Kabsch, 2010; Joosten *et al.*, 2011; Collaborative Computational Project, Number 4, 1994).

Here, we present *FragMAXapp*, our solution to facilitate the scientific analysis, presentation and storage of large data sets from CFS and similar high-throughput crystallographic experiments (<https://maxiv.lu.se/fragmax/fragmaxapp>). *FragMAXapp* is a web-based expert system that links all of the steps of crystallographic ligand screening from experimental design and sample preparation to the deposition of the final structures in the Protein Data Bank. The application focuses on user accessibility, flexibility of the data analysis and high performance, linking the MAX IV HPC infrastructure with a selection of automated data-processing and analysis workflows (Fig. 1a). *FragMAXapp* was designed within the operation framework of the BioMAX beamline (Ursby *et al.*, 2020) and connects to other beamline applications, such as the ISPyB sample database (Beteva *et al.*, 2006) and the *MXCuBE* experiment-control software (Mueller *et al.*, 2017; Oscarsson *et al.*, 2019). However, the abstraction level of *FragMAXapp* provides modularity that allows it to be ported and adapted to other sites. In fact, at the time of this publication, working versions of *FragMAXapp* are deployed on BESSY II at the Helmholtz-Zentrum Berlin (Mueller *et al.*, 2015; Wollenhaupt *et al.*, 2021) and at the University of São Paulo using data from the SIRIUS Light Source (Noske *et al.*, 2021).

2. Software design and operation

FragMAXapp has been developed using Python3/Django (Django version 2.2.1; <https://djangoproject.com>) as its back end and using JavaScript for front-end features (Fig. 1b). The project definition, including protein models, the structures of the small molecules and user information, is stored in an SQLite3 database.

As a web application, *FragMAXapp* requires a web browser with the support of WebGL (Web Graphics Library API, Mozilla). This feature allows data analysis to be performed using virtually any device connected to the internet. The user interface is designed using HTML5, CSS3 and JavaScript for responsiveness and functionality. User access can be obtained using the ISPyB authentication system, which provides access to data and results based on the MAX IV user-account system, or with a local account with access privileges granted by the system administrator.

The application offers a variety of choices for data processing, including *DIALS* (Winter *et al.*, 2018) and *XDS* (Kabsch, 2010) through *xia2* (Winter *et al.*, 2018; Evans, 2006; Gildea *et al.*, 2011), *autoPROC* (Vonnrhein *et al.*, 2011), *XDSAPP* (Krug *et al.*, 2012; Sparta *et al.*, 2016), *fastdp* (Winter & McAuley, 2011) and *EDNAproc* (Incardona *et al.*, 2009). Automated structure refinement is performed by *DIMPLE*, *BUSTER* (Bricogne *et al.*, 2017; Smart *et al.*, 2012) and *fspipeline* (Schiebel *et al.*, 2016). The ligand-finding step is performed by *Rhofit* (Smart *et al.*, 2014), *Phenix LigandFit* (Terwilliger *et al.*, 2006, 2007) and *PanDDA* (Pearce *et al.*, 2017), with ligand coordinates and restraints generated by

Phenix eLBOW (Moriarty *et al.*, 2009), *GRADE* (Smart *et al.*, 2011) or *AceDRG* (Long *et al.*, 2017). Within the application interface, structural representation of the ligands is created using the *RDKit* (Landrum, 2012) and *3Dmol* (Rego & Koes, 2015) libraries. Protein atomic coordinates, electron-density maps and reciprocal space are displayed using the *UglyMol* (Wojdyr, 2016) library with minor modifications that allow the visualization and comparison of multiple models. Interactive plots for data statistics are created using the *D3* library (Bostock, 2017). The X-ray diffraction representations are created using *Advx* (Arvai, 2012) with sufficient images to cover 1° of crystal oscillation. The selection of software available in a specific facility is defined during the deployment of the application.

The MAX IV HPC infrastructure manages its workload using Slurm, which is interfaced with the *FragMAXapp* server using a Celery workload manager and a Redis server. These two services run together with the Django server to enable promise-like requests for data while maintaining fluid navigation in the web app.

All services run from a Docker container allocated in a virtual machine configured with GitHub's continuous integration feature, allowing easy deployment and updating of the tools. The Docker deployment uses NGINX/uWSGI to serve the application.

2.1. Project management

The application organizes experimental data and their analysis into application-specific projects. A *FragMAXapp* project is defined by a unique combination of four parameters: the protein name, a common identifier of the compounds (for example the name of the fragment library), the BioMAX proposal number and the corresponding data-collection sessions (Supplementary Fig. S1). These values can be updated to incorporate follow-up experiments related to the same protein within the same proposal.

To facilitate work with sensitive information and to allow users to follow up specific policies for data handling, the project can be created in an encrypted mode. With this option the project-related data are encrypted, and are decrypted only temporarily when the user or the processing software require access. For increased security, the user can manage the

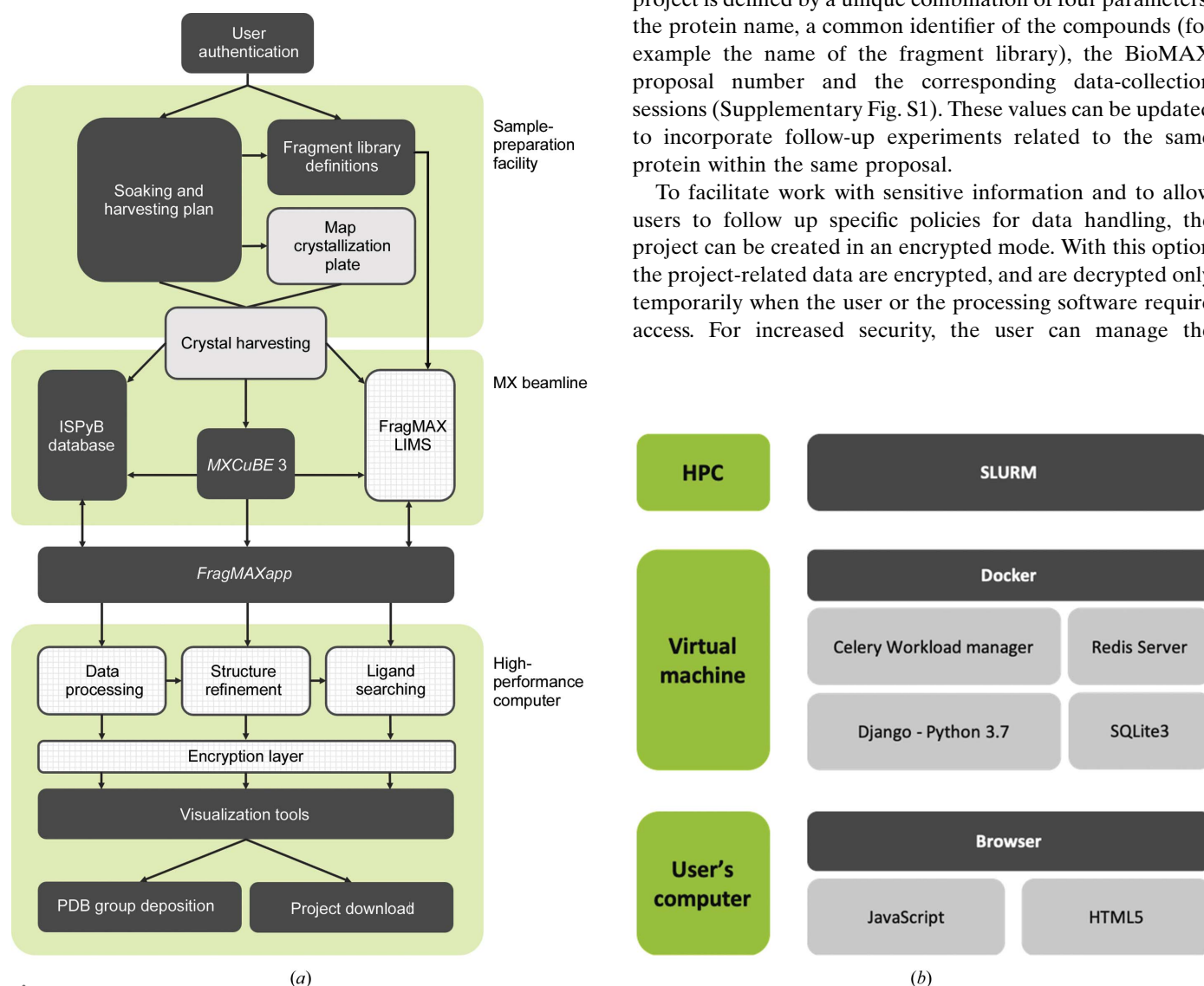


Figure 1
(a) *FragMAXapp* schematic diagram. The diagram shows connected steps and the relationship between each interaction, process and database entry. It is divided into three groups based on where the actions take place. Steps coloured dark grey use the HTML interface for interaction, light grey steps are tasks performed in the experimental laboratory and chequered steps are tasks performed in the background (for example HPC jobs). **(b)** *FragMAXapp* development design. Processing jobs are controlled by a Slurm workload manager installed in the front end of the MAX IV HPC. The three servers and the database necessary for *FragMAXapp* are deployed inside a Docker container inside a virtual machine, ensuring performance and stability of the installation. The user interface is designed using HTML5, CSS3 and JavaScript, which are available in all modern browsers.

research papers

availability of the encryption key. It is possible to download, remove and re-upload the encryption key that is used for the project. When the encryption key is removed, the project data effectively become inaccessible unless the user re-uploads the key.

Beyond meta-management of the projects, *FragMAXapp* allows the examination of individual data sets and other experimental details. Often, this information is stored in sample databases such as ISPyB and can be retrieved at any time by the user. With *FragMAXapp*, browsing this information is also possible (Fig. 2).

2.2. Sample management

Definition of the screening collection is an essential step in project management. *FragMAXapp* requires the user to upload a comma-separated file (CSV) with library specifications, including compound identifiers, which must be identical to the sample names, and optional SMILES strings for the chemical structures of the ligands. The library view page displays updated information about the sample database, with 2D and/or 3D representations of each molecule, and the SMILES code for each compound ID in the project (Fig. 3). Missing entries in the sample database are displayed in the page header, based on the sample names available and the compound IDs in the database. It is possible to update the definition of the samples, which allows work with follow-up experiments within the same project.

2.3. Data processing and structure solution

The most powerful feature of *FragMAXapp* is its capability to handle numerous data-analysis methods in a parallel and multiplexed manner (Fig. 4). It is widely accepted that no single combination of processing and analysis software outperforms any available solution, meaning that only after testing all possible combinations of automated pipelines is it possible to identify the best strategy for a particular data set (Powell, 2017). As we focus on fully automated data-analysis schemes, *FragMAXapp* offers a combination of six data-processing software packages: *autoPROC*, *xia2/DIALS*, *xia2/XDS*, *EDNAproc*, *fastdp* and *XDSAPP*, three structure-refinement pipelines, *BUSTER*, *DIMPLE* and *fspipeline*, and three automatic ligand-searching programs, *Phenix LigandFit*, *Rhofit* and *PanDDA*. Selecting all available software can generate a total of 54 combinations for the evaluation of each diffraction data set. For each software, *FragMAXapp* suggests options to optimize its function, such as the definition of resolution cutoffs for the data, alternative frame ranges for the data sets, the application of Friedel's law during processing, the water-placement method in structure-refinement stages and many more, efficiently acting as a unified GUI front end. Providing customized parameters for data analyses is optionally possible; otherwise the application chooses default values when no input is given. Using a built-in selection table, only selected data sets can be analysed or re-analysed. To ensure that all software will have the necessary information to run properly, *FragMAXapp* keeps a set of mandatory inputs.

Data Collection Overview

Search for Sample names..

Info/Process	Protein	Sample name	Run	Proc Status	Ref Status	Lig Status	Resolution [Å]	Frames	Crystal Picture	Ligand
INFO	PrtK	A10a	1	<ul style="list-style-type: none">● autoPROC● XIA2/DIALS● EDNA_proc● fastdp● XDSAPP● XIA2/XDS● Pipedream	<ul style="list-style-type: none">● BUSTER● DIMPLE● fspipeline● Pipedream	<ul style="list-style-type: none">● LigandFit● RhoFit● Pipedream	1.24	3600		
INFO	PrtK	A11a	1	<ul style="list-style-type: none">● autoPROC● XIA2/DIALS● EDNA_proc● fastdp● XDSAPP● XIA2/XDS● Pipedream	<ul style="list-style-type: none">● BUSTER● DIMPLE● fspipeline● Pipedream	<ul style="list-style-type: none">● LigandFit● RhoFit● Pipedream	1.24	3600		

● Successful
● Not started
● Failed

Figure 2

FragMAXapp Project Overview page. It displays information about the data collection and the status of data processing. A full representation of the user interface is available in Supplementary Fig. S2.

A dedicated tab is available to check all currently running data-analysis jobs and inspect the processing logs. This feature displays live information from the computing nodes and the queueing system. Within this page, it is possible to remove jobs from the processing queue or to kill running analyses.

FragMAXapp integrates *PanDDA* analysis, with several features that allow the sensitivity of the screening experiment to be increased. Under the *PanDDA* tab on the Data Analysis page, it is possible to choose which data-analysis output will be used to run *PanDDA*. If the result is missing, *FragMAXapp* will check whether a solution was offered by other combinations of analysis software. Alternatively, the application offers

the *FragPLEX* option, a method to decide which results to submit for *PanDDA* analysis. The method is based on the evaluation of processing and refinement statistics. The first step of the selection process discards all processed data sets with an overall R_{meas} of $>10\%$. The second step ranks solutions by higher ISa value and resolution and lower $R_{\text{free}}/R_{\text{work}}$. The selection is saved in the *FragMAXapp* database. In our tests, applying this selection logic allows the retrieval of all hits found by combining hits from all individual analysis (*i.e.* a single combination of data processing and structure refinement). Additionally, *FragMAXapp* offers the option to build ground-state models using exclusively known apo structures

or to use data sets without peaks in the Z-maps (the latter requires at least one *PanDDA* analysis before use). In the final steps of *PanDDA* analysis, the evaluation of modelled ligands can be performed through the *PanDDA* Giant scoring scripts (Pearce *et al.*, 2017), and the radar plots of fitting quality generated in this step are shown in the web app.

JBS Library

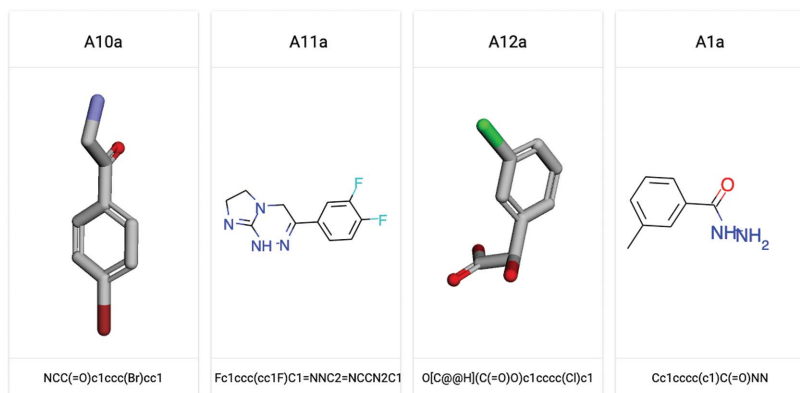


Figure 3
FragMAXapp sample-management page. The 2D representations of the ligands are generated using *RDKit*, while 3D representations are displayed using the *3Dmol* library. This page allows additions or updates to the sample definitions within the project. A full representation of the user interface is available in Supplementary Fig. S3.

Figure 4
FragMAXapp Data Analysis page. An extensive selection of software and pipelines is available to analyse the data. A full representation of the user interface is available in Supplementary Fig. S4.

2.4. Data visualization

The results generated by the automatic analysis pipelines can be visualized in the web browser without any external tools or plugins. It is possible to open and read the output logs from each software used, visualize reciprocal-lattice representation and load refined models with electron density with a *Coot*-like look and feel using *UglyMol* (Fig. 5). Besides, the web app provides interactive and comprehensive overview plots for many statistical parameters such as resolution, ISa, $R_{\text{work}}/R_{\text{free}}$ and unit-cell parameters. This allows the whole CFS data evaluation to be compared within one view, including averaged values and standard deviations for various combinations of methods. These plots can be extremely helpful to obtain a quick overview of the data analysis and to identify data sets that require reprocessing or manual examination.

The *PanDDA* results visualization within *FragMAXapp* displays the event map and the average map obtained from the ground-state model side by side (Supplementary Fig. S6). Using this visual feature, it is easier to interpret unexplained or modelled densities that are also present in the averaged ligand-free model.

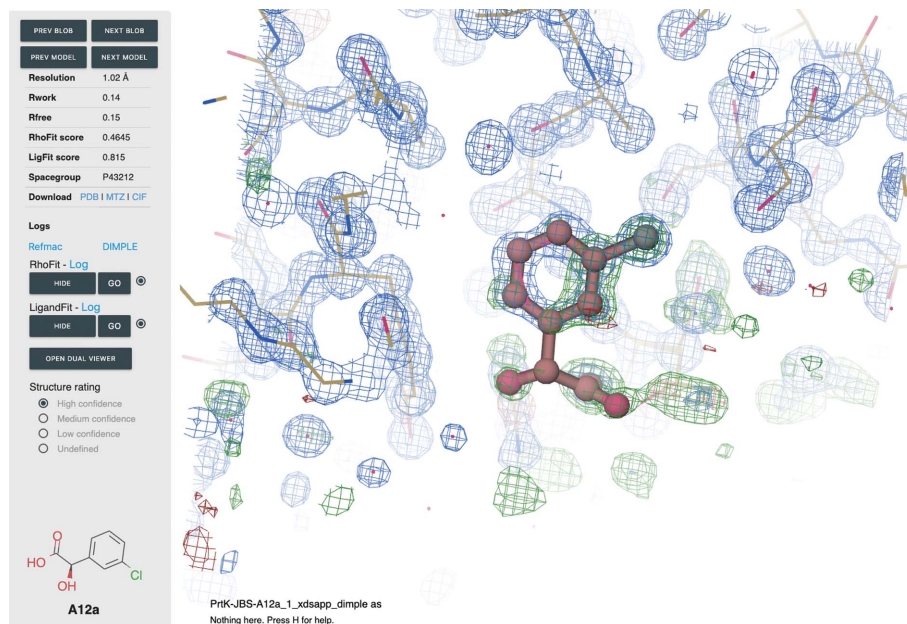


Figure 5
FragMAXapp density viewer. The viewer features unexplained blobs and navigation between the models, information about the result, log access, ligand navigation for automatic ligand-fitting results and structure scores. The density viewer is based on *UglyMOL*. A full representation of the user interface is available in Supplementary Fig. S5.

2.5. Export tools

After the analyses are finished, the web app provides tools to download the data. A variety of choices for which data to export is available, combining process data, log files, structure files and electron densities. Pre-processing results using *PDB-REDO* (Joosten *et al.*, 2012) before deposition is optional but is highly recommended as it improves the quality of automatically refined structures and prepares them for subsequent deposition in the PDB. Using this feature, users are no longer required to manually deposit every model; instead, a compressed archive file with structures, reflection data and the index file is generated, ready to be uploaded using the

OneDep system. The compressed file can be downloaded from *FragMAXapp* or using regular tools available at MAX IV. At the time of publication, SFTP and Globus (Foster, 2011) are offered to users. A final option supported by *FragMAXapp* is uploading of the raw data to public diffraction databases such as *proteindiffraction.org* (Grabowski *et al.*, 2016).

3. Fragment screening using *FragMAXapp*

To test our implementation of the automated methods, a fragment-screening campaign was performed using proteinase K (PROK) from *Tritirachium album* (UniProt accession code

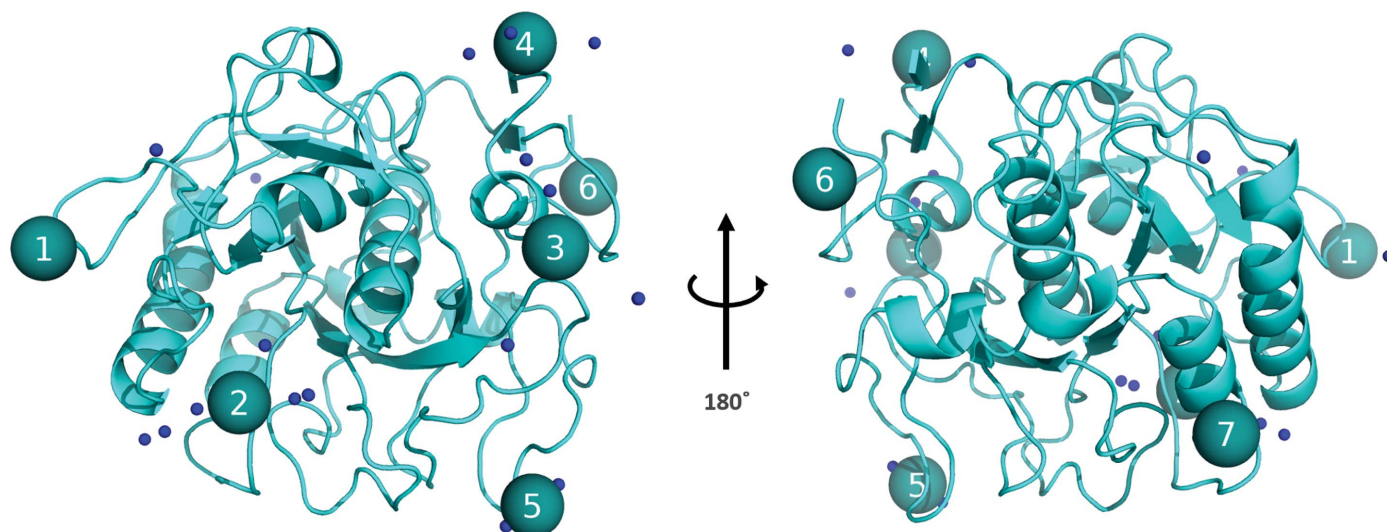
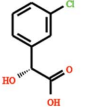
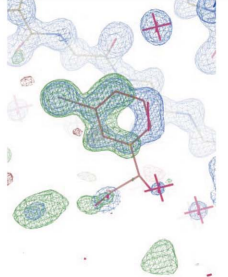
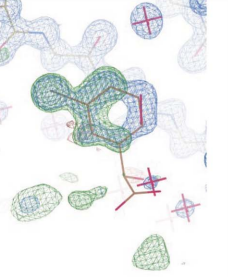
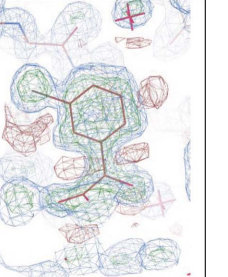
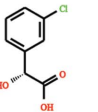
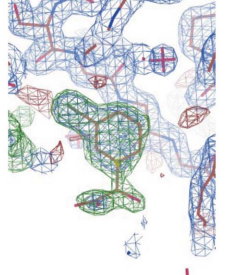
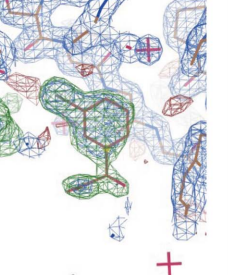
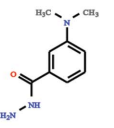
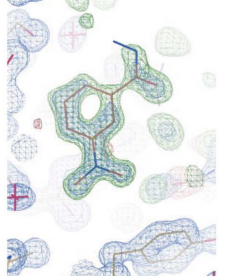
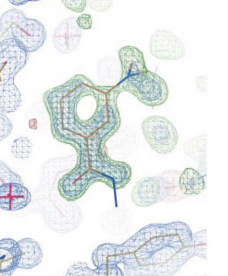
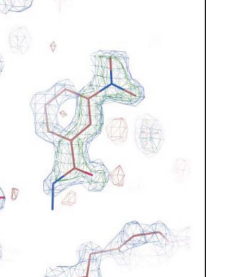
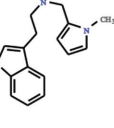
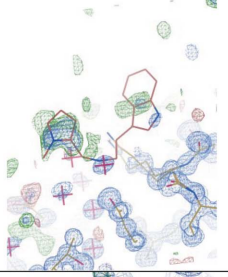
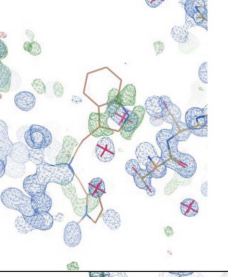
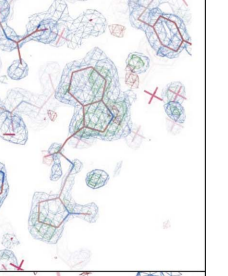
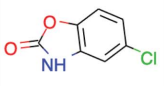
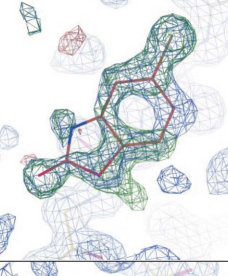
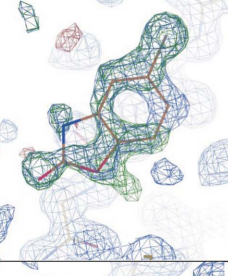
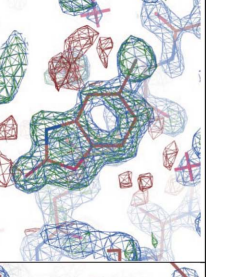
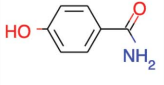
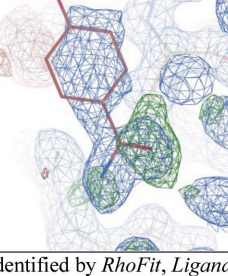
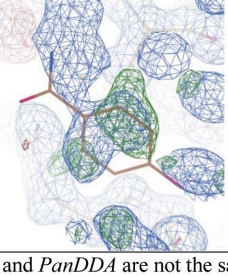
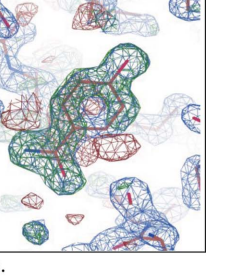


Figure 6
 Proteinase K sites numbered from 1 to 7 as identified by *PanDDA* analysis.

Data set information	<i>LigandFit</i>	<i>Rhofit</i>	<i>PanDDA</i>
A12  <i>autoPROC fspipeline</i> 5roq			
A12 RT  <i>xia2/DIALS fspipeline</i> 5rop			No solution
B5  <i>EDNAproc fspipeline</i> 5rol			
C11*  <i>autoPROC fspipeline</i> 5rof			
F6  <i>xia2/DIALS fspipeline</i> 5row			
H5  <i>EDNAproc fspipeline</i> 5rpm			

* The binding sites identified by *RhoFit*, *LigandFit* and *PanDDA* are not the same.

P06873). PROK is a study model of serine proteases (Larson *et al.*, 2009). The materials and methods used to obtain PROK crystals, including the soaking and harvesting procedures, are described in Section S1. Data-collection parameters are given in Table 1.

Fragment screening of the 96 samples prepared as ligand soaks using the Frag Xtal Screen library, and an extra 40 apo crystals, was performed and data analysis was performed using *FragMAXapp*. The plots, with an overview of data processing and structure refinement, are available in Supplementary Table S1. Applying different automated ligand-fitting (*LigandFit* and *Rhofit*) and ligand-searching (*PanDDA* with manual inspection of Z-maps) methods led to very time-efficient hit identification in two weeks of data analysis, including post-refinement and preparation for deposition. This is remarkable when compared with the expected time required to perform the analysis of 136 data sets. At the time of writing this article (GitHub commit <https://github.com/FragMAX/FragMAXapp/commit/14199af229a43b25a5994c5d58bc92df523-cb31d>), ligand modelling requires *Coot* (Emsley *et al.*, 2010), which can be achieved by accessing a virtual computer at synchrotron facility. *PanDDA* analysis was performed by allowing *FragMAXapp* to select the best combination of data processing and structure refinement; the selection is available in Section S1. *PanDDA* identified events and clustered them into seven sites (Fig. 6), which were compared with standard difference density map analysis and automated ligand-fitting methods (Fig. 7). In general, when the unexplained density was sufficiently defined to fit a large portion of the corresponding fragment, all methods modelled the bound ligand with a certain degree of precision and would

Figure 7

Automated ligand-searching comparison. The *LigandFit* and *Rhofit* columns take the highest scored ligand from each method. The *PanDDA* column shows ligands that were manually modelled. Ligands with a different binding site for different methods are annotated. All structures and density maps are publicly available.

Table 1
Data-collection parameters for PROK crystals at BioMAX.

	Cryo temperature	Room temperature
Wavelength (Å)	0.976	0.976
No. of images	3600	1800
Oscillation (°)	0.1	0.20 (helical scan)
Exposure (ms)	11	10
Photon flux (photons s ⁻¹)	~1 × 10 ¹²	~1 × 10 ¹¹
Beam size (µm)	50 × 50	50 × 50
Resolution (at the edge) (Å)	1.24	1.24

require little real-space refinement to make it ready for deposition. However, using the information from the *PanDDA* event map is likely to improve the model, as shown for fragments A12 and B5 in Fig. 7. In several cases, such as fragments B7 and C11, automated fitting methods disagree on the site in which to locate the ligand, making the *PanDDA* event map decisive in ligand modelling. To highlight how each of these methods compares with each other and what the user can typically expect from the automated methods in *FragMAXapp*, no further modelling was performed on the protein–ligand complexes. Overall, *PanDDA* led to an increase in the hit rate, from ten to 18 identified binding events, and a better modelling of bound fragments with partial occupancy.

The results from the PROK fragment screening and the apo structures used to build the ground-state model in *PanDDA* were submitted to the Protein Data Bank (PDB) using the export tool built into *FragMAXapp*. Overall, this fragment-screening campaign is an example of the application of the software and its capabilities, demonstrating what the user can expect from the web app. The fragment-screening platform of MAX IV, including *FragMAXapp*, has also been used for the validation of a structurally diverse fragment library collection, F2X-Entry, on two protein targets (Wollenhaupt *et al.*, 2020).

4. Conclusions

Crystallographic fragment screening is becoming a common method in the early stages of drug-discovery projects, sometimes replacing other biophysical methods as the primary screening technique. Advances in light-source facilities have shifted the bottleneck from generating data to analysing data, requiring equal efforts to improve software support to enable CFS. The strong automation provided at modern MX beamlines, not only for data collection but also for data processing, enables medicinal chemistry-focused groups to utilize structural biology within their research.

FragMAXapp was developed with a focus on providing the necessary tools to manage, process, analyse and export fragment-screening projects from a web browser. The host facility benefits from software that is easy to deploy and develop, based on popular programming languages such as Python and JavaScript. With our plugin-based system, adding or removing functionalities is possible, thus adapting *FragMAXapp* to a variety of scenarios. The user will benefit from a zero-installation setup that is 'ready to go' from any internet-

connected device. Furthermore, remote data analysis significantly decreases the time required to obtain results by skipping data transfer and using highly parallelized computing architecture. Similar solutions available at other sites such as *XChemExplorer* and *CRIMS* lack the extensive data-analysis features and/or easy access to the application through a web browser. Therefore, *FragMAXapp* will set the standard for CFS project-management and data-processing tools with its full web-stack implementation, parallel computing support and user-friendly interface.

5. Availability

Documentation and source code for *FragMAXapp* can be obtained from the FragMAX project GitHub page (<https://github.com/FragMAX/FragMAXapp>) free of charge for academic use or noncommercial applications. Detailed tutorial and other instructions can be found on the *FragMAXapp* project page (<https://fragmax.github.io/>).

Acknowledgements

The FragMAX group would like to thank all of the BioMAX staff for supporting and improving the beamline that makes the project possible. The authors would like to thank Lionel Trésaugues (Sprint Biosciences AB, Sweden) and Yousif Ali (Linköping University, Sweden) for collaboration during the development of the *FragMAXapp* encryption mode and for testing the software. The authors would like to thank the PReSTO group for providing the scientific software infrastructure, especially Martin Moche (Karolinska Institutet, Sweden) and Sebastian Thorarensen (Linköping University, Sweden) for promptly providing help when needed. The authors would also like to thank Zdenek Matej (MAX IV Laboratory) for valuable help with the MAX IV computing environment. Finally, the authors would like to thank Nicholas Pearce (Utrecht University, Netherlands) for his insights regarding *PanDDA* usage and web implementation. The computations and software access were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at MAX IV, partially funded by the Swedish Research Council.

Funding information

The following funding is acknowledged: Vetenskapsrådet (grant No. 2018-06454; grant No. 2018-05973); VINNOVA (grant No. 2019-02567). We acknowledge the support by iNEXT-Discovery, project number 871037, funded by the Horizon 2020 program of the European Commission.

References

- Arvai, A. (2012). *Adxv – A Program to Display X-ray Diffraction Images*. <http://www.scripps.edu/~arvai/adxv.html>.
- Barthel, T., Huschmann, F. U., Wallacher, D., Feiler, C. G., Klebe, G., Weiss, M. S. & Wollenhaupt, J. (2021). *J. Appl. Cryst.* **54**, 376–382.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

- Beteva, A., Cipriani, F., Cusack, S., Delageniere, S., Gabadinho, J., Gordon, E. J., Guijarro, M., Hall, D. R., Larsen, S., Launer, L., Lavault, C. B., Leonard, G. A., Mairs, T., McCarthy, A., McCarthy, J., Meyer, J., Mitchell, E., Monaco, S., Nurizzo, D., Pernot, P., Pieritz, R., Ravelli, R. G. B., Rey, V., Shepard, W., Spruce, D., Stuart, D. I., Svensson, O., Theveneau, P., Thibault, X., Turkenburg, J., Walsh, M. & McSweeney, S. M. (2006). *Acta Cryst.* **D62**, 1162–1169.
- Bostock, M. (2017). *D3: Bring Data to Life with SVG, Canvas and HTML*. <https://github.com/d3/d3>.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vornrhein, C. & Womack, T. O. (2017). *BUSTER* version 2.10.3. Global Phasing Ltd, Cambridge, UK.
- Cipriani, F., Röwer, M., Landret, C., Zander, U., Felisaz, F. & Márquez, J. A. (2012). *Acta Cryst.* **D68**, 1393–1399.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Davies, T. G. & Tickle, I. J. (2011). *Top. Curr. Chem.*, **317**, 33–59.
- Davies, T. G., Van Montfort, R. L. M., Williams, G. & Jhoti, H. (2006). *Fragment-based Approaches in Drug Discovery*, edited by W. Jahnke & D. A. Erlanson, pp. 193–214. Weinheim: Wiley-VCH.
- Di Castro, M., Pazos, A., Ridoutt, F., Ristau, U. & Fiedler, S. (2008). *PCaPAC 2008 – International Workshop on Personal Computers and Particle Accelerator Controls*, pp. 192–194.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Foster, I. (2011). *IEEE Internet Comput.* **15**, 70–73.
- Fuchs, M. R., Bhogadi, D. K., Jakoncic, J., Myers, S., Sweet, R. M., Berman, L. E., Skinner, J., Idir, M., Chubar, O., McSweeney, S. & Schneider, D. K. (2016). *AIP Conf. Proc.* **1741**, 030006.
- Gildea, R. J., Bourhis, L. J., Dolomanov, O. V., Grosse-Kunstleve, R. W., Puschmann, H., Adams, P. D. & Howard, J. A. K. (2011). *J. Appl. Cryst.* **44**, 1259–1263.
- Grabowski, M., Langner, K. M., Cymborowski, M., Porebski, P. J., Sroka, P., Zheng, H., Cooper, D. R., Zimmerman, M. D., Elsliger, M.-A., Burley, S. K. & Minor, W. (2016). *Acta Cryst.* **D72**, 1181–1193.
- Grimes, J. M., Hall, D. R., Ashton, A. W., Evans, G., Owen, R. L., Wagner, A., McAuley, K. E., von Delft, F., Orville, A. M., Sorensen, T., Walsh, M. A., Ginn, H. M. & Stuart, D. I. (2018). *Acta Cryst.* **D74**, 152–166.
- Incardona, M.-F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N. & Svensson, O. (2009). *J. Synchrotron Rad.* **16**, 872–879.
- Ito, S., Ueno, G. & Yamamoto, M. (2019). *J. Synchrotron Rad.* **26**, 1361–1366.
- Joosten, R. P., Joosten, K., Cohen, S. X., Vriend, G. & Perrakis, A. (2011). *Bioinformatics*, **27**, 3392–3398.
- Joosten, R. P., Joosten, K., Murshudov, G. N. & Perrakis, A. (2012). *Acta Cryst.* **D68**, 484–496.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Krojer, T., Talon, R., Pearce, N., Collins, P., Douangamath, A., Brandao-Neto, J., Dias, A., Marsden, B. & von Delft, F. (2017). *Acta Cryst.* **D73**, 267–278.
- Krug, M., Weiss, M. S., Heinemann, U. & Mueller, U. (2012). *J. Appl. Cryst.* **45**, 568–572.
- Landrum, G. (2012). *RDKit. The Official Sources for the RDKit Library*. <https://github.com/rdkit/rdkit>.
- Larson, S. B., Day, J. S., Nguyen, C., Cudney, R. & McPherson, A. (2009). *Acta Cryst.* **F65**, 192–198.
- Long, F., Nicholls, R. A., Emsley, P., Gražulis, S., Merkys, A., Vaitkus, A. & Murshudov, G. N. (2017). *Acta Cryst.* **D73**, 112–122.
- Martiel, I., Buntschu, D., Meier, N., Gobbo, A., Panepucci, E., Schneider, R., Heimgartner, P., Müller, D., Bühlmann, K., Birri, M., Kaminski, J. W., Leuenberger, J., Oliéric, V., Gletting, W. & Wang, M. (2020). *J. Synchrotron Rad.* **27**, 860–863.
- McCarthy, A. A., Barrett, R., Beteva, A., Caserotto, H., Dobias, F., Felisaz, F., Giraud, T., Guijarro, M., Janocha, R., Khadrrouche, A., Lentini, M., Leonard, G. A., Lopez Marrero, M., Malbet-Monaco, S., McSweeney, S., Nurizzo, D., Papp, G., Rossi, C., Sinoir, J., Sorez, C., Surr, J., Svensson, O., Zander, U., Cipriani, F., Theveneau, P. & Mueller-Dieckmann, C. (2018). *J. Synchrotron Rad.* **25**, 1249–1260.
- Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. (2009). *Acta Cryst.* **D65**, 1074–1080.
- Mueller, U., Förster, R., Hellmig, M., Huschmann, F. U., Kastner, A., Malecki, P., Pühringer, S., Röwer, M., Sparta, K., Steffien, M., Ühlein, M., Wilk, P. & Weiss, M. S. (2015). *Eur. Phys. J. Plus*, **130**, 141.
- Mueller, U., Thunnissen, M., Nan, J., Eguiraun, M., Bolmsten, F., Milàn-Otero, A., Guijarro, M., Oscarsson, M., de Sanctis, D. & Leonard, G. (2017). *Synchrotron Radiat. News*, **30**, 22–27.
- Murakami, H., Hasegawa, K., Ueno, G., Yagi, N., Yamamoto, M. & Kumasaka, T. (2020). *Acta Cryst.* **D76**, 155–165.
- Nascimento, A. F. Z. (2020). *11th International Workshop on X-ray Radiation Damage to Biological Samples*.
- Noske, G. D., Nakamura, A. M., Gawriljuk, V. O., Fernandes, R. S., Lima, G. M. A., Rosa, H. V. D., Pereira, H. D., Zerí, A. C. M., Nascimento, A. F. Z., Freire, M. C. L. C., Oliva, G. & Godoy, A. S. (2021). *bioRxiv*, 2020.12.23.424149.
- Nurizzo, D., Bowler, M. W., Caserotto, H., Dobias, F., Giraud, T., Surr, J., Guichard, N., Papp, G., Guijarro, M., Mueller-Dieckmann, C., Flot, D., McSweeney, S., Cipriani, F., Theveneau, P. & Leonard, G. A. (2016). *Acta Cryst.* **D72**, 966–975.
- Oscarsson, M., Beteva, A., Flot, D., Gordon, E., Guijarro, M., Leonard, G., McSweeney, S., Monaco, S., Mueller-Dieckmann, C., Nanao, M., Nurizzo, D., Popov, A., von Stetten, D., Svensson, O., Rey-Bakaikoa, V., Chado, I., Chavas, L., Gadea, L., Gourhant, P., Isabet, T., Legrand, P., Savko, M., Sirigu, S., Shepard, W., Thompson, A., Mueller, U., Nan, J., Eguiraun, M., Bolmsten, F., Nardella, A., Milàn-Otero, A., Thunnissen, M., Hellmig, M., Kastner, A., Schmuckermaier, L., Gerlach, M., Feiler, C., Weiss, M. S., Bowler, M. W., Gobbo, A., Papp, G., Sinoir, J., McCarthy, A., Karpics, I., Nikolova, M., Bourenkov, G., Schneider, T., Andreu, J., Cuní, G., Juanhuix, J., Boer, R., Fogh, R., Keller, P., Flensburg, C., Paciorek, W., Vornrhein, C., Bricogne, G. & de Sanctis, D. (2019). *J. Synchrotron Rad.* **26**, 393–405.
- Pearce, N. M., Krojer, T., Bradley, A. R., Collins, P., Nowak, R. P., Talon, R., Marsden, B. D., Kelm, S., Shi, J., Deane, C. M. & von Delft, F. (2017). *Nat. Commun.* **8**, 15123.
- Powell, H. R. (2017). *Biosci. Rep.* **37**, BSR20170227.
- Rego, N. & Koes, D. (2015). *Bioinformatics*, **31**, 1322–1324.
- Samara, Y. N., Brennan, H. M., McCarthy, L., Bollard, M. T., Laspina, D., Wlodek, J. M., Campos, S. L., Natarajan, R., Gofron, K., McSweeney, S., Soares, A. S. & Leroy, L. (2018). *Acta Cryst.* **D74**, 986–999.
- Sanchez-Weatherby, J., Sandy, J., Mikolajek, H., Lobley, C. M. C., Mazzorana, M., Kelly, J., Preece, G., Littlewood, R. & Sørensen, T. L.-M. (2019). *J. Synchrotron Rad.* **26**, 291–301.
- Schiebel, J., Krimmer, S. G., Röwer, K., Knörlein, A., Wang, X., Park, A. Y., Stieler, M., Ehrmann, F. R., Fu, K., Radeva, N., Krug, M., Huschmann, F. U., Glöckner, S., Weiss, M. S., Mueller, U., Klebe, G. & Heine, A. (2016). *Structure*, **24**, 1398–1409.
- Schurmann, J., Lindhe, I., Janneck, J. W., Lima, G. & Matej, Z. (2019). *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pp. 978–983. Piscataway: IEEE.
- Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vornrhein, C. & Bricogne, G. (2012). *Acta Cryst.* **D68**, 368–380.
- Smart, O. S., Womack, T. O., Sharff, A., Flensburg, C., Keller, P., Paciorek, W., Vornrhein, C. & Bricogne, G. (2011). *GRADE*. Global Phasing Ltd, Cambridge, UK.
- Smart, O. S., Womack, T. O., Sharff, A., Flensburg, C., Keller, P., Paciorek, W., Vornrhein, C. & Bricogne, G. (2014). *Rhofit*. Global Phasing Ltd, Cambridge, UK.

- Sparta, K. M., Krug, M., Heinemann, U., Mueller, U. & Weiss, M. S. (2016). *J. Appl. Cryst.* **49**, 1085–1092.
- Stetten, D. von, Carpentier, P., Flot, D., Beteva, A., Caserotto, H., Dobias, F., Guijarro, M., Giraud, T., Lentini, M., McSweeney, S., Royant, A., Petitdemange, S., Sinoir, J., Surr, J., Svensson, O., Theveneau, P., Leonard, G. A. & Mueller-Dieckmann, C. (2020). *J. Synchrotron Rad.* **27**, 844–851.
- Svensson, O., Andreu, J., Karpics, I. & Nan, J. (2020). *Lucid3: A Computer Vision Python Library that Detects Crystallography Sample Holders*. <https://github.com/mxcube/lucid3>.
- Svensson, O., Gilski, M., Nurizzo, D. & Bowler, M. W. (2018). *Acta Cryst.* **D74**, 433–440.
- Terwilliger, T. C., Adams, P. D., Moriarty, N. W. & Cohn, J. D. (2007). *Acta Cryst.* **D63**, 101–107.
- Terwilliger, T. C., Klei, H., Adams, P. D., Moriarty, N. W. & Cohn, J. D. (2006). *Acta Cryst.* **D62**, 915–922.
- Ursby, T., Åhnberg, K., Appio, R., Aurelius, O., Barczyk, A., Bartalesi, A., Bjelčić, M., Bolmsten, F., Cerenius, Y., Doak, R. B., Eguiraun, M., Eriksson, T., Friel, R. J., Gorgisyan, I., Gross, A., Haghighat, V., Hennies, F., Jagudin, E., Norsk Jensen, B., Jeppsson, T., Kloos, M., Lidon-Simon, J., de Lima, G. M. A., Lizatovic, R., Lundin, M., Milan-Otero, A., Milas, M., Nan, J., Nardella, A., Rosborg, A., Shilova, A., Shoeman, R. L., Siewert, F., Sondhauss, P., Talibov, V. O., Tarawneh, H., Thånell, J., Thunnissen, M., Unge, J., Ward, C., Gonzalez, A. & Mueller, U. (2020). *J. Synchrotron Rad.* **27**, 1415–1429.
- Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T. & Bricogne, G. (2011). *Acta Cryst.* **D67**, 293–302.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* **D67**, 235–242.
- Winter, G. & McAuley, K. E. (2011). *Methods*, **55**, 81–93.
- Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Cryst.* **D74**, 85–97.
- Wojdyr, M. (2016). *UglyMol – View Protein Models and Electron Density*. <https://uglymol.github.io/>.
- Wollenhaupt, J., Barthel, T., Lima, G. M. A., Metz, A., Wallacher, D., Jagudin, E., Huschmann, F. U., Hauss, T., Feiler, C. G., Gerlach, M., Hellmig, M., Förster, R., Steffien, M., Heine, A., Klebe, G., Mueller, U. & Weiss, M. S. (2021). *J. Vis. Exp.* **3**, doi:10.3791/62208.
- Wollenhaupt, J., Metz, A., Barthel, T., Lima, G. M. A., Heine, A., Mueller, U., Klebe, G. & Weiss, M. S. (2020). *Structure*, **28**, 694–706.
- Wright, N. D., Collins, P., Koekemoer, L., Krojer, T., Talon, R., Nelson, E., Ye, M., Nowak, R., Newman, J., Ng, J. T., Mitrovich, N., Wiggers, H. & von Delft, F. (2021). *Acta Cryst.* **D77**, 62–74.
- Yin, X., Scalia, A., Leroy, L., Cuttitta, C. M., Polizzo, G. M., Ericson, D. L., Roessler, C. G., Campos, O., Ma, M. Y., Agarwal, R., Jackimowicz, R., Allaire, M., Orville, A. M., Sweet, R. M. & Soares, A. S. (2014). *Acta Cryst.* **D70**, 1177–1189.



STRUCTURAL
BIOLOGY

Volume 77 (2021)

Supporting information for article:

***FragMAXapp*: crystallographic fragment screening data analysis
and project management system**

**Gustavo M. A. Lima, Elmir Jagudin, Vladimir O. Talibov, Laila S. Benz, Marullo
Costantino, Tatjana Barthel, Jan Wollenhaupt, Manfred S. Weiss and Uwe
Mueller**

Project Settings

Set experiment definitions

Protein

Library Name

Library Fragments No file chosen

Proposal Number

Shifts

Data Processing Encrypted Mode

Figure S1 Project Setting page. The minimum required information is validated before project creation.

Data Collection Overview

UPDATE PROJECT DATABASE

Search for Sample names...

Info/Process	Protein	Sample name	Run	Proc Status	Ref Status	Lig Status	Resolution [Å]	Frames	Crystal Picture	Ligand
INFO	PrtK	A10a	1	<ul style="list-style-type: none">PipedreamXIAZ/DIALSXIAZ/XDSXDSAPPautoPROCEDNA_procfastdp	<ul style="list-style-type: none">PipedreamDIMPLEBUSTERfastpipeline	<ul style="list-style-type: none">RhoFitLigandFit	1.24	3600		
INFO	PrtK	A11a	1	<ul style="list-style-type: none">PipedreamXIAZ/DIALSXIAZ/XDSXDSAPPautoPROCEDNA_procfastdp	<ul style="list-style-type: none">PipedreamDIMPLEBUSTERfastpipeline	<ul style="list-style-type: none">RhoFitLigandFit	1.24	3600		

Figure S2 Fragment screening project Home Page, displaying information about the data collection parameters, the crystal, fragment and data processing progress.

The screenshot shows the 'FragMAX - Library view' web interface. On the left is a dark sidebar with navigation options: PrtK - JBS, Project, PDBe, Library view, Project info, Home, Data Analysis, PanDDa, Results, Download, PDB export tool, Check HPC, and Logout. The main content area has a 'New Library CSV' section with a 'Choose file' button and an 'UPDATE' button. Below this is a message: 'Fragment missing from your project based on dataset names'. The 'JBS Library' section displays a grid of eight chemical structures labeled A10a, A11a, A12a, A1a, A2a, A3a, A4a, and A5a. Each structure is accompanied by its SMILES string: A10a: NCC(=O)c1ccc(Br)cc1; A11a: Fc1ccc(cc1F)C1=NNC2=NCCN2C1; A12a: O[C@@H](C(=O)O)c1ccccc1)c1; A1a: Cc1ccccc1C(=O)NN. The interface also includes a 'MAXIV LABORATORY' logo at the bottom left.

Figure S3 Full reproduction of sample management page interface.

The screenshot shows the 'FragMAX - Data Analysis' web interface. The sidebar on the left includes: hCAII - FMLv03, Project, Home, Data Analysis, PanDDa, Results, Download, PDB export tool, Check HPC, and Logout. The main content area is titled 'hCAII Data Analysis' and features a notification: 'New actions will process/reprocess your datasets using at MAX IV HPC and replace your current result'. There are three checkboxes: 'Process all datasets' (checked), 'Process new datasets', and 'Select ALL datasets'. A 'Select dataset' dropdown is present. The interface is divided into three main sections: 'Data processing', 'Structure Refinement', and 'Ligand fitting'. Each section has a 'Software' list with checkboxes, a 'Space group' field (set to P43212), and a 'Cell parameters' field (set to [a, b, c, α , β , γ]). The 'Data processing' section has checkboxes for Pipedream, XIA2/DIALS, XIA2/XDS, XDSAPP, and autoPROC. The 'Structure Refinement' section has checkboxes for Pipedream, DIMPLE, BUSTER, and fspipeline. The 'Ligand fitting' section has checkboxes for Pipedream, RhoFit, LigandFit, and PanDDA, and radio buttons for GRADE, AceDRG, and eLBOW. At the bottom of each section is a 'RUN' button. A footer bar contains tabs for each software package: XIA2/DIALS, XIA2/XDS_XSCALE, XDSAPP, autoPROC, DIMPLE, BUSTER, fspipeline, RhoFit, LigandFit, PanDDA, and Pipedream. The 'MAXIV LABORATORY' logo is at the bottom left.

Figure S4 Full reproduction Data Analysis interface. In this view, the submission buttons are enabled due to the selection of pipelines.

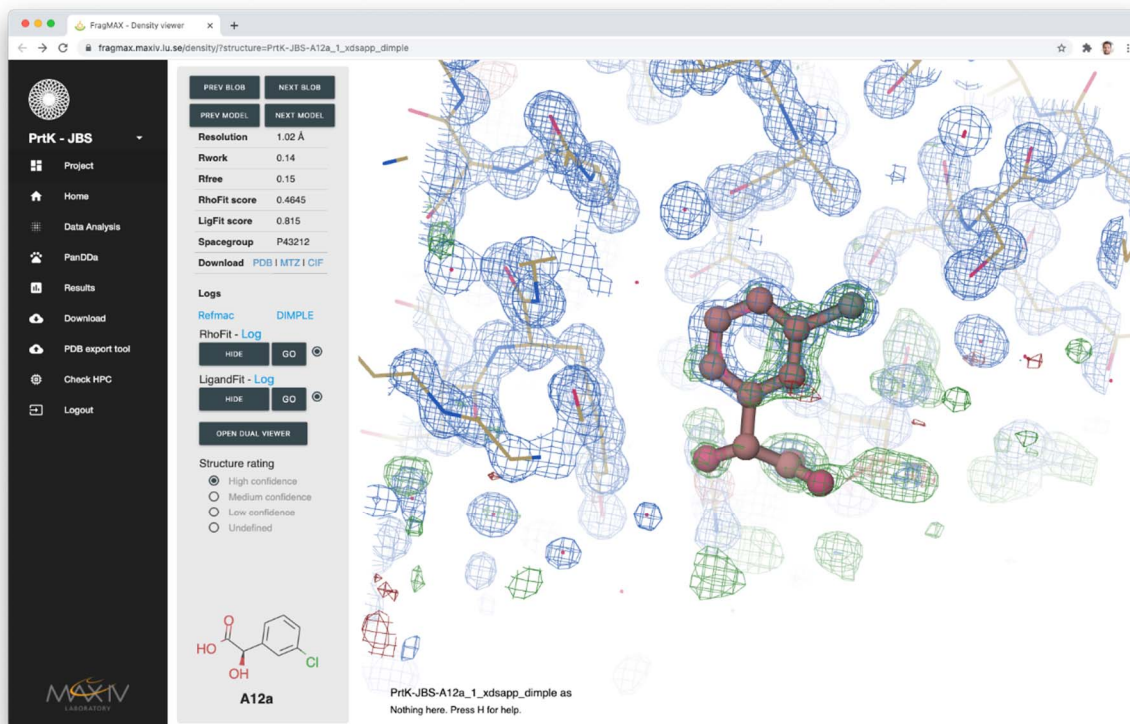


Figure S5 Full reproduction of Results view.

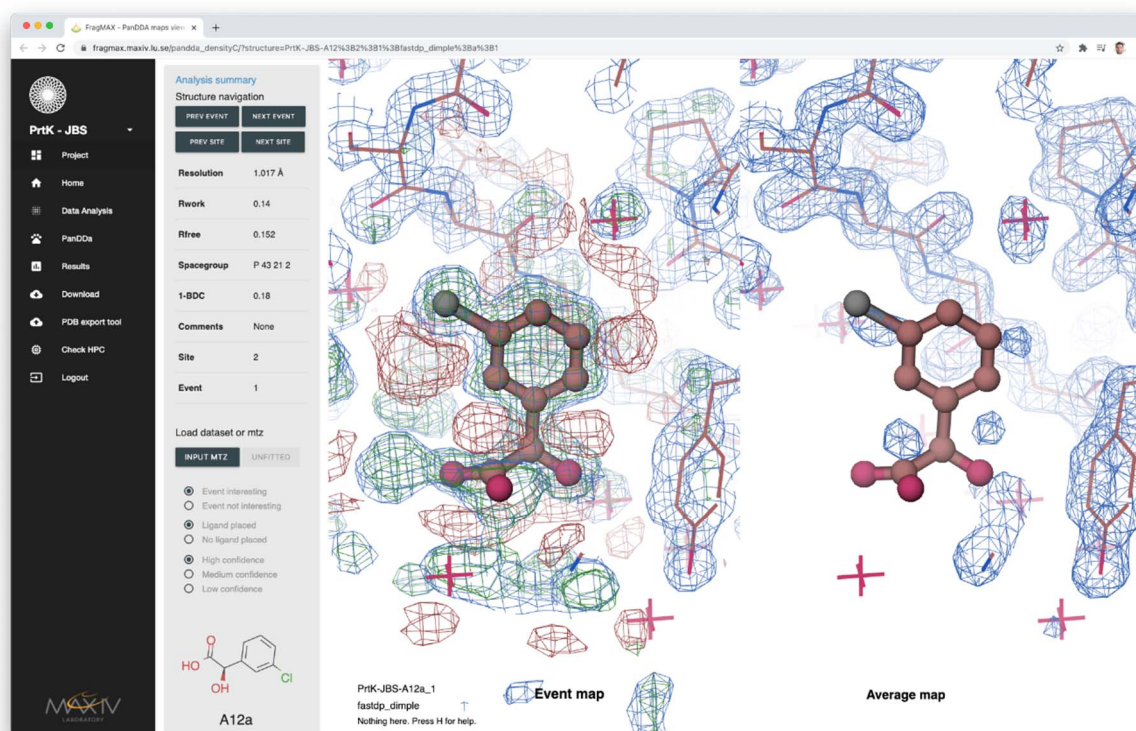


Figure S6 PanDDA events viewer. The side-by-side comparison between the event and average maps from PanDDA help identify false-positive events during the analysis. From the webapp, it is

possible to see information about the event and the user-annotated confidence level. The navigation function moves the view to different site centromeres.

S1. Proteinase K fragment screening

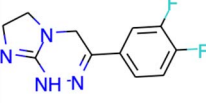
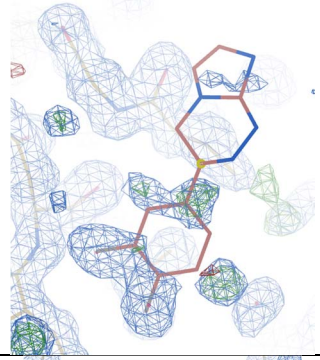
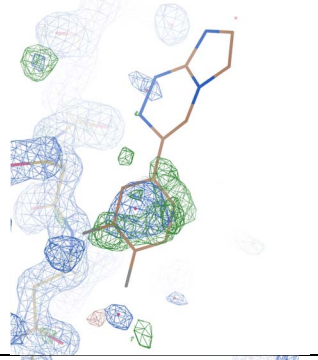
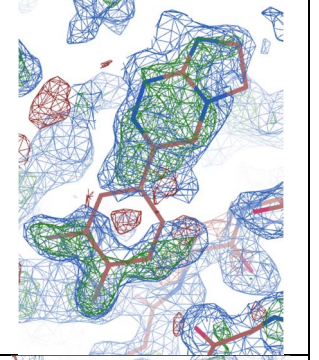
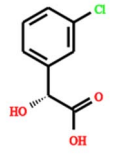
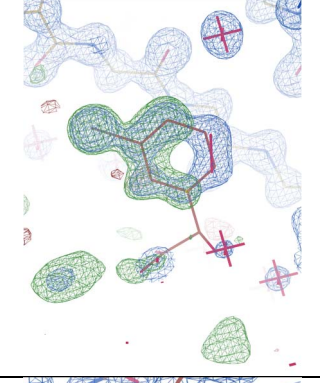
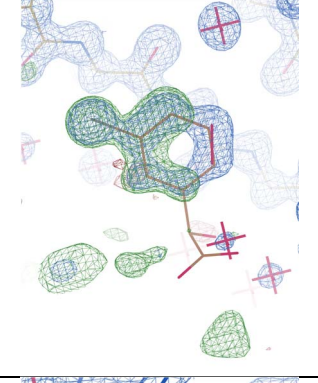
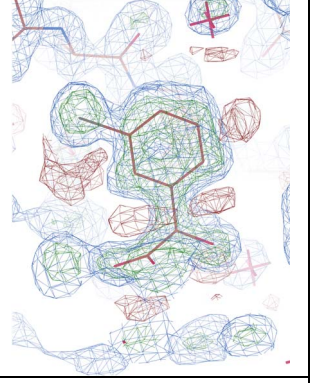
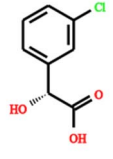
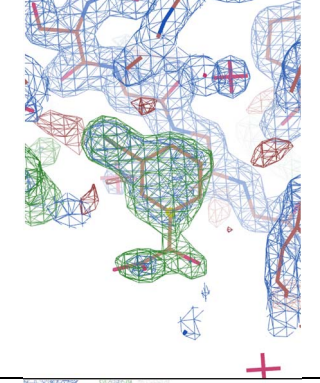
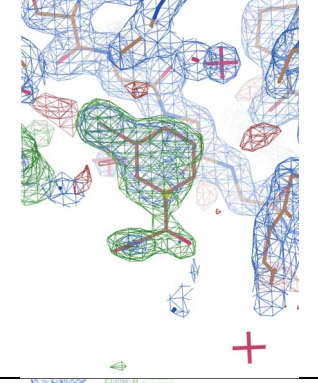
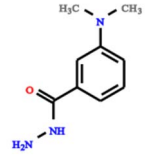
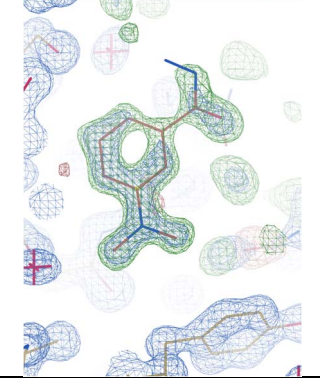
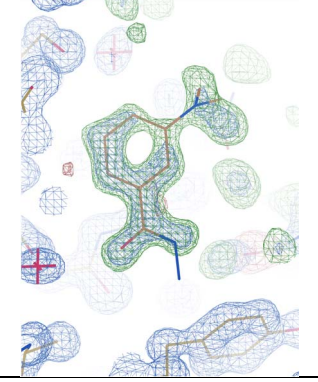
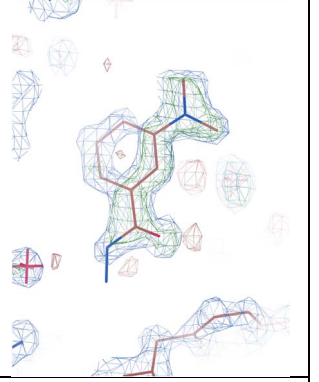
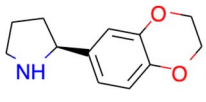
S1.1. Material and Methods

The protein was purchased (Jena Bioscience, Germany) and crystallised using sitting drop method as described in previous work (Larson *et al.*, 2009) on MRC3 plates (SWISSCI, Switzerland) using a Mosquito liquid handling system (TTP Labtech, UK). The fragment library Frag Xtal Screen was purchased (Jena BioScience, Germany) and handled according to the manual instructions. For this experiment, we renamed the fragments from its original convention (1 to 96) to an alphanumeric convention equivalent to its position in the 96-well storage plate (A1 to H12). The crystallisation drops were 300 nL containing a mixture of 150 nL of precipitating solution (1.2 M Ammonium Sulfate, 0.1 M Tris-HCl, pH 8.0) and 150 nL of 10 mg.mL⁻¹ Proteinase K solution. Suitable single crystals were obtained after 48 h, with sizes ranging from 50 to 100 µm. The crystals were soaked for 2 hours using 300 nL at 50 nM of each fragment dissolved in a precipitant solution containing 20% DMSO, for a final DMSO concentration of 10% after addition to the crystallisation drop. To transfer the fragment solutions on top of the crystallisation drop, the Crystal Shifter (Oxford Labtec, UK) was used with the crystal plate in one of its plate holders and the solubilised fragment plate in the second one. After soaking, crystals were harvested with Crystal Shifter assistance and cryo-cooled in LN2. Diffraction data of the identified hits from the cryo-experiment were also collected at room temperature (RT). The samples for RT experiment were prepared using the same protocol described in this section, mounting the crystal loops inside MicroRT capillaries (MiTeGen, USA) instead of cryo-cooling in LN2. Data collection was performed at BioMAX beamline, with experimental parameters specified in Table 1.

S1.2. Results

Using FragPLEX selection and PanDDA to analyse the data, 18 fragments out of 96 were found, with an 18,75% hit rate (Table S1).

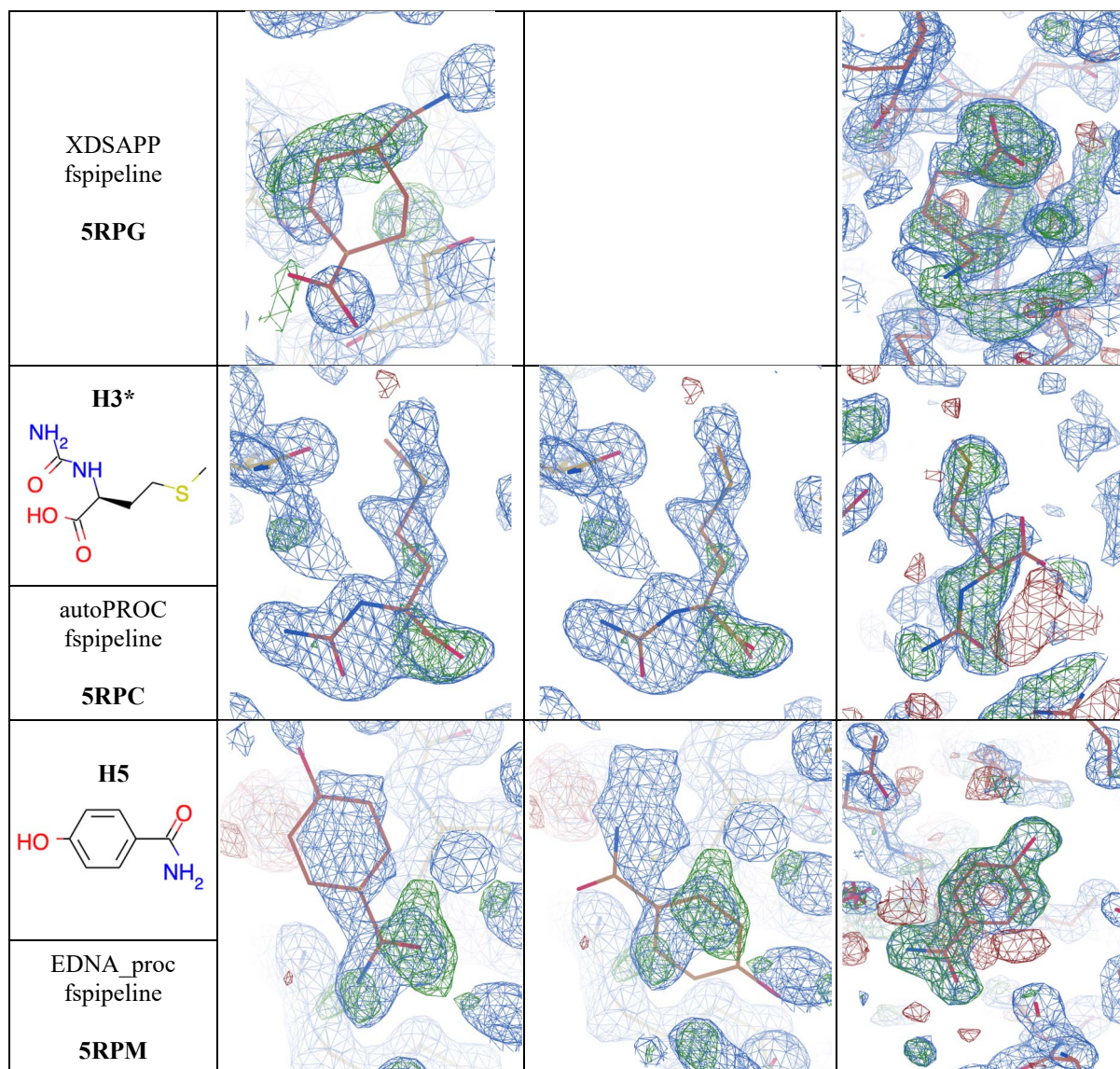
Table S1 Complete automated ligand searching comparison. Every hit found using PanDDA is compared to other automated methods. The dataset information column displays the fragment used during soaking, the methods used to process and refine the dataset, and the PDB ID of the final structure.

Dataset Information	LigandFit	RhoFit	PanDDA
<p>A11*</p> 			
<p>fastdp fspipeline</p> <p>5RPH</p>			
<p>A12</p> 			
<p>autoPROC fspipeline</p> <p>5ROQ</p>			
<p>A12 RT</p> 			<p>No Solution</p>
<p>xia2/DIALS fspipeline</p> <p>5ROP</p>			
<p>B5</p> 			
<p>EDNA_proc fspipeline</p> <p>5ROL</p>			
<p>B7</p> 			

<p>fastdp fspipeline</p> <p>5RP9</p>			
<p>B9</p> <chem>N=C(N)N1CCCCC1</chem>			
<p>xia2/XDS DIMPLE</p> <p>5RPK</p>			
<p>B12</p> <chem>CN1C=NC2=C1S=C(C2)CC(=O)NC3=CC=CN3</chem>			
<p>fastdp fspipeline</p> <p>5RPJ</p> <p>C2</p>			
<p>fastdp fspipeline</p> <p>5RP6</p>			
<p>C8</p> <chem>COC(=O)c1ccc(CN)cc1</chem>	No solution	No solution	

<p>autoPROC fspipeline</p> <p>5RPL</p>			
<p>C11*</p>			
<p>autoPROC fspipeline</p> <p>5ROF</p>			
<p>D12</p>			
<p>xia2/DIALS fspipeline</p> <p>5RP7</p>			
<p>E3*</p>			
<p>autoPROC fspipeline</p> <p>5RPA</p>			
<p>E4</p>	No Solution	No Solution	

<p>xia2/DIALS fspipeline</p> <p>5RON</p>			
<p>F1</p>	No solution	No solution	
<p>XDSAPP fspipeline</p> <p>5ROR</p>			
<p>F6</p>			
<p>xia2/DIALS fspipeline</p> <p>5ROW</p>			
<p>F12*</p>			
<p>xia2/DIALS BUSTER</p> <p>5RPD</p>			
<p>H2</p>		No solution	



* The binding site identified by RhoFit, LigandFit and PanDDA is not the same.

- III. Wollenhaupt, J.; Metz, A.; **Barthel, T.**; Lima, G. M. A.; Heine, A.; Mueller, U.; Klebe, G.; Weiss, M. S., F2X-Universal and F2X-Entry: Structurally Diverse Compound Libraries for Crystallographic Fragment Screening. *Structure* **2020**, *28* (6), 694-706.e5. <https://doi.org/10.1016/j.str.2020.04.019>.

This publication must be purchased online.



Frag4Lead: growing crystallographic fragment hits by catalog using fragment-guided template docking

Alexander Metz,^{a,‡} Jan Wollenhaupt,^{b,‡} Steffen Glöckner,^a Niki Messini,^a Simon Huber,^a Tatjana Barthel,^b Ahmed Merabet,^a Hans-Dieter Gerber,^a Andreas Heine,^a Gerhard Klebe^{a*} and Manfred S. Weiss^{b*}

Received 3 June 2021

Accepted 9 August 2021

Edited by M. Rudolph, F. Hoffmann-La Roche Ltd, Switzerland

‡ These authors contributed equally to this work.

Keywords: crystallographic fragment screening; fragment-based lead discovery; template docking; structure-based drug design; pose validation.

PDB references: endothiapepsin, 5sak; 5sal; 5sam; 5san; 5sao; 5sap; 5saq; 5sar; 5sas; 5sat

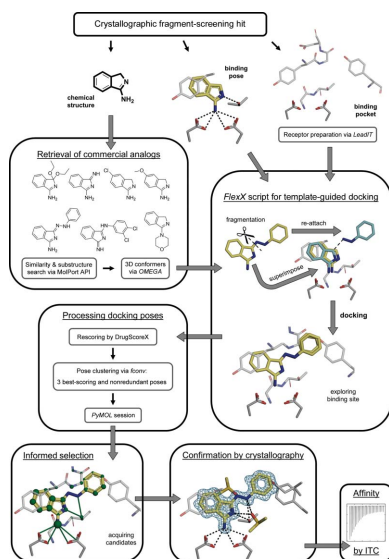
Supporting information: this article has supporting information at journals.iucr.org/d

^aDepartment of Pharmaceutical Chemistry, Philipps-University Marburg, Marbacher Weg 6, D-35032 Marburg, Germany, and ^bMacromolecular Crystallography, Helmholtz-Zentrum Berlin, Albert-Einstein-Straße 15, D-12489 Berlin, Germany. *Correspondence e-mail: klebe@staff.uni-marburg.de, msweiss@helmholtz-berlin.de

In recent years, crystallographic fragment screening has matured into an almost routine experiment at several modern synchrotron sites. The hits of the screening experiment, *i.e.* small molecules or fragments binding to the target protein, are revealed along with their 3D structural information. Therefore, they can serve as useful starting points for further structure-based hit-to-lead development. However, the progression of fragment hits to tool compounds or even leads is often hampered by a lack of chemical feasibility. As an attractive alternative, compound analogs that embed the fragment hit structurally may be obtained from commercial catalogs. Here, a workflow is reported based on filtering and assessing such potential follow-up compounds by template docking. This means that the crystallographic binding pose was integrated into the docking calculations as a central starting parameter. Subsequently, the candidates are scored on their interactions within the binding pocket. In an initial proof-of-concept study using five starting fragments known to bind to the aspartic protease endothiapepsin, 28 follow-up compounds were selected using the designed workflow and their binding was assessed by crystallography. Ten of these compounds bound to the active site and five of them showed significantly increased affinity in isothermal titration calorimetry of up to single-digit micromolar affinity. Taken together, this strategy is capable of efficiently evolving the initial fragment hits without major synthesis efforts and with full control by X-ray crystallography.

1. Introduction

In a drug-discovery project, the hits obtained by fragment screening are typically smaller than the lead-like molecules obtained from a high-throughput screening (HTS) campaign. Nonetheless, fragments constitute excellent starting points for lead discovery as they usually explore the hotspots of binding, where a large part of the binding affinity can be obtained. It is clear, however, that owing to their small size and their weak binding affinity, fragments need to be improved with respect to affinity and specificity. Also, due to their rather small number of interactions with the protein surface, fragments are often promiscuous binders. Subsequent optimization can usually be achieved more efficiently compared with HTS hits, as fragments leave sufficient space and options for exit vectors to expand and improve binding upon optimization. In the past, numerous fragment-screening methods have been established to detect such starting points for follow-up lead discovery (Erlanson *et al.*, 2016). The increasing popularity of these approaches is reflected by the growing number of reported fragment-to-lead campaigns (Mortenson *et al.*, 2019) and,



consequently, a large number of candidates have entered clinical trials (Erlanson *et al.*, 2016). Meanwhile, four approved drugs developed by fragment-based lead discovery (FBLD) have been launched to market. To efficiently accomplish such hit-to-lead-to-drug developments, the support of X-ray crystal structure analysis is essential, as validated binding modes allow the immediate application of structure-based design concepts to the subsequent optimization process (Murray & Rees, 2016; Schmidt & Rademann, 2009). Therefore, crystallographic fragment screening (CFS), if applicable, has major advantages over alternatives such as HTS based on biochemical or biophysical assays, which are mostly in need of target-binding validation and binding-mode characterization before moving forward into efficient structure-based optimization (Schiebel, Krimmer *et al.*, 2016; Schiebel, Radeva *et al.*, 2016). Recent improvements in instrumentation at several synchrotron beamline facilities, as well as in automated data-handling procedures, have greatly improved the capabilities of CFS. Consequently, CFS can be performed with relatively little effort, also enabling access for academic groups experienced in crystallographic methods (Schiebel, Krimmer *et al.*, 2016; Lamoree & Hubbard, 2017; Wollenhaupt *et al.*, 2021; Lima *et al.*, 2020; Krojer *et al.*, 2017). Based on screening collections of some 100–1000 compounds, hit rates of 0.5–10% have been achieved with CFS (Hartshorn *et al.*, 2005). More recently, however, improved libraries have elevated the hit rates to 15–30% (Schiebel, Krimmer *et al.*, 2016; Wollenhaupt *et al.*, 2020), or even above 40% for very low molecular mass fragments (O'Reilly *et al.*, 2019). As a matter of fact, these developments have shifted the initial bottleneck of finding starting points from hit detection per se towards the subsequent progression of the fragment hits into ligands with improved affinity and selectivity.

Fragment-hit optimization towards higher affinity compounds usually involves elaborate chemical synthesis with follow-up medicinal chemistry at a relatively early stage (Murray & Rees, 2016). To facilitate this process, fragment libraries can be designed and assembled in such a way that discovered hits can be easily expanded to provide entry points into larger chemical spaces (Cox *et al.*, 2016; Keserű *et al.*, 2016). In the case where a strong and experienced medicinal chemistry synthesis group is not within reach, the further progress of drug-development projects, particularly in academic settings, is easily hindered or sometimes even completely stalled, mostly in the initial phase of a lead-finding process (Murray & Rees, 2016; Chevillard & Kolb, 2015).

As the first step of a fragment-to-lead campaign, the initial fragment hits require some validation in order to ensure that reasonably close analogs of the identified hits bind in a similar fashion. If no such analogs can be identified, the fragment may be hard to optimize or may present a case with binding modes that easily swap upon minor chemical modification. The determination of the binding poses of structurally closely related fragments provides confidence in the reliability and relevance of an observed fragment hit and its pose, and often allows the development of an initial crude structure–activity relationship (SAR). This can be achieved by simply exploring

readily available analogs in an ‘SAR-by-catalog’ approach (Erlanson *et al.*, 2019; Schulz *et al.*, 2011). In fortunate cases, suitable analogs can be further evaluated by structure-based computational methods, in particular by molecular docking (Yuriev & Ramsland, 2013). However, the identification of a promising scaffold with the correct binding pose among a large variety of possibilities via docking still remains a challenging problem, especially for molecules as small as fragments that form only a few interactions and can easily alter their binding poses upon modulation of their substitution patterns (Lamoree & Hubbard, 2017; Oebbeke *et al.*, 2021).

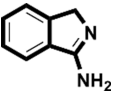
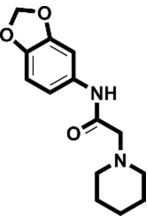
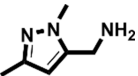
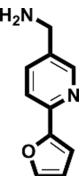
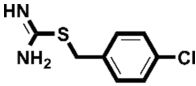
To efficiently exploit hits from CFS, methods are needed to either suggest easily accessible structural and chemical analogs of a given hit to validate its binding pose or to retrieve commercially available larger compounds embedding the initial hit. Such analogs can be retrieved by web interfaces that are often provided by the vendors or vendor aggregators themselves. Among the most used aggregators are MolPort, ChEMspace, eMolecules, Molecule, Enamine and LabNetwork. These catalogs are now also conveniently interfaced by overarching tools such as *Manifold* (<https://postera.ai/manifold/>), which is free for academic use. Other approaches to visualize the search for effective SARs have also been reported recently (Hall *et al.*, 2017). Nonetheless, efficient CFS hit exploitation requires strategies to prioritize the list of suitable follow-up candidates from the possibly vast number of commercially available analogs. This prioritization of potential follow-up compounds is best supported by computational tools and ideally exploits the crystallographic knowledge of the bound fragment as a template to guide the next design steps by the virtual screening of candidates (de Souza Neto *et al.*, 2020). In our approach, the additional chemical groups of a putative follow-up candidate are tethered to the original fragment hit in its bound state. In this regard, the information obtained by CFS is combined with a computational growth strategy.

In order to demonstrate the concept of our developments, the aspartyl protease endoprotease (EP), an enzyme frequently used to develop principles and novel strategies in inhibitor design, was used as the target protein. Five hits from a previous CFS campaign (Radeva, Krimmer *et al.*, 2016; Radeva, Schiebel *et al.*, 2016; Schiebel, Krimmer *et al.*, 2016) were used to emulate a real-case scenario with only a few and potentially non-optimal fragment hits available. This means that these hits do not reflect a prioritized selection of the 41 binders that address the catalytic dyad of EP, but instead contain direct and indirect dyad binders and span a wide range of affinities (100 μM to 8.8 mM). Additionally, only a limited number of commercially available follow-up candidates were tested. From the selected 28 follow-up compounds, ten binders could be identified by X-ray crystallography. Several of these follow-up hits have affinities increased by more than one order of magnitude compared with the original fragment hit. The best case exhibited a 266-fold improvement in affinity. In conclusion, the presented approach can successfully identify commercially available follow-up candidates in one step, thereby circumventing laborious chemical synthesis in the early stage of fragment hit advancement.

Table 1

EP–fragment complexes chosen for optimization.

The fragment nomenclature was adopted from Köster *et al.* (2011). K_d is the dissociation constant of the compound from EP and LE is the respective ligand efficiency, which is the binding energy per non-H atom. The K_d and LE values are taken from Schiebel, Radeva *et al.* (2016). The number of successfully docked analogs refers to docked analogs for which *FlexX* generated a meaningful pose.

Fragment	PDB code	Chemical structure	K_d (μM)	LE (kcal mol ⁻¹ per atom)	No. of identified follow-up candidates	No. of successfully docked follow-up candidates
F005	4y3e		1700	0.38	556	67
F041	4y3z		900	0.22	1013	88
F058	4y56		8800	0.31	10022	>1000†
F066	5dq4		400	0.35	615	395
F290	4y35		100	0.45	267	32

† Only the 1000 highest-ranking fragment analogs were considered.

2. Materials and methods

2.1. Retrieval of commercially available fragment analogs

Using the MolPort Chemical Search node (SIA MolPort, Latvia) within the Konstanz Information Miner (KNIME) version 3.4.0 (Berthold *et al.*, 2008), commercially available fragment analogs were retrieved. Three types of search were carried out: searching for analogs (i) that contain the initially discovered fragment as a substructure, (ii) that are a substructure of this fragment or (iii) that are reasonably similar to the corresponding fragment based on a MACCS fingerprint Tanimoto coefficient of ≥ 0.7 (Willett *et al.*, 1986). An increased upper limit of 10 000 retrievable structures per search type was granted by MolPort. Duplicate analogs were removed based on their MolPort IDs. Likewise, analogs containing atoms other than C, H, N, O, P, S, F, Cl, Br, I or Se were removed using the Chemistry Development Kit (CDK) Element Filter node (Beisken *et al.*, 2013). Very small molecules (molecular weight of <50 Da or containing less than four non-H atoms) were excluded from the similarity-search results based on calculations with the Standard Properties node of the LigandScout extensions for KNIME (Inte:Ligand GmbH). Also, a secondary similarity filter was applied requiring a Tanimoto coefficient of ≥ 0.4 to the corresponding fragments using Indigo 2 structural fingerprints within KNIME (EPAM

Systems Inc., Newtown, Pennsylvania, USA). Molecular formats were converted using the MolConverter node of ChemAxon LCC. The 3D conformers of the follow-up candidates were then generated by *OMEGA* (Hawkins *et al.*, 2010) version 2.5.1.4 from OpenEye Scientific Software.

2.2. Selection of EP–fragment complexes for optimization

In order to test the intended optimization, five EP–fragment complexes (Table 1) were selected from the CFS campaign carried out by Radeva, Schiebel *et al.* (2016). The nomenclature of the starting fragments **F005**, **F041**, **F058**, **F066** and **F290** is defined as in Radeva, Schiebel *et al.* (2016). The follow-up compounds are named **FU_x-y**, where the subscript *x* denotes the respective starting fragment and *y* denotes the number of the follow-up compound of this series.

2.3. Preparation of receptors for docking of follow-up molecules

Each fragment-bound EP structure was treated separately for docking and a separate list of analogs was retrieved. Here, only superstructures, *i.e.* structures containing the exact scaffold of the fragment as a substructure, of the used starting fragments were docked because at the time that the workflow was applied, to the best of our knowledge, no procedure was

available to superimpose fragments with different scaffolds. Prior to the template-docking procedure developed in this work, the fragment-bound EP crystal structures were prepared manually with the *LeadIT* software (version 2.1.8), considering only amino-acid residues within 10 Å of the bound fragment and using default settings. Water molecules, fragments and other solutes were removed.

2.4. Scoring of docked poses

For rescoring the binding poses after the customized *FlexX* template docking, DrugScoreX (DSX) was used (Neudert & Klebe, 2011a). The program *DSX* can be downloaded freely from <https://agklebe.pharmazie.uni-marburg.de/>. DSX was chosen as it is somewhat tolerant of the close atomic contacts that may arise due to the geometric constraints of template docking to a rigid crystal structure. More specifically, the DrugScore (Gohlke *et al.*, 2000) per-contact score (PCS) is used from the DrugScore^{PDB} scoring function implemented in DSX (Neudert & Klebe, 2011a). This PCS is the genuine DrugScore score divided by the number of atom–atom interactions within 6 Å of the ligand that contribute to the overall score. Thus, the PCS is a measure of interaction efficiency and sorting poses by PCS aims to enrich small but efficiently binding analogs that largely retain or improve the ligand efficiency (LE) of the corresponding fragment hit. In FBLD, a high LE is an indicator of well anchored fragments that bind efficiently with respect to their size and thus are good starting points for further optimization.

2.5. Protein purification and crystallization

EP was isolated from Suparen (kindly provided by DSM Food Specialties, Heerlen, the Netherlands) in 0.1 M sodium acetate buffer pH 4.6 as described previously (Köster *et al.*, 2011). The sample was then subjected to size-exclusion chromatography using a Superdex S200 26/60 column (GE) and the same batch of buffer as for isolation. Protein-containing fractions were pooled, concentrated and flash-cooled in liquid nitrogen. The protein was then crystallized in a vapor-diffusion experiment in 48-well format using 250 µl reservoir solution consisting of 0.1 M sodium acetate pH 4.6, 0.1 M ammonium acetate pH 7.0, 24–33% (w/v) PEG 4000. 1.5 µl protein solution at a concentration of 5 mg ml⁻¹ was mixed with an equal amount of reservoir solution. Trays were incubated at 20°C. Crystals appeared after 5–6 days and were then crushed using a seed-bead kit (Douglas Instruments) to prepare crystal seeds, which were then used in a second crystallization experiment, here using 27% (w/v) PEG 4000 in the reservoir and adding 0.1 µl of seed dilutions of 1:15–1:45 (seed stock:reservoir) to the freshly mixed drop of protein and reservoir. The seeded crystals appeared after three days.

2.6. Compound-soaking experiments

The follow-up compounds (a full list, including providers and purities, if known, is given in Supplementary Table S1) were directly dissolved in a soaking solution consisting of 68.2 mM sodium acetate pH 4.6, 68.2 mM ammonium acetate

pH 7.0, 16.9% (w/v) PEG 4000, 19.3% (v/v) glycerol, 9.09% (v/v) DMSO to a concentration of 100 mM. For poorly soluble follow-up compounds, crystals were soaked in the supernatant of the solution. After incubation for 16–22 h the crystals were flash-cooled in liquid nitrogen and stored for diffraction data collection.

2.7. Diffraction data collection and processing

All data collections were carried out on beamlines BL14.1 and BL14.2 of the BESSY II electron-storage ring operated by the Helmholtz-Zentrum Berlin (HZB; Mueller *et al.*, 2015). Typically, 360° of data were collected in 0.1° increments using an exposure time of 0.1 s. Data were automatically processed using *XDSAPP* (Sparta *et al.*, 2016). All relevant data-collection and processing statistics are listed in Table 2.

2.8. Structure refinement and hit identification

All structures were refined using the automated script *fspipeline*, which is based on Schiebel, Krimmer *et al.* (2016). The starting model was based on PDB entry 4y5l (Schiebel, Krimmer *et al.*, 2016), from which all ligands and water molecules were removed. Electron-density maps and coordinate files obtained from the automated refinement were inspected manually for each experiment to judge the presence or absence of the expected ligand in the difference electron density. Subsequently, the identified hits were subjected to several rounds of alternating model building in *Coot* (Emsley *et al.*, 2010) and refinement in *Phenix* (Liebschner *et al.*, 2019) before and after ligand placement. For all follow-up ligands, occupancy refinement was carried out. Refined models and the corresponding electron-density maps were submitted to the PDB under group deposition ID G_1002201. PDB codes for the single entries and all relevant structure-refinement and validation parameters are shown in Table 3.

2.9. Isothermal titration calorimetry (ITC)

ITC experiments were conducted similarly to the procedure described by Schiebel, Radeva *et al.* (2016) on a MicroCal ITC200 (Malvern) instrument. All buffer solutions for ITC were prepared with the same batch of buffer as used to isolate the batch of EP. Details of the ITC experiments, including the protein and ligand concentrations for each experiment, are listed in Supplementary Table S2. In brief, the affinities of the weakly binding follow-up ligands were determined by displacement ITC titrations using the strongly enthalpic EP inhibitor SAP114 (Kuhnert *et al.*, 2015) as the displacement ligand. For this, 500 µM SAP114 in a buffer solution consisting of 0.1 M sodium acetate pH 4.6, 3% (v/v) DMSO was titrated into the same buffer additionally containing 50 µM EP and 2.0 mM of the respective follow-up ligand to a final stoichiometry of $N = 2$ (SAP114:EP). As a reference for calculating the affinities of the weakly binding follow-up ligands (Rühmann *et al.*, 2015), 500 µM SAP114 was titrated into the buffer solution without follow-up ligand using the same protocol. All displacement titrations were conducted as single measurements, except for that of **FU5-2** ($n = 3$). The affinities of the

Table 2
Data-collection and processing statistics.

Values in parentheses are for the outer shell.

Ligand ID	FU ₅ -1	FU ₅ -2	FU ₅ -3	FU ₅ -4	FU ₅₈ -1	FU ₅₈ -2	FU ₅₈ -3	FU ₂₉₀ -1	FU ₂₉₀ -2	FU ₆₆ -1
PDB code	5sak	5sal	5sam	5san	5sao	5sap	5saq	5sar	5sas	5sat
X-ray source	BESSY II	BESSY II	BESSY II	BESSY II	BESSY II	BESSY II	BESSY II	BESSY II	BESSY II	BESSY II
Beamline	BL14.1	BL14.1	BL14.1	BL14.1	BL14.1	BL14.1	BL14.1	BL14.1	BL14.1	BL14.1
Wavelength (Å)	0.9184	0.9184	0.9184	0.9184	0.9184	0.9184	0.9184	0.9184	0.9184	0.9184
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M
No. of crystals	1	1	1	1	1	1	1	1	1	1
Temperature (K)	100	100	100	100	100	100	100	100	100	100
Detector distance (mm)	149.208	149.196	210.510	142.610	165.251	165.237	174.598	142.604	165.249	210.509
Rotation range per image (°)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Total rotation range (°)	360	360	360	360	360	360	360	360	360	360
Space group	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁
Unit-cell parameters										
<i>a</i> (Å)	45.33	45.34	45.34	44.96	45.31	45.41	45.43	45.23	45.30	45.17
<i>b</i> (Å)	73.69	73.50	73.27	72.61	72.91	73.49	73.38	73.15	73.06	73.40
<i>c</i> (Å)	52.74	53.12	52.97	51.63	52.62	53.15	53.06	52.77	52.64	52.56
α (°)	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
β (°)	109.70	110.21	109.76	108.61	109.78	110.15	109.90	109.53	109.45	109.37
γ (°)	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
Resolution range (Å)	42.68–1.10 (1.17–1.10)	42.55–1.00 (1.06–1.00)	42.67–1.20 (1.27–1.20)	48.93–0.94 (1.00–0.94)	42.64–1.00 (1.06–1.00)	42.63–1.04 (1.10–1.04)	42.72–1.02 (1.08–1.02)	42.62–0.98 (1.04–0.98)	42.72–1.17 (1.24–1.17)	49.59–1.60 (1.70–1.60)
Total No. of reflections	468582	638330	338920	711675	611157	574223	586735	667615	401189	144529
Unique reflections	127768	173459	99863	198823	158879	152537	159315	175062	107431	42081
Multiplicity	3.67	3.68	3.39	3.58	3.85	3.76	3.68	3.81	3.73	3.43
Mean <i>I</i> / σ (<i>I</i>)	8.9 (0.7)	7.6 (0.8)	10.8 (0.9)	8.4 (0.7)	13.6 (1.1)	8.2 (0.9)	7.7 (0.8)	5.5 (0.8)	9.6 (0.8)	14.0 (1.9)
<i>R</i> _{meas} (%)	8.2 (172.0)	8.8 (143.3)	7.8 (136.8)	7.4 (157.7)	4.7 (117.9)	8.6 (137.3)	9.5 (130.3)	13.5 (121.7)	8.0 (164.5)	6.9 (73.3)
Completeness (%)	97.9 (93.2)	98.4 (97.8)	98.0 (97.0)	97.6 (91.8)	91.6 (78.0)	97.1 (95.0)	95.8 (85.7)	94.4 (90.9)	98.0 (95.1)	98.3 (98.1)
Wilson <i>B</i> factor (Å ²)	16.19	13.16	17.16	12.03	13.87	13.60	12.73	12.12	17.56	24.79
Mosaicity (°)	0.137	0.055	0.157	0.072	0.083	0.093	0.062	0.053	0.135	0.341
CC _{1/2}	99.9 (31.6)	99.7 (35.6)	99.9 (38.8)	99.7 (40.2)	99.9 (58.4)	99.8 (41.1)	99.7 (40.9)	98.7 (46.6)	99.9 (36.6)	99.9 (30.9)
ISa	21.6	14.6	39.2	16.6	30.3	17.8	15.8	7.8	23.1	37.7

stronger binding follow-up ligands were determined by direct ITC titrations. For this, 1 mM compound in a buffer solution consisting of 0.1 M sodium acetate pH 4.6, 3% (v/v) DMSO was titrated into the same buffer additionally containing 50 μM EP to a final stoichiometry of *N* = 4. Due to the poor solubility of FU₅-1, its affinity was determined in triplicate using the same protocol but titrating 500 μM FU₅-1 against 25 μM EP in the presence of 0.1% (v/v) Tween 20 in all buffers to a final stoichiometry of *N* = 4 (FU₅-1:EP). For FU₅-1, the available amount (1 mg) was used up by soaking experiments, so we resynthesized FU₅-1·HCl (98.5% purity) for use in ITC experiments. Further details of the synthesis of FU₅-1·HCl, including experimental data and NMR spectra, can be found in the supporting information (Supplementary Figs. S1–S3).

The obtained thermogram peaks of all titrations (Supplementary Fig. S4) were integrated with *Nitpic* 1.1.8 (Keller *et al.*, 2012). Subsequently, fitting of a single-site binding-model isotherm was performed using *SEDPHAT* 10.58d (Houtman *et al.*, 2007). For the errors of the fit, see Supplementary Table S3. For FU₅-1·HCl, we used the *AFFINImeter* suite (version 2.1710; Muñoz & Piñeiro, 2018) to perform a global fit over three independent measurements to derive common values for *K*_a and ΔH (fits are also depicted in Supplementary Fig. S4). During the fit, ΔH was corrected for the heat of dilution, which was individually fitted for each experiment. The stoichiometry was arbitrarily fixed at the anticipated stoichiometry of *N* = 1 as appropriate for the present low *c*-value titrations (Rühmann *et al.*, 2015). The obtained goodness of fit was consistent for all three experiments (77.2%, 72.6% and 74.2%) and with the global goodness of fit (74.7%). Furthermore, the local minima table showed that the obtained fit was independent of the initial seed value of the algorithm in 20 independent rounds of fitting. Results from the global fit were comparable to those from individually fitting each experiment, yet were more robust in terms of numerical stability when using different seed values.

metry of *N* = 1 as appropriate for the present low *c*-value titrations (Rühmann *et al.*, 2015). The obtained goodness of fit was consistent for all three experiments (77.2%, 72.6% and 74.2%) and with the global goodness of fit (74.7%). Furthermore, the local minima table showed that the obtained fit was independent of the initial seed value of the algorithm in 20 independent rounds of fitting. Results from the global fit were comparable to those from individually fitting each experiment, yet were more robust in terms of numerical stability when using different seed values.

2.10. Restrospective and unbiased docking of crystallographically determined follow-up poses

SeeSAR (version 11.0.0; BioSolveIT; license required) was used to prepare the receptors, perform the docking and score the resulting poses. For receptor preparation, the automatic pocket identification of *SeeSAR* was used on the complexes of **F005**, **F058**, **F066** and **F290** with EP (PDB codes are given in Table 1). The *FlexX* (Rarey *et al.*, 1996) functionality of *SeeSAR* was used for docking and a maximum number of poses of 500 was chosen. The docked poses were scored using the implemented HYDE scoring function (Reulecke *et al.*, 2008; Schneider *et al.*, 2013). The structural models of the follow-up compounds were aligned with the respective receptor structure. R.m.s.d.s of the scored poses versus the

Table 3
Structure-refinement and validation statistics.

Values in parentheses are for the outer shell. Refinement was performed with *phenix.refine* version 1.19.

Ligand ID	FU ₅₋₁	FU ₅₋₂	FU ₅₋₃	FU ₅₋₄	FU ₅₈₋₁	FU ₅₈₋₂	FU ₅₈₋₃	FU ₂₉₀₋₁	FU ₂₉₀₋₂	FU ₆₆₋₁
PDB code	5sak	5sal	5sam	5san	5sao	5sap	5saq	5sar	5sas	5sat
Resolution limits (Å)	42.68–1.10 (1.13–1.10)	42.55–1.00 (1.02–1.00)	42.67–1.20 (1.23–1.20)	48.93–0.94 (0.96–0.94)	42.64–1.00 (1.02–1.00)	42.63–1.04 (1.06–1.04)	41.26–1.02 (1.04–1.02)	41.13–0.98 (1.00–0.98)	42.72–1.17 (1.19–1.17)	49.59–1.40 (1.42–1.40)
Completeness (%)	98.0	98.5	98.1	97.6	91.6	97.2	95.9	94.4	98.1	98.1
Data cutoff	$F > 1.33\sigma(F)$	$F > 1.33\sigma(F)$	$F > 1.35\sigma(F)$	$F > 1.35\sigma(F)$	$F > 1.36\sigma(F)$	$F > 1.35\sigma(F)$	$F > 1.35\sigma(F)$	$F > 1.36\sigma(F)$	$F > 1.35\sigma(F)$	$F > 1.35\sigma(F)$
No. of reflections										
Working set	125640 (7260)	171349 (11272)	97759 (6355)	196698 (10913)	156756 (7204)	150436 (9664)	157202 (8192)	172802 (10946)	105317 (6461)	59494 (2686)
Test set	2101 (122)	2101 (138)	2101 (137)	2100 (116)	2101 (97)	2101 (135)	2100 (109)	2098 (133)	2100 (129)	3132 (141)
R_{work}	0.143 (0.3678)	0.144 (0.3424)	0.138 (0.2883)	0.139 (0.4070)	0.131 (0.3680)	0.134 (0.3391)	0.139 (0.3717)	0.164 (0.3784)	0.140 (0.3328)	0.161 (0.3104)
R_{free}	0.159 (0.3818)	0.160 (0.3579)	0.164 (0.2800)	0.147 (0.4078)	0.146 (0.3764)	0.149 (0.3445)	0.157 (0.3809)	0.175 (0.3676)	0.156 (0.3391)	0.205 (0.3315)
No. of non-H atoms										
Protein	2389	2389	2389	2389	2389	2389	2389	2389	2389	2389
Ligand	240	76	46	39	39	38	16	31	42	31
Solvent	322	399	328	330	341	394	383	302	350	197
R.m.s. deviations										
Bonds (Å)	0.010	0.006	0.007	0.006	0.005	0.005	0.009	0.010	0.006	0.014
Angles (°)	1.21	0.98	1.00	0.94	0.92	0.92	1.10	1.15	0.99	1.28
Average B factors (Å ²)										
Protein	15.2	13.3	14.5	12.7	13.8	13.6	12.1	12.1	16.9	19.5
Ligand	18.1	32.5	29.6	35.2	33.0	39.6	17.8	35.0	49.9	144.4
Waters	28.8	29.8	32.9	30.3	29.3	30.1	28.0	25.4	35.7	27.6
Ramachandran plot (%)										
Outliers	0	0	0	0	0	0	0	0	0	0
Allowed	1	1	1	1	1	1	1	1	1	1
Favored	99	99	99	99	99	99	99	99	99	99

crystal structure pose of each follow-up were determined using *fconv* (Neudert & Klebe, 2011b), which can be downloaded freely from <https://agklebe.pharmazie.uni-marburg.de/>.

3. Results

3.1. Workflow for fragment growth using template docking

Elaborating fragment hits into more potent binders using commercially available compounds by exploiting the 3D structural information of the binding pose of a fragment is a very promising and at the same time a very cost-effective strategy in FBLD. Despite several advances and example campaigns, to the best knowledge of the authors this approach is not readily available as a routine or a (semi)-automated procedure. In order to fill this gap, such an optimization workflow was designed, developed and evaluated here. The different steps of the entire workflow, which is termed Frag4Lead, are presented graphically in Fig. 1. Based on a crystal structure of a fragment hit, structurally homologous compounds are retrieved from the catalog of commercially available compounds, in this case MolPort. For this, a convenient search function either via a web interface or an application programming interface (for example the MolPort KNIME node) is employed. The next step and central part of the workflow utilizes the *FlexX* docking algorithm (Rarey *et al.*, 1996) to cleave analogs into ‘FlexX fragments’, which are then superimposed onto the crystallographically bound fragment. The FlexX fragment that best matches the template fragment structurally is then used as the ‘base fragment’ to

reattach the remaining FlexX fragments in the environment of the binding pocket. In doing so, flexibly attaching moieties to the base fragment generates up to 100 docking poses so that thorough exploration of the binding pocket is ensured. Further on, the workflow includes a specific way to process and filter the docking results. For this, the *FlexX* docking poses were rescored by the DrugScoreX (DSX) per-contact score (Neudert & Klebe, 2011a). Only high-scoring unique poses identified by r.m.s.d. clustering were retained and ranked. An informed selection of follow-up candidates was then performed in a *Pymol* session (*Pymol* version 2.0; Schrödinger), highlighting favorable contact distances, per-atom contributions to the overall DSX score and molecular properties that are relevant for FBLD. Selected follow-up candidates are then purchased and validated by soaking and crystallographic structure determination. Endothiapepsin crystals usually diffract to high resolution, which is certainly beneficial for identifying the exact binding pose of the compounds. The affinities of successfully confirmed follow-up compounds are then measured via ITC. In this way, one can complete an entire round of optimization without applying any chemistry or ordering customized synthesis.

3.2. Starting fragments and follow-up compounds

In order to evaluate the power and success rate of Frag4Lead, five fragment hits that were previously discovered for EP (Radeva, Schiebel *et al.*, 2016) were used as starting points. Particular attention was paid to emulate a real-case scenario with only a few and potentially non-optimal fragment

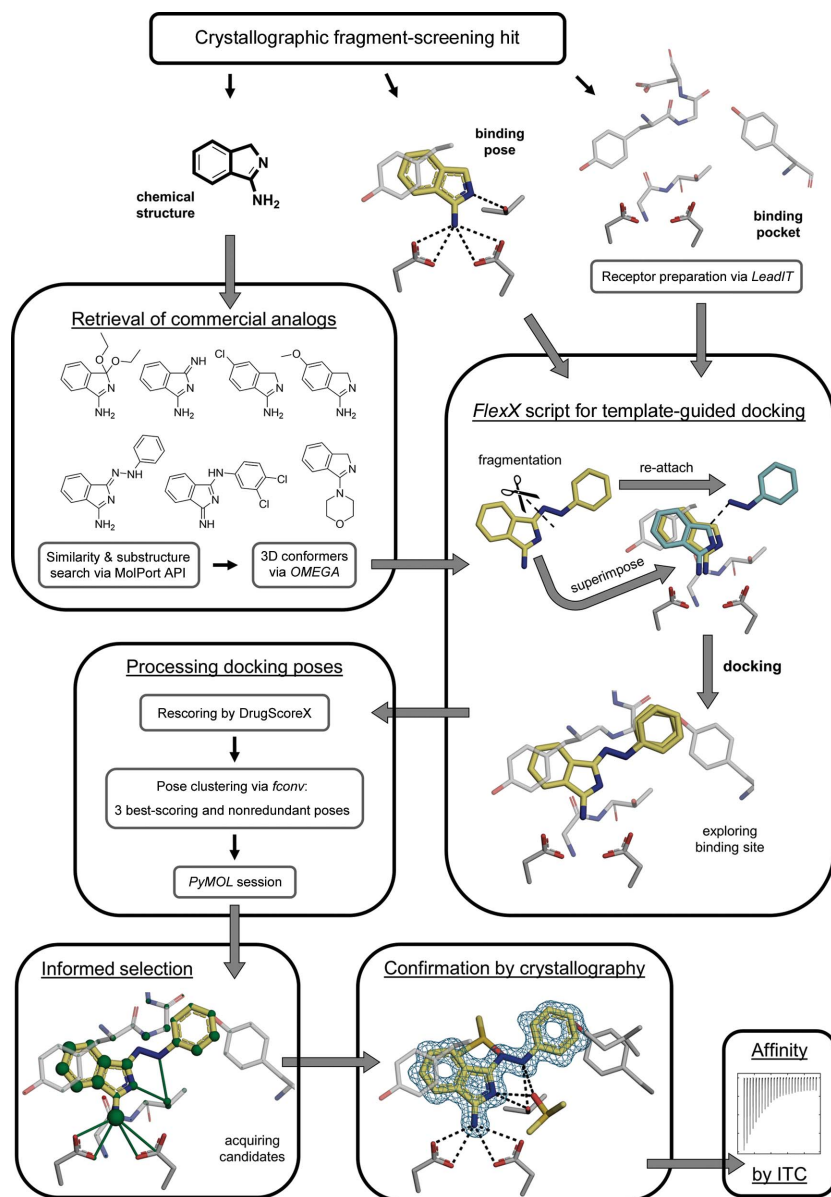


Figure 1

Frag4Lead workflow for fragment growing. The starting point of the workflow is a crystallographically detected fragment hit. It provides two types of information. The first is the identity, *i.e.* the chemical structure, of the fragment hit, based on which potential follow-up candidates are retrieved from the commercial catalog of MolPort and 3D conformers generated by *OMEGA*. The second is the 3D information of the binding pose of the fragment hit inside the binding pocket. The binding pocket is then prepared as a docking receptor via the *LeadIT* software (see Section 2.3 for details). Template-guided docking is then employed via a customized script using *FlexX* (Rarey *et al.*, 1996) using the crystallographic binding pose of the fragment as a starting point. Specifically, the *FlexX* algorithm cleaves each analog into internally rigid fragments (referred to as ‘FlexX fragments’). The FlexX fragment most similar to the starting fragment is then superimposed on the latter. Finally, each analog is incrementally reassembled by flexibly attaching its constituent FlexX fragments to the superimposed base fragment and the binding site is explored by *FlexX* docking. A maximum of 100 docking poses for each analog are generated and only the 1000 highest-scoring analogs are considered. In rare cases this process needs manual intervention, for example pruning of the docking template. The next vital step in the Frag4Lead workflow is the processing of the docking results. The docking poses generated by *FlexX* are rescored by the DrugScoreX per-contact score (see Section 2.4 for details). Next, redundant docking poses that are very similar to a better scored retained pose and would otherwise complicate the assessment of relevant poses are removed. To this end, the following procedure is applied to the docking poses of each analog. Firstly, all poses are clustered by hierarchical complete-linkage clustering with an r.m.s.d. threshold of 2.0 Å as implemented in *fconv* (Neudert & Klebe, 2011b). Only the three best-scoring, nonredundant and internally sorted poses are kept. This efficiently eliminates redundant poses and allows the direct comparison of unique poses of each analog. In order to present the ranked hit list for interactive evaluation in a way that is also amenable to non-expert users, a *PyMOL* session is created that highlights the interactions and per-atom contributions (green spheres) to the overall DrugScoreX score of the pose. Unfavorable interactions and contributions are likewise highlighted. This enables a convenient and informed selection of follow-up compounds to be acquired based on the following criteria: (i) the ability of an analog to bind in the corresponding fragment pose, (ii) the location of most of its structure in a favorable environment, indicated by high but evenly distributed per-atom contributions to the overall DrugScoreX score, (iii) the formation of additional or alternative interactions compared with the starting fragment and (iv) the adoption of a realistic conformation. The binding of acquired compounds is then investigated by X-ray crystallography. The blue mesh shows the $2mF_o - DF_c$ electron-density map for the follow-up ligand contoured at $\sigma = 1.0$. Observed binders are then further evaluated by ITC to assess their binding affinity.

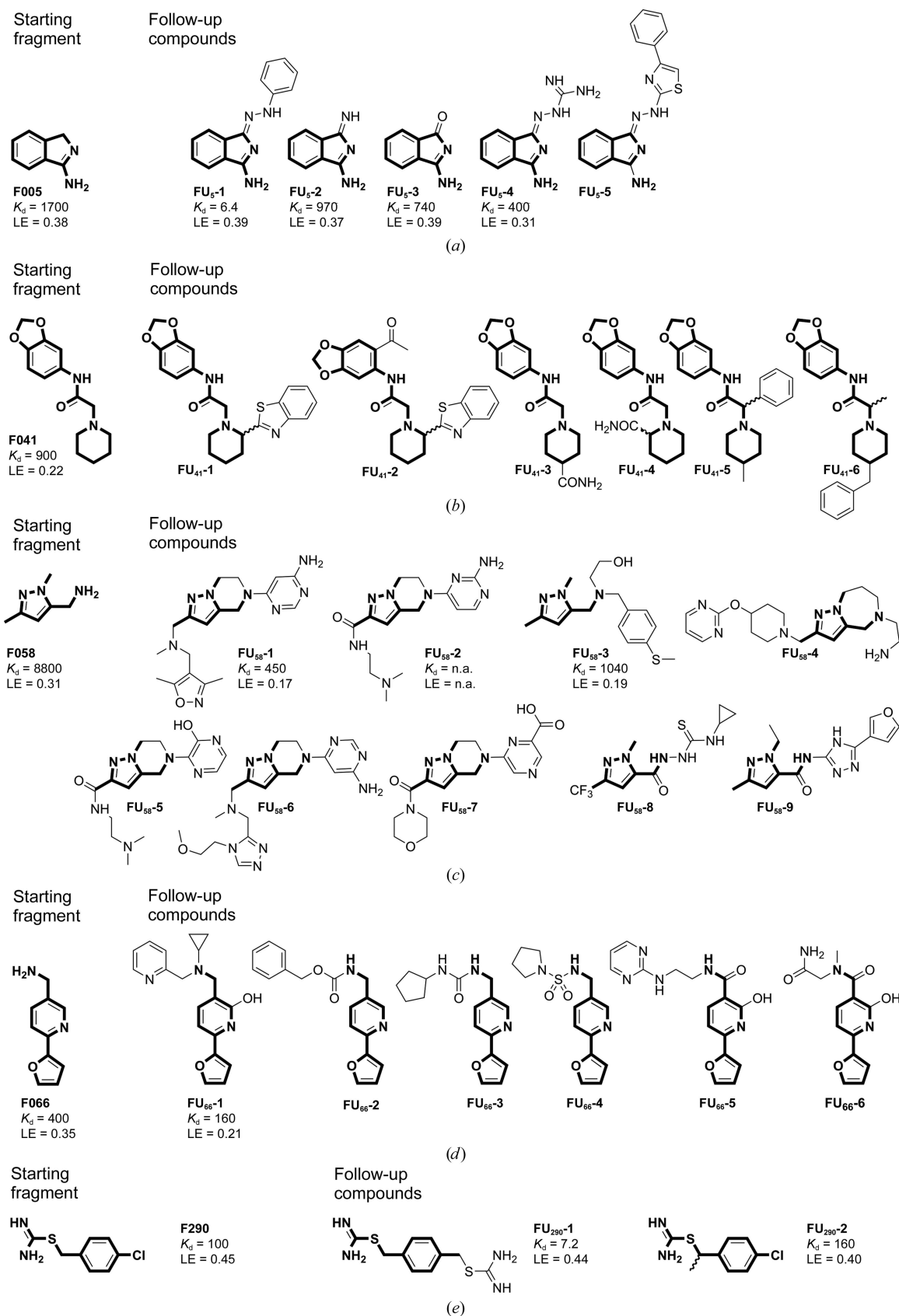


Figure 2

Starting fragments and follow-up compounds. The 2D chemical formulae of the five starting fragments of this work and the acquired follow-up candidates are given in (a)–(e). K_d is the dissociation constant of the compound from EP in μM and LE is the respective ligand efficiency in kcal mol^{-1} per atom. All crystallographic binders were evaluated by ITC, except for **FU₅₈-2**, where sufficient material for this purpose was not available. The K_d and LE values for the starting fragments were obtained in previous work (Schiebel, Radeva *et al.*, 2016).

hits available and by testing only a limited number of commercially available follow-up candidates. Only about 25–30 follow-up compounds were aimed to be acquired in order to mimic an economically realistic scenario in a typical academic setting. Table 1 depicts the selected five starting fragments and the number of potential follow-up compounds retrieved from the catalog searches. Typically, such a search reveals several hundred potential follow-up compounds, and in this campaign between 267 and 10 022 compounds were obtained. These were then narrowed down to 28 compounds highly ranked by the docking, filtering and visual inspection in the Frag4Lead workflow. Fig. 2 lists all of the selected follow-up compounds of the five starting fragments.

3.3. Validating the fragment pose

As a very first step in the optimization process, ideally even before employing the described workflow, it is important that the initial fragment-binding pose is thoroughly validated. This means that it needs to be assessed whether the binding pose observed by X-ray crystallography is retained for other highly similar analogs embedding the parent scaffold of the initial fragment hit, in order to minimize the risk of unexpected binding-mode changes during compound development. This is exemplified for starting fragment **F005** (Radeva, Schiebel *et al.*, 2016) and the follow-up compounds **FU₅-2** and **FU₅-3** (Fig. 3*a*). **F005** binds to the catalytic center of EP and establishes charge-assisted hydrogen bonds to the two catalytic aspartate residues (Fig. 3*a*). The two closely related analog fragments **FU₅-2** and **FU₅-3** differ from **F005** only by one additional atom at the 3 position. Crystal structure determination confirmed that these two follow-up fragments indeed retained the binding pose of **F005**. There are no additional directional interactions with the protein. However, in both structures an additional interaction with a DMSO molecule is observed which is not present in the original fragment **F005** structure. Strictly speaking, with such similar compounds a template-based docking approach is not needed. However, the docking was applied to all follow-up compounds irrespective of similarity and size. In this way, candidates are eliminated by the automated

workflow if they contain minimal modifications of fragments that are incompatible with the binding mode, either sterically or due to mismatched interactions. This allows the identification of close analogs that are suitable for pose validation. However, in the subsequent rapid fragment growing performed in this work, the other four starting fragments were not as stringently subjected to an experimental validation step as **F005** and were more directly used for elaboration with the objective of fast affinity improvement (Table 1).

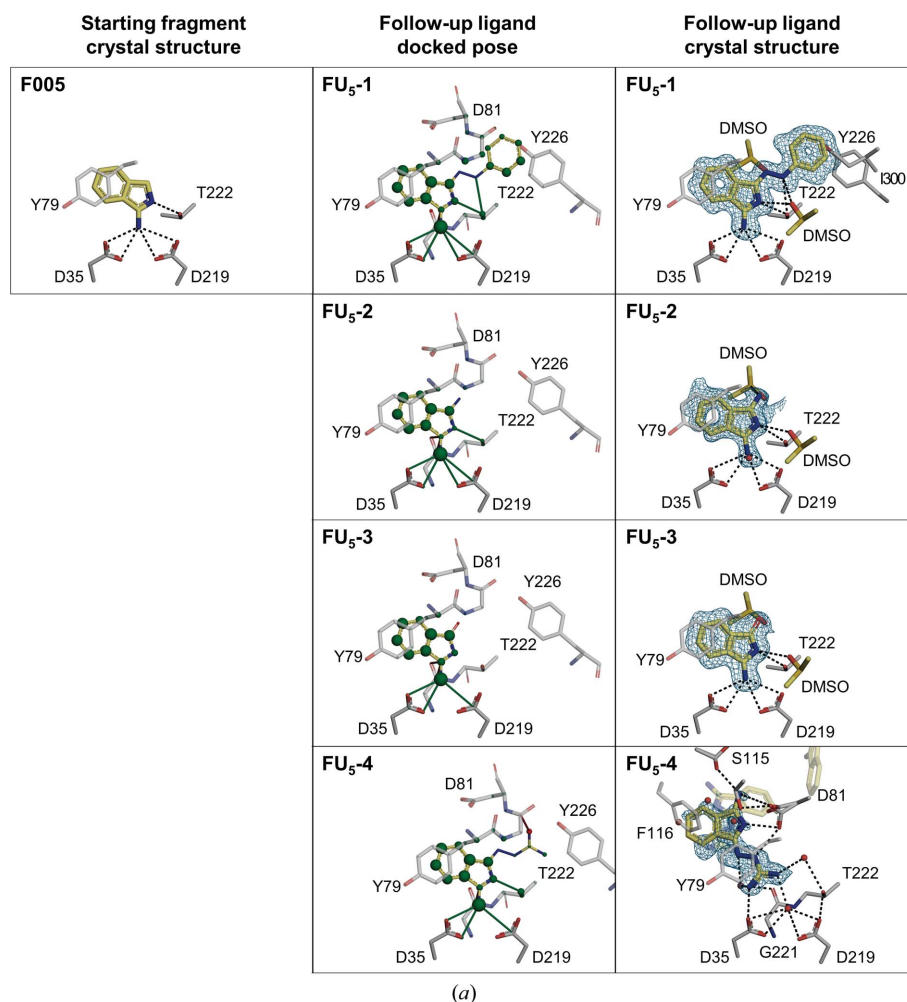
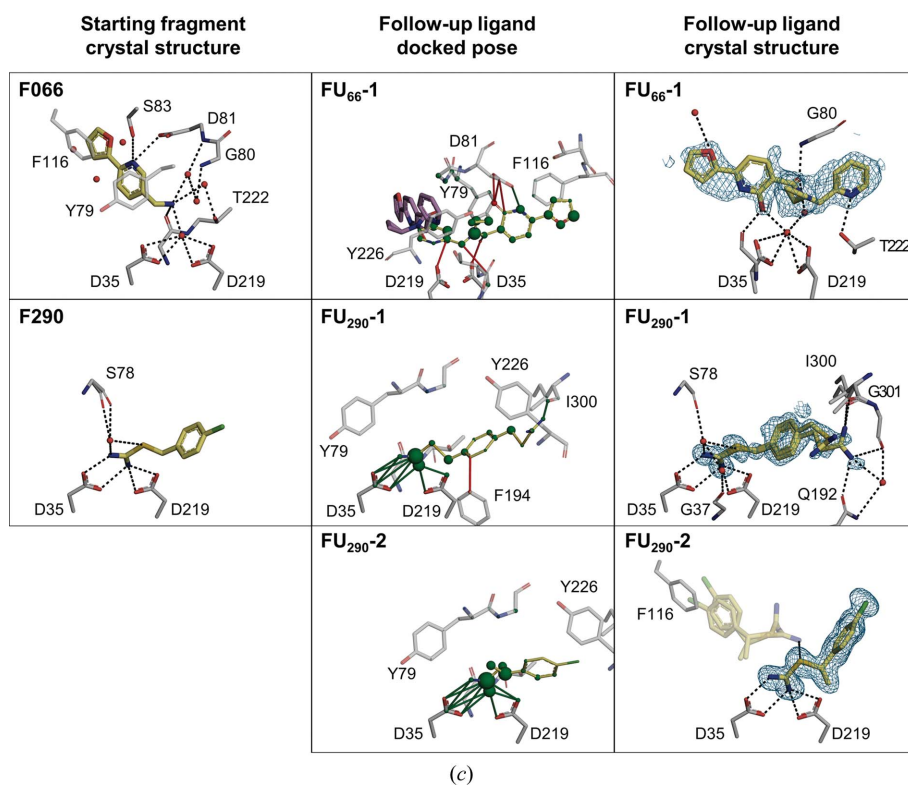
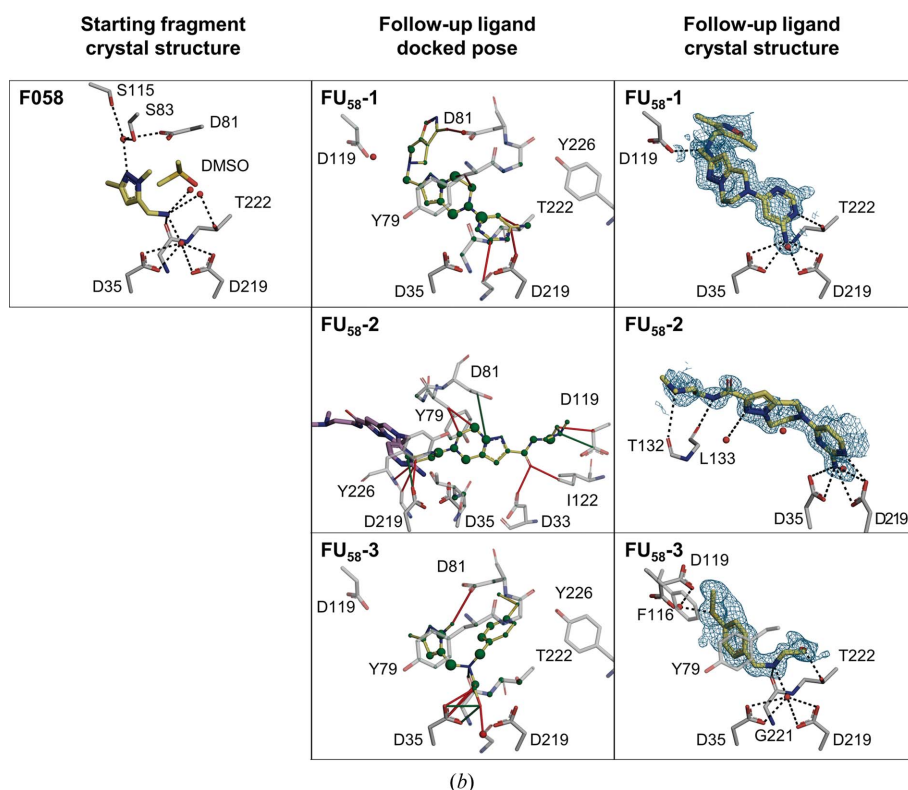


Figure 3

Side-by-side view of experimental and predicted binding poses. Shown are the binding poses of the starting fragments (left column), the docked poses of the follow-up ligands (middle column) and the binding poses of the follow-up ligands superimposed on polder OMIT $mF_o - DF_c$ electron-density maps (Liebschner *et al.*, 2017) contoured at $\sigma = 3.0$ (right column) as observed in the crystal structures for all ten follow-up ligand structures. (a) Fragment **F005** and follow-up ligands. (a, b, c) For comparison of the docking poses to the original crystallographic fragment pose, all views are identical, except for **FU₅₈-2** and **FU₆₆-1**. For the latter, the crystallographic binding poses are also shown (purple sticks) to allow a comparison of the deviating binding poses. For the docking poses, favorable and unfavorable contact distances (green and red lines) and per-atom contributions to the overall DrugScoreX score (green and red spheres, with a radius approximating the score contribution), as predicted by *DSX* (Neudert & Klebe, 2011*a*), are highlighted. For the crystal structures, polar interactions are shown as dashed lines. Ligands (yellow) and interacting residues (gray) are depicted as sticks with standard color-coding for heteroatoms and are labeled in single-letter code. Only primary binding poses near the catalytic dyad are depicted.


Figure 3 (continued)

Side-by-side view of experimental and predicted binding poses. Shown are the binding poses of the starting fragments (left column), the docked poses of the follow-up ligands (middle column) and the binding poses of the follow-up ligands superimposed on polder OMIT $mF_o - DF_c$ electron-density maps (Lieschner *et al.*, 2017) contoured at $\sigma = 3.0$ (right column) as observed in the crystal structures for all ten follow-up ligand structures. (b) Fragment **F058** and follow-up ligands. (c) Fragments **F066** and **F290** and the respective follow-up ligands.

3.4. Applying the Frag4Lead workflow to EP

Table 1 lists the potential follow-up compounds that could be found in the catalog for each of the five starting fragments and the number that are left after template-based docking has been applied as a filter. Typically, template-based docking reduces the number of candidate follow-up molecules by roughly one order of magnitude from several hundred to several dozen candidates. These were inspected visually in *PyMOL*. Based on this, 28 follow-up candidate molecules were selected and acquired for further testing by X-ray crystallography. Successful binders were subjected to ITC in order to retrieve information about the improvement in affinity compared with the starting fragment (Fig. 2). In the next paragraphs the crystallographic results will be described in detail and in the context of the obtained affinity measurements, grouped by the respective starting fragments.

3.5. Follow-up compounds for starting fragment F005

For **F005** five follow-up candidates were selected (Fig. 2a), four of which were observed in crystal structures (Fig. 3a). The strongest affinity improvement was obtained with **FU₅-1**. In this case, the pose of the starting fragment is retained (r.m.s.d. = 0.41 Å) and the additional phenylhydrazone group led to a 266-fold affinity increase from 1.7 mM to 6.4 μ M, while maintaining the ligand efficiency (LE). **FU₅-1** is a rigid molecule that does not interfere with the geometry of the binding pocket. Consequently, **FU₅-1** binds while maintaining its minimal energy conformation in the protein environment. The affinity increase of **FU₅-1** compared with **F005** is accompanied by the following additional interactions. **FU₅-1** forms a hydrogen bond between its hydrazone NH group and the hydroxyl O atom of Thr222 ($d_{N-H...O} = 3.1$ Å). Furthermore, the phenyl ring of **FU₅-1** forms hydrophobic and π -stacking interactions with the side chain of Tyr226, the amide bonds of Gly80 and Asp81, and the side chain of

Ile300 (Fig. 4a). The latter undergoes an induced fit to contact **FU₅-1**, concomitantly stabilizing the adjacent sequence segment (Ala298–Ile302). Compared with the **F005** complex, **FU₅-1** displaces no additional structural water molecules, yet it is in close contact with two DMSO molecules recruited to the binding site. Each of these DMSO molecules displaces a structural water molecule present in either the **F005** complex or the apo structure. Apparently, there is no well formed hydrogen bond between **FU₅-1** and the DMSO molecules. Even though the O atom of a DMSO molecule is close to the hydrazone NH group ($d_{\text{N-H}\cdots\text{O}} = 3.1 \text{ \AA}$), both form a non-ideal angle of $\beta(\text{N-H}\cdots\text{O}) = 120^\circ$ and the NH group of **FU₅-1** already forms a hydrogen bond to Thr222. Thus, as expected, soaking in the absence of DMSO did not alter the pose of **FU₅-1**; the originally present DMSO binding site turns out to be occupied by an acetate ion from the buffer instead (data not shown). This suggests that DMSO, which had to be included in the ITC experiments for sufficient solubilization, does not alter the apparent affinity of **FU₅-1**. The follow-up compounds **FU₅-2** and **FU₅-3** have already been described above. They each differ from **F005** merely by one atom, which does not engage in any new hydrogen bonds. Also, **FU₅-2** and **FU₅-3** exhibit nearly the same dissociation constant and ligand efficiency values as **F005**. Hence, the additional atom also does not seem to influence the strength of the hydrogen bonds compared with the **F005**–EP complex. **FU₅-4**, however, is surprisingly bound in a reversed orientation, forming a salt bridge to Asp81 via its isoindole N atom while its aminoguanidine moiety forms a salt bridge to the catalytic dyad that is partially mediated by the catalytic water. A similar binding mode to the catalytic dyad was found in an

earlier screening for fragments bound via their guanidine and amidine groups, but none of them utilized the catalytic water (Radeva, Schiebel *et al.*, 2016). It may be speculated that the strong interaction between the additional guanidinium group and the catalytic dyad led to the reversal of the orientation of **FU₅-4**. This allegedly non-optimal pose of **FU₅-4** is accompanied by only a slight increase in affinity ($K_d = 400 \text{ \mu M}$) and by a significant decrease in LE (LE = $0.31 \text{ kcal mol}^{-1}$ per atom). A possible explanation for the minor affinity enhancement could be a presumably strong increase in the desolvation costs of this more polar fragment upon binding. However, conclusive reasoning in the case of such large changes of binding mode is difficult in general.

3.6. Follow-up compounds for starting fragment F041

For **F041** six follow-up candidates were selected (Fig. 2b), none of which was observed in a crystal structure. It may be the case that the low ligand efficiency of **F041** (LE = $0.22 \text{ kcal mol}^{-1}$ per atom) already indicated weak binding, and a preceding binding-pose validation using closer analogs would have been highly advisable.

3.7. Follow-up compounds for starting fragment F058

For **F058** nine follow-up candidates were selected (Fig. 2c), which is the largest number for all five starting fragments in this work. Three of them (**FU₅₈-1**, **FU₅₈-2** and **FU₅₈-3**) were observed in crystal structures (Fig. 3b), but none of them maintained the binding pose of the original fragment. However, **FU₅₈-1** bound with the corresponding portion still in the proximity of the original position of **F058** in the S1 pocket.

The diazole ring is flipped and located roughly two bond lengths further away from the catalytic dyad. Notably, this shift enables the formation of a salt bridge between the 4-aminopyrimidine moiety of **FU₅₈-1** and the catalytic dyad, with the amino N atom displacing the catalytic water. Additionally, a salt bridge is formed from the tertiary amine of **FU₅₈-1** to the carboxylate O atom of Asp119 ($d_{\text{N-H}\cdots\text{O}} = 2.9 \text{ \AA}$). However, the intricate network of water-mediated interactions between **F058** and Asp81, Ser83, Ser115 and Thr222, as well as the catalytic dyad, was largely not formed in the **FU₅₈-1**–EP complex, most likely due to the missing primary amine of **F058**, which in **FU₅₈-1** was replaced by a tertiary amine connecting to the 4-aminopyrimidine moiety. **FU₅₈-1** exhibits an about 20-fold higher affinity than the weakly

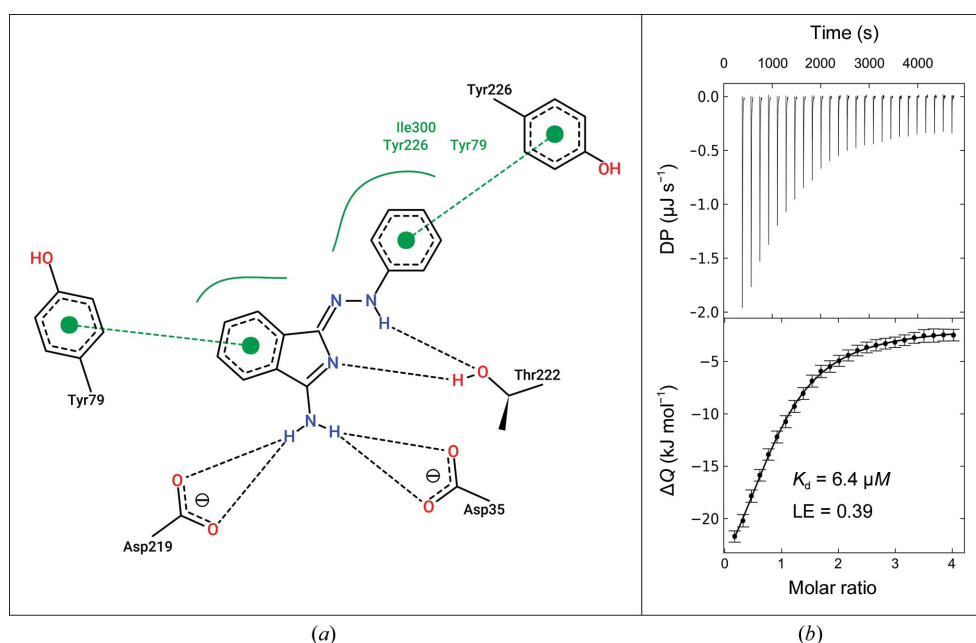


Figure 4
Details of the interaction of **FU₅-1** with EP and the corresponding ITC results. (a) For the highest affinity binder identified with the applied workflow, **FU₅-1**, the atomic interaction network is shown. The picture was generated with *PoseView* (Stierand & Rarey, 2010). (b) Representative ITC thermogram of the direct titration of **FU₅-1** against EP.

bound **F058** in an ITC experiment ($K_d = 450 \mu\text{M}$ versus 8.8 mM), but it also contains 17 more non-H atoms than the starting fragment. Consequently, the ligand efficiency is decreased drastically compared with **F058** (LE = 0.17 versus $0.31 \text{ kcal mol}^{-1}$ per atom). Similarly, **FU₅₈-2** binds displacing the catalytic water with its 4-aminopyrimidine moiety, yet mirrored at the plane spanned by the carboxylate groups of the catalytic dyad. Thus, the remainder of **FU₅₈-2** is oriented in the S1' direction, also occupying the S2' pocket of the substrate-binding cleft. Here, **FU₅₈-2** forms two direct and one water-mediated hydrogen bonds in addition to the salt bridge with the catalytic dyad, yet it does not form a salt bridge via its terminal tertiary amine. Unfortunately, **FU₅₈-2** could not be characterized by ITC because sufficient material was not available. In the structure obtained by soaking **FU₅₈-3**, only its substructure analog **FU₅₈-3b** (2-[[4-(methylthio)benzyl]-amino]ethan-1-ol) could unambiguously be identified in the electron density and built into the crystal structure after verifying its presence as an impurity in the obtained sample of **FU₅₈-3** (purity of >90% according to the provider) by mass spectrometry (see supporting information, including Supplementary Fig. S5). Given the lack of electron density for **FU₅₈-3** in our crystallographic experiment, it may be speculated that either **FU₅₈-3** does not bind in solution as well or that **FU₅₈-3b** efficiently competes with **FU₅₈-3** in the crystal structure. One might also speculate that the true concentration of the active species in the displacement ITC experiment is underestimated, so that the apparent K_d of $1040 \mu\text{M}$ must be considered an upper limit. However, despite the identification of **FU₅₈-3b** by mass spectrometry, the presence of other, potential nonspecific species in the impure sample cannot be excluded, so that attributing the apparent K_d to any specific compound is highly unreliable.

3.8. Follow-up compounds for starting fragment F066

For **F066** six follow-up candidates were selected (Fig. 2*d*), of which one was observed in the crystal structure (Fig. 3*c*). **FU₆₆-1** did not maintain the original binding pose observed for **F066**. Instead, it bridges the catalytic dyad, thereby accessing both directions of the peptide-binding cleft. This new pose is facilitated by the additional hydroxyl group of **FU₆₆-1**, which is located vicinal to the pyridine N atom of **F066**. Although this additional hydroxyl group was predicted to be compatible with the fragment pose, it unexpectedly forms new hydrogen bonds to the catalytic water as well as the carbonyl O atom of Gly80 in the pose of **FU₆₆-1**.

3.9. Follow-up compounds for starting fragment F290

F290 is a special case for follow-up candidate selection. Compounds that contain an isothioureia moiety as part of a ring were not properly matched to the starting fragment. This problem was solved by pruning **F290** down to its isothioureia moiety for follow-up compound identification. In this way, two follow-up candidates were selected (Fig. 2*e*), both of which were observed in crystal structures and maintained the original binding pose (Fig. 3*c*). In addition, for both a second

alternative binding pose was observed. The affinity of **FU₂₉₀-1** is increased 14-fold ($K_d = 7.2 \mu\text{M}$) compared with **F290** ($K_d = 100 \mu\text{M}$). At the same time, the LE was left essentially unchanged (0.44 and $0.45 \text{ kcal mol}^{-1}$ per atom, respectively). This means that the affinity of **FU₂₉₀-1** increased proportional to its size. Thus, **FU₂₉₀-1** may be another good starting point for further optimization, although the affinity determination was hampered by a noisy baseline in ITC experiments (Supplementary Fig. S4), allegedly due to its low purity (>90% according to the provider). The primary binding site of **FU₂₉₀-1** is occupied by two conformers, which bind very similarly to **F290**. While conformer *A* [r.m.s.d. of the maximum common substructure (r.m.s.d._{MCS}) = 0.20 \AA , 42% occupancy] forms no additional direct polar interactions, conformer *B* (r.m.s.d._{MCS} = 0.29 \AA , 53% occupancy) donates two additional hydrogen bonds from its guanidine NH group to the side-chain amide O atom of Gln192 ($d_{\text{N-H}\cdots\text{O}} = 3.3 \text{ \AA}$) and to the equidistant backbone carbonyl O atom of Ile300 ($d_{\text{N-H}\cdots\text{O}} = 3.4 \text{ \AA}$), which also adopts two alternative conformations. Soaking in racemic **FU₂₉₀-2** resulted in (*R*)-**FU₂₉₀-2** bound with the isothioureia moiety closely maintaining the pose in **F290**. However, the affinity was unchanged ($K_d = 160 \mu\text{M}$ for the racemic mixture) and the additional methyl group at the stereocenter coincides with a shift of the *p*-chlorobenzyl moiety away from its original position (r.m.s.d._{MCS} = 2.9 \AA) towards the flap loop. This could be due to a steric clash or alteration of the torsional preference within **FU₂₉₀-2**. The flap loop itself is displaced as well, and presumably this is the reason why docking did not produce the correct pose even before filtering. In addition, a nearby secondary site is weakly occupied by overlapping poses of (*R*)- and (*S*)-**FU₂₉₀-2**, both of which form a π -stacking interaction with Phe116, while one donates a weak hydrogen-bond to the isothioureia S atom of the primary fragment.

All in all, the workflow assembled and tested here for filtering commercially available analogs of fragment hits via template-based docking proved to be successful in the EP campaign. From only five starting fragments and a limited number of 28 follow-up compounds acquired, ten binders were identified by crystallography. Five of the follow-up binders bound in the pose of the original fragment and four of them exhibited a significantly increased affinity. Two of them, **FU₅-1** and **FU₂₉₀-1**, even reached single-digit micromolar affinity and **FU₅-1** showed a remarkable 266-fold improvement in affinity.

4. Discussion

Fragment screening by crystallography typically provides multiple fragment hits as potential starting points for FBLD. For each promising hit, it is advisable to first test close analogs in order to validate the binding pose of a given fragment hit. In a next step, the fragment needs to be grown into a larger molecule with substantial affinity improvement. This is still the most challenging step in FBLD (for a review of such methods, see de Souza Neto *et al.*, 2020). One approach that ensures rapid progress of the project, especially in a typical academic

setting with limited financial resources, is to exploit follow-up candidates that are readily available via vendor catalog databases. This limits the number of molecules compared with exploring large virtual chemical spaces, but still returns too many for the manual selection of promising compounds. Completely unbiased docking may help in an automated fashion, although this often generates binding poses that deviate from the original fragment, thus contradicting the idea of rational fragment-based design and complicating the comparison of the suggested poses.

Here, we demonstrate our Frag4Lead workflow, which is based on template-based virtual screening of commercially available follow-up compounds. It utilizes the fragment pose found in a crystal structure, for example from a crystallographic fragment-screening campaign, as additional information. Frag4Lead was validated on the model system EP using a limited number of both starting fragments and compounds to be acquired. Of the more than 70 fragment hits identified by CFS against EP (41 addressing the catalytic dyad), five were selected for this study (Köster *et al.*, 2011; Radeva, Schiebel *et al.*, 2016; Radeva, Krimmer *et al.*, 2016). These five starting fragments were selected to emulate a real-case scenario with only a few and potentially non-optimal fragment hits to follow up on. From the five starting fragments, 28 follow-up compounds were identified and purchased. Out of the 28 selected follow-up compounds, ten binders could be identified. Even though the original fragment pose was retained for only five of them, two follow-up compounds exhibited a very successful advancement to an affinity of less than 10 μM . An earlier study of docking-supported fragment growth performed similarly (Marchand *et al.*, 2016). There, six out of 16 selected candidates were binders (*i.e.* a similar success rate). However, the best affinity of 279 μM reached is two orders of magnitude lower than in the campaign presented here. Also, compared with alternative approaches for the rapid elaboration of fragment hits, for example by screening diverse fragment follow-up compounds in crude reaction mixtures from fast chemistry (Baker *et al.*, 2020; Bentley *et al.*, 2020), the presented example campaign via Frag4Lead ended up with a similar number of hits and better affinity improvement. It seems rather obvious that these approaches could complement each other. For instance, a relatively large virtual chemical space of close fragment analogs with suitable reaction handles combined with building blocks available in-house could be constructed and filtered by template docking to identify the most promising candidates and the building blocks required for their synthesis.

The follow-up compound with the highest affinity in the campaign presented here, **FU₅-1**, seems to be suitable for further ligand development for two reasons. Firstly, it forms a tight cluster of interactions, with the starting fragment substructure acting as an anchor. In addition, this anchor has an excellent growth vector along which the phenylhydrazone moiety of **FU₅-1** is oriented, forming additional hydrogen bonds and π -stacking interactions with residue Tyr226. Most importantly, however, the simple and fast synthesizability of **FU₅-1** and derivatives thereof (see supporting information)

enables efficient exploration of this growth vector, thus making a rapid elaboration of possible interactions and structure–activity relationships feasible. Yet, despite its low micromolar affinity and favorable interactions, it may seem questionable whether the hydrazone structure of **FU₅-1** is suitable for drug development. Reportedly, hydrazones may form hydrazines and other reactive or toxic derivatives (Smith, 2011). Indeed, decomposition of the synthesized **FU₅-1** was observed when the compound was exposed to air at room temperature over a longer period of time. Nonetheless, **FU₅-1** unambiguously bound in the crystal structure after soaking for 24 h at 18°C under slightly acidic conditions, indicating the stability of its protonated form in solution. Moreover, the existence of bioactive hydrazones, some of which are approved drugs (Rollas & Küçükgüzel, 2007), and the potential for bioisosteric replacement of the hydrazone, for example by amides or ureas, demonstrates that **FU₅-1** and its derivatives may well be reasonable starting points for the development of lead or tool compounds.

Other follow-up compounds in the presented EP campaign did not maintain the anticipated binding pose. In fact, a change of binding mode upon chemical variation is not uncommon, and adding substitutions that enable new but competing interactions is reportedly a major cause of this (Malhotra & Karanicolas, 2017; Oebbecke *et al.*, 2021). In the case of **FU₅-4**, for example, the changed binding pose could supposedly have been anticipated or predicted, as interactions of the guanidine moiety with the catalytic dyad are very plausible. Also, other fragments with a guanidine moiety were found to bind to the catalytic dyad, for example PDB entry 4ycy (Radeva, Schiebel *et al.*, 2016). One may also test for such possibilities via the template-docking approach in order to assess whether follow-up candidates are also compatible with the poses of other known fragment hits. However, approaches to estimate the absolute and relative stability of binding poses are difficult and laborious, so that crystallographic verification is often easier and more straightforward in the presence of a suitable crystal system and soaking condition. Therefore, these findings strongly encourage the validation of binding poses of fragment hits from a primary crystallographic screening using close analogs prior to embarking on growth strategies such as Frag4Lead.

Another follow-up ligand that did not maintain the binding pose of the starting fragment is **FU₅₈-1**. This may not be surprising because although **FU₅₈-1** is a superstructure of **F058**, its primary amino group, which forms a direct hydrogen bond to the catalytic water, is replaced by a tertiary anilinic nitrogen, thus losing its hydrogen-bonding capacity. Instead, the additional 4-aminopyrimidine moiety of **FU₅₈-1** was anticipated to replace the catalytic water and address the catalytic dyad directly (although not observed in the docking pose). Indeed, this interaction was observed in the crystal structure but required a flip of the central heterocycle as well as a slight shift away from the catalytic dyad. However, in the docked pose of **FU₅₈-1** the amino group pointing away from the catalytic dyad could have been interpreted as an indicator of a suboptimal interaction geometry of the central

part of the ligand. For this reason, future improvements of the Frag4Lead workflow should aim at identifying unstable predicted binding poses in order to focus on the most promising starting fragments and their respective follow-up candidates.

One approach for future improvements may come from a better assessment of observed and predicted interactions, for example via descriptors based on the statistical occurrence of protein–ligand contacts in the PDB (Tosstorff *et al.*, 2020). Another improvement would be to predict the unexpected or flipped poses that were observed in the crystallographic experiments of the chosen follow-ups (*i.e.* **FU₅-4**, **FU₅₈-1**, **FU₅₈-2**, **FU₅₈-3** and **FU₆₆-1**) with high confidence in order to deprioritize those compounds in the selection process. However, in a retrospective, unbiased docking experiment of the successful crystallographic binders using the newest GUI version of *SeeSAR* (version 11.0.0), unbiased *FlexX* docking and HYDE scoring did not produce any pose within an r.m.s.d. of 2 Å of the experimentally observed unexpected poses. Additionally, for the follow-up compounds that retained the binding pose of the fragment, none of the predictions turned up within the three highest scored poses. Only the pose of **FU₂₉₀-2** showed up in the ten highest scored poses (see Supplementary Table S4). This again underlines the advantage of the template-based docking employed in the Frag4Lead workflow, making use of the obtained structural information of the fragment hits to improve the follow-up compound selection.

In view of the large and constantly growing space of reliably synthesizable compounds (van Hilten *et al.*, 2019), the presented template-guided docking approach enables the rapid early discovery of improved ligands without custom synthesis requirements. In addition, the underlying docking functionality has recently been developed and implemented similarly in the *SeeSAR* software (BioSolveIT GmbH) with further improved substructure-matching algorithms that allow the guided docking of close non-substructure analogs. However, for large virtual chemical spaces with billions of compounds, the computational cost will increase. This might require more efficient prefiltering to remove sterically incompatible follow-up candidates prior to docking, for example by employing the recently described shape-based descriptors (Penner *et al.*, 2020).

The presented generic strategy is able to identify suitable follow-up candidates from any source of analogs to exploit fragment-bound structures. Supposedly, it will be more efficient in combination with fragment libraries that are designed to comprise starting points for the easy exploration of large chemical spaces (Cox *et al.*, 2016). However, the presented concept also harmonizes with our newly introduced, structurally diverse F2X-Universal Library, which is based on 3D shape and pharmacophore clustering of a large, readily available fragment space and achieves high hit rates (Wollenhaupt *et al.*, 2020). For each member, *i.e.* cluster representative, of the F2X-Universal Library, there is a high likelihood that similar and readily purchasable compounds exist.

The Frag4Lead workflow evaluated here serves as a first attempt to automate initial fragment-hit expansion for non-expert users and projects with limited resources for laborious follow-up chemistry. This limitation is even more pronounced in academic settings and often provides the most critical bottleneck in academic compound development. A key reason for this is that funding for professional compound synthesis is much harder to acquire than for personnel and equipment. However, saving costs or ensuring faster progress through more efficient fragment expansion is also highly desirable in an industrial setting.

Clearly, the concepts employed in the Frag4Lead workflow need to be optimized further, in particular with respect to transferability to different sites. For now, the Frag4Lead workflow is available to all users of the HZB fragment screening facility. However, its successful application demonstrates its clear potential to contribute to more efficient structure-based ligand design, especially in academia, in the initial stage of drug development.

5. Related literature

The following references are cited in the supporting information for this article: Biiitseva *et al.* (2013), Krimmer & Klebe (2015) and Wolf & Vollmann (1956).

Acknowledgements

We are grateful to KNIME.com AG, Chemical Computing Group Inc., MolPort SIA, Inte:Ligand GmbH and OpenEye Scientific Software Inc. for granting no-cost academic licenses for their software. We are grateful to BioSolveIT GmbH, and in particular to Marcus Gastreich, for technical support in using *LeadIT* and *FlexX*. We thank Nina Zitzer (University of Marburg) for conducting mass-spectrometry experiments and Marko Kljajic (University of Marburg) for valuable discussion. Open access funding enabled and organized by Projekt DEAL.

Funding information

This study was supported by the German Ministry of Science and Education (BMBF) via the projects Frag2Xtal (No. 05K13RM1) and Frag4Lead (No. 05K16RM1), as well as by iNEXT-Discovery, project No. 871037, funded by the Horizon 2020 program of the European Commission.

References

- Baker, L. M., Aimon, A., Murray, J. B., Surgenor, A. E., Matassova, N., Roughley, S. D., Collins, P. M., Krojer, T., von Delft, F. & Hubbard, R. E. (2020). *Commun. Chem.* **3**, 122.
- Beisken, S., Meinl, T., Wiswedel, B., de Figueiredo, L. F., Berthold, M. & Steinbeck, C. (2013). *BMC Bioinformatics*, **14**, 257.
- Bentley, M. R., Ilyichova, O. V., Wang, G., Williams, M. L., Sharma, G., Alwan, W. S., Whitehouse, R. L., Mohanty, B., Scammells, P. J., Heras, B., Martin, J. L., Totsika, M., Capuano, B., Doak, B. C. & Scanlon, M. J. (2020). *J. Med. Chem.* **63**, 6863–6875.
- Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K. & Wiswedel, B. (2008). *Data Analysis*,

- Machine Learning and Applications*, edited by C. Preisach, H. Burkhardt, L. Schmidt-Thieme & R. Decker, pp. 319–326. Berlin, Heidelberg: Springer.
- Bitseva, A., Groth, U. & Hordiyenko, O. (2013). *J. Heterocycl. Chem.* **50**, 1140–1145.
- Chevillard, F. & Kolb, P. (2015). *J. Chem. Inf. Model.* **55**, 1824–1835.
- Cox, O. B., Krojer, T., Collins, P., Monteiro, O., Talon, R., Bradley, A., Fedorov, O., Amin, J., Marsden, B. D., Spencer, J., von Delft, F. & Brennan, P. E. (2016). *Chem. Sci.* **7**, 2322–2330.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Erlanson, D. A., Davis, B. J. & Jahnke, W. (2019). *Cell Chem. Biol.* **26**, 9–15.
- Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. (2016). *Nat. Rev. Drug Discov.* **15**, 605–619.
- Gohlke, H., Hendlich, M. & Klebe, G. (2000). *J. Mol. Biol.* **295**, 337–356.
- Hall, R. J., Murray, C. W. & Verdonk, M. L. (2017). *J. Med. Chem.* **60**, 6440–6450.
- Hartshorn, M. J., Murray, C. W., Cleasby, A., Frederickson, M., Tickle, I. J. & Jhoti, H. (2005). *J. Med. Chem.* **48**, 403–413.
- Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T. (2010). *J. Chem. Inf. Model.* **50**, 572–584.
- Hilten, N. van, Chevillard, F. & Kolb, P. (2019). *J. Chem. Inf. Model.* **59**, 644–651.
- Houtman, J. C. D., Brown, P. H., Bowden, B., Yamaguchi, H., Appella, E., Samelson, L. E. & Schuck, P. (2007). *Protein Sci.* **16**, 30–42.
- Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C. A. & Schuck, P. (2012). *Anal. Chem.* **84**, 5066–5073.
- Keserü, G. M., Erlanson, D. A., Ferenczy, G. G., Hann, M. M., Murray, C. W. & Pickett, S. D. (2016). *J. Med. Chem.* **59**, 8189–8206.
- Köster, H., Craan, T., Brass, S., Herhaus, C., Zentgraf, M., Neumann, L., Heine, A. & Klebe, G. (2011). *J. Med. Chem.* **54**, 7784–7796.
- Krimmer, S. G. & Klebe, G. (2015). *J. Comput. Aided Mol. Des.* **29**, 867–883.
- Krojer, T., Talon, R., Pearce, N., Collins, P., Douangamath, A., Brandao-Neto, J., Dias, A., Marsden, B. & von Delft, F. (2017). *Acta Cryst.* **D73**, 267–278.
- Kuhnert, M., Köster, H., Bartholomäus, R., Park, A. Y., Shahim, A., Heine, A., Steuber, H., Klebe, G. & Diederich, W. E. (2015). *Angew. Chem. Int. Ed.* **54**, 2849–2853.
- Lamoree, B. & Hubbard, R. E. (2017). *Essays Biochem.* **61**, 453–464.
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* **D75**, 861–877.
- Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C. & Adams, P. D. (2017). *Acta Cryst.* **D73**, 148–157.
- Lima, G. M. A., Talibov, V. O., Jagudin, E., Sele, C., Nyblom, M., Knecht, W., Logan, D. T., Sjögren, T. & Mueller, U. (2020). *Acta Cryst.* **D76**, 771–777.
- Malhotra, S. & Karanicolos, J. (2017). *J. Med. Chem.* **60**, 128–145.
- Marchand, J. R., Lolli, G. & Caffisch, A. (2016). *J. Med. Chem.* **59**, 9919–9927.
- Mortenson, P. N., Erlanson, D. A., de Esch, I. J. P., Jahnke, W. & Johnson, C. N. (2019). *J. Med. Chem.* **62**, 3857–3872.
- Mueller, U., Förster, R., Hellmig, M., Huschmann, F. U., Kastner, A., Malecki, P., Pühlinger, S., Röwer, M., Sparta, K., Steffien, M., Ühlein, M., Wilk, P. & Weiss, M. S. (2015). *Eur. Phys. J. Plus*, **130**, 141.
- Muñoz, E. & Piñeiro, A. (2018). *J. Appl. Bioanal.* **4**, 124–139.
- Murray, C. W. & Rees, D. C. (2016). *Angew. Chem. Int. Ed.* **55**, 488–492.
- Neudert, G. & Klebe, G. (2011a). *J. Chem. Inf. Model.* **51**, 2731–2745.
- Neudert, G. & Klebe, G. (2011b). *Bioinformatics*, **27**, 1021–1022.
- Oebbecke, M., Siefker, C., Wagner, B., Heine, A. & Klebe, G. (2021). *Angew. Chem. Int. Ed.* **60**, 252–258.
- O'Reilly, M., Cleasby, A., Davies, T. G., Hall, R. J., Ludlow, R. F., Murray, C. W., Tisi, D. & Jhoti, H. (2019). *Drug Discov. Today*, **24**, 1081–1086.
- Penner, P., Martiny, V., Gohier, A., Gastreich, M., Ducrot, P., Brown, D. & Rarey, M. (2020). *J. Chem. Inf. Model.* **60**, 6269–6281.
- Radeva, N., Krimmer, S. G., Stieler, M., Fu, K., Wang, X., Ehrmann, F. R., Metz, A., Huschmann, F. U., Weiss, M. S., Mueller, U., Schiebel, J., Heine, A. & Klebe, G. (2016). *J. Med. Chem.* **59**, 7561–7575.
- Radeva, N., Schiebel, J., Wang, X., Krimmer, S. G., Fu, K., Stieler, M., Ehrmann, F. R., Metz, A., Rickmeyer, T., Betz, M., Winquist, J., Park, A. Y., Huschmann, F. U., Weiss, M. S., Mueller, U., Heine, A. & Klebe, G. (2016). *J. Med. Chem.* **59**, 9743–9759.
- Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. (1996). *J. Mol. Biol.* **261**, 470–489.
- Reulecke, I., Lange, G., Albrecht, J., Klein, R. & Rarey, M. (2008). *ChemMedChem*, **3**, 885–897.
- Rollas, S. & Küçükgülzel, Ş. G. (2007). *Molecules*, **12**, 1910–1939.
- Rühmann, E., Betz, M., Fricke, M., Heine, A., Schäfer, M. & Klebe, G. (2015). *Biochim. Biophys. Acta*, **1850**, 647–656.
- Schiebel, J., Krimmer, S. G., Röwer, K., Knörlein, A., Wang, X., Park, A. Y., Stieler, M., Ehrmann, F. R., Fu, K., Radeva, N., Krug, M., Huschmann, F. U., Glöckner, S., Weiss, M. S., Mueller, U., Klebe, G. & Heine, A. (2016). *Structure*, **24**, 1398–1409.
- Schiebel, J., Radeva, N., Krimmer, S. G., Wang, X., Stieler, M., Ehrmann, F. R., Fu, K., Metz, A., Huschmann, F. U., Weiss, M. S., Mueller, U., Heine, A. & Klebe, G. (2016). *ACS Chem. Biol.* **11**, 1693–1701.
- Schmidt, M. F. & Rademann, J. (2009). *Trends Biotechnol.* **27**, 512–521.
- Schneider, N., Lange, G., Hindle, S., Klein, R. & Rarey, M. (2013). *J. Comput. Aided Mol. Des.* **27**, 15–29.
- Schulz, M. N., Landström, J., Bright, K. & Hubbard, R. E. (2011). *J. Comput. Aided Mol. Des.* **25**, 611–620.
- Smith, G. F. (2011). *Prog. Med. Chem.* **50**, 1–47.
- Souza Neto, L. R. de, Moreira-Filho, J. T., Neves, B. J., Maidana, R. L. B. R., Guimarães, A. C. R., Furnham, N., Andrade, C. H. & Silva, F. P. Jr (2020). *Front. Chem.* **8**, 93.
- Sparta, K. M., Krug, M., Heinemann, U., Mueller, U. & Weiss, M. S. (2016). *J. Appl. Cryst.* **49**, 1085–1092.
- Stierand, K. & Rarey, M. (2010). *ACS Med. Chem. Lett.* **1**, 540–545.
- Tosstorff, A., Cole, J. C., Taylor, R., Harris, S. F. & Kuhn, B. (2020). *J. Chem. Inf. Model.* **60**, 6595–6611.
- Willett, P., Winterman, V. & Bawden, D. (1986). *J. Chem. Inf. Comput. Sci.* **26**, 36–41.
- Wolf, W. & Vollmann, H. (1956). German Patent DE 941845.
- Wollenhaupt, J., Barthel, T., Lima, G. M. A., Metz, A., Wallacher, D., Jagudin, E., Huschmann, F. U., Hauss, T., Feiler, C. G., Gerlach, M., Hellmig, M., Förster, R., Steffien, M., Heine, A., Klebe, G., Mueller, U. & Weiss, M. S. (2021). *J. Vis. Exp.*, e62208.
- Wollenhaupt, J., Metz, A., Barthel, T., Lima, G. M. A., Heine, A., Mueller, U., Klebe, G. & Weiss, M. S. (2020). *Structure*, **28**, 694–706.
- Yuriev, E. & Ramsland, P. A. (2013). *J. Mol. Recognit.* **26**, 215–239.



STRUCTURAL
BIOLOGY

Volume 77 (2021)

Supporting information for article:

Frag4Lead: growing crystallographic fragment hits by catalog using fragment-guided template docking

Alexander Metz, Jan Wollenhaupt, Steffen Glöckner, Niki Messini, Simon Huber, Tatjana Barthel, Ahmed Mehrabed, Hans-Dieter Gerber, Andreas Heine, Gerhard Klebe and Manfred S. Weiss

Table S1. Acquired follow-up candidates, affinities, providers and purities.

Follow-Up Compound	$LE^{[a]}$	$K_d^{[b]}$ [μM]	CAS Number	Salt	Provider	Provider Catalog ID	Purity ^[c] [%]
FU₅-1 · HCl	0.39	6.4	40227-56-3	1xHCl	resynthesized	resynthesized	98.5 ^[d]
FU₅-1	^[e]	^[e]	46886-52-6		Vitas-M	STL231607	90 ^[f]
FU₅-2	0.37	970	3468-11-09		AK Scientific	P175	95
FU₅-3	0.39	740	14352-51-3		ENAMINE	EN300-43093	90
FU₅-4	0.31	400	879916-45-7		Vitas-M	STK594879	90 ^[f]
FU₅-5			571161-36-9		Vitas-M	STK583580	90 ^[f]
FU₄₁-1			919949-40-9		ENAMINE	Z106906918	90
FU₄₁-2			1007745-42-7		UORSY	PB106986608	n/a
FU₄₁-3			850020-65-4		ENAMINE	Z46204096	90
FU₄₁-4			1105566-03-7		ENAMINE	Z46442873	90 ^[f]
FU₄₁-5			702651-79-4		ChemDiv	C094-0062	n/a ^[f]
FU₄₁-6			793678-62-3		ENAMINE	Z44502465	90
FU₅₈-1	0.17	450	1377929-97-9		ChemBridge	57162056	90
FU₅₈-2			1331991-20-8		ChemBridge	69271435	n/a
FU₅₈-3	(0.19) ^[g]	(1040) ^[g]	1244921-80-9		ChemBridge	58887845	90
FU₅₈-4			1360217-23-7		ChemBridge	22615646	90
FU₅₈-5			1332207-28-9		ChemBridge	55009691	n/a
FU₅₈-6			1378100-25-4		ChemBridge	17104240	n/a
FU₅₈-7			1269257-22-8		ChemBridge	33257899	n/a
FU₅₈-8			1005631-30-0		Vitas-M	STL398417	n/a ^[f]
FU₅₈-9			1355873-93-6		ENAMINE	Z1156178129	90
FU₆₆-1	0.21	160	1355830-29-3		ENAMINE	Z906038686	90
FU₆₆-2			2034580-57-7		Life Chemicals	F6482-6629	90
FU₆₆-3			2034475-57-3		Life Chemicals	F6482-6686	90
FU₆₆-4			2034307-19-0		Life Chemicals	F6482-6819	90
FU₆₆-5			1375192-22-5		ENAMINE	Z908233534	90
FU₆₆-6			1375215-10-3		ENAMINE	Z907996506	90
FU₂₉₀-1	0.44	7.2	10465-18-6	2xHCl	ChemBridge	5111431	90 ^[f]
FU₂₉₀-2	0.40	160	1417634-38-8	1xHBr	Vitas-M	STL301596	n/a

[a] Ligand efficiency in kcal mol⁻¹ atom⁻¹

[b] K_d value determined by ITC

[c] Purity reported by MolPort if not stated otherwise. “n/a”: no purity information available

[d] Determined by ¹H-qNMR

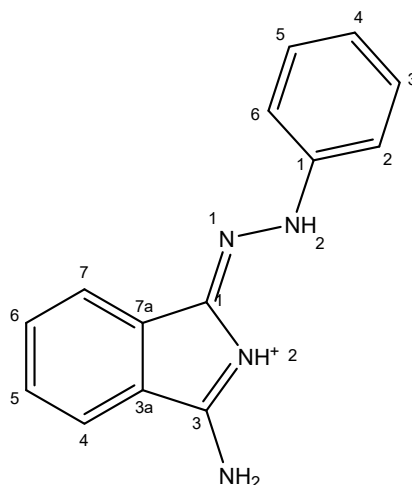
[e] Affinity only determined for resynthesized **FU₅-1** · HCl

[f] Identity verified by NMR provided by vendor

[g] Measurement is not reliable as crystallographic finding indicates only binding of impurity

Synthesis of compound FU₅-1

i.e. (Z)-3-amino-1-(2-phenylhydrazineylidene)-1*H*-isoindol-2-ium chloride



(Z)-3-amino-1-(2-phenylhydrazineylidene)-1*H*-isoindol-2-ium

Chemical Formula: C₁₄H₁₃N₄⁺

Exact Mass: 237.11

Molecular Weight: 237.29

m/z: 237.11 (100.0%), 238.12 (15.1%), 238.11 (1.5%), 239.12 (1.1%)

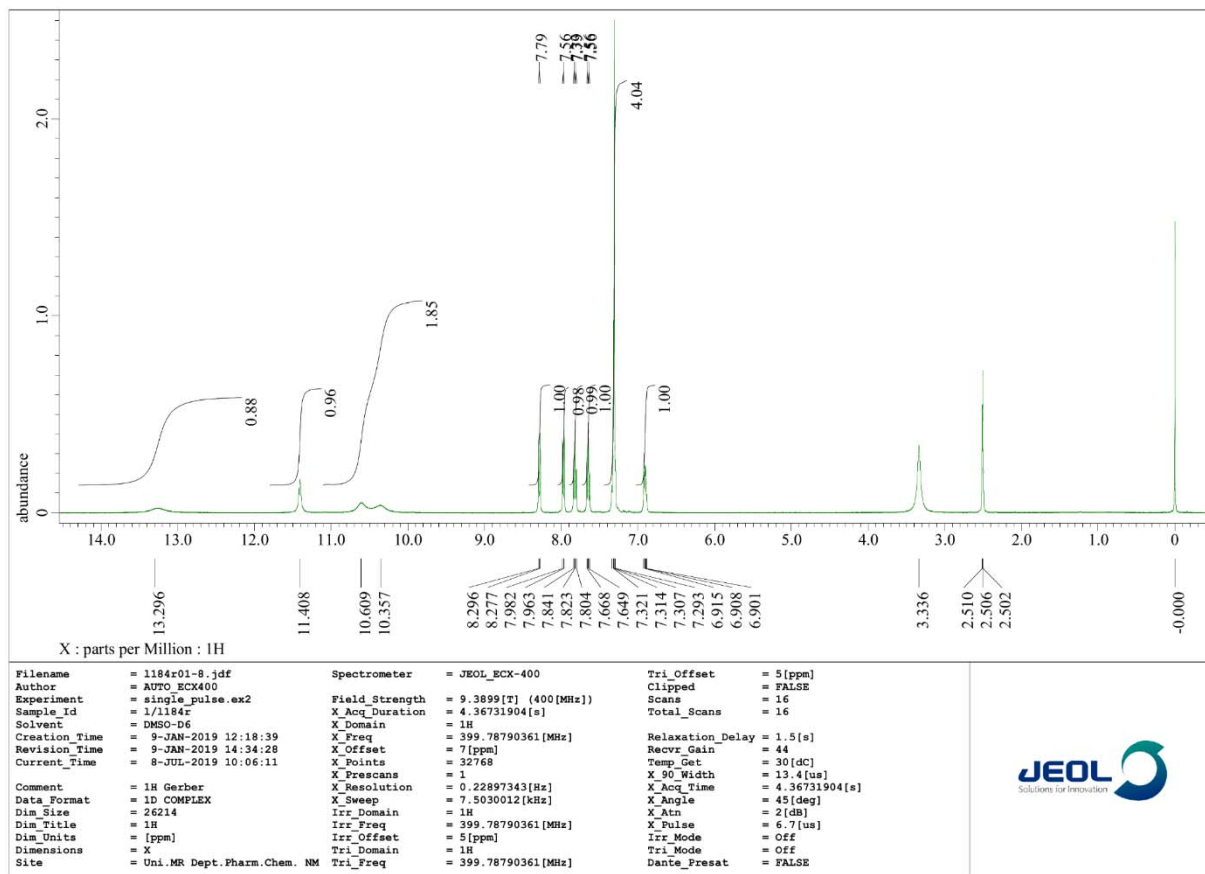
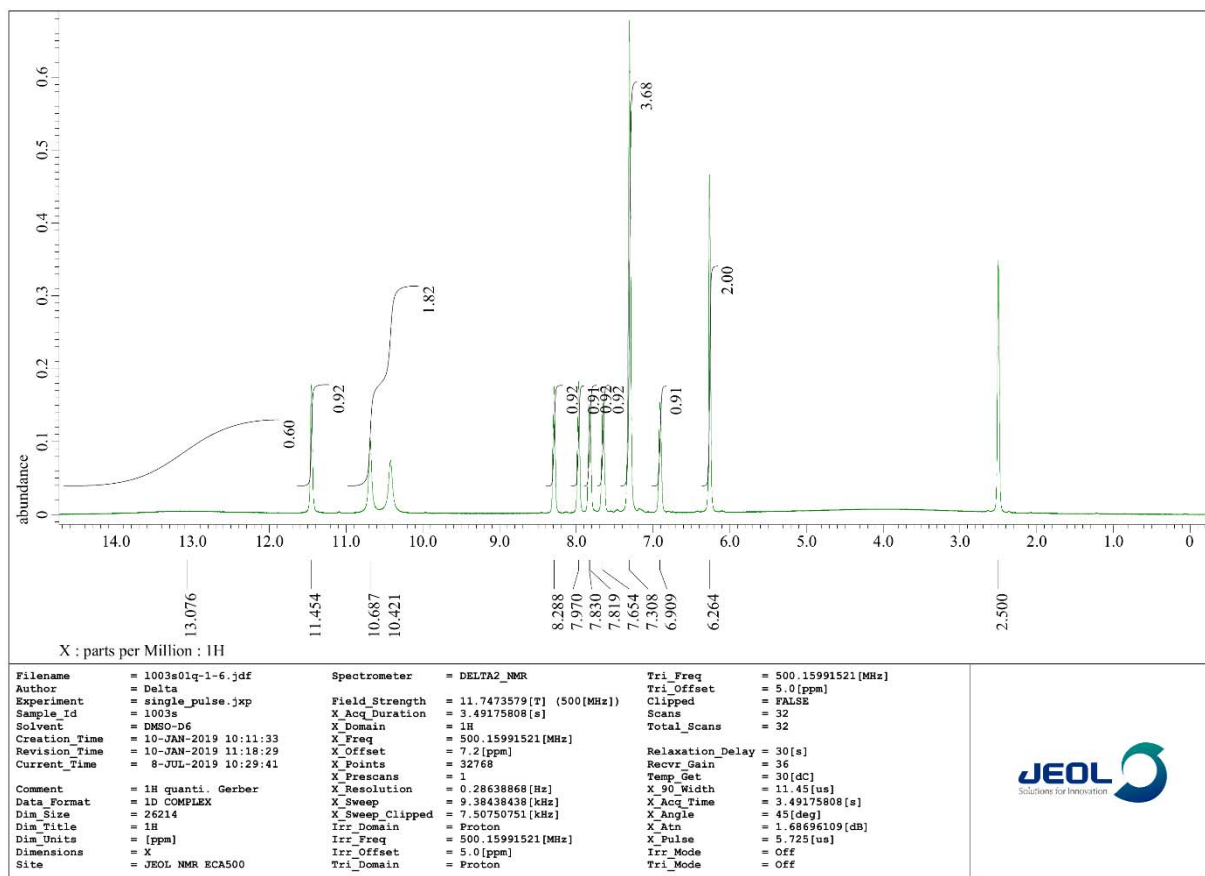
Elemental Analysis: C, 70.87; H, 5.52; N, 23.61

The target compound (FU₅-1 · HCl) was prepared by a slightly modified literature procedure (Biitseva et al. 2013, Wolf, W. & Vollmann, H. 1956) as follows:

A stirred mixture of 1,3-diiminoisoindole (0.73 g, 5.0 mmol) and phenylhydrazine (0.54 g, 5.0 mmol) in MeOH (30 mL) was refluxed for 2 h. After cooling to RT, aqueous 1 M HCl (60 mL, 60 mmol) was added and the mixture stirred for 15 min at RT. The resulting yellow-orange precipitate was filtered through a glass filter (G3), thoroughly washed with water, then with a 1:1 mixture of water/MeOH and finally dried *in vacuo* over P₄O₁₀ giving rise to 0.92 g (67.5 %) of a yellow-orange solid.

m.p. 228 °C (decomp.; (Biitseva *et al.*, 2013): 172 °C); at ca. 175 °C morphism into needles! ¹H NMR (DMSO-*d*₆, 500.1 MHz): δ = 13.30 (bs, 1H), 11.41 (bs, 1H), 10.61 and 10.36 (2 x bs, merged, 2H), 8.29 (d, ³J_{H,H} = 7.8 Hz, 1H), 7.97 (d, ³J_{H,H} = 7.6 Hz, 1H), 7.82 (t, ³J_{H,H} = 7.8 Hz, 1H), 7.65 (t, ³J_{H,H} = 7.6 Hz, 1H), 7.37-7.27 (m, 4 H), 6.95-6.87 (sm, 1H). ¹³C NMR (DMSO-*d*₆, 100.5 MHz): δ = 158.6, 144.0, 134.4, 134.2, 131.2, 129.2, 129.1, 125.7, 124.2, 120.7, 120.5, 112.7. MS (ESI+) *m/z* (%): 237 (93) [M]⁺. HRMS (ESI+): *m/z* calculated for C₁₄H₁₃N₄⁺ 237.1135 [M]⁺; found 237.1139. Elemental analysis calculated (%) for C₁₄H₁₃N₄Cl (272.74): C 61.65, H 4.80, N 20.54, Cl, 13.00; found: C, 61.13; H, 5.04; N, 20.27; Cl, 12.86

Substance assay, determined by ¹H qNMR (DMSO-*d*₆, 500.1 MHz, reference standard maleic acid TraceCert®): 98.5 %

Figure S1. ^1H NMR spectrum of $\text{FU}_5\text{-1} \cdot \text{HCl}$.Figure S2. ^1H qNMR spectrum of $\text{FU}_5\text{-1} \cdot \text{HCl}$.

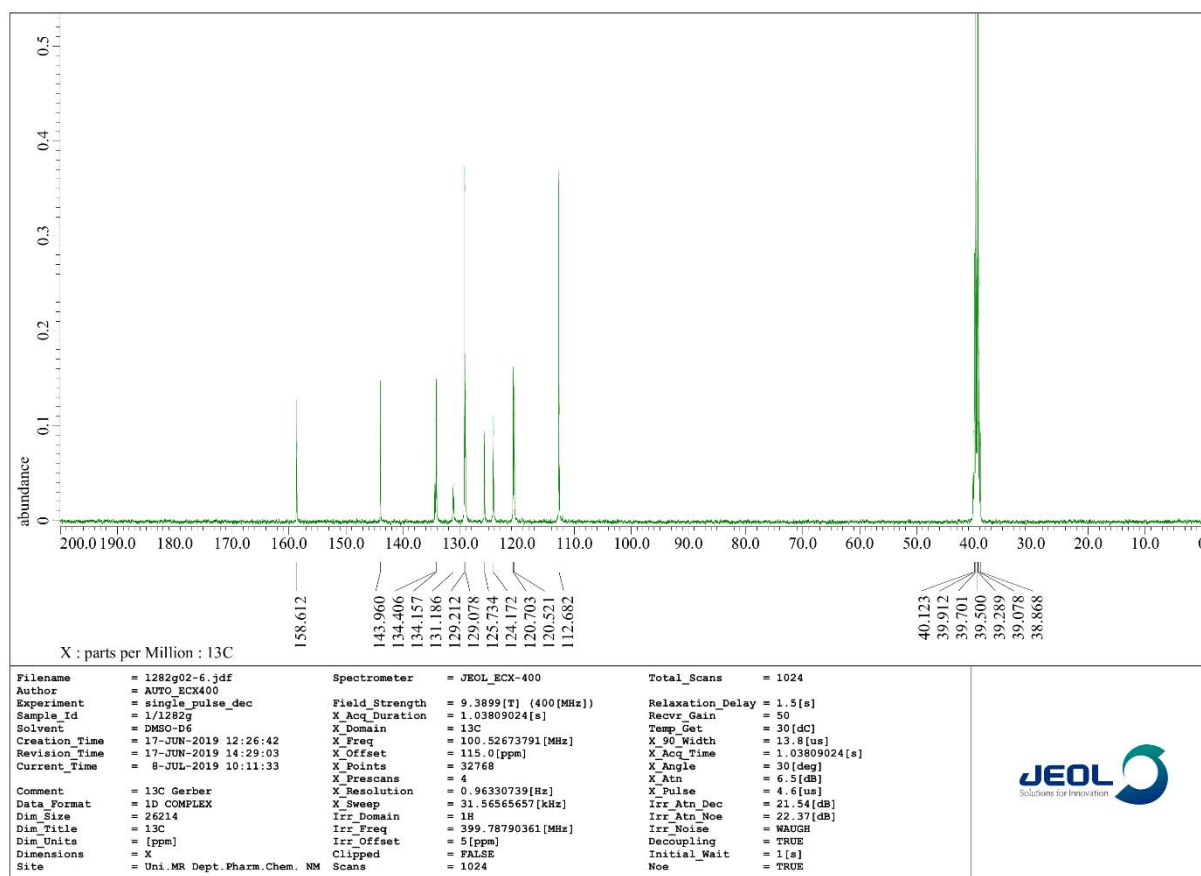


Figure S3. ¹³C NMR spectrum of FU₅-1 · HCl.

Table S2. Details of ITC measurements.

Follow-Up Compound	Titration Type	Protein Concentration in Cell	Follow-Up Ligand Concentration in Cell	Ligand Concentration in Syringe	Final Stoichiometry (Ligand:Protein)
FU ₅ -1 · HCl	direct	25 μM ^[a]	–	500 μM ^[a]	4
FU ₅ -2	displacement	50 μM	2000 μM	500 μM	2
FU ₅ -3	displacement	50 μM	2000 μM	500 μM	2
FU ₅ -4	direct	50 μM	–	2000 μM	8
FU ₅₈ -1	displacement	50 μM	2000 μM	500 μM	2
FU ₅₈ -2 ^[b]	–	–	–	–	–
FU ₅₈ -3	displacement	50 μM	2000 μM	500 μM	2
FU ₆₆ -1	displacement	50 μM	2000 μM	500 μM	2
FU ₂₉₀ -1	direct	50 μM	–	1000 μM	4
FU ₂₉₀ -2	displacement	50 μM	2000 μM	500 μM	2

[a] 0.1% (v/v) Tween 20 in syringe and cell.

[b] Not enough substance available for ITC.

Table S3. Fitting statistics for ITC titration curves

Compound	Global reduced chi-square ^[a]	Weighted local rmsd ^[a]	Stoichiometry ^[b] (B:A)	n ^[c]	K_d [μ M] (SD) ^[d]
FU₅-1 ^[e]	0.026	28.6	0.875	25	6.4 (\pm 0.06)
FU₅-1 ^[e]	0.024	38.7	0.915	25	6.4 (\pm 0.06)
FU₅-1 ^[e]	0.003	37.7	0.858	25	6.4 (\pm 0.06)
FU₅-2	0.228	135.3	0.991	25	970 (\pm 79)
FU₅-2	0.357	268.2	0.972	25	970 (\pm 79)
FU₅-2	0.284	238.2	1.018	25	970 (\pm 79)
FU₅-3	0.107	240.5	1.052	13	740
FU₅-4	0.014	18.6	1.000 ^[g]	38	400
FU₅₈-1	0.134	197.5	1.081	24	450
FU₅₈-3	0.498	396.7	1.068	24	1040
FU₆₆-1	0.052	226.0	1.142	24	160
FU₂₉₀-1	0.075	186.9	0.816	74	7.2
FU₂₉₀-2	0.065	41.4	0.871	25	160
Reference (SAP114)	0.346	193.2	1.001	24	680

Isotherm fitting statistics of ITC data, analyzed via SEDPHAT 10.58d (Houtman *et al.*, 2007). [a] Goodness of fit parameters (global reduced chi-square and weighted local rmsd) determined with SEDPHAT 10.58d are comparable to those listed for the established reference compound SAP114.

[b] Stoichiometry of the ligand protein A and the ligand B at the inflection point of the sigmoidal fitting curve

[c] Number of used data points corresponding to injections of the ITC experiment

[d] K_d values. Standard deviation (SD) is only listed for **FU₅-1** and **FU₅-2**, which were measured as triplicates.

[e] For **FU₅-1**, global reduced chi-square and weighted local rmsd are listed for comparison only and the listed K_d value was determined by global fit with the AFFINImeter suite (version 2.1710; Muñoz & Piñeiro, 2018).

[f] For **FU₅-4** the stoichiometry was fixed to a value of 1 to achieve a satisfactory fit

ΔH and $T\Delta S$ values are not reported for this experiment following the reasoning of Krimmer and Klebe (2015).

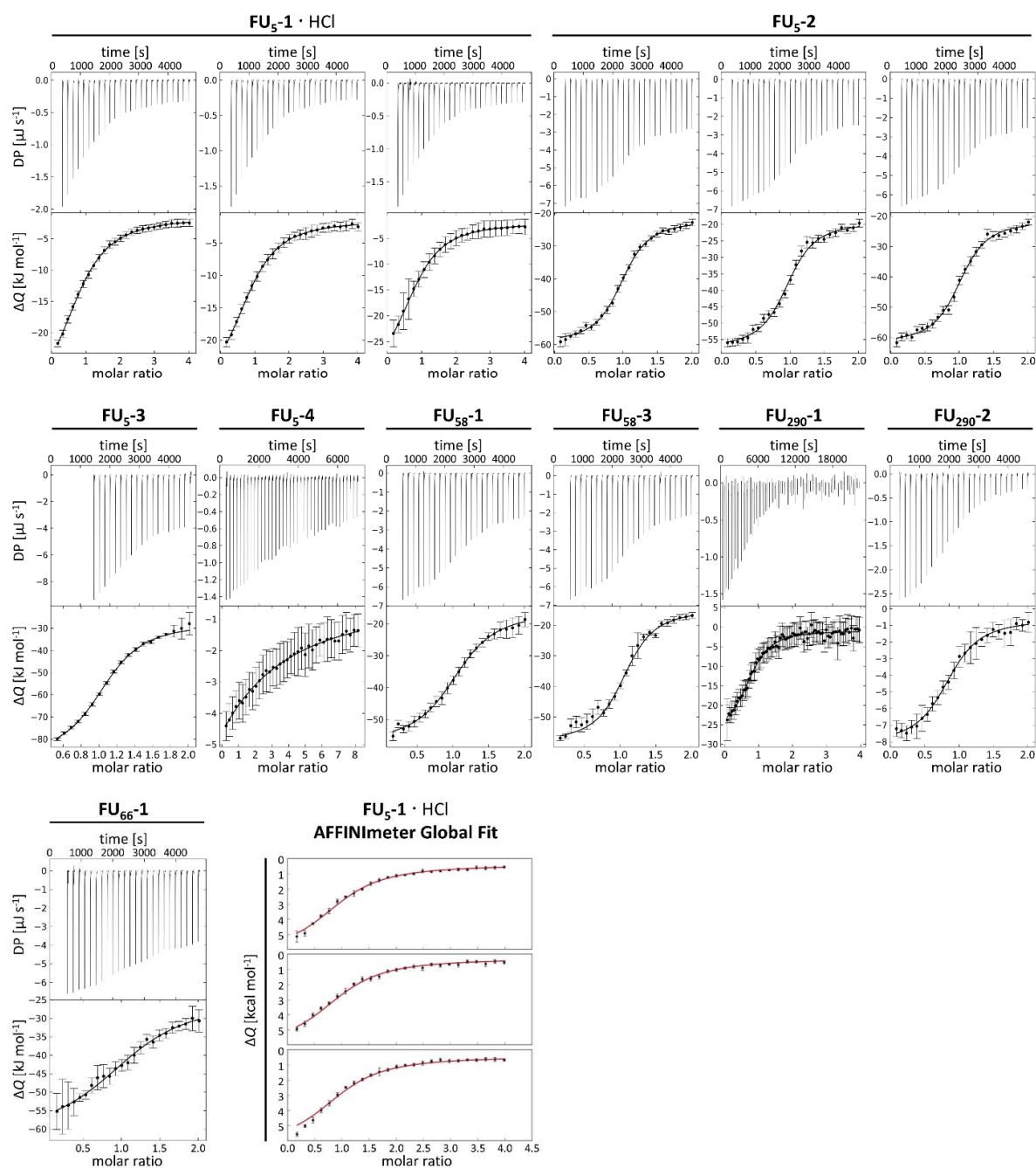


Figure S4. ITC thermograms for crystallographically identified follow-up ligands. In all cases, the heat signals DP (in $\mu\text{J s}^{-1}$) as a response by the release of the ligand into the protein solution is shown over the course of the experiment along with the integrated heat signals of the injections ΔQ (in kJ mol^{-1}). **FU₅-1 · HCl** and **FU₅-2** were measured in triplicate, all other ligands as single measurements ($n = 1$), in due regard of the limited resilience of the results of the latter. The noisy baseline of the ITC experiments for **FU₂₉₀-1** was allegedly due to its low purity ($> 90\%$ according to provider) and could not be further improved.

Mass Spectrometry Analysis of FU₅₈-3

Clear difference electron density and steric constraints in the crystal structure obtained by soaking **FU₅₈-3** showed that **FU₅₈-3b** had bound instead. The presence of **FU₅₈-3b** in the commercially obtained sample of **FU₅₈-3** (> 90% according to provider (ChemBridge, USA)) was demonstrated by mass spectrometry (MS) experiments (Figure S6) using a Q-Trap 2000 (Applied Biosystems, USA).

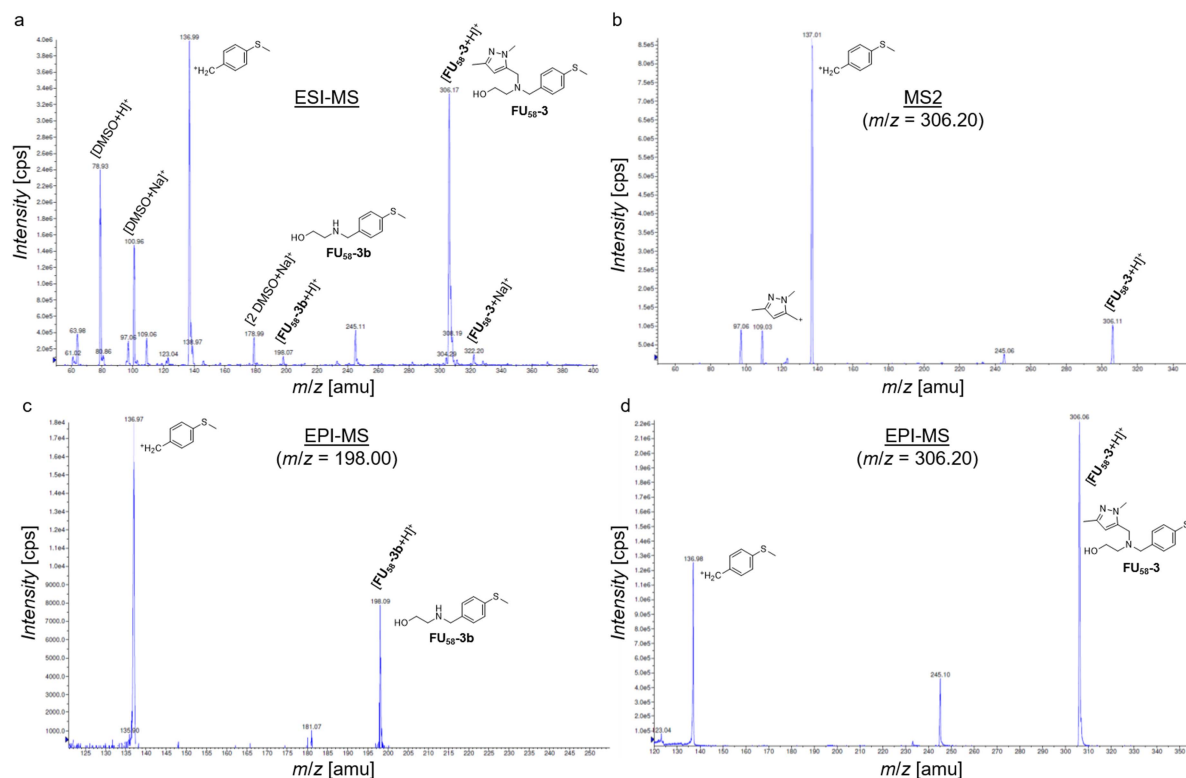


Figure S5. Mass spectrometry results of commercially obtained sample of **FU₅₈-3**. a) ESI-MS. b) MS2. c) EPI-MS targeting $m/z = 198.00$. c) EPI-MS targeting $m/z = 306.20$.

In detail, **FU₅₈-3** was dissolved in methanol from 1 M DMSO stock and characterized by electrospray ionization (ESI) MS (Figure S6a). This demonstrated the presence of **FU₅₈-3** (calcd. for $C_{16}H_{24}N_3OS$ $[M+H]^+$ 306.16, found $m/z = 306.17$) and its substructure **FU₅₈-3b** (2-((4-(methylthio)benzyl)amino)ethan-1-ol) (calcd. for $C_{10}H_{16}NOS$ $[M+H]^+$ 198.09, found $m/z = 198.07$). In addition, MS2 targeting **FU₅₈-3** ($m/z = 306.20$; Figure S6b) demonstrated its presence ($m/z = 306.11$) as well as the presence of its fragmentation products (4-(methylthio)phenyl)methyl cation (calcd. for C_8H_9S $[M]^+$ 137.04, found $m/z = 137.01$) and (1,3-dimethyl-1H-pyrazol-5-yl)methyl cation (calcd. for $C_6H_9N_2$ $[M]^+$ 109.08, found $m/z = 109.03$). However, **FU₅₈-3b** was not observed in this MS2 experiment. Finally, **FU₅₈-3b** was clearly observed in enhanced product ion (EPI) MS targeting $m/z = 306.20$ (Figure S6c) but not in the EPI MS targeting $m/z = 198.00$ (Figure S6d). This indicates **FU₅₈-3b** is not a decomposition fragment of **FU₅₈-3** but instead an impurity of the obtained substance. As a side note, 4-(methylthio)benzylamine (calcd. for $C_8H_{12}NS$ $[M+H]^+$ 154.07), which could partially have matched the observed electron density, was not observed in any of the MS experiments. Thus, only **FU₅₈-3b** was unambiguously attributed to the electron density and built into the crystal structure obtained by soaking **FU₅₈-3** (PDB ID: 5SAQ).

Table S4. Retrospective, unbiased docking of follow-up compounds.

Starting Fragment	Follow-Up	fragment core at similar position as starting fragment?	# of FlexX poses	successfully HYDE-scored poses	HYDE-scored poses without bad clashes/ torsions	RMSD to crystallographic pose < 2 Å		
						total	in top 10 scored	in top3 scored
F005	FU₅-1	yes	500	405	403	59	0	0
	FU₅-2	yes	500	450	442	68	0	0
	FU₅-3	yes	467	413	412	71	0	0
	FU₅-4	no	448	381	378	0	0	0
F058	FU₅₈-1	no	235	125	113	0	0	0
	FU₅₈-2	no	403	319	287	0	0	0
F066	FU₆₆-1	no	333	202	127	0	0	0
F290	FU₂₉₀-1	yes	386	343	236	0	0	0
	FU₂₉₀-2	yes	500	440	398	25	3	0

Supplementary references

Biitseva, A., Groth, U. & Hordiyenko, O. (2013). *J. Heterocyclic Chem.* **50**, 1140-1145

Wolf, W. & Vollmann, H. (1956). *German Patent* DE 941845

Krimmer, S. G. & Klebe, G. (2015). *J. Comput. Mol. Des.* **29**, 867–883.

Workflow and Tools for Crystallographic Fragment Screening at the Helmholtz-Zentrum Berlin

Jan Wollenhaupt¹, Tatjana Barthel^{1,2}, Gustavo M. A. Lima³, Alexander Metz⁴, Dirk Wallacher⁵, Elmir Jagudin³, Franziska U. Huschmann^{1,4}, Thomas Hauß¹, Christian G. Feiler¹, Martin Gerlach¹, Michael Hellmig¹, Ronald Förster¹, Michael Steffien¹, Andreas Heine⁴, Gerhard Klebe⁴, Uwe Mueller¹, Manfred S. Weiss¹

¹Macromolecular Crystallography, Helmholtz-Zentrum Berlin ²Structural Biochemistry Group, Institute for Chemistry and Biochemistry, Freie Universität Berlin ³BioMAX, MAX IV Laboratory ⁴Drug Design Group, Institute of Pharmaceutical Chemistry, Philipps-Universität Marburg ⁵Department Sample Environment, Helmholtz-Zentrum Berlin

Corresponding Author

Manfred S. Weiss

msweiss@helmholtz-berlin.de

Citation

Wollenhaupt, J., Barthel, T., Lima, G.M.A., Metz, A., Wallacher, D., Jagudin, E., Huschmann, F.U., Hauß, T., Feiler, C.G., Gerlach, M., Hellmig, M., Förster, R., Steffien, M., Heine, A., Klebe, G., Mueller, U., Weiss, M.S. Workflow and Tools for Crystallographic Fragment Screening at the Helmholtz-Zentrum Berlin. *J. Vis. Exp.* (169), e62208, doi:10.3791/62208 (2021).

Date Published

March 3, 2021

DOI

10.3791/62208

URL

jove.com/video/62208

Abstract

Fragment screening is a technique that helps to identify promising starting points for ligand design. Given that crystals of the target protein are available and display reproducibly high-resolution X-ray diffraction properties, crystallography is among the most preferred methods for fragment screening because of its sensitivity. Additionally, it is the only method providing detailed 3D information of the binding mode of the fragment, which is vital for subsequent rational compound evolution. The routine use of the method depends on the availability of suitable fragment libraries, dedicated means to handle large numbers of samples, state-of-the-art synchrotron beamlines for fast diffraction measurements and largely automated solutions for the analysis of the results.

Here, the complete practical workflow and the included tools on how to conduct crystallographic fragment screening (CFS) at the Helmholtz-Zentrum Berlin (HZB) are presented. Preceding this workflow, crystal soaking conditions as well as data collection strategies are optimized for reproducible crystallographic experiments. Then, typically in a one to two-day procedure, a 96-membered CFS-focused library provided as dried ready-to-use plates is employed to soak 192 crystals, which are then flash-cooled individually. The final diffraction experiments can be performed within one day at the robot-mounting supported beamlines BL14.1 and BL14.2 at the BESSY II electron storage ring operated by the HZB in Berlin-Adlershof (Germany). Processing of the crystallographic data, refinement of the protein structures, and hit identification is fast and largely automated using specialized software pipelines on dedicated servers, requiring little user input.

Using the CFS workflow at the HZB enables routine screening experiments. It increases the chances for successful identification of fragment hits as starting points to develop more potent binders, useful for pharmacological or biochemical applications.

Introduction

The first step in drug development is the screening of compounds against a target of interest. Traditionally, large compound libraries in the order of 100,000-1,000,000 entries are used in high-throughput biochemical assays in the pharmaceutical industry. This strategy was complemented by fragment-based drug design (FBDD), a newer method that took a steep rise during the last 20 years and became a mainstream strategy to generate high-quality lead candidates due to several inherent advantages of the method¹. The term "fragment" refers to a small organic molecule containing typically less than 20 non-hydrogen or heavy atoms (HAs). Thus, a fragment is significantly smaller than the drug- or lead-like molecules (usually less than 30 HAs) explored in conventional high-throughput screening. Fragments are weak-affinity binders. However, compared to larger molecules, fragments are more versatile, since even a small collection of them can better represent the respective chemical space of molecules of the same size². Also, evolving fragment screening hits into lead molecules is considerably more effective than optimizing already larger molecules^{2,3,4,5}. That means, pending sufficient sensitivity of the detection, screening of fragments can be employed efficiently and yields high-quality starting points for further compound evolution. Several biophysical methods may be applied for fragment screening, the most popular being nuclear magnetic resonance, X-ray crystallography, surface plasmon resonance and thermal shift assays. These methods are used either in a parallel or in a sequential way, with the aim to increase the confidence in the hits and reduce the numbers

of false positives or false negatives, respectively. However, a recently conducted comparative study⁶ suggested that sequential screening cascades are to be avoided due to the low overlap between the different methods.

X-ray crystallography is a well established method for structure determination at atomic detail but has recently also been developed as a tool for screening purposes^{7,8}. As protein crystals tolerate high fragment concentrations (e.g., 100 mM), crystallographic fragment screening (CFS) can compete with other biophysical methods for screening fragments or even outperform them as a first-step screening method^{6,9}. However, a vital pre-requisite for CFS is a validated crystallization system of the target protein reproducibly delivering crystals with diffraction properties to considerably high resolution, typically better than 2 Å.

An exclusive benefit of CFS compared to all other fragment screening methodologies is the provision of detailed 3D information about the binding mode of the identified fragments. This structural information is absolutely crucial for the rational optimization of the fragment hits to higher-affinity binders. Established elaboration strategies are growing, merging, and linking fragment hits⁵. Thereby relatively high ligand efficiency is provided from the start, and the introduction of unnecessary or spatially not suitable groups can be avoided, thus reducing chemical synthesis costs. All in all, CFS has unrivaled advantages as a starting strategy for drug design.

Given that a particular biological target meets the high requirements of CFS regarding crystal quality, there are some main factors that maximize the chances for a successful outcome of such screening campaigns. It depends on the quality of the fragment library used, on an efficient workflow to carry out the experiments before the diffraction experiment, on synchrotron beamlines with sufficient automation and data collection speed, as well as on ways and means for largely automated data processing and analysis. Here, the complete workflow from the crystal soaking experiments to the hit identification is presented, in the way it is successfully established at the macromolecular crystallography beamlines at BESSY II (**Figure 1**). The facility is open to academic and industrial users for collaboration. Additionally, academic users of EU-countries outside Germany can straightforwardly apply for funding via the iNEXT Discovery project.

There are indispensable prerequisites to be able to start a CFS campaign and conduct the protocol outlined in this work: well-diffracting crystals of the target protein are available that can be reproducibly grown in large numbers, that are stable at ambient temperature, and that were grown using a crystallization cocktail without highly volatile ingredients. Another prerequisite is the suitability of the crystal lattice for the experiment. In an appropriate lattice, the interesting sites of the target protein must be exposed towards the solvent channels and thus accessible. Another preceding step that is optional but nevertheless highly recommended to ensure success in the workflow of the CFS campaign is the optimization of the soaking condition for the experiment. Vital benchmark statistics here are the diffraction power of the crystal and the relevant data quality indicators, which are determined during the data scaling procedure. Typical factors to optimize are DMSO-tolerance, buffer concentration and cryo-protectant. Although not a strict prerequisite as further

detailed below, DMSO as a co-solvent can help to increase fragment solubilization. Typical tests should include soaking of 0, 3, 6, or 10% (v/v) DMSO overnight. An increase of the buffer concentration to 200 or 300 mM helps to prevent loss in diffraction quality due to occasional pH-shifting effects arising from the high fragment concentrations to be used. Finally, it is decisive to find out whether and which additional cryoprotectant is required and if it can be already included in the soaking condition. In many cases, however, an additional cryo-protectant is not needed, because DMSO itself can act as a cryo-protectant. If so, this will save one handling step in the final experiment. Most crystals need less cryo-protectant if flash-cooled on appropriately sized loops, minimizing or avoiding surrounding mother liquor as much as possible. However, in rare cases, a layer of the mother liquor is indeed necessary to prevent damage to the crystal upon flash cooling.

The number of hits obtained in a CFS campaign is not only dependent on the druggability of the target protein and the suitability of the crystal lattice (see above), but it is also dependent on the quality of the library. Library quality comprises two aspects: the selection of the compounds for the library and the confectioning of the compounds, (i.e., in which physical form they are presented for the experiment). For compound selection different strategies can be employed. Most library designs include the maximization of the chemical diversity of the fragments. A strategic focus could be to include the chemical tractability of the fragments for follow-up design, which has been applied for instance in the Diamond-SGC-iNEXT poised library¹⁰. Yet another strategic focus for library design could be to maximize the representation of commercially available chemical space of fragments by shape- and pharmacophore-based clustering, as has been exemplified by the F2X libraries developed at HZB¹¹.

More specifically, the 1103-membered F2X-Universal Library and representative 96-compound subset for initial CFS campaigns, which is called F2X-Entry Screen, have been developed and the F2X-Entry Screen has been validated successfully¹¹. The F2X-Entry Screen is the primary choice for CFS campaigns at HZB. Subsequently, larger campaigns can then be carried out using the F2X-Universal Library or the 1056-membered EU-OPENSOURCE fragment library¹² that is also being offered at HZB. At present, these libraries are available for users of the macromolecular crystallography beamlines of the BESSY II synchrotron in Berlin free-of-charge on the basis of a collaboration contract. That also applies to users via iNEXT Discovery proposals. Moreover, the F2X-Entry Screen is available to all interested scientists on the basis of a material transfer agreement.

With respect to the physical presentation of a library, two approaches are commonly adopted: the fragments are either used as DMSO stock solutions or the fragments are dried and immobilized on ready-to-use plates. At HZB, both the F2X-Entry Screen and the non-volatile compounds of the F2X-Universal Library are presented as dried-on compounds in a 3-lens 96-well MRC low profile crystallization plate. The presentation of the fragments immobilized in crystallization plates has two vital advantages: Firstly, it allows transport of the screening plates to the user's home lab. Therefore, the soaking and crystal handling steps of the workflow presented here (steps 1-3) can be carried out anywhere. Secondly, DMSO-free solution can be employed. DMSO-sensitive targets can thus be screened easily, largely retaining expected hit rates¹¹. However, DMSO does increase fragment solubility, hence it is worthwhile to check the DMSO tolerance of a crystal system of choice beforehand as outlined above.

The protocol outlined below will describe a typical experiment with a 96-compound screen such as the F2X-Entry Screen. For that, approximately 250 crystals need to be prepared in time to be used freshly. It is highly advisable to prepare the soaks for all 96 compounds in duplicate. It is recommended, but optional, to prepare additional mock-soaks that will later help with data analysis using the pan-data density analysis (PanDDA) approach for hit identification¹³. Mock-soaks are defined as soaking experiments on protein crystals using the same soaking solution as the fragment soaks for the same incubation time, but no fragments are present. If the soaking solution is equal to the crystallization condition, the crystals may be directly harvested from the crystallization plate.

Dependent on the capabilities of the robotic sample changer, different puck formats may have to be used. At the moment, samples for the HZB-operated beamline BL14.1 need to be prepared in Unipuck format, samples for the HZB-operated beamline BL14.2 need to be prepared in SPINE puck format. In this protocol, preparation in Unipuck format is assumed.

Protocol

1. Soaking crystals

1. Take the screening plate (here, an F2X-Entry Screen plate, **Figure 2**) from the -20 °C freezer and place it onto the bench/table for about 30 min to pre-warm it to room temperature, thus avoiding condensation moisture.
2. Arrange the working place with two closely arranged microscopes and all tools needed (**Figure 3A**). The materials are listed in the **Table of Materials**.
3. Choose 3-4 loops of the appropriate size for transfer of the crystals to be soaked and place them close to the microscopes.

4. Fill the glass spot plate cavities with de-ionized or distilled water.
5. Prepare 5 mL of soaking solution.
6. Cut open the bag of the screening plate pre-warmed to room temperature.
7. Remove the lid and the foil from the screening plate, while keeping the plate placed on the bench/table.
8. Decant the 5 mL of soaking solution in the reagent reservoir.
9. Fill each of the 96 reservoirs with 40 μ L of soaking solution using the 12-channel pipette.
10. Place the EasyAccess Frame on top of the screening plate and secure it with the included clamps by sliding them onto the left and right side of the device.

NOTE: The EasyAccess Frame is a special device for handling multiple crystals, which was developed at the HZB¹⁴. It enables easy access to each well by shifting the movable tiles while protecting the other wells from evaporation.

11. Place the screening plate (incl. the EasyAccess Frame) under the first microscope and the crystallization plate including the crystals to be soaked under the second microscope.
 12. Slide open well A1 of the screening plate by moving the respective acrylic glass tile of the EasyAccess Frame either with a finger or the supplied pen tool.
 13. Add 0.4 μ L of soaking solution from the reservoir to the fragment containing well (upper left lens) using a fresh pipette tip. Check through the microscope that the drop covers the dried-on fragment, so it can dissolve.
- NOTE:** Alternatively, this step can be carried out using a pipetting robot before the assembly of the EasyAccess

Frame. This way the soaking drops of all wells could be placed in one automatic procedure. However, the authors recommend adding the soaking solution directly before the soaking step as described to ensure that the fragment solubilizes slowly and in the presence of the crystal. This avoids that the crystal experiences a sudden shock upon transfer to a drop with a high fragment concentration.

14. Under the second microscope, cut open the sealing foil of the crystallization plate at one of the wells that contains the target crystals.
15. Transfer two crystals using an appropriately sized loop mounted on the crystal wand to the well A1 of the screening plate under the first microscope.
16. Wash the loop in the prepared glass spot plate and dry it by gently touching the tissue. Do this after every transfer to avoid cross contamination with fragment containing soaking solutions.
17. Use the microscope to check that the crystals have been properly placed.
18. Move on to the next well (e.g., B1).
19. Repeat steps 1.13-1.18 with all 96 wells of the screening plate until each soaking drop contains two crystals.
20. Remove the screening plate (incl. the EasyAccess Frame) from under the microscope and place it onto the bench/table.
21. Remove the EasyAccess Frame from the screening plate.
22. Seal the screening plate with sealing foil and place it in the crystallization incubator or cupboard, respectively, where the crystals were grown.
23. Incubate for the optimized soaking time. Overnight is usually convenient.

24. (optional) Preparation of approximately 40 apo crystals (i.e., mock soaking)
 1. Take an MRC 3-lens 96-well low-profile crystallization plate and fill two columns with 40 μL of soaking solution per well using the 12-channel pipette.
 2. Place the EasyAccess Frame on top of the crystallization plate and secure it with the included clamps by sliding them onto the left and right side of the device.
 3. Slide open the acrylic glass tile of well A1.
 4. Place 0.4 μL of soaking solution in each of the two left lenses of the well.
 5. Transfer 2-3 crystals to each drop. After each transfer, wash the loop in the prepared glass spot plate and dry by gently touching the tissue.
 6. Move to the next well (e.g., B1).
 7. Repeat steps 1.24.4-1.24.6 until about 40 crystals are ready for incubation.
 8. Remove the crystallization plate (incl. EasyAccess Frame) from under the microscope onto the bench/table and remove the EasyAccess Frame.
 9. Seal the crystallization plate with sealing foil and place it in the aforementioned crystallization incubator or cupboard.
 10. Incubate for the same time as the screening plate.
2. Arrange the working place with one microscope and all tools needed (**Figure 3B**). The materials are listed in the Table of Materials.
3. Prepare a Unipuck foam dewar with 3 Unipuck lids (i.e., sample enclosures) and fill it with liquid nitrogen (LN_2).

NOTE: Observe the appropriate safety precautions for working with LN_2 (i.e., wear safety goggles and use suitable protective equipment). It is best to get fresh LN_2 several times during the session in order to avoid water condensation in the LN_2 storage can. Through the entire following procedure, make sure the LN_2 level in the foam dewar is always reaching the upper edge of the dewar. Also ensure that the LN_2 is ice-free; frequently replace the LN_2 (e.g., once every 45 min), or latest if ice starts to accumulate. Then, fill the second foam dewar and transfer Unipucks to it. Empty the icy foam dewar and remove residual ice and moisture with the blow dryer.
4. Remove the foil from the screening plate and place the EasyAccess Frame on top.
5. Slide open well A1.
6. Harvest two crystals from the drop and flash-cool them in LN_2 (one by one) by plunging with a fast vertical movement into the LN_2 and then inserting the sample in the proper puck position. Take relevant notes on the sample tracking sheet.
7. Cryoprotection step (if necessary for the target crystals). In such case, perform this step instead of 2.6.
 1. Place 0.4 μL of soaking solution including cryoprotectant on the lower left lens of the well.
 2. Pull the loop with a crystal mounted from the drop in the upper left lens slowly through the solution in the lower left lens while keeping the crystal in the loop,

2. Harvesting crystals

1. Take out the incubating plate(s) from the incubator or cupboard, respectively.

and then flash-cool in LN₂. Harvest two crystals in this way.

NOTE: In steps 2.6 and 2.7, make sure that the time the crystal is in the loop and exposed to air is kept very short. The plunging (i.e., the vertical drop of the sample in the LN₂-filled dewar) should be performed as fast as possible. This ensures high sample quality and prevention of ice rings in the data. Track the samples (i.e., note if crystals have damages, etc.) to prioritize either duplicate for the following X-ray measurements, use the template for that. Even if crystals have cracks, “hairs” or other defects due to the soaking, they can still be used and should always be harvested. In case crystals broke into several pieces, two of the biggest/best looking pieces should be harvested. **Figure 5** shows some examples of how such crystals can look like. All the shown crystals gave still useful datasets in the respective campaign¹¹, underlining that it is worth to harvest crystals after soaking treatment, even if substantial morphologic changes occurred.

8. Go to the next well and repeat steps 2.5 - 2.6./2.7 until all three pucks are filled.
9. Add the Unipuck bases on top of the lids after pre-cooling them in LN₂.
10. Store the Unipucks in storage racks in a transport dewar or storage dewar.
11. Repeat the preceding steps until all the wells of the screening plate have been processed.
12. (optional) If mock-soaked crystals were prepared, harvest them in a similar fashion as described beforehand.

NOTE: If two crystals for each of the 96 conditions of the screening plate could be flash-cooled, there will be space for 32 mock-soaked apo crystals, to fill up the 14 Unipucks.

13. Store the Unipucks in LN₂ until the measurement.

3.Data collection

1. Transfer the Unipucks to beamline BL14.1. If SPINE pucks have been used in step 2, transfer them to beamline BL14.2.
2. Carry out standard measurements on the beamline using the specific recommendations given below. Details about the facility and the experiment control program MXCuBE2 have been presented previously^{15, 16}. **Figure 4** shows the interior of the experimental hutches of beamlines BL14.1 and BL14.2 as well as an example screenshot of the MXCuBE2 control software at beamline BL14.1.
 1. To maximize time efficiency and throughput, skip the collection of test images. The sample-to-detector distance will be fixed to a value that is suitable for the upper resolution limit of the crystal system determined in earlier experiments. If the data collection strategy was not optimized beforehand, collect 1800 images of 0.2 degrees each with an exposure time of 0.1 s per image.
 2. Ideally, test the data collection strategy in prior experiments using mock-soaked apo crystals. For higher symmetry space groups, 1200 images or even 900 images (i.e., 240° or 180°, respectively) will already give complete datasets with good statistics, independent of the starting angle of data collection.

NOTE: Higher redundancy and finer slicing can yield superior data quality¹⁷. However, using this “enough but not more” strategy proposed here is an excellent trade-off between quality, data collection time, as well as computational requirements for analysis later on. In the described way, 200 data collections in 24 hours are well possible at beamlines BL14.1 and BL14.2. Nevertheless, samples should be prioritized.

3. First collect diffraction datasets for one sample per fragment condition, based on the prioritization in step 2.6./2.7 (i.e., collect the data for the higher prioritized duplicate).
4. For those experiments in 3.2.3 where data collection failed, diffraction was lost or severe ice rings occurred, collect data for the second duplicate sample of the respective fragment condition.
5. Collect diffraction datasets of apo crystals (if prepared according to steps 1.24 and 2.12).
6. Collect diffraction datasets of the remaining duplicates of each fragment condition.
7. In the MXCuBE2 program, match the dataset identifiers of a CFS campaign to the following pattern: <protein>-<library>-[ABCDEFGH][01][0123456789][ab] (e.g., MyProtein-F2XEntry-B05a, where “B05” stands for the well (i.e., the fragment condition in the screen) and the following “a” for the first duplicate.)

4. Data treatment

1. For data analysis of the CFS campaign, use FragMAXapp, a web-based solution to control a multiplex analysis for processing auto-refinement and PanDDA

hit evaluation of CFS data¹⁸ (Lima et al. FragMAXapp, unpublished data). In the FragMAXapp version deployed at HZB the following programs/pipelines are available: XDSAPP¹⁹, Xia2-DIALS and Xia2-XDS²⁰, fspipeline⁷, DIMPLE²¹, Phenix LigFit²², PanDDA^{13,23}. Use a well refined input model of the target protein as input for automatic refinement; otherwise perform meticulous refinement of one high resolution mock-soaked crystal that was collected during the campaign.

NOTE: A key element for hit identification is PanDDA. Details are explained in the respective publications^{13,23}. In brief, PanDDA automatically calculates electron density maps of a set of data sets in a CFS campaign. These are then assumed as non-binding fragment conditions and averaged to generate the so-called ground state model. The ground state model is then used to derive local discrepancies between each electron density map and the ground state map, using voxel-associated Z-scores. Then, for areas of high Z-scores a so called PanDDA-map is created by fine-tuned subtraction of ground state density from the respective map. This largely enhances the visibility of fragment binding events.

2. To maximize the outcome of PanDDA, use a two-step approach. Firstly, performing a PanDDA run (pandda.analyse) with standard settings. Even if mock-soaked crystals have been collected, their identity will not be included as a parameter (which is possible nonetheless) in order to enable an unbiased generation of the ground state model by PanDDA from all available data. Afterwards, the output data is evaluated by the user via a so-called PanDDA inspection in Coot²⁴. Here, hits with relatively high confidence should be noted, concluding the first step.

3. Secondly, re-run the `pandda.analyse` step excluding the preliminary hits (determined in the first step) from the ground state model via the `--exclude_from_characterisation=<list-of-bound-dataset-ids>` command line option. Further details are described on the PanDDA help pages (<https://pandda.bitbucket.io/>). This way, datasets that are clear hits and thus would obscure the ground state model if included are disregarded. This leads to an improved ground state model and thus to improved results overall. Finally, a thorough PanDDA inspection is performed to complete the hit identification.

NOTE: FragMAXapp includes also an output option to save the modeled bound states or prepare data for PDB submission, for further detail see FragMAX webpages (<https://fragmax.github.io/>).

Representative Results

As part of the previously reported validation campaigns of the F2X-Entry Screen¹¹, three campaigns were conducted at the BioMAX beamline at MAX IV and one campaign was conducted at beamline BL14.1 at HZB. In the latter campaign, a particular set of F2X-Entry Screen conditions using a soaking condition that did not contain DMSO was screened against the protein-protein complex of yeast Aar2 and the RNaseH-like domain of yeast Prp8 (AR). The selected set of conditions comprises the hits that were found in an earlier campaign of the F2X-Entry Screen against AR in a soaking condition containing DMSO¹¹, (i.e., in the campaign performed at HZB those hits were re-screened in the absence of DMSO). **Figure 7** shows an overview of the hits obtained after analyzing the data with the FragMAXapp combination of XDSAPP for processing, fspipeline for auto-refinement and subsequent hit finding using PanDDA.

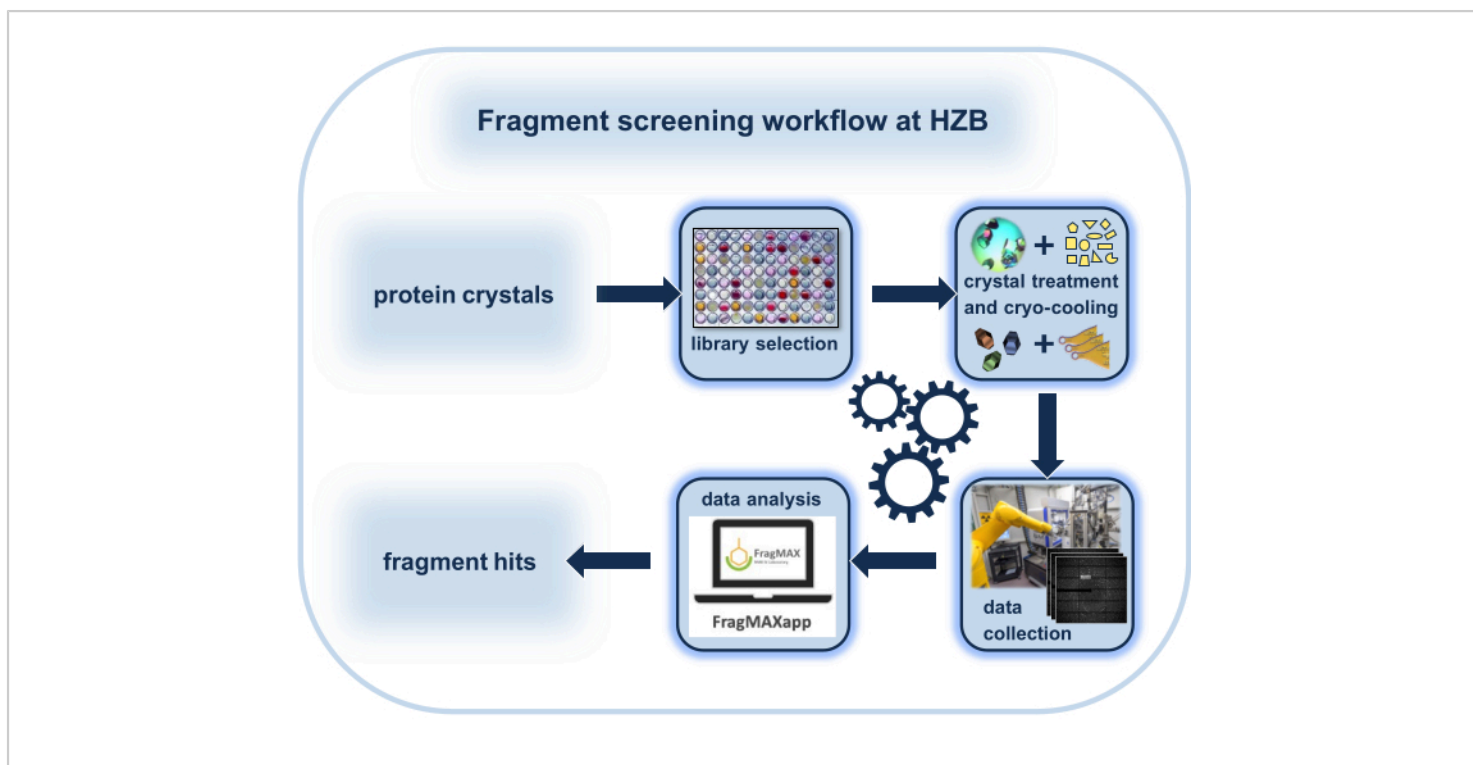


Figure 1: Schematic representation of the workflow of a crystallographic fragment-screening (CFS) experiment with a focus on the special environment at the Helmholtz-Zentrum Berlin. [Please click here to view a larger version of this figure.](#)

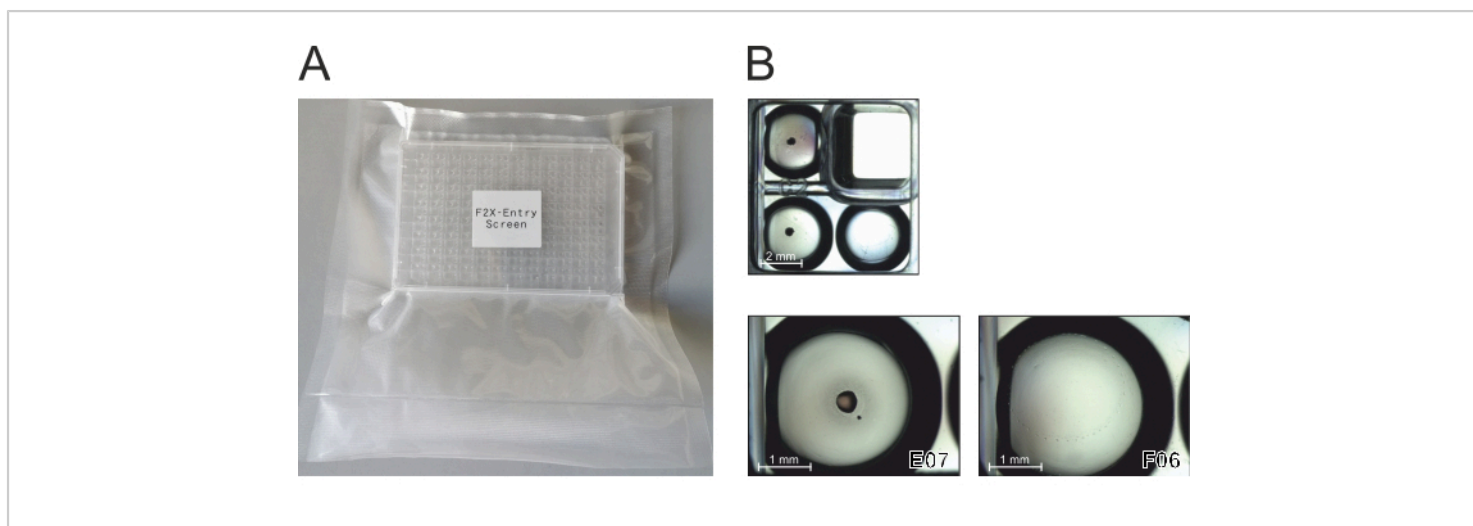


Figure 2: Formulation and packaging of the F2X-Entry Screen. The 96-compound screen is available on a 3-lens 96-well MRC low-profile plate, sealed with foil and vacuum-packed. The 96 compounds of the screen are dried from DMSO solutions in two of the three lenses of each well. [Please click here to view a larger version of this figure.](#)

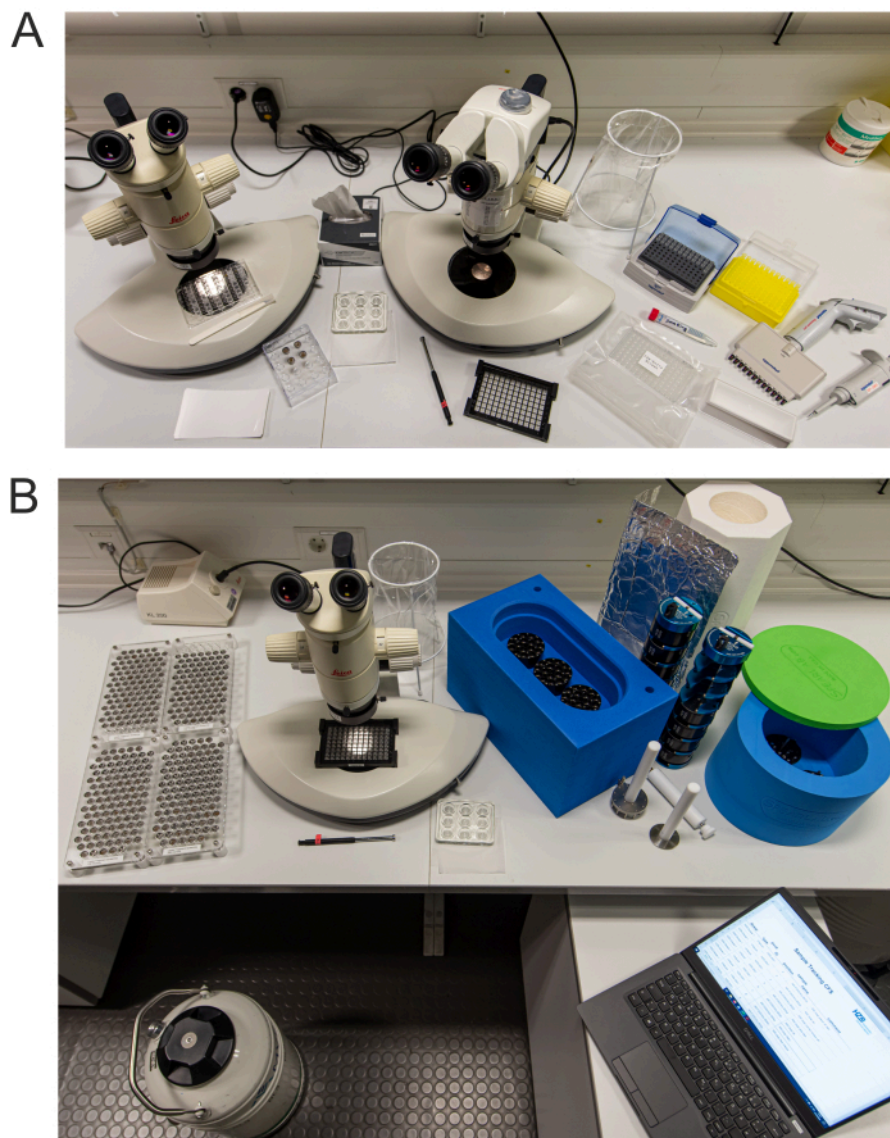


Figure 3: Photography of the CFS workbench in the HZB preparation lab. Assemblies of necessary tools for A) soaking and for B) crystal harvesting are displayed. [Please click here to view a larger version of this figure.](#)

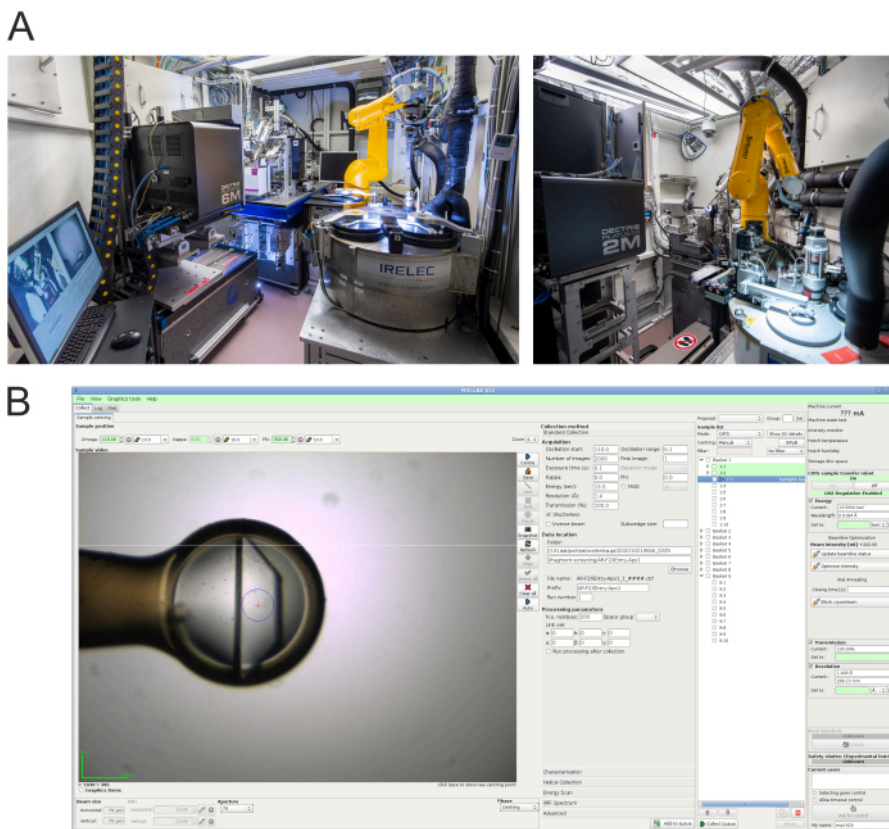


Figure 4: Data collection end stations and control software. **A)** Photograph of the experimental hutch of HZB-MX beamlines BL14.1 (left) and BL14.2 (right)¹⁵. **B)** Screenshot of the MXCuBE2 experiment control interface¹⁶ used at BL14.1 for diffraction data collection. At BL14.2 a very similar interface is used. [Please click here to view a larger version of this figure.](#)

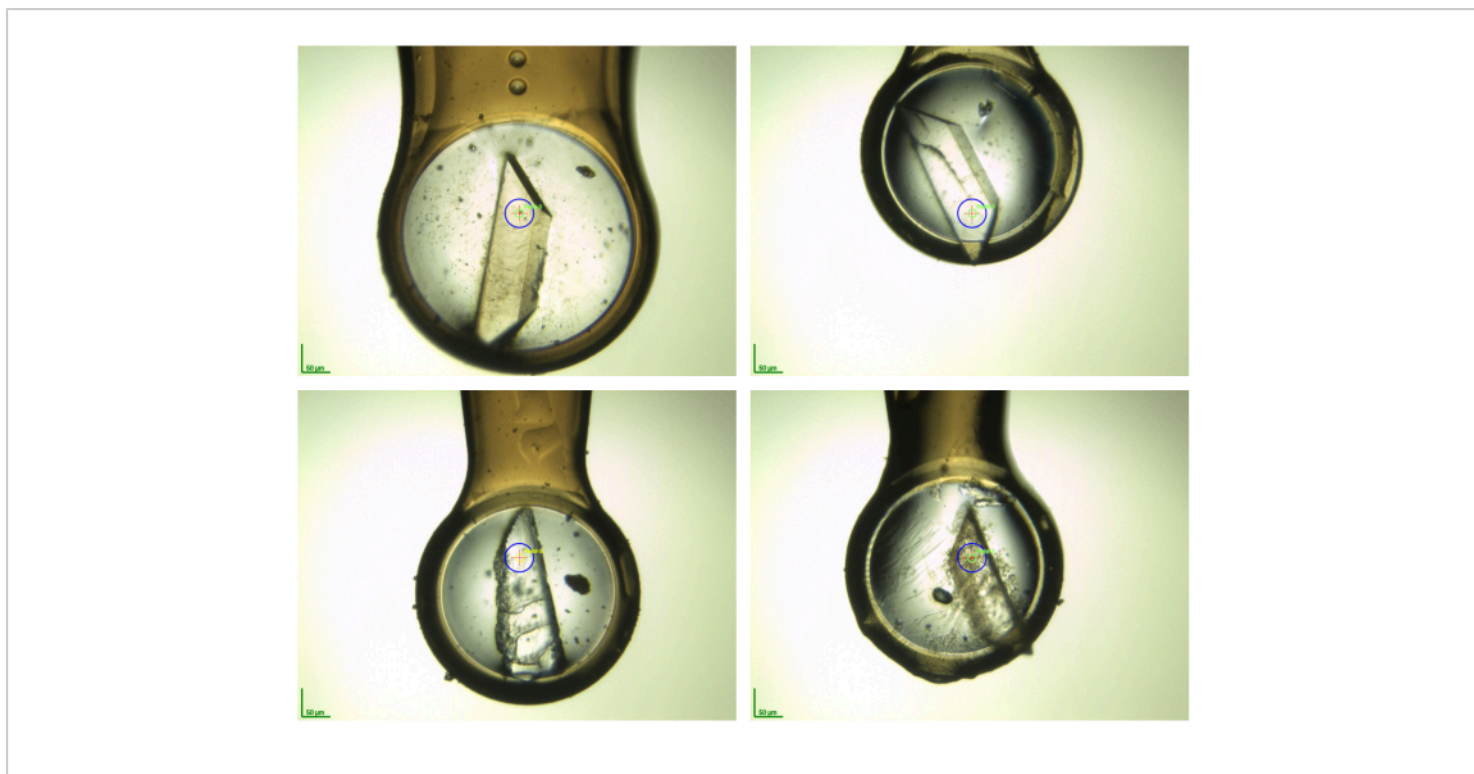


Figure 5: Photographic snapshots of some crystalline samples in cryogenic environment before data collection. This illustrates the variability of morphologies of the crystals after performing the fragment soaking and crystal harvesting. The photographs were taken on the BioMAX beamline (MAX IV synchrotron, Lund, Sweden) for AR samples collected there as part of the F2X-Entry Screen validation¹¹. [Please click here to view a larger version of this figure.](#)

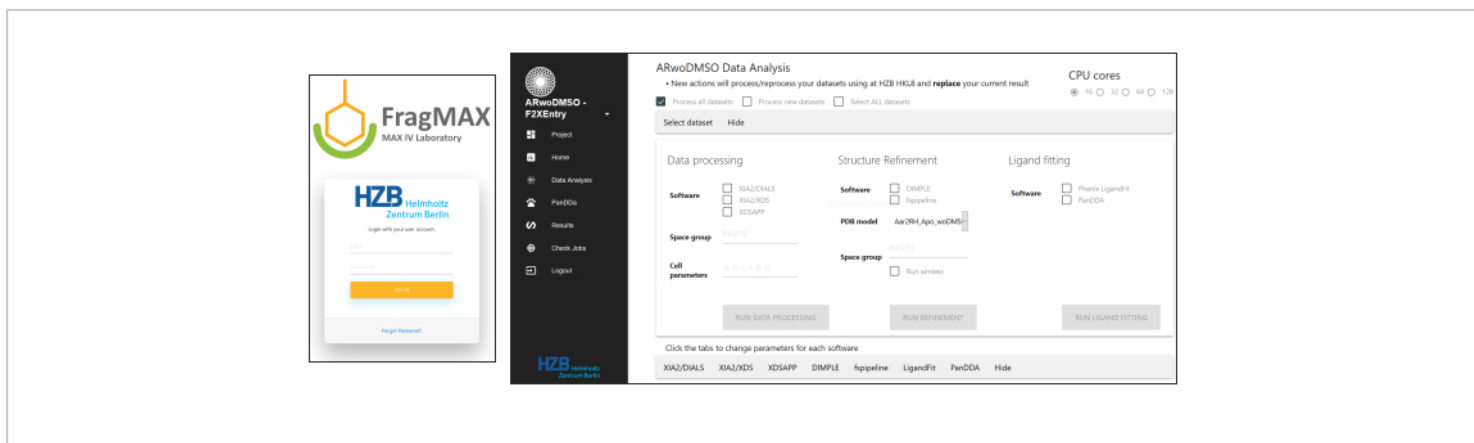


Figure 6: Screenshot of the FragMAXApp¹⁸ installed at the HZB for convenient data analysis. More details in Lima et al., FragMAXapp, unpublished data. [Please click here to view a larger version of this figure.](#)

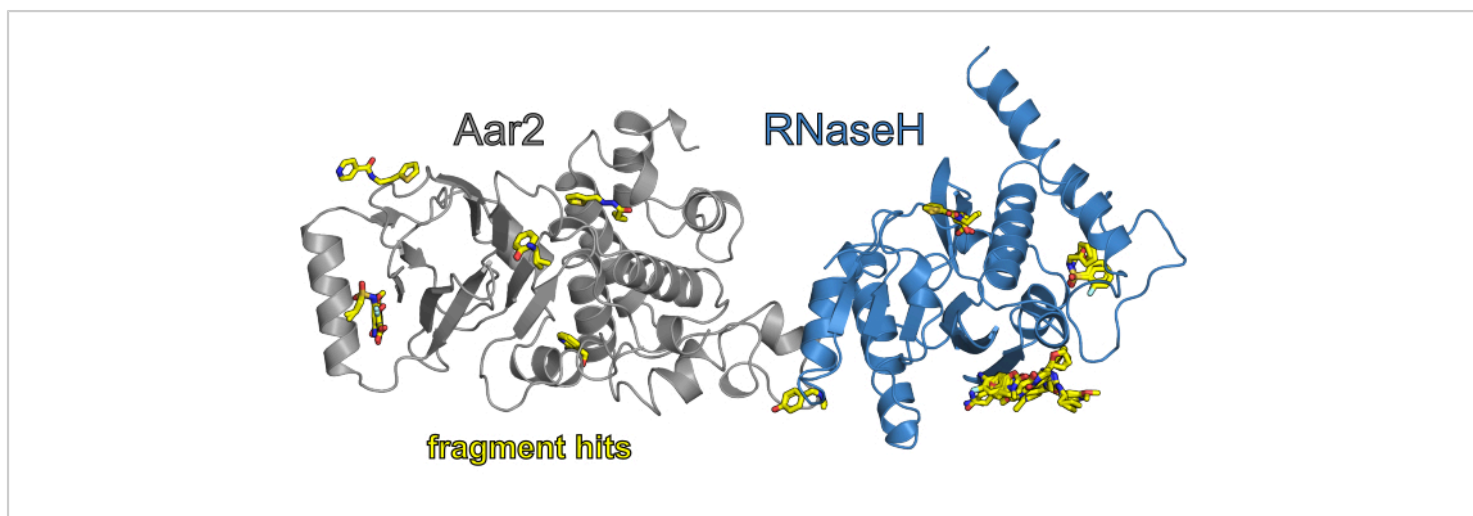


Figure 7: Overview of the results of the CFS campaign F2X-Entry vs. AR (without DMSO). The AR protein complex is shown in cartoon view, with Aar2 colored in gray and the RNaseH-like domain of Prp8 colored in blue. The fragment hits of the campaign are colored in element colors (C - yellow, O - red, N - blue, S - orange, Cl - light cyan). [Please click here to view a larger version of this figure.](#)

Discussion

For a successful CFS campaign, it is vital to adhere to the described prerequisites (see **Introduction**). A reliable crystallization system is needed for the reproducible growth of many well-diffracting crystals, and a well-refined structure is needed as the input apo model for automated refinement. It is also important to check that the target site on the protein (active site, or interface area) is accessible for fragments in the crystal lattice. It is crucial to optimize the soaking conditions beforehand to ensure that the soaking does not significantly deteriorate the crystal quality. Neglecting these aspects will very likely lead to a suboptimal experiment, which will be of limited use and will, in the worst case, require a repetition of the entire experiment.

The protocol described above outlines the procedures that are followed during a standard CFS campaign. If all prerequisites are met, at least 90% of all soaked crystals

should display diffraction to high resolution in a diffraction experiment. If this is not the case, the soaking times may be shortened to a few hours or even minutes. Due to the good solubility of most of the fragments, this should suffice to obtain decent occupancy values. Also, a typical CFS campaign will result in a hit rate of roughly 10% or above. For the F2X-Entry Screen validation campaigns¹¹ and ongoing user campaigns with the same library even higher hit rates have been observed (20% and above, data not shown).

A general caveat of crystallographic fragment screening is the presence of crystallographic contact sites. These could either occlude *a priori* known active sites (to be checked before the screening, see above), or these contact sites also often provide pockets and hot spots where fragments can bind. Such fragment hits will be artifacts of the crystallization lattice and will likely not bind to the protein in solution. These events tend to occur more often in soaking experiments than in co-crystallization experiments (probably due to the higher

fragment concentrations employed in soaking experiments). However, according to previous experience, they generally constitute only a minor portion of the hits obtained. For example, in the F2X-Entry Screen validation campaign using endothiapepsin (EP) and the spliceosomal protein-protein complex of Prp⁸^{RNaseH} and Aar2 (AR), most of the hits occurred in promising sites¹¹. For EP, 27 out of the 37 observed binding events were located in the active site (i.e., the peptide cleft of this protease). The 10 remote binding events comprise two solvent exposed binding events and eight crystal contact binding events (corresponding to five unique hits). Excluding those crystal contact hits would still reflect an overall rate of 24% unique hits for the EP campaign. It is also important to notice that binding events remote of a known active site (except crystal contact binders) could also potentially be interesting (e.g., revealing new hot spots or allosteric sites of the protein). For the AR campaign (in the same publication), of the 23 observed binding events, seven were located at crystal contacts, one was located at the direct interface of the two proteins, seven were located at known protein-protein interactions sites with other binding partners of the larger biological context (hence different assembly stages of the spliceosome), eight binding events revealed two hot spots on AR of yet unknown function and one being at a solvent exposed surface of Prp⁸^{RNaseH}. Therefore, excluding the events at crystal contacts and the Prp⁸^{RNaseH} singleton, the number of potentially useful binding events is 15 (corresponding to 14 unique hits) thus a hit rate of 15.6%. These hits can be starting points for design of protein-protein interaction modulators or for tool compounds aimed to explore the two discovered Aar2 hot spots. Taken together, also in line with conducted user campaigns, often only a minor portion of hits in crystallographic fragment screening must be

disregarded as artifacts. However, this will also be largely target dependent.

If the hit rate is significantly lower, this may indicate one of the following problems related to the target protein. For instance, in a CFS campaign against a viral cysteine protease a hit rate of only 3% was observed (data not shown). It turned out that the protein used was likely chemically modified in its active site. In such a case, a different protein preparation may solve the problem. If crystals are very DMSO intolerant, the F2X-Entry Screen may also be used without DMSO, although the results may differ to some degree. Most of the hits obtained in the presence of DMSO will also show up in its absence. There will also be some hits that cannot be observed in the absence of DMSO, even though they can be observed in its presence. And finally, there will be some that only show up in the absence of DMSO.

The most severe difficulty occurs if the protein undergoes an induced-fit motion upon substance binding. Most likely, the crystal lattice will not tolerate the protein motion and the crystals will disintegrate. In such a case, the only choice is to resort to co-crystallization of the protein and the fragments. This may, however, lead to new crystal forms. Therefore, much of the automation of the entire process will not work efficiently anymore. Luckily, in most CFS campaigns conducted at the HZB so far, this kind of problem has not been encountered. It may be, that the weak binding of a fragment, does not provide enough energy to induce a protein motion, in particular if the crystallized conformation is stabilized by crystal packing forces.

Another serious limitation of the method which the authors have encountered so far is when the crystallization cocktail (and thus the soaking solution) contains volatile compounds.

Then it becomes close to impossible to perform all the crystal handling in a meaningful way.

Different proteins may contain druggable sites to a greater or lesser extent. For example, protein-protein interactions are usually mediated by extended flat surfaces that are more difficult to target. The fragment binding hit rate will therefore likely depend on the structure of the protein's molecular surface. In an extreme case, a protein might not contain any suitable surface hot spots that serve as target sites for fragment binding. Thus, despite a meticulously performed experiment, no fragment hits will result from the screening. However, the authors have so far not encountered such a situation.

In principle, using the protocol outlined above, the crystal soaking and harvesting part of a CFS campaign can be performed in any laboratory that is equipped for crystal handling. This distinguishes the methodology at HZB from other CFS facilities and can be an advantage in some cases. For example, if the crystals cannot be easily re-produced at another site or if the travelling of the experimenters is limited (e.g., in a world-wide pandemic situation), users at HZB are therefore provided with the entire equipment (pucks, tools, EasyAccess Frame, sample holders, etc.) as a portable set.

However, the requirements for large numbers of sample holders and cryogenic storage capacities are still more conveniently met at dedicated CFS facilities. Moreover, the need for collection of many diffraction data sets strongly advocates for localizing these facilities close to beamlines which are geared towards a high sample throughput. Examples for this are the beamlines I04 at the Diamond Light Source and the associated XChem facility in UK^{8,25}, the

MASSIF beamlines at the ESRF in France²⁶ or the FragMAX facility at the BioMAX beamline at MAX IV in Sweden¹⁸.

In the future, it could be envisioned to design CFS experiments without the need for crystal handling altogether. First advances in this direction have been reported. For instance, by acoustic liquid transfer allowing the mixing of both the crystal-containing solutions and the fragment solutions directly on mesh-type sample holders²⁷. Another approach was used for XFEL-based ligand-screening. In a proof-of-principle experiment, a crystal slurry was prepared in batch, and soaking and diffraction data collection were performed on a silicon fixed target chip²⁸. However, these approaches are still under development and far from being applicable to a wide range of protein targets or feasible for CFS facilities as a routine.

With the protocol in this work detailed instructions to successfully perform CFS campaigns straight-forwardly at HZB (and elsewhere) have been outlined and general guidance and useful hands-on tips in preparing and conducting such experiments with higher chances for success have been given. Ultimately, better odds and success rates in CFS screening largely contribute to efficiently providing starting points for downstream development of tool compounds or drug candidates.

Disclosures

A patent application regarding the EasyAccess Frame has been filed by Helmholtz-Zentrum Berlin with the German Patent and Trademark Office with the registration number DE 10 2018 111 478.8. Additionally, an international patent application via the PCT route, using the priority of the German patent, has been filed.

Acknowledgments

We thank the numerous user groups that have performed CFS campaigns at the HZB. Their feedback led to the incremental improvement of our workflow. We want to thank the drug design group at the University of Marburg and the FragMAX group at MAX IV, as the close collaborations were the foundation for several developmental leaps for improved CFS. We are thankful for the support by the German Federal Ministry of Education and Science (BMBF), via the projects Frag2Xtal and Frag4Lead (numbers 05K13M1 and 05K16M1). We are additionally grateful for support via iNEXT-Discovery, project number 871037, funded by the Horizon 2020 program of the European Commission.

References

- Erlanson, D.A., Fesik, S.W., Hubbard, R.E., Jahnke, W., Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nature Reviews Drug Discovery*. **15** (9), 605-619 (2016).
- Hall, R.J., Mortenson, P.N., Murray, C.W. Efficient exploration of chemical space by fragment-based screening. *Progress in Biophysics and Molecular Biology*. **116** (2-3), 82-91 (2014).
- Erlanson, D.A. Introduction to fragment-based drug discovery. *Topics in Current Chemistry*. **317**, 1-32 (2012).
- Scott, D.E., Coyne, A.G., Hudson, S.A., Abell, C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry*. **51** (25), 4990-5003 (2012).
- Lamoree, B., Hubbard, R.E. Current perspectives in fragment-based lead discovery (FBLD). *Essays in Biochemistry*. **61** (5), 453-464 (2017).
- Schiebel, J. *et al.* Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study. *ACS Chemical Biology*. **11** (6), 1693-1701 (2016).
- Schiebel, J. *et al.* High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure*. **24** (8), 1398-1409 (2016).
- Krojer, T. *et al.* The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination. *Acta Crystallographica Section D: Structural Biology*. **73** (Pt 3), 267-278 (2017).
- Radeva, N. *et al.* Active Site Mapping of an Aspartic Protease by Multiple Fragment Crystal Structures: Versatile Warheads to Address a Catalytic Dyad. *Journal of Medicinal Chemistry*. **59** (21), 9743-9759 (2016).
- Cox, O.B. *et al.* A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chemical Science*. **7** (3), 2322-2330 (2016).
- Wollenhaupt, J. *et al.* F2X-Universal and F2X-Entry: Structurally Diverse Compound Libraries for Crystallographic Fragment Screening. *Structure*. **28** (6), 694-706.e5 (2020).
- EU OPENSREEN fragment library*. at <<https://www.eu-openscreen.eu/services/compound-collection/fragment-library.html>>. (2021).
- Pearce, N.M. *et al.* A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nature Communications*. **8**, 15123 (2017).
- Barthel, T., Huschmann, F.U., Wallacher, D., Klebe, G., Weiss, M.S., Wollenhaupt, J. Facilitated crystal handling

- using a simple device for evaporation reduction in microtiter plates. *Journal of Applied Crystallography*. **54**, <https://doi.org/10.1107/S1600576720016477> (2021).
15. Mueller, U. *et al.* The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives. *The European Physical Journal Plus*. **130** (7), 141 (2015).
 16. Oscarsson, M. *et al.* MXCuBE2: The dawn of MXCuBE collaboration. *Journal of Synchrotron Radiation*. **26** (2), 393-405 (2019).
 17. Mueller, M., Wang, M., Schulze-Briese, C. Optimal fine ϕ -slicing for single-photon-counting pixel detectors. *Acta Crystallographica Section D: Biological Crystallography*. **68** (1), 42-56 (2012).
 18. Lima, G.M.A. *et al.* FragMAX: The fragment-screening platform at the MAX IV Laboratory. *Acta Crystallographica Section D Structural Biology*. **76** (8), 771-777 (2020).
 19. Sparta, K.M., Krug, M., Heinemann, U., Mueller, U., Weiss, M.S. XDSAPP2.0. *Journal of Applied Crystallography*. **49** (3), 1085-1092 (2016).
 20. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*. **43** (1), 186-190 (2010).
 21. Winn, M.D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*. **67** (4), 235-242 (2011).
 22. Terwilliger, T.C., Klei, H., Adams, P.D., Moriarty, N.W., Cohn, J.D. Automated ligand fitting by core-fragment fitting and extension into density. *Acta Crystallographica Section D: Biological Crystallography*. **62** (8), 915-922 (2006).
 23. Pearce, N.M., Krojer, T., Von Delft, F. Proper modelling of ligand binding requires an ensemble of bound and unbound states. *Acta Crystallographica Section D Structural Biology*. **73** (3), 256-266 (2017).
 24. Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*. **66** (4), 486-501 (2010).
 25. Collins, P.M. *et al.* Chapter Eleven - Achieving a Good Crystal System for Crystallographic X-Ray Fragment Screening. *Methods in Enzymology*. **610**, 251-264 (2018).
 26. Bowler, M.W. *et al.* MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *Journal of Synchrotron Radiation*. **22** (6), 1540-1547 (2015).
 27. Cuttitta, C.M. *et al.* Acoustic transfer of protein crystals from agarose pedestals to micromeshes for high-throughput screening. *Acta Crystallographica Section D Biological Crystallography*. **71** (1), 94-103 (2015).
 28. Moreno-Chicano, T. *et al.* High-throughput structures of protein-ligand complexes at room temperature using serial femtosecond crystallography. *IUCrJ*. **6** (6), 1074-1085 (2019).

Workflow and Tools for Crystallographic Fragment Screening at the Helmholtz-Zentrum Berlin

Jan Wollenhaupt¹, Tatjana Barthel^{1,2}, Gustavo M. A. Lima³, Alexander Metz⁴, Dirk Wallacher⁵, Elmir Jagudin³, Franziska U. Huschmann^{1,4}, Thomas Hauß¹, Christian G. Feiler¹, Martin Gerlach¹, Michael Hellmig¹, Ronald Förster¹, Michael Steffien¹, Andreas Heine⁴, Gerhard Klebe⁴, Uwe Mueller¹, Manfred S. Weiss¹

¹Macromolecular Crystallography, Helmholtz-Zentrum Berlin ²Structural Biochemistry Group, Institute for Chemistry and Biochemistry, Freie Universität Berlin ³BioMAX, MAX IV Laboratory ⁴Drug Design Group, Institute of Pharmaceutical Chemistry, Philipps-Universität Marburg ⁵Department Sample Environment, Helmholtz-Zentrum Berlin

Corresponding Author

Manfred S. Weiss
msweiss@helmholtz-berlin.de

Citation

Wollenhaupt, J., Barthel, T., Lima, G.M.A., Metz, A., Wallacher, D., Jagudin, E., Huschmann, F.U., Hauß, T., Feiler, C.G., Gerlach, M., Hellmig, M., Förster, R., Steffien, M., Heine, A., Klebe, G., Mueller, U., Weiss, M.S. Workflow and Tools for Crystallographic Fragment Screening at the Helmholtz-Zentrum Berlin. *J. Vis. Exp.* (), e62208, doi:10.3791/62208 (2021).

Date Published

March 3, 2021

DOI

10.3791/62208

URL

jove.com/video/62208

Materials

Name	Company	Catalog Number	Comments
1µL pipet	Eppendorf	EP3123000012	
12 channel pipet, 100µL	Eppendorf	EP4861000791	
Blow dryer	TH-Geyer	9.106 788	
Crystal containing crystallization plates			Contains crystals to be soaked
Crystallization incubator			Providing constant temperature for crystallization experiment, at HZB: 20°C
Dual Thickness MicroLoops (LD) of different aperture sizes	MiTeGen	various, e.g. M5-L18SP-75LD	250 loops in the appropriate size needed for the protocol, can be provided by HZB
EasyAccess Frame	HZB		The EasyAccess Frame is a special device for handling multiple crystals, which was developed at the HZB (Barthel et al., 2021).
F2X-Entry Screen plate	HZB		Developed F2X-Entry Screen (Wollenhaupt et al., 2020)
Glas spot plate	VWR	MARI1406506	
Liquid nitrogen			At least a filled up 5 L can
Liquid nitrogen storage can	n.a.	n.a.	
Magentic crystal wand	MiTeGen	M-R-1013198	
Microscopes	Leica	n.a.	
MRC 3-lens 96well low profile crystallization plate	SwissCl	3W96TLP-UVP	For mock-soaked crystals (optional)
Reagent reservoir	Carl Roth	EKT6.1	25 ml volume

Sample tracking template			Provided from HZB, see supplementary information
Scalpel	B. Braun	BA825SU	
Sealing foil for microtiter plates	GreinerBioOne	676070	
Shelved puck shipping canes (for Unipucks)	MiTeGen	M-CP-111-065	2 canes made of aluminum; can be provided by the HZB
Soaking solution			At least 5 ml are needed
Soaking solution including cryo-protectant, 150 μ L			Only needed if soaking solution is not cryo-protectant already
Tissues	Roth (Kimberly Clark Professional)	AA64.1	
Transport dewar (Whartington dry shipper)	MiTeGen	TW-CX100	2 Travel dewars for storage of the 2 unipuck canes, alternatively a storage dewar of type VHC35 or similar could be used.
Unipuck foam dewars with lid	MiTeGen	M-CP-111-022	two foam dewars especially suited for unipuck handling described in the protocol if SPINE pucks are used, different foam dewars might have to be applied.
Unipuck starter set	MiTeGen	M-CP-UPSK001	Can be provided by the HZB
Unipucks	MiTeGen	M-CP-111-021	14 unipucks; can be provided by the HZB

VI. **Barthel, T.;** Wollenhaupt, J.; Lima, G. M. A.; Wahl, M. C.; Weiss, M. S., Large-Scale Crystallographic Fragment Screening Expedites Compound Optimization and Identifies Putative Protein-Protein Interaction Sites. *J. Med. Chem.* **2022**, *2022*, 14630–14641. <https://doi.org/10.1021/acs.jmedchem.2c01165>

This publication must be purchased online.