Too weak for biophysics: Prioritising and progressing fragment hits from X-ray crystallographic screening.

Yuliya Dubianok

Pembroke College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michelmas 2020

"One never notices what has been done;
one can only see what remains to be done."

Marie Curie

to Anna

Acknowledgement

I would like to thank my supervisor Frank von Delft for giving me the opportunity to join the PX team, allowing me to shape my project and enjoy the scientific freedom. I hope this journey was as educational for you as for me. I am equally grateful to my second supervisor Brian Marsden for his encouragement and support throughout the project. I further thank my Viva examiners Prof. Martin Scanlon and Prof. Paul Brennan.

I would like to thank all the PX group members, past and present, especially Elliot Nelson for all his help with debugging code, occupancy analysis, helpful discussions and his support. I thank Tobias Krojer for his help and advice in all things crystallography and beyond. I also thank Mike Fairhead for introducing me to PrepX, helping me with all lab-related things and R12. Many thanks to Ric Gillams for his kind words and motivation (it was great to have your company at the SGC). I wish to thank Beth MacLean not only for her help with orders and finances, but also her continuous encouragement. I am grateful to Frank Waters for helping me shipping my samples around the world.

Huge thanks to the XChem/I04-1 team at Diamond, especially Alice Douangamath and Romain Talon for all their help, good mood, positivity and always being able to squeeze my experiments into the schedule. I thank Patrick Collins for helping me getting started in the PX group and at Diamond and introducing me to XChem.

I thank my industrial collaborators Fabrice Moriaud, Helena Kovacs, Maksim Mayzel, Aitor Moreno and Till Kühn for welcoming me at the Bruker site at Fällanden, Switzerland, providing training in NMR and sharing their expertise and advice.

Many thanks to my industrial collaborators at Exscientia, in particular Anthony Bradley for his advice on computational approaches and fruitful discussion, Sean Robinson and Iva Navratilova for giving me the opportunity to conduct SPR experiment at Exscientia and providing their time and expertise.

Further, I thank Prof. Tim Claridge and Dr. Ben Davis for fruitful discussion on NMR experiments, and Tamas Szommer for his helpful advice on fragment elaboration.

I wish to thank my family and friends in Oxford and beyond for their support throughout this journey. I thank my fiancé Simon Nadal for his love, tireless support, patience and proof reading of this thesis.

Author declaration

I declare that the work presented in this thesis is my own except where explicitly stated otherwise in the text. No parts of this thesis have been submitted for any other degree at any other university or learning institution.

Yuliya Dubianok

Abstract

In the last decade, X-ray crystallography has matured to a powerful primary screening technique for fragment hit identification in structure-based drug discovery. Unmatched in its sensitivity, it provides valuable knowledge of precise fragment binding location in a protein-fragment complex. The comparison to ^{19}F NMR fragment campaign conducted in this work emphasises the strength of crystallographic screening by revealing additional binding sites and interactions for the breast cancer target NUDT5.

The high sensitivity of X-ray crystallography allows to identify even weakly binding fragment hits, which are often undetectable by commonly used biophysical methods. The lack of orthogonal tools with sufficient detection limits to quantify their interactions impedes prioritisation of fragment hits for elaboration. For this purpose, the use of both NMR-based techniques as well as concentration-dependent crystallographic soaking are explored. While further investigation of relaxation-based $T_{1\rho}$ NMR for rapid ranking of weak binders is necessary, the crystallographic approach demonstrates a correlation between the lowest detectable soaking concentration (LDSC) and fragment binding affinity. A low-concentration counter-soak is proposed as a simple solution to triage fragment hits according to their binding strengths, streamlining their selection for further elaboration.

Weak fragment hits can be informative for lead discovery: they can reveal unknown binding sites and can be progressed into more potent compounds. For the target NUDT5, crystallographic screening identified a novel, previously unreported binding pocket. The elaboration efforts utilising SAR by catalogue approach and crystallographic soaking as readout highlight the value of crystallography for the progression of weak binders in absence of other assays.

Contents

List of Figures

List of Tables

Abbreviations

ATP	adenosine triphosphate
ADME	absorption, distribution, metabolism and excretion
ADP	adenosine diphosphate
AMP	adenosine monophosphate
BDC	background density correction factor
CETSA	cellular thermal shift assay
CIP	calf intestinal alkaline phosphatase
Coot	crystallographic object-oriented toolkit
CORCEMA	complete relaxation and conformation exchange matrix
CPMG	Carr-Purcell-Meiboom-Gill sequence
CSA	chemical shift anisotropy
DLS	Diamond Light Source
DMSO	dimethyl sulfoxide
DOS	diversity-oriented synthesis
DSiP	Diamond SGC iNext Poised fragment library
DSPL	Diamond SGC poised library
EDTA	ethylene diamine tetraacetic acid
FBDD	fragment-based drug discovery
FBLD	fragment-based lead discovery
FBS	fragment-based screening
GEM-CRL	group epitope mapping considering relaxation of the ligand
HAC	heavy atoms count
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSA	human serum albumin

HSQC	heteronuclear single-quantum coherence
HTS	high throughput screening
IC_{50}	inhibitory concentration achieving 50% inhibition
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
ITC	isothermal titration calorimetry
K_D	dissociation constant
LDSC	lowest detectable soaking concentration
LB	lysogeny broth
LE	ligand efficiency
MD	molecular dynamics
MS	mass spectrometry
MST	microscale thermophoresis
MW	molecular weight
NAD^+	nicotinamide adenine dinucleotide
NMR	nuclear magnetic resonance
NMNAT1	nicotinamide nucleotide adenylyltransferase 1
NOE	nuclear Overhauser effect
NTA	nitrilotriacetic acid
OD_{600}	optical density, measured at 600 nm
PAGE	polyacrylamide gel electrophoresis
PanDDA	pan-dataset density analysis
PAR	poly-ADP-ribose
PARG	poly-ADP-ribose-glucohydrolase
PARP1	poly-ADP-ribose polymerase 1
PDB	protein data bank
PEG	polyethylene glycol
PK/PD	pharmacokinetics /pharmacodynamics
PLIP	protein-ligand interaction profiler
PMI	principal moments of inertia

PP$_i$	pyrophosphate
PR	progesterone receptor
PrOF	protein-observed fluorine NMR
RU	response unit
R5P	ribose-5-phosphate
RMSF	root mean square fluctuation
SAR	structure activity relationship
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
SGC	Structural Genomics Consortium
SMILES	simplified molecular input line entry system
SOC	super optimal broth
SPR	surface plasmon resonance
STD-NMR	saturation transfer difference NMR
TEV	tobacco etch virus
TSA	thermal shift assay
Tris	tris(hydroxymethyl)aminomethane
XCE	XChem Explorer

1

Introduction

Contents

1.1. Hit identification and progression in preclinical drug discovery

The development of new drugs is a complex, time-consuming and costly process.[1–3] Their identification and development follows a widely adopted pipeline comprising a preclinical and clinical phase.[4] The preclinical research involves all steps from target validation to the development of a drug candidate, while the clinical phases focus on safety and efficacy of the drug candidate in humans (Figure 1.1).[4] The drug discovery process has been reviewed extensively[5–10]; this thesis focusses on hit discovery and hit-to-lead progression in preclinical research.

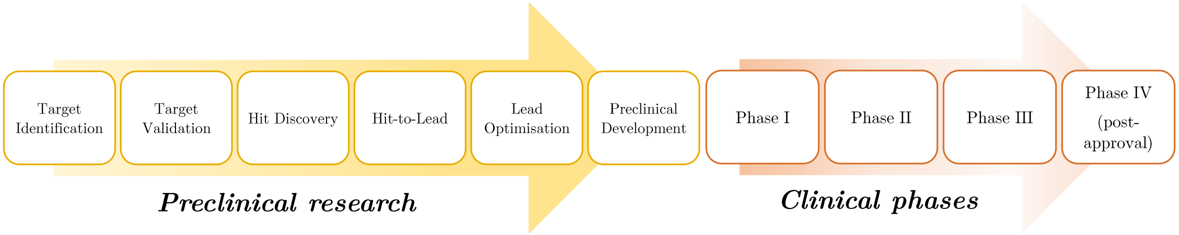


Figure 1.1. The drug discovery pipeline illustrates the phases of preclinical and clinical developments. Modified from[4].

High-throughput screening (HTS) is the dominating strategy for hit discovery, the identification of initial starting points for the modulation of target activity.[11] HTS typically involves screening of libraries containing >10000 compounds with drug-like properties in a biochemical assay.[12] Over the last two decades, fragment-based lead discovery (FBLD) emerged as an alternative. It builds on the idea of screening small building blocks – fragments – which can be assembled or expanded into larger, drug-like molecules in a rational approach (see 1.2).[13,14]

The three-dimensional structure of the target can be used to guide progression of initial hits into a lead candidate: an approach termed structure-based lead discovery.[15–18] Hereby, molecular details of the interactions and spatial constraints of the protein-ligand complex inform the design of new compounds.[18]

The hit-to-lead progression in the pharmaceutical industry involves iterative optimisation of an array of physicochemical properties, including compound affinity, target selectivity, compound solubility, cell permeability as well as toxicological and pharmacological parameters – eventually ending in a drug candidate.[4,9,19] Academic groups on the other hand often aim at the development of chemical probes – tool compounds with clearly defined mode of action used for research.[20] The identification and progression of hits using FBLD for both purposes is the central theme of this thesis.

1.2. Fragment-based lead discovery

Over the last twenty years, fragment-based lead discovery (FBLD) has become well-established as a powerful approach for finding high-quality lead candidates for drug discovery.[21,22] Four compounds originating from FBLD have entered the drug market and over 40 are currently undergoing clinical trials.[23] The FBLD approach entails screening libraries of fragments, which are defined as small molecules with molecular weight below 300 Da.[24–26]

Fragment screening offers a number of advantages over traditional lead discovery techniques such as high throughput screening (HTS).[13,27] The premise of screening fragments is based on the concept of molecular complexity: their small size and low complexity increase the probability of matching target protein binding site and allow an efficient exploration of chemical space (Figure 1.2).[27–31] The number of compounds typically screened is in the range of a few hundreds to thousands and is about three orders of magnitude smaller compared to a typical HTS experiment.[13,14,27] In addition, fragment campaigns often result in a higher hit rate than HTS, which screen larger, more functionalised molecules with drug-like properties.[27,30,32–34]

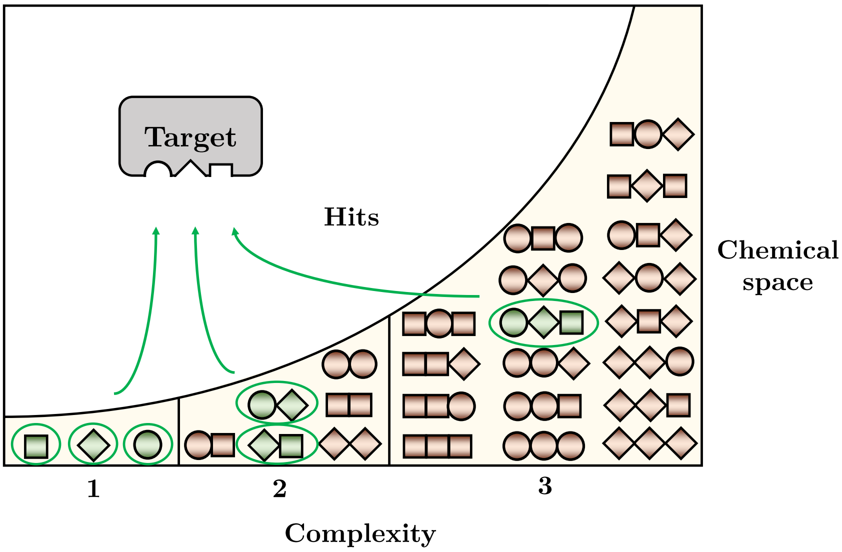


Figure 1.2. Schematic representation of efficient chemical space exploration using fragment-based drug discovery approach. Small size and low complexity of fragments increase the probability of finding a hit (green), while reducing the number of compounds to screen. Modified from the original drawing from Dr. P. Haebel.

Although small, fragments form a limited number of high quality interactions.[27,35–37] In many cases, they are likely to have a strong enthalpic contribution allowing to overcome the loss of entropy associated with ligand binding to a target.[38–40]

Despite high-quality interactions, their overall affinity remains relatively low ($K_D > 0.1-10$ mM).[38,39,41,42] Detection of weak interactions places high demands on sensitivity and robustness of screening methods.[14,36,42] Biochemical *in vitro* assays typically used in HTS approaches often do not satisfy the required needs.[39,43–45] Biophysical methods are suited to detect fragment binding, as they provide required sensitivity, reproducibility and throughput.[14,36,39,43,44] Most commonly used methods include nuclear magnetic resonance (NMR), surface plasmon resonance (SPR) and X-ray crystallography.[39] These assays are described in 1.2.3,1.2.4 and 1.2.5.

To achieve a selective chemical tool or a lead suitable for progression into clinical trials from a low affinity fragment hit, a gain in potency by several orders of magnitude is required.[39,41,46] Fragments are elaborated by adding chemical functionalities in a stepwise manner until a larger compound with improved potency and selectivity is reached.[46,47] During optimisation process, favourable interactions are expected to contribute to the affinity increase, as theoretically the initial entropic penalty of binding has been paid by the original fragment.[39] Starting chemical optimisation from a fragment with high-quality interactions is more likely to deliver a lead with molecular weight within the desired range.[27] In contrast, HTS identifies larger hits with higher affinity, however these can be more difficult to optimise.[27] Successful hit-to-lead progression of a number of compounds demonstrated that high affinity interactions are suitable for optimisation.[39] Since the first report in 1996[48], rapid developments of fragment libraries, screening methods and elaboration techniques contributed to the success of FBLD.[42]

1.2.1. Fragment libraries

The key premise of fragment screening is efficient exploration of vast chemical space by a library of small and chemically diverse fragments.[49,50] The library sizes range between 500-20000 compounds; in practice it is often limited by the throughput of the screening technique applied.[14]

Most fragment libraries adhere to the "Rule of Three"– initially defined by Astex Therapeutics[51] – which imposes restrictions on compound properties including a molecular weight below 300 Da, presence of three or fewer hydrogen bond donors and acceptors, three or fewer rotatable bonds and partition coefficient cLogP below 3 (measure of hydrophobicity).[14,51,52] The number of non-hydrogen atoms is typically <20, often decreasing lipophilicity and improving aqueous solubility.[14]

Principles for design of fragment libraries have been reviewed extensively.[26,44,49,53–59] Briefly, purity, stability and solubility are among the most important criteria to ensure reliable screening results.[14,26,53] For example, crystallographic screening (1.2.3) utilises fragment solutions at high concentration (up to 500 mM), placing particular demands on compound solubility.[34,52,60–62] Careful selection of compounds is necessary to avoid pan-assay interference compounds (PAINS), possessing reactive functionalities towards protein residues, such as alkyl halides, epoxides and Michael acceptors.[13,14,26,63] Aggregators are also common bad actors which often cause false-positives through non-specific binding.[13,14,26,64,65]

Many commercially available fragment libraries are centred on sp^2-rich aromatic compounds and tend to underrepresent chiral sp^3 molecules.[66,67] Increased 3-dimensionality in architecture of fragments provides access a larger chemical space and is arguably advantageous in probing more difficult targets and protein-protein interactions.[14,44,50,66,68,69] The application of diversity-oriented synthesis (DOS) has been utilised to incorporate more structural diversity and 3D character into novel fragment libraries.[66,70] However, retaining advantages in efficient sampling of chemical space requires balance between fragment size, complexity and diversity.[14,44,46,50]

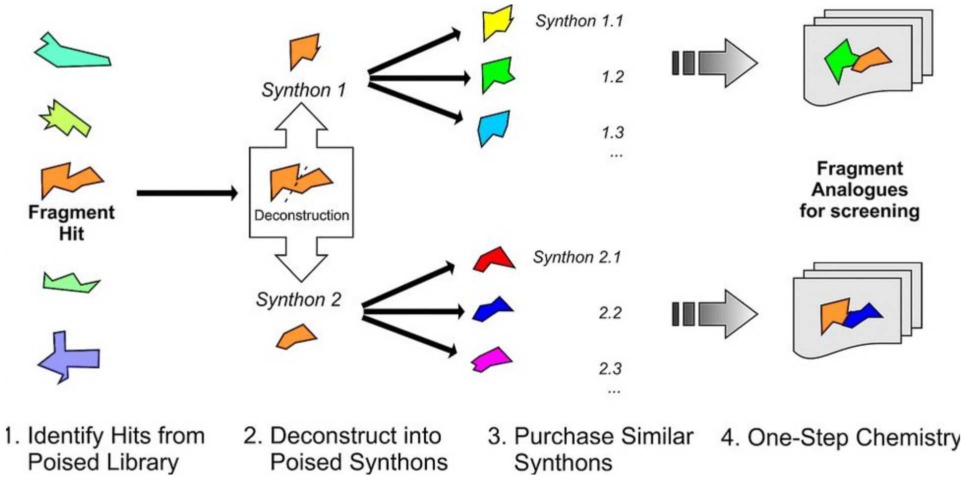


Figure 1.3. Schematic representation of poised fragment de-construction for rapid synthesis of analogues libraries. Poised fragments identified consist of at least two synthons. Commercial analogues of synthons can be combined by parallel synthesis to generate libraries for follow-up molecules. Reprinted from[71] with permission from © RSC.

A different approach to library design, centred around synthetic tractability has been suggested with the Diamond SGC iNEXT poised (DSiP) library [72] – the commercially available iteration of Diamond Sciences Poised Library (DSPL) [71]. DSiP consists of 768 poised fragments, which can be de-constructed into at least two easily accessible synthetic building blocks (synthons, see Figure 1.3). Combination of similar commercially available synthons allows for rapid elaboration of poised fragment hits into a library of analogues. [71] The underlying parallel chemistry utilises robust reactions which tolerate large substrate diversity and form high yield. [71] The scaffolds of the poised fragments are based on top ten reaction most commonly used in medicinal chemistry, twelve heterocycle forming reactions and an oxazole formation. [71,73,74]

Libraries specifically tailored to a screening method or an application have emerged. Development of fluorinated fragment libraries reflects significant attention ^{19}F NMR spectroscopy has gained in fragment-based lead discovery in recent years (2.9.1). [75] Small libraries ($<$ 100 compounds) of low molecular weight fragments (heavy atom count (HAC) 5-7, MW <100 Da)such as MiniFrags [62], FragLights [76] and very small fragments (VSFs, HAC <11) [77] aim to assess druggability by mapping protein surface to identify ligand sites and hot spots.

Development of fragment libraries remain an area of active research. While historically libraries were considered proprietary, emerging academic collections including the DSiP library have driven closer collaboration between academia and industry, promoting a more open exchange focussed on accessibility and improvement of library design. [44]

1.2.2. Methods for fragment screening

Although fragments form high-quality interactions, their binding affinities are typically in the range of 100 µM to 10 mM due their low molecular weight. [21,36] For detection of weak binding, use of sensitive and robust methods for screening is essential. [26,34] A number of biophysical techniques has been successfully used as primary screening method including high throughput X-ray crystallography [35,78,79], nuclear magnetic resonance (NMR) [48], surface plasmon resonance (SPR) [80,81], thermal shift assay (TSA) [34,82] and mass spectrometry [83] as well as more recent techniques such as microscale thermophoresis (MST) [84] and weak affinity chromatography (WAC) [85].

With a growing array of screening methods available, there is no single universal strategy for reliable identification of fragment hits. [34,86] The selection of a screening technique often depends on the nature of the target, available infrastructure and resources as well as expertise.[34,43,87] Among suitable methods for primary screening, X-ray crystallography is the only techniques which does not provide any quantitative information and requires an orthogonal method for affinity quantification.[26,88] Biophysical techniques differ in their underlying physical principle for detection of binding, sensitivity range, throughput, sample requirements and information content from the readout.[86,89]

Whereas ligand- and protein observed NMR methods are the most appropriate choice in regard to sensitivity, other techniques including surface plasmon resonance (SPR), thermal shift assay (TSA) and microscale thermophoresis (MST) provide the advantage of throughput and experimental set-up (Figure 1.4).

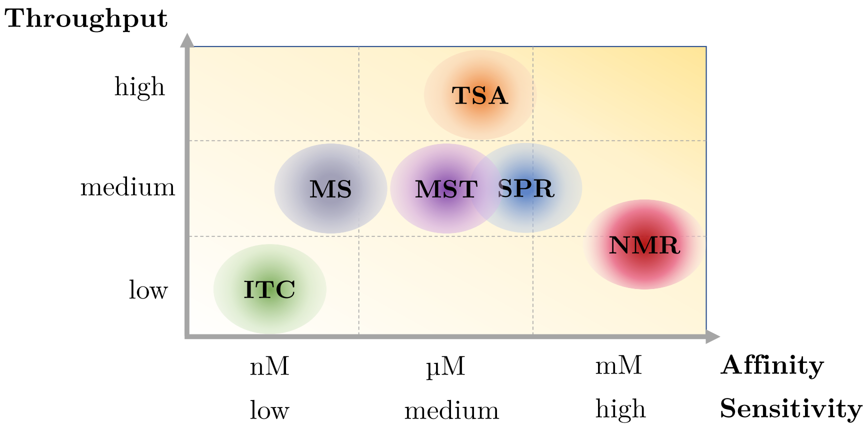


Figure 1.4. Comparison of biophysical techniques in regard to their sensitivity and throughput. Data from[86]

With growing array of primary screening techniques, multiple studies have highlighted that different methods can yield diverging sets of fragment hits against the same target with minor or no overlap.[47,86,89–93] Multiple factors can contribute to the observed discrepancy and need to be considered when comparing screening results.

First, fragment identification methods vary in their experimental requirements: NMR observes protein and ligands in solution, X-ray crystallography requires protein crystals and immobilisation-based techniques such as SPR and WAC typically involve surface capture of the target.[42,47,92] Requirements for preparation of the target can vary significantly between methods.[86] Experimental parameters such as temperature, buffer and pH may also influence hit identification.[86,90,92] In particular, biophysical methods place different demands on compound solubility. The number of applicable techniques for fragments with low solubility is restrained as methods such as NMR or X-ray crystallography require high compound concentrations.[86,90]

Assay-specific artefacts such as non-specific binding, aggregation, fluorescence quenching or solvent mismatch can lead to both false-positives and -negatives, contributing to the discrepancies in hit sets identified by different methods.[26,93]

Further, biophysical techniques differ in their underlying physical principle for detection of binding.[47,89,91,92] For example, TSA experiments identify compounds that stabilise the protein against thermal denaturation, MST monitors change in thermophoresis caused through ligand binding and SPR detects a change in refractive index due to mass increase upon ligand binding.[26,86,89,94] As such, each method uses a different readout as proxy for binding. Whereas measurements of high affinity compounds converge between the different physical principles, weak binders might be missed.

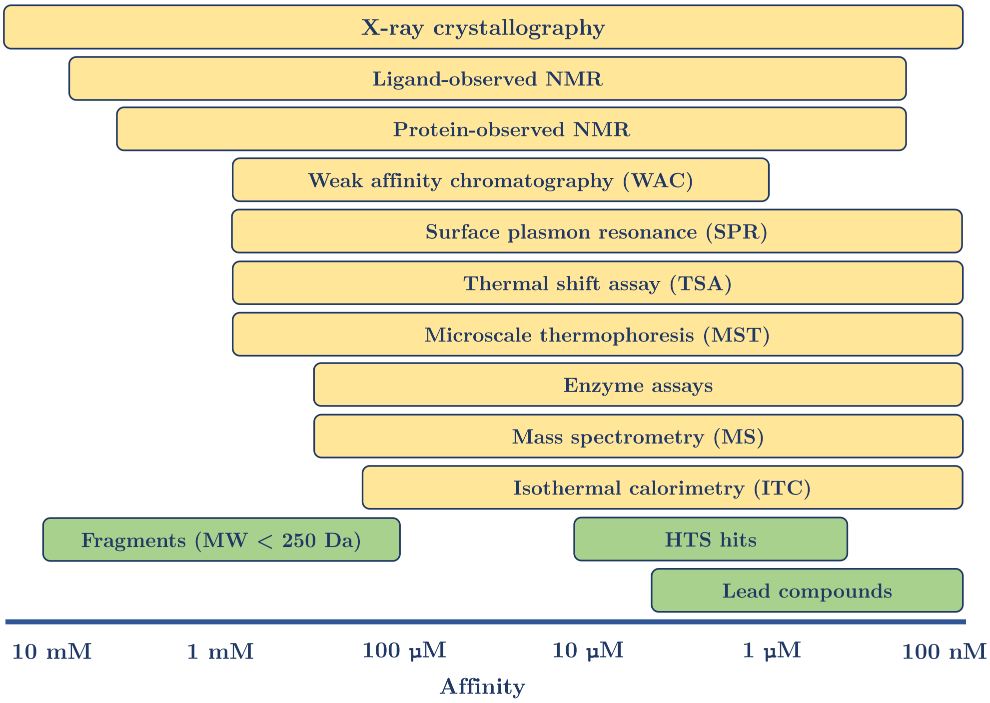


Figure 1.5. Detection sensitivity range of commonly used biophysical techniques for screening. Modified from[10,39,86].

The inherent sensitivity and detection limit of biophysical methods has to be taken into account when comparing screening results.[42] X-ray crystallography and NMR are particularly suitable for detection of weak binders (Figure 1.5).[10,86] However, high concentrations used in crystallographic screening that yield crystal structures of fragments often do not show any detectable binding in other biophysical assays.[26,92]

As illustrated in the study by Wielens and co-workers [90], the lack of overlap may also originate from selection criteria that define a hit. Different filters including threshold values and curve shapes can be applied and fortify the lack of agreement between techniques, particularly when working with very weak binders. [86]

As a result, biophysical screening methods are often applied in a cascade: approaches that typically start with fast and high throughput method to generate a primary set of hits that is further validated using more elaborate and resource-intensive techniques have been described in the literature. [36,86,87,95,96] Other cascades include screening of the same library with orthogonal techniques and consider the intersection of the fragment hits as true hits. [42] Such cascade approaches are likely to eliminate false-positives, but always bear risk of good fragments being discarded due to the least robust method in the cascade. [14,42,89]

Today, many techniques are used to identify fragment hits, each with its own strengths and limitations. [26,93,97] The subsequent section focusses on biophysical methods for screening and quantification techniques used in this work.

1.2.3. Crystallographic fragment screening

In the last decade, X-ray crystallography has matured to a powerful screening method for fragment hit identification. [79,98,99] Despite being one of the first methods applied, crystallographic screening was considered impractical due its highly resource-intensive nature and low throughput. [79,100] Advances in synchrotron technology and automation reduced logistics and increased throughput, rendering accessible the key advantages of the method. [101–103]

Unmatched in its sensitivity, X-ray crystallography provides the knowledge of precise fragment binding orientation and enables rapid assessment of hits in respect to the target, revealing where the binding occurs: in the active or an allosteric site. [35,88,100,102,104–107] The obtained structural information of protein-fragment complexes informs the elaboration process of a fragment into a potent lead compound and is considered the gold standard in drug discovery. [26,35,39] The 3-dimensional visualisation enables the identification of nearby water molecules and conformational changes in the protein structure, which provide insights into the binding landscape. [35,79] Occurrence of false-positive hits due to non-specific fragment binding – a key issue in low affinity screenings – is nearly impossible: binding artefacts at crystal contacts readily be identified. [35,39,88]

Underlying this technological advancement is the tremendous improvement of synchrotron data collection and analysis. Efficient soaking techniques such as acoustic droplet ejection reduced experimental labour and improved the reproducibility of diffraction.[108,109] Increase in X-ray beam intensities, implementation of robotic systems for automated handling of protein crystals, automation of data collection and subsequent processing increased experimental throughput allowing for rapid data analysis.[35,79,88,100,103,110–112] When relying on manual inspection of the electron density maps, identification of fragments bound to the target protein can be challenging and time-consuming.[35] The recently developed PanDDA algorithm allows for enhanced fragment identification even for low occupancy binders (1.2.3.2).[102,113,114] Taken together, over the last decade technical improvements in synchrotron technologies allowed crystallography to become a high throughput method for primary screening, placing higher demands on crystal production.[79,102]

Today, success of a crystallographic screening largely depends on obtaining a suitable crystallisation system.[35,79,100,102,115] The crystallisation procedure for the target of interest needs to be optimised to deliver a highly reproducible crystal form diffracting to the highest possible resolution and containing an accessible binding site.[88,114] Protein crystals need to display tolerance towards organic solvent which are often used to solubilise fragments.[59,79,114] Despite being a laborious and time-consuming step, establishing a robust crystallisation system has significant benefits in the context of a lead discovery project as the same crystallisation system can often be used throughout the process to drive molecular design.[106]

Protein-fragment complexes can be obtained through soaking or co-crystallisation.[79,100,116] In soaking, fragments can access the binding sites by diffusion through solvent channels in already formed crystals.[102,109,117] As all crystals can be pre-grown in the same condition, soaking tends to be experimentally simpler.[109,117,118] Co-crystallisation is often used as alternative when soaking affects crystal stability or fails to identify any binders, particularly compounds with higher molecular weight.[79,119,120] Even though it can accommodate for potential structural changes, co-crystallisation is less widely used for large screenings due to its labour-intensity, lower crystallisation success rates and throughput.[79,118]

Compounds can be introduced as a singleton per crystal or as a compound cocktail. Singleton soaking maximises the effective concentration of each fragment and simplifies the subsequent analysis.[121] To increase throughput compounds are often combined into cocktails consisting of up to ten fragments.[35,88,104,122] Cocktail design is often based on shape diversity to allow efficient deconvolution and minimise the likelihood of multiple fragment binding.[35,100]

The described developments have been incorporated into the XChem fragment screening platform at Diamond Light Source (DLS), which has been used to generate most of the crystallographic data in this work.

1.2.3.1. XChem platform at Diamond Light Source

Established in 2015, the XChem platform at Diamond Light Source (DLS) advanced high-throughput crystallographic fragment screening and offers general access as a user programme.[102,103,121] The highly streamlined workflow has substantially reduced the time from crystal to fragment hit structure and allows for screening of up to 1000 compounds in less than a week.[99,121]

The XChem workflow covers all steps from crystal soaking to automatic data collection and streamlined analysis of multiple datasets (Figure 1.6).[103,121] Starting with an optimised crystallisation system reliably providing highly diffracting crystals, crystal plates are imaged using a standard imager (Formulatrix). Subsequent analysis of the crystal drop images is performed using TeXRank.[123] Selection of suitable crystal drops and marking the location for fragment dispensing is carried out manually using the TeXRank interface.[102] Fragments from a library plate are dispensed to the targeted crystal drop locations using an acoustic liquid handler (Echo, Labcyte).[102,109] Crystal harvesting is supported by the Shifter device (Oxford Lab Technologies) – a motorised x-y-stage equipped with a microscope, while crystal information is tracked using the laboratory management system SoakDataBase (soakDB).[124] Diffraction data is collected at beamline I04-1 in an automated and unattended mode.[125] Data reduction including integration, indexing and scaling is performed by automated pipelines.[128] Subsequent data analysis of autoprocessed datasets is carried out using PanDDA algorithm for weak density interpretation[113], followed by streamlined refinement of protein-fragment structures and their deposition preparation via XChemExplorer[102,126]

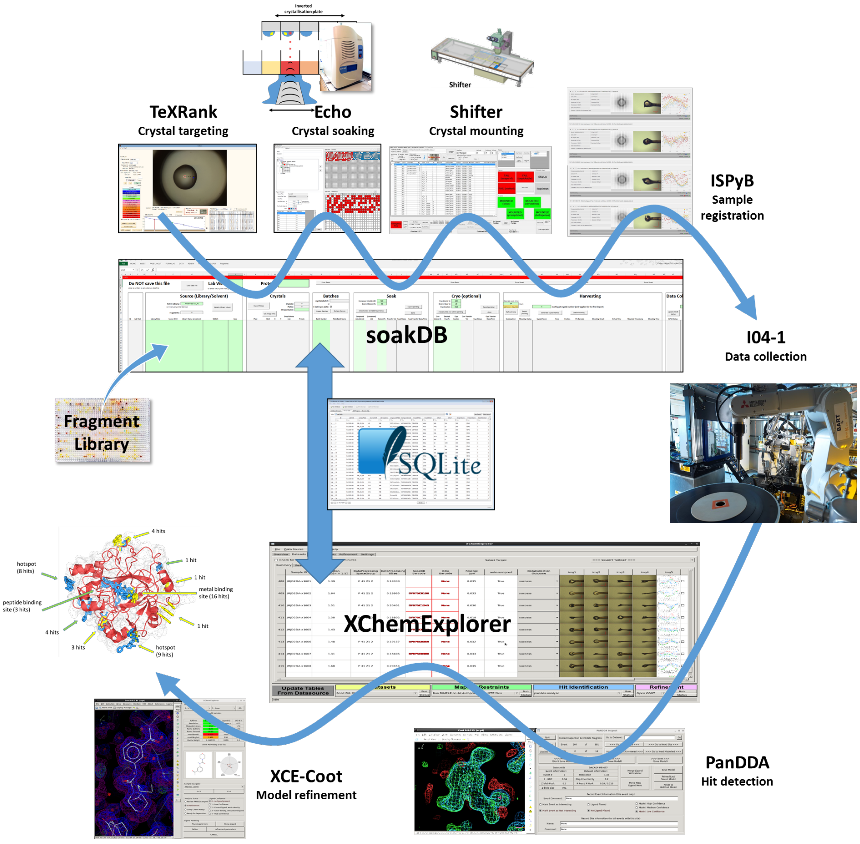


Figure 1.6. Overview of the XChem platform at Diamond Light Source. Crystals are imaged in plates and analysed using TexRank.[123] Location for fragment dispensing is selected manually prior to soaking via acoustic liquid handler[109]. Crystal harvesting is supported by the Shifter device, while all information is tracked using the SoakDataBase.[124] Diffraction data is collected at beamline I04-1 in an automated mode.[125] Data analysis of autoprocessed datasets is carried out using PanDDA algorithm[113]. Streamlined refinement of protein-fragment structures is enabled via XChemExplorer[126]. Reprinted from[127] with permission from © Diamond Light Source.

1.2.3.2. PanDDA algorithm for detection of weak binders

Crystallographic fragment screening typically generates hundreds of datasets to identify fragments that bind to a protein target.[113] The collected datasets often display only minimal differences in electron density through presence of a bound fragment and potential local conformational changes of the protein.[113,117] The inspection of electron density maps for fragment binding evidence is a time-consuming step, with superposition of multiple states due to low ligand occupancies which increase the likelihood of misinterpretation.[100,113,126,129,130] The XChem data processing pipeline at DLS utilises Pan Dataset Density Analysis (PanDDA) for streamlined and improved hit identification.[113,126]

The PanDDA protocol comprises three steps: analysis, inspection and export.[126] In the analysis step, a local alignment of electron density maps in real space is performed to bypass local conformational heterogeneity of the crystals.[113,131] The aligned datasets are voxelised and subsequently calculated distribution of electron density at each voxel reveals the outlying datasets – such as structures with bound fragments (Figure 1.7.a).[113]

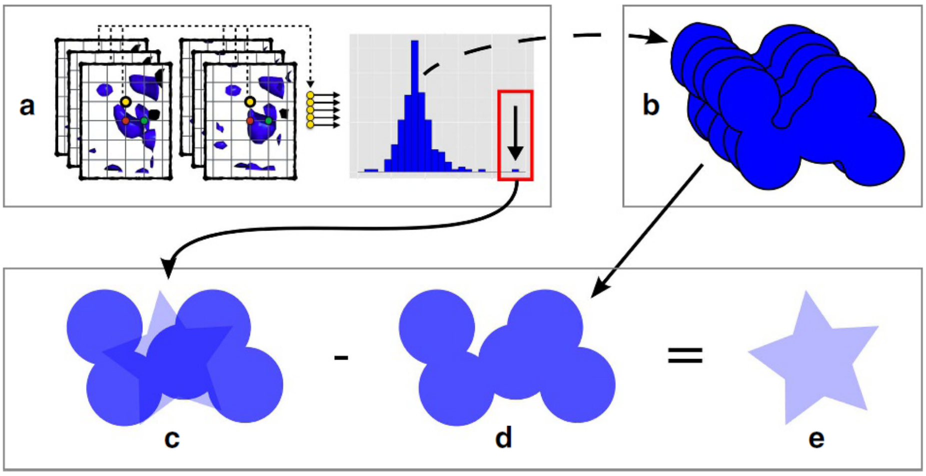


Figure 1.7. PanDDA algorithm allows identification of weak binders. **a.** Following alignment and voxelisation, distribution of electron density at each voxel is calculated to identify outliers. **b.** Multiple ground states are averaged into a mean ground-state map. **c-e.** From superposed state observed in a crystal (**c.**), a proportion of the ground state (**d.**) is subtracted to reveal the bound state (**e.**).[113] Reprinted from Pearce *et al.*[113] under CC BY licence. ©Pearce *et al.*

The datasets that do not contain any outliers, typically obtained by soaking DMSO, are averaged into a mean ground-state map to parametrise the ground-state model (Figure 1.7.b).[113,131] When analysing the screening data, the signal corresponding to the fragment binding event is often observed as a superposition of ground- and bound-state (Figure 1.7.c).[113] A background density correction removes the contribution of the ground state through weighted subtraction of the mean ground-state map by the background density correction factor (BDC) (Figure 1.7.d), allowing for deconvolution of the bound state (Figure 1.7.e):[113]

$$[\text{event map}] = [\text{observed map}] - \text{BDC x } [\text{mean ground-state map}]$$

The inspection step involves user interaction through visual assessment of the event density and modelling of the bound fragments. In the last step, the resulting models undergo a superposed refinement and validation, before they are exported for further streamlined refinement using XChem Explorer.[126,132]

The PanDDA algorithm allows to emphasise the detection of weak binders and has been widely adopted in the FBLD community.[59,76,92,112,133–135] However, its reproducibility is largely dependent on the presence of apo-structures and specialised crystallographic software, leading to an ongoing discussion how to best deposit generated fragment structures in public structure databases.[136]

1.2.4. Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) is one of the most prominent biophysical techniques for identification, validation and characterisation of protein-ligand complexes.[23,96,137–140] A number of key advantages have established NMR as an important tool at different stages of FBDD.[3,139,141] The sensitivity of NMR techniques is well suited to determine affinities of weakly binding compounds such as fragments.[23,142] In fact, NMR is the most sensitive biophysical method providing quantification of ligand binding.[23,143] Additionally, NMR experiments are carried out in solution, preserving full conformational flexibility and steric accessibility of all protein sites.[141] Further, NMR measurements inherently provide information on ligand and protein integrity as well as ligand identity and concentration, revealing degradation and aggregation which compromise measurements.[23,97,141] Most importantly, NMR comprises several types of experiments based on detection of different spectral parameters including chemical shifts, relaxation and diffusion rates which generates a wealth of information on various aspects of molecular interactions.[10,97,137,141,144]

Binding experiments can be divided into protein-observed and ligand-observed methods.[32,141] Pioneered by Fesik and co-workers[48], protein-observed experiments require isotopic labelling of the target protein with NMR-active ^{15}N and/or ^{13}C isotopes. Two-dimensional heteronuclear single-quantum coherence (HSQC) experiments are used to monitor chemical shift perturbations of protein signals upon ligand binding.[32,86,145,146] Ligand titration performed at different concentrations allows to determine dissociation constant of binding (K_D).[14] However, identification of the binding site requires resonance assignment – often a laborious and time-consuming step.[14,32,137,147] Protein-observed methods require large amounts of isotope-labelled protein and can in most cases only be applied to proteins with molecular weight below 40 kDa.[14,86,148]

Ligand-observed methods offer the advantage of not requiring isotopic labelling.[23,32,139]

In addition, they have no upper limit on the size of proteins that can be used − in fact protein with higher molecular weight tend to work better.[23,32] Also, ligand-observed experiments benefit from a lower protein consumption compared to protein-observed techniques.[14,145] Direct ligand-observed NMR methods are particularly suitable for quantification of weak binders (1 mM to 10 μM) which undergo fast chemical exchange between free and bound states, whereas competitive experiments allow for affinity quantification of strong binders (<0.1 μM) in slow exchange regime.[138,142,149] The limitations of ligand-observed methods include absence of information upon ligand location and orientation as well as relatively low throughput compared to other biophysical screening techniques.[148,150,151]

Ligand-observed NMR techniques typically measure parameters of ligands which depend on molecular rotation, these are strongly affected by protein binding. T_2 and $T_{1\rho}$ experiments monitor changes in relaxation properties; changes in cross-relaxation are exploited in STD, water-LOGSY and NOE experiments.[23,32] In the following sections, ligand-observed NMR techniques used in this work are introduced.

1.2.4.1. Saturation Transfer Difference NMR

Saturation transfer difference (STD) is one of the most popular ligand-observed methods used for screening and biophysical characterisation of protein-ligand binding.[138,152] The STD experiment observes free ligand in solution directly and relies on magnetisation transfer between protein and ligand.[141,148,153]

Through selective saturation of protein resonances by a series of high frequency pulses, magnetisation is rapidly spread across the hydrogen network of the protein via spin diffusion.[141,148,154] Any ligand that binds to the macromolecule will receive saturation of its protons as a result of the nuclear Overhauser effect (NOE).[141,148] Upon dissociation, the saturation will be carried with the ligand, resulting in accumulation of saturated ligands in solution.[148,154] Ligands that do not bind to the protein, do not experience any change in their magnetisation, allowing the detection of binders manifested as signal intensity reduction in the ^1H spectrum.[141,148,154]

For an STD experiment, two spectra are recorded: an on-resonance spectrum (I_{SAT}) and an off-resonance spectrum (I_0).[138,152] The on-resonance spectrum (I_{SAT}) is obtained by selective irradiation in the far upfield region of 0 to -1 ppm, which only contains protein resonances of hydrophobic core methyl protons.[138,141,152] The off-resonance spectrum is recorded in absence of protein saturation. The obtained difference spectrum

(I_{STD}=I_0−I_{SAT}) only contains proton signals that correspond to ligands being saturated as a result of binding to the protein (Figure 1.8).[138,152] Due to distance dependency of the saturation, only protons that are in close proximity to the macromolecular target, will appear in the difference spectrum.[152] The relative magnitude of the signals correlates approximatively with proximity to the target, providing information on the ligand position through epitope mapping.[32,138,154]

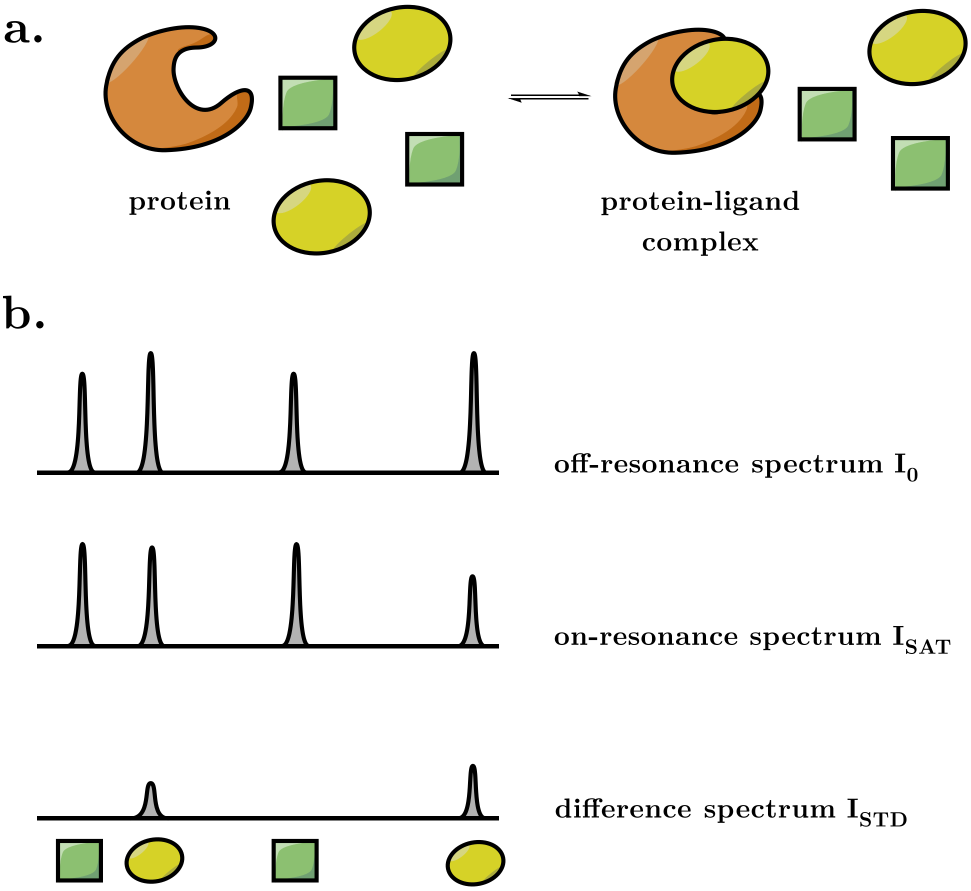


Figure 1.8. Schematic representation of the STD experiment. **a.** Upon saturation protein magnetisation is transferred to the ligand by spin diffusion. The exchange between free and bound state of the ligand results in a progressive saturation of the ligand pool (yellow), whereas non-binders do not experience any change in magnetisation (green).[148] **b.** Hit identification is carried out through subtraction of the on-resonance (I_{SAT}) from the reference (I_0) to obtain the difference spectrum (I_{STD}), which only contains signals of ligands that bind to the target. Modified from[148,152].

In theory, ligand affinity can be determined by STD through monitoring the change in STD response with increasing ligand concentration.[148] For quantification, STD amplification factor STD^{AF}, which scales the measured STD response according to the ligand excess employed in the experiment, is calculated and plotted as a function of ligand concentration.[148,149] However, K_D values determined by STD tend to be an overestimation due to strong influence of experimental parameters such as saturation time.[148,155] Indeed, a recent study reports no correlation between affinity and STD response.[154]

Accurate affinity determination by STD-NMR can be achieved as a competition experiment, if a binder with a known K_D is available.[148,156–158] The signal of the spy molecule is monitored for change in signal intensity upon displacement through a competing molecule.[148] Titration of spy molecule at constant concentrations of protein and competing ligand allow for determination of an apparent $K_{D,app}$ which is necessary to obtain the true K_D of the competing ligand.[148] Alternatively, the approach described by Dalvit and co-workers[159] yields K_D by recording the spectra with different concentrations of the competing molecule at constant spy concentration and utilising the displacement values.

In absence of a known binder, accurate measurements can be obtained using the protocol proposed by Angulo and co-workers[155]. As described by the authors[155], the main limitation of direct STD titration for affinity measurement is caused by ligand rebinding during the saturation time.[148,155] Any ligand that has been bound to the macromolecule, carries some degree of saturation back into solution. Upon rebinding of the same ligand, a reduced level of saturation is transferred, resulting in an attenuation of the STD response.[148,155] Whereas it does not play a role in screening where ligand excess is used, titrations with low ligand concentration necessary for affinity determination favour rebinding, leading to inaccurate K_D.[148,155] As suggested in the study[155], rebinding effect can be avoided by using initial growth rate factors (STD_0^{AF}) − the STD responses at the limit of zero saturation time when virtually no ligand binding occurs − instead of amplification factors (STD^{AF}). The lengthy process of determining STD_0^{AF} involves recording of STD^{AF} at a range of different saturation times and subsequent fitting for each concentration.[148] Even though it provides accurate determination of dissociating constants, this approach may not be practical for a large number of ligands.[148]

1.2.4.2. Relaxation-based experiments

Relaxation effects depend on molecular motion which is characterised by the rotation correlation time τ_c.[141] Due to rapid tumbling in solution, which corresponds to short τ_c, small molecules exhibit long relaxation times T_2.[138,139] As relaxation time T_2 is inversely proportional to the line width at half height, small molecules exhibit narrow line widths in the NMR spectrum.[138] Large molecules such as proteins are characterised through short T_2 and high transverse relaxation rates R_2 ($T_2 = 1/R_2$), which correspond to broadened line widths.[138,141,144,148] Upon binding, rotation of the ligand is decreased, as the small molecule transiently adopts rotation properties of the macromolecule.[141,142,148] The observed ligand line width in the NMR spectrum is a result of exchange-average between bound and free states.[142,148] With sufficient contribution of the bound state, line broadening is observed.[148]

When screening weakly binding ligands such as fragments, the observed effect can be rather small.[148] Binding detection can be enhanced by applying filters that attenuate signals of the bound species.[148] Sharp resonances remain unaffected as a result of slow transverse relaxation of free small molecules.[148] Such relaxation-editing allows for quantification of intensity reduction instead of measuring line width directly.[138,142,148] Both transverse relaxation parameters T_2 and $T_{1\rho}$ can be exploited for detection of ligand binding through intensity reduction.[159] In T_2, a Carr-Purcell-Meiboom-Gill (CPMG) spin echo filter is applied.[141] For experiments measuring the spin-lattice relaxation $T_{1\rho}$, a continuous spin-lock is utilised that holds magnetisation static in the rotating frame as it relaxes.[148,160,161]

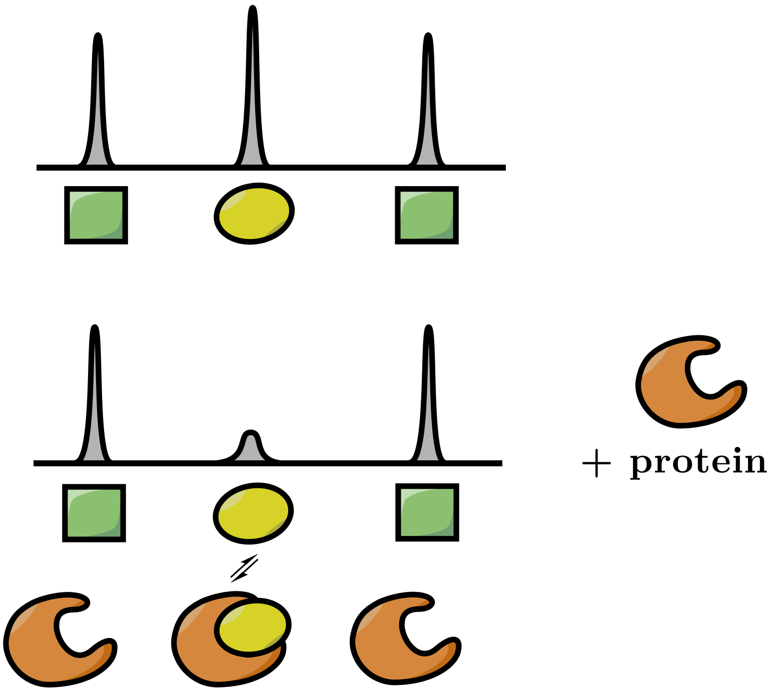


Figure 1.9. Schematic representation of a relaxation-editing experiment. Signal intesity of the ligand that binds to target protein is significantly attenuated in relaxation-editing experiments T_2 and $T_{1\rho}$. Modified from[142].

The relaxation experiments are recorded in presence and absence of protein.[141] Typically, two spectra with different relaxation periods are recorded on the same sample: one with a short (e.g. 20 ms) and the other with a long relaxation delay (100-400 ms).[141] Ligands that interact with the target protein display faster attenuation in signal intensity at long relaxation periods.[142,162,163] A comparison of spectra reveals binding events, which are manifested as a reduction in signal intensity of the fragment resonances (Figure 1.9).[139,141,142] A control experiment only containing the fragment cocktail in buffer serves as a reference spectrum.[23]

Similarly as in STD experiments, quantification measurements can also be performed in competition mode (1.2.4.1).[148,159] While ligand-observed NMR methods provide sufficient sensitivity for KD determination of weakly binding compounds, their application is resource-intensive in regard to throughput and expertise (see chapter 5).

1.2.4.3. ^{19}F NMR-based fragment screening

Screening of libraries with fluorinated fragments has gained popularity as primary screening method for identification of hits and rapid assessment of ligandability of new targets.[120,164,165] Fluorinated ligands are widely used in medicinal chemistry, with 25% of drugs approved containing at least one fluorine atom.[75,141,165]

A number of features makes ^{19}F an attractive nucleus for fragment screening.[141] With its spin 1/2, ^{19}F is a stable NMR-active nucleus with 100% isotope abundance.[164,165] ^{19}F has a comparable sensitivity to proton (83.4% relative to ^{1}H), which is enhanced by the absence of natural fluorine background in most biological samples.[138,149,164,166] Large chemical shift dispersion covering several hundreds ppm and narrow line width for ^{19}F resonances offer ideal conditions for screening of multiple ligands simultaneously, typically combined into cocktails of 20-30, without observing signal overlaps.[138,141,149] Screening of fragment mixtures enables rapid data collection with low protein consumption.[167–169] The large dynamic range − the ratio of observable parameters such as chemical shift in presence and absence of protein − is particularly suitable for detection of weakly binding ligands.[165,167]

Transverse relaxation (T_2) is the most common ^{19}F parameter utilised for fragment screening and is exploited in CPMG experiments (see 1.2.4.2).[165] Chemical shift (δ) is another indicator of binding.[141] Upon ligand binding, chemical shifts of both protein and ligand are changed as a result of local magnetic field perturbations that both particles experience.[141] However, as only a small population of the ligand is bound, the chemical shift is often less pronounced in a screening set-up with large ligand excess.[141]

Analysis of fragment screening data of different ligand mixtures can be a laborious process.[23] However, the flat baseline of the spectra, large chemical shift dispersion of the signal and their narrow line width are well suited for automation.[138,165] Multiple tools including Fragment-Based Screening (FBS, Bruker)[170], and MNova Screen (MestreLab Research)[171], have been implemented to facilitate streamlined data analysis.

In addition to direct mode, ^{19}F screening can also be performed in competition mode.[138,165] The use of a fluorinated spy molecule enables rapid binding detection through monitoring ^{19}F resonance of the spy.[138,142,172,173]

Protein-observed methods can also be utilised for fragment screening.[165] Protein-observed fluorine NMR (PrOF) requires selective labelling of the protein target with fluorinated amino acids or site-specific labelling, enabling detection of ligand binding through observation of changes in ^{19}F chemical shift.[142,165]

1.2.5. Surface Plasmon Resonance

Surface plasmon resonance (SPR) is one of the most widely used biophysical methods in fragment-based drug discovery.[120,174–176] SPR has often been used in lead optimisation as an orthogonal technique for validation and affinity determination of fragments as well as elaborated compounds.[13,81,175,177] In the last decade, its application as primary screening technique has emerged.[178]

SPR is an optical biosensor method which detects the change in refractive index that follows a mass increase resulting from a binding event between a protein target immobilised on the biosensor chip surface and a ligand in solution (Figure 1.10).[10,179] The change of the refractive index depends on the ratio of the ligand to the protein mass.[26]

Target proteins can be immobilised on the chip surface covalently through amine, thiol or aldehyde couplings or non-covalently via affinity tags such as streptavidin-biotin, monoclonal antibody tags or metal chelation.[10,14,175,178,179] Microfluidic systems allow injection of varying compound concentrations onto biosensors.[26,86] With continuous signal registration, detection of binding response on the chip surface occurs in real time.[86,120,176]

SPR measurements allows accurate quantification of kinetic parameters including association (k_a) and dissociation rate (k_d), affinity (K_D) and stoichiometry of the interactions.[86,96,120,176,179] Additionally, thermodynamic parameters of the interactions can be determined by performing measurements at multiple temperatures.[86,178] Further advantages of SPR include low sample quantities and protein consumption as well as relatively high throughput allowing to test approx. 500 compounds per day in a screening set-up.[42,176,179]

Successful quantification of binding interactions by SPR requires a careful assay development and implementation of quality controls.[179] Optimal screening conditions to maintain the target protein functional and to provide low signal-to-noise ratio are often identified through systematic variation of buffer components, detergents, pH and temperature.[23,96,175] Reference compounds with known interaction characteristics can be used to monitor protein functionality and stability on a biosensor surface over time.[23,80,180]

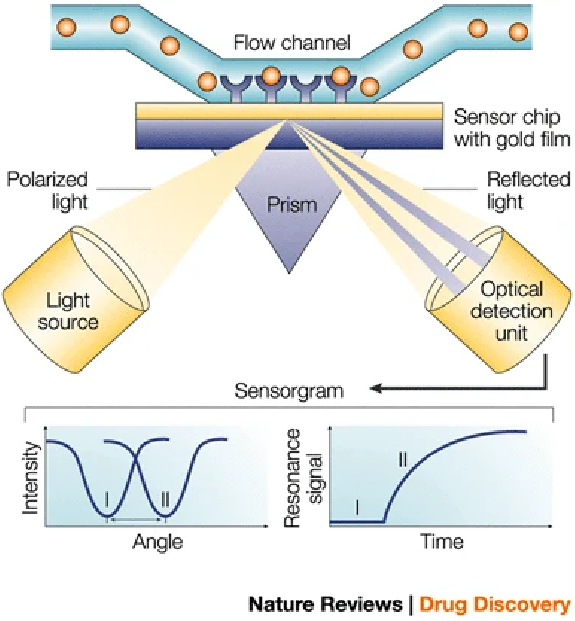


Figure 1.10. Schematic illustration of a typical SPR biosensor set-up. A beam of polarised monochromatic light is irradiated through a prism at the gold-coated surface of the sensor. The reflection is detected by an optical unit. At a certain angle, the resonance angle, light is absorbed by the electrons in the metal, creating resonating surface plasmons (collective oscillation of free electrons). This creates a dark line in the reflected beam, a dip in light intensity at the resonance angle (**I**). Any compound that binds to the immobilised target increases the mass near the metal surface, resulting in change of the refractive index, and a measurable shift in the resonance angle (**II**). This change is observed in real-time and recorded as a sensorgram.[14,174] Reprinted from[174] with permission from © Springer Nature.

For novel targets, these are often not available requiring an orthogonal method for validation.[80] Reference proteins immobilised on parallel surfaces help to deconvolute non-specific protein interactions.[80,178]

The current generation of biosensor instruments is capable of detecting small molecule binding with molecular weights as low as 50 Da.[179] However, measuring weak binders with low molecular weight such as fragments is challenging due to the small change in the refractive index. Signal amplification can be achieved through immobilisation of active and functional protein at high density on the biosensor surface (see 4.2.2).[178,179]

The minimal detection level also depends on the signal-to-noise ratio. While current biosensor instruments can obtain quality responses below 1 response unit (RU), some compounds have a high refractive index which contributes to the noise level by creating a mismatch between blanks and samples.[178,179] Screening compounds are usually solubilised in dimethylsulfoxide (DMSO) which has a high refractive index and can be a major source of false-positives, if the DMSO concentration in both samples and running buffer is not matched accurately.[81,175]

Affinity quantification of weak binders such as fragments ideally requires screening at significantly higher concentrations than their affinities to saturate the binding sites of the target protein.[179] In practice, this is often not feasible due to limited solubility of compounds at high concentrations and aggregation effects, making deconvolution of responses from specific and non-specific interactions difficult.[10,14,23] Therefore, signal responses of weak binders are *de facto* measured at concentrations far below their affinities.[80,178] When measuring weak binders, binding kinetics are rarely observed due their fast association and dissociation rates, only allowing for steady-state affinity analysis (see 4.2.2).[26,80]

With its quantitative nature and high throughput, SPR is an ideal method for fragment screening, binding validation and kinetic characterisation of elaborated compounds.[14] Its use for orthogonal validation of weak crystallographic fragment hits can be limited by their low affinity and requires careful assay optimisation, as will be further discussed in this thesis.

1.2.6. Fragment elaboration

Early success of FBDD has led to a paradigm shift in small molecule drug discovery: based on the experience with HTS, compounds with affinity above 1 µM were considered too weak for progression. Nowadays, many medicinal chemistry optimisation campaigns routinely begin from low affinity molecules (10 µM-1 mM).[43,181,182]

The elaboration cycle of fragments comprises an iterative process of design and synthesis of follow-up compounds followed by biophysical quantification to monitor the progress.[36] Three conceptual strategies to increase potency of compounds derived from fragments have been described: fragment linking, growing and merging[13,36,181,183] (Figure 1.11).

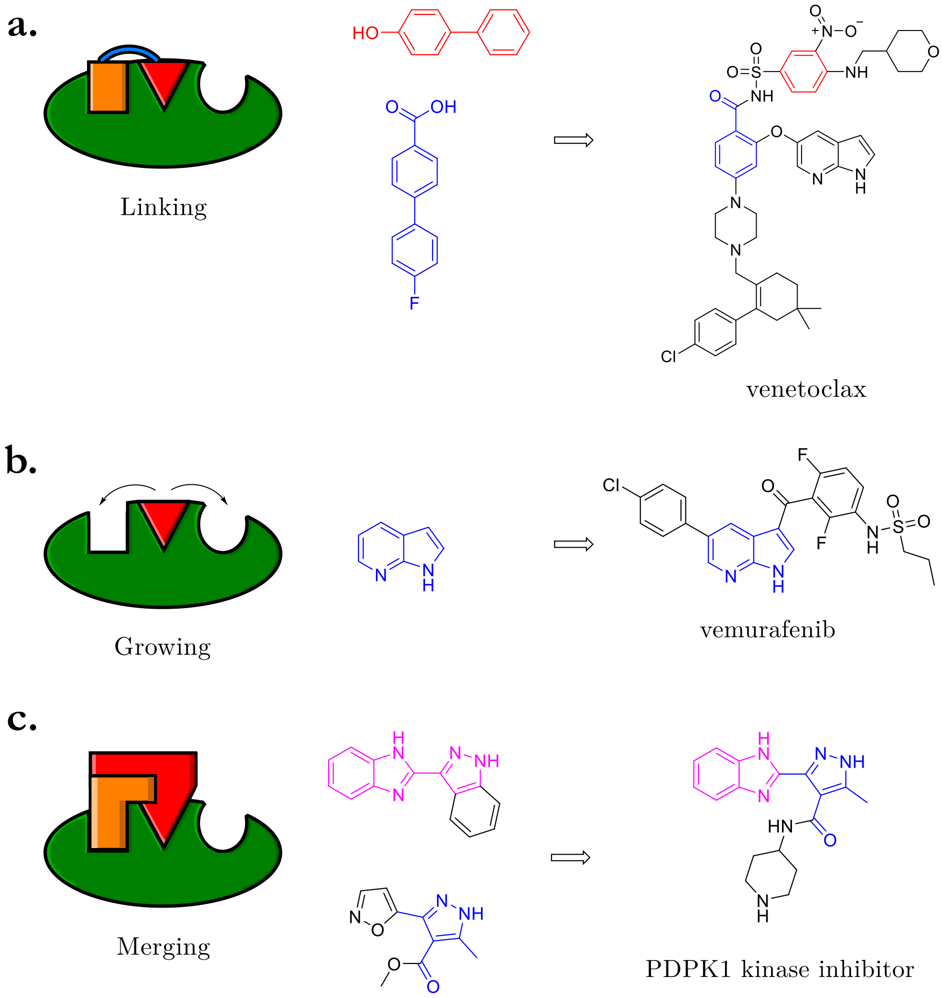


Figure 1.11. Conceptual strategies for fragment elaboration. **a.** Linking. Two fragment hits identified by NMR were linked and optimised to obtain venetoclax, a selective Bcl-2 inhibitor. **b.** Growing. The initial fragment hit identified in a biochemical screen was grown to vemurafenib, a selective inhibitor of B-Raf kinase. **c.** Merging. Both fragments identified in a SAR by catalogue approach based on a fragment hit from ligand-observed NMR were merged and optimised to the PDPK1 inhibitor. Modified from[42,184].

Fragment linking was the first approach successfully applied in a FBDD campaign.[48] This strategy entails linking together two fragments that bind at adjacent non-overlapping sites on the target protein surface.[24,47,48,185] In theory, successful linking of two fragments would result in a compound with affinity greater than the sum of individual fragments.[186,187] Although conceptually very attractive, this technique requires a binding site of appropriate size to accommodate two fragments.[42] Additionally, efficient joining without disruption of the binding mode of the fragments is necessary, placing demands on the length and rigidity of the linker.[36,42,47,188,189] Venetoclax, a drug approved in 2016 for treating certain forms of chronic lymphocytic leukaemia, demonstrates a successful application of the fragment linking approach (Figure 1.11.a)[190]

Fragment growing is the most widely applied method for elaboration of fragment hits.[42] Growing aims to increase the size of a fragment by adding functional groups to establish additional interactions with the target to achieve the required affinity and selectivity.[10,24,189] The directionality of chemical functionalisation requires identification of potential growth vectors.[189] The availability of structural data from crystallography or NMR is crucial for guiding the elaboration process.[42,189] In the development of the selective inhibitor of V600E mutant B-Raf kinase vemurafenib – the first approved drug originating from FBDD – fragment growing approach was applied (Figure 1.11.b).[13,42,191]

Fragment merging involves combination of structural features of overlapping fragments into a compound.[36,42] This approach can be attempted when overlapping functionalities share similarity in their structure or interaction profile.[189] As in fragment linking, the merged compound would ideally combine the affinities of initial fragments additively or even synergistically.[189] For this technique, fewer examples can be found in the literature.[42] The development of selective inhibitor for PDPK1 kinase at Vernalis demonstrates how fragment merging can be successfully utilised (Figure 1.11.c).[10,42,192]

Prior to initiation of the fragment elaboration process, it can be useful to explore the cores of identified fragment hits through purchase or limited synthesis of similar chemotypes.[42,112] Exploration of structure-activity relationships (SAR by catalogue) is a fast and powerful method for initial mapping of the binding site and probing how small changes in the chemical structure can affect binding.[42,46,183]

Although examples for fragment optimisation purely guided through biophysical and computational methods can be found in literature, structural knowledge of precise fragment binding mode is paramount for the progression.[18,36,46,112,120,193] Typically, this is obtained by X-ray crystallography.

Fragments are considered efficient binders as they form high quality interaction.[39,41,52] Fragment elaboration needs to ensure that additionally introduced functionalities contribute to the affinity gain and thus justify the increase of heavy atoms and in size of the molecule.[47] The ligand efficiency (LE) metric is widely utilised to rank fragment hits and to assess their optimisation.[47,194] LE is defined as the ratio of the free energy of binding to the number of heavy atoms (HAC) in the molecule and is instructive to monitor how this ratio changes while the fragment is undergoing optimisation.[195,196]

In addition to potency, fragment elaboration offers the opportunity to control and optimise a number of physicochemical properties including molecular weight, solubility, lipophilicity, selectivity and absorption, distribution, metabolism and excretion (ADME) at an early development stage into a drug candidate or a chemical probe.[183,197]

1.3. NUDIX hydrolase NUDT5 as model system

In this work, human NUDIX hydrolase NUDT5 was used to generate all datasets for method development. NUDT5 is a drug target of the Structural Genomics Consortium (SGC). The biological relevance and target rational are outlined in the subsequent section.

1.3.1. NUDIX hydrolase superfamily

The human NUDIX superfamily consists of 22 enzymes which catalyse the hydrolysis of nucleoside diphosphates linked to a variable group (X).[198–200] Their substrate preference displays high diversity and promiscuity, and includes canonical (d)NTPs, oxidised (d)NTPs, nucleotide polyphosphates and sugars as well as capped mRNA.[201,202] All members of the superfamily share a highly conserved 23-residue motif, which is often described as the NUDIX box: $GX_5EX_7REUXEEXGU$, with U being one of the bulky hydrophobic amino acids I, L and V.[198,203] This motif forms a loop-helix-loop structure and is involved in catalysis as well as binding of essential divalent metal ions.[201,202]

Historically, NUDIX hydrolases have been associated with the role of "cell sanitisers"; as the founding member of the superfamily, the enzyme MuT (NUDT1), demonstrated antimutagenic properties by hydrolysing oxidised deoxynucleoside triphosphates and prevented mutations.[203,204] More recently, the enzymes of the NUDIX superfamily have been linked with a broad variety of processes in cellular metabolism, suggesting that only a few members play a role in the prevention of DNA mutations.[200,201,205,206] NUDIX hydrolases are highly conserved and ubiquitously expressed in all species; some functional redundancies among the members have been described.[206] Despite first biochemical characterisation of the NUDIX enzymes carried out in early 1990s, some of the superfamily members have not been characterised and the biological function of many remains unknown.[205,206]

Several NUDIX hydrolases are upregulated upon cellular stress, suggesting they may be relevant for survival of the cells under stress conditions.[200,207–209] Among others, the enzymes NUDT1 and NUDT5 are indeed overexpressed in different cancer types and were identified as promising therapeutic targets.[200,210,211]

1.3.2. NUDT5 – Target information

1.3.2.1. Enzymatic activity of NUDT5

The human hydrolase NUDT5 (NUDIX5), a member of the NUDIX hydrolase superfamily, has been linked to key processes in nucleotide metabolism and cancer.[200,212,213] The enzyme hydrolyses nucleotide phosphates from a wide range of substrates, including ADP-sugars and nucleotide phosphates.[214–217] The main reactions catalysed by NUDT5 are summarised in Figure 1.12.

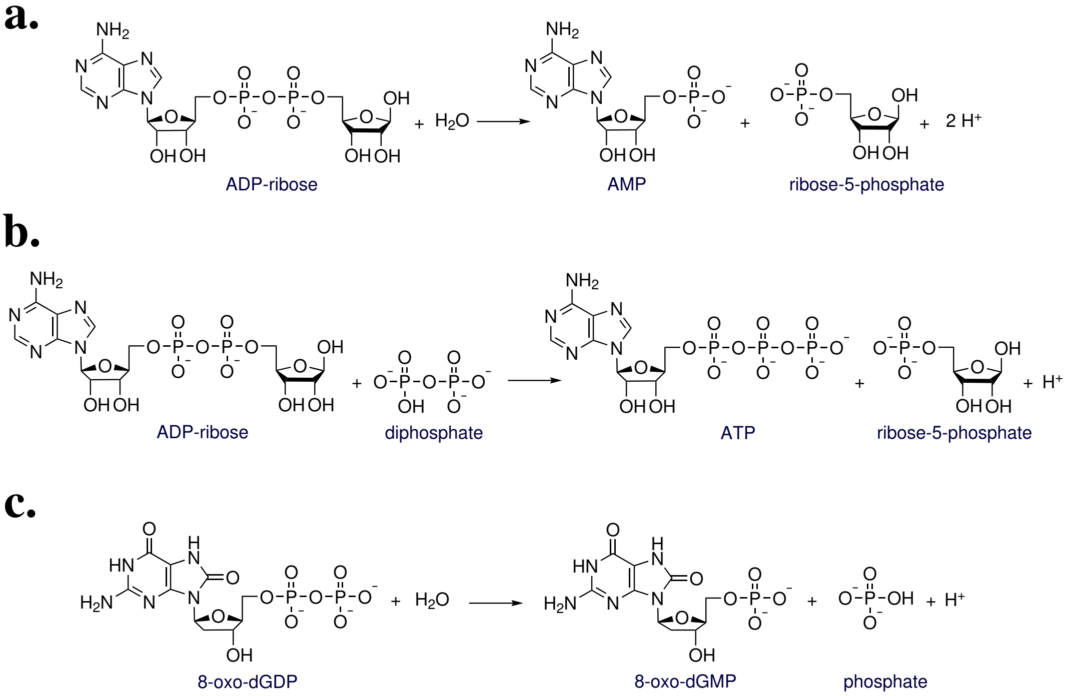


Figure 1.12. Catalytic activity of NUDT5. **a.** The main catalytic reaction of NUDT5 is hydrolysis of ADP-ribose into AMP and ribose-5-phosphate.[200,217] **b.** In presence of diphosphate, NUDT5 catalyses conversion of ADP-ribose to ATP and ribose-5-phosphate.[213] **c.** NUDT5 is capable of hydrolysing 8-oxo-dGDP into 8-oxo-dGMP and phosphate.[200]

NUDT5 as ADP-sugar pyrophosphatase

NUDT5 was first characterised as an ADP-sugar pyrophosphatase.[214] Although the enzyme has been reported to hydrolyse a variety of ADP-sugars, ADP-ribose has been identified as predominant substrate for NUDT5.[200,214] NUDT5 catalyses the hydrolysis of ADP-ribose into adenosine 5'-monophosphate (AMP) and ribose-5-phosphate (R5P) (Figure 1.12.a) and plays an important role in controlling intracellular ADP-ribose levels.[199,217,218] Generated by turnover of NAD$^+$ and cyclic, polymeric or protein-bound ADP-ribose, free ADP-ribose is a highly active metabolite which causes non-enzymatic ADP-ribosylation at high concentrations.[199,219] ADP-ribosylation of proteins is a post-translational modification that influences various cellular processes, including regulation of mitosis, gene transcription, proliferation, differentiation and apoptosis.[220]

NUDT5 as key enzyme in nuclear ATP synthesis

A recent study by Wright and co-workers[213] linked NUDT5 to synthesis of ATP in cell nuclei. The outlined non-canonical pathway describes the rapid generation of ATP to support the activity of energy-consuming enzymes involved in chromatin remodelling, DNA replication and gene regulation upon DNA-damage response.[213,221] As a consequence of the stress response, hydrolysis of transiently increased poly(ADP-ribose) generates free ADP-ribose which can be then converted by NUDT5 to ATP in presence of PP$_i$ (Figure 1.12.b).[213,221]

Whereas hydrolysis of ADP-ribose into AMP and R5P is energetically favourable, the study suggests that endergonically synthesised ATP is only generated in presence of increased ADP-ribose and PP$_i$ concentration and possibly posttranslational phosphorylation of NUDT5.[213]

NUDT5 as nucleotide pool sanitizer

Kamiya and co-workers[215] have reported that NUDT5 also displays a broad substrate specificity for oxidised desoxyribonucleoside diphosphates with highest efficiency for 8-oxo-dGDP (Figure 1.12c), indicating its role in elimination of oxidised DNA precursors from nucleotide pool to prevent replicational and translational errors.[215,217,222] Upon hydrolysis, 8-oxo-dGMP and phosphate are produced (Figure 1.12.c).

Although NUDT5 is capable of 8-oxo-dGDP hydrolysis, a recent study[200] has demonstrated that NUDT5 plays only a limited cellular role in 8-oxo-guanine metabolism due to poor hydrolysis efficiency under physiological conditions.[216,217]

1.3.2.2. Structural analysis

The first crystal structures of full-length NUDT5 in *apo* form and truncated form in complex with ADP-ribose and AMP have been reported by Zha and co-workers[199] in 2006. In the crystal structure, two NUDT5 monomers form a homodimer. As shown in Figure 1.13, each monomer can be divided in three domains: the N-terminal (green, residues 14-53), the catalytic (blue, residues 54-210) and the C-terminal domain (grey, residues 211-214).[199]

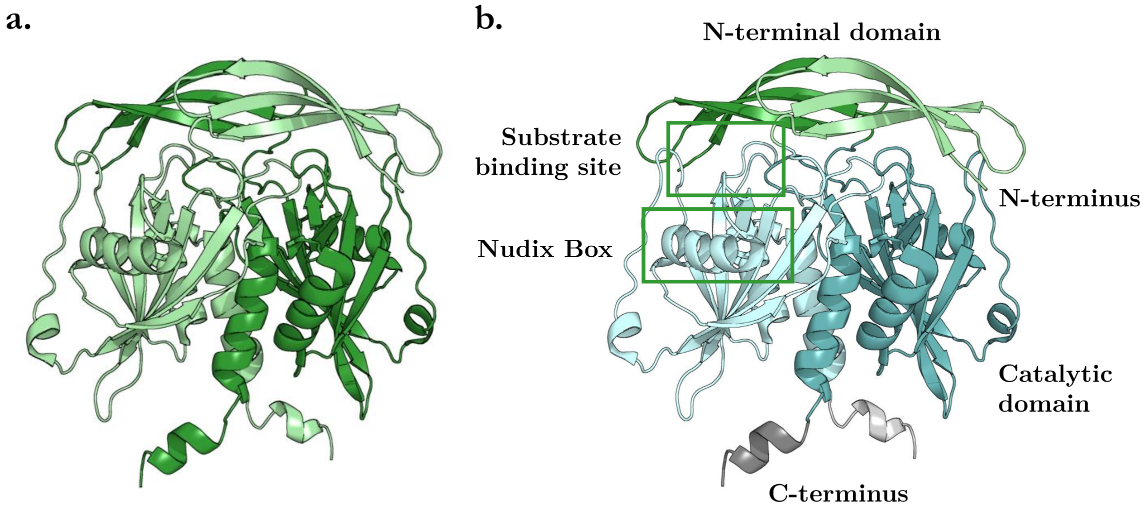


Figure 1.13. Crystal structure of the apo form of NUDT5 (PDB:2DSB). **a.** The NUDT5 enzyme forms a homodimer with substantial domain swapping. Each monomer is shown in light and dark green, respectively. **b.** Structural classification of a NUDT5 dimer: the residues 14-53 of each monomer form the N-terminal domain (shown in light and dark green), the residues 54-210 form the catalytic domain (shown in light and dark cyan), the residues 211-214 belong to the C-terminal domain (shown in light and dark grey). Adapted from[199].

The homodimer shows a substantial domain swapping as the N-terminal domain of the first monomer crosses over to interact with the N-terminal domain of the second monomer.[199] This domain swapping is stabilised by hydrophobic interactions as well as salt-bridges and hydrogen bonds.[199]

Substrate binding The substrate binding occurs at the dimer interface between the catalytic domain of first monomer and the N-terminal domain of the second monomer.[199] As shown in Figure 1.14, in the co-crystal structure of NUDT5 and ADP-ribose the adenine moiety is sandwiched between two tryptophan residues Trp28 and Trp46 originating from different chains. Compared to the apo-form, both residues Trp28 and Trp46 move closer together to allow π-π stacking with the aromatic adenine core of ADP-ribose.[199] The N1 and N6 atoms of the adenine core are specifically recognised by the amine nitrogen and the carbonyl oxygen of the conserved Glu47 backbone via hydrogen bond interactions.[199] Additional hydrogen bond is observed between N7 atom of the adenine core and the guanidinium group of Arg51 side chain.[199] The hydroxyl groups of adenosyl ribose do not show any direct interactions with the protein and only form water-mediated interactions.[199,220] The negative charge of the pyrophosphate is partially neutralised by the Arg84 side chain.[220] The terminal ribose is stabilised through hydrogen bonds between hydroxyl groups and the Asp133 side chain, and Gly135 hydrogen bond with 2'-OH-group of the ribose.[199,220]

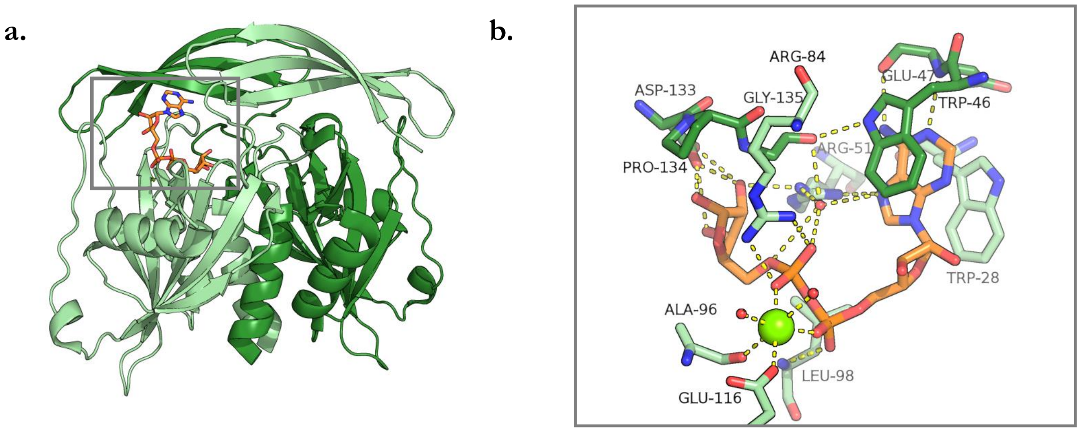


Figure 1.14. Crystal structure of NUDT5 in complex with ADP-ribose (PDB:2DSC). **a.** ADP-ribose (shown in orange) is located at the substrate binding at the dimer interface. The active site is formed by the catalytic domain of the first monomer (shown in light green) and the N-terminal domain of the second monomer (shown in dark green). **b.** Coordination of ADP-ribose inside the binding site: the adenine core of ADP-ribose is bound between Trp28 and Trp46 via π-π stacking interactions. In addition, the positions N1, N6 and N7 of the adenine core are stabilised by hydrogen bonds. Mg^{2+} ion is shown as a bright green sphere, water molecules as red spheres; hydrogen bonds as broken lines in yellow. Adapted from[199].

As observed in crystal structures in complex with AMP and non-hydrolysable analogue of ADP-ribose, α,β-methyleneadenosine, similar substrate coordination is retained, although the number of magnesium ions varies between 1 and 3.[199,220]

Mutagenesis and subsequent kinetic studies identified the key residues which are necessary for substrate binding and catalysis: Trp28 and Trp46 are crucial for recognition of adenosine moiety, Arg51 is required for both substrate binding and catalysis.[220] For the recognition of the phosphates and the terminal ribose, both Arg84 and Leu98 are important, with Arg84 affecting catalytic activity significantly and emphasising its role in charge neutralisation of the β-phosphate. As demonstrated in mutation studies, residues involved in coordination of the terminal ribose are less important for substrate recognition.[220]

The catalytic function is dramatically impaired if residues of the NUDIX box Glu112 and Glu116 directly involved in metal coordination are mutated, highlighting the importance of the metal ions for the enzyme catalysis.[220] Additionally, Glu166 was identified as another key residue necessary for catalysis. The authors suggest Glu166 acts as a catalytic base to deprotonate a conserved water molecule which then initiates the nucleophilic attack.[220]

Suggested catalytic mechanism Based on the structural and biochemical data, Zha and co-workers[199] suggest a catalytic mechanism of ADP-ribose hydrolysis in the presence of three magnesium ions and provide the structural basis for high specificity for ADP-sugars (Figure 1.15). In the active site, the phosphate moieties of the ADP-ribose are stabilised by the guanidinium group of Arg84 and a positively charged magnesium ion, respectively. A conserved water molecule is deprotonated by Glu166 acting as a catalytic base, and is further polarised by the remaining magnesium ions.[220] Acting as a hydroxyl ion, the oxygen performs a nucleophilic attack on α-phosphate of ADP-ribose resulting in its hydrolysis into AMP and ribose-5-phosphate.[220]

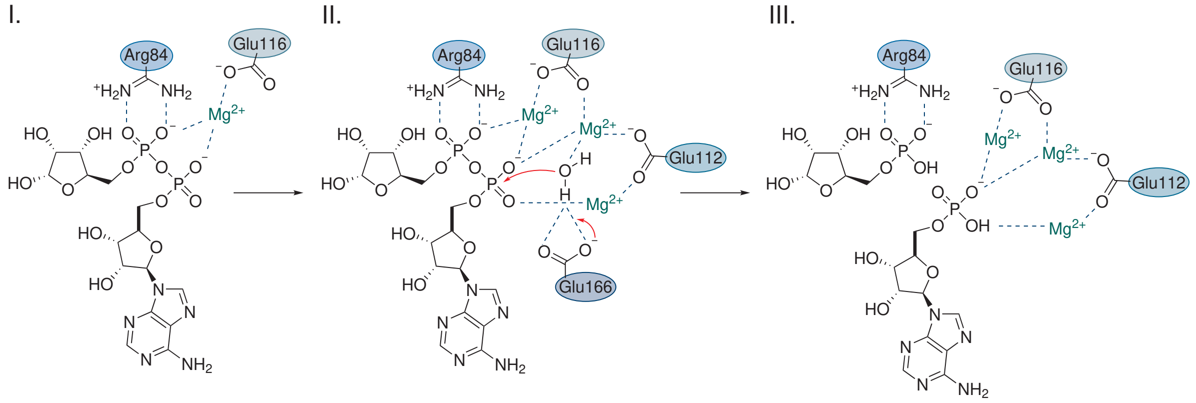


Figure 1.15. Catalytic mechanism of ADP-ribose hydrolysis by NUDT5. **I.** The negative charges of the α- and β-phosphates of ADP-ribose are neutralised by the Arg84 side chain and magnesium ions in the active site, respectively. **II.** Deprotonated by Glu166, which acts as catalytic base, and stabilised by two magnesium ions, a conserved water molecule performs a nucleophilic attack on α-phosphate of ADP-ribose. **III.** The nucleophilic attack generates the hydrolysis products ribose-5-phosphate and AMP. Adapted from[220].

Arimori and co-workers[216] reported first NUDT5 crystal structures in complex with 8-oxo-GDP and identified that 8-oxo-dGDP displays an inverted binding mode compared to ADP-ribose. As a consequence, the guanine core is stabilised via a different hydrogen bonds pattern: Glu47 backbone interacts with positions N1, N2 and O6; Arg51 sidechain with N3. In addition, the positions of a α and β-phosphates are inverted. Using isotope-labelling they confirmed that hydrolysis is carried out by a nucleophilic attack at β-phosphate.[216]

1.3.2.3. NUDT5 as therapeutic target

NUDT5 overexpression is correlated with poor prognosis, increased risk of recurrence and metastasis in several cancer types, including breast cancer.[213,223–225] In a recent study[223], the NUDIX hydrolyse NUDT5 was linked to development of a more aggressive disease phenotype through regulation of key oncogenic pathways of genes involved in cell adhesion, cancer stem cell maintenance and epithelial to mesenchyme transition.

Chromatin remodelling and transcriptional regulation are catalysed by energy-consuming enzymes that rely on a rapid ATP production.[213,221,223] Previous assumptions that energy demands are met by mitochondria, were challenged by Wright and co-workers[213] proposing a novel nuclear pathway for ATP synthesis during hormone-dependent transcription with NUDT5 playing a key role (Figure 1.16).

In response to the hormone progesterone, poly-ADP-ribose polymerase 1 (PARP1) is activated via a phosphorylation, generating poly-ADP-ribose (PAR) from NAD^+ by nicotinamide nucleotide adenylyltransferase 1 (NMNAT1).[213,221,226] Increased PARylation facilitates a more open chromatin structure, allows for recruitment of chromatin remodelling complexes and promotes histone displacement at the promoters of hormone-responsive genes.[213,226]

In the initial phase, ATP required for these processed is supplied through mitochondrial respiration.[213,221] The study by Wright and co-workers[213] suggests that the later phase is independent of mitochondrial ATP. However, a transient increase of PAR in the nucleus during the initial phase has been observed followed by a decrease in the later phase through activity of the poly-ADP-ribose-glucohydrolase (PARG).[221] PARG is the counteracting nuclear enzyme to PARP-1 and hydrolyses PAR into single ADP-ribose units.[213,223] The generated ADP-ribose is directly converted into ATP in the presence of pyrophosphate by NUDT5 (Figure 1.16).[223]

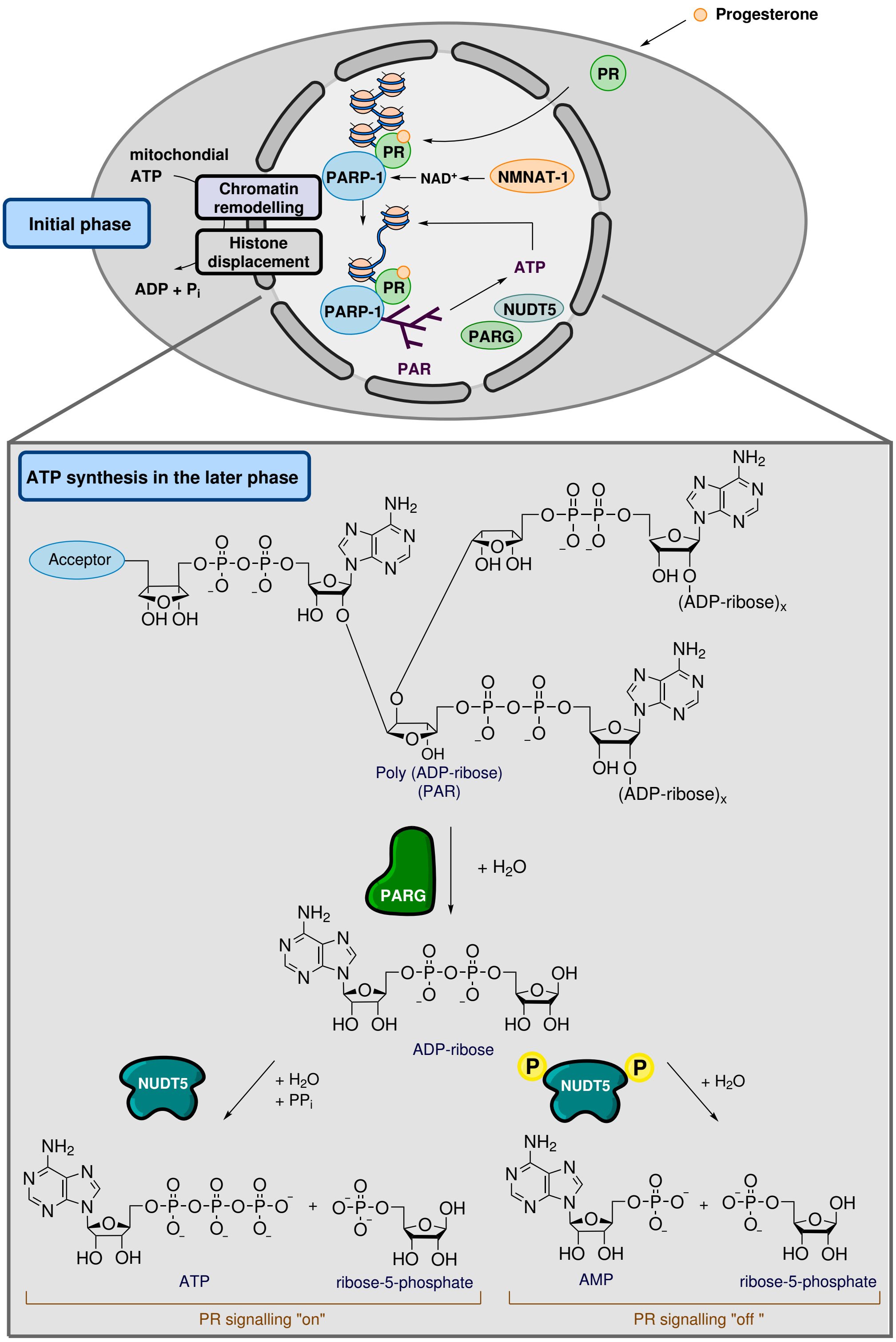


Figure 1.16. Rationale for targeting NUDT5. Upon progesterone exposure, poly-ADP-ribose polymerase 1 (PARP1) rapidly generates poly-ADP-ribose (PAR) inside the nucleus. With chromatin remodelling complexes, PARP-1 promotes displacement of histones. In the initial phase, ATP demands are covered by mitochondrial respiration. In the later phase, ATP is synthesised in the nucleus through hydrolysis of PAR to ADP-ribose, which is converted by NUDT5 into ATP in presence of PP_i. Dephosphorylation of NUDT5 during the hormone exposure favours the energetically more unfavourable ATP synthesis over AMP generation. Modified from[221].

As conversion of ADP-ribose into AMP is energetically more favourable reaction than synthesis of ATP, Wright and co-workers[213] demonstrate that enzymatic activity of NUDT5 is modulated through a posttranslational phosphorylation. Phosphoproteomic analysis revealed that NUDT5 is phosphorylated at Thr45 before hormone treatment and mostly dephosphorylated upon progesterone exposure.[213] Subsequent studies, including mutagenesis and model predictions provide evidence that phosphorylation stabilises the dimer conformation, shifting the equilibrium towards the AMP synthesis.[213] Upon hormone exposure, NUDT5 is dephosphorylated, leading to a destabilised dimer conformation allowing for PP_i to enter the substrate groove and facilitating ATP synthesis.[213,223]

Inhibition of PAR formation and depletion of PARG and NUDT5 abrogate the hormone-induced cell proliferation, confirming that the interplay between PAPR-1, PARG and NUDT5 is essential for nuclear ATP synthesis.[213]

Taken together, nuclear ATP synthesis catalysed by PARP-1, PARG and NUDT5 is required for gene regulation and chromatin remodelling and plays a role in breast cancer progression.[221] The novel pathway provides the rational for correlation of NUDT5 over-expression in cancer and poor disease outcome.[223] The enzymatic activity of NUDT5 is essential in driving the oncogenic pathways, suggesting that inhibition of nuclear ATP synthesis may result in a promising new therapeutic approach for breast cancer treatment.[200,213,227]

Inhibitor development First efforts in the development of a potent inhibitor for NUDT5 were undertaken by Page and co-workers.[200] In a screening campaign based on a enzyme-coupled malachite green assay (Figure 1.17.a), NUDT5 was tested against 72000 compounds resulting in identification of 528 hits. Based on the chemical structure of the hits, the theophylline core was identified as key for inhibition. The most potent compound TH1713 (IC_{50}=0.58 µM) was selected for further optimisation.

In a cascade approach of cellular thermal shift assay (CETSA) experiments, target engagement of multiple follow-up compounds could be confirmed including the most potent stabiliser TH5427. Its selectivity was subsequently confirmed in screening against a NUDIX and kinase panels.[200]

As shown in Figure 1.17, the co-crystal structures of the compounds TH1713 and TH5427 demonstrate a similar binding mode to the substrate ADP-ribose: The theophylline ring is anchored by Thr28 and Thr46 via $\pi-\pi$-stacking and is stabilised by hydrogen bonds

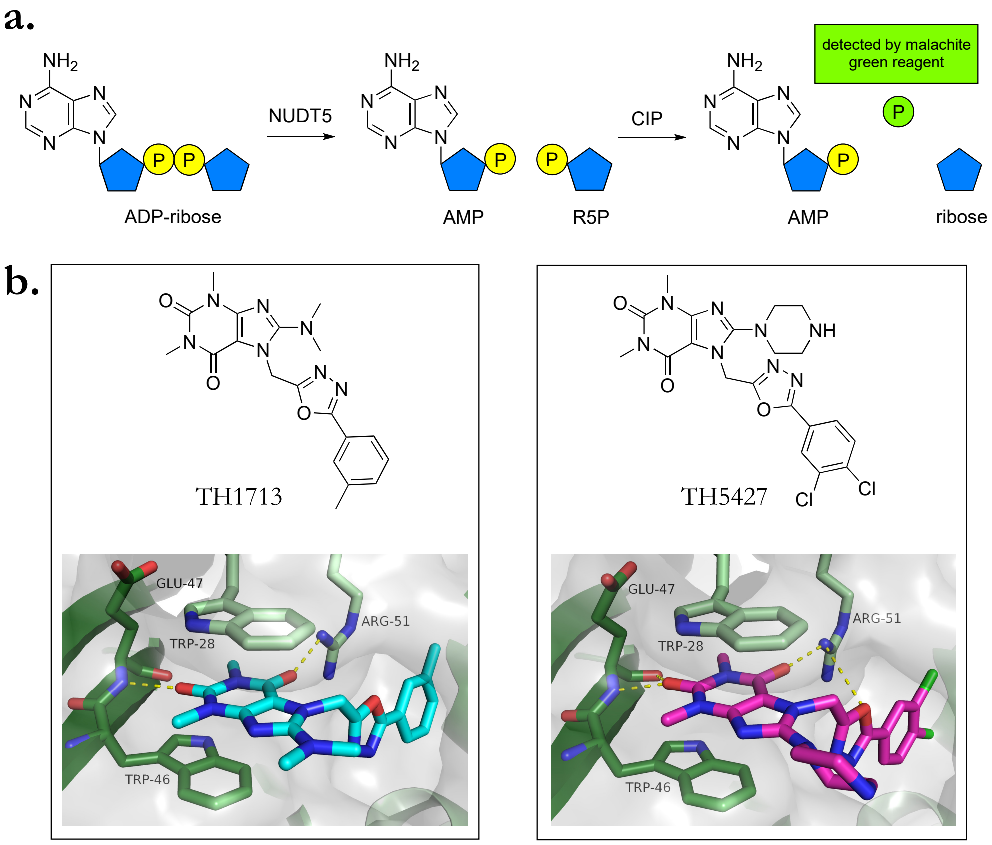


Figure 1.17. Development of NUDT5 inhibitor by Page and co-workers[200]. **a.** Representation of the enzyme-coupled malachite green assay used for the NUDT5 screening campaign. ADP-ribose is hydrolysed by NUDT5 into AMP and R5P. Subsequently, R5P is converted by calf intestinal alkaline phosphatase (CIP) to ribose and inorganic phosphate, which is then detected by malachite green reagent. Adapted from[200]. **b.** Chemical and crystal structures of NUDT5 in complex with inhibitors TH1713 (PDB: 5NQR; blue) and TH5427 (PDB: 5NWH; magenta). Both compounds are stabilised between Trp28 and 46 via $\pi-\pi$-stacking interactions and hydrogen bonds with Glu49 backbone and Arg51 side chain inside the active site of NUDT5.

to Glu49 backbone and Arg51 side chain. The 3-methylbenzene and dichlorobenzene moiety of TH1713 and Th5427, respectively, extend into the hydrophobic pocket.

After confirming the inhibitory effect of TH5427 *in vitro* first, further exploration of TH5427 in breast cancer cells demonstrated an attenuation of nuclear ATP synthesis upon progestin treatment.[200] The depletion of nuclear ATP blocked the activation of ATP-dependent enzymes required for chromatin remodelling and histone displacement and resulted in down-regulation of hormone-dependent genes and associated cell proliferation in the breast cancer cell line T47D.[200]

Taken together, Page and co-workers[200] identified the first potent and cell-active lead compound to inhibit NUDT5, demonstrating a potential therapeutic approach for breast cancer treatment.

1.4. Aims of the thesis

Crystallography is one of the most powerful fragment screening techniques: it provides precise structural information upon fragment binding and its high sensitivity allows to detect even very weak binders. This in turn presents unique challenges in selection and progression of identified hits.

This thesis addresses both method development and application of FBLD in the context of crystallographic fragment screening. The methodological work focusses on the development of tools for quantification of weak binders. At the same time, the SGC Oxford is interested in novel inhibition modes as well as chemical entities for NUDT5 modulation in breast cancer.

Chapter 3 presents the outcome of two fragment screening campaigns, namely crystallographic and 19F NMR screen. Both methods are amongst the most sensitive techniques to date, revealing even weakly binding fragment hits. Both campaigns are compared in regard to the number of identified binding sites, fragment hits and sampled interactions.

For targets bearing promiscuous sites such as NUDT5, fragment screening campaigns can yield a high number of hits in the same binding location. Their low affinity calls for novel tools for selection and prioritisation of fragment hits for further elaboration – a recurring challenge in FBLD.

This thesis aims to investigate the use of crystallographic and NMR-based methods for quantification of weakly binding fragment hits for rapid compound prioritisation and ranking. Chapter 4 introduces concentration-dependent crystallographic soaking which enables relative in-crystal ranking of fragment hits. Chapter 5 explores ligand-observed NMR methods for compound ranking in protein titration experiments.

The crystallographic screen revealed a previously unreported fragment binding site for NUDT5 bearing the promise of a novel inhibition mode. In chapter 6, efforts towards the initial site exploration and elaboration of a fragment hit are described. Strategies expanding the application of X-ray crystallography for streamlining elaboration, such as co-soaking of multiple compounds, are discussed.

2

Material and Methods

Contents

2.1. Chemicals and compounds

Chemicals for biological use were purchased from commercial sources (Acros Organics, Bayer, Biomol, Fluka, Formedium, GE Healthcare, Merck KGA, Sigma-Aldrich Co., Serva, Thermo Fischer Scientific, VWR) and were used without further purification.

All small molecule compounds for crystallographic and biophysical experiments were purchased from Enamine (Ukraine) and KeyOrganics (UK) and used without further purification.

2.2. Cell lines

For transformation, **Tuner™ (DE3) competent cells (Novagen)** were used.

Genotype: F^- *ompT* *hsdS*$_B$ (r_B^- m_B^-) *gal dcm lacY1*(DE3).

2.3. Plasmid map

The plasmid map of human His-tagged NUDT5 is shown in Figure 2.1.

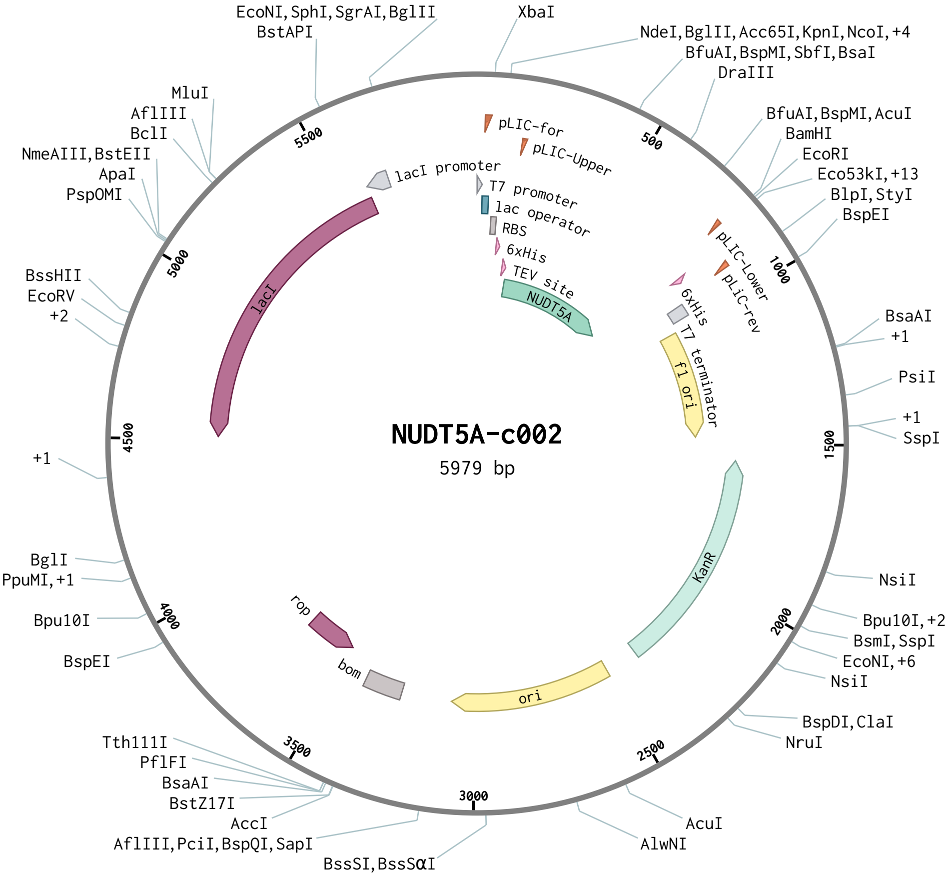


Figure 2.1. Plasmid map of human His-tagged NUDT5. Plasmid map was generated using Benchling.

2.4. Proteins

Proteins used in this work are listed in Table 2.1. All proteins were recombinantly expressed in *E. coli* cells.

Table 2.1. Proteins and their molecular weights (MW).

Proteins	MW [kDa]
His-tagged NUDT5	25.58
NUDT5 (cleaved)	23.11
NUDT7 (cleaved)	25.12

2.5. Media

Media and their composition used in this work are summarised in Table 2.2.

Table 2.2. Media and their composition.

medium	composition
SOC medium	20 g/L tryptone, 5 g/L yeast extract, 4.8 g/L $MgSO_4$, 3.603 g/L dextrose, 0.5 g/L NaCl, 0.186 g/L KCl.
LB medium	10 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract
TB medium	12 g/L tryptone, 24 g/L yeast extract, 9.4 g/L K_2HPO_4, 2.2 g/L KH_2PO_4 and 8 mL/L glycerol

2.6. Fragment libraries

2.6.1. DSiP fragment library for crystallographic screening

The DSiP fragment library for crystallographic screening[72] is the standard fragment library of the XChem fragment screening platform at DLS. Originally developed by Cox and co-workers, the library allows for rapid elaboration of fragments using standard parallel chemistry.[71] Since January 2018, an updated version of the DSiP library has been commercialised by Enamine.[72] The library contains 896 fragments at 500 mM in d_6-DMSO and is supplied in 1536-well (Echo compatible) plate and stored at − 80°C.

2.6.2. BIONET fluorine fragment library for ^{19}F NMR screening

The BIONET fluorine fragment library[228] is commercially available from KeyOrganics and is designed for ^{19}F NMR screening. The library is constructed according to the Rule of Three[51] and contains 461 fluorinated fragments which are combined into 24 compound mixtures containing 16-20 fragments each. The library is supplied at 200 μM in aqueous PBS buffer and stored at − 20°C.

2.7. Recombinant protein expression and purification

2.7.1. Expression and purification of NUDT5 for crystallisation

The construct encoding NUDT5 residues 1-208 was available in the laboratory and had previously been cloned into pNIC28-Bsa4 (see A.1). The expression plasmids were transformed into *E.coli* Tuner cells; the colonies were used to set up overnight starter cultures. Six litres of formulated TB media (24 g/L yeast extract, 12 g/L tryptone, 9.4 g/L K_2HPO_4 and 2.2 g/L KH_2PO_4) supplemented with 2% glycerol final concentration, 1 mL of 1000x*E.coli* metals (27 g/L $FeCl_2 \cdot 6\,H_2O$, 2 g/L $ZnCl_2 \cdot 4\,H_2O$, 2 g/L $CoCl_2 \cdot 6\,H_2O$, 2 g/L $Na_2MoO_4 \cdot 2\,H_2O$, 1 g/L $CaCl_2 \cdot 2\,H_2O$, 1 g/L $CuCl_2$, 0.5 g/L H_3BO_3 and 100 mL 37% HCl), 1 mL 10% antifoam 204, 1 mL 1 M magnesium sulphate, 10 mL 1 M ammonium sulphate and 20 mL 0.5% glucose per litre containing 50 μg/mL kanamycin were inoculated with 10 mL/L overnight culture and grown at 37 °C. Expression was induced with 0.2 mM IPTG at $A_{600} = 2.5$ and allowed to continue for 18 h at 18°C. The cells were collected by centrifugation (4700 rpm, 20 min, 4°C).

For lysis, cells were re-suspended in lysis buffer (50 mM HEPES, 500 mM NaCl, 20 mM imidazole, 5% glycerol, 0.5 mM TCEP, 0.5 mg/mL lysozyme and 1 μg/mL benzonase) supplemented with a cocktail of protease inhibitors. After 30 min incubation at room temperature, 1 % Triton X-100 final concentration were added before storing them at $-$ 80°C for further use.

Cell debris were removed by centrifugation (4700 rpm for 60 min at 4°C). The supernatant was loaded onto pre-equilibrated 1 mL His GraviTrap columns (GE Healthcare) and washed with 10 column volumes of wash buffer twice (50 mM HEPES, 500 mM NaCl, 20 mM imidazole, 5% glycerol and 0.5 mM TCEP). Protein was eluted with buffer containing 250 mM imidazole. A buffer exchange was performed subsequently using PD-10 columns (GE Healthcare). His-tag cleavage was performed with TEV protease (1 OD TEV for 10 OD units of the target) overnight at 4°C.

To remove His-tagged protein, a reverse IMAC step was carried out. In the final step, protein solutions were concentrated and loaded onto a size exclusion column (Yarra SEC-2000, Phenomenex) at 4 mL/min in SEC buffer (10 mM HEPES, 500 mM NaCl, 5% glycerol and 0.5 mM TCEP) on an NGC chromatography system (BioRad). The protein solution was concentrated to 25 mg/mL, and stored at $-$ 80°C.

2.7.2. Purification of NUDT5 for NMR experiments

Protein expression and purification were carried out as described in 2.7.1.

Following the purification by size exclusion chromatography, an additional buffer exchange step was performed to remove all protonated solvent. PD-10 columns were thoroughly pre-equilibrated using 30 mL of NMR buffer (pre-formulated PBS buffer tablets in D_2O). 2.5 mL of concentrated protein solution were applied to each PD-10 columns and eluted using 3.5 mL of the NMR buffer. The obtained protein solutions were concentrated to 28 mg/mL and stored at − 80°C for further use.

2.7.3. Purification of NUDT5 for SPR experiments

Protein expression and subsequent lysis were carried out as described in 2.7.1.

Supernatant was cleared by centrifugation (4700 rpm for 60 min at 4°C) and was loaded onto a 1 mL His GraviTrap columns (GE Healthcare). Following two washing steps using 10 column volumes of wash buffer (50 mM HEPES, 500 mM NaCl, 20 mM imidazole, 5% glycerol and 0.5 mM TCEP), protein was eluted from the column using wash buffer supplemented with 250 mM imidazole. Subsequently, a buffer exchange into SEC buffer (10 mM HEPES, 500 mM NaCl, 5% glycerol and 0.5 mM TCEP) was carried out using a PD-10 column (GE Healthcare). The obtained solution was concentrated and loaded onto a size exclusion column (Yarra SEC-2000, Phenomenex) at 4 mL/min in SEC buffer on an NGC chromatography system (BioRad). The protein solution was concentrated to 11 mg/mL, and stored at − 80°C.

2.7.4. Expression and purification of NUDT7 for crystallisation

The expression and purification of the NUDT7 (residues 14-235) according to the procedure described in[229] was carried out by Dr. Michael Fairhead, SGC Oxford.

2.8. X-ray crystallography

2.8.1. Crystallisation screens and solutions

Crystallisation screens used in this work are listed in Table 2.3. All solutions required for custom-designed crystallisation screens were obtained from Molecular Dimensions (UK).

Table 2.3. Commercial crystallisation screens and their suppliers.

Crystallisation screen	Supplier
Basic ChemSpace (BCS)	Molecular Dimensions
Crystal Screen HT (HCS)	Hampton Research
JCSG+	Molecular Dimensions
Index (HIN)	Hampton Research
Ligand Friend Screen (LFS)	Molecular Dimensions
SaltRx	Hampton Research

2.8.2. Crystallisation and data collection

NUDT5. For initial structure determination, NUDT5 crystals were obtained at 20 °C in sitting drops of 75 nL protein (25 mg/mL) and 75 nL of a precipitant consisting of 30% PEG 4K, 0.2 M magnesium chloride and 0.1 M Tris pH 8.5. NUDT5 crystals were cryo-protected using ethylene glycol and flash-cooled in liquid nitrogen. Diffraction data was collected at the beamlines I03, I04 and I04-1 at Diamond Light Source.

Diffraction data was processed automatically using software pipelines available at Diamond Light Source.[125] Molecular replacement was carried out using PHASER MR[230] from the CCP4 suite[231]. The structure PDB:2DSB was used as a search model. Model building was performed using COOT[232] and refinement using REFMAC[233].

NUDT7. Following the procedure described in[229], NUDT7 crystals were obtained at 20 °C in sitting drops by mixing 100 nL of 20 mg/mL protein in 10 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol and 50 nL of 0.1 M BisTris pH 5.5, 0.1 M ammonium acetate and 6-8% (w/v) PEG10000.

2.8.3. XChem fragment screening and data collection

Apo-crystals of NUDT5 were obtained at 20 °C in sitting drops of 100 nL protein (25 mg/mL) and 100 nL of a precipitant consisting of 33% PEG 4K, 0.2 M magnesium chloride and 0.1 M Tris pH 8.0.

For fragment screening at the XChem Fragment Screening Facility at Diamond Light Source, concentrated solutions (0.5 M) of fragments from the DSiP library (Enamine) were added to the crystallisation drops using an acoustic liquid dispenser[109] to 10% of the drop volume. All crystals were cryo-protected using 5% ethylene glycol after an incubation time of 30 min for NUDT5 and flash-cooled in liquid nitrogen. Diffraction data was collected at the Diamond Light Source beamline I04-1.

For validation experiments of fragment hits from the ^{19}F fragment screen, fragments were obtained from KeyOrganics and dissolved in DMSO. A library plate containing concentrated solutions of all fragment hits at 500 mM, 200 mM and 100 mM was prepared. Soaking experiments were carried out as described above.

Diffraction data were automatically processed by software pipelines available at Diamond Light Source.[125] Initial map calculation was carried out using DIMPLE[234] using PDB:6GRU for NUDT5 as search model. Hit identification was performed using PanDDA[113]. For further refinement and model building, REFMAC[233] and COOT[232] were used via the XChemExplorer data management and workflow tool.[126]

Coordinates and structure factor for all protein-fragment datasets were deposited in the Protein Data Bank (PDB) in a group deposition G_1002057. Validated fragment hits from the ^{19}F screen are deposited in a group deposition G_1002105.

2.8.4. Concentration-dependent fragment soaking

The concentration-dependent fragment soaking was carried out at the XChem Fragment Screening Facility at Diamond Light Source using two targets, NUDT5 and NUDT7.

NUDT5. 43 fragment hits originally identified in the XChem campaign were re-soaked into NUDT5 crystals (obtained as described in 2.8.3) at 10% of the drop volume from library plates containing DMSO-solution of fragments at different concentrations (500 mM, 250 mM, 100 mM, 50 mM and 25 mM), obtained from Enamine. After an incubation time of 28 min, crystals were harvested, cryo-protected using 5% ethylene glycol and flash-cooled in liquid nitrogen. Diffraction data was collected in automated mode at the Diamond Light Source beamline I04-1. For each fragment and each concentration point, duplicate datasets were collected. Experiments, which did not result in a datasets due to experimental issues (such as damaged or non-diffracting crystals) or had a diffraction limit above 2.2 Å, were repeated until at least duplicates for each fragment and concentration were obtained.

Diffraction data was processed as described in 2.8.3. The hit identification was carried out using PanDDA[113] and the NUDT5 apo-structure (PDB:6GRU) was used as reference model. In datasets containing multiple events due the presence of four protein chain copies, the ligands were only modelled in one site. The presence of the fragments were confirmed by two-cycles of refinement within the XChem Explorer[126]. Subsequently, the binding mode of the fragments across soaking concentrations were validated by structural alignment in PyMOL (version 1.8.2.0, Schroedinger, US) to the deposited protein-fragment structures (PDB group deposition G_1002057). For the downstream analysis, fragment hits that were detected in at least one of the duplicates were defined as present at a given concentration.

NUDT7. For the concentration-dependent soaking experiments, the non-covalently binding XChem hit **NUDT7-REV-XChem** and the elaborated compound **NUDT7-REV-1** were used. The NUDT7 crystals (obtained as described in 2.8.2) were soaked at 30% of the drop volume from a library plate containing DMSO-solutions of both compounds at the concentrations of 500 mM, 250 mM, 100 mM, 50 mM and 25 mM (obtained from Enamine) in duplicates and incubated overnight at room temperature. Upon incubation, crystals were harvested without additional cryoprotection and flash-cooled in liquid nitrogen. Diffraction data was collected in automated mode at the Diamond Light Source beamline I04-1. As for NUDT5, all experiments which did not result in a dataset were repeated until duplicated were obtained. The hit identification and subsequent analysis were carried out as described for NUDT5.

2.8.5. Co-soaking of multiple fragments

The combinatorial soaking experiments were performed at the XChem Fragment Screening Facility at Diamond Light Source. Concentrated solution of fragment **41** at 500 mM was added to the pre-formed NUDT5 crystals using the acoustic liquid dispenser at 5% DMSO. Following an incubation of 5 min, a second XChem fragment hit previously identified in the binding site A was added at the same concentration. In total, 40 fragment hits were tested in combination with fragment **41** in duplicates. After an incubation of 25 min, crystals were harvested, cryo-protected using 5% ethylene glycol and flash-cooled in liquid nitrogen. Diffraction data was collected at the Diamond Light Source beamline I04-1.

The automatically processed diffraction data was analysed using the PanDDA algorithm[113]. The NUDT5 structure containing fragment **41** (PDB: 5QJ5) was used as reference. The down-stream analysis was carried as described in 2.8.3.

2.9. Nuclear magnetic resonance

2.9.1. Fluorine NMR-based fragment screening

Fragment screening was performed on Avance Neo 700 MHz instrument (Bruker, USA) with a 5 mm QCI-F cryoprobe. The BIONET fluorine fragment library (Keyorganics, UK) consisting of 461 compounds was screened in 24 mixtures of 16-20 fragments at 50 μM against 1.5 μM NUDT5. The NMR samples with a total volume of 170 μL in a 3 mm NMR tube were prepared in 10 mM phosphate, 137 mM NaCl, 2.7 mM KCl, 0.001 mM sodium azide, pH 7.4, supplemented with 10% D_2O for field frequency lock.

^{19}F CPMG experiments were acquired for each mixture in the presence and absence of protein at 298 K using ^{1}H decoupling with a spectra width of 81967.211 Hz, an acquisition time of 999 ms and 128 scans with a relaxation delay of 1 s. All spectra were recorded with 20 ms and 200 ms relaxation delay.

Data were processed using Topspin 4.0.6 and the FBS tool (Bruker, USA). For hit identification, the integrated peak ratios of 200 ms to 20 ms spectra in the presence of protein were compared to the peak ratios of the 200 ms to 20 ms spectra of the reference. Fragments exhibiting at least 40% ^{19}F T_2 effect were defined as hit.

2.9.2. Determination of concentration in solution

For reference measurements, ^{1}H spectra of individual fragments in absence of protein were acquired. For each fragment hit, a sample at 2 mM in D_2O-based PBS buffer was prepared. Briefly, fragment stock solutions of 500 mM in d_6-DMSO were diluted in

PBS buffer to 10 mM, followed by a second 1:5 dilution to obtain fragment solutions of 2 mM containing 0.4% DMSO. For fragments with limited solubility (fragment **1**, **39** and **41**), an additional dilution to 250 mM in d_6-DMSO was included, resulting in the final DMSO concentration of 0.8%. To ensure accurate quantification, concentration of each fragment in solutions was verified using 2 mM solution of sucrose in 90:10% H_2O:D_2O as an external standard. For that, ^{1}H spectra for each fragment were referenced to the ^{1}H signals of the external standard. Concentration in solution for all ligands were calculated using ERETIC2 tool in TopSpin 4.0.6.[235] Following the obtained results, fresh samples were prepared with necessary adjustments to obtain concentrations of 2 mM.

2.9.3. NMR-based ranking of fragment hits

Ranking experiments were carried out on a Ascend 600 MHz instrument (Bruker, USA), equipped with a helium cryoprobe at Bruker site in Fällanden, Switzerland. Protein titrations were performed at a constant ligand concentration at 2 mM. For each fragment hit, three molar ligand-to-protein ratios of 0.5, 1.5 and 3% (0.02 mM, 0.04 mM and 0.1 mM) were tested. Each sample with a total volume of 200 μL was prepared in a 3 mm NMR tube using D_2O-PBS buffer, pH 7.5.

STD experiments. STD experiments were recorded using Bruker pulse sequence stddiffesgp.3 at 283.15 K with 64 scans, an acquisition time of 1.7 s and a relaxation delay of 2 s. Protein saturation was carried out by a train of 50 ms Gaussian pulses at 0 ppm separated by 2 ms delays. For the off-resonance spectra, a similar saturation pulse was applied at 30 ppm. ^{1}H signal assignment for each fragment was assisted through IconNMR tool (Bruker). For each protein concentration, STD effect was calculated from the integrals of each ^{1}H signal as a ratio of signals in the saturation difference and the reference spectra.

$T_{1\rho}$ experiments. $T_{1\rho}$ experiments were acquired for each fragment and concentration at 283.15 K with 64 scans, an acquisition time of 1.7 s. All spectra were recorded with a short (20 ms) and long (200 ms) relaxation delay. ^{1}H signal assignment for each fragment was assisted through IconNMR tool (Bruker). Spectra were processed using Topspin 4.0.6 (Bruker, USA). For each fragment and concentration, assigned 1H signals were integrated. Signal attenuation was calculated from integrals corresponding to short (20 ms) and long (200 ms) delay. The $T_{1\rho}$ for each ^{1}H signal was calculated as ratio of integral ratios in presence and absence of NUDT5. The $T_{1\rho}$ for the entire molecule was calculated as a sum of the effects on each ^{1}H signal weighted by the number of protons and normalised by the total number of protons (equation 5.3). The ranking was obtained using the $T_{1\rho}$ effect observed with highest protein concentration (ligand to protein ratio of 3%).

2.10. Surface plasmon resonance

2.10.1. Binding affinity quantification of NUDT5 fragment hits

SPR experiments were performed on a Biacore T200 (GE Healthcare). Experiments were conducted with immobilised His-tagged proteins on a Series S NTA sensor chip (GE Healthcare). The purification procedure of His-tagged NUDT5 is described in 2.7.3. GFP used as reference protein was kindly provided by Dr. Michael Fairhead (SGC Oxford, UK).

NTA sensor chip surface was prepared for protein immobilisation through injections of regeneration buffer (10 mM HEPES, 150 mM NaCl, 350 mM EDTA, pH 8.3), capture buffer (50 mM HEPES, 150 mM NaCl, 50 μM EDTA, 0.005% Tween 20 pH 7.4) and $NiCl_2$ solution (3.75 μM $NiCl_2$ in capture buffer) over all flow cells. The first flow cell was used as reference. Both NUDT5 and GFP were diluted in capture buffer to 3.9 μM and 4.6 μM, respectively, and captured with a 20-min injection at a flow rate of 5 µL/min each. NUDT5 and GFP were immobilised at densities 8554 RU and 9081 RU, respectively. Lower immobilisation densities of NUDT5 (791 RU, 1597 RU, 2673 RU and 4615 RU) were achieved by decreasing the protein concentration to 1.3 μM and the duration of the injection.

Compounds provided as 500 mM d_6-DMSO stocks were diluted to 20 mM in DMSO and then in the running buffer (20 mM HEPES, 150 mM NaCl, 5 mM $MgCl_2$ and 0.005% Tween 20 pH 7.5) with 3% DMSO at 2-fold concentration series from 18.75 μM to 600 μM. All experiments were run at 10 °C. Each fragment was injected for 30 s contact and 60 s dissociation time at a flowrate of 30 μL/min. Positive controls were injected at 3-fold concentration series as follows: ATP from 411 nM to 100 μM, ADP-ribose and AMP-PNP from 206 nM to 50 μM and ADP from 20.6 nM to 5 μM. In addition, AMP-PNP was injected throughout the experiment at 25 μM.

Validation experiments were repeated for selected subset of compounds at 2-fold concentration series from 3.9 μM to 500 μM. For ATP, a 3-fold concentration series from

45.7 nM to 100 μM was used and injections were carried out with 60 s dissociation time. For ADP-ribose, a 3-fold concentration series from 22.9 nM to 50 μM was prepared and injected with 60 s dissociation time. ADP was injected in a 3-fold concentration series from 2.3 nM to 5 μM with 120 s dissociation time. To monitor protein stability on the biosensor surface, AMP-PNP was injected throughout the experiment at 25 μM. In addition, the flow rate was increased from 30 to 60 μL/min.

All data was double-referenced for the blank injections and reference surface standard processing procedure. A solvent correction was performed to correct potential deviations in DMSO concentrations between samples and running buffer. The data processing, equilibrium and kinetic fitting were performed using Scrubber software version 2 (BioLogic Software, Australia). The affinity (K_D) and kinetic binding parameters for the compounds were determined by global fitting the sensorgrams using a one-site kinetic model.

2.11. Computational Chemistry

2.11.1. PLIP analysis

The enumeration of protein-fragment interactions was carried out using Protein-Ligand Interaction Profiler (PLIP).[236] For each protein-ligand structure, the freely available tool generates a list of detected non-covalent interactions as well as a corresponding PyMOL session for visualisation of the interactions. Parallel processing of both the XChem (PDB: G_1002057) and ^{19}F fragment hits that we confirmed crystallographically (PDB: G_1002105) was performed in batch mode using python source code of PLIP (version 1.4.5) from GitHub (https://github.com/pharmai/plip/releases).[236] The crystal structures of all protein-ligand complexes were provided by supplying their four-letter PDB ID. For comparison between the XChem and the ^{19}F NMR campaigns, the identified interactions across all protein-fragment structures were collated separately. The list of unique interactions for each fragment set was obtained by removing the redundancies. For structures containing multiple copies of the fragment with slight differences in their binding modes, all unique interactions were listed.

For the analysis of the concentration-dependent study, fully refined PDB-deposited structural models obtained in the XChem fragment screen (PDB: G_1002057) were used as binding modes of the fragments remain conserved across the concentration series.

2.11.2. Molecular Docking

In silico docking was performed using the Molsoft ICM software package.[237] The receptor was prepared from the structure of NUDT5 in complex with fragment **41** (PDB: 5QJ5). The protein was converted to an ICM object using the standard ICM procedure[238,239], deleting waters and optimising hydrogen atoms including the optimised orientation of His, Pro, Asn, Gln, Cys. The ligand atoms were assigned based on the MMFF94 forcefield[240], the ligand was created as a separate object. The binding site was defined from fragment **41** using the "Define Site Around Selected Ligand" function. The ligand database was prepared using ChemDraw and exporting these in SDF format to a ChemOffice spreadsheet. 2D depictions were converted to SMILES ensuring correct absolute stereochemistry was retained. The SDF file was loaded into ICM as a chemical table. A rigid docking protocol was performed using the standard parameters in ICM Dock Chemical Table. The best-scoring results for each ligand were manually inspected.

2.11.3. Molecular Dynamics

All simulations were carried out with the GROMACS 4.6 package[241] and the AMBER99-SB*-ILDN force field[242–244]. The structure of NUDT5 (PDB:6GRU) was used as initial conformation. Crystallographic ligands were deleted and pdb2gmx used for structure preparation. The structure was added to a triclinic box with periodic boundary conditions with the edge of the box being 12 Å from the nearest protein atom. The box was populated with SPC/E waters.[245] The system was neutralized by adding Cl^- and Na^+ ions to a final concentration of 150 mM. Energy minimisation using the steepest descent method was followed by equilibration with positional restraints on heavy atoms for 100 ps. The 100 ns trajectory was collected at 300 K and 1 bar with standard NPT ensemble settings. The temperature was kept at 300 K with velocity rescaling (interval 0.1 ps)[246] and the pressure was kept at 1 bar using the Parrinello-Rahman coupling scheme (interval 2 ps)[247]. Electrostatic interactions were calculated at every step with the particle-mesh Ewald method[248] and short-range repulsive and attractive dispersion interactions were simultaneously described by a Lennard-Jones potential, which was cut off at 1.0 nm. The SETTLE algorithm was used to constrain bonds and angles of water molecules[249]. LINCS was used for all other bonds[250,251]. Additionally, virtual interaction sites (VSITE) replaced the fastest angular degrees of freedom involving hydrogen atoms, allowing a time step of 4 fs. Visualisation of results was performed with

PyMOL[252] and analysis tools of GROMACS 4.6[241]. Convergence of energy terms, ensemble values and minimal distance to periodic images were monitored to ensure quality of the simulation.

3

Fragment hit identification: Crystallographic screening reveals more binding sites and interactions than ^{19}F NMR

Contents

3.1. Introduction

Fragment screening allows to identify starting points for development of novel chemical entities for target modulation. This chapter presents and critically evaluates fragment screening methods on the NUDIX hydrolase NUDT5. The catalytic activity of human NUDT5 was recently linked to a newly proposed non-canonical pathway for nuclear ATP synthesis required for chromatin remodelling and gene regulation in breast cancer.[213,221] This pathway provides the rationale for correlation between NUDT5 overexpression and development of an aggressive disease phenotype.[223] As enzymatic activity of NUDT5 plays a key role in driving this oncogenic pathway, inhibition of NUDT5 may present a promising therapeutic approach for breast cancer treatment (see 1.3.2).

Negative modulation of biological activity of NUDT5 might be achieved either through competitive or allosteric inhibition. Potent inhibitors targeting the substrate binding site have been previously described[200], however the promiscuous nature of ADP-ribose, the endogenous substrate of NUDT5, raises concerns about possible off-targets their limiting clinical applicability. As such, the identification of biologically relevant secondary sites, including allosteric sites, is of particular interest to modulate cellular activity of NUDT5 more selectively. The work in this chapter reports the identification of starting points for development of novel NUDT5 antagonists.

Two fragment-based screening techniques were utilised: a XChem crystallographic fragment screen at Diamond Light Source and a ^{19}F NMR-based screen using the BIONET fluorinated fragment library. The majority of identified hits bind in the substrate binding site, and one in a novel, previously unreported deep cavity in close proximity to the active site. Both campaigns are compared in regard to the number of binding sites, fragment hits and protein-fragment interactions, highlighting the advantages of XChem by directly providing structural information upon ligand binding and enabling a broader sampling of the chemical space on the target surface.

3.2. Results

3.2.1. Development of a crystallisation system for NUDT5

Crystallographic fragment screening requires a crystallisation system that delivers highly reproducible crystals diffracting to high resolution (below 2.5 Å) and providing accessible sites of interest.[102,114] In addition, protein crystals need to tolerate presence of organic solvent, as fragment libraries are often solubilised in DMSO.[102]

At the start of the project, five constructs (NUDT5A-c001 to NUDT5A-c005, appendix A.1) of NUDT5, with varying N- and C-terminal truncations, were already available at the laboratory. All five constructs were expressed in *E.coli* Tuner strain with N-terminal hexahistidine tag. The purification strategy for all constructs included nickel-affinity chromatography and a His-tag cleavage step by Tobacco Etch Virus (TEV) protease, followed by reverse nickel-affinity and size-exclusion chromatography. The mass and purity were confirmed by mass spectrometry. The detailed procedure is described in 2.7.1. As the protein encoded by the construct NUDT5A-c004 displayed significant degradation during the purification process, a cation-exchange chromatography step was required to remove the degradation products. Except for the construct NUDT5A-c003, all constructs could be purified successfully, yielding 3.6 mg (NUDT5A-c004) to 49 mg (NUDT5A-c001) protein per liter.

Freshly purified proteins were taken into crystallisation trials set up using sparse matrix screens at 20 °C and 4 °C (see 2.8.1).[253] NUDT5 proteins encoded by the constructs NUDT5A-c002 and c004 yielded crystals in multiple conditions diffracting up to 2.0 Å. Subsequent optimisation of the initial crystallisation conditions to improve crystal size and quality was carried out with grid screens for a systematic exploration of the initial condition, prepared with stepwise gradient of pH, PEG and precipitating agents. Further optimisation improved consistency of the crystals across the plate, achieved by varying protein concentration, drop size and protein to precipitant ratio.

The optimal conditions for the construct NUDT5A-c002 encoding residues 1-208 yielded reproducible crystals diffracting up to 1.9 Å in 33% PEG 4000, 0.2 M $MgCl_2$ and 0.1 M Tris pH 8 at 20°C using 1:1 protein to precipitant ratio in sitting drops (Figure 3.1).

The initial structure of NUDT5 was solved from a dataset diffracting up to 1.93 Å in space group P1 using molecular replacement and deposited to the PDB (PDB: 6GRU).

The residues 1-12 are disordered in the structure and were omitted from the final model. The statistics of the diffraction data and the structure refinement are given in Table 3.1.

Table 3.1. Crystallographic key parameters for refined structure of NUDT5 (PDB: 6GRU).

Crystallographic parameters for protein-ligand structures	
Space group	P1
Cell dimensions:	
a, b, c (Å)	49.6, 60.14, 80.01
α, β, γ	79.92°, 81.52°, 75.49°
Diffraction range (Å)	$32.03 - 1.93$
Mean R-free	0.26
Mean R-work	0.23

The asymmetric unit contains four copies of the chain, assembled as two homodimers. Each homodimer displays a substantial domain swapping as the N-terminal domain of one monomer crosses over to interact with the C-terminal domain of the other monomer, as previously described by Zha and co-workers[220] (Figure 3.1.b). The dimer consisting of the chain C and D is disordered in the outer regions. All chains contain two octahedrally coordinated magnesium ions bound in the substrate binding site (Figure 3.1.c-d).

Prior to the fragment screening, tolerance of NUDT5 crystals towards organic solvent was assessed. Initial characterisation was performed using varying concentrations of DMSO (5-30%) and incubation time (1-4 h). The characterisation indicated 10% (v/v) as maximal concentration of DMSO over an incubation period of approx. 30 minutes.

With this highly reproducible crystallisation system in hand, yielding up to 80 crystals per plate, all pre-requisites for crystallographic fragment screening were fulfilled.

3.2.2. Results of the crystallographic fragment screening

NUDT5 crystals were screened against a subset of 650 randomly selected compounds from the DSiP library[71] individually. All crystals were soaked at 500 mM (10% DMSO in crystal drops) for an average soaking time of 32 min. In total, 646 datasets were collected at beamline I04-1 at DLS, with 602 crystals yielding diffraction data that

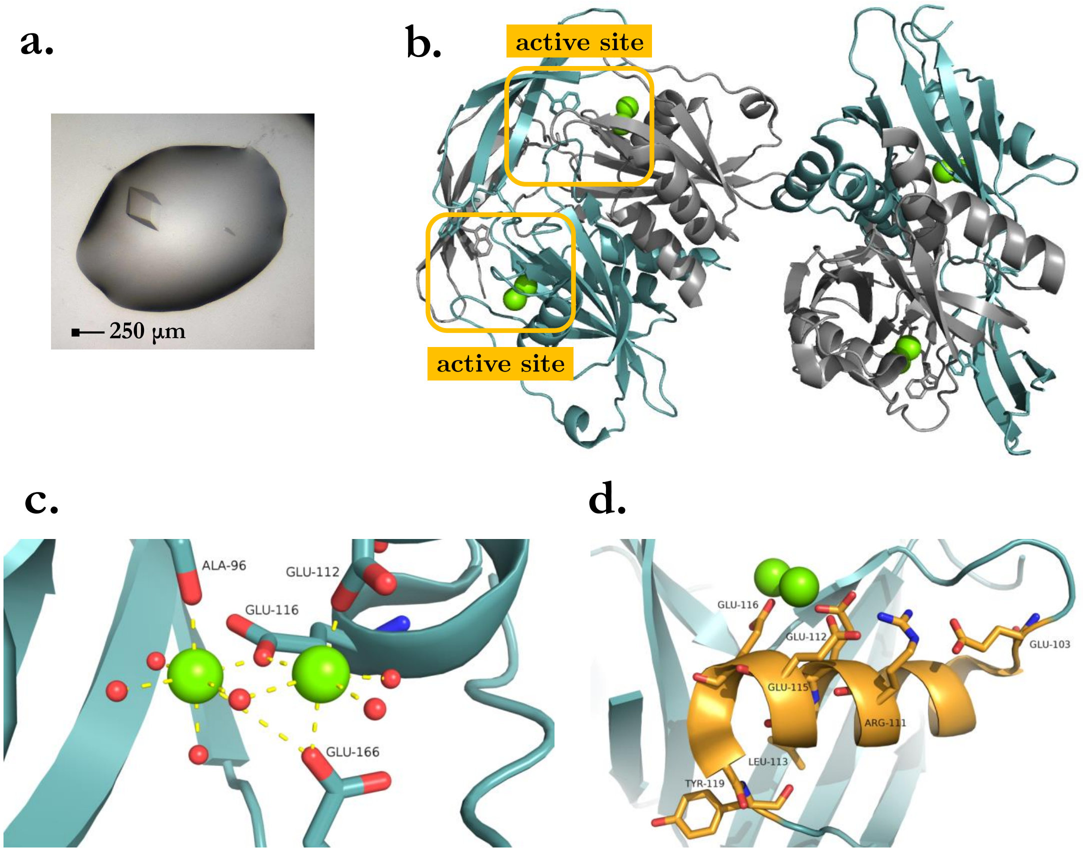


Figure 3.1. Crystallisation system for the NUDIX hydrolase NUDT5. **a.** NUDT5 crystals in 33% PEG 4000, 0.2 M $MgCl_2$ and 0.1 M Tris pH 8 at 20°C after four days. **b.** Crystal structure of NUDIX hydrolase NUDT5 (PDB: 6GRU). The asymmetric unit contains four copies of the protein chain, which form two homodimers (with chains shown in grey and teal). The active site is located at the interface of both chains. Each chain contains two octahedrally coordinated magnesium ions in the active site (shown in green). **c.** Coordination of magnesium ions in the structure: two magnesium ions are present in each monomer and are coordinated octahedrally by the residues Ala96, Glu112, Glu116 and Glu166 and water molecules. **d.** The NUDIX box with the sequence $GX_5EX_7REUXEEXGU$ (with U being a bulky hydrophobic amino acid I, L or V) is a conserved motif among the NUDIX superfamily which contains catalytic residues (shown as sticks in orange).

could be processed successfully using auto-processing pipelines at DLS. The diffraction limit ranged from 1.42 Å to 2.81 Å, with an average of 1.73 Å. All crystals belonged to the triclinic space group P1 containing two homodimers (four molecules) per asymmetric unit.

Hit identification was carried out using Pan-Dataset Density Analysis (PanDDA).[113] The previously solved apo-structure of NUDT5 in space group P1 (PDB: 6GRU) was used as reference model. In total, 580 events in 600 datasets were identified by PanDDA, which were spread across the protein structure. Many datasets contain multiple events due the presence of four protein chain copies. All identified events were clustered into 32 non-overlapping sites (S1-S32) as shown in Figure 3.2.

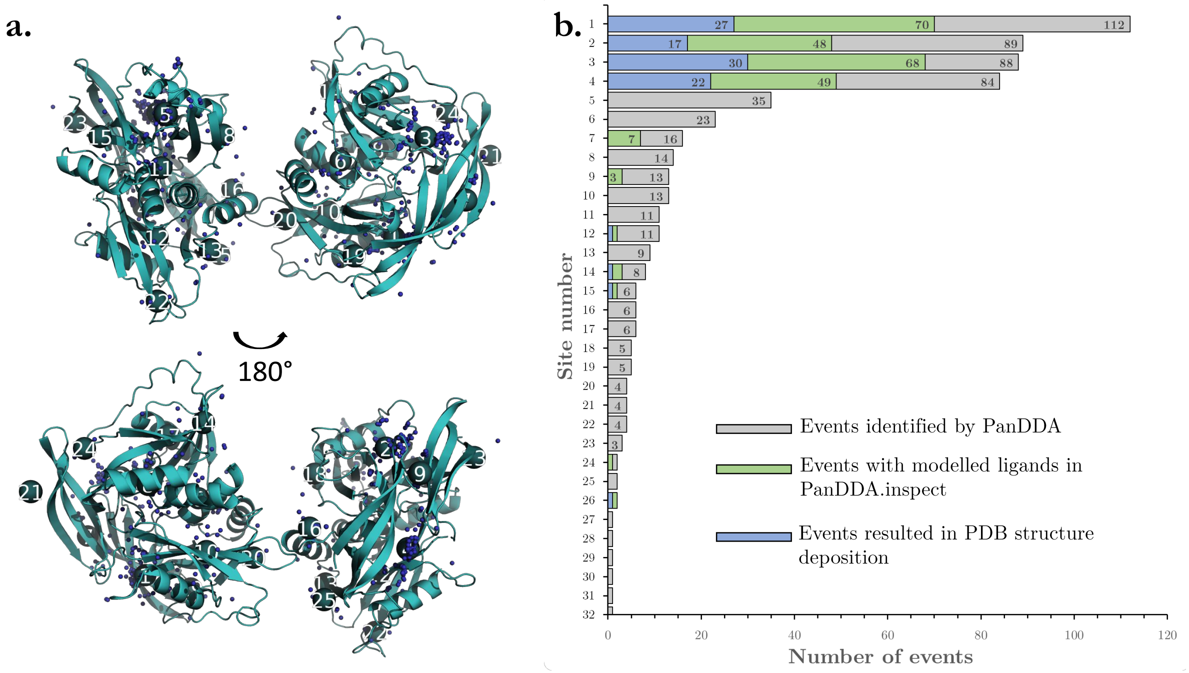


Figure 3.2. PanDDA analysis of NUDT5 fragment screen. **a.** PanDDA algorithm clustered all identified events into 32 non-overlapping sites across the protein structure. The NUDT5 structure is shown in cyan (front view (top), back view (bottom), the events are represented as dark blue dots, the event sites are numbered from 1-32. **b.** Histogram of non-overlapping sites (1-32) and corresponding number of events in each site. Total number of events identified by PanDDA algorithm is shown in grey, events with modelled ligands during PanDDA.inspect step in green, events resulted in PDB deposition of protein-ligand complex in blue.

155 ligands were modelled manually and refined within the XChem Explorer.[126] Multiple refinement cycles confirmed presence of 43 unique fragment hits (Figure 3.6). Due to the presence of four protein chain copies in the asymmetric unit, 77% of the datasets contain more than one copy of the identified fragment. Ligand occupancy ranged from 0.34 to 1.0 with an average occupancy of 0.81. The average diffraction limit of the refined protein-ligand complexes is 1.66 Å, which is slightly higher than average diffraction limit of the processed datasets of 1.73 Å. The detailed statistics of the diffraction data and structure refinement as given in Table 3.2.

In the PanDDA analysis, 43 fragments were identified, giving a total hit rate of 7.2%. Fragment hits were found across five different sites: substrate binding site A, site B – a deep pocket in close proximity to the binding site A, and surface-exposed sites C, D and E (Figure 3.3). Of all identified fragment hits, 40 fragments bind in the active site,

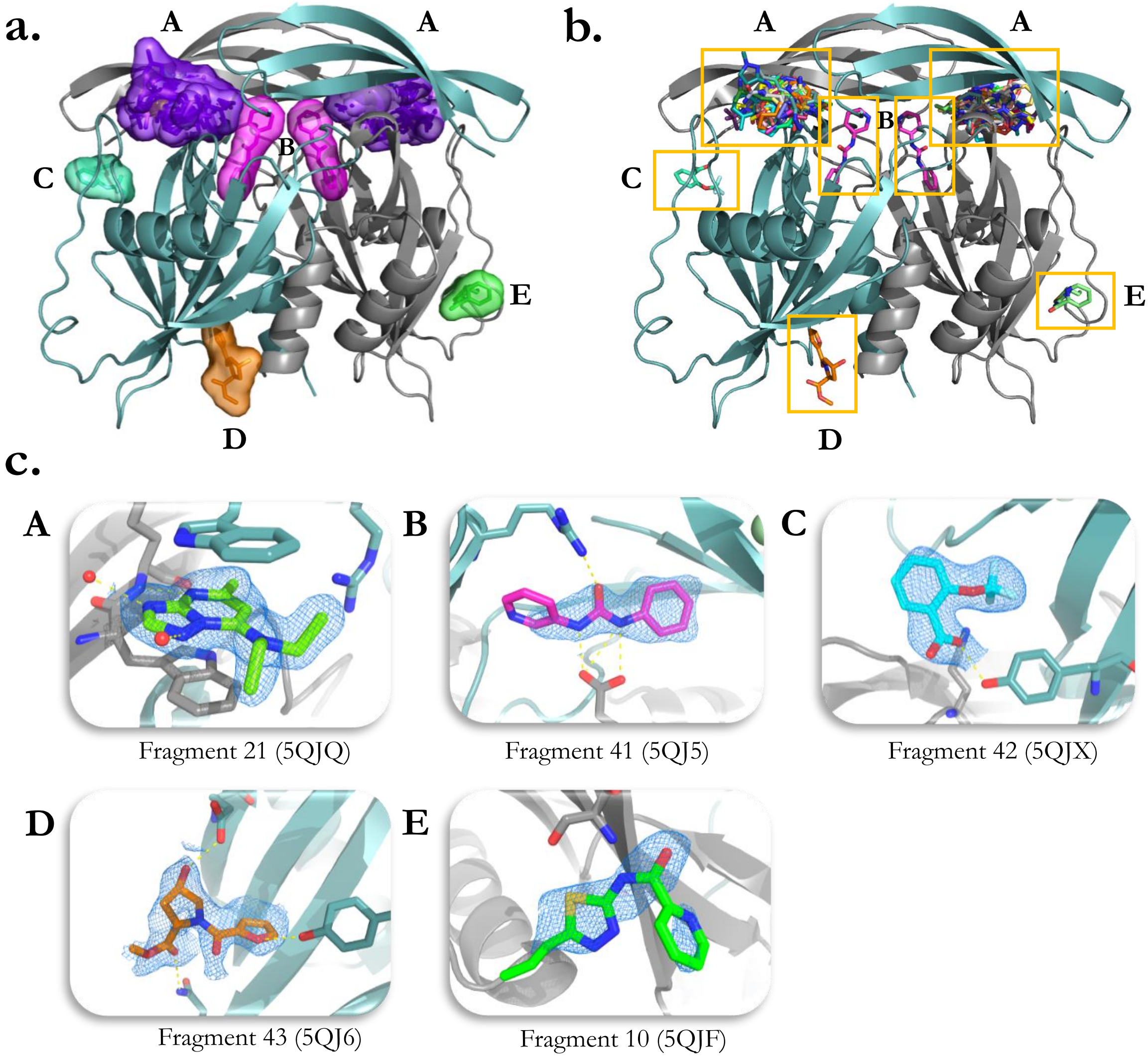
Table 3.2. Crystallographic key parameters overview for refined protein-ligand complexes for NUDT5 and fragment hits from DSiP fragment library (Enamine).

Crystallographic parameters for protein-ligand structures	
Space group	P1
Cell dimensions:	
a, b, c (Å)	49.19 (0.34), 59.79 (0.19), 80.01 (0.25)
α, β, γ	79.44° (0.21), 81.6° (0.40), 75.78° (0.35)
Diffraction range (Å)	1.44−2.03
Mean diffraction (Å)	1.66 (0.12)
Occupancy range	0.34−1.0
Mean occupancy	0.81 (0.13)
Mean R-free	0.27 (0.021)
Mean R-work	0.22 (0.010)

whereas only one fragment hit was found per site in B-E. A detailed overview of all fragments, which have been deposited to the PDB, is given in appendix, including the binding site information, fragment binding mode with refined $2F_o$-F_c electron density, event map, their chemical structure and molecular weight A.1.2.

Fragment binding site A. 40 fragment hits have been identified in the site A. The site A is the substrate binding site of NUDT5 located at the dimer interface between the catalytic domain of first monomer and the N-terminal domain of the second monomer. Fragment binding occurs between Trp28 and Trp46 originating from different monomers (Figure 3.4). Whilst both tryptophan residues are found in an open conformation in the apo-structure of NUDT5, upon fragment binding they move closer and align allowing for π-π stacking.

From 40 fragments located in the site A, the chemical structures of the fragments **15**, **16** and **17** display high similarity to adenine moiety of ADP-ribose, the endogenous substrate of NUDT5 (Figure 3.5). The adenine core of ADP-ribose is anchored between Trp28 and Trp46 through π-π stacking interactions, and specifically recognised via hydrogen bonds at the positions N1, N6 and N7, involving the amine nitrogen and the carbonyl oxygen of Glu47 backbone as well as the guanidinium moiety of the catalytic Arg51.[220]



Fragment 21 (5QJQ)

Fragment 41 (5QJ5)

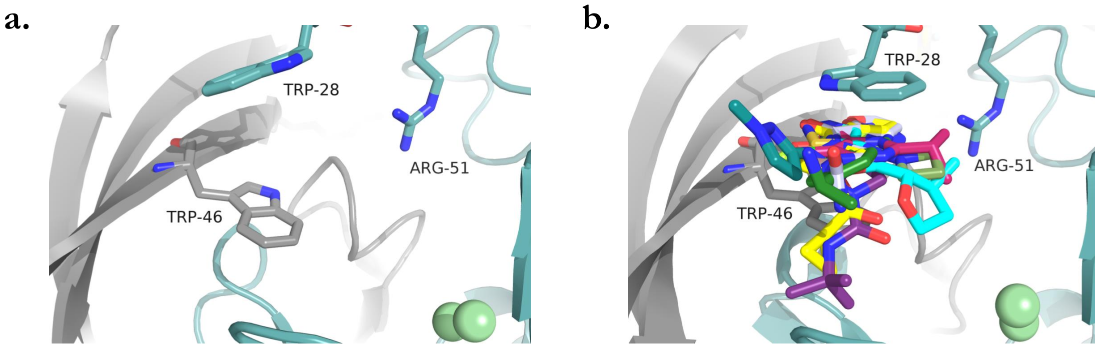
Fragment 42 (5QJX)

Fragment 43 (5QJ6)

Fragment 10 (5QJF)

Figure 3.3. NUDT5 fragment binding sites identified in the crystallographic fragment screening against DSiP library at the XChem platform, DLS. **a.** Overview of fragment binding sites for NUDT5 identified by XChem. Fragment binding sites are represented as surfaces. **b.** NUDT5 fragment binding sites – detailed view. **c.** Fragment hits identified in the sites A, B, C, D and E. For site A, only one fragment is shown as an example. The $2F_o$-F_c electron density map is shown at 1.00 σ as blue mesh.

Fragments **15**, **16** and **17** differ in their exact binding orientation from ADP-ribose (Figure 3.5). The positions of fragment **16** and **17** are inverted and fragment **15** is additionally turned by 90°. Nonetheless, all fragments are coordinated by Glu47 and Arg51 – the key residues responsible for substrate recognition and catalysis. Whereas hydrogen bond pattern of the fragment **15** exactly corresponds to the interaction profile of adenine core in ADP-ribose, fragment **16** and **17** only interact with Glu47 through



Substrate binding site of the apo-structure

Substrate binding site in presence of fragment binding

Figure 3.4. Comparison of the substrate binding site in the apo-structure and in presence of fragment binding. **a.** Substrate binding site in the apo-structure (PDB:6GRU): both tryptophan residues are found in an open conformation. The key residues Trp28, Trp46 and Arg51 are shown as sticks; Mn^{2+}-ions are shown as pale green spheres. **b.** Substrate binding site in presence of fragment binding – superimposition of multiple fragments. Upon fragment binding, Trp28 and Trp46 are aligned to interact with aromatic fragments via π-π stacking.

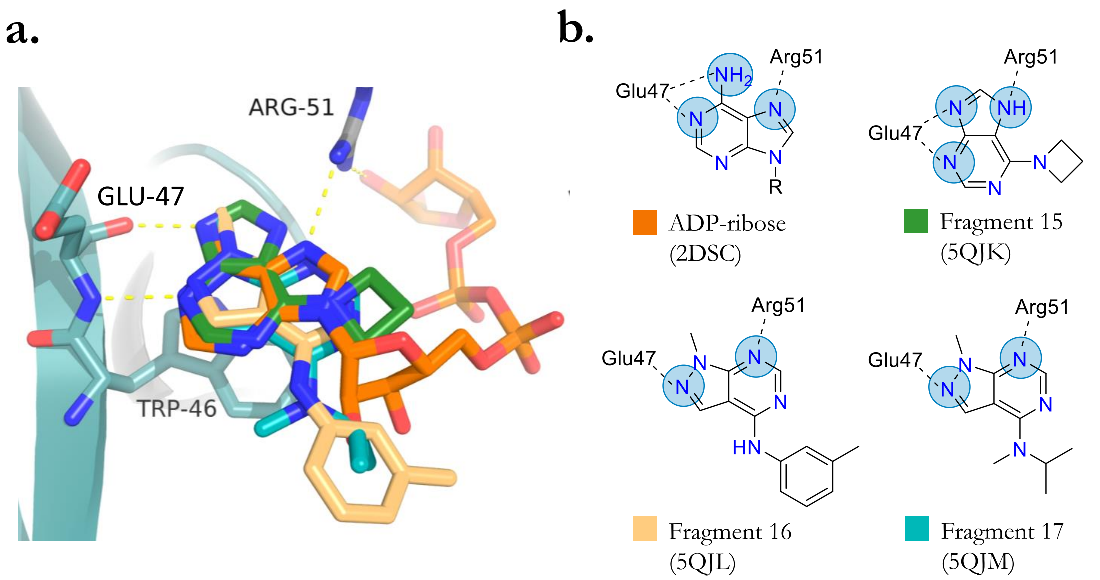
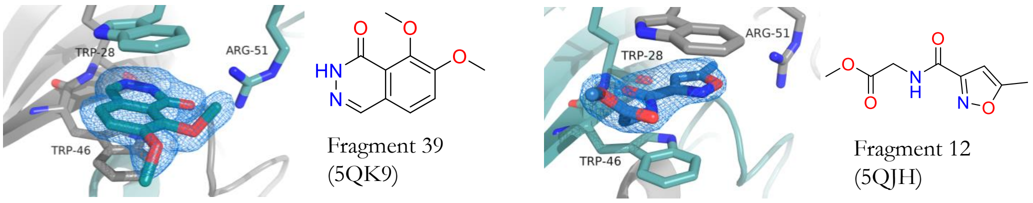


Figure 3.5. Fragment hit structures share high similarity to adenine moiety in ADP-ribose. **a.** Superimposition of crystal structures with ADP-ribose (shown in orange) and fragments **15** (green), **16** (yellow) and **17** (blue) in the substrate binding site. **b.** Chemical structures and interaction analysis of ADP-ribose and fragments **15**, **16** and **17**. All fragments display similar interactions to ADP-ribose involving the key residues Glu47 and Arg51. The atoms involved in hydrogen bond interactions are highlighted by circles in light blue.

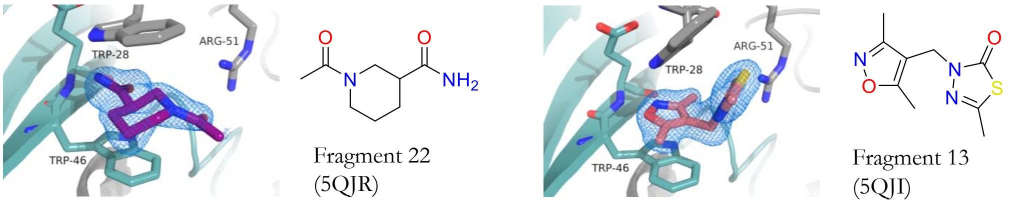
pyrazole nitrogen.

The fragment hits in site A display limited chemical diversity: their structures include various aliphatic and aromatic ring systems of different size, often containing multiple heteroatoms (appendix A.1.2). In addition, distinct conformational features associated with their binding mode inside the active site were observed (Figure 3.6). 37 contain an aromatic system consisting of at least one aromatic ring (appendix A.1.6). The observed heteroatom ring structures include 5-membered rings such as amide-conjugated thiazoles, oxazoles, pyrazoles, tri- and tetrazoles as well as isothiadiazoles. 6-membered rings and bicyclic aromatic structures comprise pyrimidine and purine derivatives, respectively. The majority of fragment hits containing an aromatic system identified in the substrate binding site A are anchored in place between Trp28 and Trp46 creating a π-π stacking interaction, as shown in Figure 3.6 a. Fragments **2**, **3** and **12** are noteworthy

a. **aromatic system with π-π stacking:** **b.** **conjugated without π-π stacking:**



c. **non-aromatic system:** **d.** **double π-π stacking:**



e. **open tryptophan sandwich:**

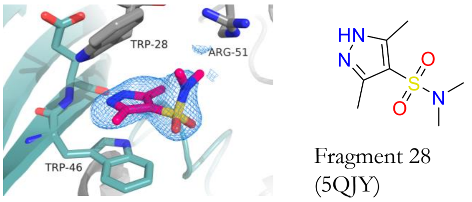


Figure 3.6. Examples of fragment hits and associated conformational observations in the binding site A. **a.** Aromatic systems displaying π-π stacking; **b.** Conjugated systems without π-π stacking interactions; **c.** Non-aromatic systems; **d.** Systems displaying double π-π by interacting with two separate aromatic rings; **e.** Open tryptophan sandwich conformation. The $2F_o$-F_c electron density map is shown at 1.00 σ as blue mesh.

exceptions as they do not display π-π stacking interactions with tryptophan residues despite containing a conjugated aromatic system (Figure 3.6 b). These fragments are primarily stabilised through hydrogen bonds.

Only three fragments (**11**, **22** and **34**) identified in the site A are non-aromatics as they only contain aliphatic ring structures including piperidine and piperazine (Figure 3.6 c). Presumably, hydrophobic Van-der-Waals interaction as well as the entropic effect drive the binding of these fragments.

Fragment **13**, which contains two aromatic rings hyperconjugated via sp^3 C-H bond, was found to interact with both tryptophan residues in the active site by creating two separate π-π stacking interactions between the isoxazole moiety and Trp46 as well as the thiadiazole ring and Trp28 (Figure 3.6 d).

The binding mode of the fragment **28** only allows for π-π stacking between Trp46 and pyrazole ring of the fragment, preventing the alignment of the tryptophan sandwich through a steric clash and leaving Trp28 in an open conformation (Figure 3.6 e).

Overall, fragment binding site A is particularly well sampled. The majority of the fragment hits represents aromatic systems with high proportion of heteroatoms. As many share high overlap in their binding position − particularly in regard to the π-π interaction with Trp28 and Trp46 − the binding site A can be described as promiscuous. This finding is unsurprising, considering the generic nature of the endogenous substrate ADP-ribose.

Fragment binding site B. Only one fragment hit (fragment **41**) was identified in the binding site B. This site is located in close proximity to the binding site A and has not been described in literature previously. Closer inspection reveals that site B is a deep and less accessible pocket inside the substrate binding site, associated with a conformational change.

The fragment **41** interacts with the protein chains through multiple hydrogen bonds (Figure 3.7 a). Both urea nitrogen atoms are stabilised by hydrogen bonds through Asp133. In addition, the carbonyl oxygen is coordinated through the catalytic Arg51 side chain.

PanDDA analysis has facilitated the identification of this binding event with low occupancy (0.4). Whereas the event map density for fragment clearly provides strong evidence for the presence of the fragment **41**, the initial $2F_o$-F_c electron density was more

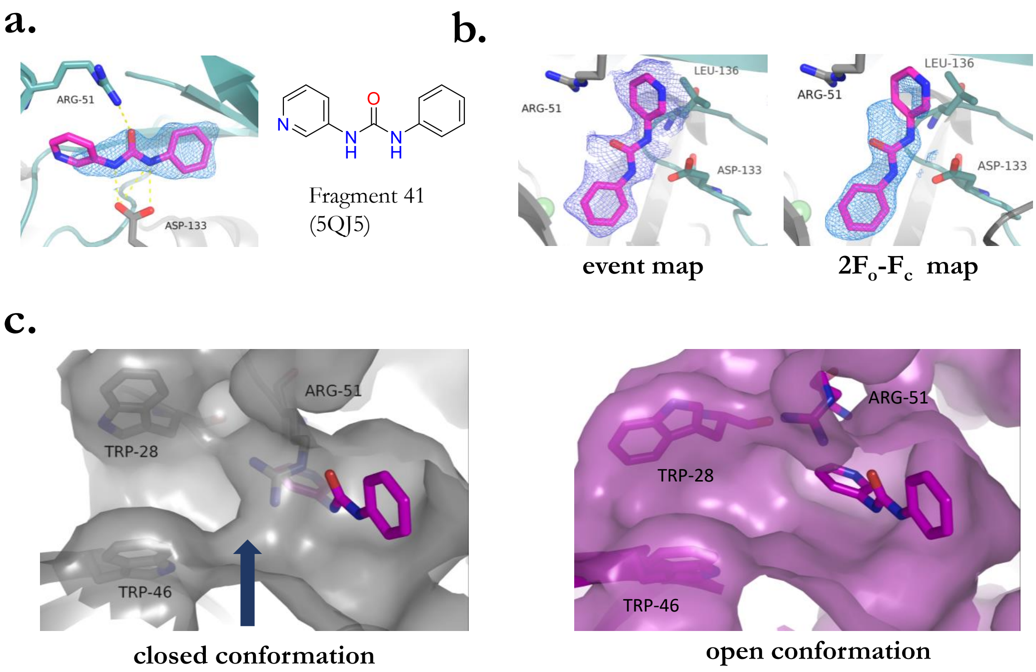


Figure 3.7. Analysis of the fragment binding site B. **a.** Crystal and chemical structure of fragment **41** identified in the site B. Fragment **41** is stabilised by hydrogen bonds to Arg51 and Asp133. **b.** Whereas the event map resembles the ligand shape very well, the density coverage of the final $2F_o$-F_c map is weaker. The event at σ corresponding to 2x(1-BDC) and $2F_o$-F_c at 1.00 σ shown as purple and blue mesh, respectively. **c.** In absence of fragment **41**, the site B is not accessible due to the position of Arg51. Upon binding of fragment **41**, Arg51 is displaced, creating a deep pocket inside the active site. The surface of closed and open conformation is shown in grey and violet, respectively. Fragment **41** is shown in magenta in both structures for comparison.

ambiguous and difficult to interpret (Figure 3.7 b). This ambiguity is a consequence of superimposition of two different states: the ground state in absence of fragment and the bound state in its presence. The deconvolution of both states revealed a conformational change associated with fragment binding.

Upon binding of fragment **41**, the catalytic Arg51 is displaced, opening up a deep cavity and connecting the site A and B into one deep pocket which allows to accommodate the fragment (Figure 3.7 c). In addition, the displacement of Arg51 causes a shift of the loop segment between the residues Arg51 and Thr58, which connects the N-terminal and catalytic domain of NUDT5.

These findings are particularly interesting for ligand design in this novel, unknown binding pocket. Arg51 appears to be a gate-keeper residue, which is opened up by the H-bond network of the ligand. This binding in turn seems to induce a conformational change of the protein. The effect is remarkable given the small size of the fragment.

Fragment binding sites C, D and E. The XChem fragment screen identified three further sites C, D and E, with one solitary binder per site. Due to the location on the protein surface and high solvent exposure, these fragment hits are of less interest as their binding is likely to be an artefact of crystallographic symmetry axes. For reasons of completeness, the structures of these hits have been refined and deposited.

In summary, crystallographic screening revealed fragment binding across five fragment binding sites. The discovery of a novel binding site (B) and particularly well characterised substrate binding site (A) both present multiple opportunities for further elaboration.

3.2.3. Results of the ^{19}F NMR-based fragment screening

In addition to the crystallographic screening campaign, NUDT5 was screened against a library of fluorinated fragments by ^{19}F NMR. The BIONET fluorine library (KeyOrganics) consists of 461 compounds which are combined into 24 cocktail mixtures with 16 to 20 fragments per mixture to allow unambiguous identification.

Prior to the screen, optimal screening ratio of ligand mixture to protein was determined. Initial characterisation was carried out with multiple ligand mixture: protein molar ratios from 10:1 to 33:1. The ligand mixture to protein ratio of 33:1 (50:1.5 μM) was determined as optimal for screening, as the initial test demonstrated signal broadening sufficient for binding detection of compounds present in the tested mixture.

Fragment screening was carried out using Avance Neo 700 MHz instrument equipped with a QCI-F cryoprobe at 298 K at the Bruker site in Fällanden, Switzerland. For the CPMG experiment, spectra for each ligand mixture at 50 μM in presence and in absence of NUDT5 at 1.5 μM were acquired. All spectra were recorded with 20 ms and 200 ms relaxation delay.

Data analysis was performed using Fragment-Based Screening tool (FBS, Bruker), allowing to generate an overlay of ^{19}F-CPMG 20 ms and 200 ms spectra to calculate the integral ratios and the resulting CPMG-effect (see 2.9.1). All fragments exhibiting a CPMG-effect of at least 40% were characterised as a hit. An example of data analysis is demonstrated in Figure 3.8. In mixture 3, signals of the compounds **AS-5627** and

MS-1061 exhibit significant line broadening accompanied by decrease of signal intensity. With a ^{19}F-CPMG effect of 87% and 95%, respectively, these compounds fulfil the hit definition and are identified as binders. Non-binding fragments, such as fragment **PS-6631**, do not experience any change in their environment and therefore retain their line width and intensity.[146]

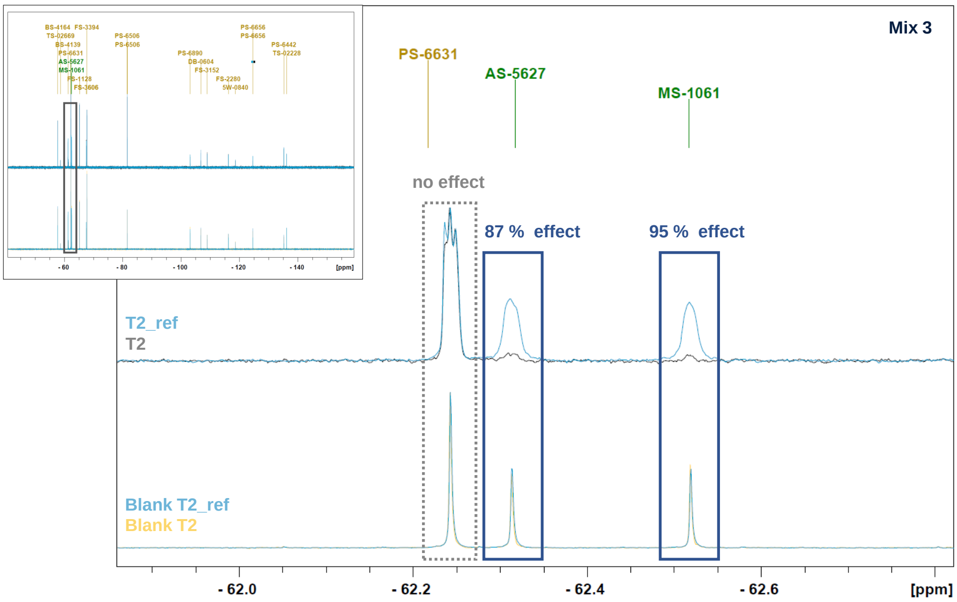


Figure 3.8. Ligand-observed NMR spectra of ^{19}F CPMG experiment with the fragment screening mixture 3. Compound reference spectra were recorded using samples with fragment mixtures at 50 μM. Protein samples contained 1.5 μM NUDT5 and 50 μM fragment mixture. The fragment screening was carried out at Avance Neo 700 MHz instrument with a QCI-F cryoprobe at 298 K. NMR spectra shown from bottom to top: overlay of T2 200 ms (yellow) and 20 ms (blue) delay experiment of compound in absence of protein; overlay of T2 200 ms (grey) and 20 ms (blue) delay experiment of compound in presence of protein; expected signals from the peaks list of the compounds present in the mixture are shown as lines. The fragments **AS-5627** and **MS-1061** exhibiting a ^{19}F-CPMG effect of 87% and 95%, respectively, are binders. The fragment **PS-6631** does not shown any signal decrease and is characterised as non-binder.

This screen identified 20 fragment binders in 11 fragment mixtures from the BIONET fluorine fragment library (4.3% hit rate). The chemical structures of the fragments with corresponding signal attenuation are shown in Figure 3.9. Out of 20 fragment hits, 10 contain aliphatic CF_3 or/and CF_2 groups and 10 aromatic CF groups. All fragment hits are aromatics: their chemical structures include one 5-membered ring, six 6-membered rings, 13 bicyclic or bis-aromatic systems.

As ^{19}F NMR screening does not provide any structural information, an orthogonal

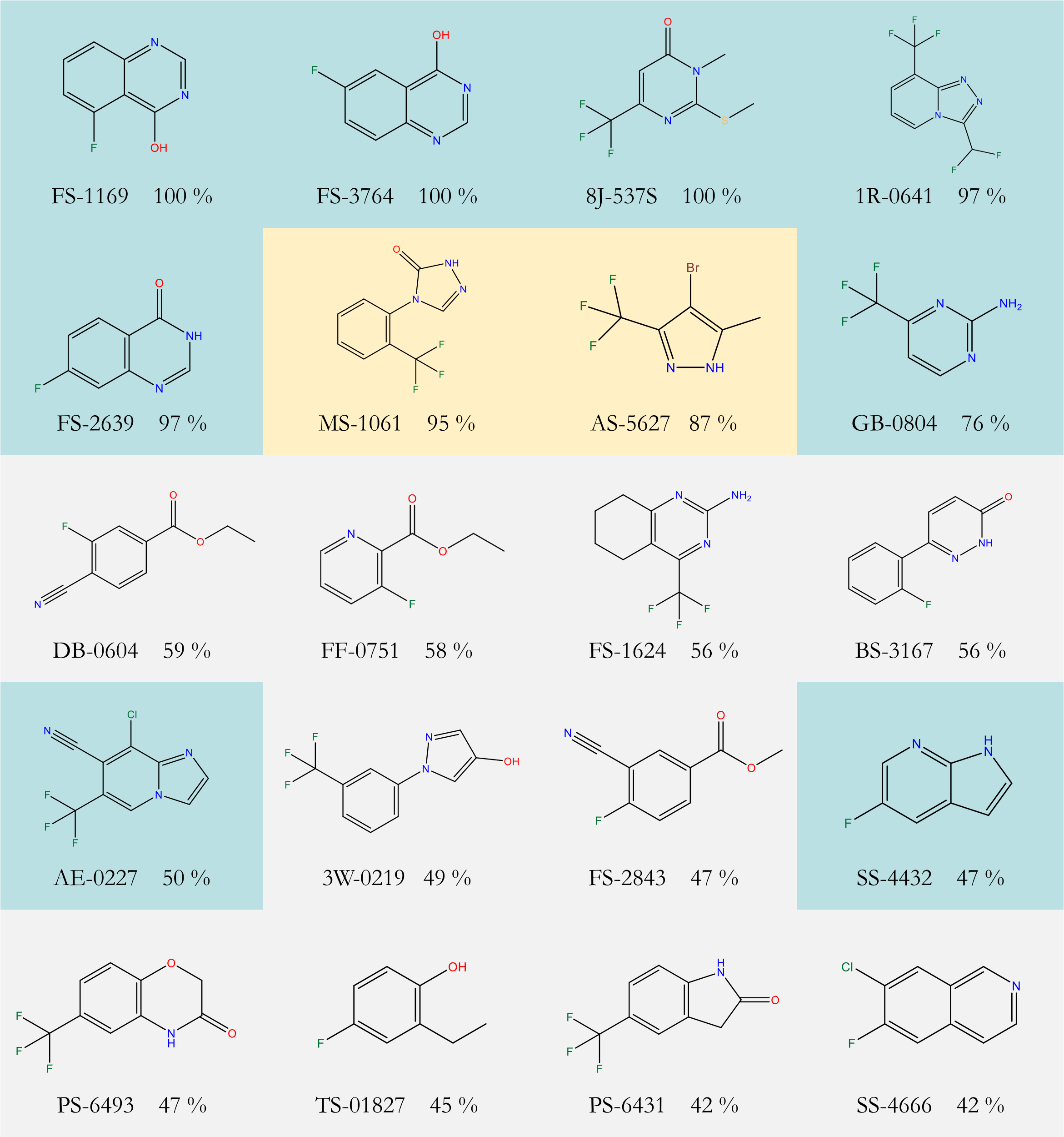


Figure 3.9. Chemical structures of fragment hits identified in ^{19}F NMR screen. The name and observed signal attenuation are indicated for each fragment. Fragment hits that have been confirmed by X-ray crystallography are shown blue. Fragment hits for which some binding evidence was present, but was insufficient for confident modelling are shown in yellow.

method for determination of the fragment binding location is required. With an already established crystallisation system for NUDT5, the binding of the compounds from this screening campaign was validated by X-ray crystallography.

3.2.4. Crystallographic validation of fragment hits from ^{19}F screen

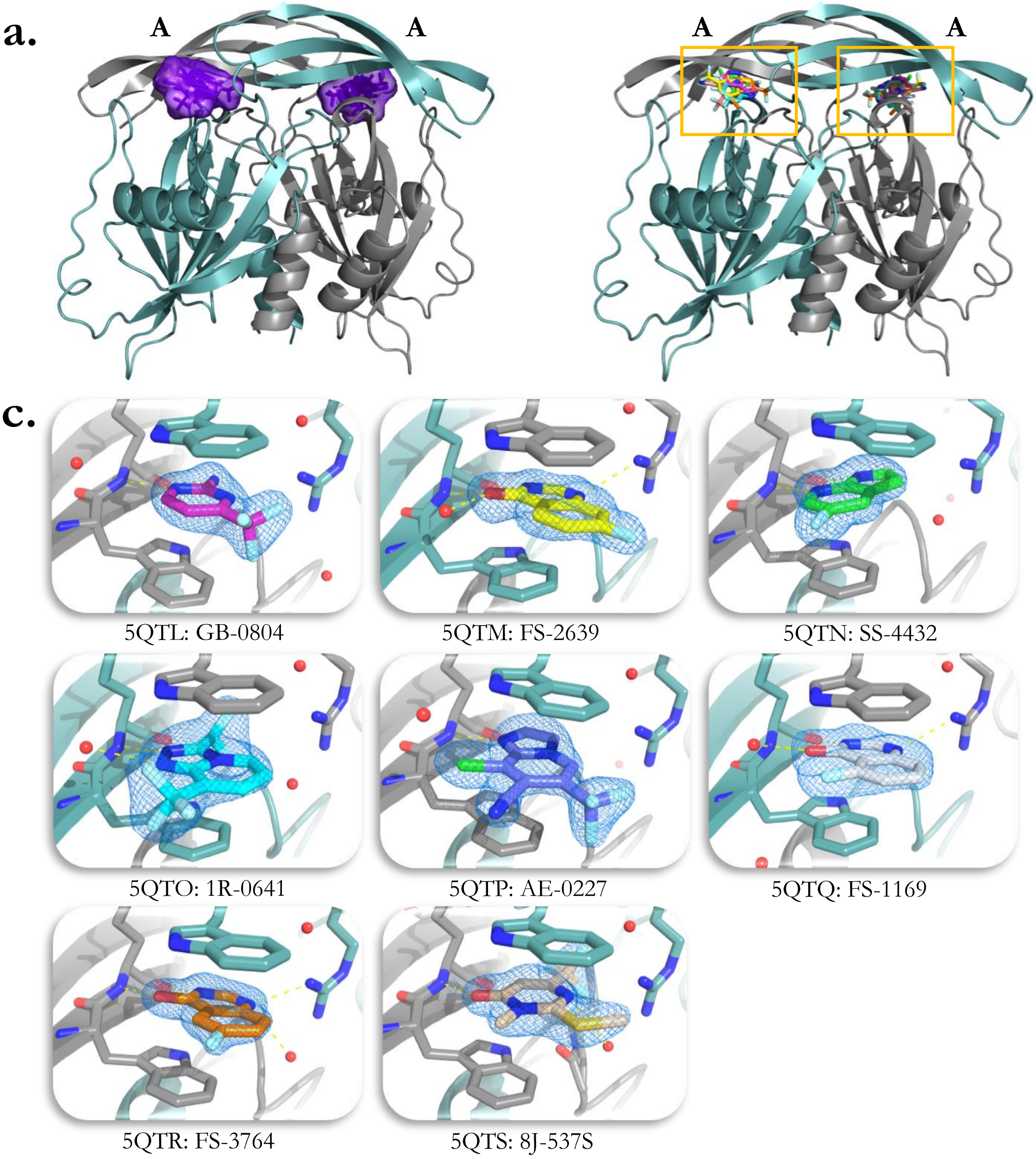
The crystallisation system established for the XChem fragment screen was used to determine the binding location of the ^{19}F fragment hits. The 20 BIONET fragment hits were soaked at concentrations of 100, 200 and 500 mM (10% DMSO in crystal drops) into NUDT5 crystals with a soaking time of 25-30 min. The experiments were performed in duplicates for each concentration point and were carried out at the XChem platform at DLS. The data were collected at beamline I04-1. From 120 datasets, 111 yielded diffraction that could be processed successfully.

In total, structures of eight unique protein-fragment complexes could be identified. As soaking experiments were performed in duplicates at multiple concentrations, 1-5 datasets for each identified fragment were obtained. The dataset with the highest diffraction limit and best electron density maps for the ligands was selected for refinement and deposition. Both average diffraction limit (1.75 Å) and average ligand occupancy (0.8) are similar to the results obtained in the XChem screen. The statistics of the diffraction data and structure refinement are summarised in Table 3.3.

Table 3.3. Crystallographic key parameters overview for refined protein-ligand complexes for NUDT5 and fragment hits from the BIONET fluorine fragment library (KeyOrganics).

Crystallographic parameters for protein-ligand structures	
Space group	P1
Cell dimensions:	
a, b, c (Å)	49.19 (0.13), 59.81 (0.06), 80.05 (0.21)
α, β, γ	79.40° (0.07), 81.49° (0.15), 75.72° (0.11)
Diffraction range (Å)	1.55−1.85
Mean diffraction (Å)	1.75 (0.094)
Occupancy range	0.68−0.9
Mean occupancy	0.8 (0.054)
Mean R-free	0.24 (0.014)
Mean R-work	0.22 (0.012)

As shown in Figure 3.10, all eight fragment hits were identified in the substrate binding site A. With one exception, each protein chain contains a copy of the fragment, resulting



5QTL: GB-0804

5QTM: FS-2639

5QTN: SS-4432

5QTO: 1R-0641

5QTP: AE-0227

5QTQ: FS-1169

5QTR: FS-3764

5QTS: 8J-537S

Figure 3.10. NUDT5 fragment binding sites identified in the ^{19}F screening using BIONET fluorine fragment library (KeyOrganics) and confirmed by crystallography. **a.** NUDT5 fragment binding sites. **b.** NUDT5 fragment binding sites – detailed view. Fragment binding was only observed in the active site. **c.** Eight protein-ligand complexes for confirmed ligands. The 2F$_o$-F$_c$ electron density map is shown at 1.00 σ as blue mesh.

in four copies per asymmetric unit. Inside the binding site, they are anchored in place between the key residues Trp28 and Trp46 displaying π-π stacking interactions as well as multiple hydrogen bonds.

The chemical diversity of the validated ^{19}F hits is limited. Fragments **FS-2639**, **FS-**

1169 and **FS-3764** share the exact same scaffold and only vary in the position of ^{19}F in the ring structure. Unsurprisingly, all three fragments bind in the same orientation. Multiple fragments probe the same interactions such as **GB-0804** and **SS-4432** both create hydrogen bonds to the amide nitrogen and the carbonyl oxygen of Glu47 backbone.

Out of 20 fragment hits identified by ^{19}F screening, binding of eight fragments was successfully confirmed by X-ray crystallography, providing information upon their binding location. Interestingly, six binders which exhibited strong signal attenuation in the GPMG experiment could be confirmed using crystallographic approach, suggesting a possible coherence between both methods (Figure 3.9), further discussed in 3.3.

3.2.5. Analysis of protein-fragment interaction using PLIP

Different fragment libraries were used in both screening campaigns: whereas NUDT5 was screened against the DSiP library (Enamine)[72] in the XChem screen, ^{19}F NMR fragment screening required a fluorinated collection (BIONET library, KeyOrganics)[228]. As both compound libraries do not overlap, identified fragments hits cannot be utilised to compare the outcome of both screens. Therefore, enumeration of interaction between NUDT5 and the fragment hits provides a reasonable metric for comparison.

While fragment scaffolds are often seen as chemical starting point for elaboration towards affinity gain[254], the molecular interactions between the target protein and its fragment hits provide by themselves valuable insights into molecular recognition and useful guidance for elaboration. A comprehensive sampling of the protein pocket thus provides a 3D map of interactions informative for ligand design.[76] Hence an argument can be made that an efficient screening technique should not only identify suitable chemical matter as starting points for elaboration, but also enumerate a maximal number of useful interactions.

The Protein-Ligand Interaction Profiler (PLIP) tool enables the systematic characterisation of protein-fragment interactions.[236] PLIP is a freely available tool for automated detection and visualisation of non-covalent protein-ligand interactions, which can be accessed both as a web-server or python source code.[236] For each structure, PLIP generates a list of detected interactions, including hydrogen and halogen bonds, π-π stacking, π-cation interactions, salt and water bridges.[236] In addition, the protein ligand patterns are visualised in corresponding PyMOL sessions.

The enumeration of the interactions of both fragment sets from XChem and ^{19}F was performed in batch mode using python source code of PLIP (version 1.4.5) on GitHub allowing to process all protein-ligand structures in parallel.[236] The crystal structures of protein-ligand complexes from both screens were provided by supplying their PDB ID. As this analysis requires structural data, only eight fragments from the ^{19}F NMR screen that could be confirmed by crystallography were used. The identified interactions across all protein-fragment structures from the XChem and from ^{19}F NMR screening campaigns were collated separately. Both fragment sets were manually examined for redundancies and contain only unique interactions (appendix A.1.1).

As shown in Figure 3.11, the XChem screen identified 37 unique interactions across five binding sites, including 17 distinct hydrogen bonds. The fragments from the ^{19}F NMR screen confirmed by crystallography sampled 11 unique interactions in the fragment binding site A. Only three hydrogen bonds were identified, which correspond to the residues that contribute to the binding of the natural substrate ADP-ribose: amide nitrogen and carbonyl oxygen of the Glu47 backbone as well as guanidinium group of the Arg51 side chain.

Unsurprisingly, the validated ^{19}F NMR hits sampled more halogen bonds (four interactions) compared to the XChem fragments (two interactions). However, the XChem screen additionally identified a chlorine mediated halogen bond to Arg51 side chain (PDB:5QJB).

Both expected π-π stacking interactions including Trp28 and Trp46 in the substrate binding site A were found by both screening techniques. All fragment hits from ^{19}F NMR and most fragment hits from the XChem screen (exceptions are shown in Figure 3.6 b) form the π-π stacking, highlighting the promiscuous nature of the site A. Consequently, these interactions are essential to generate potent binders, yet the overall diversity of the interaction found in the XChem screen allows for additional starting points towards an affinity gain.

The XChem screen revealed 17 water mediated interactions, with 15 of them present in the site A. In the ^{19}F NMR screen followed by crystallographic confirmation only one water bridge was observed. Whereas it is unclear whether these interactions involve conserved water molecules, they provide further sampling of the active site, indicating potential space for growth and capture energetically rewarding polar interactions.

In summary, except for three additional halogen bonds, the combination of the ^{19}F NMR screen and crystallographic validation does not provide any novel interactions for

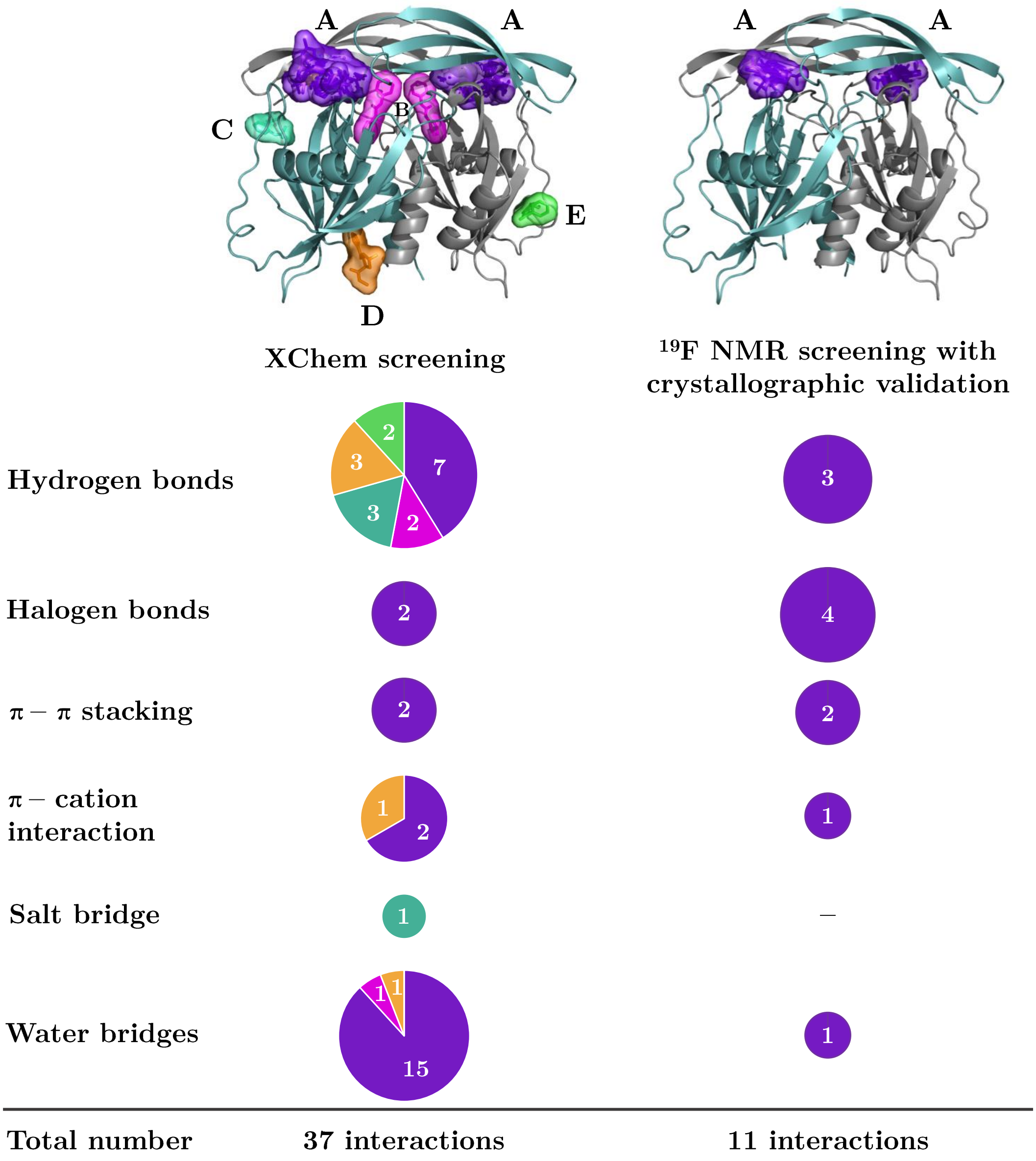


Figure 3.11. Interaction analysis of fragment hits from XChem and ^{19}F NMR screening with subsequent crystallographic validation using PLIP. The XChem campaign was conducted at 500 mM soaking concentration, translating in approx. 25 mM drop concentration in presence of 1.1 mM protein; the ^{19}F NMR screen was carried out at 50 μM in presence of 1.5 μM protein. Each type of interaction is represented as a pie chart, with colours indicating the fragment binding site.

NUDT5 beyond the sampling through the XChem screen. Although it is clear that the XChem screen performed at a high concentration reveals more fragment binding sites and demonstrates greater diversity of the interactions, many of these are presumably of weak nature. It is yet to be demonstrated whether such weak interactions can be useful in a lead discovery campaign. This will be further discussed in the chapter 4.

3.3. Discussion and Outlook

Fragment-based screening was conducted with the aim of identifying novel small molecule binders for the breast cancer target NUDT5. The crystallographic fragment screen using the XChem platform at DLS successfully revealed 43 fragment hits clustered in five binding sites. The ^{19}F NMR screening campaign identified 20 fragment binders – of which eight could be confirmed by crystallography, all clustered in the substrate binding pocket. In the following, a perspective on the advantages and limitations of both screening technologies is given and its impact on detection of weak binders discussed.

Library design impacts on variety of sampled interactions. The design of the ^{19}F fragment library is limited by synthetic accessibility of fluorinated compounds. The majority of these contain aromatic CF and aliphatic CF_2 or CF_3 groups. In consequence, the principal moments of inertia (PMI) of the BIONET library, which describes the overall shape of compounds, are mainly disc or rod-like, implying a library dominated by flat aromatic compounds.[228] The DSiP library was developed on the premise of rapid generation of analogues using standard parallel chemistry.[71] The fragments were selected to provide maximal diversity while being poised for a limited set of high yielding chemical reaction for further elaboration. As the PLIP analysis reveals, the differences in library design translate into sampling of a higher number of unique interactions for NUDT5 using the DSiP library.

Differences in hit definition prevent comparison across screening methods. Most biophysical techniques do not measure binding affinity directly. Identification of fragments in a ^{19}F screen utilises change in relaxation properties of a fragment which occurs upon binding to a macromolecular target as a proxy for its affinity. This change is quantified through integration of the signals and comparison to the reference samples. As a result of the quantitative nature of the analysis, a threshold for hit definition is chosen – a lower threshold yields more hits. As a consequence, the hit rate is a poor measure of the success of a fragment screening campaign. When comparing results, other information such as the number of identified binding sites, ligand scaffolds and molecular interactions are more informative for further elaboration. In light of these criteria, the XChem screen offers a higher information content than ^{19}F screen.

Validation cascades provide orthogonal information but reduce number of hits. While ^{19}F NMR screening is a rapid and efficient method for identification of fragment binders, it lacks the information upon their binding location, requiring an orthogonal method.[86,255] In this work, the structural information of only eight out of 20 fragments could be obtained through a validation step using crystallography. Such discrepancies in hit detection for validation are well-described in literature (see 1.2.2). These can result from either inherent sensitivity of the methods or experimental requirements such as compound solubility, impacting on the recovery of both weak and strong binders. Consequently, not all identified hits can be validated and thus potentially valuable information is lost.

Crystallographic screen directly provides structural information. Crystallographic fragment screening circumvents both limitations – hit definition and binding site determination – by providing atomic coordinates of the protein-ligand complex and is therefore a much more informative technology. Artefacts including crystal contact binders are easily spotted when considering the crystallographic axis. Crystallographic screening therefore rarely delivers false positives, independently of the binding affinity of the hit.[92]

High sensitivity and concentration of crystallographic screening increase the number of fragment hits. X-ray crystallography outmatches other biophysical methods in its sensitivity, crystallographic fragment screen can often reveal extremely weak binding hits with affinities presumably above 10 mM.[23] This results in identification of a larger number of fragment binding sites and a larger number of interactions, as demonstrated by the PLIP analysis. Therefore, it is unsurprising that the XChem fragment screening revealed a higher hit rate and sampled more protein-fragment interactions compared to the results of the ^{19}F screening campaign – keeping in mind that some ^{19}F hits are dismissed by the hit definition filter and the orthogonal validation.

In addition, soaking of fragments at high concentrations is commonly utilised practise in crystallographic screenings.[14,35,79,89] Screenings at the XChem platform are typically carried out at 500 mM. High concentration presents a large excess of the fragment and ensures detection of low affinity compounds at sufficient occupancy.[79,109,256] In addition, high soaking concentration subsequently increases the number of fragment hits identified, including weaker binders or crystal contacts artefacts.

The combination of high sensitivity and high soaking concentration in crystallographic screening therefore reveals extremely weak binders, which cannot be distinguished from higher affinity fragments due to the lack of quantitative information. This hampers prioritisation and can result in a blind, resource-intensive elaboration if unsuitable fragments are selected as starting points. On the other hand, the wealth of information including different binding sites, chemical diversity of potential starting points and sampled molecular interactions can be valuable for compound progression. Developing a quantitative approach for weak binders would allow to overcome the limitations in prioritisation and thus allow to bridge the best of both worlds. Chapter 4 and 5 approach this problematic using concentration-dependent X-ray crystallography and ligand-observed NMR methods, respectively.

Discovery of a novel binding site for NUDT5. Beyond the problem of weak binders, the example of the XChem fragment screen of NUDT5 reveals new insights for this drug target. While the majority of the fragments was found in the active site of the enzyme, the XChem screen uncovered a novel, previously unknown binding pocket B in close proximity to the active site. This deep cavity is guarded by Arg51, a key catalytic residue which functions as a gatekeeper to the deeper cavity, requiring a conformational change for the fragment to bind. This novel binding mode bears the promise of new chemical entities for the inhibition of NUDT5. Chapter 6 explores the elaboration of fragments tailored to take advantage of this new site.

4

Concentration-dependent crystallographic soaking for prioritisation of fragment hits.

Contents

4.1. Introduction

Advances in automation and synchrotron science propelled X-ray crystallography to a powerful method for fragment hit identification.[79,99,257] Crystallographic fragment screening advantageously provides information upon binding location and molecular interaction of the protein-fragment complex.[257] However, identified fragment hits lack quantitative information and are rarely detectable by commonly used biophysical methods.[14,258] Two underlying reasons are the sensitivity of X-ray crystallography and high fragment concentrations.

Among all fragment screening techniques, X-ray crystallography has the widest detection range revealing not only strong binders, but also extremely weak fragments, with binding affinities over 10 mM.[23] These weak binders are often missed by other methods, as their detection covers lower affinity ranges, creating a biophysical detection gap.[86,89]

In addition, crystallographic screening utilises fragment soaking at high concentration, often above 100 mM.[39,92,259] At the XChem platform at DLS, screening campaigns are routinely carried out using a fragment library at 500 mM, with effective concentration of 50-100 mM. Such high concentrations exceed assay limitations of other biophysical techniques, including NMR and SPR, where they cause aggregation and other assay artefacts.[92]

In absence of quantitative information, prioritisation of weak binders such as XChem fragments for elaboration and medicinal chemistry activities is challenging.[159] Decisions on the hits purely rely on structural information – this is the opposite of screening campaigns using other biophysical methods, which provide quantitative information but lack structural data. The selection of XChem fragment hits for elaboration is thus highly project-specific and often depends on the number of identified hits, their distribution across the protein structure, their synthetic tractability, but also available resources and expertise. Beyond hit selection, monitoring of the indispensable affinity gain while progressing weak hits is equally hampered by the lack of quantitative information.

Integrating quantitative techniques into the XChem platform would be a stepping stone for accelerating compound elaboration. In light of the difficulties involved with the biophysical quantitation of weak binders, using crystallography as a readout would provide the required sensitivity and seamlessly integrate into the XChem pipeline. The use of crystallography for ligand ranking has been suggested in the literature.[260,261] In early experiments, Sauter *et al.*[260] used crystallographic soaking of ligands at three concentration to estimate their relative affinity for a second binding site in influenza virus hemagglutinin. More recently, a comprehensive study by Tanaka and co-workers[261] performed combinatorial competitive soaking experiments to rank bovine pancreatic trypsin inhibitors by their *in-crystal* binding affinities. In addition, the authors demonstrate that the relative crystallographic affinities correlate with inhibitory activities in solution.[261] Using the conformational change of a residue in the active site, Wu *et al.*[262] were able to quantify crystal binding affinities of dipeptide ligands to cyclophilin 3 by concentration-dependent soaking, presenting results in agreement with biophysically determined K_D values.

In this chapter, the correlation between crystallographic soaking concentration and the number of detected fragment hits is systematically investigated, using NUDT5 as a model system. As described in chapter 3, NUDT5 presents multiple hits in the same pocket, highlighting the necessity of prioritisation. By exploring a wide range of soaking concentrations (from 25 to 500 mM) in high throughput crystallography setup, a concentration dependency in regard to the number of detected hits and their corresponding interactions is observed. In addition, this study provides evidence for correlation between lowest detectable soaking concentration (LDSC) and ligand affinity as determined by SPR, enabling ranking of fragment hits according to their binding strength.

Building on this proof of concept, a low-concentration counter-screen can be used to triage identified fragment hits in crystallographic fragment screening for further elaboration and to monitor their progression. The value of this approach becomes apparent when quantifying weakly binding fragments, whose affinities lie outside the detection range of commonly used biophysical techniques (e.g. SPR, MST, ITC). The combination of wealth of information provided by weak fragment hits detected at high soaking concentration, and the prioritised lead fragments should facilitate a streamlined and structure-guided elaboration into potent compounds.

4.2. Results

4.2.1. Crystallographic soaking concentration correlates with number of detected fragment hits

Exploiting X-ray crystallography for quantitation of ligand binding has been an ongoing effort. Beyond the relative ranking of ligands[260,261], there is no study systematically exploring how soaking concentration and ligand affinity are linked. Understanding these principles would enable to harness the unmatched sensitivity of X-ray crystallography for the ranking and quantification of weak fragment binders.

The influence of crystallographic soaking concentration on the number of identified fragments hits was investigated by conducting systematic concentration-dependent soaking experiments on NUDT5. For that, the original set of 43 fragment hits from the XChem campaign (chapter 3) was re-screened at five concentrations: 500 mM, 250 mM, 100 mM, 50 mM, 25 mM, under the same conditions as the original fragment screen, while keeping the DMSO content at 10% for each concentration. Two replicate datasets were collected for all fragments and concentration points. Experiments, which did not result in a dataset due to experimental issues or had a diffraction limit above 2.2 Å, were repeated until at least duplicates for each fragment and concentration were obtained.

494 datasets were successfully analysed using PanDDA.[113] Upon inspection of the highlighted events and model building, presence of the fragments at each concentration was confirmed by two cycles of refinement. The identical binding mode for the fragments across soaking concentrations was validated by structural alignment to the previously deposited protein-fragment structures. For further analysis, fragment hits that were detected in at least one of the duplicates were defined as being present at a given concentration.

Figure 4.1.a shows the relationship between crystallographic soaking concentration and number of fragments detected. All fragment hits display a consistent behaviour regarding their crystallographic identification: they are detectable throughout the concentration range and disappear at a specific cut-off concentration, referred to here as the lowest detectable soaking concentration (LDSC).

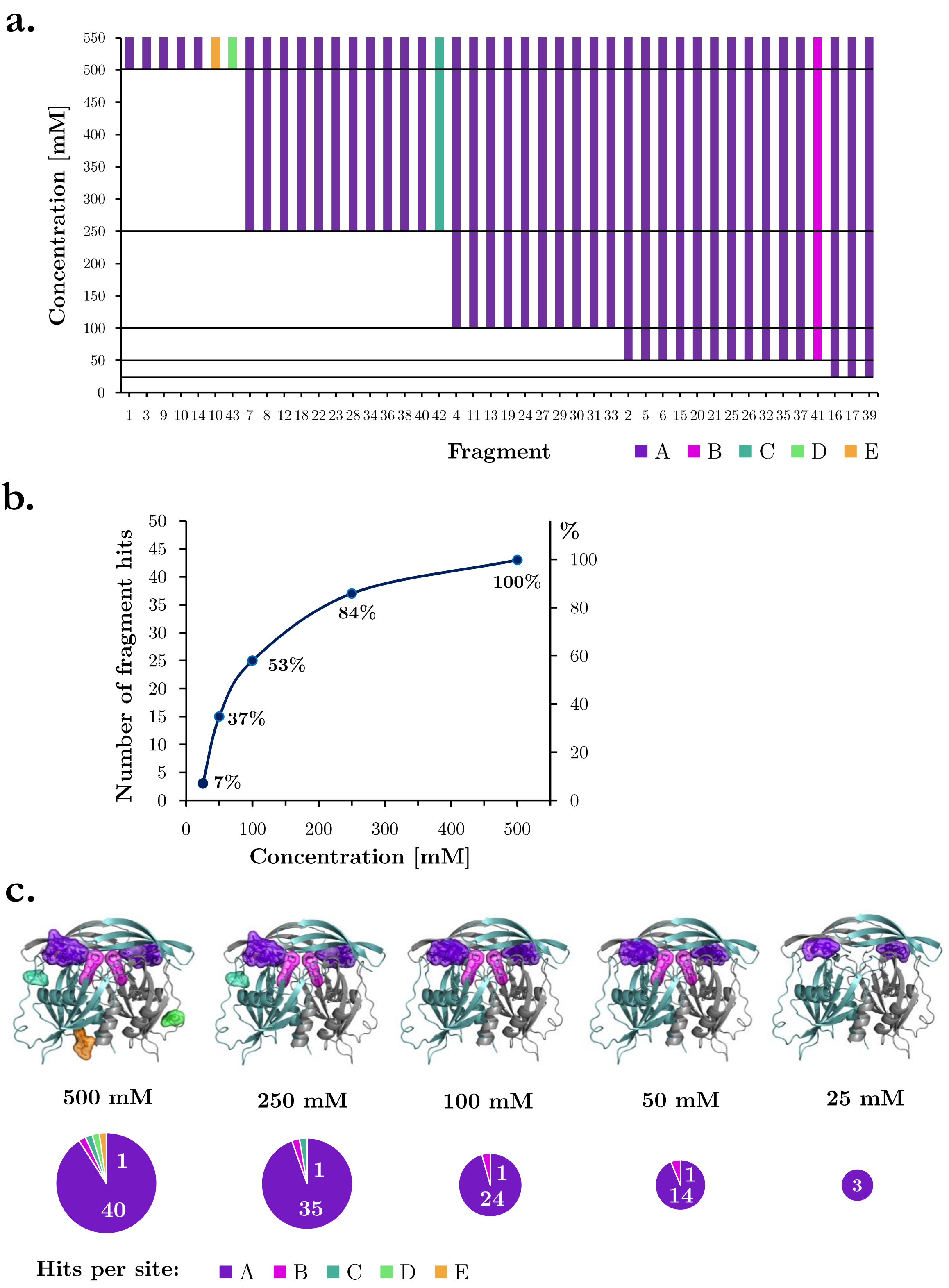


Figure 4.1. Concentration-dependent crystallographic screening of fragment hits binding to NUDT5. **a.** The crystallographic soaking concentration influences the number of fragment hits identified. Each fragment displays a specific detection cut-off, the lowest detectable soaking concentration (LDSC). All data was acquired in duplicates, with all

Figure 4.1. (***continued***) fragments exhibiting a consistent behaviour across the concentration series. **b.** Correlation between crystallographic soaking concentration and number of fragment hits identified at a given concentration. Percentages indicate the recovery of detected fragments at each concentration relative to the original number of XChem fragment hits at 500 mM. **c.** The number of fragment hits diminishes with decreasing soaking concentration. For NUDT5, crystallographic artefacts and solvent exposed surface-binders disappear first. Below 100 mM, only fragments that form high-quality interactions are detected.

It classifies fragments hits into distinct groups: some fragment hits, such as fragment **1** and **3**, are only detectable at the highest soaking concentration of 500 mM and could not be identified at any other concentration. Others, such as fragments **7** and **13**, can be detected throughout the tested concentrations until 250 mM and 100 mM, respectively. At the lowest concentration, only three fragments, **16**, **17** and **39**, are detectable. This consistent behaviour indicates underlying relationship between the lowest detectable concentration and fragment binding.

The curve in Figure 4.1.b shows a quasi-logarithmic relation between the soaking concentration and the number of detected fragments. For each concentration, a hit recovery rate was calculated relative to the original number of XChem fragment hits identified at 500 mM. By halving the concentration to 250 mM, 84% of the XChem set remain detectable in the crystallographic experiment. At the lowest soaking concentration of 25 mM, only three fragment binder can still be identified. Thus, decreasing the soaking concentration reduces the number of fragment hits for NUDT5.

To understand the causality for the observed quasi-logarithmic relationship, the identity and binding location of fragments disappearing with decreasing concentration was investigated (Figure 4.1.c). While soaking experiments at 500 mM result in identification of five fragment binding sites and 43 fragment hits, at 250 mM three of the fragments in binding site A are missing along with the surface-exposed binders in the sites D and E. This implies that crystallographic artefacts are removed first when concentration is decreased. At 100 mM, only 23 fragment hits in the active site A can still be detected. In addition, the solitary binder located in the shallow, solvent-exposed site C cannot be observed, suggesting that high-quality interactions are required to remain at 100 mM. Soaking experiments at 50 mM result in the detection of only 15 fragment hits in the site A and a singleton in the site B. All these binders show combinations of high quality interactions including key hydrogen bonds in addition to the aromatic stacking. At 25 mM – the last concentration at which fragment from the original sets are still detectable – only three fragment hits are present in the active site A, which are structural mimics of the natural substrate. An expanded analysis of the interactions formed at each concentration is described in 4.2.3.

These results suggest a link between fragment binding affinity and its lowest detectable soaking concentration (LDSC). Before further investigation of this relationship using SPR, presented in section 4.2.2, the reliability of hit detection using the XChem platform and PanDDA algorithm were assessed – thus ensuring high quality and reproducibility of the data presented here.

Quality control and reproducibility of concentration-dependent soaking experiments

In contrast to the standard use of the XChem platform, designed with high fragment concentration in mind, this study probes the limit of detection by lowering the soaking concentration. Therefore, a thorough evaluation of PanDDA configurations was carried out to ensure a reliable hit identification.

For hit detection, the PanDDA algorithm characterises a ground state by generating a ground state mean map and subtracts a proportion of it from individual datasets, allowing to deconvolute binding events (1.2.3.2).[113,263] An averaged ground state map from multiple crystals allows to capture the experimental noise and provides an accurate estimation of the background.[113] For large datasets containing a high proportion of fragment-bound datasets, it is important to ensure that the majority of datasets used for its characterisation represent the ground state and therefore do not contain bound ligands. A high proportion of fragment-bound structures used for characterisation would result in a distorted ground state mean map, leading to decreased sensitivity in fragment hit identification.

The high number of datasets containing the expected bound ligands in this concentration-dependent soaking study therefore requires a thorough definition of the ground state. Indeed, differences in hit detection between PanDDA runs have been observed when fragment-bound structures were not explicitly excluded from the characterisation of the ground state, leading to distortion of the results. For NUDT5, the algorithm did not identify the presence of six fragments in datasets corresponding to their lowest soaking concentration. The presence of these fragments could only be revealed retrospectively in a run explicitly excluding the known fragment-bound datasets from characterisation, and using datasets from DMSO soaks as well as empty datasets for the ground state definition.

Table 4.1 details the obtained number and percentage of fragment-bound datasets automatically selected by PanDDA for the characterisation of the NUDT5 ground state. For each resolution shell, between 6.7-16.6% of the selected datasets contained bound fragments. The proportion of fragment-bound datasets is especially high in the resolution shells between 1.6-1.9 Å, which correspond to the diffraction limit of NUDT5 crystals with highest distributions. The over-representation of fragment-bound datasets results in generation of a ground state mean map that does not reflect the actual ground state of the apo-protein. Failed identification of binders at low concentration impacted the results by shifting the lowest detectable soaking concentration (LDSC) towards a higher cut-off.

Table 4.1. The proportion of fragment-bound datasets in the characterisation of the ground state using default parameters, without explicit exclusion of fragment-bound datasets.

resolution shells (Å)	**> 1.6**	**1.6-1.75**	**1.75-1.8**	**1.8-1.85**	**1.85-1.95**	**1.95-2.05**	**2.05-2.3**
# of fragment-bound datasets in ground state	5	10	7	8	7	5	4
% of fragment-bound datasets in ground state	11.6	16.7	11.7	13.3	11.7	8.3	6.7

The data obtained with an accurate ground state characterisation is more consistent in hit identification, providing a lower variability in duplicates. This example highlights the importance of accurate ground state characterisation for successful and reliable hit detection by PanDDA. All data presented in this chapter is generated using an accurately characterised ground state.

The concentration-dependent study allows to further inspect the effect of soaking concentration on the quality of electron density maps ($2F_o$-F_c) and the corresponding event maps. The event maps of both high and low concentration soaks display good density coverage for the fragments and show only small differences across the concentration series (Figure 4.2). The majority of fragments hits did not show a significant drop in $2F_o$-F_c map quality across the concentration series, comparable resolution provided. However, few fragments show differences in quality of $2F_o$-F_c maps (Figure 4.2). Soaking at high

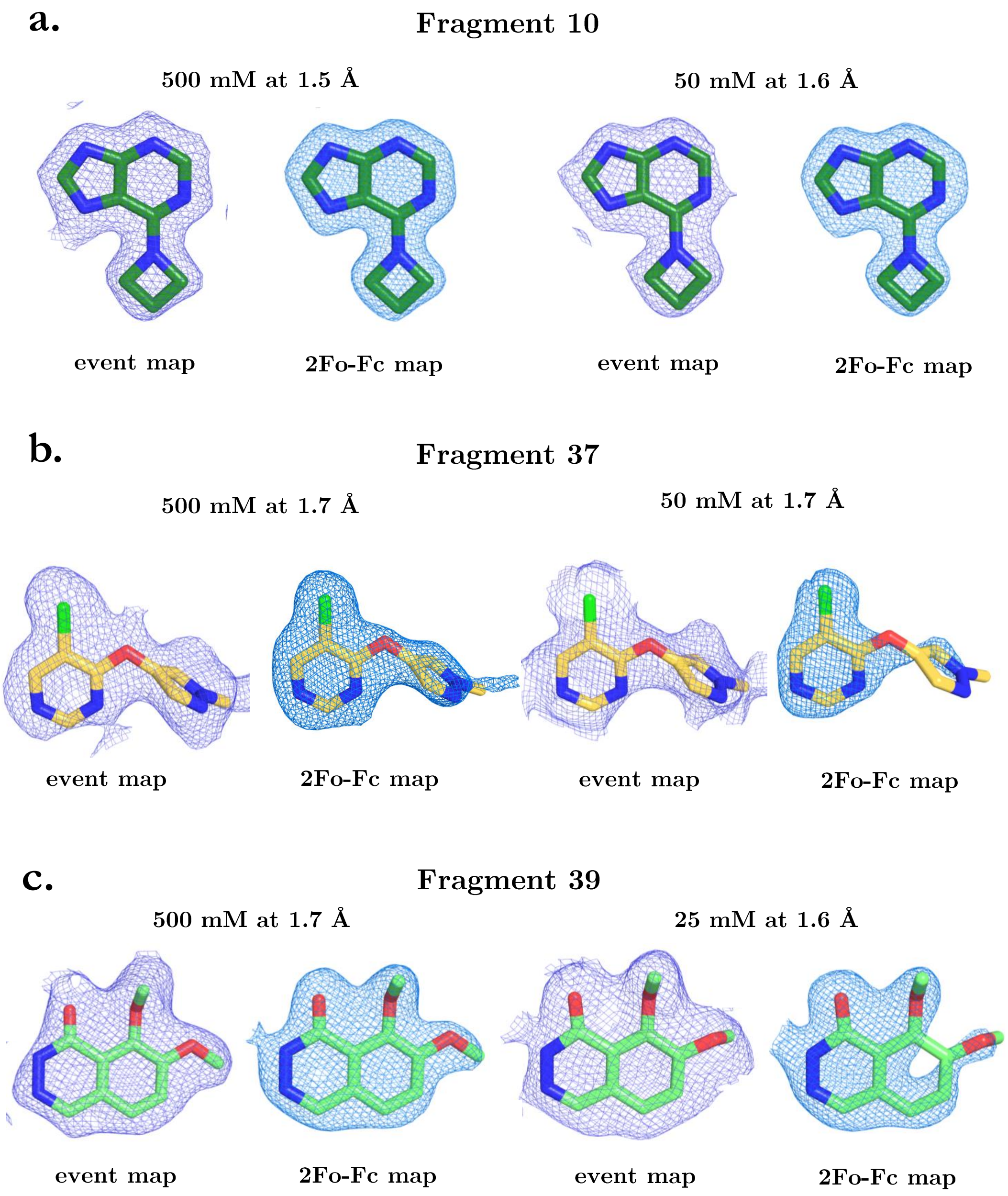
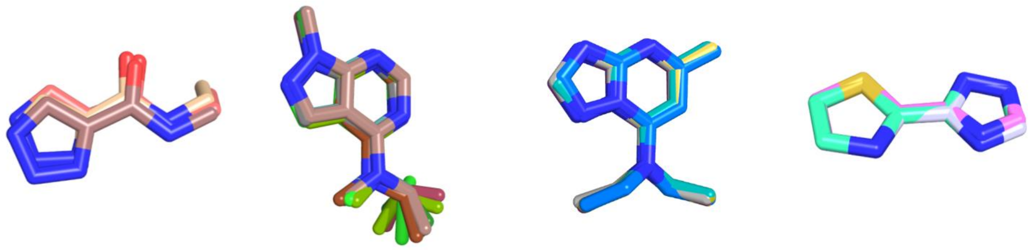


Figure 4.2. Comparison of electron density maps quality across different concentrations. While the event maps do not show significant differences across concentration, a drop of quality of the 2F_o-F_c maps at similar resolution can be observed for some fragments at low concentration. The event density map is shown are 2×(1-BDC) as purple mesh and 2F_o-F_c density at 1.00 σ as blue mesh. **a.** Fragment **10**: both the event and 2F_o-F_c maps at 500 mM and 50 mM show clear and well-defined maps. **b.** Fragment **37**: the density coverage of the 2F_o-F_c maps for the 5-membered ring connected via a rotatable bond is incomplete at a concentration of 50 mM. **c.** Fragment **39**: the density coverage of the 2F_o-F_c maps of the aromatic bicyclic system is incomplete at 25 mM.

concentrations presents a large ligand excess and ensures a higher population of binding sites, yielding clear and well-defined maps for fragments. As a result of lower occupancy, weaker density is often observed for fragments soaked at significantly lower concentrations, such as 50 mM or 25 mM. In particular, flexible groups connected via rotatable bonds are affected (Figure 4.2.b). In conclusion, while fragment hit identification benefits from high concentration, low soaking concentrations nonetheless provide sufficient evidence for detection of fragment binding.

The binding mode of the fragments was verified to ensure the comparability of datasets across the concentration series. Structural alignment of the datasets corresponding to each fragment at different concentration was carried out. As shown in Figure 4.3, the data reveals that all fragments bind in exactly the same pose at different concentrations.



Fragment 7
3 structures

Fragment 17
8 structures

Fragment 21
7 structures

Fragment 31
3 structures

Figure 4.3. Structural alignment confirms identical binding poses for fragments across different concentrations. Structural superposition of fragment poses across the concentration series is shown on selected examples.

Having ensured the quality of the obtained data, the reproducibility of ligand detection for the NUDT5 crystal system can now be assessed. For each concentration, the total number of successfully collected datasets and the proportion of fragment-bound datasets are compared. Across all fragments, the reproducibility of ligand detection varies according to the concentration: 93% at 500 mM, 88% at 250 mM, 87% at 100 mM, 76% at 50 mM and 57% at 25 mM. This can be explained by higher variability at the lowest detectable concentration for each fragment due to low occupancy. Indeed, when only considering the lowest detectable concentration for each ligand, the reproducibility lies only at 74%. In opposite, excluding the lowest concentration results in overall reproducibility of 93%. Concentration-dependent experiments conducted at the XChem platform at DLS are therefore highly reproducible and enable reliable assessment the relationship between ligand affinity and crystallographic soaking concentration.

4.2.2. Lowest detectable soaking concentration correlates with affinity determined by SPR

In order to assess the correlation between lowest detectable soaking concentration and affinity of the fragments, the ranking obtained using concentration-dependent crystallographic soaking was cross-validated by SPR as an orthogonal method.

SPR confirms strongest hits identified by crystallography

For affinity determination by SPR, His-tagged NUDT5 and His-tagged GFP, used as a reference, were immobilised on a Ni-NTA biosensor at high surface density. For assessment of assay conditions, protein stability and reproducibility of results, positive control compounds including ADP-ribose, ATP, AMP-PNP and ADP were included in the assay.[81] Figure 4.4.a-d shows the obtained binding curves on a Biacore T200 for the positive controls. The binding curves, obtained by global fitting using a one-site kinetic model, display clean sensorgrams with high signal, thus confirming the suitability of chosen assay conditions for the experiment.[175] The observed binding verifies the accessibility of the active site and functional integrity of immobilised NUDT5 on the biosensor surface.

The crystallographic fragment hits were initially measured in 2-fold concentration series from 600 μM to 18.75 μM (section 2.10). From all tested compounds, the fragments **16**, **17** and **39** displayed a concentration-dependent increased response and were selected for further validation. The remaining compounds did not show any evidence for binding. The affinity and kinetic binding parameters of the selected compounds were subsequently determined by a 2-fold concentration series from 500 μM to 3.9 μM in duplicates. With K_D values between 33 and 70 μM, the affinities of all three fragments are in low μM range (Figure 4.4.e-g).

The fragments that could successfully be quantified are equally those with the lowest detectable soaking concentration (LDSC) by X-ray crystallography. Conversely, fragments with higher LDSC could not be adequately measured by SPR (Figure 4.4.h). The binding affinity of a fragment thus correlates with its lowest detectable crystallographic soaking concentration.

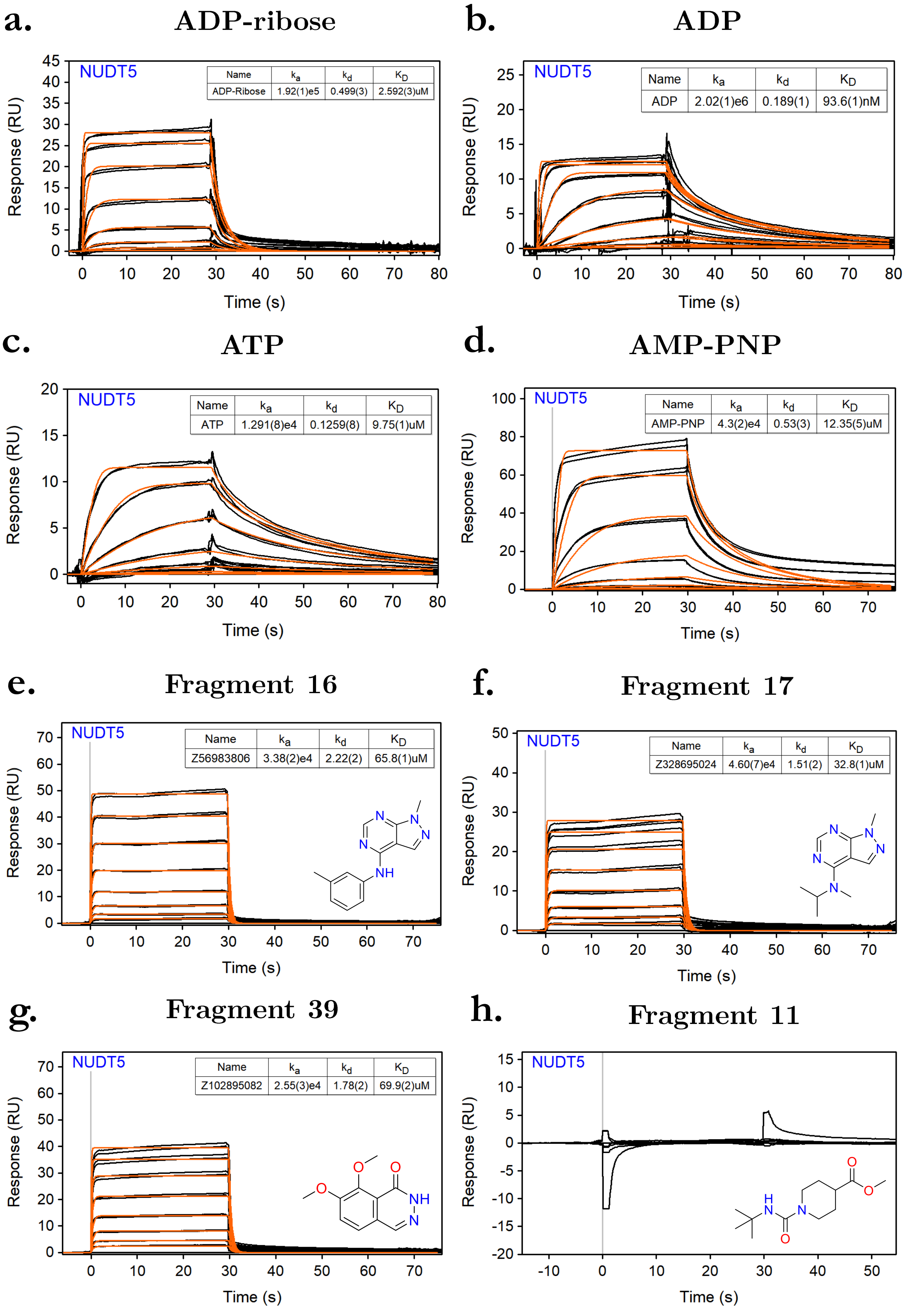


Figure 4.4. SPR analysis of positive controls and selected fragments for NUDT5. Sensorgrams of the positive controls: **a.** ADP-ribose (0.21-50 μM), **b.** ADP (20.6 nM-5 μM),

Figure 4.4. (***continued***) **c.** ATP (0.41-100 μM), **d.** AMP-PNP (0.21 nM-50 μM). Sensorgrams of binding fragments: **e.** fragment **16** (3.9-500 μM), **f.** fragment **17** (3.9-500 μM), **g.** fragment **39** (3.9-500 μM). **h.** Example of a non-binding fragment in SPR: fragment **11** (3.9-500 μM), identified by X-ray crystallography. Black lines represent measured binding curves, orange lines show 1:1 kinetic fit.

Optimisation of surface density in SPR assays for higher affinity compounds

Binding quantification of low molecular-weight compounds such as fragments typically benefits from a protein immobilisation at high density through amplification of response.[264–266] However, with high immobilisation levels, secondary effects such as mass transport and aggregation can occur, affecting the determination of kinetic parameters.[265,267–269] Mass transport is commonly observed in samples, in which association rate of the analyte is higher than the diffusion rate. This results in a local depletion of the analyte at the interaction surface as well as re-binding events, concealing the true k_a and k_d.[175,270] Whereas the effect of the mass transport for quantification of weak binders is negligible, the shape of the sensorgrams for stronger binders can be altered at high immobilisation levels, as these tend to have limited or no curvature in the association and an unstable equilibrium phase. This is observed in the previously discussed experiments and is reflected in the sensorgram shape of the positive controls ADP, ATP and AMP-PNP, clearly indicating presence of mass transport effects (Figure 4.5).

While it is possible to account to some extent for mass transport effects using advanced kinetic models, the analysis can be simplified through improvement of experimental parameters.[268,271] Mass transport limitations can be minimised by lowering the immobilisation density and/ or increasing the flow rates.[267,268] The validation experiment including the controls ADP and AMP-PNP as well as the fragments **16**, **17** and **39** was repeated at NUDT5 immobilisation densities of 8554, 4615, 2673, 1597 and 791 RU. Lower protein surface density was achieved by decreasing the protein concentration and the duration of the injection. In addition, the flow rate was increased from 30 to 60 μL/min.

Figure 4.5 demonstrates the effect of mass transport at different immobilisation levels of NUDT5 for a stronger binder, ADP, and the weakly binding fragment **39**. For ADP, the effect of mass transport is pronounced at high immobilisation level of NUDT5 (Figure 4.5.a-c) and affects the shape of the binding curves in the association and/or

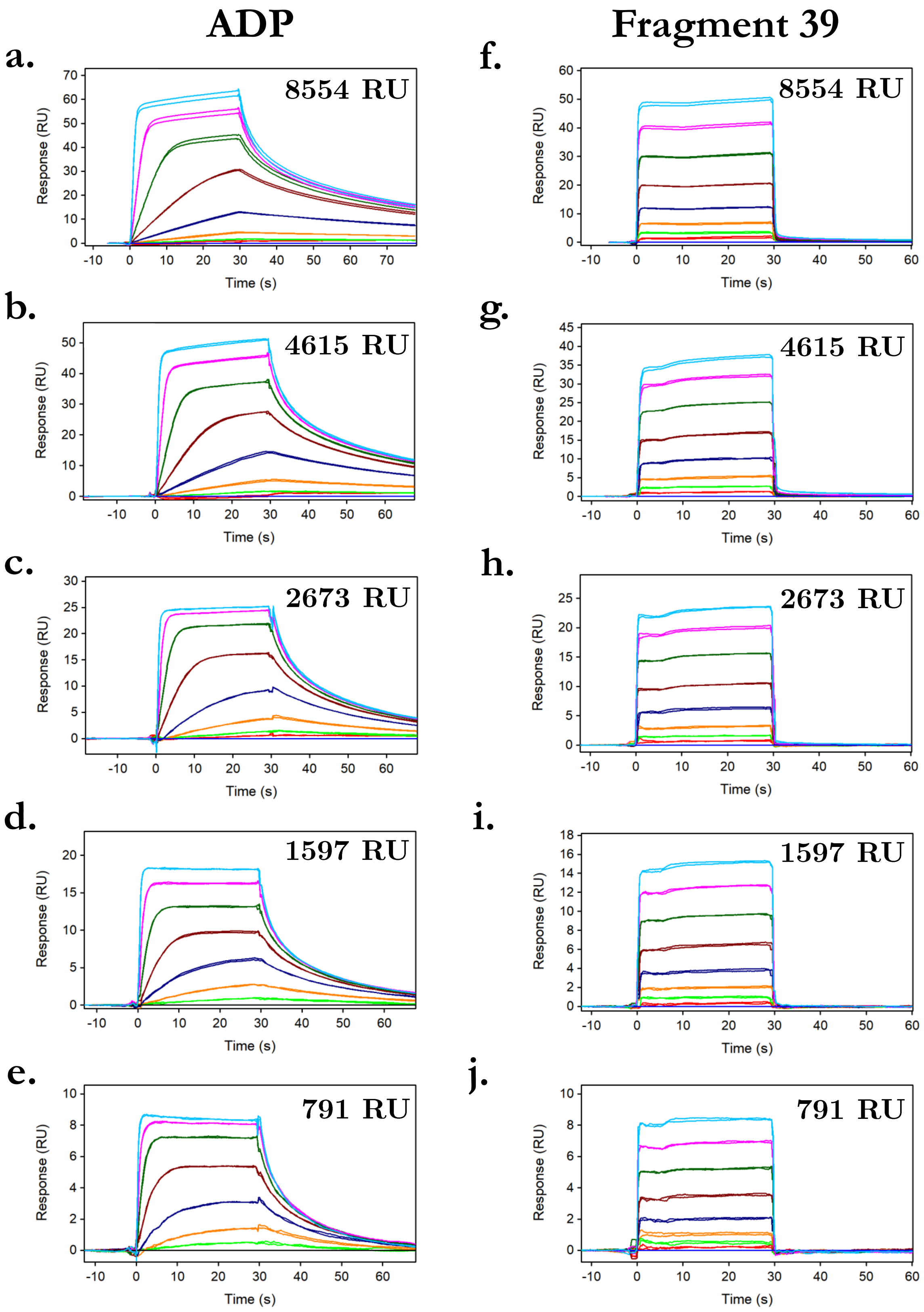


Figure 4.5. Decrease of protein immobilisation level can reduce the mass transport effect

Figure 4.5. (***continued***) observed only with strong binders. The binding sensorgrams of the NUDT5 substrate ADP show mass transport effect at high immobilisation levels of NUDT5. Reducing the surface density reveals the true kinetic parameters. **a.-e.** Binding sensorgram for ADP at different immobilisation levels of NUDT5: **a**. 8554 RU, **b**. 4615 RU, **c.** 2673 RU *d.* 1597 RU and *e.* 791 RU. Colours correspond to different fragment concentration (light blue = 5 μM, magenta = 1.67 μM, green = 555.6 nM, brown = 185.2 nM, dark blue = 61.7 nM, orange = 20.6 nM, bright green = 6.9 nM, red = 2.3 nM). **f.-j.** Binding sensorgram for the weak fragment hit **39** do not show a mass transport effect at different immobilisation levels of NUDT5: **f.** 8554 RU, **g**. 4615 RU, **h.** 2673 RU *i.* 1597 RU and *j.* 791 RU. Colours correspond to different fragment concentration (light blue = 250 μM, magenta = 125 μM, green =62.5 μM, brown = 31.25 μM, dark blue = 15.6 μM, orange = 7.8 μM, bright green = 3.9 μM, red = 1.95 μM).

equilibrium phases. Decreasing the immobilisation level of NUDT5 on the biosensor surface results in a reduction of the mass transport effect. The binding curves measured at the lowest protein densities, 791 and 1597 RU (Figure 4.5.d-e), yield significantly improved results and can be used for kinetic analysis with a simple bimolecular model.[272]

As expected, the binding sensorgrams of fragment **39** at different immobilisation densities of NUDT5 confirm that the mass transport effect is negligible for weak binders, as their affinities are too weak to observe any kinetics (Figure 4.5.f-j).[175]

With decreasing immobilisation levels of NUDT5, the intensity of the obtained response for both ADP and fragment **39** reduces proportionally to the surface density.[273] For fragment **39**, the highest concentration measured at the protein density of 8554 RU yields a signal of 49.7 RU. However, the same concentration only produces a signal of 8.5 RU when measured at the lowest protein immobilisation level (791 RU). Whereas a signal of 8.5 RU is sufficient for detection of binding conducted on a modern SPR instrument, a much lower signal-to-noise ratio might hamper identification of a weak binder, especially as these rarely show any kinetics.[175,179,273] Thus, quantification of weak interactions is best performed at high densities.

Improved SPR assay enables accurate quantification of fragment hits

The binding curves obtained at optimised immobilisation levels of NUDT5 (791 RU) were analysed using both kinetic and equilibrium fit. While ideally both methods should provide the same quantification, their application is determined by compound affinity and can be mutually exclusive.[269] Kinetic analysis of a sensorgram displaying an association and dissociation phase allows to determine k$_{on}$ (k$_a$) and k$_{off}$ (k$_d$) of strong binders, in addition to the K_D value.[176,269] In presence of fast association and dissociation rates, such as displayed by weak fragment hits, kinetic fit cannot be applied. Equilibrium analysis is a simple alternative to obtain only the K_D value for weak binders.[264,269]

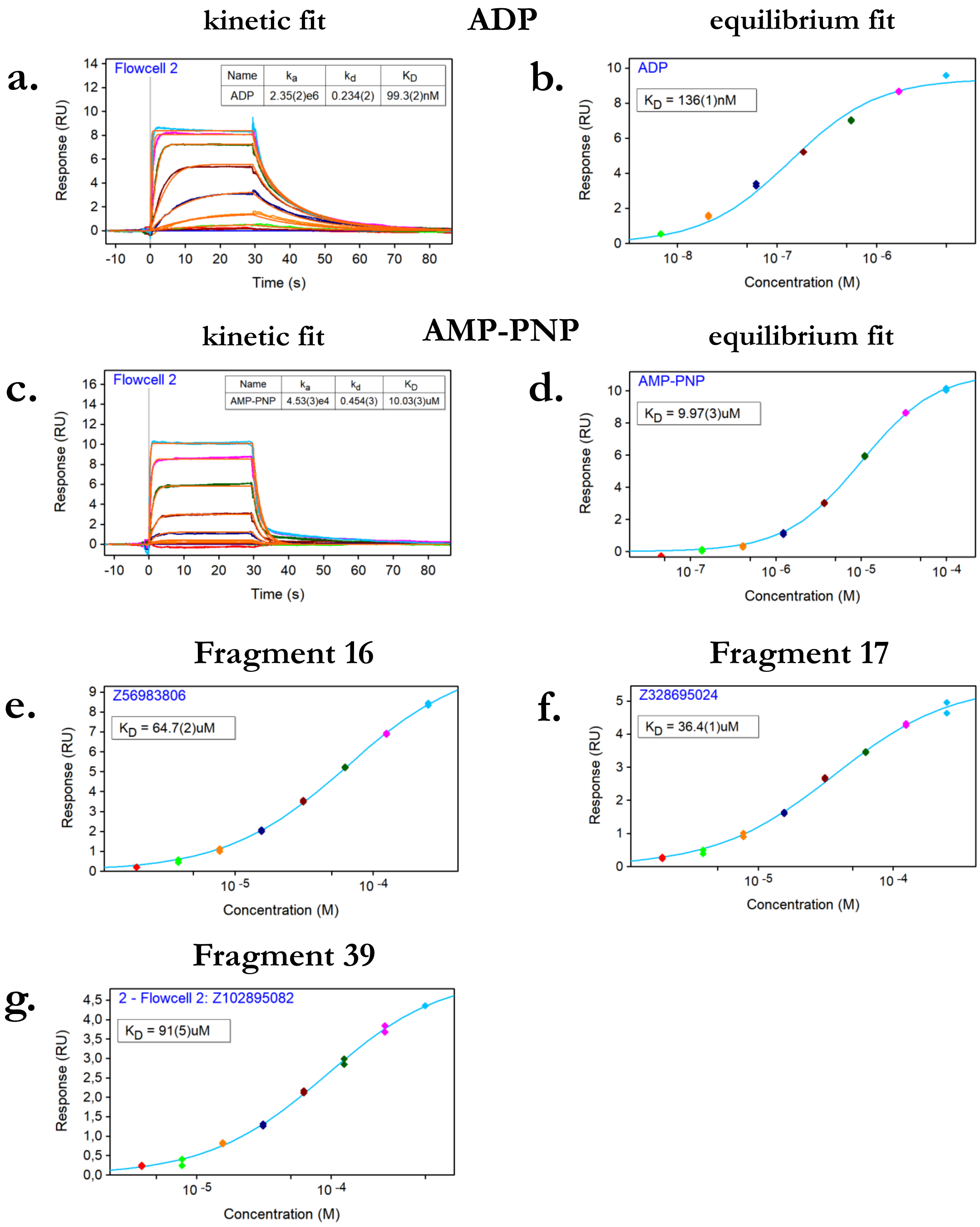


Figure 4.6. Quantification of NUDT5 binders using the kinetic and/or equilibrium fit: **a-b.** ADP (2.27 nM-5 μM), **c-d.** AMP-PNP (22.8 nM-50 μM), **e-f.** Fragment **16** (1.95-250 μM), **g-h.** Fragment **17** (1.95-250 μM), **i-j.** Fragment **39** (3.91-500 μM). Black lines represent measured binding curves, orange lines show 1:1 kinetic fit.

For the strong binder ADP, the 1:1 kinetic model provides an excellent fit to the measured data clearly exhibiting an association and a dissociation phase and yield a K_D of 99 nM. In comparison, the equilibrium fit results in a K_D of 136 nM (Figure 4.6.a-b). As the binding curves of the lowest ADP concentrations do not reach a stable equilibrium, the quantification using steady-state analysis is less accurate.

For AMP-PNP, a low μM binder, both analysis methods provide the same K_D values of 10 μM (Figure 4.6.c-d). The simulated binding responses demonstrate a consistent fit to the experimental data. As all binding curves except for the lowest concentration reach a stable equilibrium, the steady-state analysis also provides a reliable quantification.

For all three fragment hits, the results of the equilibrium fit provide the K_D values of 38 μM, 66 μM and 92 μM for fragment **17**, **18** and **39**, respectively (Figure 4.6.e-j). To our knowledge, these data represent the first reported characterisation of NUDT5 binders using biophysical methods. Previous studies reported biochemical IC_{50} values or enzymatic assays.

In summary, the SPR analysis of the fragments from the XChem campaign yielded a successful quantification of three fragment hits. As the quantified fragments are the same as the ones that can be observed at the lowest concentration in crystallographic soaking experiments, the SPR data corroborates the correlation between compound affinity and lowest detectable soaking concentration (LDSC). The successful cross-validation suggests that concentration-dependent X-ray crystallographic experiments can indeed be used for triage or ranking of weakly binding fragment hits.

4.2.3. Concentration-dependent interaction analysis using PLIP

While the scaffolds of fragment hits are often seen as starting points for their elaboration, the molecular interactions that fragments sample provide valuable information for guidance of further fragment evolution. This concentration-dependent study allows to compare the information content provided at each concentration through sets of identified fragments.

The enumeration of all protein-ligand interaction at each concentration was carried out using the Protein Ligand Interaction Profiler (PLIP), previously described in 3.2.5.[236] Its automated detection and visualisation yield a qualitative list of all non-covalent interactions per fragment. As verified by structural alignment in 4.2.1, the binding

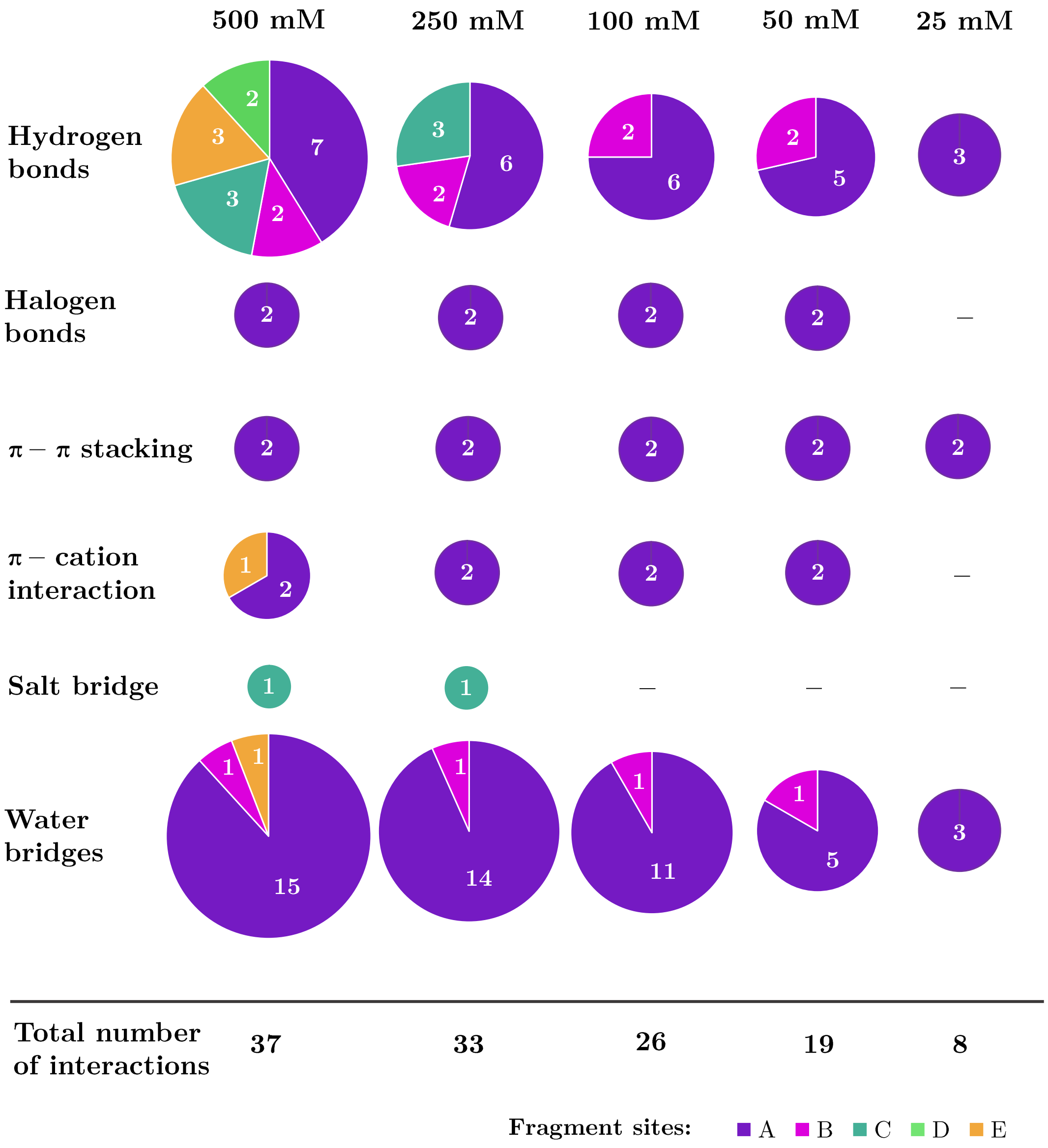


Figure 4.7. Comparison of all unique protein-fragment interactions observed at soaking concentrations from 500 mM to 25 mM. The number of sampled interactions increases with increasing soaking concentration. Each type of interaction is represented as a pie chart, with colours indicating the fragment binding site.

modes of the fragments across different concentrations remain conserved. Therefore, the fully refined PDB-deposited structural models at 500 mM are representative for the binding of each fragment across concentration series and were used for the PLIP analysis.

Figure 4.7 presents a comparison of all unique molecular interactions observed in the concentration range between 500 mM and 25 mM. The results clearly reveal that soaking at the highest concentration samples the highest number of interactions: while at 25 mM only eight interactions are observed, soaking at 500 mM yields 37 unique protein-fragment interactions. A detailed breakdown of all observed interactions and their concentration-dependency is given in Figure 4.8.

At 500 mM and 250 mM, the high number of interactions results from detection of additional fragment binding sites with unique interactions (C-E). As mentioned in 3.2.2, these are surface-exposed binders which may not necessary bear promise of lead discovery.

When restricting the analysis to the interactions in the active site (A and B), an increase in number of hydrogen bonds and water-mediated interactions is observed, while the numbers in the remaining categories remain relatively stable across the concentration series. The increase is particularly pronounced between 25 mM and 100 mM, the number of interactions found at 100, 250 and 500 mM essentially remains the same for hydrogen bonds. Although water-mediated interactions can potentially inform fragment elaboration by indicating energetically rewarding polar interactions, their asset for fragment evolution is unclear. Putting water-mediated interactions aside, this result in turn suggest that for NUDT5, screening at 100 mM would provide equivalent number of interactions as at 500 mM.

Appropriate concentration for crystallographic fragment screening has been the topic of an ongoing discussion in the crystallography community. The fragment screening facility at DLS routinely uses 500 mM for soaking, whereas a preference for 100 mM or lower concentration is often indicated in literature reports.[39,92,259] The NUDT5 data presented here suggest that 100 mM is sufficient for sampling the active site. However, secondary site including allosteric sites and weak binders might be missed if the soaking concentration is too low.

Interactions	*Binding site*	*protein residue (atom)*	*location*	*interaction type*	*Concentration (mM)*				
					500	*250*	*100*	*50*	*25*
Hydrogen bonds	A	Trp28 (N)	side chain	acceptor	■	■	■	■	■
	A	Tyr36 (O)	side chain	donor	■				
	A	Thr45 (O)	side chain	acceptor	■	■	■	■	
	A	Glu47 (N)	backbone	donor	■	■	■	■	■
	A	Glu47 (O)	backbone	acceptor	■	■	■	■	■
	A	Arg51 (N)	side chain	donor	■	■	■	■	■
	A	Asp164 (O)	backbone	acceptor	■	■			
	B	Asp133 (O)	side chain	acceptor	■	■	■	■	
	B	Ser137 (O)	side chain	donor	■	■	■	■	
	C	Lys14 (N)	backbone	donor	■	■			
	C	Gln15 (N)	side chain	donor	■	■			
	C	Tyr90 (O)	side chain	donor	■	■			
	D	Gln169 (N)	side chain	donor	■				
	D	Lys205 (N)	side chain	donor	■				
	D	Lys205 (N)	side chain	acceptor	■				
	E	Glu47 (O)	side chain	donor	■				
	E	Ser172 (N)	backbone	donor	■				
Halogen bonds	A	Thr45 (O)	backbone	acceptor	■	■	■	■	
	A	Arg51 (N)	side chain	donor	■	■	■	■	
π-π stacking	A	Trp28	side chain	-	■	■	■	■	■
	A	Trp46	side chain	-	■	■	■	■	■
π-cation stacking	A	Trp28	side chain	-	■	■	■	■	
	A	Trp46	side chain	-	■	■	■	■	
	D	His206	sidechain	-	■				
Salt bridges	C	Lys14	side chain	-	■	■			
Water bridges	A	Trp28 (N)	side chain	donor	■	■	■	■	■
	A	Trp28 (O)	backbone	acceptor	■	■	■		
	A	Trp28 (N)	side chain	acceptor	■	■	■		
	A	Tyr36 (O)	side chain	acceptor	■	■	■		
	A	Tyr36 (O)	side chain	donor	■	■			
	A	Arg44 (N)	side chain	donor	■	■	■		
	A	Thr45 (O)	side chain	acceptor	■	■	■	■	
	A	Thr45 (O)	side chain	donor	■	■			
	A	Thr45 (N)	backbone	donor	■	■			
	A	Thr46 (N)	side chain	acceptor	■				
	A	Glu47 (O)	backbone	acceptor	■	■	■	■	■
	A	Arg51 (N)	side chain	donor	■	■	■	■	
	A	Leu98 (N)	backbone	donor	■	■	■		
	A	Gly135 (O)	backbone	acceptor	■	■	■	■	■
	A	Glu166 (O)	side chain	donor	■	■	■		
	B	Asp133 (O)	side chain	donor	■	■	■	■	
	D	Lys175 (N)	side chain	donor	■				

Figure 4.8. Concentration-dependency of molecular interaction between fragment hits and NUDT5. The colours of the bars corresponds to the fragment binding site.

At the lowest concentration of 25 mM, only three fragment hits are detected, sampling eight unique interactions in site A. Fragment hits **17**, **18** and **39** are aromatic and bind between Trp28 and Trp46 via π-π stacking. In addition, all three fragments interact with both the amide nitrogen of Glu47 backbone and the guanidinium nitrogen of Arg51 side chain through hydrogen bonds (Figure 4.9). A comparison to the interaction pattern of ADP-ribose, the endogenous substrate of NUDT5, reveals that both protein residues are involved in the recognition of ADP-ribose. The probing of endogenous interactions of the NUDT5 substrate by these fragments confers their affinity. In addition to the interactions sampled by the fragments, the adenine core of ADP-ribose creates additional hydrogen bond to amide oxygen which acts as an acceptor. This could contribute in part to the higher affinity of the ADP-ribose, along with the phosphates and the ribose units.

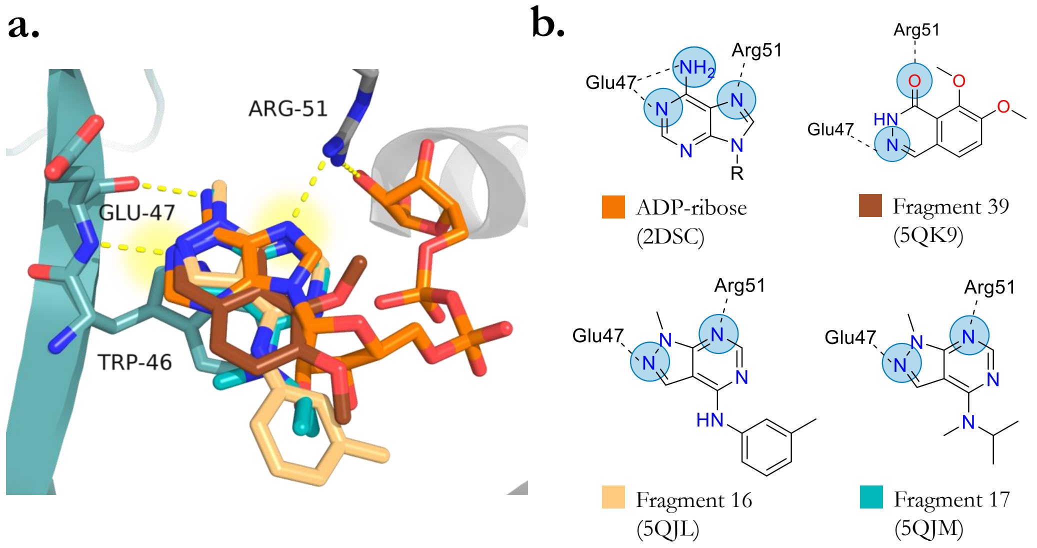


Figure 4.9. Fragments identified at the lowest soaking concentration display similar interaction pattern as the adenine-moiety of ADP-ribose. **a.** Superimposition of crystal structures with ADP-ribose (shown in orange) and fragments **39** (brown), **17** (yellow) and **18** (blue) in the substrate binding site. The key interaction with Glu47 and Arg51 are highlighted with yellow circles. **b.** Chemical structures and interaction analysis of ADP-ribose and the three fragments. In addition to the π-π stacking, all fragments share the interactions with the key residues Glu47 and Arg51. The atoms involved in hydrogen bond are highlighted with light blue circles.

Although formation of each hydrogen bond is generally associated with an affinity increase, the contribution of non-polar interactions to the binding can be significant: compounds that form multiple hydrogen bonds can still display low affinities.[274,275] As binding affinity is a result of both enthalpic and entropic contributions, the number of interactions is not predictive for binding affinity of a fragment.[274] However, the presence of interactions at the lowest soaking concentration involving the key residues and resembling the binding pattern of ADP-ribose may be indicative of their strength, highlighting the importance to retain these particular interactions in further structure-based elaboration cycles.

4.2.4. Weak binders are informative for elaboration and can be monitored by concentration-dependent soaking

The results of the concentration-dependent study with NUDT5 presented in the chapter indicate that soaking at high concentration is particularly beneficial to identify secondary fragment binding sites and weak binders. Hence, the question of appropriate concentration for crystallographic fragment screening goes hand in hand with the question of whether weak binders are informative for elaboration.

In absence of sufficiently progressed compounds based on the structures of the XChem fragment hits for NUDT5, a different example for retrospective analysis was selected: the NUDIX pyrophosphohydrolase NUDT7. This example will demonstrate that weak fragment binders identified at high soaking concentration are indeed useful starting points and can be progressed to potent compounds using the lowest detectable soaking concentration (LDSC) to monitor the affinity gain during elaboration.

In the efforts to target the NUDT7 pyrophosphohydrolase, the SGC Oxford previously solved the first crystal structure of NUDT7 and delivered the first-in-class covalent small molecule inhibitor.[276] The design of this compound was based on the results of two fragment screening campaigns, namely a non-covalent XChem fragment screening at DLS and a covalent screening using an electrophile-fragment library.[229]

As described by Resnick *et al.*[229], the MS-based screening of covalent electrophiles yielded 24 fragment hits, including **PCM-0102716** forming a covalent bond to the catalytic Cys73 in the active site (Figure 4.10.a). The XChem screening identified 18 fragment hits, including the compound **NUDT7-REV-XChem** (Figure 4.10.b). While the event maps provided sufficient evidence for the presence of this fragment, the 2F$_o$-F$_c$ density suggested a rather weak binder. In the initial elaboration efforts, a series of follow-up compounds retaining the chlorophenyl-*N*-amide core motif were designed. The binding of **NUDT7-REV-1** was confirmed in co-crystallisation experiments (Figure 4.10.b). A comparison of the structures containing the covalent **PCM-0102716** and non-covalent

NUDT7-REV-1 revealed high overlap of one of the phenyl rings, suggesting a perfect merging opportunity.[229] The merged compound **NUDT7-COV-1** did not only adopt the expected binding pose, but also displayed enzymatic activity with an IC_{50} value of 2.7 μM.[229]

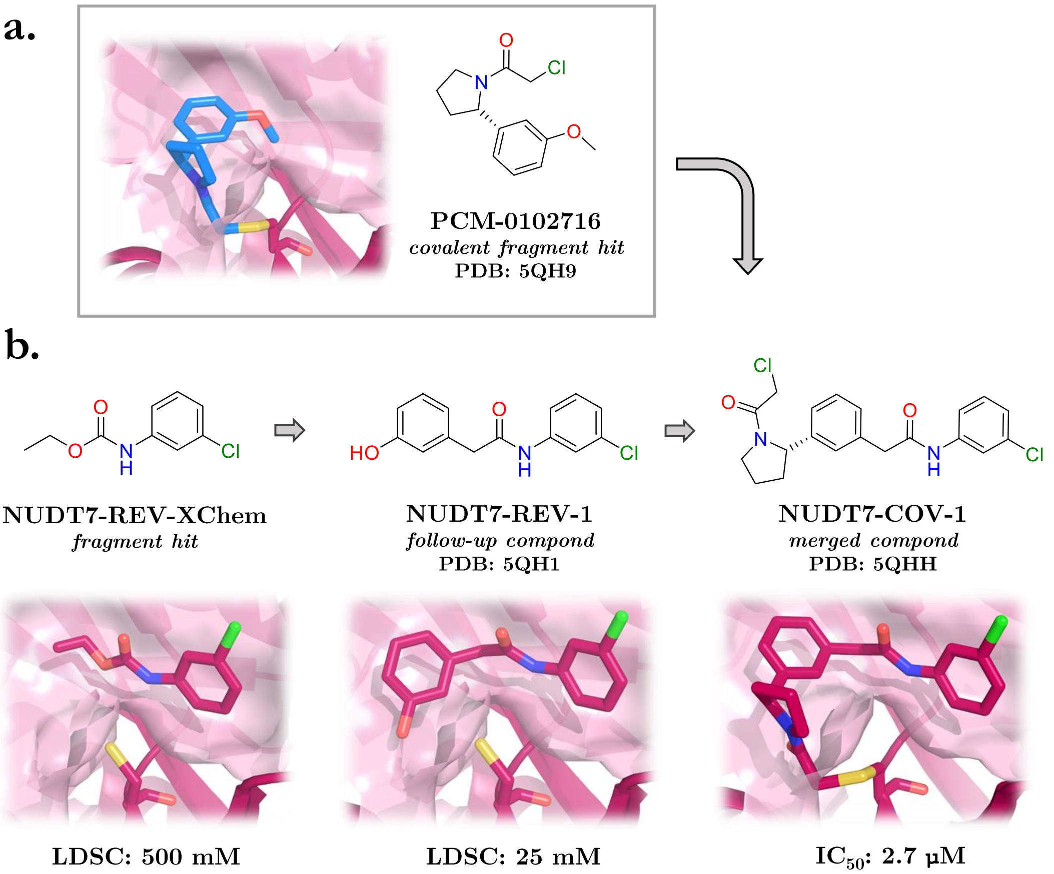


Figure 4.10. Fragment merging yields the first covalent NUDT7 inhibitor, as described by Resnick *et al.*[229]. **a.** MS-based electrophile screening revealed the covalent hit **PCM-0102716**. **b.** Elaboration of the non-covalent XChem fragment hit **NUDT7-REV-XChem** yielded the follow-up compound **NUDT7-REV-1**. Merging of both fragments **NUDT7-REV-1** and **PCM-0102716** delivered a potent inhibitor **NUDT7-COV-1** with enzymatic activity.[229] Using the concentration-dependent soaking, LDSC values were added retrospectively in the present work.

As both non-covalent compounds **NUDT7-REV-XChem** and **NUDT7-REV-1** did not show any detectable enzymatic activity and lack biophysical quantification, they provide a real-case scenario for retrospective assessment of concentration-dependent crystallographic soaking for prioritisation.

Concentration-dependent experiments were carried out using both the original XChem fragment hit **NUDT7-REV-XChem** and the follow-up compound **NUDT7-REV-1**.

Both were re-screened in a concentration series from 500 mM to 25 mM. The overnight soaking experiments were conducted under the same conditions as in the original XChem campaign, while keeping the DMSO content at 30%. For each fragment and concentration, duplicates were obtained. The detailed procedure is described in 2.8.4.

For the original XChem fragment hit **NUDT7-REV-XChem**, the concentration-dependent soaking confirms the presence of the fragment at the highest concentrations of 500 mM only, providing an LDSC of 500 mM. This finding confirms the weak nature of the XChem fragment hit and highlights the requirement of high screening concentration for its discovery. It is important to emphasise that this compound, which could be progressed into the fist NUDT7 inhibitor, would not have been detected in screening at 100 mM or lower – a concentration widely used for crystallographic screening.[39,92,259]

The elaborated compound **NUDT7-REV-1** is present throughout the concentration series to the lowest concentration of 25 mM (LDSC: 25 mM). This finding suggests a 10-fold affinity gain through addition of the phenol group. The NUDT7 crystallisation system is highly reproducible, providing the same conditions in regard to added concentration and volume of the compounds, receptor accessibility via solvent channels and kinetic effects associated with diffusion for both compounds. This allows comparison of the affinity gain based on the LDSC value for each compound in the same crystal system. Whether the LDSC value for the same compound obtained in different crystal systems and proteins is comparable, remains unknown.

This example delivers the proof of concept for use of concentration-dependent soaking to monitor the elaboration efforts in absence of biophysical quantification data. The LDSC can be used to monitor affinity gain of weak binders and offers a semi-quantitative metric to accompany medicinal chemistry efforts. In the case of NUDT7, the elaborated compound **NUDT7-REV-1** was then merged with a hit from a covalent screening, resulting in an inhibitor with enzymatic and cellular activity. This example shows that weak binders are useful for drug discovery and can be progressed.

4.3. Discussion and Outlook

The concentration-dependent study, presented in this chapter, clearly demonstrates a correlation between crystallographic soaking concentration and number of detected fragment hits. The fragments detectable at the lowest crystallographic soaking concentration (LDSC) have the highest affinity as suggested by SPR. While high soaking concentration is beneficial for detection of weak binders, probing secondary binding site and sampling more protein-ligand interactions, low concentration soaks yield the strongest binders identifying the key interactions in the protein-ligand complex.

This study complements previous literature demonstrating that crystallography can be used for qualitative and quantitative readout for fragment affinity. Whereas previous studies have shown a relative ranking of ligands in competitive[261] and comparative[260] experiments, this work relates compound affinity to the lowest detectable soaking concentration (LDSC) in a proof of principle approach. As will be described in the following, these findings highlight how concentration-dependent soaking can be instrumented not only to triage fragment hits, but also to monitor the progression of fragment elaboration. In addition, the results open up a ground for further investigation of binding quantification using X-ray crystallography.

Introducing a quantitative dimension in XChem

In a lead development campaign using crystallographic screening as primary method for fragment identification, initial prioritisation of fragment hits for further elaboration is challenging. The lack of biophysical data that could inform fragment prioritisation results in choices based on compound scaffolds, availability of follow-up compounds and user preference. The progression of follow-up compounds relies on crystallographic detection until sufficient affinity is reached to allow binding quantification using less sensitive methods such as NMR and SPR (Figure 4.11.a).

The extension of the workflow by an additional step, namely counter-soaking of previously identified hits at low concentration, would aid the prioritisation by identifying the strongest binders among the fragment hits and characterising the key interactions (Figure 4.11.b). It is particularly suitable for fragment hits, whose affinities are too weak to be measured by commonly used biophysical methods such as SPR. Whereas the re-soaking step will not be required for a target with high dispersion of fragments across the protein structure or with a low hit rate, this step is particularly useful when dealing with a highly promiscuous binding site such as the active site of NUDT5. From a practical point of view, the counter-soak does not require any additional resources and can be implemented easily e.g. as an addition following an XChem or other crystallographic fragment screening campaign. This simple extension of the current workflow would allow to extract valuable insights to guide the elaboration by providing an estimation for the affinity of observed interactions.

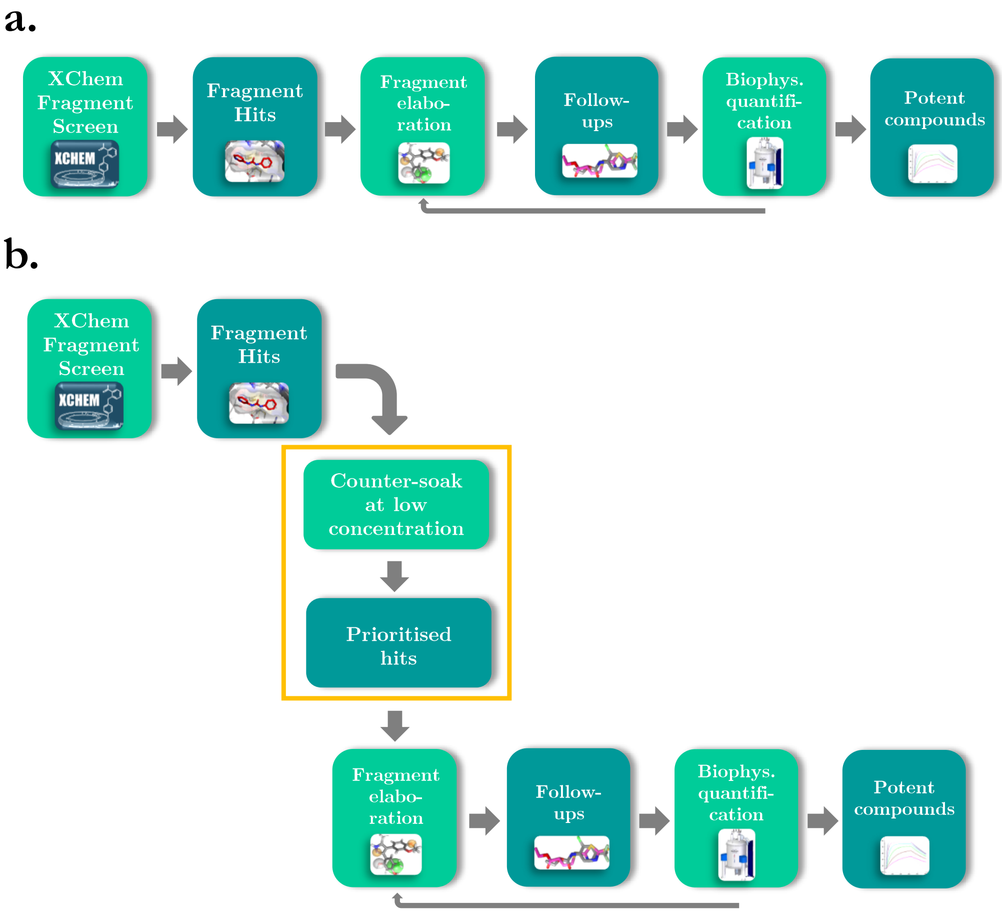


Figure 4.11. Introducing a quantitative dimension into the XChem elaboration workflow. **a.** The current XChem workflow lacks quantitative data for fragment hits, resulting in a challenging initial elaboration only guided by structural information. **b.** The newly proposed workflow includes a counter-soaking step of fragment hits at low concentration to identify the strongest binder and aid the prioritisation.

Bridging the biophysics gap

The retrospective analysis of the NUDT7 fragments demonstrated that weak binders are informative for fragment elaboration and can be progressed into potent leads, highlighting the use of concentration-dependent soaking beyond the initial triage. In absence of biophysical and biochemical quantification data, the early elaboration of weak binders lacks a reliable read-out to monitor the affinity change.

Soaking of follow-up compounds in a concentration series enables determination of their lowest detectable soaking concentration (LDSC) and thus to compare binding affinities of the fragments and their follow-up compounds. Concentration-dependent soaking therefore allows to bridge the biophysical detection gap: once weak hits are sufficiently progressed using the LDSC as a read-out, other biophysical methods for binding quantification can take over for conventional lead elaboration.

Towards a crystallographic K_D?

This work presents proof of concept demonstrating the correlation between crystallographic soaking concentration and affinity of the fragment hits. While determining the LDSC values allows for a semi-quantitative ranking of the fragment hits, this approach could potentially be expanded by taking ligand occupancy into account and enabling a more precise quantification using X-ray crystallography, yielding crystal binding affinities of ligands.

Crystallographic detection of a binder relies on sufficiently high occupancy of the binding site, which is required to be at least 50% and ideally between 70-90%.[86,277] For an equilibrium reaction between the receptor-ligand complex [RL] and its components, receptor [R] and ligand [L]:

$$[R] + [L] \rightleftharpoons [RL] \tag{4.1}$$

the fraction of occupied receptor site X_{RL} is determined by the concentration of the protein-receptor complex [RL] and the amount of the receptor [R] present[277]:

$$X_{RL} = \frac{[RL]}{[R] + [RL]} \tag{4.2}$$

For the same reaction, the dissociation constant K_D is described as:

$$K_D = \frac{[R][L]}{[RL]} \tag{4.3}$$

Assuming the presence of the ligand in excess, the equations 5.2 and 5.3 can be summarised into the following expression:

$$X_{RL} = \frac{[L]}{K_D + [L]} \tag{4.4}$$

This equation directly correlates all three parameters, the ligand fraction bound, the ligand concentration and the binding constant, allowing to estimate the soaking concentration required to achieve sufficient occupancy of the ligand in the binding site for crystallographic interpretation.[277,278]

Table 4.2. Soaking concentration and compound affinity are decisive for crystallographic detection. Example calculations with varying soaking concentrations and binding affinities demonstrate the influence on the ligand occupancy. As for the NUDT5 crystallisation system, the calculations were carried out using 10% DMSO content, reflected in the drop concentration.

Stock conc.	**fragment conc. in the drop**	**K_D of the fragment**	**theoretical fraction bound X_{RL}**
500 mM	50 mM	0.1 mM	99.8%
		1 mM	98.0%
		10 mM	83.3%
100 mM	10 mM	0.1 mM	99%
		1 mM	90.9%
		10 mM	50%
10 mM	1 mM	0.1 mM	90.9%
		1 mM	50%
		10 mM	9.1%

The example calculations shown in Table 4.2 demonstrate that while for detection of stronger binders a lower soaking concentration is sufficient, identification of weak binders requires high soaking concentration.[277,278]

Although the calculated fraction bound is good estimate for the required soaking concentration, it only provides the theoretical equivalent of the occupancy. The crystallographic occupancy, on the other hand, is a experimentally measurable parameter, reflecting the true occupancy of the binding site. In addition to ligand affinity (represented by the theoretical fraction bound), other factors such as solubility and crystal lattice play a role in determining the occupancy of the ligand.[132,278]

An important aspect of the quantification is therefore the correct determination of the ligand occupancy. Due to compensation between ligand occupancy and B-factors during refinement, the actual ligand occupancy values can only be obtained when B-factors are restrained.[277] The practical realisation of such restraints is however challenging, as methods for parametrising and performing reliable occupancy refinement remain an ongoing area of research.[132]

The translation of occupancy into a crystal binding affinity (K_c) succeeded in the case of cyclophilin 3 and its dipeptide binders.[262,279] In this particular case, the distinct conformational switch of catalytic Arg62 between native and ligand-bound form enabled to bypass ligand occupancy refinement and define the fraction bound in relation to the conformational state of this residue.[262]

In summary, whereas it is theoretically possible to determine the crystal binding affinity of ligands, practical limitations in occupancy refinement have prevented their realisation. Resolving these limitations would enable a generalisable determination of crystal binding affinities and access their untapped potential for accurate quantification of weak binders. Software development to solve this problem is limited to date by absence of publicly available datasets that provide concentration-dependent ligand binding in a reproducibly diffracting and well-characterised crystal system. The work and data presented in this chapter on NUDT5 and NUDT7 provide such a dataset and could therefore potentially contribute to the improvement of occupancy determination, ultimately enabling future quantification of crystal binding affinities.

5

Fragment hit ranking by ^{1}H NMR using relaxation-based methods

Contents

5.1. Introduction

Crystallographic fragment screening does not provide any quantitative information upon ligand binding, requiring an orthogonal method for determination of binding affinity. A robust technique allowing to rapidly rank fragment hits according to their binding strengths would allow for streamlined prioritisation of fragments for further elaboration.[159]

As crystallographic fragment hits typically have low affinities, a sensitive yet reliable technique is pre-requisite for affinity determination. NMR spectroscopy has been reported as particularly suitable technique for quantification of weak binders due to its high detection sensitivity.[173,280] In particular, ligand-observed NMR techniques are practical to implement as they do not require any isotope labelling, place lower demands on sample quantity and are not limited by protein size .[280–282]

Multiple ligand-observed NMR techniques have been employed for quantification of weak binders, including saturation transfer difference (STD) and relaxation-based methods ($T_{1\rho}$ and T_2).[155,159] Generally, ligand affinity can be determined either indirectly using competition experiments or in the direct mode. Competition experiments utilise a reporter molecule with known binding affinity and deliver the most reliable results.[283,284] However, for novel targets, such a reporter molecule is often not available.[173,285] Direct ligand-observed NMR experiments can also yield accurate determination of dissociation constants.[155,159] However, these typically involve laborious ligand titration to obtain a complete binding curve and result in long acquisition and low throughput.[286]

Relative ranking of fragment binders can provide a faster alternative to the precise quantification. As reported by Shortridge and co-workers[280], protein titration into a fixed ligand concentration in ^{1}H NMR line broadening experiments can be used to rapidly rank ligands according to their binding affinities. For fluorinated ligands, an approach exploiting chemical shift anisotropy for ranking has been introduced recently.[286]

In this chapter, efforts to implement ligand-observed NMR measurements for relative ranking are described, using protein titration to allow prioritisation of fragment hits according to their binding strengths. These approaches bypass the need of protein labelling or a reporter molecule.[286] For that, two different ligand-observed NMR methods, STD and $T_{1\rho}$, were assessed using the crystallographic fragment hits obtained for NUDT5 (chapter 3).

For STD NMR, the results demonstrate high variation in STD response across protons in the same molecule, making its utility for relative ranking of weak binders questionable. Indeed, the recently published work of Aretz *et al.*[154] denies the correlation between STD response and binding affinity.

Implementation of $T_{1\rho}$ relaxation for prioritisation using protein titration yielded a preliminary ranking of NUDT5 XChem fragment hits. However, the analysis of this ranking could not deconvolute ligand affinity from other contribution to the displayed $T_{1\rho}$ effect. Further investigations are required to enable NMR-based fragment ranking through protein titrations.

5.2. Results

5.2.1. Fragment hit ranking using ligand-based NMR techniques

For rapid and high throughput ranking of crystallographic fragment hits, two ligand-observed techniques, STD NMR and $T_{1\rho}$ were assessed. Quantification attempts were carried out with a subset of 25 fragments from the original XChem campaign. Fragment hits were selected based on five criteria: the site of interest, chemical redundancy and commercial availability of follow-up compounds for further elaboration, solubility in buffer and presence of impurities. The selected subset is summarised in appendix A.2.1.

Briefly, for all ranking experiments ligand concentration was kept constant at 2 mM, while protein concentration was varied. Samples with molar ligand-to-protein ratios of 0.5 (10 μM), 1.5 (30 μM) and 3% (60 μM) were prepared. These ratios were identified as suitable, as ligand signal is still detectable in presence of the highest protein concentration. The detailed sample preparation is described in 2.9.3. The data was acquired on Ascend 600 MHz instrument with a 5 mm helium cryoprobe at Bruker site in Fällanden, Switzerland. For each fragment, STD and $T_{1\rho}$ experiments were acquired using standard pulse sequences. For each fragment and concentration, all assigned signals were integrated in TopSpin and used for calculation of intensity decrease. The results for each technique are discussed separately in the following sections.

5.2.2. STD NMR is not suitable for fragment ranking

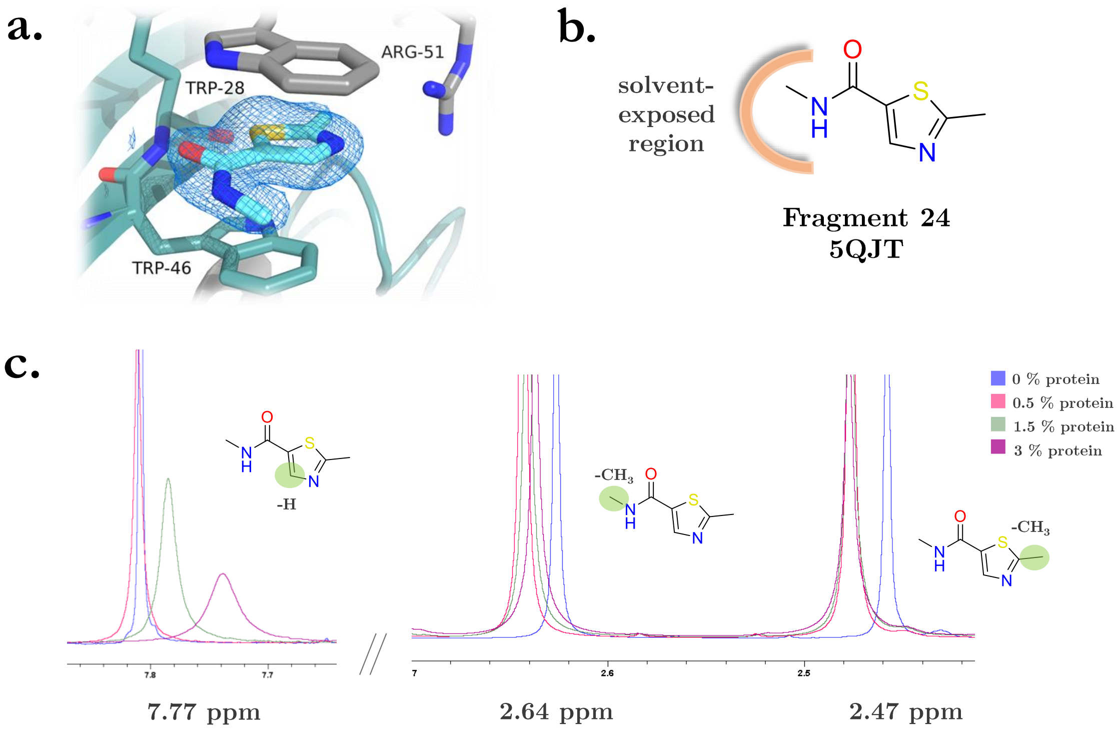
In STD experiments, ligand detection is based on saturation transfer from protein to binding ligands.[287] A correlation of affinity and saturation transfer seems plausible, as a ligand with higher affinity displays a longer residential time in the binding site, which results in higher magnetisation transfer.[154] This assumption holds true for binders in the fast exchange regime, such as weakly binding fragment hits.

This section investigates whether the STD effect can be harnessed in protein titration experiments for ranking of low affinity fragments. Example of the obtained data for fragment **24** and **36** are given in Figure 5.1 and Figure 5.2.

Both fragment **24** and **36** are anchored in the active site of NUDT5 by Trp28 and Trp46. For fragment **24**, the aromatic thiazole ring is pointing into the protein cavity, while the methyl amide group is exposed to the solvent (Figure 5.1.a-b). As shown in Figure 5.1.c, with increasing protein concentration, line broadening and change of the chemical shift are observed for all three characteristic ^{1}H signals of the fragment hit. However, the signals display high variability in the saturation transfer (Figure 5.1.d-e). While the aromatic proton at 7.77 ppm exhibits the most pronounced STD effect with increasing protein concentration, the effect on solvent-exposed methyl amide moiety at 2.64 ppm is the lowest (Figure 5.1.e).

These findings are observed with the majority of tested fragment hits. Fragment **36** provides another example, illustrating the same trends: the methyl group is highly solvent-exposed and points into the solvent channel (Figure 5.2.a-b), while the aromatic protons are anchored between both tryptophan residues. The binding location is again reflected in the strengths of the STD effect, with ^{1}H signal corresponding to solvent-exposed methyl group showing a reduced effect (Figure 5.2.c-e)

The variation in STD signal strength among different protons inside the same molecule may result from protein-ligand proximity: saturation transfer from protein to ligand depends on the distance, with ligand hydrogen in close proximity to protein receiving more saturation.[154,156,288]



d.

STD effect [%] with:

Signals [ppm]	0 % protein	0.5 % protein	1.5 % protein	3 % protein
7.77	0	27	44	58
2.64	0	4	11	23
2.47	0	11	25	43

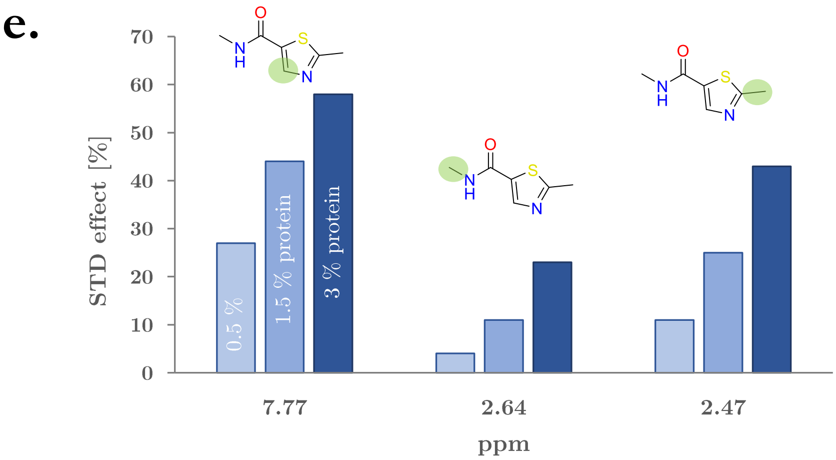


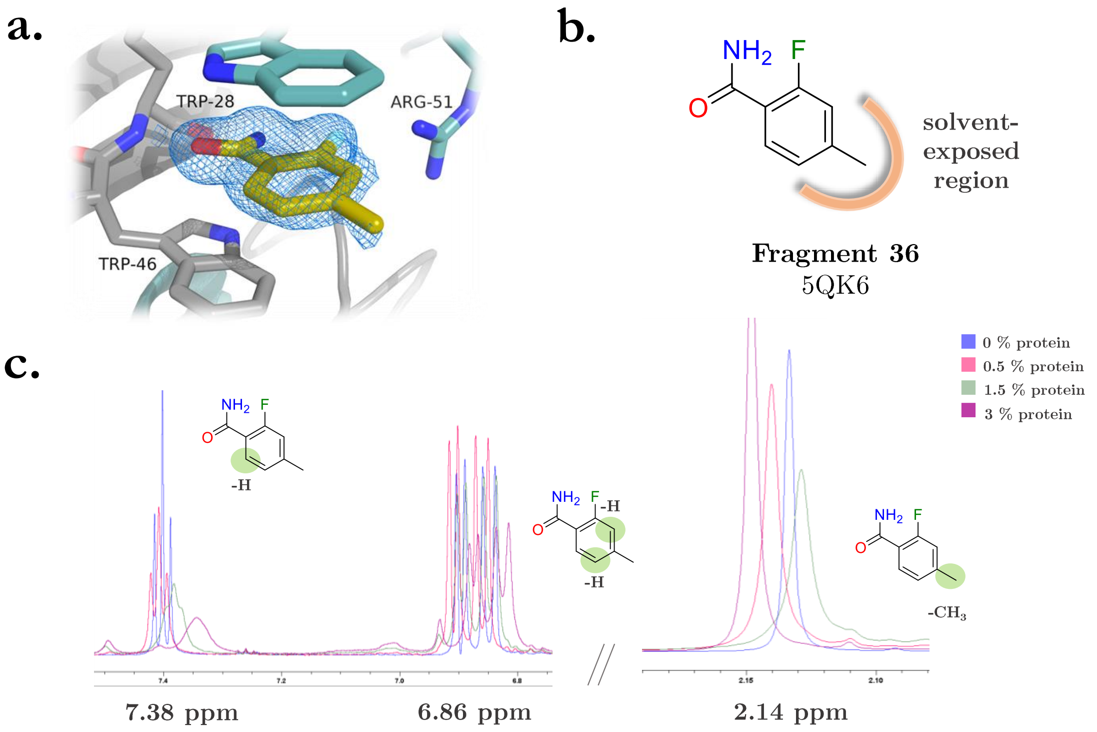
Figure 5.1. Analysis of the STD effect on ^{1}H signals of fragment **24**. **a.** Binding mode of the fragment **24** inside the active site of NUDT5. **b.** Chemical structure of fragment **24**. The methyl amide group is pointing towards a solvent channel (highlighted in orange). **c.** ^{1}H signal assignment of fragment **24**. The spectrum displays three signals, which corres-

Figure 5.1. (***continued***) pond to the aromatic proton (7.77 ppm), and amide methyl (2.64) and aromatic (2.47 ppm) methyl group. Changes in intensity, line width and chemical shift are observed upon with increased protein concentration. **d-e.** The STD effect displays high variability between ^{1}H signal of fragment **24**. Calculated STD effect (**d.**) and corresponding histogram (**e.**) at given protein concentration for each ^{1}H signal of fragment **24**.

The relaxation properties of the ligand can also contribute to the observed variability of the STD effect.[289] The T_1 relaxation time of each proton is determined by its local chemical environment and can demonstrate strong variations between protons of the same compound.[281] Consequently, the observed STD response is a product of multiple effects, including ligand orientation and relaxation, impeding the deconvolution of binding affinity contribution.

This lack of uniformity in response to saturation transfer requires a selection of a signal for comparison to other ligands without any concrete criteria. Any ranking building on this selection would be entirely arbitrary. Although all 25 fragment structures are aromatic, they do not share any recurring chemical pattern, and thus no common ^{1}H signal that could be exploited for qualitative ranking by STD.

In the course of this study, the use of STD NMR for rapid ranking of binders has been denied by Aretz *et al.*[154]. Comparison of theoretical STD effects calculated using COmplete Relaxation and Conformation Exchange MAtrix (CORCEMA) theory with experimental measurements demonstrated a lack of correlation to binding affinity concluding that STD effect is not predictive of the binding strength of a compound.[154] The authors provide evidence that even ranking by single concentration measurements of structurally similar compounds is not feasible.[154,286] Our findings are in line with the report of Aretz *et al.*[154], demonstrating a high variation across ^{1}H signals presumably caused through ligand orientation and T_1 relaxation, preventing the use of STD effect for compound ranking.



d.

STD effect [%] with:

Signals [ppm]	**0 % protein**	**0.5 % protein**	**1.5 % protein**	**3 % protein**
7.38	0	12	26	38
6.86	0	11	25	37
2.14	0	6	15	23

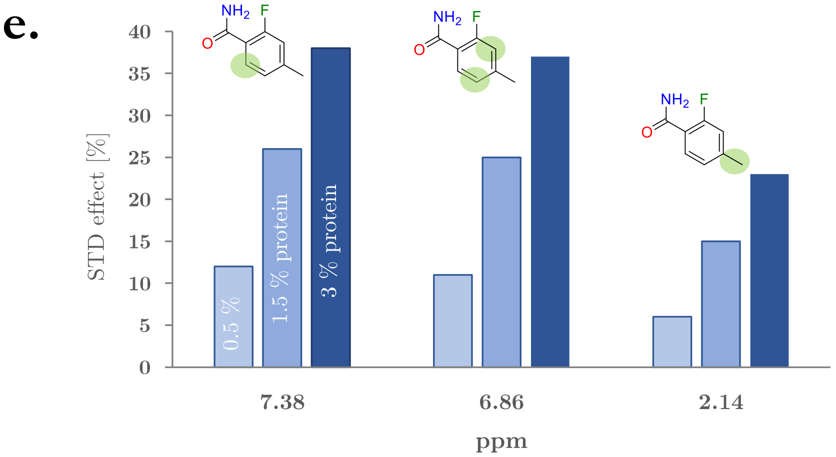


Figure 5.2. Analysis of the STD effect on ^{1}H signals of fragment **36**. **a.** Binding mode of the fragment **36** inside the active site of NUDT5. **b.** Chemical structure of fragment **36**. The methyl group is pointing into a solvent channel (highlighted in orange). **c.** ^{1}H signal

Figure 5.2. (***continued***) assignment of fragment **36**. The spectrum displays three signals, which correspond to the aromatic proton in *ortho*-position (7.38 ppm), aromatic protons in *meta*-positions (6.86) and methyl group (2.14 ppm). Changes in intensity, line width and chemical shift are observed with increased protein concentration. **d-e.** The STD effect displays high variability between ^{1}H signal of fragment **36**. Calculated STD effect (**d.**) and corresponding histogram (**e.**) at given protein concentration for each ^{1}H signal of fragment **36**.

5.2.3. Towards a $T_{1\rho}$-based fragment ranking

Shortridge *et al.*[280] described a method for ranking of compounds according to their binding affinities using 1D ^{1}H NMR line broadening experiments. In contrast to traditional K_D determination, their approach is based on protein titration to a constant ligand concentration.[280] By correlating signal height of free and bound ligand to the fraction of bound ligand, the authors present a good agreement in quantification of twelve small molecule binders to human serum albumin (HSA) with literature values for compounds >300 Da in the low μM affinity range.[280]

^{1}H line broadening experiments are not suitable for weak affinity fragments as the observed broadening can be difficult to detect due to low population of the bound fraction. Therefore, the use of relaxation-based experiments for ranking of the NUDT5 fragment hits was explored using protein titration. In $T_{1\rho}$ experiments, ligands bound to a protein adopt its properties and relax faster than free ligands in solution. Consequently, the signal attenuation is indicative of ligand binding.[159,287,290] As affinity is one of the main contributors to the extent of signal decrease, the use of $T_{1\rho}$ for qualitative ranking and triage of fragment hits was explored.

The NMR peak intensity ratio represents an easily measurable response of ligand binding. For $T_{1\rho}$ experiments, two spectra – with short and long delay – are recorded. The attenuation for each ^{1}H signal is calculated as a ratio R of integrals for each long (I_L) and short (I_S) delay:

$$R = 1 - \frac{I_L}{I_S} \tag{5.1}$$

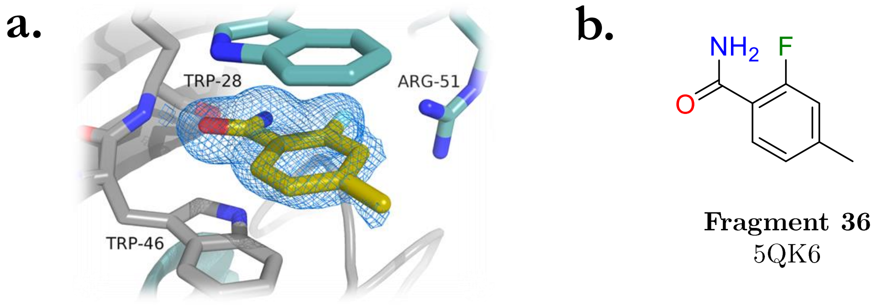
This ratio is calculated separately in presence (bound ligand, R_B) and absence of protein (free ligand, R_F). The $T_{1\rho}$ effect for a given ^{1}H signal (B) is therefore a ratio of both integral ratios in presence and absence of protein:

$$B = \frac{R_F}{R_B} \tag{5.2}$$

The T$_{1\rho}$ effect for the entire molecule is defined here as a sum of the effects on each ^{1}H signal weighted by the number of protons (n_H) and normalised by the total number of protons ($n_{H,tot}$):

$$B_{tot} = \frac{\sum\limits_H n_H \cdot B}{n_{H,tot}} = \frac{\sum\limits_H n_H \cdot \frac{R_B}{R_F}}{n_{H,tot}} \tag{5.3}$$

These calculations were carried out for the three molar ligand-to-protein ratios of 0.5, 1.5 and 3% for each fragment hit. An example of obtained data is shown in Figure 5.3.



Fragment 36
5QK6

c.

Signals [ppm]	**Signal assignment**	**R$_F$ [%]** **0 % protein**	**R$_B$ [%]** **0.5 % protein**	**R$_B$ [%]** **1.5 %protein**	**R$_B$ [%]** **3 % protein**
7.38	H$_{ortho}$	7	66	96	100
6.86	H$_{meta}$	6	37	70	89
2.14	CH$_3$	14	45	74	88
T$_{1rho}$ effect (B$_{tot}$)			**5.2**	**8.8**	**10.5**

Figure 5.3. Quantification of the T$_{1\rho}$ effect on ^{1}H signals of fragment **36**. **a.** Binding mode of the fragment **36** inside the active site of NUDT5. **b.** Chemical structure of fragment **36**. **c.** Calculated T$_{1\rho}$ effect on ^{1}H signals of fragment **36**. Integral ratios of three ^{1}H signals in presence of different protein concentrations were calculated as shown in equation 5.1. The T$_{1\rho}$ effect for the entire molecule, obtained using equation 5.3, increases with ascending protein concentration.

The qualitative ranking shown in Figure 5.4 was obtained using T$_{1\rho}$ effect in presence of the highest protein concentration of the series (3%) as it is most pronounced. The ranking of ligands was overall the same across all three protein concentrations. Ten fragment hits that displayed the strongest $T_{1\rho}$ effect are presented in Figure 5.4, the complete ranking is shown in appendix A.2.2.

The detected $T_{1\rho}$ effect used for the ranking ranges from 10.5 to 2.2, indicating fragment **36** as presumably the strongest and fragment **11** as the weakest binder in the tested subset. The $T_{1\rho}$ values for some positions are very close, particularly for the position 3, 4 and 5, implying that even smallest experimental errors could have affected the ranking.

The compounds that occupy the first five positions are rather small, they have a molecular weight below 190 Da and low number of protons. However, considering all obtained results, no correlation in regard to molecular weight or number of protons can be seen.

The comparison to the crystallographic ranking obtained in chapter 4 using the lowest detectable soaking concentration (LDSC) reveals no correlation to the $T_{1\rho}$ NMR results (appendix Figure A.10). From eight fragments displaying low LDSC values (50 mM) in the crystallographic study, only three (fragment **26**, **35** and **41**) were ranked among the top ten compounds in the NMR-based ranking. When considering the fragments ranked last, the $T_{1\rho}$ effect obtained for fragment **20** and **21** is unexpected as both compounds have LDSC values of 50 mM. Equally, fragments displaying a similar $T_{1\rho}$ effect, such as fragment **26** and **7**, were found to exhibit very diverging LDSC values (50 and 250 mM, respectively).

Except for fragment **41** and **42**, all ligands assessed bind in exactly the same site of NUDT5, implying the $T_{1\rho}$ effect observed here is inherently determined by the chemical properties of the ligand itself. The $T_{1\rho}$ effect could not be related to ligand affinity, suggesting other effects impacting signal attenuation and questioning the utilisation of peak height ratios for fragment ranking.

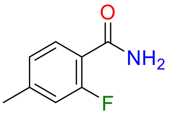
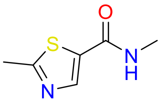
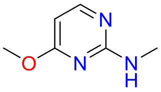
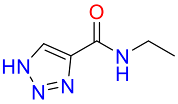
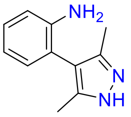
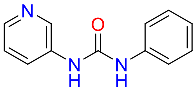
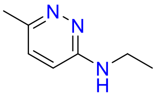
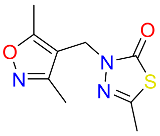
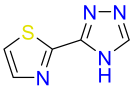
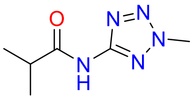
Rank	**Fragment** **PDB ID**	**Chemical structure** **MW**	**T1rho effect**	**LDSC [mM]**
1	**Fragment 36** 5QK6	 MW: 153.16 Da	10.5	250
2	**Fragment 24** 5QJT	 MW: 156.20 Da	9.1	100
3	**Fragment 26** 5QJV	 MW: 139.16 Da	7.3	50
4	**Fragment 7** 5QJC	 MW: 140.15 Da	7.2	250
5	**Fragment 30** 5QK0	 MW: 187.25 Da	7.1	100
6	**Fragment 41** 5QJ5	 MW: 213.24 Da	6.9	50
7	**Fragment 35** 5QK5	 MW: 137.19 Da	6.8	50
8	**Fragment 13** 5QJI	 MW: 225.27 Da	6.4	100
9	**Fragment 31** 5QK1	 MW: 152.18 Da	6.0	100
10	**Fragment 19** 5QJO	 MW: 169.19 Da	5.7	100

Figure 5.4. Fragment hit ranking using $T_{1\rho}$ effect – top ten positions.

5.3. Discussion and Outlook

As NMR-based techniques provide high sensitivity in detection of protein-ligand binding, their utilisation for ranking of weak binders, such as crystallographic fragment hits, was assessed. In this chapter, attempts to implement two methods, namely STD and $T_{1\rho}$, for fragment ranking using protein titration were undertaken. While STD response displayed high variation across protons in the same molecule and could not provide a ranking, a preliminary quantification using $T_{1\rho}$ effect could be obtained. However, a correlation to the affinity observed by concentration-dependent crystallographic soaking could not be detected.

STD response depends on ligand orientation The STD experiments do not provide additional information for fragment hits from crystallographic screening. However, in absence of structural information, binding epitope mapping, which allows to distinguish between protons buried in the protein and solvent-exposed protons, may provide valuable structural insights for further fragment elaboration.[156,282,291] As an example, Kemper *et al.*[289] developed a method for determination of more precise group epitope maps (GEM-CRL) by taking ligand relaxation into account.

Overcoming limitations of $T_{1\rho}$-based ranking Recently, it has been demonstrated that reduction in intensity of NMR signal in ligand-observed experiments is not proportional to its affinity, meaning that weakly binding ligands can exhibit larger signal decrease than stronger binders.[141,286] This includes both ^{1}H-observed $T_{1\rho}$ as well as ^{19}F-observed T_2 experiments[286] and is corroborated by the data obtained in this thesis.

Relating relaxation-based rankings with fragment affinity is therefore a complex endeavour. Ligand-observed NMR experiments often exploit difference in transverse relaxation rate between free and bound state of the ligand. According to Rudisser *et al.*[286], in the fast exchange regime (weak binders) the transverse relaxation (R_2^{eff}) comprises three contributions, namely the population-weighted average of free ($R_{2,f}$) and bound ($R_{2,b}$) states as well as the chemical exchange term (R_{ex}):

$$R_2^{eff} = p_f R_{2,f} + p_b R_{2,b} + R_{ex} \tag{5.4}$$

with p_f and p_b corresponding to fractions of free and bound ligand, respectively. The fraction bound p_b is a function of ligand affinity and is required for ranking. As ligand-observed NMR techniques exclusively provide information upon the free state, only one of the contributions given in the expression 5.4 is experimentally accessible, namely $R_{2,f}$.[286,292]

Therefore, the ability to use NMR-based relaxation methods for ranking amounts to calculating or suppressing additional contributions to isolate the $p_bR_{2,b}$ term (equation 5.4). The exchange term R_{ex}, depends on two parameters: resonance frequency of the free and bound state, and the exchange rates between these states.[286] The values of both parameters can display a high variation and cannot be measured.[286] However, their contribution can be suppressed using high power spin lock pulses, as demonstrated by Rudisser *et al.* on ^{19}F.[286]

The contribution corresponding to the bound state is influenced by two factors: the dipolar relaxation (R_{DD}) and relaxation from chemical shift anisotropy (R_{CSA}):

$$R_{2,b} = R_{DD} + R_{CSA} \tag{5.5}$$

While relaxation from chemical shift anisotropy can be estimated with high accuracy, dipolar relaxation is not experimentally accessible.[286] In ^{19}F NMR, this can be bypassed by measuring dipolar relaxation effects at two different magnetic fields and using the difference in chemical shift anisotropy to calculate the fraction bound p_b which is proportional to ligand affinity.[286]

As the contribution of chemical shift anisotropy can vary significantly between functional groups, a normalisation is required. Rudisser *et al.*[286] performed quantum chemical calculations to obtain values for the chemical shielding tensors.

While the chosen approach by Rudisser *et al.*[286] is elegant solution of fluorinated ligands, its translation onto ^1H NMR is more challenging. The quantum chemical calculations of tensors for the broad functionalities of ^1H atoms are more delicate. In addition, the strengths of chemical shift anisotropy is less pronounced for ^1H compared to ^{19}F nuclei.[141] This can be explained by the high electronegativity of fluorine, leading to a very anisotropic electronic environment.[141] Instead, the transverse relaxation is dominated by dipole-dipole relaxation.[141] Further investigations are thus required to realise the full potential of $T_{1\rho}$-based relaxation method for fragment affinity ranking using ^1H NMR.

6

Targeting NUDT5: Elaboration of fragment hits in the novel binding mode

Contents

6.1. Introduction

The recently discovered role of NUDT5 in nuclear ATP synthesis presents a potential therapeutic approach for breast cancer treatment as NUDT5 maintains the energy demands of chromatin remodelling and gene regulation.[213] The development of first potent and cell-active NUDT5 inhibitors has been reported by Page *et al.*[200] These share high similarity to the endogenous substrate ADP-ribose and target the same active site.

The SGC Oxford and its collaborators are primarily interested in modulation of the target using novel inhibition modes and chemical matter. In light of the confirmed druggability of the target and with already existing inhibitors, a fragment-based approach was selected to identify new starting points for ligand design. The XChem fragment screening campaign presented in this work (chapter 3), revealed a novel fragment site with a solitary binder, fragment **41** (Figure 3.3). Although the fragment binding is presumably rather weak, the discovery of this previously unknown site bears the promise of a novel inhibition mode for NUDT5.

This chapter presents the elaboration of NUDT5 fragment hits towards more potent compounds in a novel binding mode. In absence of a biochemical assay for NUDT5, the initial fragment elaboration was purely guided by the structural information. Different strategies were outlined aiming to explore the binding site and generate structure activity relationships (SAR) for the original fragment hit.[12,293] In fragment elaboration, SAR by catalogue is often utilised as a generally accessible and rapid method to explore the chemical space around the fragment and test potential elaboration hypothesis.[56,192,293–295] This approach involves a substructure- or similarity-based search of analogues among compounds from in-house collections or commercially available sources, allowing for fast elaboration turn-over that does not require any synthetic chemistry efforts.[56,293,296] To facilitate compound prioritisation, SAR by catalogue is often combined with molecular docking.[295,297,298] By predicting the binding pose of the compound, docking can be used as a filter for selection of most promising compounds for subsequent experimental testing.[295]

In this work, a hit exploration cycle comprising three steps was utilised: initial compound selection in a SAR by catalogue approach, followed by rigid docking and subsequent experimental testing using X-ray crystallography. In addition, to understand the dynamic nature of the ligand binding, hit exploration was supported through comparative analysis of the crystal structures and molecular dynamics simulations.[293]

The initial exploration established the structure-activity relationships for fragment **41** binding in the site B and identified the key features informing further ligand design. However, the introduction of larger substituents to the original fragment was largely unsuccessful, yielding only one structure with a follow-up compound in an unexpected binding mode. An alternative approach aiming to obtain a crystal structure with simultaneously occupied binding sites A and B using combinatorial soaking, yielded a structure with two bound fragments, providing a concrete starting point for linking and further optimisation. Along with describing the findings for NUDT5, strengths and pitfalls of the applied rational structure-guided ligand design are discussed.

6.2. Results

6.2.1. Strategies for fragment elaboration

In efforts to identify novel inhibition modes for NUDT5, crystallographic screening revealed two fragment binding sites – site A and B – that could potentially be targeted (chapter 3). As both sites were located in close proximity, three potential strategies for fragment elaboration were outlined, including separate elaboration of site A and B as well as targeting both sites simultaneously (Figure 6.1).

Elaboration in the active site A. The active site of NUDT5 binds ADP-ribose, the main substrate of the enzyme, as well as several oxidised desoxyribonucleoside diphosphates. The recognition of these substrates involves two tryptophan residues, Trp28 and Trp48, which sandwich the ribonucleotide core in a π-π stacking interaction. In addition, the Glu47 backbone and the Arg51 side chain play a key role by stabilising the substrates through multiple hydrogen bonds (see 1.3).[199,220]

The XChem screening campaign identified 40 fragment hits binding in the site A, providing a variety of options for a novel inhibitor scaffold. In addition, the fragment hits display multiple options for fragment growing by sampling different directions – potential exit vectors – inside the active site (Figure 6.1.a). Due to high sampling of the site A, the main challenge for elaboration is the prioritisation of the fragment hits and selection of a suitable scaffold.

Elaboration of the singleton binder in the site B. The site B is a novel, previously unreported pocket which forms an additional cavity inside the active site. Interestingly,

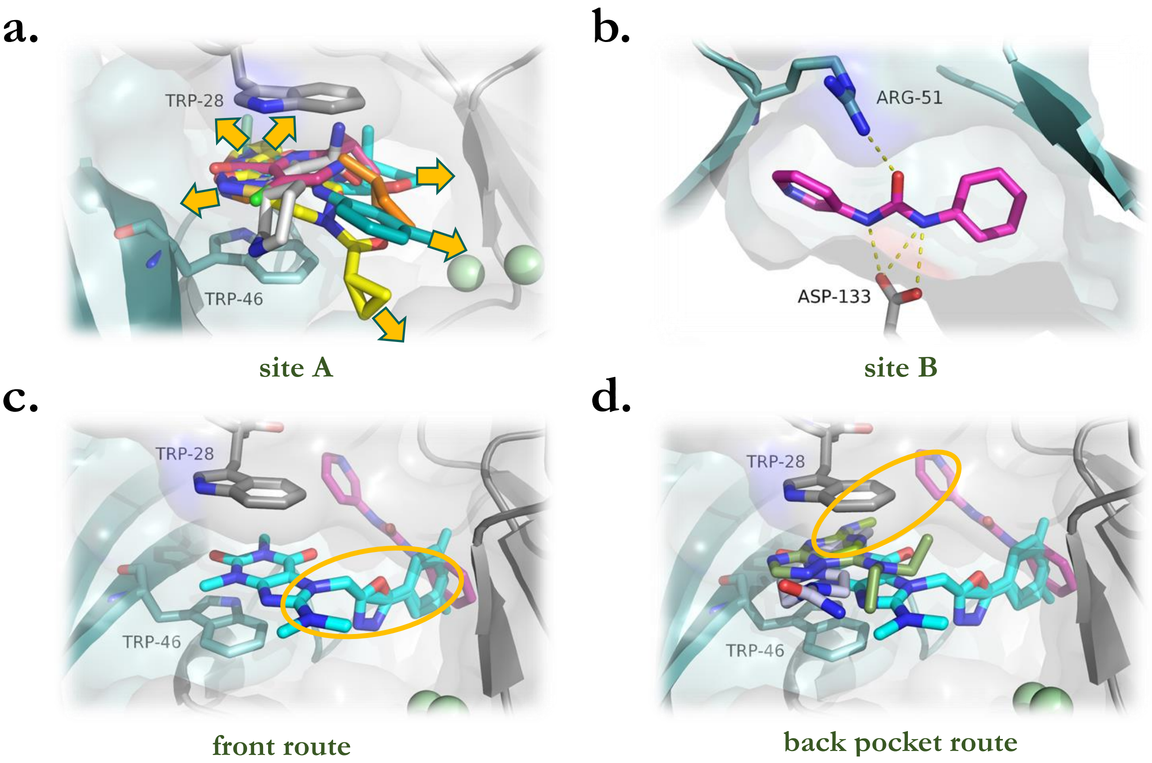


Figure 6.1. Potential strategies for fragment elaboration for NUDT5. **a.** In the site A, fragment hits identified in the XChem screening provide multiple options for fragment growing as indicated by the arrows. **b.** Elaboration in the site B. The solitary binder in the site B which is located in close proximity to the active site is coordinated by key residues Arg51 and Asp133 bearing the promise of a novel inhibition mode (PDB: 5QJ5). **c.-d.** Two strategies for joining the site A and B are potentially feasible: **c.** following the front route demonstrated by Page *et al.*[200] (shown in blue, PDB: 5NQR) or **d.** the novel back pocket route entering deeper into the active site.

fragment **41** is interacting with the catalytic Arg51 through a hydrogen bond (Figure 6.1.b). In addition, the urea moiety of the fragment is coordinated by Asp133 – a residue that stabilises the hydroxyl groups of the endogenous substrate ADP-ribose. Although sterically restricted, this novel site is a suitable starting point for further exploration and bears the promise of a novel inhibition mode for NUDT5.

Joining the sites A and B. The connectivity of both fragment binding site A and B suggests a third potential strategy for elaboration – the design of a compound joining both sites. For that, fragment **41** or its analogues identified in the site B could be linked with one of the fragment hits binding in the active site. However, to assess the feasibility of this approach, simultaneous occupation of both sites A and B needs to be confirmed.

Theoretically, the sides could be bridged in two ways – the front and the back pocket route. The front route would involve linking of a binder in the site A with fragment **41** or its analogues in the site B, following a similar path as the first NUDT5 inhibitor described by Page *et al.* (Figure 6.1.c).[200] These binders follow the surface topology of the protein in its active conformation. The alternative route, covering shorter distance would involve linking of the fragments following the back of the pocket, as shown in Figure 6.1.d. A consequence of this binding mode would involve a distinct conformational change of the catalytic Arg51, preventing its function in substrate recognition.

Given a good sampling of the site, the approach aiming to join both sites seems particularly promising for delivering a new inhibition mode of NUDT5. The undertaken efforts toward the described approaches are summarised in the subsequent sections.

6.2.2. Structural analysis of the site B reveals a conformational change of the catalytic Arg51 upon fragment binding

The site B was identified through binding of a single fragment hit – an urea derivative with a phenyl and a pyridine ring (Figure 6.1.b). As mentioned in 6.2.1, the urea core is coordinated through two distinct hydrogen bonds: to Arg51 and to Asp133. Both residues are involved in recognition of the endogenous substrate ADP-ribose.[220] In addition, kinetic studies carried out by Zha and co-workers[220] demonstrated that a mutation R51Q causes a 17-fold decrease of k_{cat}, indicating a lower turnover and confirming the role of Arg51 in catalysis. The aromatic rings of the fragment hit adopt a staggered conformation and interact with the protein surface purely through hydrophobic interactions.

A structural analysis explains the low sampling of site B in the XChem screening campaign. Both apo- and fragment-bound structures of site A demonstrate that Arg51 adopts the same conformation and is positioned closely to both residues Trp28 and Trp46 involved in substrate recognition (Figure 6.2.a).[220]

Upon binding of fragment **41**, Arg51 is displaced as the loop segment including the residues Arg51 and Thr58 shifts into a different conformation. As a consequence, an additional cavity – the site B – opens up elongating the active site and allowing to accommodate fragment **41** (Figure 6.2.b). Arg51 appears to be a gatekeeper residue regulating the access to this deep cavity inside the active site.

Following ligand binding in site B, the conformational change of the key residue Arg51 may prevent its function in substrate recognition and catalysis, presumably by locking

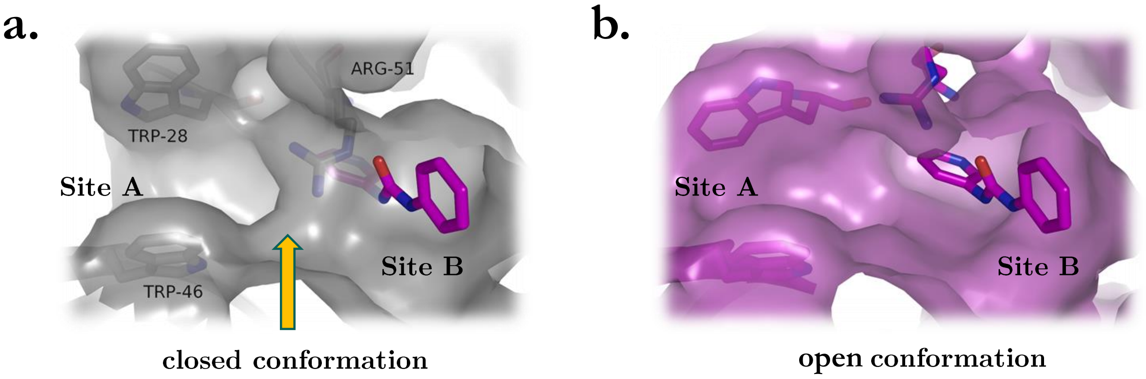


Figure 6.2. The binding of fragment **41** induces a conformational change retracting Arg51 from the binding site. **a.** In the apo-structure of NUDT5 (PDB: 6GRU), Arg51 is positioned in close proximity to key residues Trp28 and Trp46 to carry out its function in substrate recognition and catalysis. **b.** Binding of fragment **41** reveals a deep cavity elongating the active site. This suggests a gatekeeper function for Arg51.

the enzyme in an inactive conformation. Such an inhibition would allow to by-pass the promiscuous and generic nature of the substrate. Exploration of the site B and validation of the fragment binding mode could therefore lead to novel chemical matter for selective NUDT5 inhibition.

6.2.3. Fragment hit exploration in site B

Fragment **41** is supported by weak evidence in the 2F$_o$-F$_c$ map. Replicated high concentration soaks at 500 mM validated its presence in site B and the conformational change of the loops segment (Arg51-Thr58) observed in the initial structure.

Structure-activity relationships of fragment 41 provide understanding of structural elements responsible for binding.[4,299] In absence of a quantitative assay, initial exploration of fragment site B was carried out with rigid docking and subsequent crystallographic validation. 46 close structurally related analogues of fragment **41** were selected using the Fragalysis platform.[300] Fragalysis is a cloud-based tool for fragment analysis developed at DLS which allows to visualise protein-fragment interactions and to map elaboration vectors onto 3D structures.[127] This tool aims to accelerate fragment elaboration by presenting commercially available follow-ups compounds in the Enamine REAL database[301] for each elaboration vector of a given compound, allowing a rapid and specific exploration of the binding site.[127]

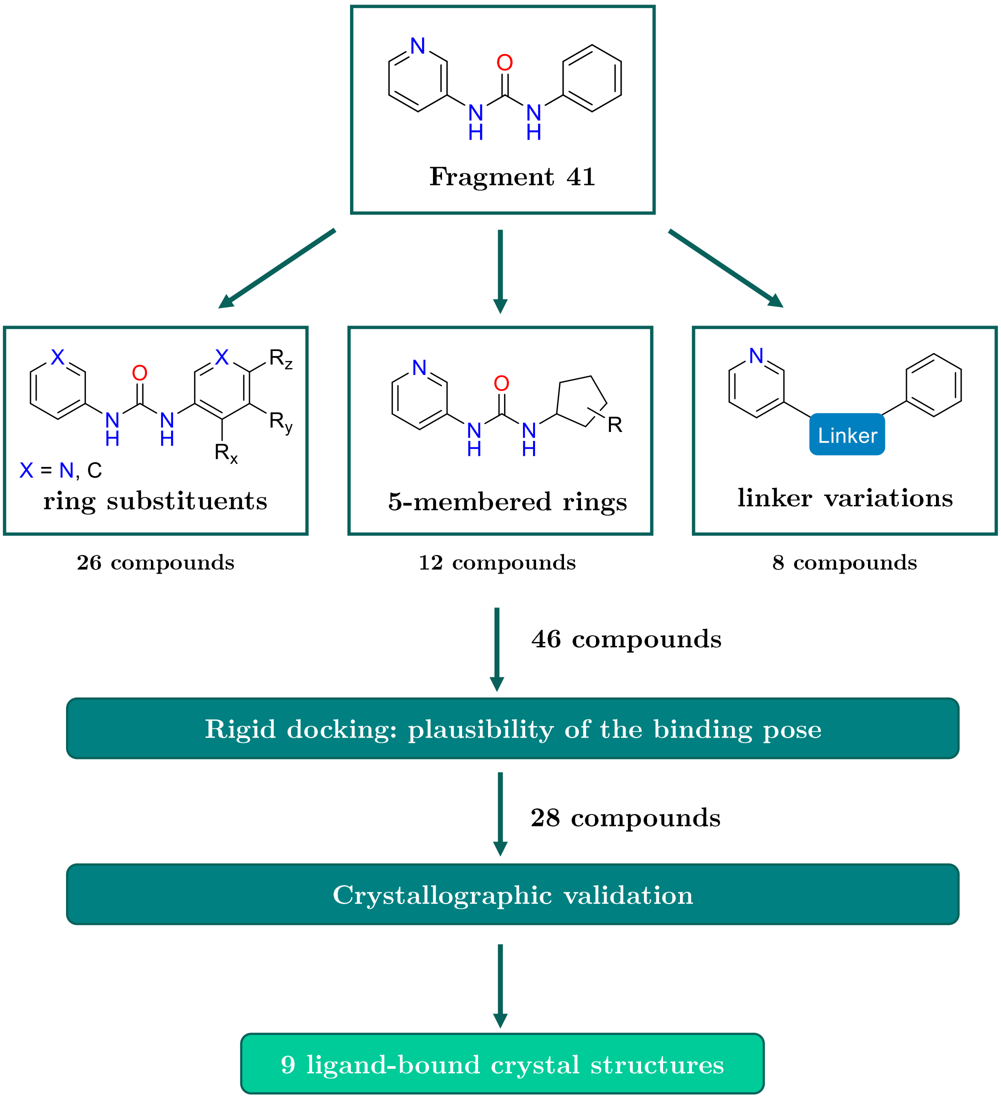
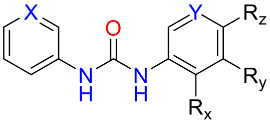


Figure 6.3. Workflow summarising the exploration of the fragment hit in the site B.

Close analogues of the initial fragment hit were selected to explore three elaboration possibilities: substitution of a 6-membered by a 5-membered ring, addition of ring substituents and variation of the urea linker (Figure 6.3). Introduction of substituents of varying size in *ortho*-, *meta*- and *para*-position of the aromatic ring aimed to provide a better understanding of the pocket size and topology, and identify possibilities to form specific interactions beyond the urea core. For the replacement of a 6-membered ring, a variety of aromatic and aliphatic 5-membered rings containing heteroatoms at different positions were selected, exploring formation of a hydrogen bond.[8] In addition, 5-membered rings may provide more suitable geometry for potential joining of the site A and B, as mentioned in 6.2.1. The variation of the urea linker intended to confirm the importance of the identified key interaction to Arg51 and Asp133 for binding and explore the size of the site B.

For prioritisation of experimental efforts, rigid docking experiments were carried out. The selected set of the follow-up compounds was docked into the crystal structure in which fragment **41** was identified (PDB: 5QJ5). For evaluation of the results not only the top ranking but all suggested binding poses were inspected manually. The detailed procedure is described in 2.11.2.

The obtained docking results clearly demonstrated that both interactions with Arg51 and Asp133 are crucial for binding, as none of the linker substitutions retains a similar



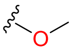
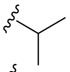
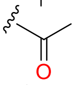
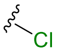
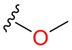
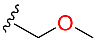
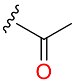
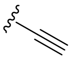
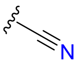
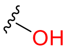
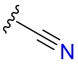
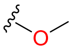
Compound	R_x	R_y	R_z	X	Y	MW [g/mol]	crystallographic validation
FU001/2		-H	-H	N, C	C, N	243.3	✓ –
FU003		-H	-H	N	C	255.3	–
FU004		-H	-H	N	C	255.3	–
FU005		-H	-H	C	N	247.7	–
FU006	-H		-H	N	C	243.3	✓
FU007	-H		-H	N	C	257.3	–
FU008	-H		-H	N	C	255.3	–
FU009/10	-H		-H	N, C	C, N	292.1	✓ ✓
FU011	-H		-H	N	C	237.3	✓
FU012	-H		-H	N	C	238.3	✓
FU013	-H	-H		N	C	229.2	–
FU014	-H	-H		N	C	243.3	–
FU015	-H	-H		C	C	242.3	–

Figure 6.4. Variations of ring substituents on the core of fragment **41** selected for experimental validation after filtering by rigid docking.

binding pose as the original fragment. In addition, an extended linker as in the compound **FU032** (appendix Figure A.11) cannot be accommodated due to space limitations. As a consequence of their poor performance in the rigid docking, the compounds exploring substitution for the urea linker were excluded from the experimental validation.

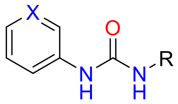
Further, the rigid docking results show that a variety of substituents in different positions can be accommodated. While the *ortho*- and *meta*-positions allow to incorporate larger substituents such as acetyl- and bromide groups on both ring sites, the *para*-position seems to be less favourable as it is sterically hindered. Therefore, the majority of the compounds with a *para*-substituent were removed from the crystallographic validation. The analogues selected for experimental testing are shown in Figure 6.4.

The docking poses obtained with compounds containing a 5-membered ring replacing one of the aromatic rings shown that most of the analogues yield plausible docking poses retaining the key hydrogen bonds. Interestingly, substitutions for both the pyridine and the phenyl rings were observed equally often. The selection of the compounds for experimental testing is shown in Figure 6.5.

Based on the described filtering by rigid docking, 28 follow-up compounds were taken into crystallographic validation. Similarly as in the XChem screening campaign, the compounds were soaked into the pre-formed NUDT5 using compound concentrations of 500 mM and 200 mM, while keeping the DMSO content at 10%. The detailed procedure is described in 2.8.3. In total, 9 unique protein-ligand structures could be obtained (Figure 6.6). The structure with **FU021** and **FU022** were obtained from soaking experiments at 500 mM, the majority of the structures were obtained at 200 mM due to limited compound solubility.

As expected, the structures with analogues are coordinated by the key residues Arg51 and Asp133 and demonstrate the same binding mode as the original fragment **41**. The only exception is the compound **FU029** which lacks a hydrogen bond to Arg51 and adopts a different conformation. Indeed, the substitution of the urea oxygen by a sulfur atom results in loss of H-bond acceptor functionality, confirming the specificity of the Arg51 interaction to the urea core.

In all structures, the aromatic ring pointing into the deep cavity of the site B – shown on the left in Figure 6.6 – adopts the same distinct conformation, highlighting the limited space of the pocket that can only accommodate a flat aromatic system with potentially



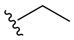
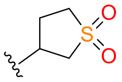
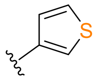
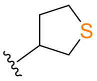
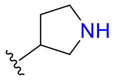
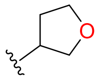
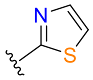
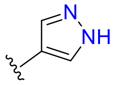
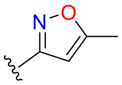
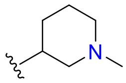
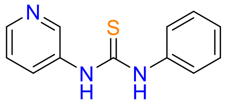
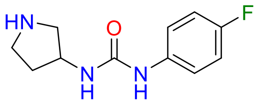
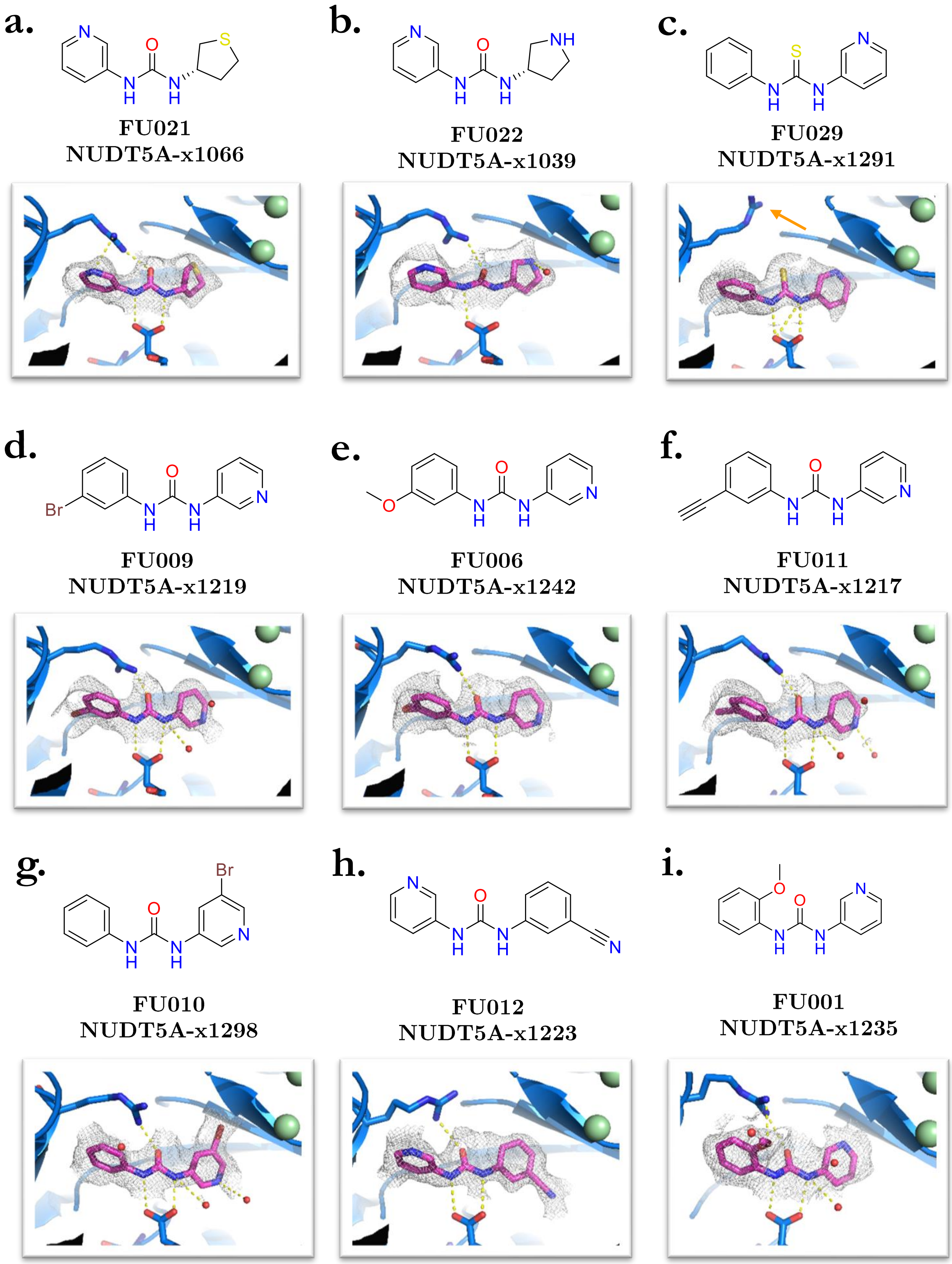
Compound	R	X	MW [g/mol]	crystallographic validation
FU016		N	165.1	–
FU017/18		N, C	255.3/ 254.3	–
FU019/20		N, C	219.3/ 218.3	–
FU021		N	223.3	✓
FU022		N	206.2	✓
FU023		N	207.2	–
FU024/25		N, C	220.3/ 219.3	–
FU026		N	203.2	–
FU027		C	217.1	–
FU028		C	233.3	–
FU029			229.3	✓
FU030			223.3	–

Figure 6.5. Variation of ring structures selected for crystallographic validation after filtering by rigid docking. Compounds were obtained as racemic mixtures.



a. **FU021** **NUDT5A-x1066**

b. **FU022** **NUDT5A-x1039**

c. **FU029** **NUDT5A-x1291**

d. **FU009** **NUDT5A-x1219**

e. **FU006** **NUDT5A-x1242**

f. **FU011** **NUDT5A-x1217**

g. **FU010** **NUDT5A-x1298**

h. **FU012** **NUDT5A-x1223**

i. **FU001** **NUDT5A-x1235**

Figure 6.6. Crystallographically validated follow-up compounds based on fragment **41**. The event density is shown at 2x(1-BDC) σ as grey mesh. The orientation of the pyrimidine in **FU029** (**c.**) cannot be determined due to the symmetry of the molecule.

meta-substituted position. While the docking results suggested that 5-membered rings could potentially substitute both the phenyl- or the pyridine ring, the experimental data only supports the presence of a 5-membered ring on the more open side located further away from the active site – the right site in Figure 6.6.b-c. From the analogues with ring substitutions shown in Figure 6.5, only the 5-membered aliphatic rings yielded co-structures. This suggests that the right site of the pocket favours the accommodation of lipophilic ring systems with large hydrophobic surface.

In three of the obtained structures, analogues containing various ring substituents in *meta*-position were identified (Figure 6.6.d-f). The obtained data suggests that substituents with limited steric constraints in ring proximity such as bromide, alkyne and methoxy-groups can be accommodated in the deep cavity occupied by the aromatic ring (left side in Figure 6.7), whereas acetyl- and methoxymethyl-moieties (**FU007** and **FU008** in Figure 6.4) require more space as they are larger and rotatable. This distal *meta*-position on the ring pointing into the deep cavity can therefore be exploited for affinity increase.

The substituents in *meta*-position on the right ring, shown in Figure 6.6.j-h, allow for more flexibility and demonstrate two confirmations exploring different directions in fragment site B. These suggest an avenue for larger substitution and further exploration, especially since these finding may be informative for joining the site A and B.

Interestingly, the only analogue carrying a substitution in *ortho*-position (Figure 6.6.i) points towards the site A and could be used as a possible route to link both fragment binding sites.

The experimentally validated features informing further elaboration process are summarised in a pharmacophore model, as shown in Figure 6.7.

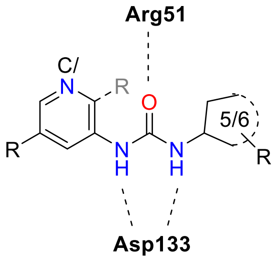


Figure 6.7. The identified pharmacophore for hit exploration in the site B.

In summary, compounds building on the initial fragment should retain the key interactions with Arg51 and Asp133, while conserving a 6-membered aromatic ring for binding into the deep cavity. On this side, small substituents in *meta*-position can be accommodated. The right side offers more options for introduction of larger substituents and is suited for a fragment growing approach. Due to size limitations of site B, joining of site A and B appear to be a promising approach for gaining specificity and affinity.

6.2.4. Understanding the dynamics of the binding site

Comprehensive understanding of accessible conformations of the binding site is required to assess the feasibility to join sites A and B in a fragment linking approach.

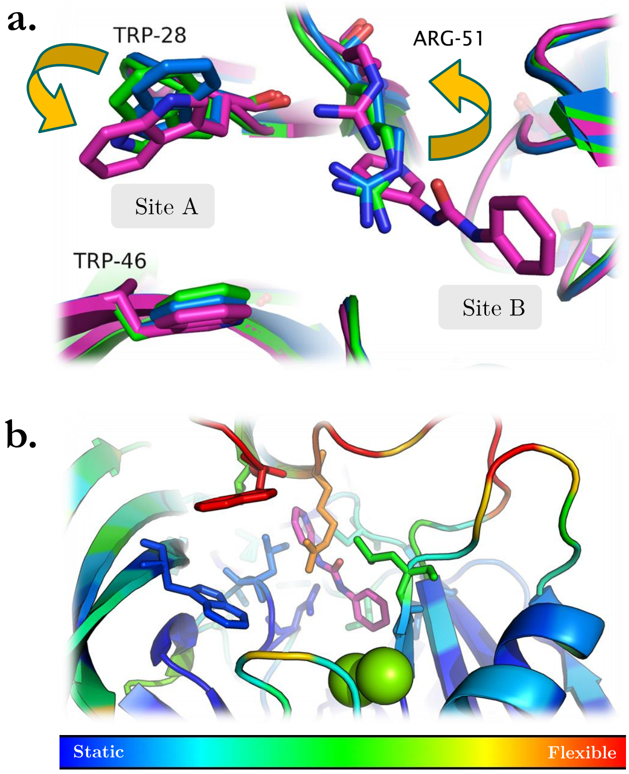
The comparative analysis of the NUDT5 crystal structures in absence and presence of fragment **41** already framed two distinct conformations of the binding site. In presence of the site A binders, both Trp28 and Trp46 are aligned creating a π-π interaction. As shown in Figure 6.8.a, the catalytic Arg51 is then positioned closely to the tryptophan residues, pointing toward the active site. In this conformation, represented by most of the crystal structures, the deep cavity comprising the fragment site B is not accessible. In the second conformation, revealed upon binding of fragment **41** or its analogues, the catalytic Arg51 shifts away from the binding site, thus opening up the site B. In this confirmation, the tryptophan residues are not aligned.

Designing a ligand joining sites A and B requires bridging of two conformations. Therefore, knowledge of the dynamics in the active site, including flexibility of the key residues involved in binding, is essential.

For understanding the dynamics of the active site, all atom molecular dynamics simulations using the apo-structure of NUDT5 (PDB: 6GRU) in explicit water were carried out using Gromacs and the Amber99sb forcefield over 100 ns. The simulation reveals a contrasting flexibility for tryptophan residues in the active site: while Trp46 is rather static, Trp28 displays high flexibility (Figure 6.8.b).

In the course of the simulation, the structure switches between two conformations of the tryptophan sandwich: aligned and shifted, induced by a 180° flip of the Trp28. Similarly, influenced by the position of Trp28, Arg51 shows high flexibility. The loop segment ranging from Arg51 to Thr58 is equally highly variable, thus influencing the space available in the binding site (Figure 6.8.c).

Taken together, the dynamics of the binding site as evidenced by MD appear to support the joining strategy of both fragment sites. The highly flexible and dynamic character



c.

residue	**RMSF** [Å]
Trp28	1.99
Trp46	0.87
Arg51	1.67
Thr52	1.54
Thr53	1.97
Arg54	3.01
Lys55	3.02
Glu56	2.67
Gln57	2.18
Thr58	1.75
Ala59	0.99
protein residue average	1.06

Figure 6.8. Understanding the dynamics of the NUDT5 binding site. **a.** Comparative analysis of the crystal structures reveals a distinct conformation in presence of fragment **41**. The structure of the complex with fragment **41** (PBD: 5QJ5) is shown in magenta, with ADP-ribose (PDB: 2DSC) in blue and with 8-oxo-dGDP (PDB: 3AC9) in green. **b.** Flexibility analysis from molecular dynamics simulations using the apo-structure (PDB: 6GRU) reveals highly dynamic Trp28 and Arg51. The heap map colouring by root mean square fluctuation (RMSF per residue) visualises the flexibility of each residue. The fragment **41** was inserted after the simulation. **c.** RMSF values (averaged over both NUDT5 chains) for the key residues and the flexible loop segment containing Arg51.

of Trp28 and Arg51 should allow to accommodate ligands as these conformations are populated in the MD simulations at 298 K. However, careful ligand design is required as the tryptophan residues might not be aligned in the resulting structure, creating a weaker π-π interaction. Locking Arg51 into the open conformation bears the promise of inhibition without stark demands on the compound affinity.

6.2.5. Towards joining sites A and B: growing of the B site hit

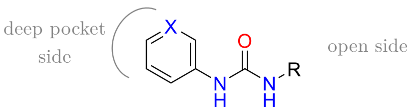
Building on the pharmacophore for site B and the insights into the active site dynamics of NUDT5, the binding of larger follow-up compounds with additional ring systems or substituents was explored. To enable a rapid elaboration cycle, commercially available compound from the Enamine REAL[301] database were selected by a substructure search using the original fragment hit and its validated analogues. As for previous rounds of follow-up compounds, the selection was filtered by rigid docking, all ligands with plausible binding pose were retained.

For experimental validation using X-ray crystallography, 24 follow-up compounds were selected. These ligands can be grouped in two categories. The first series shown in Figure 6.9 explores binding of larger substituents including various ring systems as well as aromatic and aliphatic rings connected by a flexible linker in the open side − while retaining the pyridine pointing toward the deep pocket side (left side). The second series shown in Figure 6.10, the aromatic ring in the open side (right side) is substituted by a pyrrolidine / piperidine, while larger ring systems are introduced in the deep pocket side. This series aims to assess previous observation of 5-membered rings only binding in the open and more accessible side. In addition, it explores the possibility of growing from the deep pocket into site A starting from *meta*-position.

The binding of the compounds was validated using X-ray crystallography. The compounds were soaked into NUDT5 crystals at different concentrations including 200 mM, 100 mM and 50 mM. The experiments were carried out in triplicates for each compound and concentration to account for any variability such as low solubility.

The crystallographic validation could only confirm binding of one follow-up compound containing a octahydroindolizine group, **FU038**, at 200 and 50 mM (Figure 6.11). However, the experimentally obtained binding mode does not correspond to the expected. According to the docking results, the compound **FU038** was predicted to adopt a similar pose as the original fragment **41**, with pyridine ring pointing into the deep cavity and octahydroindolizine group occupying the open side (Figure 6.11.a). In the obtained crystal structure, the flexible loop segment containing Arg51 does not display the same shift as observed upon binding of the original fragment hit (Figure 6.11.b). Consequently, the deep cavity remains inaccessible. Compound **FU038** occupies the elongated side of the active site in close proximity to the magnesium ions. While the urea oxygen is coordinated by one of the magnesium ions, the aliphatic octahydroindolizine ring points towards the open site. The overlap with the structure containing the original fragment **41** reveals that both the phenyl and the octahydroindolizine rings are located perpendicularly to each other (Figure 6.11.b).

This finding suggests a possible entry route for fragment **41** and its analogues into the



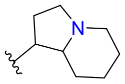
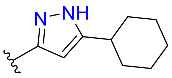
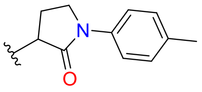
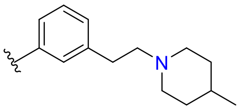
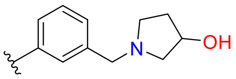
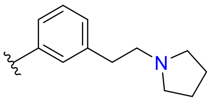
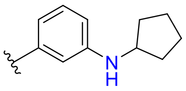
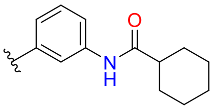
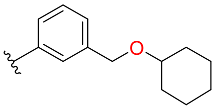
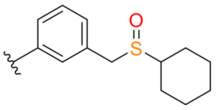
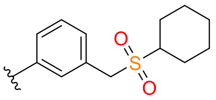
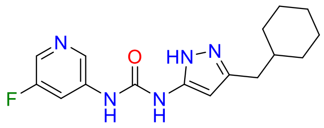
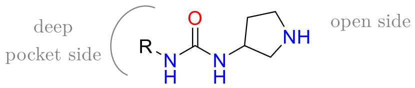
Compound	R	X	MW [g/mol]	crystallographic validation
FU038		N	260.3	✓
FU039		N	285.4	–
FU040		N	310.4	–
FU041		N	338.5	–
FU042		N	312.4	–
FU043		N	310.4	–
FU044		C	295.4	–
FU045		N	338.4	–
FU046		N	299.4	–
FU047		N	357.5	–
FU048		N	373.4	–
FU049			317.4	–

Figure 6.9. Variation of substituents replacing the pyridine ring in fragment **41** structure selected for crystallographic validation. Compounds were obtained as racemic mixtures.



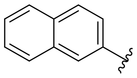
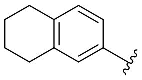
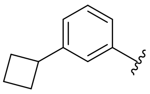
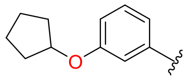
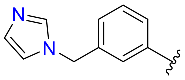
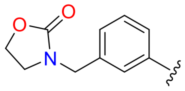
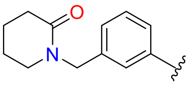
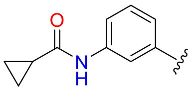
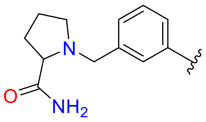
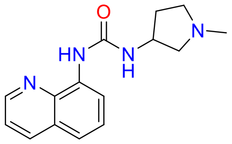
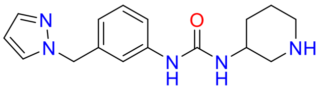
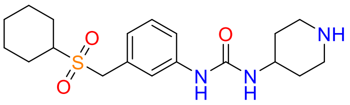
Compound	R	MW [g/mol]	crystallographic validation
FU050		291.8	–
FU051		259.4	–
FU052		259.4	–
FU053		289.4	–
FU054		285.4	–
FU055		304.4	–
FU056		316.4	–
FU057		288.4	–
FU058		331.4	–
FU059		270.3	–
FU060		299.4	–
FU061		379.5	–

Figure 6.10. Variation of substituents replacing the pyridine ring in the follow-up compound **FU022** structure selected for crystallographic validation. Compounds were obtained as racemic mixtures.

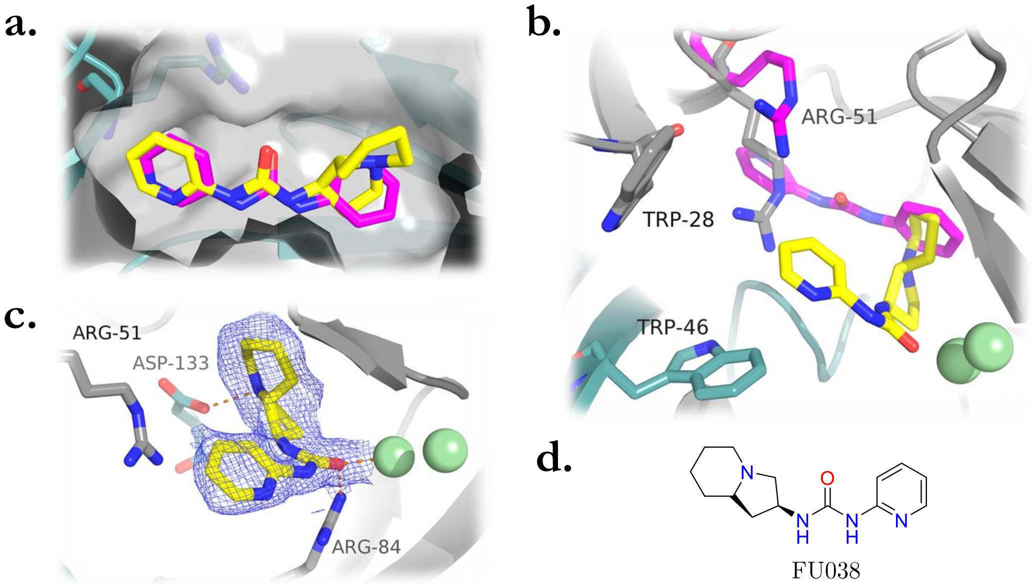
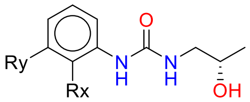


Figure 6.11. Comparison of predicted and experimental results for binding of for the compound **FU038**. **a.** Binding pose of the follow-up compound **FU038** (yellow) as predicted by rigid docking in comparison to the binding mode of the original fragment **41** (magenta). **b.** Overlay of the structures with **FU038** (yellow) and fragment **41** (magenta) reveal that **FU038** does not enter the deep cavity of the site B and occupies the elongation of the active site. **c.** The binding mode of the compound **FU038** in the obtained crystal structure demonstrates that the urea oxygen is coordinated by the guanidinium group of Arg84 and Mg^{2+}. In addition, the ring nitrogen is stabilised by Asp133. The event density is shown at 1-(2xBDC) as purple mesh. **d.** The chemical structure of the follow-up compound **FU038**.

deep cavity in soaking experiments: supposedly, the compound binds in proximity of the two magnesium ions first and subsequently slides behind Arg51, allowing the conformational shift. However, the follow-up compound **FU038** entered the active with the octahydroindolizine ring first, blocking its insertion into the deep cavity. The retention of the compound at the Mg^{2+} site, facilitated by the coordination through Arg84 and Asp 133, enabled the identification of this intermediate binding pose (Figure 6.11.c). This observation implies that derivatives bulkier than initial fragments may not be able to enter the deep cavity in soaking experiments when the crystal is already formed.

Steric constrains impaired on the progression of compounds joining sites A and B through the deep pocket route. Growing of the fragment hit from the deep pocket towards the site A using the proximal *meta*-position of the phenyl ring did not lead to the identification of any binders in the soaking experiments. This suggests that the deep cavity of the site B is too narrow for such steric demands and the joining of the two sites would require a connection in *ortho*-position, if possible. However, initial tests with compounds from the SGC collection containing *ortho*-substituted phenylureas were unsuccessful (Figure 6.12), which could be related to the isopropyl group being unsuited or the amide linker in *ortho*-position introducing unfavourable steric constraints. Therefore, further exploration of the *ortho*-position is required.



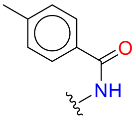
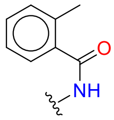
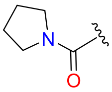
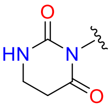
Compound	**R$_x$**	**R$_y$**	**R$_z$**	**MW [g/mol]**	**crystallographic validation**
NU000828a		-H	-H	327.4	–
NU000829a		-H	-H	327.4	–
NU000825a	-H		-H	291.4	–
NU000824a	-H		-H	306.3	–

Figure 6.12. Variation of substituents in *ortho*- and *meta*-positions in analogues available from the SGC compound collection.

In summary, the follow-up compounds have a substantial increased size compared to the original fragment **41**. Binding of larger ligands could potentially require a significant conformational change.[116,118] Compound soaking into pre-formed crystals may not allow for induction of such conformational changes, preventing the compounds from binding. While co-crystallisation could be an alternative validation, it is less amenable for testing in a high-throughput manner as each condition may require individual optimisation.[116] Therefore, instead of growing from one site into another, an alternative approach using co-soaking of two compound into different sites was attempted.

6.2.6. Co-soaking of multiple fragments towards linking

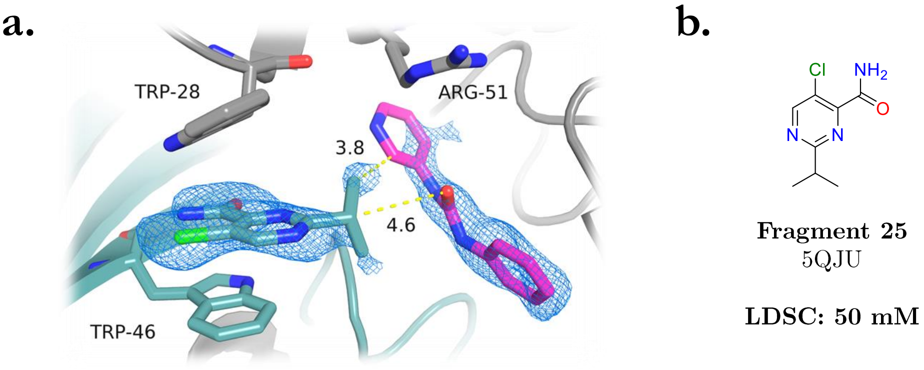
Following the comparative analysis of the crystal structure and the MD simulation, an experimental validation of simultaneous occupation of the fragment binding sites A and

B was sought.[302] Upon initial binding of the fragment hit inside the deep cavity in site B, the flexible loop segment would undergo a conformational change shifting the Arg51 away from the active site. The altered conformation of the binding site would allow to evaluate, whether a different fragment can bind between both tryptophan residues. The XChem fragment hits obtained for the site A differ in their size and binding mode and provide a suitable test set for such exploration. A co-structure containing two bound fragments in both sites A and B would not only demonstrate the feasibility of this approach, but also facilitate fragment linking proving concrete starting points.[302]

First, NUDT5 crystals were soaked with fragment **41** at the concentration of 500 mM. Following a short incubation, a second fragment previously identified in the active site A was added. In total, 40 fragment hit combinations were tested in triplicates. For all soaks, high concentration solutions of fragments at 500 mM were used, while keeping the total DMSO content at 10%. The detailed experimental procedure is outlined in 2.8.5.

From 120 soaking experiments, 118 datasets could be successfully processed and analysed by the PanDDA algorithm.[113] While the majority of the obtained structures contained fragment **41** only, nine exclusively presented binding of site A fragments. Among the triplicates, reproducibility was low, with only one out of three replicates containing a site A binder. From the conducted experiment, it remains unclear whether these active site binders outcompeted fragment **41** or whether experimental variations caused its absence.

The co-soaking experiments yielded five structures presenting evidence for binding of



Fragment 25
5QJU

LDSC: 50 mM

Figure 6.13. Co-soaking experiments reveal a simultaneous binding of two fragment hits. **a.** While fragment **41** (magenta) binds in the deep cavity of the site B, fragment **25** (green) occupies in the active site A. The 2F$_o$-F$_c$ density at 1 σ is shown as blue mesh. **b.** The chemical structure of fragment **25**.

both fragment **41** and a second binder occupying the active site. The structure presenting the most compelling evidence contained fragment **41** in site B and fragment **25** in site A (Figure 6.13), which was observed in one of three replicates. Crystallography intrinsically presents an average of the observed states in the crystal. Therefore, the possibility cannot be excluded that the observed simultaneous binding of these compounds is an artefact resulting from crystallographic averaging over two distinct states binding each a single ligand.

The identified fragment **25** has a LDSC of 50 mM and is therefore a stronger binder. Both fragments were observed in their previously described binding modes and are located closely to each other. The distance between the isopropyl carbon of fragment **25** and *ortho*-position of the pyridine ring is only 3.8 Å, suggesting a possible site for connecting the two molecules. The fragments adopt a perpendicular orientation to each other, creating less than ideal geometry for linking. However, screening of close fragment **41** analogues revealed the possibility to replace one of the aromatic rings through a 5-membered ring which may provide a more suitable conformation. As suggested by the MD simulation, the tryptophan rings do not align perfectly, with the static Trp46 interacting with the aromatic core of fragment **25** to create π-π stacking.

The ultimate feasibility of the linking both site A and B following the back pocket route can only be assessed through additional bespoke synthesis and experimental validation of the linked compound. Upon success, the desired inhibition effect on NUDT5 needs to be validated in an activity assay.[200]

6.3. Discussion and Outlook

In the efforts to target NUDT5 in a novel inhibition mode, the elaboration was focussed on the newly discovered and previously unreported fragment site B. Possible elaboration strategies including growing of the solitary fragment hit and potential joining of the sites A and B through fragment linking were explored.

The initial exploration of fragment **41** using SAR by catalogue approach, supported by rigid docking and validated by crystallographic soaking experiments yielded a pharmacophore model summarising the key features for further ligand design. The attempt of further elaboration aiming to grow fragment **41** towards site A using larger substituents was unsuccessful, only resulting in a structure containing a follow-up compound in an unexpected binding mode. However, co-soaking of fragment hits for both sites A and B yielded an interesting starting point for joining the two sites through the back pocket of the active site, while displacing the catalytic gatekeeper residue Arg51.

Combining SAR by catalogue and high-throughput crystallography allows for rapid hit exploration. The SAR by catalogue approach allows for rapid exploration of the binding site using databases of readily available compounds. This approach is particularly suitable when access to medicinal chemistry support is limited or not available. The successful application of SAR by catalogue has been demonstrated in numerous examples of hit to lead optimisations.[303–305] Selection of close analogues for fragment **41** enabled precise hypothesis testing, pinpointing steric constraints, sites for potential functional group additions and key interactions of the novel fragment site B. However, the exploration of a binding site may require bespoke synthetic routes and therefore be limited by the availability of suitable compounds in commercial databases.[46] As such, the exploration of substituents in *ortho*-position of fragment **41** was constrained by lack of suitable commercially available compounds.

Docking is a helpful filter for selection of close analogues. In the described hit exploration cycles, rigid docking was used as a filter for selection of follow-up compounds for subsequent crystallographic validation. The compound analogues were docked into a crystal structure with accessible binding site B following a rigid docking protocol, using fragment **41** as positive control. When introducing moderate changes to the original fragment such as ring substituents, the observations from docking experiments could largely be confirmed by the obtained experimental validation. Thereby, rigid docking provided a useful filter to reduce the experimental efforts.[306]

Beyond simple functional group variations, the success of binding mode predictions using rigid docking for follow-up compounds with significantly larger substituents was rather low. Only one structure in complex with a follow-up compound could be obtained, however showing an unexpected binding mode. While fragment and large drug-like molecules do not display significant differences in docking performance, failures to predict the binding mode mainly occur for two reasons: undersampling of the conformational space and inaccurate scoring.[36,306,307] The first point particularly applies to larger molecules and explains their poor performance. More reliable binding pose predictions can be achieved through induced fit docking.[308] Using an ensemble of different receptor conformations, such protocols take information upon binding site flexibility into account[308,309] However, induced fit docking protocols are typically more time- and resource-intensive compared to rigid docking.[310]

Choosing an adequate experimental validation for weak binders. The sensitivity of X-ray crystallography combined with the high reproducibility of the XChem implementation at DLS allows for experimental validation of small weak binders and their follow-up compounds. However, it does not provide a functional readout. In absence of a biochemical assay, the impact of site B binders on enzyme activity, such as potential inhibition, could not be validated yet. On the other hand, larger follow-up compounds have a sufficient size to display modulation effects in a biochemical assay, yet are not always suited for use in soaking experiments. The follow-up compounds with larger ring systems presented here did not show binding in soaking experiments, but might be detected in site B in co-crystallisation experiments as it allows for larger conformational changes. Therefore, it is necessary to establish a co-crystallisation system for characterisation of follow-up compounds.

Co-soaking is promising for fragment joining. Tailored soaking experiments can be another experimental method for elaboration of compounds in absence of a biochemical assay. This work describes co-soaking of fragments to identify simultaneous binding in the sites A and B of NUDT5 and to discover concrete possibilities for joining. Although at least one suitable combination was found, the practicalities of co-soaking experiments require further optimisation to fulfil their promise. In order to fix the protein in the desired conformation, NUDT5 crystals were incubated with fragment **41** first, before the second compound was added. While a longer incubation time with first fragment could potentially benefit the experiment, the short-lived nature of NUDT5 crystals upon prolonged solvent exposure does not allow significant time or concentration increase. The diffusion time of a small molecule binder into a protein crystal largely depends on physical and chemical properties of both the protein crystal and the small molecule including protein density, size and binding affinity.[311] Therefore, a systematic exploration of soaking time to reach sufficient occupancy could determine optimal incubation.

A more robust alternative to access the desired protein conformation would be a prior co-crystallisation of NUDT5 with fragment **41**, followed by soaking of fragments binding in site A. As the co-crystallisation conditions may not be identical to the soaking condition in which site A binders were identified, their solubility may be impacted and would require further optimisation.

In this work, the fragments used in the co-soaking experiment were limited to the hits previously identified in the site A. Screening of smaller fragments such as MiniFrags[62] or FragLites[76] in combination with the fragment **41** could result in identification of new joining opportunities.

While the co-soaking of fragment **41** and **25** presented sufficient evidence for binding, the remaining crystallographic datasets displayed more ambiguous data containing partial occupancy for both Arg51 states. This, in turn, makes it unclear whether both fragments bind simultaneously or the observed crystallographic state results from superposition of different protein conformations. Data analysis methods enabling further deconvolution of the bound state might resolve such ambiguous cases.

Towards a novel inhibition mode for NUDT5. The discovery of fragment **41**, displacing catalytic Arg51 as gatekeeper and opening a deep pocket in the binding site called for a deeper investigation whether this structural feature can be harnessed for therapeutic modulation of NUDT5. Conceptually, two routes were designed for inhibition, taking advantage of this new pocket: the front route, prolonging the site occupied by the inhibitors of Page *et al.*[200], targets the known and easily accessible conformation of the binding site. The back pocket route requires a conformational shift in solution, but would allow to design smaller inhibitors displaying higher selectivity for NUDT5, bypassing the promiscuity of site A.

Initial attempts at growing fragment **41** did not yield any concrete indication for joining of both sites. Encouragingly, the co-soaking experiment delivered an interesting co-structure containing both fragment **41** and **25** − which are among the strongest binders (LDSC: 50 mM). This crystal structure provides a promising starting point for further medicinal chemistry efforts to target NUDT5 for breast cancer.

7

Summary & Outlook

The high sensitivity of crystallographic fragment screening is both a blessing and a curse. As demonstrated in this work, it allows to detect novel binding sites and fragment hits sampling additional protein-ligand interactions that can provide starting points for further exploration.

As a consequence of the high sensitivity of X-ray crystallography as well as low affinity and small size of fragments, crystallographic hits can rarely be detected in other commonly used biophysical methods. However, quantitative binding information is essential for both rapid selection of fragment hits and monitoring of their affinity gain in initial elaboration cycles. This presents particular challenges for prioritisation and quantification of crystallographic fragment hits. Novel tools for ranking and affinity determination are necessary to streamline the elaboration of weak binders into potent lead compounds. In this work, two methods for ranking of weak binders were explored: concentration-dependent crystallographic soaking and relaxation-based $T_{1\rho}$ NMR.

The crystallographic proof of principle study demonstrates a correlation between lowest detectable soaking concentration and binding affinity of a fragment hit. Low concentration counter-soak is proposed to triage fragment hits according to their binding strengths. This additional step can easily be implemented following a screening campaign at high soaking concentration using the already existing infrastructure at screening platforms, such as the XChem facility at DLS. The combination of the wealth of information provided by weak fragment hits detected at high soaking concentration combined with prioritised fragments enables a streamlined and structure-guided elaboration into potent compounds.

Therefore, its application is recommended for prioritisation of fragments hits for targets with a high hit rate, whose affinities cannot be measured by commonly used biophysical methods (e.g. SPR, MST). The consistent use of X-ray crystallography for both fragment identification and ranking does not require the development of an orthogonal assay, bypassing the discrepancies in assay results.

The correlation between fragment affinity and its lowest detectable soaking concentration (LDSC) has an impact beyond fragment screening. This study shows that a quantitative dimension can be introduced to X-ray crystallography of protein-ligand complexes, by using the detection limit of fragment binding as a value for their affinity. Although previous studies have applied concentration-dependent crystallographic soaking, these are limited to relative ranking of compounds in competition[261] or comparison of fragment binding at secondary sites of the same target[260]. In contrast, concentration-dependent crystallography can be directly applied to obtain an affinity metric (LDSC) to rank fragments binding to the same target in different sites, or to assess the strengths of selected interactions.

This thesis is the first to systematically explore the relationship between ligand affinity and crystallographic soaking concentration, which only became possible through high reproducibility (93 %) of the high-throughput crystallography infrastructure at DLS. As such the dataset generated for NUDT5 in this thesis provides a stepping stone for further developments in quantitative macromolecular crystallography. Following the theoretical framework outlined in chapter 4, this dataset can be used for advancing algorithms enabling the calculation of apparent crystallographic KD directly from the crystallographic occupancy. Further, it can provide experimental validation for computational predictions of fragment binding affinity, such as dynamic undocking.[312,313]

The detection range of NMR comes closest to crystallography. Quantification of ligand binding using currently available ligand-observed NMR methods is typically a lengthy process requiring multiple ligand titrations. A relative NMR-based ranking of weak binders would allow to accelerate their prioritisation, focussing medicinal chemistry activities on the most promising starting points. Ligand-observed NMR experiments with protein titration were investigated as practicable ranking method for hit triage. Although measuring $T_{1\rho}$ has the most promising potential from the NMR methods assessed in this work, further investigations are required to determine the contribution of binding affinity to the observed ^1H signal.

Fragment elaboration typically comprises an iterative cycle of compound design followed by their testing in a biochemical or biophysical assay, which allows to monitor the progress of elaboration. The low affinity of crystallographic fragment hits impedes their detection in orthogonal quantitative methods. The initial cycles of elaboration can nonetheless be monitored by X-ray crystallography until sufficient affinity is reached. Indeed, this unconventional approach has proven valuable in the development of a kRAS inhibitor.[314] By obtaining structural information first, the precise understanding of the binding mode guided the design of inhibitors by rationalising affinity readouts – illustrating the power of structure-based approaches on challenging drug targets.[314]

In absence of a quantitative assay, initial exploration of the fragment hit in the newly discovered NUDT5 binding site was carried out in a SAR by catalogue approach using X-ray crystallography as a readout. Various substituents and alternative ring structures exploiting different exit vectors from the pocket were explored, leading to a possible route to connect the new binding site with the known active site. Hereby, the application of co-soaking provided concrete starting points for fragment linking. While the linking could not be completed, this thesis provides the groundwork for the development of novel NUDT5 modulators.

Bibliography

[1] Khanna, I. Drug Discovery in Pharmaceutical Industry: Productivity Challenges and Trends. *Drug Discovery Today* **2012**, *17*, 1088–1102.

[2] Dahlin, J. L.; Inglese, J.; Walters, M. A. Mitigating Risk in Academic Preclinical Drug Discovery. *Nature Reviews Drug Discovery* **2015**, *14*, 279–294.

[3] Norton, R. S.; Jahnke, W. NMR in Pharmaceutical Discovery and Development. *Journal of Biomolecular NMR* **2020**, *74*, 473–476.

[4] Stocks, M.; Alcaraz, L.; Griffen, E. *On Medicinal Chemistry*; Sci-ink Ltd: Great Britain, 2007.

[5] Tamimi, N. A. M.; Ellis, P. Drug Development: From Concept to Marketing! *Nephron Clinical Practice* **2009**, *113*, c125–c131.

[6] Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L. Principles of Early Drug Discovery. *British Journal of Pharmacology* **2011**, *162*, 1239–1249.

[7] Klebe, G. *Drug Design: Methodology, Concepts, and Mode-of-Action*; Springer-Verlag: Berlin Heidelberg, 2013.

[8] Patrick, G. L. *An Introduction to Medicinal Chemistry*, 5th ed.; OUP Oxford: Oxford, 2013.

[9] Davis, A.; Ward, S. E. *The Handbook of Medicinal Chemistry*; 2014.

[10] Davis, B. J.; Hubbard, R. E. *Structural Biology in Drug Discovery*; John Wiley & Sons, Ltd, 2020; Chapter 4, pp 79–98.

[11] Bleicher, K. H.; Böhm, H.-J.; Müller, K.; Alanine, A. I. Hit and Lead Generation: Beyond High-Throughput Screening. *Nature Reviews Drug Discovery* **2003**, *2*, 369–378.

[12] Holenz, J.; Stoy, P. Advances in Lead Generation. *Bioorganic & Medicinal Chemistry Letters* **2019**, *29*, 517–524.

[13] Erlanson, D. A. In *Fragment-Based Drug Discovery and X-Ray Crystallography*; Davies, T. G., Hyvönen, M., Eds.; Topics in Current Chemistry; Springer: Berlin, Heidelberg, 2012; pp 1–32.

[14] 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* **2016**, *15*, 605–619.

[15] Anderson, A. C. The Process of Structure-Based Drug Design. *Chemistry & Biology* **2003**, *10*, 787–797.

[16] Hubbard, R. E. *Structure-Based Drug Discovery*; 2006.

[17] Tari, L. W., Ed. *Structure-Based Drug Discovery*; Methods in Molecular Biology; Humana Press, 2012.

[18] van Montfort, R. L. M.; Workman, P. Structure-Based Drug Design: Aiming for a Perfect Fit. *Essays In Biochemistry* **2017**, *61*, 431–437.

[19] Keserü, G. M.; Makara, G. M. The Influence of Lead Discovery Strategies on the Properties of Drug Candidates. *Nature Reviews Drug Discovery* **2009**, *8*, 203–212.

[20] Arrowsmith, C. H. *et al.* The Promise and Peril of Chemical Probes. *Nature Chemical Biology* **2015**, *11*, 536–541.

[21] Murray, C. W.; Verdonk, M. L.; Rees, D. C. Experiences in Fragment-Based Drug Discovery. *Trends in Pharmacological Sciences* **2012**, *33*, 224–232.

[22] Jacquemard, C.; Kellenberger, E. A Bright Future for Fragment-Based Drug Discovery: What Does It Hold? *Expert Opinion on Drug Discovery* **2019**, *14*, 413–416.

[23] Coyle, J.; Walser, R. Applied Biophysical Methods in Fragment-Based Drug Discovery. *SLAS DISCOVERY: Advancing the Science of Drug Discovery* **2020**, 2472555220916168.

[24] Alex, A. A.; Flocco, M. M. Fragment-Based Drug Discovery: What Has It Achieved so Far? *Current Topics in Medicinal Chemistry* **2007**, *7*, 1544–1567.

[25] Chessari, G.; Woodhead, A. J. From Fragment to Clinical Candidate - a Historical Perspective. *Drug Discovery Today* **2009**, *14*, 668–675.

[26] Davis, B. J.; Erlanson, D. A. Learning from Our Mistakes: The 'unknown Knowns' in Fragment Screening. *Bioorganic & Medicinal Chemistry Letters* **2013**, *23*, 2844–2852.

[27] Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-Based Lead Discovery. *Nature Reviews Drug Discovery* **2004**, *3*, 660–672.

[28] Hann, M. M.; Leach, A. R.; Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *Journal of Chemical Information and Computer Sciences* **2001**, *41*, 856–864.

[29] Fink, T.; Reymond, J.-L. Virtual Exploration of the Chemical Universe up to 11 Atoms of C, N, O, F: Assembly of 26.4 Million Structures (110.9 Million Stereoisomers) and Analysis for New Ring Systems, Stereochemistry, Physicochemical Properties, Compound Classes, and Drug Discovery. *Journal of Chemical Information and Modeling* **2007**, *47*, 342–353.

[30] Nilar, S. H.; Ma, N. L.; Keller, T. H. The Importance of Molecular Complexity in the Design of Screening Libraries. *Journal of Computer-Aided Molecular Design* **2013**, *27*, 783–792.

[31] Giordanetto, F.; Jin, C.; Willmore, L.; Feher, M.; Shaw, D. E. Fragment Hits: What Do They Look Like and How Do They Bind? *Journal of Medicinal Chemistry* **2019**, *62*, 3381–3394.

[32] Jhoti, H.; Cleasby, A.; Verdonk, M.; Williams, G. Fragment-Based Screening Using X-Ray Crystallography and NMR Spectroscopy. *Current Opinion in Chemical Biology* **2007**, *11*, 485–493.

[33] Barelier, S.; Pons, J.; Gehring, K.; Lancelin, J.-M.; Krimm, I. Ligand Specificity in Fragment-Based Drug Design. *Journal of Medicinal Chemistry* **2010**, *53*, 5256–5266.

[34] Mashalidis, E. H.; Sledz, P.; Lang, S.; Abell, C. A Three-Stage Biophysical Screening Cascade for Fragment-Based Drug Discovery. *Nature Protocols* **2013**, *8*, 2309–2324.

[35] Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. Fragment-Based Lead Discovery Using X-Ray Crystallography. *Journal of Medicinal Chemistry* **2005**, *48*, 403–413.

[36] Scott, D. E.; Coyne, A. G.; Hudson, S. A.; Abell, C. Fragment-Based Approaches in Drug Discovery and Chemical Biology. *Biochemistry* **2012**, *51*, 4990–5003.

[37] Chodera, J. D.; Mobley, D. L. Entropy-Enthalpy Compensation: Role and Ramifications in Biomolecular Ligand Recognition and Design. *Annual Review of Biophysics* **2013**, *42*, 121–142.

[38] Ferenczy, G. G.; Keseru, G. M. On the Enthalpic Preference of Fragment Binding. *MedChemComm* **2016**, *7*, 332–337.

[39] Price, A. J.; Howard, S.; Cons, B. D. Fragment-Based Drug Discovery and Its Application to Challenging Drug Targets. *Essays In Biochemistry* **2017**, *61*, 475–484.

[40] Canning, P.; Birchall, K.; Kettleborough, C. A.; Merritt, A.; Coombs, P. J. Fragment-Based Target Screening as an Empirical Approach to Prioritising Targets: A Case Study on Antibacterials. *Drug Discovery Today* **2020**,

[41] Ferenczy, G. G.; Keseru, G. M. How Are Fragments Optimized? A Retrospective Analysis of 145 Fragment Optimizations. *Journal of Medicinal Chemistry* **2013**, *56*, 2478–2486.

[42] van Montfort, R. L. M.; Workman, P.; Lamoree, B.; Hubbard, R. E. Current Perspectives in Fragment-Based Lead Discovery (FBLD). *Essays in Biochemistry* **2017**, *61*, 453–464.

[43] Murray, C. W.; Rees, D. C. The Rise of Fragment-Based Drug Discovery. *Nature Chemistry* **2009**, *1*, 187–192.

[44] Keseru, G. M.; Erlanson, D. A.; Ferenczy, G. G.; Hann, M. M.; Murray, C. W.; Pickett, S. D. Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *Journal of Medicinal Chemistry* **2016**, *59*, 8189–8206.

[45] Keseru, G. M.; Hann, M. M. What Is the Future for Fragment-Based Drug Discovery? *Future Medicinal Chemistry* **2017**, *9*, 1457–1460.

[46] Murray, C. W.; Rees, D. C. Opportunity Knocks: Organic Chemistry for Fragment-Based Drug Discovery (FBDD). *Angewandte Chemie International Edition* **2016**, *55*, 488–492.

[47] Doak, B. C.; Norton, R. S.; Scanlon, M. J. The Ways and Means of Fragment-Based Drug Design. *Pharmacology & Therapeutics* **2016**, *167*, 28–37.

[48] Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534.

[49] Chen, I.-J.; Hubbard, R. E. Lessons for Fragment Library Design: Analysis of Output from Multiple Screening Campaigns. *Journal of Computer-Aided Molecular Design* **2009**, *23*, 603–620.

[50] Hall, R. J.; Mortenson, P. N.; Murray, C. W. Efficient Exploration of Chemical Space by Fragment-Based Screening. *Progress in Biophysics and Molecular Biology* **2014**, *116*, 82–91.

[51] Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'Rule of Three' for Fragment-Based Lead Discovery? *Drug Discovery Today* **2003**, *8*, 876–877.

[52] Carr, R. A. E.; Congreve, M.; Murray, C. W.; Rees, D. C. Fragment-Based Lead Discovery: Leads by Design. *Drug Discovery Today* **2005**, *10*, 987–992.

[53] Baurin, N.; Aboul-Ela, F.; Barril, X.; Davis, B.; Drysdale, M.; Dymock, B.; Finch, H.; Fromont, C.; Richardson, C.; Simmonite, H.; Hubbard, R. E. Design and Characterization of Libraries of Molecular Fragments for Use in NMR Screening against Protein Targets. *Journal of Chemical Information and Computer Sciences* **2004**, *44*, 2157–2166.

[54] Siegal, G.; Ab, E.; Schultz, J. Integration of Fragment Screening and Library Design. *Drug Discovery Today* **2007**, *12*, 1032–1039.

[55] Boyd, S. M.; de Kloe, G. E. Fragment Library Design: Efficiently Hunting Drugs in Chemical Space. *Drug Discovery Today: Technologies* **2010**, *7*, 173–180.

[56] Schulz, M. N.; Landström, J.; Bright, K.; Hubbard, R. E. Design of a Fragment Library That Maximally Represents Available Chemical Space. *Journal of Computer-Aided Molecular Design* **2011**, *25*, 611.

[57] *Fragment-Based Drug Discovery Lessons and Outlook*; John Wiley & Sons, Ltd, 2016; pp 487–500.

[58] Ray, P. C.; Kiczun, M.; Huggett, M.; Lim, A.; Prati, F.; Gilbert, I. H.; Wyatt, P. G. Fragment Library Design, Synthesis and Expansion: Nurturing a Synthesis and Training Platform. *Drug Discovery Today* **2017**, *22*, 43–56.

[59] 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*, 694–706.e5.

[60] Zartler, E. R.; Shapiro, M. J. Fragonomics: Fragment-Based Drug Discovery. *Current Opinion in Chemical Biology* **2005**, *9*, 366–370.

[61] Bruns, R. F.; Watson, I. A. Rules for Identifying Potentially Reactive or Promiscuous Compounds. *Journal of Medicinal Chemistry* **2012**, *55*, 9763–9772.

[62] 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 Discovery Today* **2019**, *24*, 1081–1086.

[63] Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *Journal of Medicinal Chemistry* **2010**, *53*, 2719–2740.

[64] Seidler, J.; McGovern, S. L.; Doman, T. N.; Shoichet, B. K. Identification and Prediction of Promiscuous Aggregating Inhibitors among Known Drugs. *Journal of Medicinal Chemistry* **2003**, *46*, 4477–4486.

[65] Feng, B. Y.; Shelat, A.; Doman, T. N.; Guy, R. K.; Shoichet, B. K. High-Throughput Assays for Promiscuous Inhibitors. *Nature Chemical Biology* **2005**, *1*, 146–148.

[66] Hung, A. W.; Ramek, A.; Wang, Y.; Kaya, T.; Wilson, J. A.; Clemons, P. A.; Young, D. W. Route to Three-Dimensional Fragments Using Diversity-Oriented Synthesis. *Proceedings of the National Academy of Sciences* **2011**, *108*, 6799–6804.

[67] Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *Journal of Medicinal Chemistry* **2009**, *52*, 6752–6756.

[68] Barker, A.; Kettle, J. G.; Nowak, T.; Pease, J. E. Expanding Medicinal Chemistry Space. *Drug Discovery Today* **2013**, *18*, 298–304.

[69] Taylor, R. D.; MacCoss, M.; Lawson, A. D. G. Rings in Drugs. *Journal of Medicinal Chemistry* **2014**, *57*, 5845–5859.

[70] Kidd, S. L.; Osberger, T. J.; Mateu, N.; Sore, H. F.; Spring, D. R. Recent Applications of Diversity-Oriented Synthesis Toward Novel, 3-Dimensional Fragment Collections. *Frontiers in Chemistry* **2018**, *6*.

[71] 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. *Chemical Science* **2016**, *7*, 2322–2330.

[72] DSI-Poised Library - Enamine. https://enamine.net/fragments/plated-libraries/dsi-poised-library, accessed on 16.09.2020.

[73] Roughley, S. D.; Jordan, A. M. The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. *Journal of Medicinal Chemistry* **2011**, *54*, 3451–3479.

[74] Hartenfeller, M.; Eberle, M.; Meier, P.; Nieto-Oberhuber, C.; Altmann, K.-H.; Schneider, G.; Jacoby, E.; Renner, S. A Collection of Robust Organic Synthesis Reactions for In Silico Molecule Design. *Journal of Chemical Information and Modeling* **2011**, *51*, 3093–3098.

[75] Vulpetti, A.; Dalvit, C. Design and Generation of Highly Diverse Fluorinated Fragment Libraries and Their Efficient Screening with Improved 19F NMR Methodology. *ChemMedChem* **2013**, *8*, 2057–2069.

[76] 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. *Journal of Medicinal Chemistry* **2019**, *62*, 3741–3752.

[77] Draxler, S. W.; Bauer, M.; Eickmeier, C.; Nadal, S.; Nar, H.; Rangel Rojas, D.; Seeliger, D.; Zeeb, M.; Fiegen, D. Hybrid Screening Approach for Very Small Fragments: X-Ray and Computational Screening on FKBP51. *Journal of Medicinal Chemistry* **2020**, *63*, 5856–5864.

[78] Blundell, T. L.; Jhoti, H.; Abell, C. High-Throughput Crystallography for Lead Discovery in Drug Design. *Nature Reviews Drug Discovery* **2002**, *1*, 45–54.

[79] Patel, D.; Bauman, J. D.; Arnold, E. Advantages of Crystallographic Fragment Screening: Functional and Mechanistic Insights from a Powerful Platform for Efficient Drug Discovery. *Progress in Biophysics and Molecular Biology* **2014**, *116*, 92–100.

[80] Danielson, U. H. Fragment Library Screening and Lead Characterization Using SPR Biosensors. *Current Topics in Medicinal Chemistry* **2009**, 1725–1735.

[81] Navratilova, I.; Hopkins, A. L. Fragment Screening by Surface Plasmon Resonance. *ACS Medicinal Chemistry Letters* **2010**, *1*, 44–48.

[82] Niesen, F. H.; Berglund, H.; Vedadi, M. The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions That Promote Protein Stability. *Nature Protocols* **2007**, *2*, 2212–2221.

[83] Maple, H. J.; Garlish, R. A.; Rigau-Roca, L.; Porter, J.; Whitcombe, I.; Prosser, C. E.; Kennedy, J.; Henry, A. J.; Taylor, R. J.; Crump, M. P.; Crosby, J. Automated Protein Ligand Interaction Screening by Mass Spectrometry. *Journal of Medicinal Chemistry* **2012**, *55*, 837–851.

[84] Linke, P.; Amaning, K.; Maschberger, M.; Vallee, F.; Steier, V.; Baaske, P.; Duhr, S.; Breitsprecher, D.; Rak, A. An Automated Microscale Thermophoresis Screening Approach for Fragment-Based Lead Discovery:. *Journal of Biomolecular Screening* **2015**,

[85] Meiby, E.; Simmonite, H.; le Strat, L.; Davis, B.; Matassova, N.; Moore, J. D.; Mrosek, M.; Murray, J.; Hubbard, R. E.; Ohlson, S. Fragment Screening by Weak Affinity Chromatography: Comparison with Established Techniques for Screening against HSP90. *Analytical Chemistry* **2013**, *85*, 6756–6766.

[86] Renaud, J.-P.; Chung, C.-w.; Danielson, U. H.; Egner, U.; Hennig, M.; Hubbard, R. E.; Nar, H. Biophysics in Drug Discovery: Impact, Challenges and Opportunities. *Nature Reviews Drug Discovery* **2016**, *15*, 679–698.

[87] Hubbard, R. E.; Davis, B.; Chen, I.; Drysdale, M. J. The SeeDs Approach: Integrating Fragments into Drug Discovery. *Current Topics in Medicinal Chemistry* **2007**, *7*, 1568–1581.

[88] Badger, J. In *Structure-Based Drug Discovery*; Tari, L. W., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2012; pp 161–177.

[89] 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 Chemical Biology* **2016**, *11*, 1693–1701.

[90] Wielens, J.; Headey, S. J.; Rhodes, D. I.; Mulder, R. J.; Dolezal, O.; Deadman, J. J.; Newman, J.; Chalmers, D. K.; Parker, M. W.; Peat, T. S.; Scanlon, M. J. Parallel Screening of Low Molecular Weight Fragment Libraries: Do Differences in Methodology Affect Hit Identification? *Journal of Biomolecular Screening* **2013**, *18*, 147–159.

[91] Schiebel, J. *et al.* One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists. *ChemMedChem* **2015**, *10*, 1511–1521.

[92] Hassaan, E.; Eriksson, P.-O.; Geschwindner, S.; Heine, A.; Klebe, G. Fragments as Novel Starting Points for tRNA-Guanine Transglycosylase Inhibitors Found by Alternative Screening Strategies. *ChemMedChem* **2020**, *15*, 324–337.

[93] Kutchukian, P. S.; Wassermann, A. M.; Lindvall, M. K.; Wright, S. K.; Ottl, J.; Jacob, J.; Scheufler, C.; Marzinzik, A.; Brooijmans, N.; Glick, M. Large Scale Meta-Analysis of Fragment-Based Screening Campaigns: Privileged Fragments and Complementary Technologies. *Journal of Biomolecular Screening* **2015**, *20*, 588–596.

[94] Duhr, S.; Braun, D. Why Molecules Move along a Temperature Gradient. *Proceedings of the National Academy of Sciences* **2006**, *103*, 19678.

[95] Silvestre, H. L.; Blundell, T. L.; Abell, C.; Ciulli, A. Integrated Biophysical Approach to Fragment Screening and Validation for Fragment-Based Lead Discovery. *Proceedings of the National Academy of Sciences* **2013**, *110*, 12984–12989.

[96] Bowling, J. J.; Shadrick, W. R.; Griffith, E. C.; Lee, R. E. Going Small: Using Biophysical Screening to Implement Fragment Based Drug Discovery. *Special Topics in Drug Discovery* **2016**,

[97] Keiffer, S.; Carneiro, M. G.; Hollander, J.; Kobayashi, M.; Pogoryelev, D.; AB, E.; Theisgen, S.; Müller, G.; Siegal, G. NMR in Target Driven Drug Discovery: Why Not? *Journal of Biomolecular NMR* **2020**, *74*, 521–529.

[98] Verlinde, C. L. *et al.* Fragment-Based Cocktail Crystallography by the Medical Structural Genomics of Pathogenic Protozoa Consortium. *Current topics in medicinal chemistry* **2009**, *9*, 1678–1687.

[99] Douangamath, A. *et al.* Crystallographic and Electrophilic Fragment Screening of the SARS-CoV-2 Main Protease. *Nature Communications* **2020**, *11*, 5047.

[100] Tickle, I.; Sharff, A.; Vinkovic, M.; Yon, J.; Jhoti, H. High-Throughput Protein Crystallography and Drug Discovery. *Chemical Society Reviews* **2004**, *33*, 558–565.

[101] Joachimiak, A. High-Throughput Crystallography for Structural Genomics. *Current Opinion in Structural Biology* **2009**, *19*, 573–584.

[102] Collins, P. M.; Douangamath, A.; Talon, R.; Dias, A.; Brandao-Neto, J.; Krojer, T.; von Delft, F. In *Methods in Enzymology*; Lesburg, C. A., Ed.; Modern Approaches in Drug Discovery; Academic Press, 2018; Vol. 610; pp 251–264.

[103] 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. Where Is Crystallography Going? *Acta Crystallographica Section D: Structural Biology* **2018**, *74*, 152–166.

[104] Nienaber, V. L.; Richardson, P. L.; Klighofer, V.; Bouska, J. J.; Giranda, V. L.; Greer, J. Discovering Novel Ligands for Macromolecules Using X-Ray Crystallographic Screening. *Nature Biotechnology* **2000**, *18*, 1105–1108.

[105] Mooij, W. T. M.; Hartshorn, M. J.; Tickle, I. J.; Sharff, A. J.; Verdonk, M. L.; Jhoti, H. Automated Protein-Ligand Crystallography for Structure-Based Drug Design. *ChemMedChem* **2006**, *1*, 827–838.

[106] Murray, C. W.; Blundell, T. L. Structural Biology in Fragment-Based Drug Design. *Current Opinion in Structural Biology* **2010**, *20*, 497–507.

[107] Ludlow, R. F.; Verdonk, M. L.; Saini, H. K.; Tickle, I. J.; Jhoti, H. Detection of Secondary Binding Sites in Proteins Using Fragment Screening. *Proceedings of the National Academy of Sciences* **2015**, *112*, 15910–15915.

[108] Zander, U. *et al.* Automated Harvesting and Processing of Protein Crystals through Laser Photoablation. *Acta Crystallographica Section D: Structural Biology* **2016**, *72*, 454–466.

[109] 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 Crystallographica Section D: Structural Biology* **2017**, *73*, 246–255.

[110] Chilingaryan, Z.; Yin, Z.; Oakley, A. J. Fragment-Based Screening by Protein Crystallography: Successes and Pitfalls. *International Journal of Molecular Sciences* **2012**, *13*, 12857–12879.

[111] Owen, R. L.; Juanhuix, J.; Fuchs, M. Current Advances in Synchrotron Radiation Instrumentation for MX Experiments. *Archives of Biochemistry and Biophysics* **2016**, *602*, 21–31.

[112] 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 (REFiLX). *Journal of Medicinal Chemistry* **2020**, *63*, 6863–6875.

[113] 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. *Nature Communications* **2017**, *8*, 1–8.

[114] Maveyraud, L.; Mourey, L. Protein X-Ray Crystallography and Drug Discovery. *Molecules* **2020**, *25*, 1030.

[115] Shaw Stewart, P.; Mueller-Dieckmann, J. Automation in Biological Crystallization. *Acta Crystallographica Section F: Structural Biology Communications* **2014**, *70*, 686–696.

[116] Hassell, A. M. *et al.* Crystallization of Protein-Ligand Complexes. *Acta Crystallographica Section D: Biological Crystallography* **2007**, *63*, 72–79.

[117] Carvalho, A. L.; Trincão, J.; Romão, M. J. In *Ligand-Macromolecular Interactions in Drug Discovery: Methods and Protocols*; Roque, A. C. A., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2010; pp 31–56.

[118] Muller, I. Guidelines for the Successful Generation of Protein-Ligand Complex Crystals. *Acta Crystallographica. Section D, Structural Biology* **2017**, *73*, 79–92.

[119] Ehrmann, F. R.; Stojko, J.; Metz, A.; Debaene, F.; Barandun, L. J.; Heine, A.; Diederich, F.; Cianférani, S.; Reuter, K.; Klebe, G. Soaking Suggests "Alternative Facts": Only Co-Crystallization Discloses Major Ligand-Induced Interface Rearrangements of a Homodimeric tRNA-Binding Protein Indicating a Novel Mode-of-Inhibition. *PLoS ONE* **2017**, *12*.

[120] Osborne, J.; Panova, S.; Rapti, M.; Urushima, T.; Jhoti, H. Fragments: Where Are We Now? *Biochemical Society Transactions* **2020**, *48*, 271–280.

[121] Bradley, A. R.; Echalier, A.; Fairhead, M.; Strain-Damerell, C.; Brennan, P.; Bullock, A. N.; Burgess-Brown, N. A.; Carpenter, E. P.; Gileadi, O.; Marsden, B. D.; Lee, W. H.; Yue, W.; Bountra, C.; von Delft, F. The SGC beyond Structural Genomics: Redefining the Role of 3D Structures by Coupling Genomic Stratification with Fragment-Based Discovery. *Essays In Biochemistry* **2017**, *61*, 495–503.

[122] Robson-Tull, J. Biophysical Screening in Fragment-Based Drug Design: A Brief Overview. *Bioscience Horizons: The International Journal of Student Research* **2018**, *11*.

[123] Ng, J. T.; Dekker, C.; Kroemer, M.; Osborne, M.; von Delft, F. Using Textons to Rank Crystallization Droplets by the Likely Presence of Crystals. *Acta Crystallographica Section D: Biological Crystallography* **2014**, *70*, 2702–2718.

[124] 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. The Low-Cost Shifter Microscope Stage Transforms the Speed and Robustness of Protein Crystal Harvesting. *Acta Crystallographica Section D: Structural Biology* **2021**, *77*, 62–74.

[125] Winter, G.; McAuley, K. E. Automated Data Collection for Macromolecular Crystallography. *Methods* **2011**, *55*, 81–93.

[126] 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 Crystallographica Section D: Structural Biology* **2017**, *73*, 267–278.

[127] Diamond Light Source, The XChem Pipeline - Diamond Light Source. https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening/New-The-XChem-Pipeline.html, accessed on 22.06.2020.

[128] Winter, G. Xia2: An Expert System for Macromolecular Crystallography Data Reduction. *Journal of Applied Crystallography* **2010**, *43*, 186–190.

[129] Schiebel, J. *et al.* High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure* **2016**, *24*, 1398–1409.

[130] Pearce, N. M.; Bradley, A. R.; Krojer, T.; Marsden, B. D.; Deane, C. M.; von Delft, F. Partial-Occupancy Binders Identified by the Pan-Dataset Density Analysis Method Offer New Chemical Opportunities and Reveal Cryptic Binding Sites. *Structural Dynamics* **2017**, *4*.

[131] Nelson, E. Multiple Crystallographic Dataset Analysis to Explore Crystallographic Occupancy and the Conformational States of Proteins. http://purl.org/dc/dcmitype/Text, University of Oxford, 2019.

[132] 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* **2017**, *73*, 256–266.

[133] McIntyre, P. J.; Collins, P. M.; Vrzal, L.; Birchall, K.; Arnold, L. H.; Mpamhanga, C.; Coombs, P. J.; Burgess, S. G.; Richards, M. W.; Winter, A.; Veverka, V.; von Delft, F.; Merritt, A.; Bayliss, R. Characterization of Three Druggable Hot-Spots in the Aurora-A/TPX2 Interaction Using Biochemical, Biophysical, and Fragment-Based Approaches. *ACS Chemical Biology* **2017**, *12*, 2906–2914.

[134] Keedy, D. A.; Hill, Z. B.; Biel, J. T.; Kang, E.; Rettenmaier, T. J.; Brandão-Neto, J.; Pearce, N. M.; von Delft, F.; Wells, J. A.; Fraser, J. S. An Expanded Allosteric Network in PTP1B by Multitemperature Crystallography, Fragment Screening, and Covalent Tethering. *eLife* **2018**, *7*, e36307.

[135] 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. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2019**, *377*, 20180422.

[136] Wlodawer, A.; Dauter, Z.; Shabalin, I.; Gilski, M.; Brzezinski, D.; Kowiel, M.; Minor, W.; Rupp, B.; Jaskolski, M. Ligand-Centered Assessment of SARS-CoV-2 Drug Target Models in the Protein Data Bank. *The FEBS Journal* **2020**, *n/a*.

[137] Jahnke, W. Perspectives of Biomolecular NMR in Drug Discovery: The Blessing and Curse of Versatility. *Journal of Biomolecular NMR* **2007**, *39*, 87–90.

[138] Campos-Olivas, R. NMR Screening and Hit Validation in Fragment Based Drug Discovery. *Current Topics in Medicinal Chemistry* **2011**, *11*, 43–67.

[139] Bentley, M.; Doak, B. C.; Mohanty, B.; Scanlon, M. J. In *Modern Magnetic Resonance*; Webb, G. A., Ed.; Springer International Publishing: Cham, 2017; pp 1–22.

[140] Horst, R.; Farley, K. A.; Kormos, B. L.; Withka, J. M. NMR Spectroscopy: The Swiss Army Knife of Drug Discovery. *Journal of Biomolecular NMR* **2020**, *74*, 509–519.

[141] Gossert, A. D.; Jahnke, W. NMR in Drug Discovery: A Practical Guide to Identification and Validation of Ligands Interacting with Biological Macromolecules. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2016**, *97*, 82–125.

[142] Sugiki, T.; Furuita, K.; Fujiwara, T.; Kojima, C. Current NMR Techniques for Structure-Based Drug Discovery. *Molecules* **2018**, *23*, 148.

[143] Dalvit, C. NMR Methods in Fragment Screening: Theory and a Comparison with Other Biophysical Techniques. *Drug Discovery Today* **2009**, *14*, 1051–1057.

[144] Craik, D. J.; Scanlon, M. J. *Annual Reports on NMR Spectroscopy*; Academic Press, 2000; Vol. 42; pp 115–174.

[145] Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T. L.; Homans, S. W.; Kessler, H.; Luchinat, C.; Meyer, B.; Oschkinat, H.; Peng, J.; Schwalbe, H.; Siegal, G. Perspectives on NMR in Drug Discovery: A Technique Comes of Age. *Nature Reviews Drug Discovery* **2008**, *7*, 738–745.

[146] Jordan, J. B.; Whittington, D. A.; Bartberger, M. D.; Sickmier, E. A.; Chen, K.; Cheng, Y.; Judd, T. Fragment-Linking Approach Using 19F NMR Spectroscopy To Obtain Highly Potent and Selective Inhibitors of Beta-Secretase. *Journal of Medicinal Chemistry* **2016**, *59*, 3732–3749.

[147] Namanja, A. T.; Xu, J.; Wu, H.; Sun, Q.; Upadhyay, A. K.; Sun, C.; Van Doren, S. R.; Petros, A. M. NMR-based Fragment Screening and Lead Discovery Accelerated by Principal Component Analysis. *Journal of Biomolecular NMR* **2019**, *73*, 675–685.

[148] Claridge, T. D. *High-Resolution NMR Techniques in Organic Chemistry*; Elsevier, 2016.

[149] Lepre, C. A.; Moore, J. M.; Peng, J. W. Theory and Applications of NMR-Based Screening in Pharmaceutical Research. *Chemical Reviews* **2004**, *104*, 3641–3676.

[150] Dalvit, C.; Gossert, A. D.; Coutant, J.; Piotto, M. Rapid Acquisition of 1H and 19F NMR Experiments for Direct and Competition Ligand-Based Screening. *Magnetic Resonance in Chemistry* **2011**, *49*, 199–202.

[151] Kleckner, I. R.; Foster, M. P. An Introduction to NMR-based Approaches for Measuring Protein Dynamics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2011**, *1814*, 942–968.

[152] Viegas, A.; Manso, J.; Nobrega, F. L.; Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *Journal of Chemical Education* **2011**, *88*, 990–994.

[153] Diethelm-Varela, B. Using NMR Spectroscopy in the Fragment-Based Drug Discovery of Small-Molecule Anticancer Targeted Therapies. *ChemMedChem* **2020**, *15*, 1–19.

[154] Aretz, J.; Rademacher, C. Ranking Hits From Saturation Transfer Difference Nuclear Magnetic Resonance-Based Fragment Screening. *Frontiers in Chemistry* **2019**, *7*.

[155] Angulo, J.; Enriquez-Navas, P. M.; Nieto, P. M. Ligand-Receptor Binding Affinities from Saturation Transfer Difference (STD) NMR Spectroscopy: The Binding Isotherm of STD Initial Growth Rates. *Chemistry – A European Journal* **2010**, *16*, 7803–7812.

[156] Mayer, M.; Meyer, B. Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *Journal of the American Chemical Society* **2001**, *123*, 6108–6117.

[157] Wang, Y.-S.; Liu, D.; Wyss, D. F. Competition STD NMR for the Detection of High-Affinity Ligands and NMR-based Screening. *Magnetic Resonance in Chemistry* **2004**, *42*, 485–489.

[158] Szczepina, M. G.; Zheng, R. B.; Completo, G. C.; Lowary, T. L.; Pinto, B. M. STD-NMR Studies Suggest That Two Acceptor Substrates for GlfT2, a Bifunctional Galactofuranosyltransferase Required for the Biosynthesis of Mycobacterium Tuberculosis Arabinogalactan, Compete for the Same Binding Site. *ChemBioChem* **2009**, *10*, 2052–2059.

[159] Dalvit, C.; Parent, A.; Vallée, F.; Mathieu, M.; Rak, A. Fast NMR Methods for Measuring in the Direct and/or Competition Mode the Dissociation Constants of Chemical Fragments Interacting with a Receptor. *ChemMedChem* **2019**, *14*, 1115–1127.

[160] Deverell, C.; Morgan, R. E.; Strange, J. H. Studies of Chemical Exchange by Nuclear Magnetic Relaxation in the Rotating Frame. *Molecular Physics* **1970**, *18*, 553–559.

[161] Hajduk, P. J.; Olejniczak, E. T.; Fesik, S. W. One-Dimensional Relaxation- and Diffusion-Edited NMR Methods for Screening Compounds That Bind to Macromolecules. *Journal of the American Chemical Society* **1997**, *119*, 12257–12261.

[162] Jahnke, W.; Widmer, H. Protein NMR in Biomedical Research. *Cellular and Molecular Life Sciences CMLS* **2004**, *61*, 580–599.

[163] Peng, C.; Frommlet, A.; Perez, M.; Cobas, C.; Blechschmidt, A.; Dominguez, S.; Lingel, A. Fast and Efficient Fragment-Based Lead Generation by Fully Automated Processing and Analysis of Ligand-Observed NMR Binding Data. *Journal of Medicinal Chemistry* **2016**, *59*, 3303–3310.

[164] Norton, R. S.; Leung, E. W. W.; Chandrashekaran, I. R.; MacRaild, C. A. Applications of 19F-NMR in Fragment-Based Drug Discovery. *Molecules* **2016**, *21*, 860.

[165] Dalvit, C.; Vulpetti, A. Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *Journal of Medicinal Chemistry* **2019**, *62*, 2218–2244.

[166] Howe, P. W. A. Recent Developments in the Use of Fluorine NMR in Synthesis and Characterisation. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2020**, *118–119*, 1–9.

[167] Vulpetti, A.; Dalvit, C. Fluorine Local Environment: From Screening to Drug Design. *Drug Discovery Today* **2012**, *17*, 890–897.

[168] Dalvit, C.; Piotto, M. 19F NMR Transverse and Longitudinal Relaxation Filter Experiments for Screening: A Theoretical and Experimental Analysis. *Magnetic Resonance in Chemistry* **2017**, *55*, 106–114.

[169] Skora, L.; Jahnke, W. 19F-NMR-Based Dual-Site Reporter Assay for the Discovery and Distinction of Catalytic and Allosteric Kinase Inhibitors. *ACS Medicinal Chemistry Letters* **2017**, *8*, 632–635.

[170] Bruker, Fragment Based Screening | Bruker NMR Systems with TopSpin Include a Fragment Based Screening Tool for the Acquisition, Analysis, and Reporting of Screening Data. https://www.bruker.com/products/mr/nmr/software/fragment-based-screening-with-nmr.html, accessed on 22.06.2020.

[171] Mestrelab, Mnova Screen: Automatic Analysis Tool for Fragment-Based Drug Screening by NMR. https://mestrelab.com/software/mnova/screen/, accessed on 22.06.2020.

[172] Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. 19 F NMR: A Valuable Tool for Studying Biological Events. *Chemical Society Reviews* **2013**, *42*, 7971–7982.

[173] de Castro, G. V.; Ciulli, A. Spy vs. Spy: Selecting the Best Reporter for 19F NMR Competition Experiments. *Chemical Communications* **2019**, *55*, 1482–1485.

[174] Cooper, M. A. Optical Biosensors in Drug Discovery. *Nature Reviews Drug Discovery* **2002**, *1*, 515–528.

[175] Giannetti, A. M. In *Methods in Enzymology*; Kuo, L. C., Ed.; Fragment-Based Drug DesignTools, Practical Approaches, and Examples; Academic Press, 2011; Vol. 493; pp 169–218.

[176] Navratilova, I.; Aristotelous, T.; Picaud, S.; Chaikuad, A.; Knapp, S.; Filappakopoulos, P.; Hopkins, A. L. Discovery of New Bromodomain Scaffolds by Biosensor Fragment Screening. *ACS Medicinal Chemistry Letters* **2016**, *7*, 1213–1218.

[177] Perspicace, S.; Banner, D.; Benz, J.; Müller, F.; Schlatter, D.; Huber, W. Fragment-Based Screening Using Surface Plasmon Resonance Technology. *Journal of Biomolecular Screening* **2009**, *14*, 337–349.

[178] Navratilova, I.; Hopkins, A. L. Emerging Role of Surface Plasmon Resonance in Fragment-Based Drug Discovery. *Future Medicinal Chemistry* **2011**, *3*, 1809–1820.

[179] Shepherd, C. A.; Hopkins, A. L.; Navratilova, I. Fragment Screening by SPR and Advanced Application to GPCRs. *Progress in Biophysics and Molecular Biology* **2014**, *116*, 113–123.

[180] Opassi, G.; Nordstrom, H.; Lundin, A.; Napolitano, V.; Magari, F.; Dzus, T.; Klebe, G.; Danielson, U. H. Establishing Trypanosoma Cruzi Farnesyl Pyrophosphate Synthase as a Viable Target for Biosensor Driven Fragment-Based Lead Discovery. *Protein Science* **2020**, *29*, 991–1003.

[181] Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-Based Drug Discovery. *Journal of Medicinal Chemistry* **2004**, *47*, 3463–3482.

[182] de Esch, I. J. P. The HTS Dogma Is Shattered, and Now There Are Fragments Everywhere. *Drug Discovery Today: Technologies* **2010**, *7*, 147–148.

[183] Warr, W. A. Fragment-Based Drug Discovery. *Journal of Computer-Aided Molecular Design* **2009**, *23*, 453–458.

[184] Schulz, M. N.; Hubbard, R. E. Recent Progress in Fragment-Based Lead Discovery. *Current Opinion in Pharmacology* **2009**, *9*, 615–621.

[185] Hajduk, P. J.; Greer, J. A Decade of Fragment-Based Drug Design: Strategic Advances and Lessons Learned. *Nature Reviews Drug Discovery* **2007**, *6*, 211–219.

[186] Jencks, W. P. On the Attribution and Additivity of Binding Energies. *Proceedings of the National Academy of Sciences* **1981**, *78*, 4046–4050.

[187] Ichihara, O.; Barker, J.; Law, R. J.; Whittaker, M. Compound Design by Fragment-Linking. *Molecular Informatics* **2011**, *30*, 298–306.

[188] Foloppe, N. The Benefits of Constructing Leads from Fragment Hits. *Future Medicinal Chemistry* **2011**, *3*, 1111–1115.

[189] Kirsch, P.; Hartman, A. M.; Hirsch, A. K. H.; Empting, M. Concepts and Core Principles of Fragment-Based Drug Design. *Molecules* **2019**, *24*, 4309.

[190] Souers, A. J. *et al.* ABT-199, a Potent and Selective BCL-2 Inhibitor, Achieves Antitumor Activity While Sparing Platelets. *Nature Medicine* **2013**, *19*, 202–208.

[191] Bollag, G.; Tsai, J.; Zhang, J.; Zhang, C.; Ibrahim, P.; Nolop, K.; Hirth, P. Vemurafenib: The First Drug Approved for BRAF -Mutant Cancer. *Nature Reviews Drug Discovery* **2012**, *11*, 873–886.

[192] Hubbard, R. E. Fragment Approaches in Structure-Based Drug Discovery. *Journal of Synchrotron Radiation* **2008**, *15*, 227–230.

[193] Erlanson, D. A.; Davis, B. J.; Jahnke, W. Fragment-Based Drug Discovery: Advancing Fragments in the Absence of Crystal Structures. *Cell Chemical Biology* **2019**, *26*, 9–15.

[194] Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand Efficiency: A Useful Metric for Lead Selection. *Drug Discovery Today* **2004**, *9*, 430–431.

[195] Schultes, S.; de Graaf, C.; Haaksma, E. E. J.; de Esch, I. J. P.; Leurs, R.; Krämer, O. Ligand Efficiency as a Guide in Fragment Hit Selection and Optimization. *Drug Discovery Today: Technologies* **2010**, *7*, e157–e162.

[196] Murray, C. W.; Erlanson, D. A.; Hopkins, A. L.; Keserü, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H.; Richmond, N. J. Validity of Ligand Efficiency Metrics. *ACS Medicinal Chemistry Letters* **2014**, *5*, 616–618.

[197] Johnson, C. N.; Erlanson, D. A.; Murray, C. W.; Rees, D. C. Fragment-to-Lead Medicinal Chemistry Publications in 2015. *Journal of Medicinal Chemistry* **2017**, *60*, 89–99.

[198] Bessman, M. J.; Frick, D. N.; O'Handley, S. F. The MutT Proteins or "Nudix" Hydrolases, a Family of Versatile, Widely Distributed, "Housecleaning" Enzymes. *Journal of Biological Chemistry* **1996**, *271*, 25059–25062.

[199] Zha, M.; Zhong, C.; Peng, Y.; Hu, H.; Ding, J. Crystal Structures of Human NUDT5 Reveal Insights into the Structural Basis of the Substrate Specificity. *Journal of Molecular Biology* **2006**, *364*, 1021–1033.

[200] Page, B. D. G. *et al.* Targeted NUDT5 Inhibitors Block Hormone Signaling in Breast Cancer Cells. *Nature Communications* **2018**, *9*, 250.

[201] Mildvan, A. S.; Xia, Z.; Azurmendi, H. F.; Saraswat, V.; Legler, P. M.; Massiah, M. A.; Gabelli, S. B.; Bianchet, M. A.; Kang, L. W.; Amzel, L. M. Structures and Mechanisms of Nudix Hydrolases. *Archives of Biochemistry and Biophysics* **2005**, *433*, 129–143.

[202] McLennan, A. G. The Nudix Hydrolase Superfamily. *Cellular and Molecular Life Sciences CMLS* **2006**, *63*, 123–143.

[203] Sakumi, K.; Furuichi, M.; Tsuzuki, T.; Kakuma, T.; Kawabata, S.; Maki, H.; Sekiguchi, M. Cloning and Expression of cDNA for a Human Enzyme That Hydrolyzes 8-Oxo-dGTP, a Mutagenic Substrate for DNA Synthesis. *Journal of Biological Chemistry* **1993**, *268*, 23524–23530.

[204] Fujikawa, K.; Kamiya, H.; Yakushiji, H.; Fujii, Y.; Nakabeppu, Y.; Kasai, H. The Oxidized Forms of dATP Are Substrates for the Human MutT Homologue, the hMTH1 Protein. *Journal of Biological Chemistry* **1999**, *274*, 18201–18205.

[205] Rudd, S. G.; Valerie, N. C. K.; Helleday, T. Pathways Controlling dNTP Pools to Maintain Genome Stability. *DNA Repair* **2016**, *44*, 193–204.

[206] Carreras-Puigvert, J. *et al.* A Comprehensive Structural, Biochemical and Biological Profiling of the Human NUDIX Hydrolase Family. *Nature Communications* **2017**, *8*, 1541.

[207] Okamoto, K.; Toyokuni, S.; Kim, W.-J.; Ogawa, O.; Kakehi, Y.; Arao, S.; Hiai, H.; Yoshida, O. Overexpression of Human mutT Homologue Gene Messenger RNA in Renal-Cell Carcinoma: Evidence of Persistent Oxidative Stress in Cancer. *International Journal of Cancer* **1996**, *65*, 437–441.

[208] Kennedy, C. H.; Cueto, R.; Belinsky, S. A.; Lechner, J. F.; Pryor, W. A. Overexpression of hMTH1 mRNA: A Molecular Marker of Oxidative Stress in Lung Cancer Cells. *FEBS Letters* **1998**, *429*, 17–20.

[209] Nakayama, S.; Kajiya, H.; Okabe, K.; Ikebe, T. Effects of Oxidative Stress on the Expression of 8-Oxoguanine and Its Eliminating Enzymes in Human Keratinocytes and Squamous Carcinoma Cells. *Oral Science International* **2011**, *8*, 11–16.

[210] Huber, K. V. M. *et al.* Stereospecific Targeting of MTH1 by (S)-Crizotinib as an Anticancer Strategy. *Nature* **2014**, *508*, 222–227.

[211] Gad, H. *et al.* MTH1 Inhibition Eradicates Cancer by Preventing Sanitation of the dNTP Pool. *Nature* **2014**, *508*, 215–221.

[212] Zhang, L.-Q.; Dai, D.-P.; Gan, W.; Takagi, Y.; Hayakawa, H.; Sekiguchi, M.; Cai, J.-P. Lowered Nudix Type 5 (NUDT5) Expression Leads to Cell Cycle Retardation in HeLa Cells. *Molecular and Cellular Biochemistry* **2012**, *363*, 377–384.

[213] Wright, R. H. G. *et al.* ADP-ribose-derived Nuclear ATP Synthesis by NUDIX5 Is Required for Chromatin Remodeling. *Science* **2016**, *352*, 1221–1225.

[214] Gasmi, L.; Cartwright, J. L.; McLennan, A. G. Cloning, Expression and Characterization of YSA1H, a Human Adenosine 5'-Diphosphosugar Pyrophosphatase Possessing a MutT Motif. *Biochemical Journal* **1999**, *344*, 331–337.

[215] Kamiya, H.; Hori, M.; Arimori, T.; Sekiguchi, M.; Yamagata, Y.; Harashima, H. NUDT5 Hydrolyzes Oxidized Deoxyribonucleoside Diphosphates with Broad Substrate Specificity. *DNA Repair* **2009**, *8*, 1250–1254.

[216] Arimori, T.; Tamaoki, H.; Nakamura, T.; Kamiya, H.; Ikemizu, S.; Takagi, Y.; Ishibashi, T.; Harashima, H.; Sekiguchi, M.; Yamagata, Y. Diverse Substrate Recognition and Hydrolysis Mechanisms of Human NUDT5. *Nucleic Acids Research* **2011**, *39*, 8972–8983.

[217] Ito, R.; Sekiguchi, M.; Setoyama, D.; Nakatsu, Y.; Yamagata, Y.; Hayakawa, H. Cleavage of Oxidized Guanine Nucleotide and ADP Sugar by Human NUDT5 Protein. *The Journal of Biochemistry* **2011**, *149*, 731–738.

[218] Yang, H.; Slupska, M. M.; Wei, Y.-F.; Tai, J. H.; Luther, W. M.; Xia, Y.-R.; Shih, D. M.; Chiang, J.-H.; Baikalov, C.; Fitz-Gibbon, S.; Phan, I. T.; Conrad, A.; Miller, J. H. Cloning and Characterization of a New Member of the Nudix Hydrolases from Human and Mouse. *Journal of Biological Chemistry* **2000**, *275*, 8844–8853.

[219] Jacobson, E. L.; Cervantes-Laurean, D.; Jacobson, M. K. Glycation of Proteins by ADP-ribose. *Molecular and Cellular Biochemistry* **1994**, *138*, 207–212.

[220] Zha, M.; Guo, Q.; Zhang, Y.; Yu, B.; Ou, Y.; Zhong, C.; Ding, J. Molecular Mechanism of ADP-Ribose Hydrolysis By Human NUDT5 From Structural and Kinetic Studies. *Journal of Molecular Biology* **2008**, *379*, 568–578.

[221] Ryu, K. W.; Kraus, W. L. Who Put the "A" in ATP? Generation of ATP from ADP-Ribose in the Nucleus for Hormone-Dependent Gene Regulation. *Molecular Cell* **2016**, *63*, 349–351.

[222] Hori, M.; Satou, K.; Harashima, H.; Kamiya, H. Suppression of Mutagenesis by 8-Hydroxy-2'-Deoxyguanosine 5'-Triphosphate (7,8-Dihydro-8-Oxo-2'-Deoxyguanosine 5'-Triphosphate) by Human MTH1, MTH2, and NUDT5. *Free Radical Biology and Medicine* **2010**, *48*, 1197–1201.

[223] Pickup, K. E.; Pardow, F.; Carbonell-Caballero, J.; Lioutas, A.; Villanueva-Canas, J. L.; Wright, R. H. G.; Beato, M. Expression of Oncogenic Drivers in

3D Cell Culture Depends on Nuclear ATP Synthesis by NUDT5. *Cancers* **2019**, *11*, 1337.

[224] Tong, X.-Y.; Quan, Y.; Zhang, H.-Y. NUDT5 as a Novel Drug Target and Prognostic Biomarker for ER-positive Breast Cancer. *Drug Discovery Today* **2020**,

[225] Li, D.-N.; Yang, C.-C.; Li, J.; Ou Yang, Q.-G.; Zeng, L.-T.; Fan, G.-Q.; Liu, T.-H.; Tian, X.-Y.; Wang, J.-J.; Zhang, H.; Dai, D.-P.; Cui, J.; Cai, J.-P. The High Expression of MTH1 and NUDT5 Promotes Tumor Metastasis and Indicates a Poor Prognosis in Patients with Non-Small-Cell Lung Cancer. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2020**, *1868*, 118895.

[226] Wright, R. H. G.; Castellano, G.; Bonet, J.; Dily, F. L.; Font-Mateu, J.; Ballare, C.; Nacht, A. S.; Soronellas, D.; Oliva, B.; Beato, M. CDK2-dependent Activation of PARP-1 Is Required for Hormonal Gene Regulation in Breast Cancer Cells. *Genes & Development* **2012**, *26*, 1972–1983.

[227] Sultana, R.; Islam, M.; Haque, M.; Evamoni, F. Z.; Imran, Z. M.; Khanom, J.; Munim, M. Molecular Docking Based Virtual Screening of the Breast Cancer Target NUDT5. *Bioinformation* **2019**, *15*, 784–789.

[228] Bionet Fragment Libraries, Keyorganics. https://www.keyorganics.net/services/bionet-products/fragment-libraries/, accessed on 16.9.2020.

[229] Resnick, E. *et al.* Rapid Covalent-Probe Discovery by Electrophile-Fragment Screening. *Journal of the American Chemical Society* **2019**, *141*, 8951–8968.

[230] McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *Journal of Applied Crystallography* **2007**, *40*, 658–674.

[231] Project, C. C.; 4, N. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallographica Section D: Biological Crystallography* **1994**, *50*, 760–763.

[232] Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Cryst D* **2010**, *66*, 486–501.

[233] Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallographica Section D: Biological Crystallography* **2011**, *67*, 355–367.

[234] Wojdyr, M.; Keegan, R.; Winter, G.; Ashton, A. DIMPLE - a Pipeline for the Rapid Generation of Difference Maps from Protein Crystals with Putatively Bound Ligands. *Acta Crystallographica Section A: Foundations of Crystallography* **2013**, *69*, 299–299.

[235] Wider, G.; Dreier, L. Measuring Protein Concentrations by NMR Spectroscopy. *Journal of the American Chemical Society* **2006**, *128*, 2571–2576.

[236] Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. PLIP: Fully Automated Protein-Ligand Interaction Profiler. *Nucleic Acids Research* **2015**, *43*, W443–W447.

[237] Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM - A New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from the Distorted Native Conformation. *Journal of Computational Chemistry* **1994**, *15*, 488–506.

[238] Abagyan, R.; Totrov, M. Biased Probability Monte Carlo Conformational Searches and Electrostatic Calculations for Peptides and Proteins. *Journal of Molecular Biology* **1994**, *235*, 983–1002.

[239] Orry, A. J. W.; Abagyan, R. *Homology Modeling: Methods and Protocols*; Humana Press: Totowa, NJ, 2012; pp 351–373.

[240] Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *Journal of Computational Chemistry* **1996**, *17*, 490–519.

[241] Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.

[242] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712–725.

[243] Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix - Coil Transition of Polypeptides. *The Journal of Physical Chemistry B* **2009**, *113*, 9004–9015.

[244] Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.

[245] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *The Journal of Physical Chemistry* **1987**, *91*, 6269–6271.

[246] Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *The Journal of Chemical Physics* **2007**, *126*, 014101.

[247] Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52*, 7182–7190.

[248] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *The Journal of Chemical Physics* **1995**, *103*, 8577–8593.

[249] Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *Journal of Computational Chemistry* **1992**, *13*, 952–962.

[250] Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 116–122.

[251] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 435–447.

[252] PyMOL. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

[253] Ng, J. T.; Dekker, C.; Reardon, P.; von Delft, F. Lessons from Ten Years of Crystallization Experiments at the SGC. *Acta Crystallographica Section D: Structural Biology* **2016**, *72*, 224–235.

[254] 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. Rapid Optimisation of Fragments and Hits to Lead Compounds from Screening of Crude Reaction Mixtures. *Communications Chemistry* **2020**, *3*, 1–11.

[255] Harner, M. J.; Frank, A. O.; Fesik, S. W. Fragment-Based Drug Discovery Using NMR Spectroscopy. *Journal of biomolecular NMR* **2013**, *56*, 65–75.

[256] Patel, D.; Antwi, J.; Koneru, P. C.; Serrao, E.; Forli, S.; Kessl, J. J.; Feng, L.; Deng, N.; Levy, R. M.; Fuchs, J. R.; Olson, A. J.; Engelman, A. N.; Bauman, J. D.; Kvaratskhelia, M.; Arnold, E. A New Class of Allosteric HIV-1 Integrase Inhibitors Identified by Crystallographic Fragment Screening of the Catalytic Core Domain. *Journal of Biological Chemistry* **2016**, *291*, 23569–23577.

[257] Collins, P. M.; Douangamath, A.; Talon, R.; Dias, A.; Brandao-Neto, J.; Krojer, T.; von Delft, F. In *Methods in Enzymology*; Lesburg, C. A., Ed.; Modern Approaches in Drug Discovery; Academic Press, 2018; Vol. 610; pp 251–264.

[258] Georgiou, C.; McNae, I.; Wear, M.; Ioannidis, H.; Michel, J.; Walkinshaw, M. Pushing the Limits of Detection of Weak Binding Using Fragment-Based Drug Discovery: Identification of New Cyclophilin Binders. *Journal of Molecular Biology* **2017**, *429*, 2556–2570.

[259] Fischer, M.; Shoichet, B. K.; Fraser, J. S. One Crystal, Two Temperatures: Cryocooling Penalties Alter Ligand Binding to Transient Protein Sites. *ChemBioChem* **2015**, *16*, 1560–1564.

[260] Sauter, N. K.; Glick, G. D.; Crowther, R. L.; Park, S. J.; Eisen, M. B.; Skehel, J. J.; Knowles, J. R.; Wiley, D. C. Crystallographic Detection of a Second Ligand Binding Site in Influenza Virus Hemagglutinin. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 324–328.

[261] Yamane, J.; Yao, M.; Zhou, Y.; Hiramatsu, Y.; Fujiwara, K.; Yamaguchi, T.; Yamaguchi, H.; Togame, H.; Tsujishita, H.; Takemoto, H.; Tanaka, I. In-Crystal Affinity Ranking of Fragment Hit Compounds Reveals a Relationship with Their Inhibitory Activities. *Journal of Applied Crystallography* **2011**, *44*, 798–804.

[262] Wu, S.-y.; Dornan, J.; Kontopidis, G.; Taylor, P.; Walkinshaw, M. D. The First Direct Determination of a Ligand Binding Constant in Protein Crystals. *Angewandte Chemie International Edition* **2001**, *40*, 582–586.

[263] Glöckner, S.; Heine, A.; Klebe, G. A Proof-of-Concept Fragment Screening of a Hit-Validated 96-Compounds Library against Human Carbonic Anhydrase II. *Biomolecules* **2020**, *10*, 518.

[264] Retra, K.; Irth, H.; van Muijlwijk-Koezen, J. E. Surface Plasmon Resonance Biosensor Analysis as a Useful Tool in FBDD. *Drug Discovery Today: Technologies* **2010**, *7*, e181–e187.

[265] Chavanieu, A.; Pugnière, M. Developments in SPR Fragment Screening. *Expert Opinion on Drug Discovery* **2016**, *11*, 489–499.

[266] Huber, S.; Casagrande, F.; Hug, M. N.; Wang, L.; Heine, P.; Kummer, L.; Plückthun, A.; Hennig, M. SPR-based Fragment Screening with Neurotensin Receptor 1 Generates Novel Small Molecule Ligands. *PLOS ONE* **2017**, *12*, e0175842.

[267] Morton, T. A.; Myszka, D. G. *Methods in Enzymology*; Energetics of Biological Macromolecules Part B; Academic Press, 1998; Vol. 295; pp 268–294.

[268] Myszka, D. G. Improving Biosensor Analysis. *Journal of Molecular Recognition* **1999**, *12*, 279–284.

[269] Williams, M., Daviter, T., Eds. *Protein-Ligand Interactions: Methods and Applications*, 2nd ed.; Methods in Molecular Biology; Humana Press, 2013.

[270] Glaser, R. W. Antigen-Antibody Binding and Mass Transport by Convection and Diffusion to a Surface: A Two-Dimensional Computer Model of Binding and Dissociation Kinetics. *Analytical Biochemistry* **1993**, *213*, 152–161.

[271] Myszka, D. G.; He, X.; Dembo, M.; Morton, T. A.; Goldstein, B. Extending the Range of Rate Constants Available from BIACORE: Interpreting Mass Transport-Influenced Binding Data. *Biophysical Journal* **1998**, *75*, 583–594.

[272] BioLogic Software. http://www.biologic.com.au/scrubber.html, 2020.

[273] Karlsson, R.; Stahlberg, R. Surface Plasmon Resonance Detection and Multispot Sensing for Direct Monitoring of Interactions Involving Low-Molecular-Weight Analytes and for Determination of Low Affinities. *Analytical Biochemistry* **1995**, *228*, 274–280.

[274] Klebe, G.; Böhm, H.-J. Energetic and Entropic Factors Determining Binding Affinity in Protein-Ligand Complexes. *Journal of Receptors and Signal Transduction* **1997**, *17*, 459–473.

[275] Lafont, V.; Armstrong, A. A.; Ohtaka, H.; Kiso, Y.; Amzel, L. M.; Freire, E. Compensating Enthalpic and Entropic Changes Hinder Binding Affinity Optimization. *Chemical Biology & Drug Design* **2007**, *69*, 413–422.

[276] Velupillai, S. *et al.* Human Peroxisomal Coenzyme A Diphosphatase (NUDT7); A Target Enabling Package. 2018.

[277] Rupp, B. In *Encyclopedia of Biophysics*; Roberts, G. C. K., Ed.; Springer: Berlin, Heidelberg, 2013; pp 1346–1353.

[278] Danley, D. E. Crystallization to Obtain Protein-Ligand Complexes for Structure-Aided Drug Design. *Acta Crystallographica Section D: Biological Crystallography* **2006**, *62*, 569–575.

[279] McNae, I. W.; Kan, D.; Kontopidis, G.; Patterson, A.; Taylor, P.; Worrall, L.; Walkinshaw, M. D. Studying Protein-Ligand Interactions Using Protein Crystallography. *Crystallography Reviews* **2005**, *11*, 61–71.

[280] Shortridge, M. D.; Hage, D. S.; Harbison, G. S.; Powers, R. Estimating Protein-Ligand Binding Affinity Using High-Throughput Screening by NMR. *Journal of Combinatorial Chemistry* **2008**, *10*, 948–958.

[281] Cala, O.; Guilliere, F.; Krimm, I. NMR-based Analysis of Protein-Ligand Interactions. *Analytical and Bioanalytical Chemistry* **2014**, *406*, 943–956.

[282] Cala, O.; Krimm, I. Ligand-Orientation Based Fragment Selection in STD NMR Screening. *Journal of Medicinal Chemistry* **2015**, *58*, 8739–8742.

[283] Dalvit, C.; Flocco, M.; Knapp, S.; Mostardini, M.; Perego, R.; Stockman, B. J.; Veronesi, M.; Varasi, M. High-Throughput NMR-Based Screening with Competition Binding Experiments. *Journal of the American Chemical Society* **2002**, *124*, 7702–7709.

[284] Dalvit, C. Theoretical Analysis of the Competition Ligand-Based NMR Experiments and Selected Applications to Fragment Screening and Binding Constant Measurements. *Concepts in Magnetic Resonance Part A* **2008**, *32A*, 341–372.

[285] Jose, R. A.; Voet, A.; Broos, K.; Jakobi, A. J.; Bruylants, G.; Egle, B.; Zhang, K. Y. J.; Maeyer, M. D.; Deckmyn, H.; Borggraeve, W. M. D. An Integrated Fragment Based Screening Approach for the Discovery of Small Molecule Modulators of the VWF-GPIba Interaction. *Chemical Communications* **2012**, *48*, 11349–11351.

[286] Rüdisser, S. H.; Goldberg, N.; Ebert, M.-O.; Kovacs, H.; Gossert, A. D. Efficient Affinity Ranking of Fluorinated Ligands by 19F NMR: CSAR and FastCSAR. *Journal of Biomolecular NMR* **2020**, *74*, 579–594.

[287] Singh, M.; Tam, B.; Akabayov, B. NMR-Fragment Based Virtual Screening: A Brief Overview. *Molecules* **2018**, *23*, 233.

[288] Jayalakshmi, V.; Krishna, N. R. Complete Relaxation and Conformational Exchange Matrix (CORCEMA) Analysis of Intermolecular Saturation Transfer Effects in Reversibly Forming Ligand-Receptor Complexes. *Journal of Magnetic Resonance* **2002**, *155*, 106–118.

[289] Kemper, S.; Patel, M. K.; Errey, J. C.; Davis, B. G.; Jones, J. A.; Claridge, T. D. W. Group Epitope Mapping Considering Relaxation of the Ligand (GEM-CRL): Including Longitudinal Relaxation Rates in the Analysis of Saturation Transfer Difference (STD) Experiments. *Journal of Magnetic Resonance* **2010**, *203*, 1–10.

[290] Maity, S.; Gundampati, R. K.; Suresh Kumar, T. K. NMR Methods to Characterize Protein-Ligand Interactions. *Natural Product Communications* **2019**, *14*, 1934578X19849296.

[291] Aretz, J.; Anumala, U. R.; Fuchsberger, F. F.; Molavi, N.; Ziebart, N.; Zhang, H.; Nazaré, M.; Rademacher, C. Allosteric Inhibition of a Mammalian Lectin. *Journal of the American Chemical Society* **2018**, *140*, 14915–14925.

[292] Clarkson, J.; Campbell, I. D. Studies of Protein-Ligand Interactions by NMR. *Biochemical Society Transactions* **2003**, *31*, 1006–1009.

[293] Hall, R. J.; Murray, C. W.; Verdonk, M. L. The Fragment Network: A Chemistry Recommendation Engine Built Using a Graph Database. *Journal of Medicinal Chemistry* **2017**, *60*, 6440–6450.

[294] Lamoree, B.; Hubbard, R. E. Using Fragment-Based Approaches to Discover New Antibiotics. *SLAS DISCOVERY: Advancing the Science of Drug Discovery* **2018**, *23*, 495–510.

[295] 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* **2020**, *8*.

[296] Desjarlais, R. L. Using Computational Techniques in Fragment-Based Drug Discovery. *Methods in Enzymology* **2011**, *493*, 137–155.

[297] Rudling, A.; Gustafsson, R.; Almlöf, I.; Homan, E.; Scobie, M.; Warpman Berglund, U.; Helleday, T.; Stenmark, P.; Carlsson, J. Fragment-Based Discovery and

Optimization of Enzyme Inhibitors by Docking of Commercial Chemical Space. *Journal of Medicinal Chemistry* **2017**, *60*, 8160–8169.

[298] Borysko, P.; Moroz, Y. S.; Vasylchenko, O. V.; Hurmach, V. V.; Starodubtseva, A.; Stefanishena, N.; Nesteruk, K.; Zozulya, S.; Kondratov, I. S.; Grygorenko, O. O. Straightforward Hit Identification Approach in Fragment-Based Discovery of Bromodomain-Containing Protein 4 (BRD4) Inhibitors. *Bioorganic & Medicinal Chemistry* **2018**, *26*, 3399–3405.

[299] Ciulli, A.; Abell, C. Fragment-Based Approaches to Enzyme Inhibition. *Current Opinion in Biotechnology* **2007**, *18*, 489–496.

[300] Fragalysis - a Cloud-Based Tool for Fragment Analysis. Diamond Light Source. https://fragalysis.diamond.ac.uk/viewer/react/landing, accessed on 09.11.2020.

[301] REAL Database - Enamine. https://enamine.net/library-synthesis/real-compounds/real-database, accessed on 09.11.2020.

[302] Nair, P. C.; Malde, A. K.; Drinkwater, N.; Mark, A. E. Missing Fragments: Detecting Cooperative Binding in Fragment-Based Drug Design. *ACS Medicinal Chemistry Letters* **2012**, *3*, 322–326.

[303] Shipe, W. D.; Sharik, S. S.; Barrow, J. C.; McGaughey, G. B.; Theberge, C. R.; Uslaner, J. M.; Yan, Y.; Renger, J. J.; Smith, S. M.; Coleman, P. J.; Cox, C. D. Discovery and Optimization of a Series of Pyrimidine-Based Phosphodiesterase 10A (PDE10A) Inhibitors through Fragment Screening, Structure-Based Design, and Parallel Synthesis. *Journal of Medicinal Chemistry* **2015**, *58*, 7888–7894.

[304] Deaton, D. N. *et al.* The Discovery of Quinoline-3-Carboxamides as Hematopoietic Prostaglandin D Synthase (H-PGDS) Inhibitors. *Bioorganic & Medicinal Chemistry* **2019**, *27*, 1456–1478.

[305] Kessler, D. *et al.* Drugging an Undruggable Pocket on KRAS. *Proceedings of the National Academy of Sciences of the United States of America* **2019**, *116*, 15823–15829.

[306] Chen, Y.; Shoichet, B. K. Molecular Docking and Ligand Specificity in Fragment-Based Inhibitor Discovery. *Nature Chemical Biology* **2009**, *5*, 358–364.

[307] Verdonk, M. L.; Giangreco, I.; Hall, R. J.; Korb, O.; Mortenson, P. N.; Murray, C. W. Docking Performance of Fragments and Druglike Compounds. *Journal of Medicinal Chemistry* **2011**, *54*, 5422–5431.

[308] Totrov, M.; Abagyan, R. Flexible Ligand Docking to Multiple Receptor Conformations: A Practical Alternative. *Current Opinion in Structural Biology* **2008**, *18*, 178–184.

[309] Sledz, P.; Caflisch, A. Protein Structure-Based Drug Design: From Docking to Molecular Dynamics. *Current Opinion in Structural Biology* **2018**, *48*, 93–102.

[310] Chen, Y.-C. Beware of Docking! *Trends in Pharmacological Sciences* **2015**, *36*, 78–95.

[311] Geremia, S.; Campagnolo, M.; Demitri, N.; Johnson, L. N. Simulation of Diffusion Time of Small Molecules in Protein Crystals. *Structure* **2006**, *14*, 393–400.

[312] Ruiz-Carmona, S.; Schmidtke, P.; Luque, F. J.; Baker, L.; Matassova, N.; Davis, B.; Roughley, S.; Murray, J.; Hubbard, R.; Barril, X. Dynamic Undocking and the Quasi-Bound State as Tools for Drug Discovery. *Nature Chemistry* **2017**, *9*, 201–206.

[313] Majewski, M.; Ruiz-Carmona, S.; Barril, X. In *Rational Drug Design: Methods and Protocols*; Mavromoustakos, T., Kellici, T. F., Eds.; Methods in Molecular Biology; Springer: New York, NY, 2018; pp 195–215.

[314] Kessler, D.; Bergner, A.; Böttcher, J.; Fischer, G.; Döbel, S.; Hinkel, M.; Müllauer, B.; Weiss-Puxbaum, A.; McConnell, D. B. Drugging All RAS Isoforms with One Pocket. *Future Medicinal Chemistry* **2020**, *12*, 1911–1923.

A

Appendix

A.1. Appendix for chapter 3

A.1.1. NUDT5A constructs

Construct name	**NUDT5A-c001**
Construct protein sequence	MHHHHHHSSGVDLGTENLYFQSMESQEPTESSQNGKQYIISEELISEGKWVKLEKTTYMDPTGKTRTWESVKRTTRKEQTADGVAVIPVLQRTLHYECIVLVKQFRPPMGGYCIEFPAGLIDDGETPEAAALRELEEETGYKGDIAECSPAVCMDPGLSNCTIHIVTVTINGDDAENARPKPKPGDGEFVEVISLPKNDLLQRLDALVAEEHLTVDARVYSYALALKHANAKPFEVPFLKF
Number of residues	241
Molecular weight (Da)	26880.4
Extinction coef. (M^{-1} cm^{-1})	22920
Estimated pI	4.95
Construct protein sequence, cleaved	SMESQEPTESSQNGKQYIISEELISEGKWVKLEKTTYMDPTGKTRTWESVKRTTRKEQTADGVAVIPVLQRTLHYECIVLVKQFRPPMGGYCIEFPAGLIDDGETPEAAALRELEEETGYKGDIAECSPAVCMDPGLSNCTIHIVTVTINGDDAENARPKPKPGDGEFVEVISLPKNDLLQRLDALVAEEHLTVDARVYSYALALKHANAKPFEVPFLKF
Number of residues, cleaved	220
Molecular weight (Da), cleaved	24414.8
Extinction coef. (M^{-1} cm^{-1}), cleaved	21430
Estimated pI	4.65

Figure A.1. The sequence of the construct NUDT5A-c001 and its characteristics.

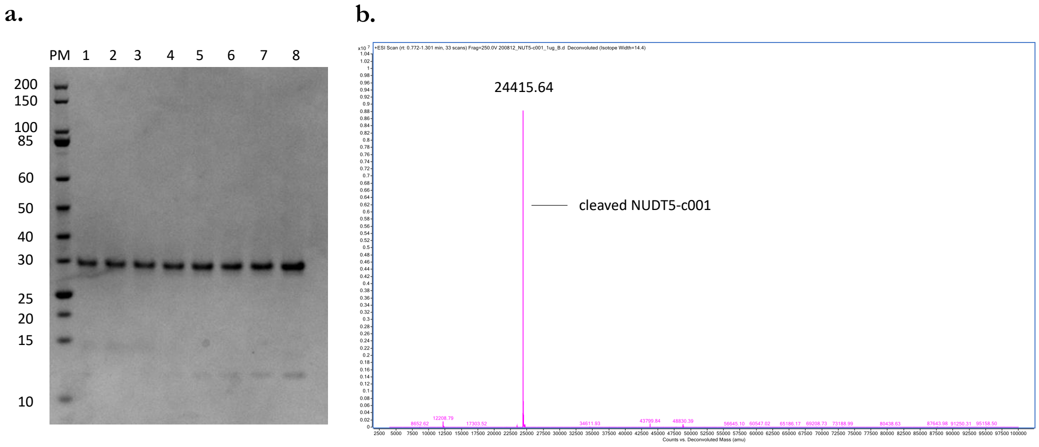


Figure A.2. SDS-PAGE and MS analysis of the construct NUDT5A-c001. **a.** SDS PAGE analysis of NUDT5A-c001 fractions after size exclusion chromatography. PM lane is molecular weight marker; lanes 1 to 8 correspond to protein fractions collected. **b.** MS analysis confirms presence of the cleaved construct encoded by NUDT5A-c001.

Construct name	**NUDT5A-c002**
Construct protein sequence	MHHHHHHSSGVDLGTENLYFQSMESQEPTESSQNGKQYIISEELISEGKWVKLEKTTYMDPTGKTRTWESVKRTTRKEQTADGVAVIPVLQRTLHYECIVLVKQFRPPMGGYCIEFPAGLIDDGETPEAAALRELEEETGYKGDIAECSPAVCMDPGLSNCTIHIVTVTINGDDAENARPKPKPGDGEFVEVISLPKNDLLQRLDALVAEEHLTVDARVYSYALALKHAN
Number of residues	230
Molecular weight (Da)	25575.8
Extinction coef. (M^{-1} cm^{-1})	22920
Estimated pI	4.85
Construct protein sequence, cleaved	SMESQEPTESSQNGKQYIISEELISEGKWVKLEKTTYMDPTGKTRTWESVKRTTRKEQTADGVAVIPVLQRTLHYECIVLVKQFRPPMGGYCIEFPAGLIDDGETPEAAALRELEEETGYKGDIAECSPAVCMDPGLSNCTIHIVTVTINGDDAENARPKPKPGDGEFVEVISLPKNDLLQRLDALVAEEHLTVDARVYSYALALKHAN
Number of residues, cleaved	209
Molecular weight (Da), cleaved	23110.2
Extinction coef. (M^{-1} cm^{-1}), cleaved	21430
Estimated pI	4.55

Figure A.3. The sequence of the construct NUDT5A-c002 and its characteristics.

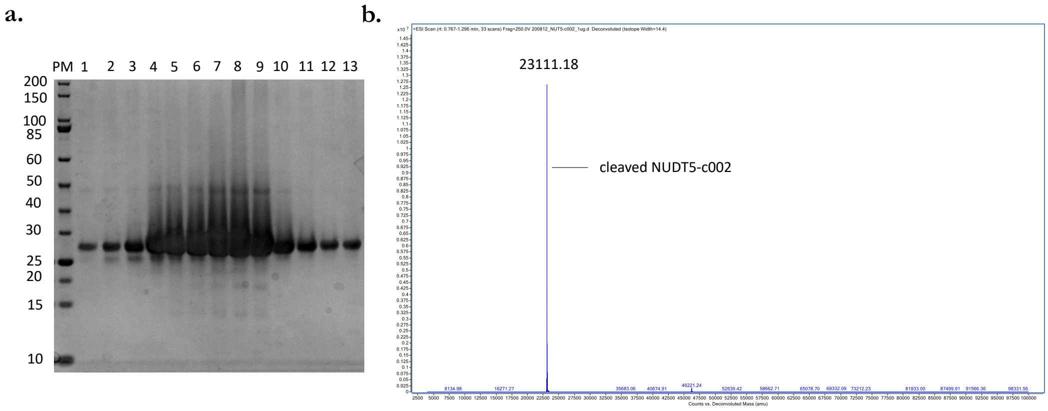


Figure A.4. SDS-PAGE and MS analysis of the construct NUDT5A-c002. **a.** SDS PAGE analysis of NUDT5A-c002 fractions after size exclusion chromatography. PM lane is molecular weight marker; lanes 1 to 13 correspond to protein fractions collected. **b.** MS analysis confirms presence of the cleaved construct encoded by NUDT5A-c002.

Construct name	**NUDT5A-c003**
Construct protein sequence	MHHHHHHSSGVDLGTENLYFQSMKQYIISEELISEGKWVKLEKTTYMDPTGKTRTWESVKRTTRKEQTADGVAVIPVLQRTLHYECIVLVKQFRPPMGGYCIEFPAGLIDDGETPEAAALRELEEETGYKGDIAECSPAVCMDPGLSNCTIHIVTVTINGDDAENARPKPKPGDGEFVEVISLPKNDLLQRLDALVAEEHLTVDARVYSYALALKHANAKPFEVPFLKF
Number of residues	229
Molecular weight (Da)	25606.2
Extinction coef. (M^{-1} cm^{-1})	22920
Estimated pI	5.25
Construct protein sequence, cleaved	SMKQYIISEELISEGKWVKLEKTTYMDPTGKTRTWESVKRTTRKEQTADGVAVIPVLQRTLHYECIVLVKQFRPPMGGYCIEFPAGLIDDGETPEAAALRELEEETGYKGDIAECSPAVCMDPGLSNCTIHIVTVTINGDDAENARPKPKPGDGEFVEVISLPKNDLLQRLDALVAEEHLTVDARVYSYALALKHANAKPFEVPFLKF
Number of residues, cleaved	208
Molecular weight (Da), cleaved	23140.5
Extinction coef. (M^{-1} cm^{-1}), cleaved	21430
Estimated pI	4.85

Figure A.5. The sequence of the construct NUDT5A-c003 and its characteristics.

Construct name	**NUDT5A-c004**
Construct protein sequence	MHHHHHHSSGVDLGTENLYFQSMKQYIISEELISEGKWVKLEKTTYMDPTGKTRTWESVKRTTRKEQTADGVAVIPVLQRTLHYECIVLVKQFRPPMGGYCIEFPAGLIDDGETPEAAALRELEEETGYKGDIAECSPAVCMDPGLSNCTIHIVTVTINGDDAENARPKPKPGDGEFVEVISLPKNDLLQRLDALVAEEHLTVDARVYSYALALKHAN
Number of residues	218
Molecular weight (Da)	24301.6
Extinction coef. (M^{-1} cm^{-1})	22920
Estimated pI	5.05
Construct protein sequence, cleaved	SMKQYIISEELISEGKWVKLEKTTYMDPTGKTRTWESVKRTTRKEQTADGVAVIPVLQRTLHYECIVLVKQFRPPMGGYCIEFPAGLIDDGETPEAAALRELEEETGYKGDIAECSPAVCMDPGLSNCTIHIVTVTINGDDAENARPKPKPGDGEFVEVISLPKNDLLQRLDALVAEEHLTVDARVYSYALALKHAN
Number of residues, cleaved	197
Molecular weight (Da), cleaved	21835.9
Extinction coef. (M^{-1} cm^{-1}), cleaved	21430
Estimated pI	4.75

Figure A.6. The sequence of the construct NUDT5A-c004 and its characteristics.

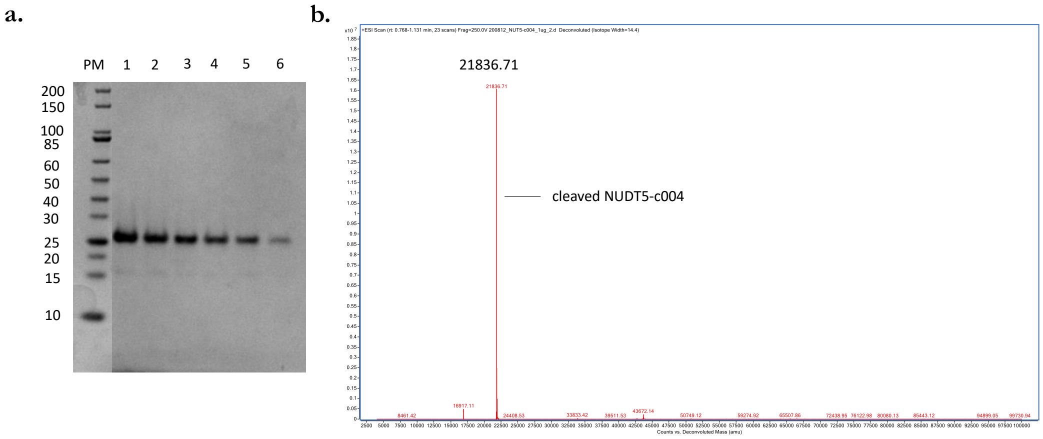


Figure A.7. SDS-PAGE and MS analysis of the construct NUDT5A-c004. **a.** SDS PAGE analysis of NUDT5A-c004 fractions after ion exchange chromatography. PM lane is molecular weight marker; lanes 1 to 6 correspond to protein fractions collected. **b.** MS analysis confirms presence of the cleaved construct encoded by NUDT5A-c004.

Construct name	**NUDT5A-c005**
Construct protein sequence	MHHHHHHSSGVDLGTENLYFQSMESQEPTESSQNGKQYIISEELISEGKWVKLEKTTYMDPTGKTRTWESVKRTTRKEQTADGVAVIPVLQRTLHYECIVLVKQFRPPMGGYCIEFPAGLIDDGETPEAAALRELEEETGYKGDIAECSPAVCMDPGLSNCTIHIVTVTINGDDAENARPKPKPGDGEFVEVISLPKNDLLQRLDALVAEEHLTVDARVYSYALALKHANAKPFEVPFLKFSSKGGYGLNDIFEAQKIEWHE
Number of residues	262
Molecular weight (Da)	29271.0
Extinction coef. (M^{-1} cm^{-1})	29910
Estimated pI	5.1
Construct protein sequence, cleaved	SMESQEPTESSQNGKQYIISEELISEGKWVKLEKTTYMDPTGKTRTWESVKRTTRKEQTADGVAVIPVLQRTLHYECIVLVKQFRPPMGGYCIEFPAGLIDDGETPEAAALRELEEETGYKGDIAECSPAVCMDPGLSNCTIHIVTVTINGDDAENARPKPKPGDGEFVEVISLPKNDLLQRLDALVAEEHLTVDARVYSYALALKHANAKPFEVPFLKFSSKGGYGLNDIFEAQKIEWHE
Number of residues, cleaved	241
Molecular weight (Da), cleaved	26805.4
Extinction coef. (M^{-1} cm^{-1}), cleaved	28420
Estimated pI	4.86

Figure A.8. The sequence of the construct NUDT5A-c005 and its characteristics.

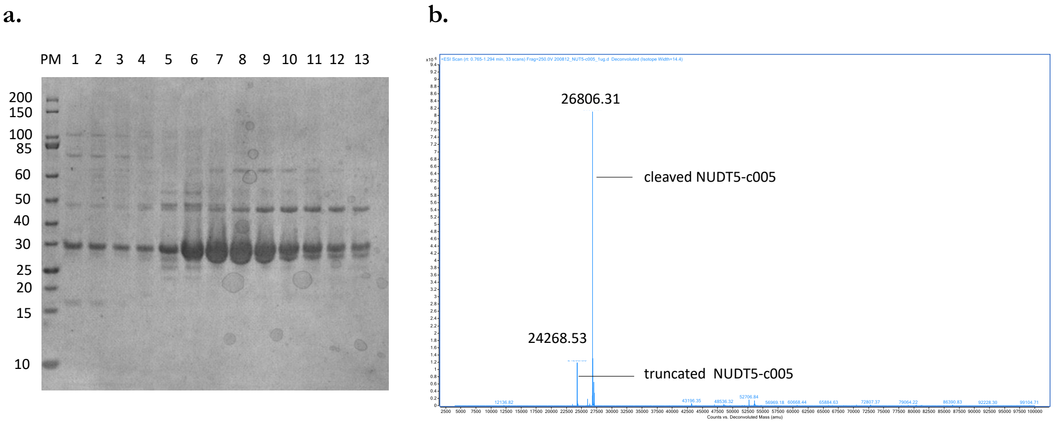


Figure A.9. SDS-PAGE and MS analysis of the construct NUDT5A-c005. **a.** SDS PAGE analysis of NUDT5A-c005 fractions after size exclusion chromatography. PM lane is molecular weight marker; lanes 1 to 13 correspond to protein fractions collected. **b.** MS analysis confirms presence of the cleaved construct encoded by NUDT5A-c005 as well as truncated NUDT5A-c005.

A.1.2. Overview of NUDT5 XChem fragment hits.

#	Data Set #	PDB ID Crystal	PDB ID Ligand	Smiles string	MW (Da)	Site	# of copies	Occupancy	Soaking time (hh:mm)	Diffraction (Å)
1	NUDT5A-x0114	5QJ4	K04	CC1CCOC1C(=O)NC2=CN=NS2	213.26	1/2/3/4	4	0.8/0.94/0.86/0.89	00:35	1.89
2	NUDT5A-x0122	5QJ5	K0G	O=C(NC=1C=CC=CC1)NC=2C=CC=NC2	213.24	1/2/3/4	4	0.34/0.42/0.40/0.40	00:38	1.72
3	NUDT5A-x0125	5QJ6	GQJ	COC(=O)[C@@H]1C[C@@H](O)CN1C(=O)C2=CC=CO2	239.23	12/14	2	0.6/0.56	00:39	1.65
4	NUDT5A-x0131	5QJ7	JMM	O=C(C1CC1)N2CCN(CC2)C(=O)C3=CC=CO3	248.28	1/2/3/4	4	0.68/0.82/0.74/0.83	00:35	1.56
5	NUDT5A-x0158	5QJ8	ELQ	CN1CCN(CC1)C(=O)C=2C=CC(F)=C(F)C2	240.25	1	1	0.70	00:18	1.76
6	NUDT5A-x0169	5QJ9	K0D	CC1CN(C(C)CO1)C(=O)C=2C=NSN2	227.28	1/2/3/4	4	0.72/0.86/0.92/0.86	00:22	1.85
7	NUDT5A-x0171	5QJA	K0A	CC1=NC(C(=O)N2CCCC2)=C(C)S1	210.30	1/2/3/4	4	0.8/0.86/0.82/0.81	00:51	1.64
8	NUDT5A-x0176	5QJB	K07	CN(CC1CCOC1)C=2N=CN=CC2Cl	227.69	1/2/3	2	0.82/0.94/0.8/	00:52	1.66
9	NUDT5A-x0177	5QJC	K0J	CCNC(=O)C1=CNN=N1	140.15	1/4	2	0.64/0.72	00:53	1.47
10	NUDT5A-x0195	5QJD	K0M	CC=1C=C(ON1)C(=O)NC=2C=CN(C)N2	206.21	3/4	2	0.84/0.87	00:34	1.61
11	NUDT5A-x0197	5QJE	K0P	CCN1C=C(CNC(=O)C=2C=CC=NC2)C=N1	230.27	1/4	2	0.74/0.77	00:34	1.75
12	NUDT5A-x0210	5QJF	K0S	CCCC1=NN=C(NC(=O)C=2C=CC=CN2)S1	248.30	1/12	2	0.74/1.0	00:41	1.63
13	NUDT5A-x0235	5QJG	K0V	COC(=O)C1CCN(CC1)C(=O)NC(C)(C)C	242.32	1/2/3	4	0.68/0.71/0.68/0.71	00:34	1.57
14	NUDT5A-x0244	5QJH	K0Y	COC(=O)CNC(=O)C=1C=C(C)ON1	198.18	1	1	0.68	00:37	2.03
15	NUDT5A-x0251	5QJI	K1A	CC1=NN(CC=2C(C)=NOC2C)C(=O)S1	225.27	1/2	2	0.74/0.98	00:39	1.69
16	NUDT5A-x0256	5QJJ	6SU	COC(=O)C=1C=CC=C(NS(=O)(=O)C)C1	229.25	2/4	2	0.86/0.86	00:41	1.71
17	NUDT5A-x0262	5QJK	K1D	C1CN(C1)C=2N=CN=C3N=CNC23	175.20	1/2/3/4	4	0.88/0.86/0.9/0.86	00:25	1.65
18	NUDT5A-x0286	5QJL	JJM	CC=1C=CC=C(NC=2N=CN=C3N(C)N=CC23)C1	239.28	1/2/3/4	4	0.82/0.77/0.8/0.77	00:33	1.66
19	NUDT5A-x0299	5QJM	K1G	CC(C)N(C)C=1N=CN=C2N(C)N=CC12	205.27	1/2/3/4	4	0.8/0.9/0.9/0.92	00:32	1.75
20	NUDT5A-x0319	5QJN	K1J	CC1OCCN(C1C)C2=NC(C)=NS2	213.30	1/3	2	0.72/0.9	00:40	1.77
21	NUDT5A-x0320	5QJO	K41	CC(C)C(=O)NC=1N=NN(C)N1	169.19	1/2/3/4	4	0.86/0.88/0.86/0.92	00:40	1.98

22	NUDT5A-x0333	5QJP	K1P	CC1=NSC(=N1)N2CCCNCC2	198.29	1/2/3/4	4	0.76/0.98/0.92/0.83	00:21	1.50
23	NUDT5A-x0339	5QJQ	K1S	CCN(CC)C1=CC(C)=NC2=NC=NN12	205.27	1/2/3/4	4	0.94/0.93/0.94/0.93	00:24	1.55
24	NUDT5A-x0392	5QJR	K1V	CC(=O)N1CCCC(C1)C(=O)N	170.21	1	1	0.72	00:30	1.62
25	NUDT5A-x0403	5QJS	K1Y	COC(=O)NC=1SC(C)=NC1C=2C=CC=CC2	248.30	3	1	0.80	00:32	1.58
26	NUDT5A-x0412	5QJT	K24	CNC(=O)C1=CN=C(C)S1	156.20	1/4	2	0.80/0.91	00:35	1.62
27	NUDT5A-x0463	5QJU	K2G	CC(C)C=1N=CC(Cl)=C(N1)C(=O)N	199.64	1/2/3/4	4	0.84/0.91/0.96/0.91	00:24	1.77
28	NUDT5A-x0469	5QJV	K2J	CNC=1N=CC=C(OC)N1	139.16	1/2/3	3	0.78/0.9/0.78	00:26	1.61
29	NUDT5A-x0525	5QJW	K2M	NC(=O)N1CCN(CC1)C(=O)C2=CC=CO2	223.23	1/2/3	3	0.72/0.90/0.96	00:23	1.57
30	NUDT5A-x0526	5QJX	K2P	[O-]C(=O)C=1C=CC=CC1OC(F)(F)F	205.11	24	1	0.98	00:23	1.73
31	NUDT5A-x0552	5QJY	K2S	CN(C)S(=O)(=O)C=1C(C)=NNC1C	203.26	1/3/4	2	0.56/0.56 /0.65	00:31	1.77
32	NUDT5A-x0553	5QJZ	K2V	CC(N)C=1C=CC(=CC1)C=2C=NC=NC2	199.26	1/3	2	0.7/0.78	00:32	1.52
33	NUDT5A-x0554	5QK0	K31	CC1=NNC(C)=C1C=2C=CC=CC2N	187.25	1/3	2	0.80/0.94	00:32	1.44
34	NUDT5A-x0574	5QK1	K34	C1=CSC(=N1)C2=NN=CN2	152.18	1/2/3/4	4	0.74/0.89/0.88/0.89	00:31	1.49
35	NUDT5A-x0600	5QK2	JH4	CNC=1N=CC=CN1	109.13	1	1	0.78	00:28	1.65
36	NUDT5A-x0605	5QK3	K3A	CC(=O)NC1=CC(C)=NN1	139.16	3	1	0.76	00:29	1.71
37	NUDT5A-x0627	5QK4	K3D	CCNC(=O)N1CCN(CC1)C(=O)C	199.25	1/2/3	3	0.7/0.92/0.78	00:38	1.62
38	NUDT5A-x0637	5QK5	K3J	CCNC=1C=CC(C)=NN1	137.19	1/2/3/4	4	0.84/0.94/0.86/0.7	00:41	1.63
39	NUDT5A-x0651	5QK6	K3M	CC=1C=CC(C(=O)N)=C(F)C1	153.16	3	1	0.92	00:28	1.51
40	NUDT5A-x0673	5QK7	K3P	CN1C=C(OC=2N=CN=CC2Cl)C=N1	210.62	1/2/3/4	4	0.78/0.96/0.96/0.83	00:29	1.66
41	NUDT5A-x0681	5QK8	K3S	CCNCC1=CN(C)N=N1	140.19	3	1	0.8	00:31	1.71
42	NUDT5A-x0685	5QK9	K3V	COC=1C=CC=2C=NNC(=O)C2C1OC	206.20	1/2/3/4	3	0.82/0.91/0.88/0.91	00:33	1.56
43	NUDT5A-x0692	5QKA	K3Y	CN1C=C(C(=O)N)C(=N1)C2CCCNC2	208.27	1	1	0.8	00:35	1.55

A.1.3. Overview of NUDT5 ^{19}F NMR fragment hits.

#	Data Set #	PDB ID	PDB ID	Smiles string	MW (Da)	Site	# of copies	Occupancy	Soaking time (hh:mm)	Diffraction (Å)
		Crystal	Ligand							
1	NUDT5A-x1072	5QTL	PW7	FC(F)(C1=NC(N)=NC=C1)F	163.10	1/2/3/4	4	0.68/0.8/0.8/0.84	00:25	1.73
2	NUDT5A-x1078	5QTM	PWA	FC(C=C1)=CC(N=CN2)=C1C2=O	164.14	1/2/3/4	4	0.75/0.8/0.88/0.86	00:27	1.79
3	NUDT5A-x1083	5QTN	PWD	FC1=CN=C2NC=CC2=C1	136.13	1/2/4	3	0.74/0.74/0.9	00:29	1.83
4	NUDT5A-x1093	5QTO	PWG	FC(F)(C1=CC=CN2C1=NN=C2C(F)F)F	237.13	1/2/3/4	4	0.76/0.8/0.8/0.82	00:33	1.67
5	NUDT5A-x1211	5QTP	PWJ	ClC(C1=NC=CN1C=C2C(F)(F)F)=C2C#N	245.59	1/2/3/4	4	0.7/0.8/0.87/0.83	00:27	1.73
6	NUDT5A-x1212	5QTQ	PWM	FC1=CC=CC2=NC=NC(O)=C12	164.14	1/2/3/4	4	0.84/0.86/0.86/0.86	00:27	1.82
7	NUDT5A-x1213	5QTR	PWP	FC(C=CC1=NC=N2)=CC1=C2O	164.14	1/2/3/4	4	0.79/0.8/0.8/0.79	00:27	1.55
8	NUDT5A-x1214	5QTS	PWS	CSC(N1C)=NC(C(F)(F)F)=CC1=O	224.20	1/2/3/4	4	0.74/0.74/0.74/0.74	00:28	1.85

A.1.4. Data collection and refinement statistics of NUDT5 XChem hits.

PDB ID	**5QJ4**	**5QJ5**	**5QJ6**	**5QJ7**
PDB ID Ligand	K04	K0G	GQJ	JMM
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1	P 1
Cell: a b c (Å)	49.38 59.83 80.14	49.15 59.77 80.09	49.25 59.78 80.09	49.34 59.79 79.91
Cell: α β γ (°)	79.36 81.67 75.75	79.47 81.89 75.78	79.42 81.6875.64	79.38 81.48 75.77
Resolution range (Å)	78.33 - 1.89 (1.94 - 1.89)	78.34-1.72 (1.76-1.72)	78.30-1.65 (1.69-1.65)	78.08-1.56 (1.60-1.56)
Total reflections	116615 (8581)	155834 (11326)	178103 (11806)	204321 (25390)
Unique reflections	67122 (4949)	88549 (6499)	100263 (7347)	117075 (17002)
Multiplicity	1.70 (1.70)	1.80 (1.70)	1.80 (1.60)	1.70 (1.50)
Completeness (%)	96.40 (95.40)	96.50 (95.60)	96.30 (94.80)	94.90 (94.00)
Mean I/σ(I)	7.50 (-)	9.70 (-)	10.10 (-)	8.80 (-)
R-merge	0.050 (0.557)	0.031 (0.720)	0.038 (0.582)	0.034 (0.981)
R-rim	0.050 (0.557)	0.031 (0.720)	0.038 (0.582)	0.034 (0.981)
CC-half	0.995 (0.297)	0.998 (0.635)	0.997 (0.719)	0.999 (0.422)
R-factor	0.213 (0.340)	0.215 (0.363)	0.212 (0.369)	0.237 (0.410)
R-free	0.253 (0.328)	0.253 (0.347)	0.249 (0.373)	0.274 (0.406)
Number of total atoms	6090	6064	6053	6141
atoms for ligands	85	81	66	105
atoms for waters	254	249	261	279
Number of polymer residues	770	770	770	770
Average B-factor (Å^2)	34.0	38.6	33.7	33.7
B-factor for ligands (Å^2)	54.4	43.7	45.3	40.7
B-factor for solvent (Å^2)	34.9	40.5	37.5	35.5
RMS: bonds (Å)	0.015	0.016	0.018	0.017
RMS: bond angles (°)	1763	1823	1951	1930
RMS: dihedral angles (°)	7233	7383	7177	7184
Ramachandran favoured (%)	96.56	96.69	97.62	96.56
Ramachandran outliers (%)	0.93	0.66	0.53	0.79
Clashscore	5.27	5.54	7.66	7.65

PDB ID	**5QJ8**	**5QJ9**	**5QJA**	**5QJB**
PDB ID Ligand	ELQ	K0D	K0A	K07
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1	P 1
Cell: a b c (Å)	49.53 59.99 79.65	49.59 59.88 79.67	49.44 59.84 80.19	49.60 60.13 79.86
Cell: α β γ (°)	79.56 81.09 75.49	79.45 81.18 75.37	79.33 81.51 75.50	79.31 80.90 75.37
Resolution range (Å)	77.81-1.76 (1.81-1.76)	77.83-1.85 (1.90-1.85)	78.35-1.64 (1.68-1.64)	77.93-1.66 (1.70-1.66)
Total reflections	149802 (11257)	124530 (9217)	176125 (11333)	172357 (11489)
Unique reflections	83225 (6086)	71290 (5223)	102140 (7345)	98938 (7225)
Multiplicity	1.80 (1.80)	1.70 (1.80)	1.70 (1.50)	1.70 (1.60)
Completeness (%)	96.80 (96.00)	96.50 (95.40)	95.90 (93.80)	96.00 (94.00)
Mean I/σ(I)	8.60 (-)	7.10 (-)	12.20 (-)	9.60 (-)
R-merge	0.043 (0.699)	0.048 (0.573)	0.026 (0.407)	0.031 (0.596)
R-rim	0.043 (0.699)	0.048 (0.573)	0.026 (0.407)	0.031 (0.596)
CC-half	0.996 (0.687)	0.997 (0.674)	0.997 (0.786)	0.998 (0.693)
R-factor	0.234 (0.383)	0.242 (0.375)	0.225 (0.367)	0.250 (0.385)
R-free	0.278 (0.408)	0.288 (0.350)	0.257 (0.373)	0.291 (0.403)
Number of total atoms	5997	6060	6119	6070
atoms for ligands	34	81	85	66
atoms for waters	231	247	275	256
Number of polymer residues	770	770	770	769
Average B-factor (Å^2)	40.6	37.7	32.4	40.1
B-factor for ligands (Å^2)	55.4	49.6	38.6	50.1
B-factor for solvent (Å^2)	39.3	35.8	34.2	39.0
RMS: bonds (Å)	0.016	0.015	0.016	0.013
RMS: bond angles (°)	1892	1787	1861	1671
RMS: dihedral angles (°)	7424	7236	6974	6963
Ramachandran favoured (%)	94.58	95.37	96.96	95.62
Ramachandran outliers (%)	1.06	0.79	0.53	1.06
Clashscore	12.44	10.60	3.79	13.58

PDB ID	**5QJC**	**5QJD**	**5QJE**	**5QJF**
PDB ID Ligand	K0J	K0M	K0P	K0S
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1	P 1
Cell: a b c (Å)	49.26 59.77 80.26	48.73 59.59 80.19	48.97 59.71 79.93	48.39 59.42 79.92
Cell: α β γ (°)	79.25 81.78 75.66	79.65 82.00 76.23	79.66 81.83 75.99	79.85 82.34 76.55
Resolution range (Å)	78.45-1.47 (1.51-1.47)	78.48-1.61 (1.65-1.61)	78.21-1.75 (1.79-1.75)	78.31-1.63 (1.67-1.63)
Total reflections	232452 (26446)	179910 (11312)	143103 (10747)	172182 (24143)
Unique reflections	137149 (18887)	105862 (7621)	82932 (6017)	96807 (14701)
Multiplicity	1.70 (1.40)	1.70 (1.50)	1.70 (1.80)	1.80 (1.60)
Completeness (%)	93.60 (88.30)	95.30 (92.80)	95.60 (94.90)	91.10 (94.60)
Mean I/σ(I)	8.50 (-)	9.80 (-)	10.50 (-)	9.70 (-)
R-merge	0.027 (0.960)	0.033 (0.595)	0.030 (0.434)	0.025 (1.135)
R-rim	0.027 (0.959)	0.033 (0.595)	0.030 (0.434)	0.025 (1.134)
CC-half	0.998 (0.496)	0.994 (0.672)	0.999 (0.794)	0.999 (0.469)
R-factor	0.221 (0.417)	0.210 (0.335)	0.210 (0.339)	0.214 (0.431)
R-free	0.254 (0.414)	0.237 (0.336)	0.245 (0.346)	0.249 (0.430)
Number of total atoms	6097	6198	6099	6069
atoms for ligands	53	59	67	80
atoms for waters	290	319	269	246
Number of polymer residues	770	770	770	770
Average B-factor (Å^2)	33.6	44073	34.9	40.4
B-factor for ligands (Å^2)	38.7	43.6	47.2	65.4
B-factor for solvent (Å^2)	37.5	34.4	36.9	44.8
RMS: bonds (Å)	0.018	0.011	0.017	0.017
RMS: bond angles (°)	1979	1569	1907	1866
RMS: dihedral angles (°)	7188	6439	7066	7323
Ramachandran favoured (%)	97.09	97.49	97.09	97.22
Ramachandran outliers (%)	0.66	0.40	0.40	0.53
Clashscore	6.21	7.92	4.89	5.99

PDB ID **PDB ID Ligand**	**5QJG** K0V	**5QJH** K0Y	**5QJI** K1A	**5QJG** 6SU
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1	P 1
Cell: a b c (Å)	48.85 59.53 80.12	49.01 59.63 79.83	48.98 59.77 79.38	48.86 59.60 79.86
Cell: α β γ (°)	79.71 82.19 76.21	79.41 81.64 75.79	79.84 81.43 75.99	79.70 81.81 76.08
Resolution range (Å)	78.46-1.57 (1.61-1.57)	78.03-2.03 (2.08-2.03)	77.65-1.69 (1.73-1.69)	78.14-1.71 (1.75-1.71)
Total reflections	384331 (15171)	93805 (6534)	302983 (11594)	155108 (11088)
Unique reflections	115731 (8335)	53231 (3867)	92935 (6722)	88844 (6544)
Multiplicity	3.30 (1.80)	1.80 (1.70)	3.30 (1.70)	1.70 (1.70)
Completeness (%)	96.50 (93.90)	96.20 (94.80)	97.10 (95.50)	96.00 (94.90)
Mean I/σ(I)	11.40 (-)	7.00 (-)	12.30 (-)	12.40 (-)
R-merge	0.043 (0.537)	0.076 (0.522)	0.048 (0.643)	0.030 (0.520)
R-rim	0.027 (0.446)	0.076 (0.522)	0.030 (0.643)	0.030 (0.520)
CC-half	0.998 (0.713)	0.990 (0.746)	0.998 (0.638)	0.997 (0.689)
R-factor	0.206 (0.330)	0.224 (0.335)	0.257 (0.429)	0.214 (0.378)
R-free	0.233 (0.356)	0.273 (0.374)	0.301 (0.438)	0.247 (0.368)
Number of total atoms	6215	5980	6010	6072
atoms for ligands	101	35	51	59
atoms for waters	323	225	229	251
Number of polymer residues	770	770	770	770
Average B-factor (Å^2)	32.6	39.9	41.9	35.8
B-factor for ligands (Å^2)	36.2	48.0	51.0	56.5
B-factor for solvent (Å^2)	35.6	40.0	41.4	37.7
RMS: bonds (Å)	0.014	0.014	0.015	0.016
RMS: bond angles (°)	1741	1665	1865	1791
RMS: dihedral angles (°)	6413	7312	7560	7056
Ramachandran favoured (%)	98.02	96.69	95.11	96.83
Ramachandran outliers (%)	0.66	0.53	0.40	0.40
Clashscore	4.87	6.12	7.27	3.36

PDB ID **PDB ID Ligand**	**5QJK** K1D	**5QJL** JJM	**5QJM** K1G	**5QJN** K1J
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1	P 1
Cell: a b c (Å)	48.96 59.77 80.22	48.89 59.51 79.98	49.64 59.95 80.27	49.45 59.88 79.89
Cell: α β γ (°)	79.30 81.69 75.95	79.68 82.07 76.20	79.09 81.28 75.44	79.29 81.26 75.44
Resolution range (Å)	78.39-1.65 (1.69-1.65)	78.29-1.66 (1.70-1.66)	78.34-1.75 (1.79-1.75)	78.02-1.77 (1.82-1.77)
Total reflections	341941 (19946)	174437 (11848)	138886 (22794)	142252 (10038)
Unique reflections	100346 (7328)	97841 (7176)	77189 (12283)	80967 (5843)
Multiplicity	3.40 (2.70)	1.8 (1.70)	1.80 (1.90)	1.80 (1.70)
Completeness (%)	96.70 (94.60)	96.60 (95.10)	87.30 (94.80)	95.90 (93.60)
Mean I/σ(I)	11.20 (-)	9.70 (-)	6.20 (-)	8.10 (-)
R-merge	0.048 (0.580)	0.041 (0.690)	0.052 (0.880)	0.070 (0.716)
R-rim	0.031 (0.412)	0.041 (0.690)	0.052 (0.880)	0.070 (0.716)
CC-half	0.995 (0.775)	0.996 (0.612)	0.995 (0.471)	0.987 (0.589)
R-factor	0.207 (0.301)	0.208 (0.348)	0.223 (0.376)	0.229 (0.379)
R-free	0.237 (0.349)	0.241 (0.346)	0.271 (0.389)	0.273 (0.389)
Number of total atoms	6130	6195	6090	6043
atoms for ligands	85	109	81	57
atoms for waters	281	289	235	245
Number of polymer residues	770	770	770	769
Average B-factor (Å^2)	31.0	34.5	37.4	38.2
B-factor for ligands (Å^2)	44074	43.2	43.6	47.1
B-factor for solvent (Å^2)	35.4	37.5	39.3	39.0
RMS: bonds (Å)	0.015	0.018	0.016	0.016
RMS: bond angles (°)	1822	1958	1809	1858
RMS: dihedral angles (°)	6696	7022	6924	7338
Ramachandran favoured (%)	97.75	97.49	97.62	96.30
Ramachandran outliers (%)	0.53	0.53	0.66	1.06
Clashscore	5.43	6.57	8.24	7.12

PDB ID	**5QJO**	**5QJP**	**5QJQ**	**5QJR**
PDB ID Ligand	K41	K1P	K1S	K1V
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1	P 1
Cell: a b c (Å)	49.06 59.57 79.93	49.13 59.63 79.62	48.92 59.63 79.84	49.54 60.03 79.99
Cell: α β γ (°)	79.55 81.90 75.89	79.86 81.64 76.07	79.67 82.01 76.26	79.25 81.06 75.33
Resolution range (Å)	78.20-1.98 (2.03-1.98)	77.92-1.50 (1.54-1.50)	78.14-1.55 (1.59-1.55)	78.08-1.62 (1.66-1.62)
Total reflections	100312 (6748)	203554 (26603)	200660 (12199)	188829 (12106)
Unique reflections	57683 (4157)	119825 (18692)	119114 (8696)	106603 (7744)
Multiplicity	1.70 (1.60)	1.70 (1.40)	1.70 (1.40)	1.80 (1.60)
Completeness (%)	96.60 (94.60)	88.00 (93.80)	95.50 (93.60)	96.30 (94.50)
Mean I/σ(I)	7.10 (-)	7.60 (-)	11.20 (-)	9.00 (-)
R-merge	0.070 (0.417)	0.031 (1.358)	0.025 (0.389)	0.054 (0.666)
R-rim	0.070 (0.417)	0.031 (1.358)	0.025 (0.389)	0.054 (0.665)
CC-half	0.964 (0.643)	0.998 (0.399)	0.998 (0.814)	0.989 (0.632)
R-factor	0.210 (0.292)	0.223 (0.415)	0.211 (0.355)	0.242 (0.408)
R-free	0.249 (0.309)	0.255 (0.443)	0.240 (0.371)	0.277 (0.403)
Number of total atoms	6111	6100	6177	6026
atoms for ligands	69	69	89	49
atoms for waters	264	262	288	245
Number of polymer residues	770	770	770	770
Average B-factor (Å^{2})	32.9	37.7	32.3	36.2
B-factor for ligands (Å^{2})	39.3	49.5	33.2	44.3
B-factor for solvent (Å^{2})	33.8	41.5	36.7	37.4
RMS: bonds (Å)	0.015	0.016	0.016	0.015
RMS: bond angles (°)	1730	1867	1834	1753
RMS: dihedral angles (°)	6900	6880	6866	7017
Ramachandran favoured (%)	97.62	97.09	97.09	95.90
Ramachandran outliers (%)	0.53	0.53	0.66	0.93
Clashscore	4.34	4.78	6.82	9.02

PDB ID **PDB ID Ligand**	**5QJS** K1Y	**5QJT** K24	**5QJU** K2G	**5QJV** K2J
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1	P 1
Cell: a b c (Å)	49.46 59.94 79.92	49.45 59.87 80.23	48.92 59.65 80.05	48.68 59.55 79.79
Cell: α β γ (°)	79.53 81.32 75.46	79.16 81.53 75.51	79.55 81.88 76.15	79.84 82.09 76.39
Resolution range (Å)	78.11-1.58 (1.62-1.58)	78.36-1.62 (1.66-1.62)	78.31-1.77 (1.81-1.76)	78.15-1.61 (1.65-1.61)
Total reflections	198056 (25177)	186282 (24949)	144424 (21533)	182950 (11526)
Unique reflections	113202 (16464)	104799 (15343)	80398 (11699)	105949 (7636)
Multiplicity	1.70 (1.50)	1.80 (1.60)	1.80 (1.80)	1.70 (1.50)
Completeness (%)	94.80 (94.20)	94.80 (94.60)	95.10 (94.50)	96.00 (93.40)
Mean $I/\sigma(I)$	8.30 (-)	9.30 (-)	7.50 (-)	7.80 (-)
R-merge	0.032 (1.027)	0.032 (0.782)	0.045 (0.821)	0.042 (0.593)
R-rim	0.032 (1.027)	0.032 (0.781)	0.045 (0.821)	0.042 (0.593)
CC-half	0.998 (0.392)	0.998 (0.533)	0.997 (0.483)	0.994 (0.620)
R-factor	0.235 (0.445)	0.220 (0.366)	0.210 (0.366)	0.213 (0.360)
R-free	0.272 (0.485)	0.251 (0.380)	0.248 (0.398)	0.247 (0.380)
Number of total atoms	6016	6058	6106	6091
atoms for ligands	42	45	81	51
atoms for waters	250	258	249	260
Number of polymer residues	770	770	770	770
Average B-factor (Å^2)	37.6	34.1	38.0	35.2
B-factor for ligands (Å^2)	48.9	43.2	47.8	43.3
B-factor for solvent (Å^2)	38.7	37.8	41.3	39.6
RMS: bonds (Å)	0.017	0.016	0.016	0.015
RMS: bond angles (°)	1905	1824	1745	1762
RMS: dihedral angles (°)	6909	6927	6832	6650
Ramachandran favoured (%)	96.69	97.49	97.09	98.41
Ramachandran outliers (%)	0.79	0.53	0.66	0.66
Clashscore	9.99	6.81	6.87	7.40

PDB ID	**5QJW**	**5QJX**	**5QJY**	**5QJZ**
PDB ID Ligand	K2M	K2P	K2S	K2V
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1	P 1
Cell: a b c (Å)	49.53 59.99 80.05	48.52 59.52 80.62	49.51 59.92 80.14	49.43 59.82 79.90
Cell: α β γ (°)	79.16 81.16 75.31	79.68 82.36 76.36	79.30 81.43 75.55	79.61 81.70 75.99
Resolution range (Å)	78.13-1.57 (1.61-1.57)	78.96-1.73 (1.77-1.73)	78.29-1.77 (1.81-1.77)	78.15-1.52 (1.56-1.52)
Total reflections	201268 (25283)	154603 (11599)	113492 (21113)	215289 (26443)
Unique reflections	115029 (16672)	86383 (6385)	63650 (11550)	125392 (18365)
Multiplicity	1.70 (1.50)	1.80 (1.80)	1.80 (1.80)	1.70 (1.40)
Completeness (%)	94.70 (94.00)	96.40 (95.80)	74.50 (92.60)	94.10 (94.20)
Mean I/σ(I)	10.00 (-)	8.90 (-)	10.60 (-)	10.70 (-)
R-merge	0.030 (0.945)	0.051 (0.623)	0.028 (0.876)	0.019 (1.258)
R-rim	0.030 (0.944)	0.051 (0.623)	0.028 (0.876)	0.019 (1.258)
CC-half	0.998 (0.515)	0.993 (0.699)	0.999 (0.493)	0.999 (0.401)
R-factor	0.232 (0.374)	0.220 (0.367)	0.214 (0.377)	0.226 (0.441)
R-free	0.267 (0.392)	0.256 (0.379)	0.259 (0.398)	0.263 (0.460)
Number of total atoms	6067	5985	6012	6010
atoms for ligands	73	35	64	51
atoms for waters	250	226	222	231
Number of polymer residues	770	770	770	770
Average B-factor (Å^2)	34.7	37.9	40.6	39.8
B-factor for ligands (Å^2)	45.4	60.2	50.0	48.8
B-factor for solvent (Å^2)	36.6	42.9	40.9	42.0
RMS: bonds (Å)	0.017	0.018	0.015	0.018
RMS: bond angles (°)	1983	1890	1777	1922
RMS: dihedral angles (°)	7172	7085	7483	6864
Ramachandran favoured (%)	96.56	97.75	96.56	96.83
Ramachandran outliers (%)	0.93	0.53	0.53	0.79
Clashscore	8.89	5.79	4.40	9.47

PDB ID	**5QK0**	**5QK1**	**5QK2**	**5QK3**
PDB ID Ligand	K31	K34	JH4	K3A
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1	P 1
Cell: a b c (Å)	49.67 60.15 79.78	49.30 59.89 80.36	48.51 59.47 80.09	49.49 60.02 80.13
Cell: α β γ (°)	79.37 80.65 75.23	79.21 81.58 75.63	79.55 82.12 76.18	79.15 81.35 75.43
Resolution range (Å)	77.83-1.44 (1.48-1.44)	78.50-1.49 (1.53-1.49)	78.38-1.65 (1.69-1.65)	78.23-1.71 (1.75-1.71)
Total reflections	216147 (21832)	225572 (26641)	129272 (15361)	163557 (11727)
Unique reflections	130398 (16151)	132526 (18829)	77041 (10297)	91119 (6665)
Multiplicity	1.70 (1.40)	1.70 (1.40)	1.70 (1.50)	1.80 (1.80)
Completeness (%)	82.50 (70.00)	93.50 (91.00)	75.20 (68.70)	96.80 (95.90)
Mean I/σ(I)	6.30 (-)	7.50 (-)	7.50 (-)	5.60 (-)
R-merge	0.041 (1.119)	0.033 (1.164)	0.041 (1.305)	0.090 (0.615)
R-rim	0.041 (1.119)	0.033 (1.163)	0.041 (1.305)	0.090 (0.614)
CC-half	0.995 (0.464)	0.997 (0.484)	0.997 (0.358)	0.976 (0.667)
R-factor	0.306 (0.514)	0.218 (0.410)	0.223 (0.410)	0.228 (0.357)
R-free	0.338 (0.476)	0.252 (0.444)	0.267 (0.412)	0.266 (0.363)
Number of total atoms	5990	6128	6014	5999
atoms for ligands	53	69	33	31
atoms for waters	221	275	248	246
Number of polymer residues	770	770	770	770
Average B-factor (Å^2)	40.1	34.5	33.7	35.1
B-factor for ligands (Å^2)	47.3	41.8	39.0	41.7
B-factor for solvent (Å^2)	39.5	38.4	37.1	36.6
RMS: bonds (Å)	0.006	0.019	0.015	0.017
RMS: bond angles (°)	0.972	1999	1772	1884
RMS: dihedral angles (°)	4950	7037	7077	7141
Ramachandran favoured (%)	96.30	97.49	97.22	97.22
Ramachandran outliers (%)	0.66	0.66	0.79	0.66
Clashscore	3.94	5.72	8.01	8.15

PDB ID	**5QK4**	**5QK5**	**5QK6**	**5QK7**
PDB ID Ligand	K3D	K3J	K3M	K3P
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1	P 1
Cell: a b c (Å)	49.41 59.93 80.08	49.44 59.96 79.98	49.26 59.80 80.31	48.58 59.46 80.40
Cell: α β γ (°)	79.27 81.33 75.46	79.36 81.22 75.51	79.26 81.50 75.67	79.39 82.31 76.19
Resolution range (Å)	78.20-1.62 (1.66-1.62)	78.11-1.63 (1.67-1.63)	78.45-1.51 (1.55-1.51)	78.68-1.66 (1.71-1.66)
Total reflections	187778 (24775)	185217 (11897)	223045 (13175)	155480 (23708)
Unique reflections	105652 (15367)	104461 (7608)	130552 (9448)	88015 (13924)
Multiplicity	1.80 (1.60)	1.80 (1.60)	1.70 (1.40)	1.80 (1.70)
Completeness (%)	95.20 (94.40)	96.30 (94.00)	95.80 (93.60)	87.70 (94.80)
Mean I/σ(I)	6.90 (-)	6.20 (-)	8.60 (-)	7.50 (-)
R-merge	0.059 (0.852)	0.083 (0.684)	0.069 (0.682)	0.036 (1.245)
R-rim	0.059 (0.851)	0.083 (0.681)	0.069 (0.681)	0.036 (1.245)
CC-half	0.991 (0.543)	0.985 (0.560)	0.982 (0.511)	0.998 (0.389)
R-factor	0.223 (0.390)	0.231 (0.398)	0.227 (0.358)	0.217 (0.412)
R-free	0.255 (0.400)	0.266 (0.410)	0.255 (0.372)	0.255 (0.395)
Number of total atoms	6113	6086	6054	6095
atoms for ligands	67	69	44	81
atoms for waters	276	270	269	249
Number of polymer residues	770	770	770	770
Average B-factor (Å²)	34.4	36.1	43921	39.2
B-factor for ligands (Å²)	44.5	39.7	38.4	51.5
B-factor for solvent (Å²)	35.7	37.3	35.4	41.9
RMS: bonds (Å)	0.018	0.017	0.017	0.016
RMS: bond angles (°)	1962	1898	1877	1768
RMS: dihedral angles (°)	7129	7123	6823	7110
Ramachandran favoured (%)	97.49	96.56	97.62	96.69
Ramachandran outliers (%)	0.66	0.66	0.40	0.66
Clashscore	6.16	7.85	6.78	5.58

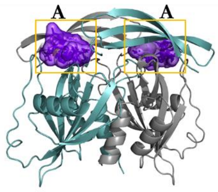
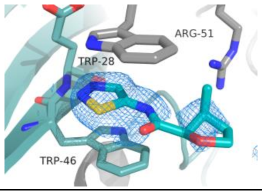
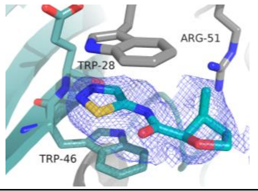
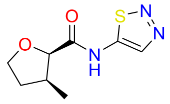
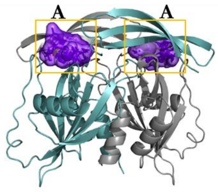
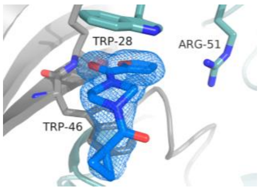
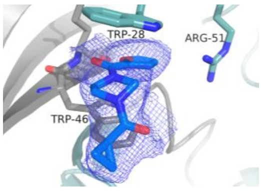
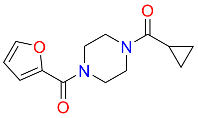
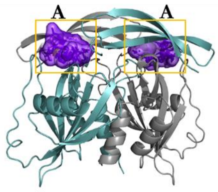
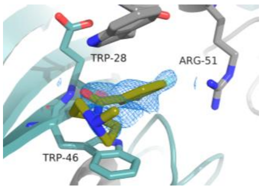
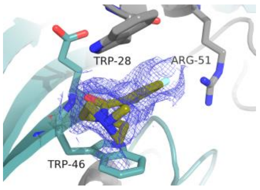
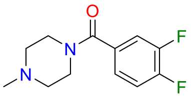
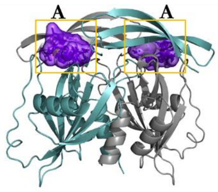
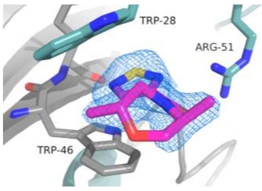
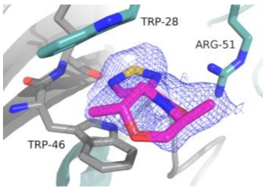
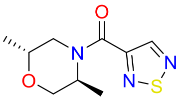
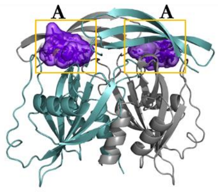
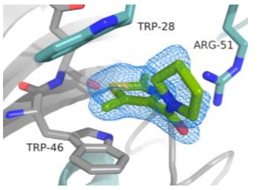
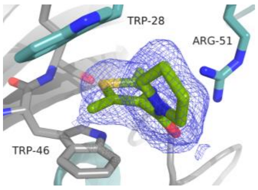
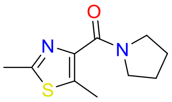
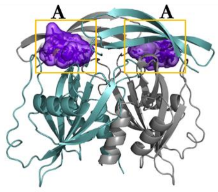
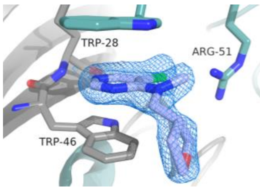
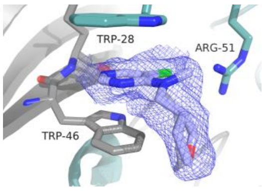
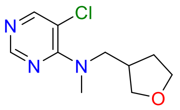
PDB ID **PDB ID Ligand**	**5QK8** K3S	**5QK9** K3V	**5QKA** K3Y
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1
Cell: a b c (Å)	49.27 59.91 80.01	49.33 59.84 80.49	49.75 60.12 79.52
Cell: α β γ (°)	79.55 81.38 75.74	79.07 81.63 75.64	79.24 81.03 75.23
Resolution range (Å)	78.20-1.71 (1.75-1.71)	78.61-1.56 (1.60-1.56)	77.61-1.55 (1.58-1.54)
Total reflections	153246 (10747)	196670 (25656)	205200 (25919)
Unique reflections	89368 (6536)	113723 (17205)	118439 (17553)
Multiplicity	1.70 (1.60)	1.70 (1.50)	1.70 (1.50)
Completeness (%)	95.40 (94.20)	91.50 (94.20)	92.90 (94.30)
Mean I/σ(I)	11.70 (-)	9.00 (-)	9.30 (-)
R-merge	0.026 (0.416)	0.033 (0.904)	0.024 (0.933)
R-rim	0.026 (0.416)	0.033 (0.904)	0.024 (0.932)
CC-half	0.999 (0.811)	0.998 (0.441)	0.999 (0.513)
R-factor	0.225 (0.360)	0.214 (0.393)	0.245 (0.425)
R-free	0.260 (0.335)	0.251 (0.404)	0.286 (0.419)
Number of total atoms	5994	6170	5986
atoms for ligands	35	89	40
atoms for waters	239	287	230
Number of polymer residues	770	770	770
Average B-factor (Å^2)	34.6	33.2	41.5
B-factor for ligands (Å^2)	43.2	42.0	49.9
B-factor for solvent (Å^2)	36.7	36.2	42.0
RMS: bonds (Å)	0.019	0.018	0.016
RMS: bond angles (°)	1981	1881	1811
RMS: dihedral angles (°)	7257	7014	7422
Ramachandran favoured (%)	97.09	97.49	95.63
Ramachandran outliers (%)	0.79	0.53	0.93
Clashscore	8.07	4.28	5.48

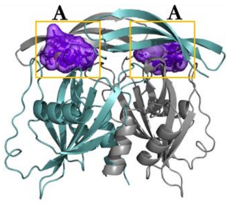
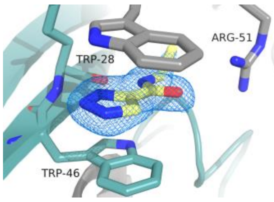
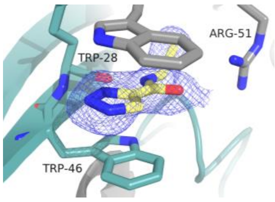
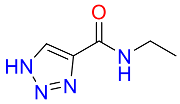
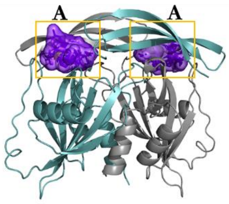
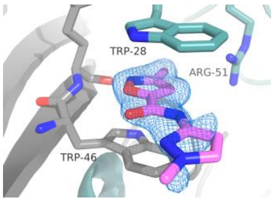
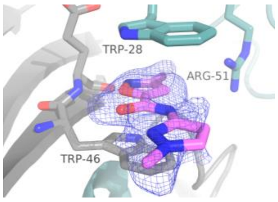
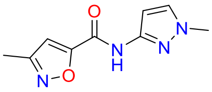
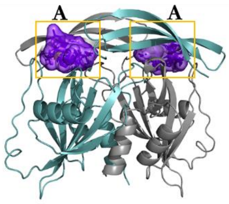
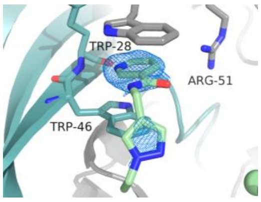
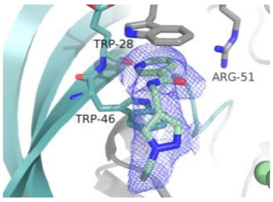
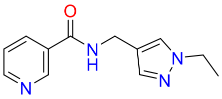
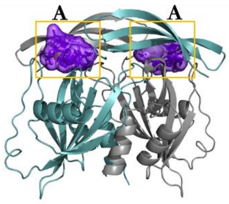
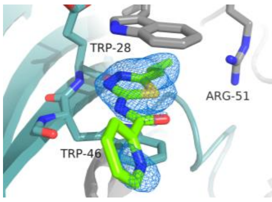
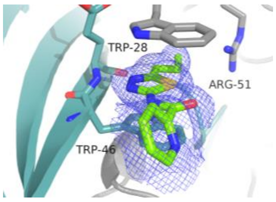
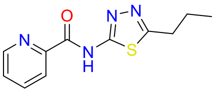
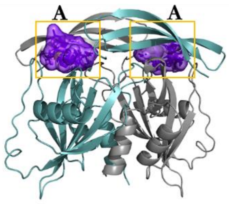
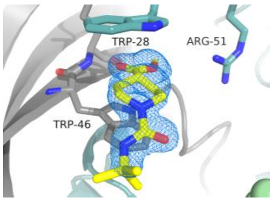
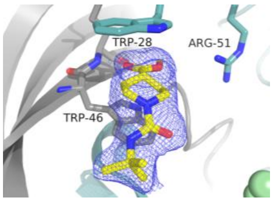
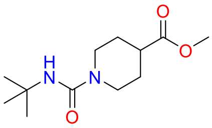
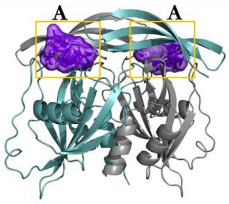
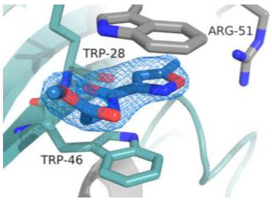
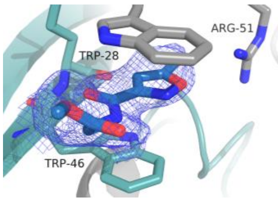
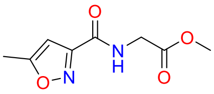
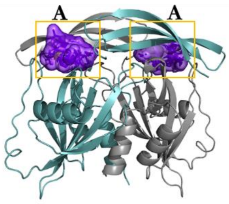
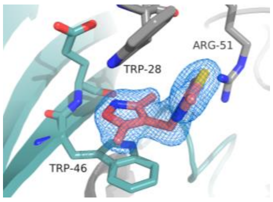
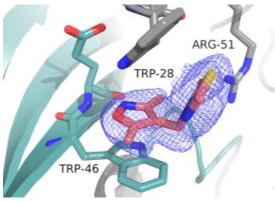
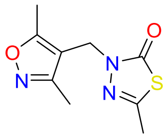
A.1.5. Data collection and refinement statistics of NUDT5 ^{19}F NMR fragment hits.

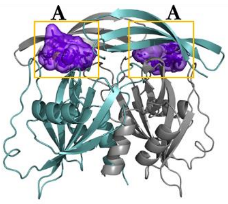
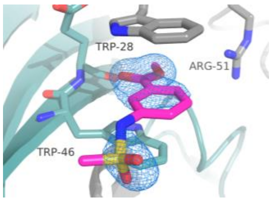
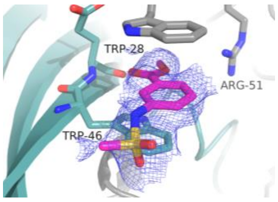
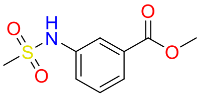
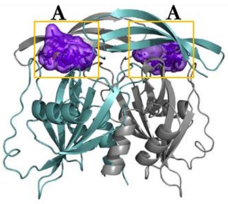
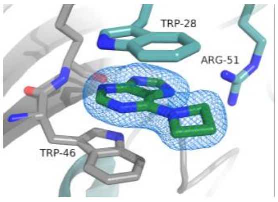
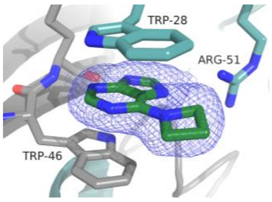
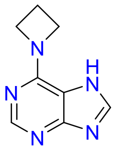
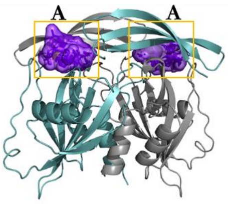
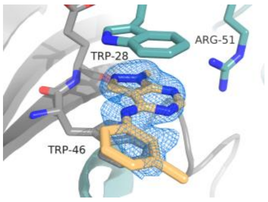
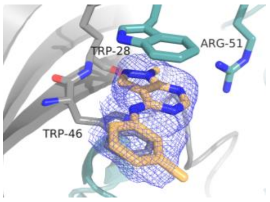
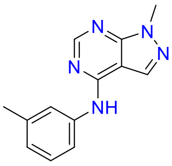
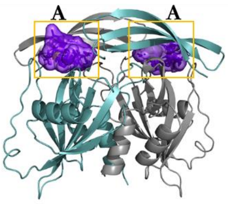
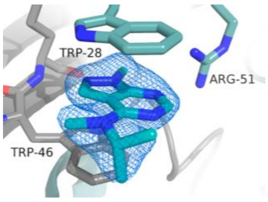
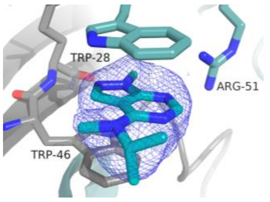
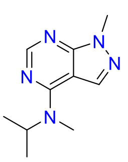
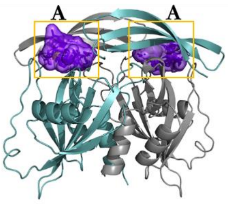
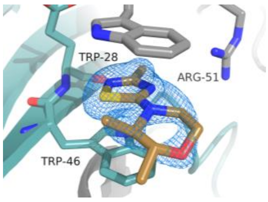
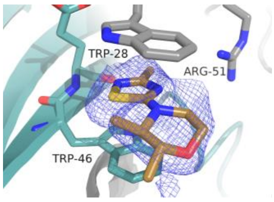
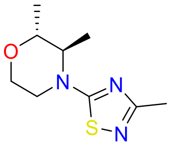
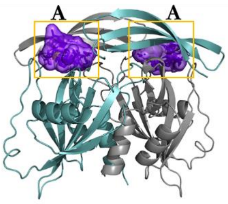
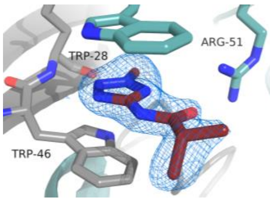
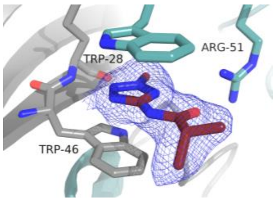
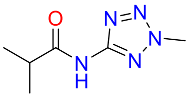
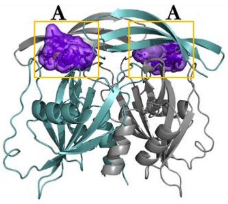
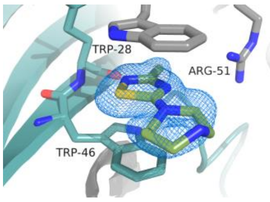
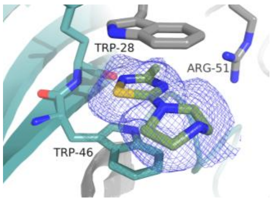
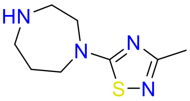
PDB ID **PDB ID Ligand**	**5QTL** PW7	**5QTM** PWA	**5QTN** PWD	**5QTO** PWG
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1	P 1
Cell: a b c (Å)	49.33 59.82 80.18	49.27 59.81 80.07	49.08 59.81 79.59	49.26 59.87 79.89
Cell: α β γ (°)	79.34 81.47 75.65	79.33 81.48 75.66	79.55 81.29 75.78	79.47 81.46 75.59
Resolution range (Å)	78.34 - 1.73 (1.77 - 1.73)	47.50 - 1.79 (1.83 - 1.79)	49.91 - 1.83 (1.88 - 1.83)	41.52 - 1.67 (1.71 - 1.67)
Total reflections	155169 (10161)	137971 (9624)	128429 (9618)	166019 (21069)
Unique reflections	87469 (6388)	78713 (5770)	72935 (5388)	95026 (13864)
Multiplicity	1.80 (1.60)	1.80 (1.70)	1.80 (1.80)	1.70 (1.50)
Completeness (%)	96.60 (95.20)	96.50 (95.60)	96.50 (95.60)	94.20 (93.90)
Mean I/σ(I)	8.30 (-)	8.80 (-)	6.30 (-)	9.50 (-)
R-merge	0.068 (0.726)	0.047 (0.966)	0.056 (0.494)	0.033 (0.885)
R-rim	0.068 (0.726)	0.047 (0.966)	0.056 (0.494)	0.033 (0.885)
CC-half	0.986 (0.639)	0.997 (0.322)	0.979 (0.723)	0.999 (0.432)
R-factor	0.221 (0.358)	0.214 (0.357)	0.238 (0.374)	0.229 (0.455)
R-free	0.253 (0.361)	0.247 (0.377)	0.275 (0.370)	0.262 (0.459)
Number of total atoms	6223	6182	6070	6144
atoms for ligands	77	89	51	93
atoms for waters	288	275	230	256
Number of polymer residues	776	774	775	774
Average B-factor (Å^2)	36.4	32.5	39.1	38.3
B-factor for ligands (Å^2)	43.2	37.8	42.8	48.7
B-factor for solvent (Å^2)	38.1	34.9	38.7	38.3
RMS: bonds (Å)	0.008	0.008	0.007	0.008
RMS: bond angles (°)	1.502	1.477	1.474	1.518
RMS: dihedral angles (°)	7.296	7.531	7.512	7.547
Ramachandran favoured (%)	95.95	97.08	95.63	96.15
Ramachandran outliers (%)	0.78	0.53	0.53	0.40
Clashscore	8.29	6.22	6.52	6.12

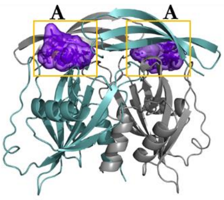
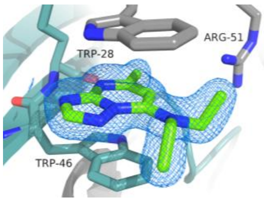
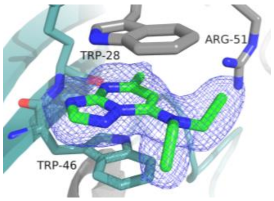
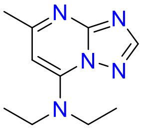
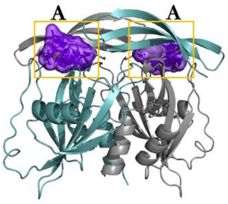
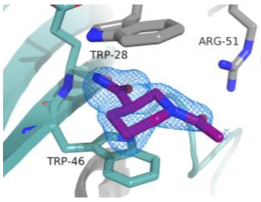
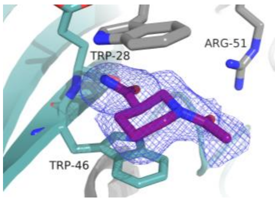
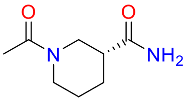
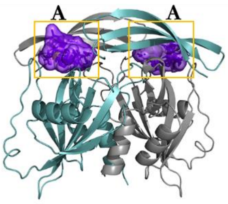
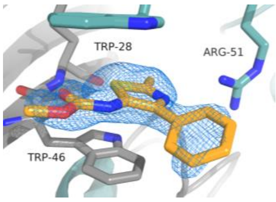
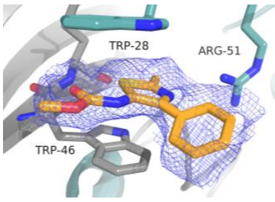
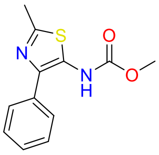
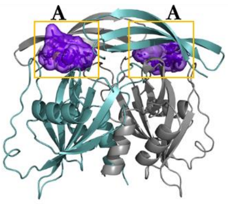
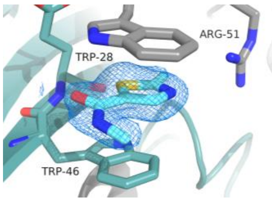
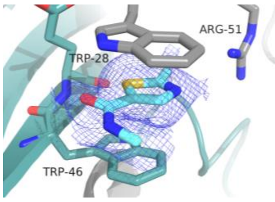
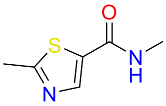
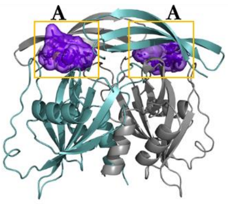
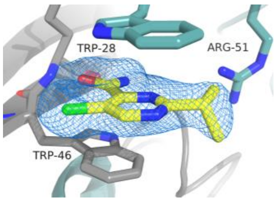
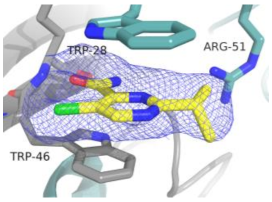
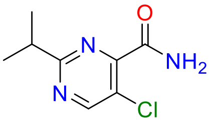
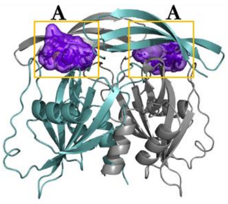
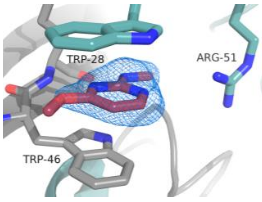
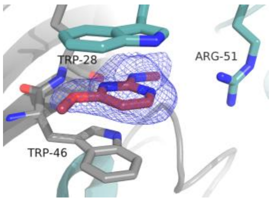
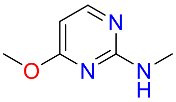
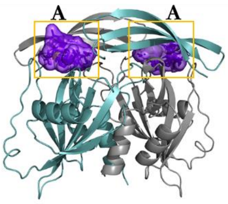
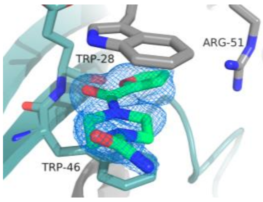
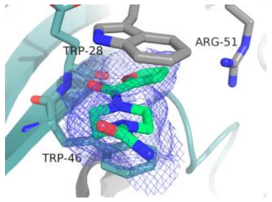
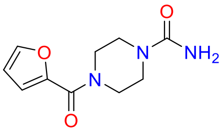
PDB ID **PDB ID Ligand**	**5QTP** PWJ	**5QTQ** PWM	**5QTR** PWP	**5QTS** PWS
X-ray source	I04-1, DLS	I04-1, DLS	I04-1, DLS	I04-1, DLS
Wavelength (Å)	0.9159	0.9159	0.9159	0.9159
Space group	P 1	P 1	P 1	P 1
Cell: a b c (Å)	49.25 59.83 80.05	49.00 59.70 80.19	48.99 59.76 80.34	49.31 59.91 80.08
Cell: α β γ (°)	79.42 81.55 75.70	79.38 81.67 75.86	79.35 81.73 75.91	79.34 81.27 75.59
Resolution range (Å)	42.67 - 1.73 (1.77 - 1.73)	47.30 - 1.82 (1.87 - 1.82)	35.06 - 1.55 (1.59 - 1.55)	50.11 - 1.85 (1.90 - 1.85)
Total reflections	151077 (9788)	130622 (9595)	195220 (10119)	125372 (9561)
Unique reflections	87066 (6333)	74407 (5484)	118554 (7390)	71454 (5283)
Multiplicity	1.70 (1.50)	1.80 (1.70)	1.60 (1.40)	1.80 (1.80)
Completeness (%)	96.30 (94.50)	96.40 (95.80)	94.40 (79.00)	96.40 (95.50)
Mean I/σ(I)	8.90 (-)	8.80 (-)	12.20 (-)	9.60 (-)
R-merge	0.041 (0.659)	0.043 (0.567)	0.023 (0.454)	0.040 (0.718)
R-rim	0.041 (0.659)	0.043 (0.567)	0.023 (0.454)	0.040 (0.718)
CC-half	0.991 (0.461)	0.997 (0.651)	0.999 (0.783)	0.998 (0.643)
R-factor	0.229 (0.352)	0.211 (0.347)	0.198 (0.336)	0.216 (0.364)
R-free	0.261 (0.381)	0.250 (0.345)	0.224 (0.328)	0.251 (0.351)
Number of total atoms	6177	6228	6366	6191
atoms for ligands	105	81	84	97
atoms for waters	258	287	406	265
Number of polymer residues	775	776	776	775
Average B-factor (Å^2)	33.6	33.0	29.9	35.9
B-factor for ligands (Å^2)	42.3	36.2	35.8	44.8
B-factor for solvent (Å^2)	35.4	34.0	35.5	34.7
RMS: bonds (Å)	0.009	0.008	0.010	0.009
RMS: bond angles (°)	1.615	1.501	1.606	1.524
RMS: dihedral angles (°)	7.530	7.272	7.163	7.413
Ramachandran favoured (%)	96.16	97.48	96.74	95.75
Ramachandran outliers (%)	0.66	0.40	0.78	0.53
Clashscore	7.27	6.89	10.60	5.64

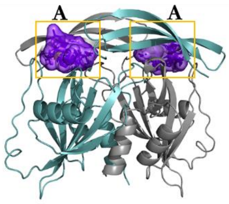
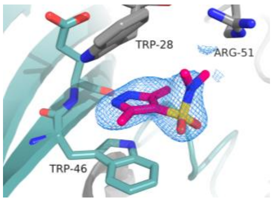
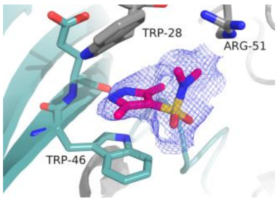
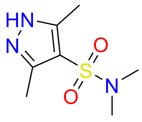
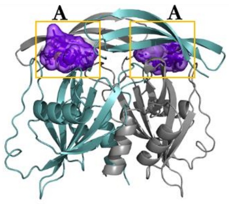
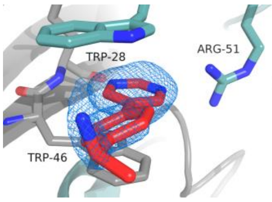
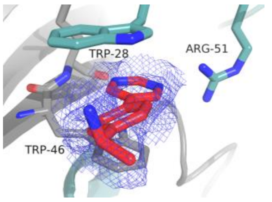
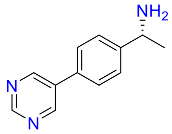
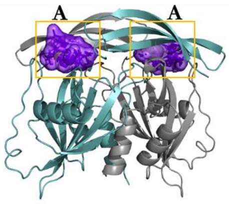
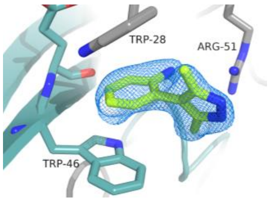
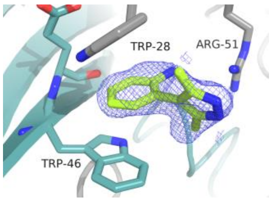
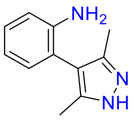
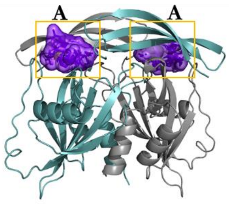
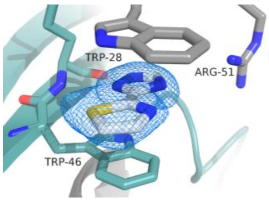
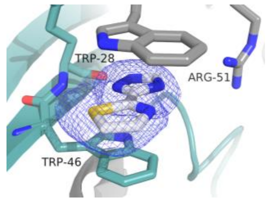
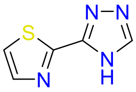
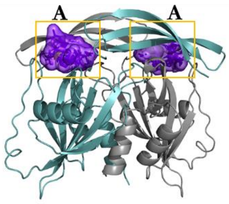
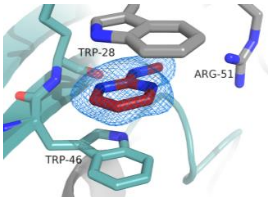
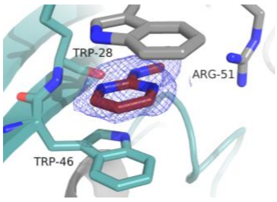
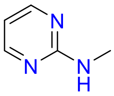
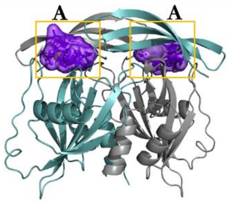
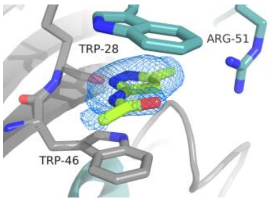
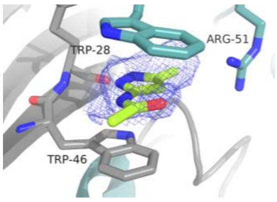
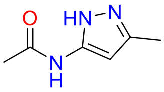
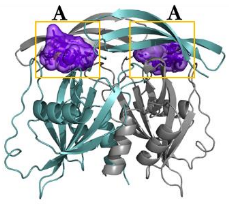
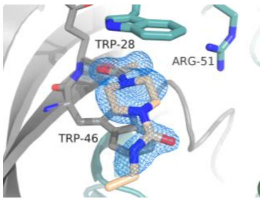
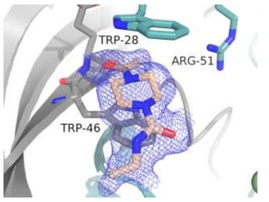
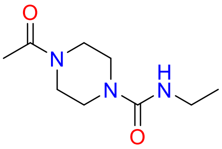
A.1.6. NUDT5 XChem fragment hits.

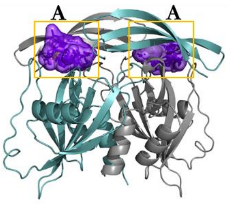
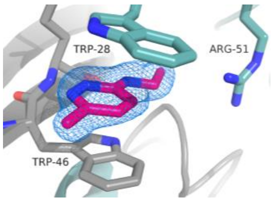
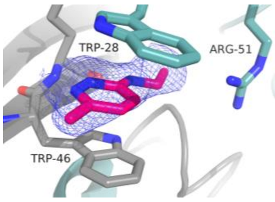
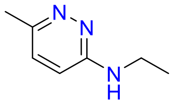
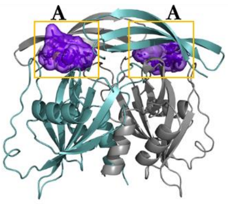
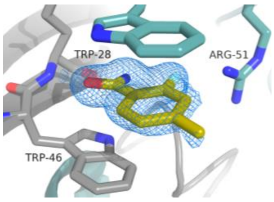
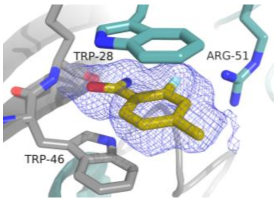
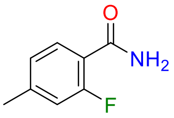
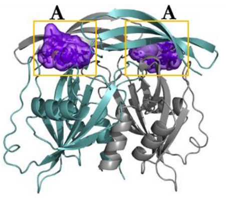
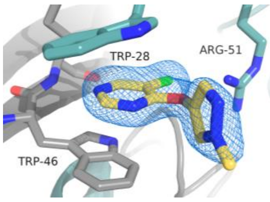
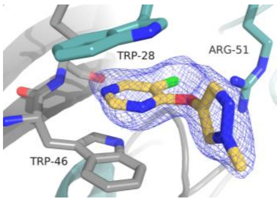
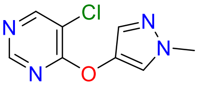
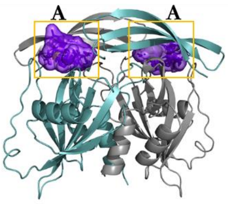
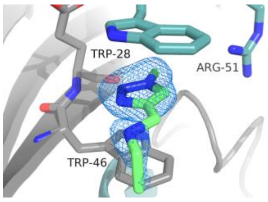
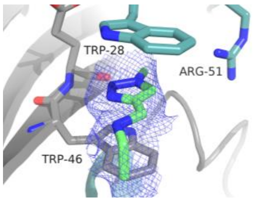
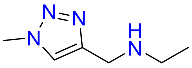
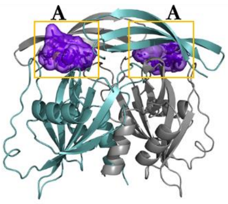
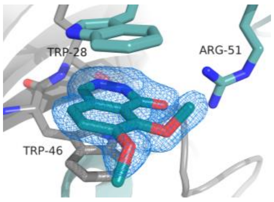
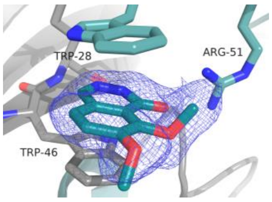
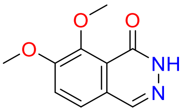
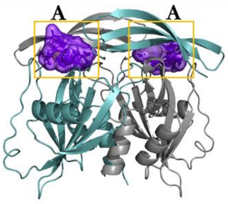
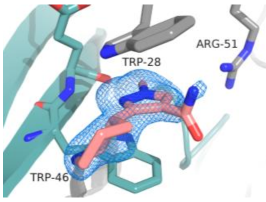
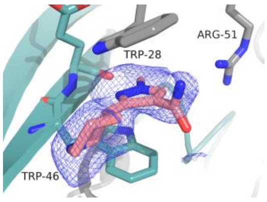
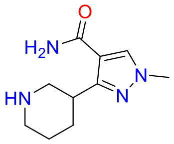
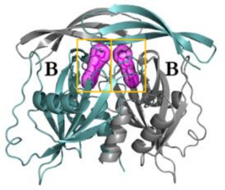
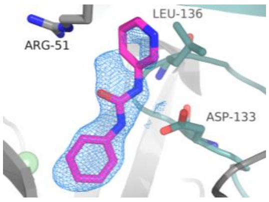
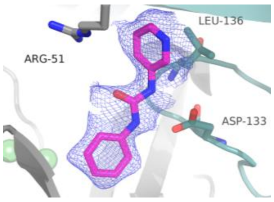
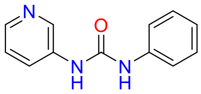
Fragment binding site	Data set # PDB ID	Structural data: $2F_o$-F_c map	Structural data: event map	Chemical structure
	**Fragment 1** NUDT5A-x0114 5QJ4 DSiP library 1-BDC: 0.28			 MW: 213.26 Da
	**Fragment 2** NUDT5A-x0131 5QJ7 DSiP library 1-BDC: 0.28			 MW: 248.28 Da
	**Fragment 3** NUDT5A-x0158 5QJ8 DSiP library 1-BDC: 0.18			 MW: 240.25 Da
	**Fragment 4** NUDT5A-x0169 5QJ9 DSiP library 1-BDC: 0.43			 MW: 227.28 Da
	**Fragment 5** NUDT5A-x0171 5QJA DSiP library 1-BDC: 0.34			 MW: 210.30 Da
	**Fragment 6** NUDT5A-x0176 5QJB DSiP library 1-BDC: 0.39			 MW: 227.69 Da

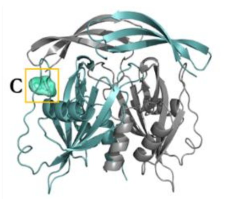
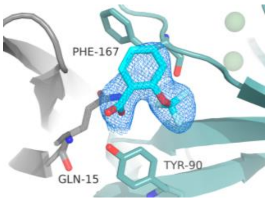
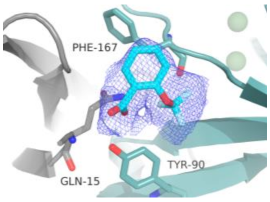
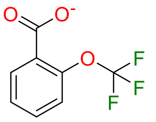
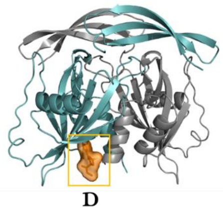
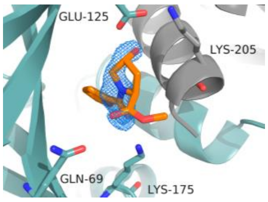
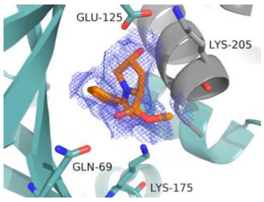
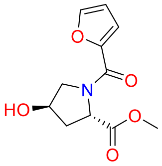
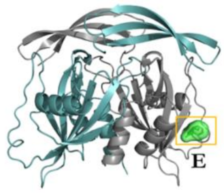
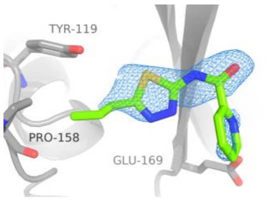
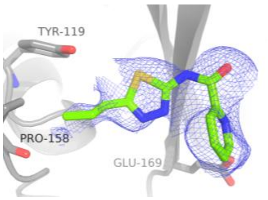
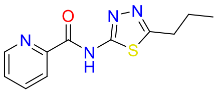
	**Fragment 7** NUDT5A-x0177 5QJC DSiP library 1-BDC: 0.21			 MW: 140.15 Da
	**Fragment 8** NUDT5A-x0195 5QJD DSiP library 1-BDC: 0.35			 MW: 206.21 Da
	**Fragment 9** NUDT5A-x0197 5QJE DSiP library 1-BDC: 0.21			 MW: 230.27 Da
	**Fragment 10** NUDT5A-x0210 5QJF DSiP library 1-BDC: 0.28			 MW: 248.30 Da
	**Fragment 11** NUDT5A-x0235 5QJG DSiP library 1-BDC: 0.31			 MW: 242.32 Da
	**Fragment 12** NUDT5A-x0244 5QJH DSiP library 1-BDC: 0.34			 MW: 198.18 Da
	**Fragment 13** NUDT5A-x0251 5QJI DSiP library 1-BDC: 0.37			 MW: 225.27 Da

	**Fragment 14** NUDT5A-x0256 5QJJ DSiP library 1-BDC: 0.21			 MW: 229.25 Da
	**Fragment 15** NUDT5A-x0262 5QJK DSiP library 1-BDC: 0.39			 MW: 175.20 Da
	**Fragment 16** NUDT5A-x0286 5QJL DSiP library 1-BDC: 0.32			 MW: 239.28 Da
	**Fragment 17** NUDT5A-x0299 5QJM DSiP library 1-BDC: 0.41			 MW: 205.27 Da
	**Fragment 18** NUDT5A-x0319 5QJN DSiP library 1-BDC: 0.28			 MW: 213.30 Da
	**Fragment 19** NUDT5A-x0320 5QJO DSiP library 1-BDC: 0.37			 MW: 169.19 Da
	**Fragment 20** NUDT5A-x0333 5QJP DSiP library 1-BDC: 0.31			 MW: 198.29 Da

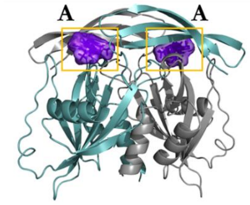
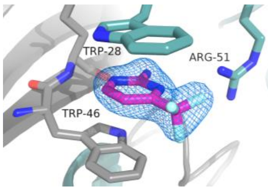
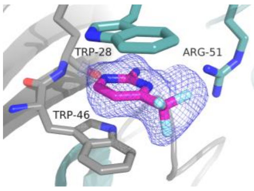
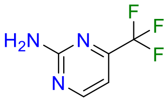
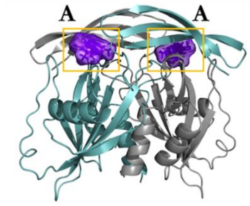
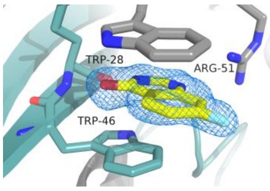
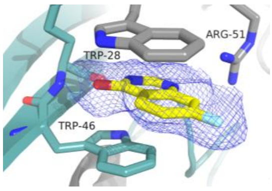
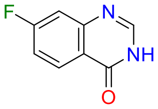
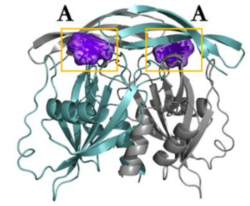
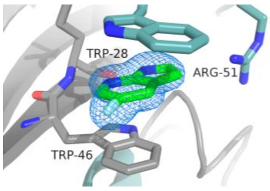
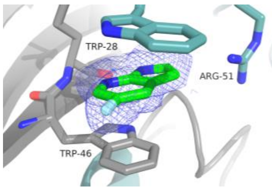
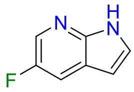
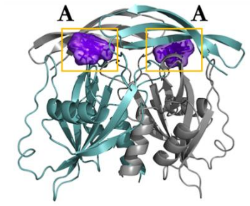
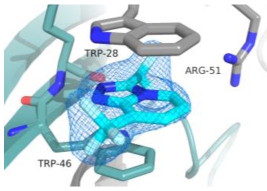
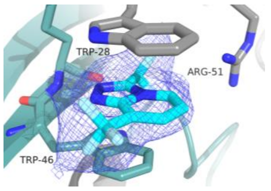
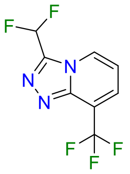
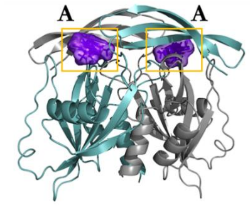
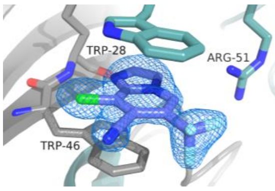
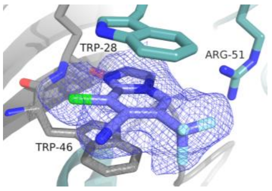
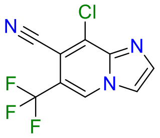
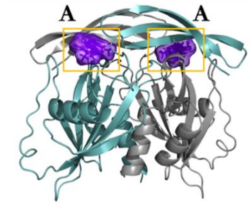
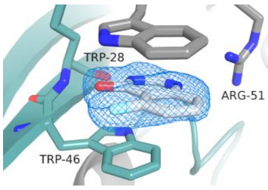
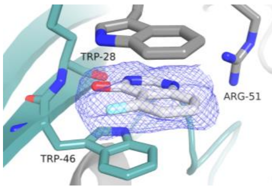
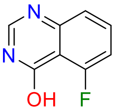
	**Fragment 21** NUDT5A-x0339 5QJQ DSiP library 1-BDC: 0.36			 MW: 205.27 Da
	**Fragment 22** NUDT5A-x0392 5QJR DSiP library 1-BDC: 0.3			 MW: 170.21 Da
	**Fragment 23** NUDT5A-x0403 5QJS DSiP library 1-BDC: 0.24			 MW: 248.30 Da
	**Fragment 24** NUDT5A-x0412 5QJT DSiP library 1-BDC: 0.31			 MW: 156.20 Da
	**Fragment 25** NUDT5A-x0463 5QJU DSiP library 1-BDC: 0.36			 MW: 199.64 Da
	**Fragment 26** NUDT5A-x0469 5QJV DSiP library 1-BDC: 0.34			 MW: 139.16 Da
	**Fragment 27** NUDT5A-x0525 5QJW DSiP library 1-BDC: 0.25			 MW: 223.23 Da

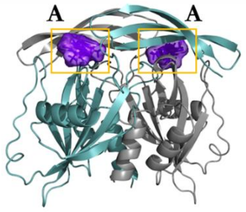
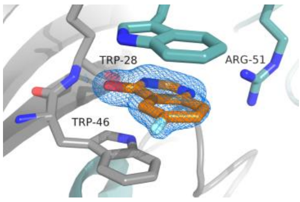
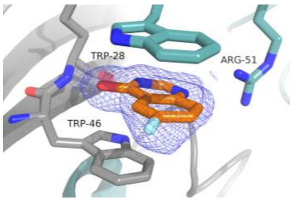
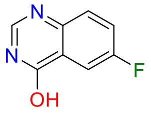
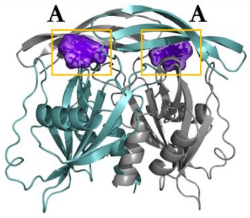
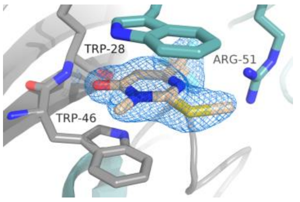
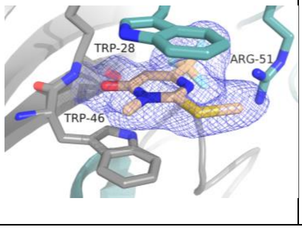
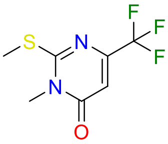
	**Fragment 28** NUDT5A-x0552 5QJY DSiP library 1-BDC: 0.23			 MW: 203.26 Da
	**Fragment 29** NUDT5A-x0553 5QJZ DSiP library 1-BDC: 0.23			 MW: 199.26 Da
	**Fragment 30** NUDT5A-x0554 5QK0 DSiP library 1-BDC: 0.51			 MW: 187.25 Da
	**Fragment 31** NUDT5A-x0574 5QK1 DSiP library 1-BDC: 0.28			 MW: 152.18 Da
	**Fragment 32** NUDT5A-x0600 5QK2 DSiP library 1-BDC: 0.39			 MW: 109.13 Da
	**Fragment 33** NUDT5A-x0605 5QK3 DSiP library 1-BDC: 0.28			 MW: 139.16 Da
	**Fragment 34** NUDT5A-x0627 5QK4 DSiP library 1-BDC: 0.25			 MW: 199.25 Da

	**Fragment 35** NUDT5A-x0637 5QK5 DSiP library 1-BDC: 0.32			 MW: 137.19 Da
	**Fragment 36** NUDT5A-x0651 5QK6 DSiP library 1-BDC: 0.32			 MW: 153.16 Da
	**Fragment 37** NUDT5A-x0673 5QK7 DSiP library 1-BDC: 0.41			 MW: 210.62 Da
	**Fragment 38** NUDT5A-x0681 5QK8 DSiP library 1-BDC: 0.16			 MW: 140.19 Da
	**Fragment 39** NUDT5A-x0685 5QK9 DSiP library 1-BDC: 0.32			 MW: 206.20 Da
	**Fragment 40** NUDT5A-x0692 5QKA DSiP library 1-BDC: 0.36			 MW: 208.27 Da
	**Fragment 41** NUDT5A-x0122 5QJ5 DSiP library 1-BDC: 0.21			 MW: 213.24 Da

	**Fragment 42** NUDT5A-x0526 5QJX DSiP library 1-BDC: 0.35			 MW: 205.11 Da
	**Fragment 43** NUDT5A-x0125 5QJ6 DSiP library 1-BDC: 0.1			 MW: 239.23 Da
	**Fragment 10** NUDT5A-x0210 5QJF DSiP library 1-BDC: 0.18			 MW: 248.30 Da

A.1.7. NUDT5 ^{19}F NMR fragment hits confirmed by X-ray crystallography.

Fragment binding site	Data set # PDB ID	Structural data: $2F_o$-F_c map	Structural data: event map	Chemical structure
	**GB-0804** NUDT5A-x1072 5QTL ^{19}F BIONET library 1-BDC: 0.32			 MW: 163.10 Da
	**FS-2639** NUDT5A-x1078 5QTM ^{19}F BIONET library 1-BDC: 0.28			 MW: 164.14 Da
	**SS-4432** NUDT5A-x1083 5QTN ^{19}F BIONET library 1-BDC: 0.38			 MW: 136.13 Da
	**1R-0641** NUDT5A-x1093 5QTO ^{19}F BIONET library 1-BDC: 0.26			 MW: 237.13 Da
	**AE-0227** NUDT5A-x1211 5QTP ^{19}F BIONET library 1-BDC: 0.3			 MW: 245.59 Da
	**FS-1169** NUDT5A-x1212 5QTQ ^{19}F BIONET library 1-BDC: 0.42			 MW: 164.14 Da

	**FS-3764** NUDT5A-x1213 5QTR ^{19}F BIONET library 1-BDC: 0.42			 MW: 164.14 Da
	**8J-537S** NUDT5A-x1214 5QTS ^{19}F BIONET library 1-BDC: 0.22			 MW: 224.20 Da

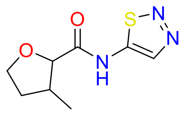
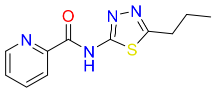
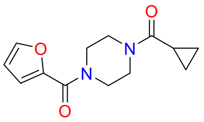
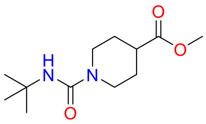
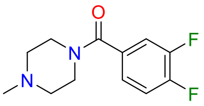
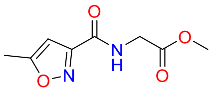
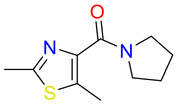
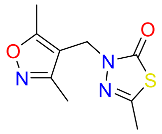
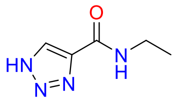
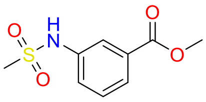
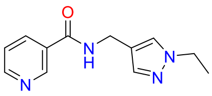
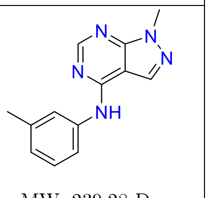
A.1.8. Interaction analysis of protein-fragment complexes using PLIP.

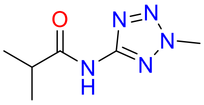
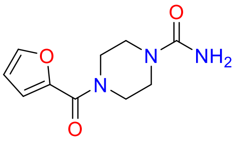
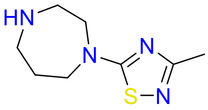
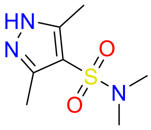
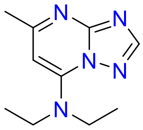
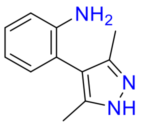
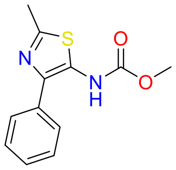
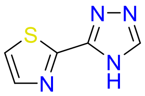
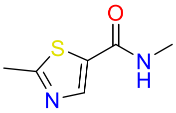
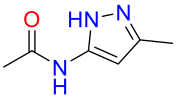
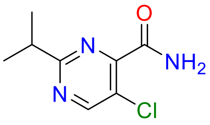
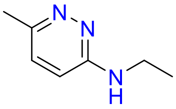
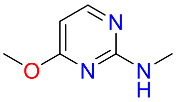
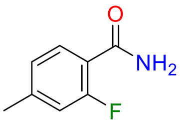
	XChem screening				**^{19}F screening followed by crystallographic conformation**		
Interactions	*Binding site*	*protein residue (atom)*	*location*	*interaction type*	*protein residue (atom)*	*location*	*interaction type*
Hydrogen bonds	A	Trp28 (N)	side chain	acceptor			
	A	Tyr36 (O)	side chain	donor			
	A	Thr45 (O)	side chain	acceptor			
	A	Glu47 (N)	backbone	donor	Glu47 (N)	backbone	donor
	A	Glu47 (O)	backbone	acceptor	Glu47 (O)	backbone	acceptor
	A	Arg51 (N)	side chain	donor	Arg51 (N)	side chain	acceptor
	A	Asp164 (O)	backbone	acceptor			
	B	Asp133 (O)	side chain	acceptor			
	B	Ser137 (O)	side chain	donor			
	C	Lys14 (N)	backbone	donor			
	C	Gln15 (N)	side chain	donor			
	C	Tyr90 (O)	side chain	donor			
	D	Gln169 (N)	side chain	donor			
	D	Lys205 (N)	side chain	donor			
	D	Lys205 (N)	side chain	acceptor			
	E	Glu47 (O)	side chain	donor			
	E	Ser172 (N)	backbone	donor			
Halogen bonds	A				Trp28 (O)	backbone	acceptor
	A	Thr45 (O)	backbone	acceptor	Thr45 (O)	backbone	acceptor
	A				Glu47 (O)	backbone	acceptor
	A	Arg51 (N)	side chain	donor			
	A				Glu135 (O)	backbone	acceptor
π-π stacking	A	Trp28	side chain	-	Trp28	side chain	-
	A	Trp46	side chain	-	Trp46	side chain	-
π-cation stacking	A	Trp28	side chain	-			
	A	Trp46	side chain	-	Trp46	side chain	-
	D	His206	sidechain	-			
Salt bridges	C	Lys14	side chain	-			

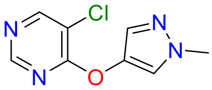
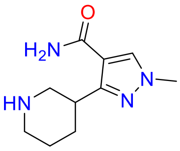
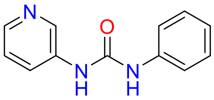
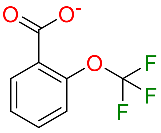
Water bridges	A	Trp28 (N)	side chain	donor			
	A	Trp28 (O)	backbone	acceptor			
	A	Trp28 (N)	side chain	acceptor			
	A	Tyr36 (O)	side chain	acceptor			
	A	Tyr36 (O)	side chain	donor			
	A	Arg44 (N)	side chain	donor			
	A	Thr45 (O)	side chain	acceptor			
	A	Thr45 (O)	side chain	donor			
	A	Thr45 (N)	backbone	donor			
	A	Thr46 (N)	side chain	acceptor			
	A	Glu47 (O)	backbone	acceptor			
	A	Arg51 (N)	side chain	donor	Arg51 (N)	side chain	donor
	A	Leu98 (N)	backbone	donor			
	A	Gly135 (O)	backbone	acceptor			
	A	Glu166 (O)	side chain	donor			
	B	Asp133 (O)	side chain	donor			
	D	Lys175 (N)	side chain	donor			

A.2. Appendix for chapter 5

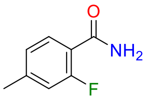
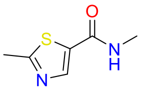
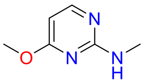
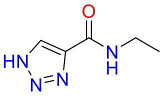
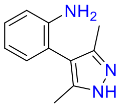
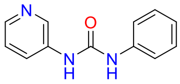
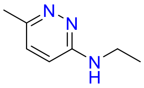
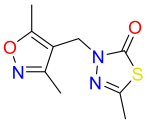
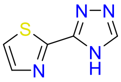
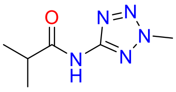
A.2.1. Subset of NUDT5 XChem fragments for ranking by NMR.

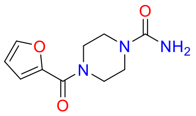
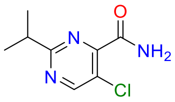
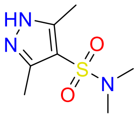
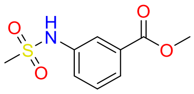
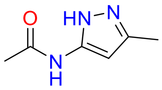
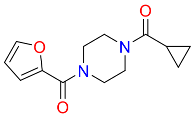
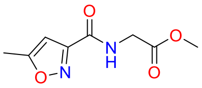
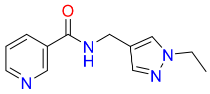
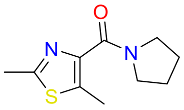
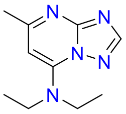
Fragment # **Data set #** **PDB ID**	**Chemical structure** **MW**	**Fragment #** **Data set #** **PDB ID**	**Chemical structure** **MW**
Fragment 1 NUDT5A-x0114 5QJ4	 MW: 213.26 Da	Fragment 10 NUDT5A-x0210 5QJF	 MW: 248.30 Da
Fragment 2 NUDT5A-x0131 5QJ7	 MW: 248.28 Da	Fragment 11 NUDT5A-x0235 5QJG	 MW: 242.32 Da
Fragment 3 NUDT5A-x0158 5QJ8	 MW: 240.25 Da	Fragment 12 NUDT5A-x0244 5QJH	 MW: 198.18 Da
Fragment 5 NUDT5A-x0171 5QJA	 MW: 210.30 Da	Fragment 13 NUDT5A-x0251 5QJI	 MW: 225.27 Da
Fragment 7 NUDT5A-x0177 5QJC	 MW: 140.15 Da	Fragment 14 NUDT5A-x0256 5QJJ	 MW: 229.25 Da
Fragment 9 NUDT5A-x0197 5QJE	 MW: 230.27 Da	Fragment 16 NUDT5A-x0286 5QJL	 MW: 239.28 Da

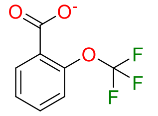
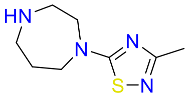
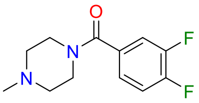
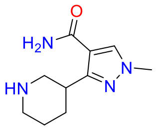
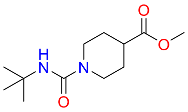
Fragment 19 NUDT5A-x0320 5QJO	 MW: 169.19 Da	Fragment 27 NUDT5A-x0525 5QJW	 MW: 223.23 Da
Fragment 20 NUDT5A-x0333 5QJP	 MW: 198.29 Da	Fragment 28 NUDT5A-x0552 5QJY	 MW: 203.26 Da
Fragment 21 NUDT5A-x0339 5QJQ	 MW: 205.27 Da	Fragment 30 NUDT5A-x0554 5QK0	 MW: 187.25 Da
Fragment 23 NUDT5A-x0403 5QJS	 MW: 248.30 Da	Fragment 31 NUDT5A-x0574 5QK1	 MW: 152.18 Da
Fragment 24 NUDT5A-x0412 5QJT	 MW: 156.20 Da	Fragment 33 NUDT5A-x0605 5QK3	 MW: 139.16 Da
Fragment 25 NUDT5A-x0463 5QJU	 MW: 199.64 Da	Fragment 35 NUDT5A-x0637 5QK5	 MW: 137.19 Da
Fragment 26 NUDT5A-x0469 5QJV	 MW: 139.16 Da	Fragment 36 NUDT5A-x0651 5QK6	 MW: 153.16 Da

Fragment 37 NUDT5A-x0673 5QK7	 MW: 210.62 Da	Fragment 40 NUDT5A-x0692 5QKA	 MW: 208.27 Da
Fragment 41 NUDT5A-x0122 5QJ5	 MW: 213.24 Da	Fragment 42 NUDT5A-x0526 5QJX	 MW: 205.11 Da

A.2.2. $\mathrm{T}_{1\rho}$ ranking of NUDT5 XChem fragments.

Rank	Fragment PDB ID	Chemical structure MW	$\mathrm{T}_{1\mathrm{rho}}$ effect ligand: protein 3 %	LDSC [mM]
1	**Fragment 36** 5QK6	 MW: 153.16 Da	10.5	250
2	**Fragment 24** 5QJT	 MW: 156.20 Da	9.1	100
3	**Fragment 26** 5QJV	 MW: 139.16 Da	7.3	50
4	**Fragment 7** 5QJC	 MW: 140.15 Da	7.2	250
5	**Fragment 30** 5QK0	 MW: 187.25 Da	7.1	100
6	**Fragment 41** 5QJ5	 MW: 213.24 Da	6.9	50
7	**Fragment 35** 5QK5	 MW: 137.19 Da	6.8	50
8	**Fragment 13** 5QJI	 MW: 225.27 Da	6.4	100
9	**Fragment 31** 5QK1	 MW: 152.18 Da	6.0	100
10	**Fragment 19** 5QJO	 MW: 169.19 Da	5.7	100

11	**Fragment 27** 5QJW	 MW: 223.23 Da	5.6	100
12	**Fragment 25** 5QJU	 MW: 199.64 Da	5.5	50
13	**Fragment 28** 5QJY	 MW: 203.26 Da	5.3	250
14	**Fragment 14** 5QJJ	 MW: 229.25 Da	5.3	500
15	**Fragment 33** 5QK3	 MW: 139.16 Da	5.3	100
16	**Fragment 2** 5QJ7	 MW: 248.28 Da	4.8	50
17	**Fragment 12** 5QJH	 MW: 198.18 Da	4.4	250
18	**Fragment 9** 5QJE	 MW: 230.27 Da	4.2	500
19	**Fragment 5** 5QJA	 MW: 210.30 Da	4.2	50
20	**Fragment 21** 5QJQ	 MW: 205.27 Da	3.8	50

21	**Fragment 42** 5QJX	 MW: 205.11 Da	3.5	250
22	**Fragment 20** 5QJP	 MW: 198.29 Da	3.4	50
23	**Fragment 3** 5QJ8	 MW: 240.25 Da	3.2	500
24	**Fragment 40** 5QKA	 MW: 208.27 Da	3.0	250
25	**Fragment 11** 5QJG	 MW: 242.32 Da	2.2	100

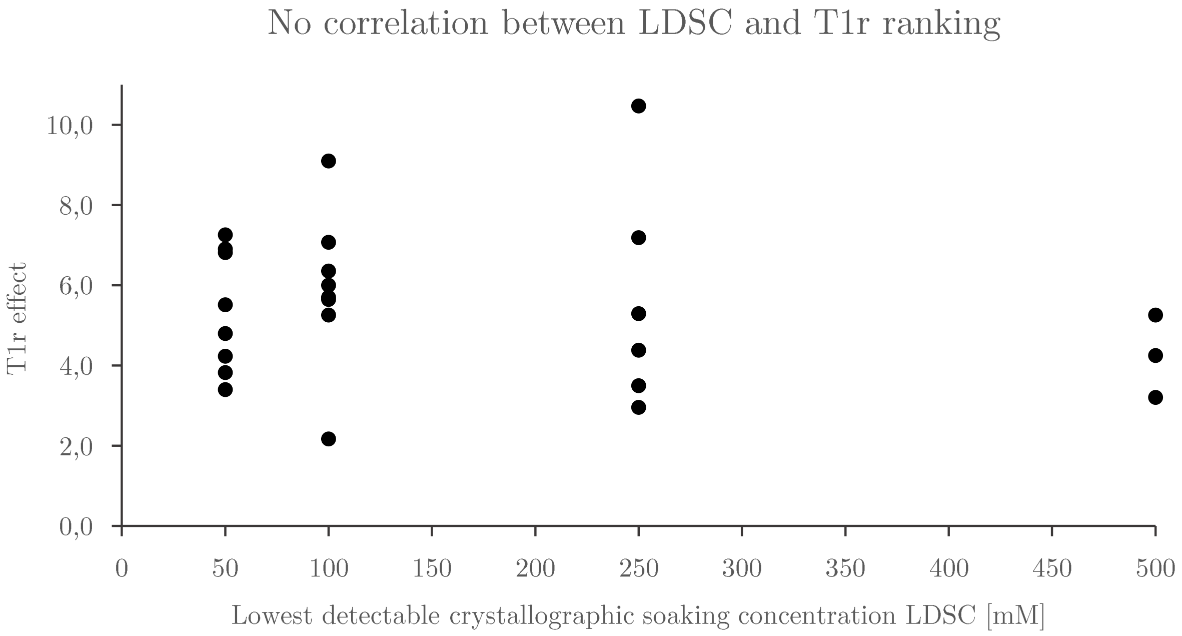
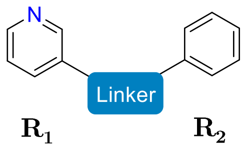


Figure A.10. No correlation between LDSC and $T_{1\rho}$. Each dot represents a fragment hit.

A.3. Appendix for chapter 6



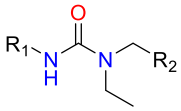
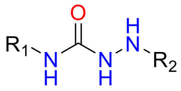
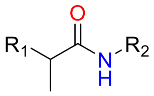
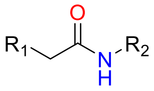
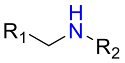
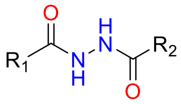
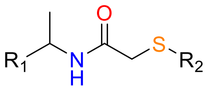
Compound	Linker	MW [g/mol]
FU031		255.3
FU032		228.3
FU033		226.3
FU034		212.3
FU035		184.3
FU036		241.3
FU037		272.4

Figure A.11. Variation of linkers to substitute the urea core of fragment **41**.