



UNIVERSITAT DE
BARCELONA

Fragment-to-Lead Optimization with Automated and Iterative Virtual Screening

Moira Rachman

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Universitat de Barcelona
Facultat de Farmàcia i Ciències de l'Alimentació

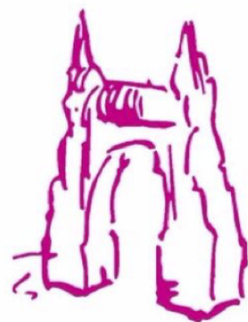
Fragment-to-Lead Optimization with Automated and Iterative Virtual Screening

Moira Rachman

2020



UNIVERSITAT DE
BARCELONA



FACULTAT DE FARMÀCIA I
CIÈNCIES DE L'ALIMENTACIÓ

Programa de Doctorat en Biomedicina, Universitat de Barcelona

Director de Tesi: Dr. Xavier Barril Alonso

UNIVERSITAT DE BARCELONA

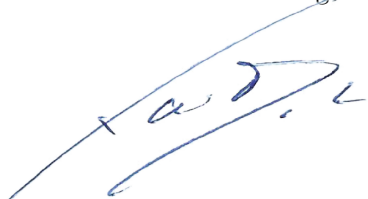
FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

DEPARTAMENT DE FARMÀCIA I TECNOLOGIA FARMACÈUTICA I
FISICOQUÍMICA

PROGRAMA DE DOCTORAT EN BIOMEDICINA

FRAGMENT-TO-LEAD OPTIMIZATION
WITH AUTOMATED AND ITERATIVE VIRTUAL SCREENING

Aquesta tesi ha estat realitzada per Moira Rachman sota la direcció del Dr. Xavier Barril Alonso, Professor d'Investigació ICREA en el Departament de Farmàcia i Tecnologia Farmacèutica i Físicoquímica de la Facultat de Farmàcia i Ciències de l'Alimentació de la Universitat de Barcelona. Es presenta aquesta memòria per optar al títol de doctor per la Universitat de Barcelona en el Programa de Doctorat en Biomedicina.



Xavier Barril Alonso
Director de tesi



Moira Rachman
Doctorand

*Moira Rachman
July 2020*

*To My Beloved Family,
Thank you!*

Acknowledgements

Being in Xavi's lab showed me the type of environment in which one could thrive. With that I'd like to thank Miriam, Salvo, Andrea, Serena, Maciej, Alex, Alvaro, Dylan, Juan, Marina and Guillermo, for the help, support and good times! Then I have to acknowledge the FragNet ITN, Iwan, Jacqueline and all other PI's, thank you for making all this possible. I'd like to also especially thank Gyuri at the RCNS, as well as Helena, for being wonderful supervisors during my secondments. Then the ESRs, Lorena, Angelo, Lena, Andrea, Ted, Bas, Eleni, David, Darius, Aaron, Sebas, Hanna and Pierre for the great memories in York, Amsterdam, Barcelona, Budapest, Cambridge and Basel.

With some I've had the chance to collaborate on fruitful projects, and I hope to continue on those in the future. Then Carles as Xavi's right hand, and finally, Xavi for not only ensuring that my projects remained on track, but also for being a mentor. You went above and beyond, for each and every one of us, and it's a miracle that you still have the time to do EVERYTHING else that you do. Thank you!

Finally, my greatest support system, I am very lucky to have you. Mama, papa, Ryan and Jair, you know how much you mean to me, and how happy it makes me to make you proud. Gina and Wayne who are practically family, the same counts for you. You don't know how much it helps to actually know that I have you as my safety net.

Table of contents

List of figures	xiii
List of tables	xv
List of Abbreviations	xvii
1 Introduction	1
1.1 The Road to a Drug	3
1.2 Lead Candidate Discovery From Fragments	4
1.3 <i>In Silico</i> Fragment-to-Lead Optimization	5
1.3.1 Chemical Space Navigation	6
1.3.1.1 <i>De Novo</i> Design	6
1.3.1.2 Assessing Synthetic Accessibility	7
1.3.1.3 Structure-based Virtual Screening	8
1.3.1.4 Ligand-based Design	9
1.3.1.5 Secondary Objectives	10
1.3.2 Fragment-to-lead <i>In Silico</i> Design Strategies	10
1.3.2.1 Scaffold Hopping	17
1.3.2.2 Highlighted Examples	17
1.3.3 Towards an integrated framework for F2L	19
1.4 Summary	20
2 Objectives	23
3 Methods	27
3.1 Overview	29
3.2 Background rDock	30
3.3 Background Schrödinger's MMGBSA	32
3.4 Background Dynamic Undocking	33

3.5	Automated Pipeline Protocol	35
3.5.1	Structure Preparation	35
3.5.2	Interaction Analysis with MDMix	35
3.5.3	Pose Clustering	35
3.5.4	Library Preparation	36
3.5.5	Similarity Searching	36
3.5.6	Ligand Preparation	37
3.5.7	MCS Extraction and Superposition	37
3.5.8	Tethered Docking	38
3.5.9	Diversity Selection	38
3.5.10	MMGBSA	39
3.5.11	DUck	40
3.5.12	Subsequent Iterations	40
4	Results	41
4.1	The Automated Pipeline	43
4.2	The Platform Applied to NUDT21	46
4.2.1	The Putative Binding Site	46
4.2.2	Comparing VS Strategies	50
4.2.3	Experimental Results	53
4.2.4	Future Perspectives	57
4.3	Exploratory Strategies	58
4.3.1	Scaffold Hopping Strategies	58
4.3.2	Chemical Space Exploration	61
4.3.2.1	Experimental Results	62
4.3.2.2	Scaffold Analysis	63
4.4	Blinded Fragment Optimization	65
4.4.1	Target Selection	67
4.4.2	Pose Selection	67
4.4.3	Case Study 1 – BRD4	68
4.4.4	Case Study 2 – HSP90 – Fragment 1	70
4.4.5	Case Study 2 – HSP90 – Fragment 2	74
4.4.6	Case Study 4 – DYRK1A	76
4.4.7	Summary	77
5	Discussion	79

Table of contents	xi
6 Conclusions	85
Bibliography	89
Appendix Supplementary Information	105
Appendix Additional Publications	143

List of figures

1.1	Lead Candidate Discovery in the Road to a Drug.	3
1.2	HTS vs. FBDD.	4
1.3	LE in F2L Optimization	5
1.4	Highlighted Examples <i>In Silico</i> F2L Optimization	18
1.5	Proposed pipeline for integrated F2L optimization.	20
3.1	Concept of Structural Stability	34
3.2	HTVS Protocols	39
4.1	Concept of Iterative VS	43
4.2	Orthogonal Structure-based Screening	45
4.3	The developed pipeline.	45
4.4	NUDT21 Fragment Hits	47
4.5	Consensus Interaction Hotspots Ferived from MDmix.	48
4.6	NUDT21 Complex	49
4.7	Poised fragment deconstruction.	51
4.8	rDock Score Distributions VS Comparison	53
4.9	SPR hits evolved from fragment X-0581.	55
4.10	SPR hits evolved from fragment X-0404	56
4.11	Further Design Strategies	57
4.12	Scaffold Hopping Strategies	59
4.13	rDock Score Distribution Exploratory Strategies	62
4.14	X-0404 In-Stock SPR Hits Binding Modes	64
4.15	X-0404 On-Demand SPR Hits Binding Modes	65
4.16	Scaffold Analysis	66
4.17	Overview of Case Studies and Predicted Binding Modes.	69
4.18	BRD4 Hits	71
4.19	BRD4 X-ray vs. Predicted.	72

4.20 HSP90 Fragment 1 Hits	73
4.21 HSP90 Fragment 2 Hits.	75
4.22 DYRK1A Hits.	78

List of tables

1.1	Automated F2L Programs	10
1.2	MedChem Design Examples	12
1.3	Fragment-Assisted Drug Design Examples	14
1.4	Automated Tailored F2I <i>In Silico</i> Design Examples	16
4.1	Attrition Rates VS Comparison	52
4.2	NUDT21 SPR Result Summary VS Comparison	54
4.3	The platform's user defined parameters.	60
4.4	Attrition Rates In-Stock and On-Demand Space.	61
4.5	SPR Results Summary In-Stock and On-Demand Space	63
4.6	Attrition Rates VS Comparison	68
4.7	NMR and SPR Results HSP90 Fragment 1.	73
4.8	NMR and SPR Results HSP90 Fragment 2.	74
4.9	NMR and SPR Results DYRK1A	76

List of Abbreviations

ADMET Absorption **D**istribution **M**etabolism **E**xcretion **T**oxicity. [6](#)

BB Building **B**lock. [6](#)

DND *De Novo* **D**esign. [6](#)

F2L Fragment-**to**-**L**ead. [5](#)

FADD Fragment-**A**ssisted **D**rug **D**esign. [8](#)

FBDD Fragment-**B**ased **D**rug **D**esign. [5–7](#), [9](#), [10](#), [12](#), [19–21](#)

GPU Graphical **P**rocessing **U**nits. [9](#)

HTS High-Throughput **S**creening. [4](#)

LE Ligand **E**fficiency. [5](#)

MD Molecular **D**ynamics. [8](#)

ML Machine **L**earning. [9](#)

MOO Multi-Objective **O**ptimization. [10](#)

QSAR Quantitative **S**tructure-**A**ctivity **R**elationship. [9](#)

SA Synthetic **A**ccessibility. [6](#)

VS Virtual **S**creening. [8](#)

Chapter 1

Introduction

1.1 The Road to a Drug

Now it should be more clear than ever that the world needs safe and effective drugs. Drugs that could cure or prevent disease, or alleviate symptoms can be delivered world-wide through established protocols that ensure the safety and efficacy of their use. However, the road to a drug is not straightforward. The start a drug discovery project reflects an unmet medical need for a particular disease, whether it be a neglected disease, a rare disease or one that is extremely prevalent with mild or deadly outcomes. The decision to commence on this path also greatly depends on a sound scientific basis that indicates that such a high risk pursuit will be worth it. This research is the foundation of a drug discovery project that may take years, or even decades in special cases where the mechanism of action is entirely novel. Once the foundation is laid, the drug discovery pipeline involving rigorous validation ensues, which if successful can take 12-15 years, and cost over 1 billion dollars.[1]

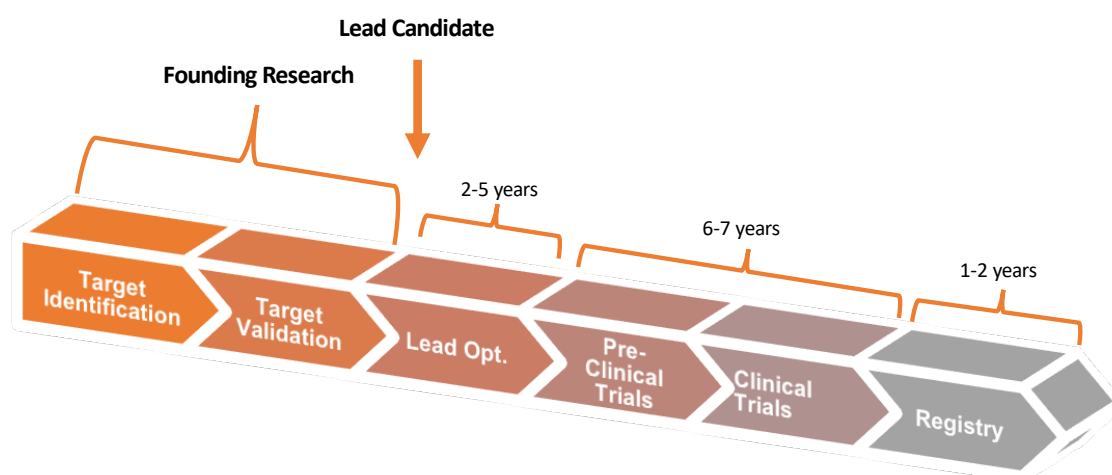


Figure 1.1: Lead Candidate Discovery in the Road to a Drug.

The founding research (Figure 1.1) usually involves the identification of small molecule probes to bind to a target. Like a key that is made for a lock, this binding may elicit a desired effect, namely, a therapeutic response that is linked to a disease state. If this response is effective in cells (*in vitro*) and in animal models (*in vivo*), the target has been essentially validated to be druggable. But before the search for the key or probe can take place, a target must be selected. This selection is based on proof that it is linked to a disease state by e.g. knockout models i.e. removing the protein out of the equation *in vitro* or *in vivo*, and checking whether the disease state changes. At the same time, significant efforts are being made to explore the druggability of all targets known in the human body.[2] In Figure 1.1, a clear distinction is made between the founding research, and the rest of the pipeline, because

this thesis is focused on strategies leading up to the identification of lead candidates. After lead optimization the clinical candidate can proceed to clinical trials. The likelihood of a clinical candidate from Phase I being approved as a drug is estimated to be a fair 10% [3][3], however, the basic research that precedes this must be substantial.

1.2 Lead Candidate Discovery From Fragments

Finding a lead candidate is like finding a needle in a haystack. The potential size of this “haystack” consisting of realistic and drug-like molecules, often referred to as chemical space has been estimated to be 10^{33} based on an extrapolation [4] of the size of GDB-17 [5] (enumeration of stable molecules up to 17 heavy atoms containing C, N, O, S and halogen atoms). Not more than a decade ago, the dominant strategy for lead candidate discovery (“exposing the needles”) was **High-Throughput Screening (HTS)** i.e. to individually screen every single molecule in the massive chemical collections that were available. The main goal of HTS was to identify hits as close in potency as possible to the sought drug. Conceptually, the fragment-based approach represented a Copernican turn, as it promoted the idea that starting from a small compound (a fragment) might ultimately be a more efficient strategy to obtain lead-like compounds (Figure 1.2). [6]

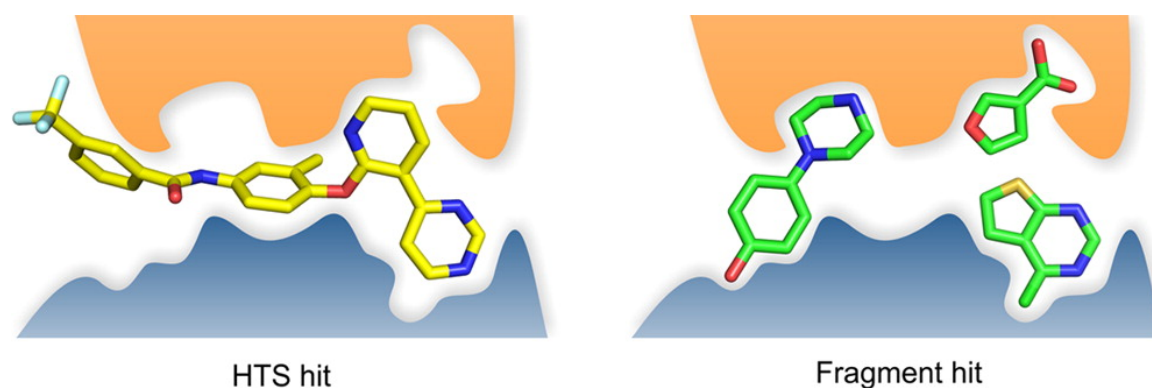


Figure 1.2: HTS vs. FBDD. Adapted from [6].

The first reason for this being that diverse fragment screening collections can represent a much larger chemical space than an equally sized library of lead-like compounds (the size of the “haystack” is reduced). Secondly, because fragments are small, they tend to have high intrinsic hit rates, which sometimes also results in hits for targets that were previously deemed as challenging by HTS (more “needles” are present). [7] However, fragments bind with low affinity, and must be optimized to lead-like compounds [8]. With this, comes the

advantage that a fragment hit can be optimized to optimally fit the binding pocket, while still having control over the pharmacokinetic and pharmacodynamic properties.

1.3 *In Silico* Fragment-to-Lead Optimization

In **Fragment-Based Drug Design (FBDD)** **Ligand Efficiency (LE)**[8][9] becomes the main decision criteria in lead optimization (as opposed to potency with HTS hit optimization), where the aim is to achieve large potency jumps with the smallest changes in molecular weight (Figure 1.3). The process involves multiple iterations of design-synthesis-testing cycles that are both time-consuming and resource-intensive, which often constituting the bottleneck in . Computational methods can significantly impact this process, because a) automatic approaches can efficiently exploit the abundance of initial fragment hits; b) computation is essential to explore the - what seem like - endless opportunities offered by chemical space; and c) design precision is less emphasized, as **Fragment-to-Lead (F2L)** optimization aims to improve binding constants by orders of magnitude, which should give preference to exploratory (non-obvious chemical modifications) rather than conservative (trivial chemical modifications) optimization strategies. Thus, computational methods can speed up the process, but also aid in generating ideas that would otherwise not be considered.

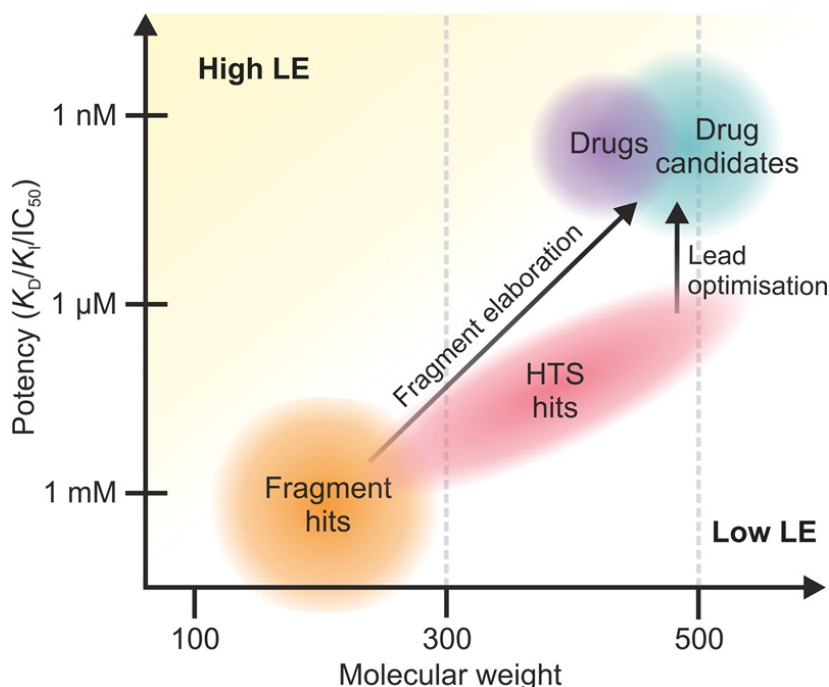


Figure 1.3: HTS hit-to-lead optimization vs. fragment-to-lead optimization. Adapted from [6].

1.3.1 Chemical Space Navigation

As noted above, computational methods can aid in exposing the “needles” (i.e. drug candidates) that are less obvious. Over thirty years ago, the idea was to navigate through chemical space by automatically assembling atoms or fragments in the binding site through *De Novo Design* (DND). Over the years DND methods have aided in the design of diverse and potent leads for various targets[10][11]. When FBDD was conceived about 10 years later, many of these methods were adopted to accommodate FBDD, laying the foundation for many *in silico* F2L optimization methods. In Section 1.3.1.1 it’s described how DND has been applied to FBDD and discuss how *Synthetic Accessibility* (SA) has been addressed Section 1.3.1.2.

Another approach involves navigating through portions of chemical space that can be acquired quickly and at low cost. Commercial chemical space has experienced continuous growth over the last decades. For instance, the ZINC database (an aggregator of chemical catalogues) started with less than 1 million molecules in 2005,[12] reaching 120 million in 2015.[13] More recently, the emergence of carefully designed virtual chemical libraries representing compounds that can be synthesized rapidly and at low cost has expanded the purchasable chemical space into the tens of billions, which creates opportunities[14][15], as well as some challenges.[16][17] Virtual screening methods applied to F2L are set to benefit the most from this trend, which is described in Section 1.3.1.3 and Section 1.3.2.

1.3.1.1 *De Novo Design*

There are two primary strategies in computational drug design. The first – structure-based – is to assess the fit of a molecule to the binding site through receptor-based scoring functions. The second – ligand-based – is to determine how well a ligand can recapitulate features of known active binders, evaluated by ligand-based scoring functions. Structure-based methods can maximize the possibility of finding binders by exploiting experimentally obtained or modelled structural information. However, when a structure is not available, ligand-based methods have also shown to be successful. Secondary objectives such as *Absorption Distribution Metabolism Excretion Toxicity* (ADMET), selectivity, synthetic accessibility and patentability/novelty can be considered as explained in Section 1.3.1.4 and Section 1.3.1.5.

DND methods can be both structure- or ligand-based. Other notable differences are 1) the *Building Block* (BB)s that are used for construction, 2) how the BBs are assembled and 3) the optimization algorithm.[18] Although using individual atoms as BBs has the advantage of exploring all theoretically possible molecules, the use of fragments as BBs meaningfully reduces the search space. BB fragments are generated by decomposing known

ligands into ring and linker moieties. The cleavage of the molecules can be done at every single bond[19] or with retrosynthetic rules.[20] For receptor-based approaches, BB assembly strategies include linking (more than one BB placed in the binding site)[21][22][23] and growing (one BB placed in the binding site).[20,21][24] The most straightforward way to join the BBs is through a single bond. However, additional rules to guide the assembly can be applied, for example creating links based on the frequency of connections between BBs extracted from a database of known drugs.[25] On the other hand, ligand-based approaches do not typically employ growing or linking. In the context of FBDD they are better suited for merging and “scaffold-replacement” strategies, often using a lead or a potent ligand as a template to guide the optimization. The optimization algorithm can be described as iterative BB assembly and scoring. The most exhaustive example is the breadth-first search strategy, which systematically explores all possible solutions. Although this can be applied to small BB libraries, heuristic algorithms offer a faster alternative with a good compromise between exhaustive exploration and efficiency.[6] For example, evolutionary algorithms can apply (crossover)[26] mutations, deletions and additions of BBs to the original molecule. [17,22] The resulting molecule with the best fit to the function describing known ligands (the initial pool) are selected for further optimization. Many ligand- and receptor- based DND methods that apply an evolutionary algorithm can be used for “scaffold hopping” purposes from starting fragments. An exception of a program that performs scaffold hopping but does not utilize an evolutionary algorithm is SPROUT,[21] which uses a template-based strategy where all-carbon chemical graphs are substituted by heteroatoms in a stepwise manner. Other algorithms include random sampling with Monte Carlo search[23], where the Metropolis criterion[27] determines the probability of rejection of a newly assembled molecule based on its score. Popular examples of DND programs used for F2L optimization include: LUDI[20] that links pre-docked fragments and has led to inhibitors of Demethylase of Fungi,[28] Aurora Kinase,[29][30] thrombin,[31] DNA Gyrase[32] and CDK4,[33] as well as LigBuilder[22] that adds fragments based on seed structures, aiding campaigns against HCV[34] and Cyclophilin A.[35]

1.3.1.2 Assessing Synthetic Accessibility

DND is a way to generate plausible and novel ideas. Unfortunately, they will not be realized if the designed molecules are considered difficult to synthesize. Furthermore, it is generally accepted that *in silico* designed molecules will require further optimization before it can be considered a *bona fide* lead. The importance of SA of *in silico* designs is projected by the numerous strategies that have been developed. The first group of DND methods implicitly considers SA by evaluating the difficulty of synthesis based on chemical complexity.[36]

Alternative approaches identify derivatives containing key features of the *de novo* designs among commercially available or synthetically accessible compounds. The program SEEDS was developed to perform this task and when combined with LEGEND[23] (one of the first DND programs that performs atom-by-atom growing based on intra- and inter- molecular stability), it led to novel potent CDK4 inhibitors.[37] The second group of DND methods implement synthetic rules to either I) perform retrosynthetic analysis to derive building blocks for possible synthesis routes or II) perform virtual synthesis where molecules are built up from commercially available building blocks. An example of a program that uses the former approach is SQUIRREL,[38] which selects BB fragments from a library of decomposed drug-like molecules using RECAP[19] rules (11 bond cleavage reaction types). The virtual synthesis approach is the most recent and seems to be most accepted as it explicitly accounts for SA. An example of such a chemistry-driven DND approach is DOGS.[39] It uses a set of 58 unique reactions[40] to link building blocks derived from a commercial library. Both SQUIRREL and DOGS were successfully applied for bioisosteric substitution of known ligands to generate novel and potent binders of PPAR and GPCR, respectively, using the synthetic routes proposed. On the other hand, the fragment-based counterpart of this design approach is the use of chemically “poised” fragment libraries[41] which are amenable to retrosynthesis by robust reactions.[42] A crystallographic screen of a poised fragment library allows rapid follow-up synthesis of analogues based on the reaction used for decomposition.

1.3.1.3 Structure-based Virtual Screening

An alternative approach to generating ligands “from scratch” is **Virtual Screening (VS)** of libraries of computationally enumerated or available compounds. For F2L purposes VS libraries can be biased towards the starting fragment either through similarity or substructure searches in available libraries, or enumeration using synthetic rules. This can be termed as **Fragment-Assisted Drug Design (FADD)**. [43] The most high-throughput structure-based VS strategy is flexible ligand docking or docking of pre-generated conformers to a rigid receptor. Alternatively, docking can be performed against multiple conformations of the protein or a protein where the residues of the binding site are flexible. Despite its many advantages (e.g. low cost and high speed), docking is known to suffer from overly simplistic and inaccurate scoring functions.[44] Therefore, VS is often only the first step of computational workflows and can be complemented by more expensive methods that consider protein flexibility and solvation effects. Examples of such methods are different variations of **Molecular Dynamics (MD)** simulations that can serve as a more accurate predictor of binding affinity.[45] In addition to FADD, structure-based approaches are often used to prioritize fragment hits before or in parallel to experimental screening.[46][47][48][49] Furthermore, if a fragment

is found to be a hit, but crystallography fails to obtain its structure, docking[50][51][52] or MD based applications[53] can be used to derive the binding mode. In either case, pharmacophore restraints (known or predicted by hotspot detection methods[54] can be applied to bias the ligands towards more likely poses.[55][56] This is especially useful for fragments, as there are many more binding opportunities in a protein for smaller molecules than lead-like molecules.

1.3.1.4 Ligand-based Design

In the absence of information about the receptor, ligand-based approaches have proven successful. They rely on the assumption that active molecules encode the properties of the binding site. The most popular and well-established methods in ligand-based drug design are **Quantitative Structure-Activity Relationship (QSAR)** and pharmacophore modeling. In recent years, however, the attention has been shifted to more advanced **Machine Learning (ML)** approaches that take advantage of “big data” and advancements in computer hardware such as **Graphical Processing Units (GPU)s**. [57] QSAR searches for quantitative correlations between structural molecular descriptors and properties (e.g. biological activities and toxicity) for a series of compounds.[58] It assumes that the structure of a molecule must contain the features responsible for its chemical and biological activities. Pharmacophore modeling extracts a set of structural features (e.g. aromatic rings, hydrogen bond donors and acceptors) of a molecule that are recognized by receptors and are responsible for its biological activity. Its speed allows fast screening of large datasets, which makes it particularly useful in the early stages of drug design.[59] ML methods, especially deep learning are able to predict properties of unknown molecules, as well as generate novel molecules of properties similar to input molecules.[60][61] Deep generative models are based on recurrent neural networks and trained on either string-based[62] or graph-based representations[63] of molecules to learn the relation between ligand representation and structure. Then models can be used to generate novel and chemically valid ligands of desired properties.[64] However, the majority of these methods do not account for synthetic feasibility. To address this problem some efforts have been made in the field of computer-aided synthesis planning[65][66][67][68] most recently by directly embedding synthetic knowledge into *de novo* drug design with forward synthesis model powered by reinforcement learning.[69] Additionally artificial intelligence systems for retrosynthesis planning can suggest possible synthetic pathways for the new chemical entities. Implicit SA can be considered through rule-based filtering[70] as done in an example for 3D linker design in **FBDD**. [71] It can also be argued that models trained on real examples will more likely supply realistic synthetic examples.

1.3.1.5 Secondary Objectives

The vast majority of drug designs will fail due to ADMET issues. General filters as described by Lipinski[72] can be applied on the generated designs to remove problematic structures (for example done in a study on Aurora Kinase A[44]). There are plenty of *in silico* programs that can be used for filtering out ligands with undesirable ADMET properties.[73] Another approach is **Multi-Objective Optimization (MOO)** that may address ADMET, selectivity or multi-target design on top of potency. This can be performed manually, for example in a study where an **FBDD** strategy was employed to design ligands that selectively modulate all PPARs, which led to designs that proceeded to clinical trials.[74] *De novo* methods that leverage off of ML could be of particular interest with respect to MOO as it may also consider scaffold hopping.[75] A nice example of *de novo* MOO design using ML is presented by Besnard et al,[76] where secondary objectives included blood-brain barrier penetration, polypharmacology and designs against unwanted targets for GPCR's. *De novo* methods that utilize deep learning can be aimed at F2L purposes.[61][77] When handling deep learning methods in a multidimensional way, MOO in automated F2L optimization can be achieved.[63]

1.3.2 Fragment-to-lead *In Silico* Design Strategies

Regardless of the hypothesis or theory behind a computational methods, their validation only comes from successful prospective applications. Table 1.1 presents an overview of some F2L tailored in silico programs that have been prospectively used (not necessarily developed) after 2005 in the context of **FBDD**. Selected examples are highlighted below and in Figure 1.4.

Table 1.1: Examples of in silico programs for which prospective examples are available in the context of **FBDD** after the year 2005.

Program	Year	Primary Objective	Ob-	Method Description
CHEMISTRY-DRIVEN				
AutoCouple[78]	2018	Receptor-based		Virtual synthesis from 1 fragment with available building blocks Tethered docking with rDock Minimization with CHARMM implicit solvent model
PINGUI[79]	2018	Receptor-based		Virtual synthesis from 1 fragment with available building blocks that can fit pocket

Continuation of Table 1.1			
DOTS[80]	2018	Receptor-based	Docked separately Docking coupled ligand For growing, linking & merging Fragment hit optimized based on possible one-step reactions that can be performed by robots Constrained docking with S4MPLE
FRAGMENT REPLACEMENT			
BREED[25]	2004	Ligand-based	Alignment ligand-bound structures Matching bonds used to split ligands Re-links all that match
DE NOVO FRAGMENT OPTIMIZATION			
Auto T&T2[81]	2019	Receptor-based	Fragments from decomposed ligands Derivation interaction sites Docking of fragments to interaction sites Linking of docked fragments Similarity search performed in commercial libraries
LigBuilder[22]	2000	Receptor-based	Fragment database of linkers and rings Link or grow based on receptor complementarity User must consider synthetic accessibility
LUDI[20]	1992	Receptor-based	Fragment database of linkers and rings Interaction sites are identified Fragments docked according to sites Linking of docked fragments User must consider synthetic accessibility
SCAFFOLD HOPPING			
TOPAS[17]	2000	Ligand-based	Building blocks from fragmentation based on cleavage reactions Template derived pharmacophores

Continuation of Table 1.1			
SPROUT[21]	1994	Receptor-based	Random addition of synthetically compatible building blocks Selection based on pharmacophore similarity Database of fragment skeletons (no atom types) Interaction sites are identified Addition of fragment skeletons to template skeleton Mutation of skeletons to atoms types complementary to the binding site
FRAGMENT ASSISTED DRUG DESIGN			
ALTA[82]	2008	Receptor-based	Decomposed ligands to fragments Docking of decomposed fragments Substructure search best scoring fragments Docking of substructure hits Molecular dynamics of best scoring ligands

Despite the development of plenty of automatic F2L tailored programs, *in silico* methods in FBDD are most often manually implemented. In the most traditional approach to F2L, medicinal chemists make modifications to a fragment by eye. In this way, synthetic feasibility is explicitly considered, but relatively few chemical transformations are used, due to subjective training and experience of the chemists. These strategies benefit from *in silico* scoring of the modified fragments based on their likelihood of improved binding affinity. Docking-guided F2L optimization, sometimes combined with more expensive methods like MD, seems to be the most common computational strategy. Table 1.2 gives an overview of examples using traditional “medchem” design with *in silico* methods after 2015.

Table 1.2: MedChem Design Examples

Entry	Year	Target	<i>In silico</i> Strategy
1[83]	2020	InhA	Docking guided optimization towards site specific regions
2[84]	2018	PKC-iota	Docking optimized fragment hit

Continuation of Table 1.2

			H-bond interaction energies on the scaffold were calculated with DFT to suggest modifications
3[85]	2018	ERK1/2	Docking guided optimization towards site specific regions
4[86]	2018	WDR5-WIN-Site	Docking guided optimization towards site specific regions
5[87]	2018	AMPA	Docking of linked analogues
6[88]	2017	NAMPT	Docked fragment overlaid with other x-rays and merged
7[89]	2017	PDK1	Docking & MD guided optimization towards site specific regions
8[90]	2017	MTase	Docking of linked analogues
9[91]	2016	GyrB	Docking guided optimization towards site specific regions
10[92]	2016	beta-Secretase	Docking of linked analogues
11[93]	2016	CYP	Docking guided optimization towards site specific regions
			Docking of linked analogues
12[94]	2016	KEAP-1	Docking guided optimization towards site specific regions
13[95]	2015	PDE10	Docking guided optimization towards selectivity handles
14[96]	2015	JAK	Docking guided optimization towards known inhibitors
15[97]	2015	Plasmepsins	Homology modeling Multiple conformation docking Docking guided optimization towards site specific regions
			Docking guided optimization towards selectivity handles
16[98]	2015	ErK1/2	Docking guided optimization towards site specific regions
17[99]	2015	PKC-iota	Docking guided optimization towards site specific regions
18[100]	2015	Factor XI	Docking of linked analogues

Continuation of Table 1.2			
19[101]	2015	BCATm	Docking guided optimization towards in silico derived hotspots
20[102]	2015	Mcl-1	NMR guided docking of merged analogues into sub-pocket
21[103]	2015	mGlu5	Homology modeling Docking guided optimization towards known inhibitors
22[104]	2015	BRD9	Induced fit docking guided optimization from dual binding modes
23[105]	2015	Factor VIIa	VS fragment library in multiple conformations Hits merged with existing ligand

Sometimes, computational methods take a more prominent role, directing the whole fragment evolution exercise. In the simplest case, hotspot detection methods can suggest the most promising modifications and vectors for growth. In one example, SiteMap[106] was used to probe the binding site and 3D-RISM[107] was used to probe possible water sites to guide optimization of ligands for ROR γ . [108] While 3D-RISM was also used for studying water displacement in Gyrase B inhibitors, [109] others used WaterMap[110] to guide optimization for pyruvate kinase M2. [111] Information from crystallized fragments can also be used for ligand design for example with the FADD approach. Table 1.3 gives an overview of studies using this approach after 2015.

Table 1.3: Fragment-Assisted Drug Design Examples

Entry	Year	Target	Strategy
1[112]	2019	NSD3	Ligand-based VS of fragment hit Docking guided optimization towards site specific regions
2[113]	2018	DRYK1A	Fragment hit deconstructed with the aim to improve LE Docking guided optimization of new (docked) fragment
3[114]	2018	3C Protease	Scaffold hopping on validated fragment hit Pharmacophores translated to SMARTS pattern Substructure search commercial vendors for new scaffold

Continuation of Table 1.3			
			Virtual synthesis of experimentally validated hit and VS
4[115]	2017	PEX14	NMR screen fragment library VS drug-like compounds VS hits merged with NMR fragment hits
5[116]	2017	BCL6	VS fragments different sub-pockets Virtual fragment hits were merged Merged ligand used as a template for ligand-based screening
6[117]	2017	USP7	Fragment hit used as a template for similarity search VS of similarity hits
7[118]	2017	Lp-PLA2	Fragment hit used as a template for similarity search VS of similarity hits
8[119]	2017	MTH1	VS commercial analogues of fragment hit
9[120]	2016	MKK3/6	SAR-by-catalogue followed by docking. Docking guided optimization for novelty
10[121]	2016	FabH	Experimental fragment screening Common chemotype docked for optimization
11[122]	2016	Lp-PLA2	X-ray fragment hits used to guide VS Substructure search analogues around fragment hit Pharmacophore based VS
12[123]	2016	Lp-PLA2	VS of substructure search hits around fragment hit
13[124]	2015	MMP-13	Docking guided optimization of extracted motif in known binders
14[120]	2015	cIAP1 & XIAP	Docking of virtual library around fragment targeting sub-pocket
15[125]	2015	CK	Docking of virtual library around fragment hits for SAR

Although such bespoke strategies work well and are fairly common, they lack the systematics, reproducibility, broad applicability, automatization and validation of some of the tools listed in Table 1.1. As an example, ALTA is an automated tool for FADD, and was successfully used for the design of BRD4,[126] EphB4,[127] and most recently, CREBBP[128][129]

inhibitors. The 38 examples listed in Table 1.2 and Table 1.3 also make apparent that *in silico* methods are most often used to manually rank modified fragments. *In silico* methods have the potential to both guide the design strategy and score the candidate molecules in an automated way. The most recent prospective examples of automated F2L tailored *in silico* methods are shown in Table 1.4.

Table 1.4: Automated Tailored F2L *In Silico* Design Examples

Entry	Year	Target	Program	Strategy
1[76]	2019	GPCRs	Auto T&T2	<i>De Novo</i> Growing
2[73]	2018	CBP	AutoCouple	Chemistry-driven
3[74]	2018	2AR	PINGUI	Chemistry-driven
4[126]	2018	2AR	PINGUI	Chemistry-driven
5[40]	2018	AKA	LigBuilder	VS building block libraries identified core LigBuilder (<i>De Novo</i> Optimization) Fragment library made to include AKA moieties Definition of core attachment point FEP on <i>de novo</i> designs
6[75]	2018	BRD4	DOTS	Chemistry-driven
7[127]	2016	GSK3	LigBuilder	Site-directed VS fragment library LigBuilder (<i>De Novo</i> Optimization) Link virtual fragment hits in distinct sites Known binder as a template for linker placement VS of <i>de novo</i> designs
8[24]	2016	AKA	LUDI	Substructure search common scaffold Inhibition assay revealed hit with common scaffold Binding mode prediction with docking LUDI (<i>De Novo</i> Optimization)
9[128]	2015	DDR1/2	BREED (version by Astex)	X-ray fragment hit merged with known inhibitor (Fragment Replacement)
10[124]	2015	CREBBP	ALTA	Fragment-Assisted Drug Design

1.3.2.1 Scaffold Hopping

Only in five out of the 49 examples listed in Table 1.2, Table 1.3 and Table 1.4 the initial fragment was not maintained, as scaffold hopping strategies were employed.[111,113,115–117] Scaffold hopping can be imperative in cases when a chosen series is potent but selectivity or ADMET issues arise.[130] Furthermore, scaffold hopping methods can generate novel ideas that are far from obvious, and could likely not be predicted by eye. Some examples not listed above utilized SPROUT to generate novel scaffolds for *Plasmodium falciparum* dihydroorotate dehydrogenase,[131][132] as well as for BACE-1,[133] where SPROUT-HitOpt and SPROUT-LeadOpt were used to generate synthetically accessible derivatives of the scaffolds. Another study used TOPAS for the Identification of novel CB-1 ligands by scaffold hopping based on a known template, followed by retrosynthetic analysis and subsequent design of focused libraries for testing.[134]

1.3.2.2 Highlighted Examples

Despite the limited use of automated F2L tailored *in silico* methods, their value has clearly been demonstrated. Here three examples are highlighted, all of which used a significant amount of *in silico* guidance for the design. Although in the first example the methods were not entirely streamlined, it also displays that the methods can be tailored to a specific aim in a project.

In the first study,[113] the researchers identified inhibitors for enteroviral 3C protease through a fragment-based approach (Figure 1.4A). Inhibitors that could potentially covalently bind a reactive cysteine in the pocket were sought. Based on known ligands, pharmacophores were derived with LigandScout.[135] The pharmacophore model included a covalent attachment point (Michael acceptor warhead). Pharmacophore-guided ligand-based VS of a library containing 3000 fragments led to 47 virtual hits, of which one showed concentration- and time-dependent activity. The time dependency indicated covalent binding. Next, a substructure search using the pharmacophore model (translated into a SMARTS pattern) identified five additional structures with scaffold hops in commercial fragment libraries, with one classified as covalent binder. The next step was virtual synthesis which encoded a nucleophilic substitution reaction at the pyrrole nitrogen with alkyl halide building blocks. This resulted in 2000 N-alkylated analogues. Subsequent ligand- and structure-based VS led to an improved, albeit not by much, optimized fragment. Nonetheless, this strategy identified novel scaffolds that could covalently bind the 3C protease.

In a related example, an automated chemistry-driven approach was taken to identify CBP bromodomain inhibitors[77] (Figure 1.4B). First, a known CBP inhibitor was deconstructed

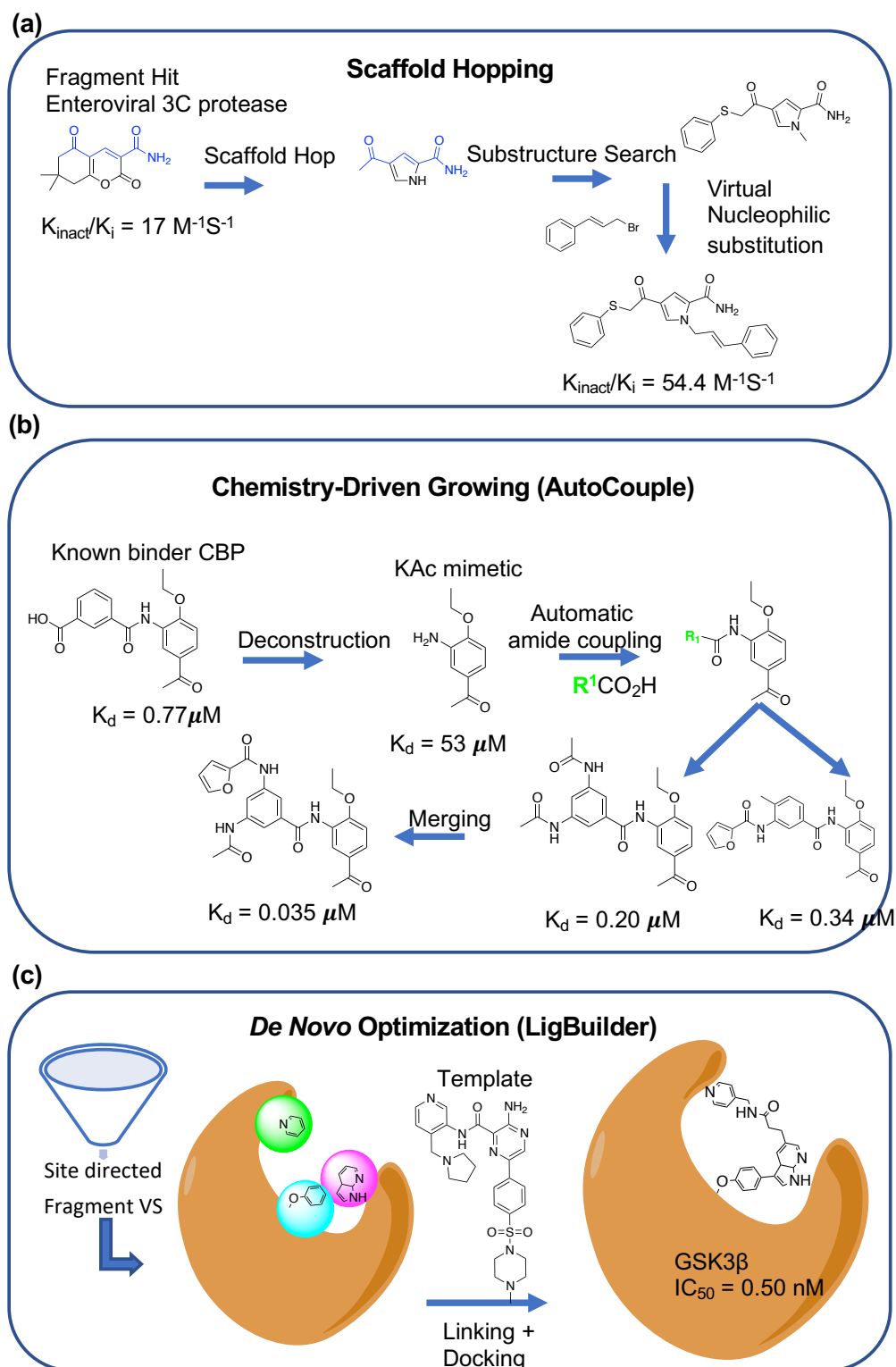


Figure 1.4: (a) Scaffold hopping approach[109] from Table 1.3. (b) Chemistry-driven approach with AutoCouple[73] from Table 1.4. (c) *De Novo* Optimization with LigBuilder[127] from Table 1.4

to a scaffold that is an acetyl-lysine mimetic. Three separate coupling reactions were encoded for virtual synthesis, namely, amide, Buchwald-Hartwig and Suzuki cross-coupling. After the successful synthesis of 53 compounds by the proposed reactions, six compounds showed nanomolar affinity. The most potent binder was derived from an amide coupling, which presented optimal growth vectors. The two compounds showing more selectivity for CBP with respect to another bromodomain, BRD4 were then selected for merging. This example highlights the use of different reaction types for site-directed optimization. The last highlighted example in Figure 1.4C describes a fragment linking approach with

LigBuilder, a DND program developed 20 years ago that is still used (not limited to FBDD) and updated regularly.[136] In this example[45], the researchers performed site-directed VS of a fragment library to identify substructures for the targeted sub-pockets. The BB library in LigBuilder was adapted to contain the substructures that were identified to be likely binders for each sub-pocket during the first virtual screening. It generated 285,000 designs using a template of a known binder, which were subsequently docked using a modified scoring function that accounted for hydration effects. Synthesis of the best scoring ligands led to picomolar inhibitors for GSK3 β .

1.3.3 Towards an integrated framework for F2L

Ultimately, automatic F2L methods could have a major impact on FBDD. The trend for automatization and open platforms for drug discovery, exemplified by X-Chem[137] and other streamlined drug discovery platforms,[138][139] introduces fast, low-cost and reproducible computational pipelines to evolve the fragments into suitable chemical tools, performing a systematic exploration of chemical space that does not neglect the scaffold hopping opportunities. However, as it has been noted several times, SA of the molecules must be appropriately considered. In fragment-based DND, similarity searches in libraries of commercially available compounds, virtual (retro)synthesis, or chemical complexity filters can address SA. In regards to scaffold hopping strategies, it is less straightforward to ensure SA. Indeed, some efforts have been made in the field of computer-aided (retro)synthesis planning.[64–67] If properly validated, this can also contribute to the continuously growing chemical libraries available through on-demand synthesis. Therefore, we propose an automated framework utilizing the ever-growing commercial on-demand libraries (Figure 1.5). It performs structure-based scaffold hopping, to exploit structural information necessary in fragment evolution, and systematically explore as much of the available chemical space as possible. The steps include identifying analogues of the fragment in the search space, identifying common features and performing tethered docking that samples different binding modes of the analogues while maintaining key features. The analogues that remain are

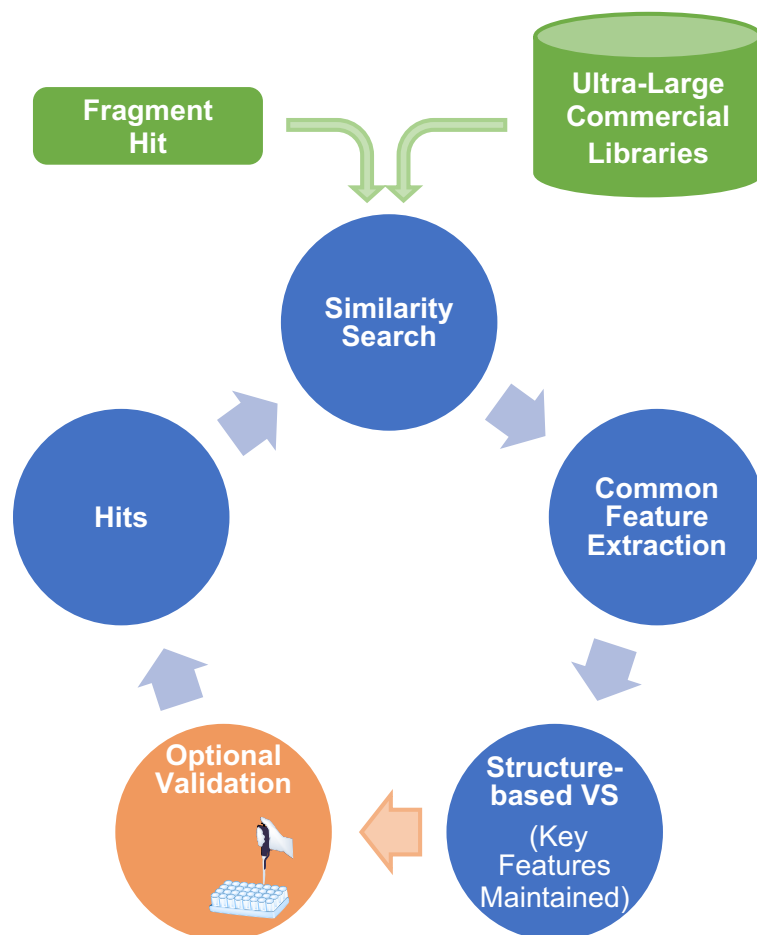


Figure 1.5: Proposed pipeline for integrated F2L optimization.

predicted to be the most likely to bind. These can be validated experimentally, after which a next round of optimization ensues. The devised protocol is iterative in line with the criteria of LE maintenance in F2L evolution.

1.4 Summary

It is clear that *in silico* methods are prominent in FBDD. There are an abundant number of programs that have been useful in FBDD. However, *in silico* methods are most often used to

rank optimized fragments, whereas the design strategy is guided by expertise and intuition that ensure the synthesizability of the designs. This common practice may unknowingly lead to biased and scaffold centric designs, thereby limiting the chemical search space. *In silico* strategies for chemical space exploration involve de novo drug design and virtual screening methods that address SA implicitly or explicitly. Scaffold hopping is one key strategy to explore as much chemical space as possible, however, ensuring SA has no easy solution. To circumvent this, my thesis develops a framework of iterative VS for F2L evolution utilizing the continuously growing commercial libraries (as of now in the billions of molecules) that does not neglect scaffold hopping opportunities. Some limitations that remain are inherent to high-throughput computational methods which include receptor flexibility, although addressed by several fragment optimization programs,[140][141]. Nevertheless, integrated in silico methods have the potential of systematically and automatically exploring all areas of chemical space, which can be applied at every stage in FBDD. Along with development, a lot of emphasis has been placed on prospective validation of the platform, in order to demonstrate that they can be fully integrated into streamlined drug discovery platforms, accelerating the release of fragment originated drugs.

Chapter 2

Objectives

General objective

The main objective of my work comprised the development and validation of an automated and iterative fragment-to-lead optimization pipeline as described in Section 1.3.3, including its application to diverse scenarios. The pipeline can be fully integrated into a streamlined drug-discovery platforms to accelerate the release of fragment derived drugs.

Specific objectives

The specific objectives were as follows:

1. Development:

- The developed pipeline must be sufficiently fast and scalable to screen increasingly large databases.
- The developed pipeline must be automated and robust so that it can be used by external parties.
- The developed pipeline must provide non-obvious analogues with a good chance of having superior binding affinity.

2. Validation and application to diverse scenarios:

- The developed pipeline must be proven to be at least equally successful as other virtual screening strategies.
- The developed pipeline must provide equally good results even when the binding mode of a fragment has not been experimentally determined.
- The platform must be fast and adaptable to various discovery settings.

Chapter 3

Methods

3.1 Overview

As proposed in Figure 1.5, the developed pipeline performs iterative structure-based scaffold hopping, to exploit structural information and systematically explore as much of the available chemical space as possible. This is done by imposing two main requirements, namely, the hits must be similar to the initial fragment, and the hits must contain the key features for binding.

A general summary of steps in the pipeline (depicted in Section 4.1):

- (i) A similarity search between the starting fragment and the search library is performed.
- (ii) The maximum common substructure (MCS) is derived.
- (iii) The MCS between fragment and analogues is superposed.
- (iv) Tethered docking is performed so that the MCS cannot deviate, while the remaining parts of the molecule can adjust to the binding pocket.
- (v) Molecular Mechanics using the Generalized Born model and Solvent Accessibility (MMGBSA) is performed to account for solvation effects.
- (vi) Dynamic Undocking (DUck) is performed to assess the structural robustness.
- (vii) The best scoring analogues become the new parents for a next iteration, starting from step one.

A summary of filters used to guide the platform to give results that are expected:

- The MCS must be at least a percentage of the number of atoms in the analogues or the cores of the analogues as defined by the user.
- The MCS must be contained in the area that has been defined as key for binding.
- After tethered docking, the MCS is restricted by a Root Mean Square Deviation (RMSD) cutoff defined by the user.
- After minimization using the MMGBSA formalism, the pose obtained must be similar to that of tethered docking evaluated by an RMSD value defined by the user.
- During tethered docking, MMGBSA and DUck, the analogues must score better than a user defined cutoff in iteration 1, or must score better than the parents in iterations over 1.

A summary of further restrictions to ensure scalability:

- After tethered docking, MMGBSA and DUck a diverse set of compounds is selected, if the number of compounds to proceeding to the next step is too large.
- For iterations over 1, only the top similarity hits, ranked by the number of parents and their scores are selected.
- Ligands are prepared on-the-fly, so that preparation of ultra-large libraries is not a bottleneck, and so that the libraries can be updated when regularly.

Because the platform is iterative, the most promising hits must be obtained to become the new parents for the subsequent iterations. For this reason, three orthogonal structure-based approaches were implemented. The first approach is docking, which is sufficiently fast but suffers from overly simplistic and inaccurate scoring functions due to insufficient consideration of desolvation penalties and entropic effects. Therefore, in the virtual screening (VS) strategy used here, docking is complemented by more expensive methods. The first being MMGBSA to explicitly consider solvation effects. Then, DUck is performed as it is complementary to both docking and MMGBSA, as it considers structural stability through H-bonds rather than thermodynamic stability.

3.2 Background rDock

Docking is the most high-throughput way to assess if a molecule might bind to a target, and therefore likely the most widely recognized. It involves a two-step process, where in the first, it attempts to predict the conformation and orientation of the molecule in the binding pocket (the binding mode or “pose”), and in the second, it attempts to predict the binding affinity by assessing the molecule’s fit through the evaluation of molecular interactions. There are different ways of sampling the binding modes that a molecule may adopt, which can be categorized as systematic or stochastic approaches. Systematic approaches aim to sample all degrees of freedom in a molecule, which can become infeasible for larger molecules. Stochastic approaches make random changes to the binding mode that are later evaluated by a predefined probability function. The developed pipeline uses the docking program rDock[142], which implements a combined approach with different stages of a genetic algorithm to aid sampling efficiency by not over penalizing bad contacts in the initial, randomized population, and by encouraging the formation of intermolecular hydrogen bonds (H-bonds). For example, the algorithm makes random mutations in a first stage where less pronounced intermolecular H-bonds are still accounted for, while in later stages, the

definition of H-bonds becomes more strict. The algorithm will make mutations until there is a convergence in the score, which promotes early termination of the run for bad poses. This is done for the generation of a single pose, for which multiple “runs” can be performed to obtain multiple poses.

Besides specifying the number of runs, the cavity has to be defined. This can be done with the “two sphere” method (cavity accessible by small spheres, but not by large spheres), and the reference ligand method (with a defined volume of a given size). If a docked molecule lies more than 0.1 Å outside of this cavity, the scoring function imposes a linear penalty based on the distance from the cavity. Furthermore, the scoring function penalizes binding modes that do not contain features where pharmacophores are defined. Pharmacophores are defined as radii around coordinates, and if the pose does not contain the specified feature within the radius, a quadratic penalty is applied based on the distance from the radius. Lastly, defined atoms can be tethered so that rDock applies mutations where defined atoms can only deviate to a degree defined for dihedral angles and rotatable bonds or according to a distance defined for translational deviations.

In general, three types of scoring functions exist, all of which attempt to approximate the binding free energy, namely force field (quantify inter- and intra- interactions e.g. AutoDock[143]), empirical (sum of parametrized functions that reproduce experimental data e.g. rDock), and knowledge-based (sum of parametrized functions that reproduce observed atomic distributions; i.e. potentials of mean force[146]) scoring functions. The scoring function of rDock is given in Equations 3.1, which is a weighted sum of intermolecular (S^{inter} , Equation 3.2), ligand intramolecular (S^{intra} , Equation 3.3), site intramolecular (S^{site} , Equation 3.4), and external restraint terms if provided ($S^{restraint}$, Equation 3.5). The weights have been extrapolated from known experimental data of 102 protein-ligand and RNA-ligand complexes. When comparing different ligands, typically S^{inter} is used for ranking, because the intramolecular scores are not sufficiently accurate and they lack an absolute scale (they have not been referenced to the lowest-energy conformation of each ligand). Furthermore, the best fit of S^{intra} to the data did not include a desolvation potential, therefore it was decided to account for solvation effects through a complementary approach that is computationally more expensive but still sufficiently high-throughput, namely, Schrödinger’s MMGBSA as described in Section 3.3 and Section 3.5.10. Despite its limitations, rDock was used due to its useful implementations of constrained docking (tethered and pharmacophoric) and due to its high-throughput VS capabilities.

$$S^{total} = S^{inter} + S^{intra} + S^{site} + S^{restraint} \quad (3.1)$$

$$S^{inter} = W_{vdw}^{inter} S_{vdw}^{inter} + W_{polar}^{inter} S_{polar}^{inter} + W_{repul}^{inter} S_{repul}^{inter} + W_{arom}^{inter} S_{arom}^{inter} + W_{solv} S_{solv} + W_{rot} N_{rot} + W_{const} \quad (3.2)$$

$$S^{intra} = W_{vdw}^{intra} S_{vdw}^{intra} + W_{polar}^{intra} S_{polar}^{intra} + W_{repul}^{intra} S_{repul}^{intra} + W_{dihedral}^{intra} S_{dihedral}^{intra} \quad (3.3)$$

$$S^{site} = W_{vdw}^{site} S_{vdw}^{site} + W_{polar}^{site} S_{polar}^{site} + W_{repul}^{site} S_{repul}^{site} + W_{dihedral}^{site} S_{dihedral}^{site} \quad (3.4)$$

$$S^{total} = W_{cavity} S_{cavity} + W_{tether} S_{tether} + W_{nmr} S_{nmr} + W_{ph4} S_{ph4} \quad (3.5)$$

3.3 Background Schrödinger's MMGBSA

Unlike empirical scoring functions that are parameterized to correlate with binding free energies, MMGBSA uses force fields to study the movement of atoms relative to each other based on the laws of physics, after which the binding free energy can be derived through statistical thermodynamics. On the other hand, MMGBSA is an end-state method, which is a crude approximation of the binding free energy compared to more rigorous methods such as thermodynamic integration (TI) or free energy perturbation (FEP), where multiple trajectories of the complex, receptor and ligand are simulated. Nonetheless, for VS applications such as the one described in this thesis, computational efficiency is favored. The MM energy function describes bonded (stretching of bonds, bending of angles and torsions) and non-bonded (van der Waals and electrostatic) interactions. In principle these energy calculations should be performed separately for the complex, the apo form of the receptor and dissociated ligand in order to obtain the relative free binding energy (Equation 3.6). However, with Schrödinger's MMGBSA, only geometry of the complex is used after which the corresponding groups are separated, as it's known to produce less noise due to insufficient cancellation of intramolecular interactions when using independent geometries.[144] Furthermore, the calculations are performed in an implicit solvent model that represents solvent as a continuous medium.

$$\Delta G = G_{complex} - (G_{receptor} + G_{ligand}) \quad (3.6)$$

Schrödinger's MMGBSA uses the OPLS force field[145], along with the VSGB 2.0[146] solvent model to estimate the binding free energy according to Equation 3.7. The force field accounts for the enthalpic contribution (ΔH) to ΔG_{bind} and is meant to describe the energetic gain or loss when the protein and the ligand interact. The solvent model is composed of a GB (electrostatic) and an SA (nonpolar) part (Equation 3.8). The GB model calculates the electrostatic terms using a lowered dielectric constants of the solvent molecules, while the accessible surface area (ASA) accounts for non-electrostatic solvation terms. As the

ASA term is empirically adjusted to fit the experimental free energy data, the GBSA term implicitly accounts for entropic solvation effects. The entropic term (S) can be calculated through the calculation of vibrational frequencies, but is often neglected as it typically gives the largest statistical errors.[147]. Together, these terms are meant to describe the gain or loss of free energy when solvent interacting with the protein is replaced by interactions with the ligand.

$$\Delta G = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (3.7)$$

$$\Delta G_{sol} = \Delta G_{GB} + \Delta G_{SA} \quad (3.8)$$

Though the method has important limitations, it has been widely used in structure-based drug discovery, with mixed results.[147] Nonetheless, in regards to the development of the pipeline, MMGBSA is used as a crude filter that accounts solvation effects and only account for the solvation term, not accounted for by rDock.

3.4 Background Dynamic Undocking

As stated in Section 3.1, structural stability is equally important to consider, as known ligands are not only known to bind with high selectivity and affinity, but are also known to form structurally stable complexes.[147][148] Figure 3.1 is meant to illustrate this, namely that two high affinity ligands will both have low ΔG_{binds} , however, the more robust complex will be that with a steeper energy minimum. This can be quantified by displacing the ligand from equilibrium in the energy minimum to the exact state in which measured key interactions have been broken, namely, the quasi-bound (QB) state. This quantification of structural robustness by DUck has been shown to work extremely well in VS, and is meant to be used as an orthogonal approach to thermodynamic-based methods. For the method, H-bonds are used to assess structural robustness as they provide strict distance and angular dependencies, and are abundant in protein-ligand complexes. Steered Molecular Dynamics (SMD) is used to displace the ligand from its optimal position. SMD uses the distance from bound to the QB state between the H-bond interacting atoms as the collective variable, during which the force is recorded and is used to calculate the work (W_{QB}). The reported W_{QB} values are thus a measure of the structural robustness of the measured H-bonds in the complex, which can be related to structural stability.

The DUck protocol starts with an equilibration to ensure a stable simulation. The protocol performs minimization for 1000 cycles, gradual warming of the system (100 K to 300 K for

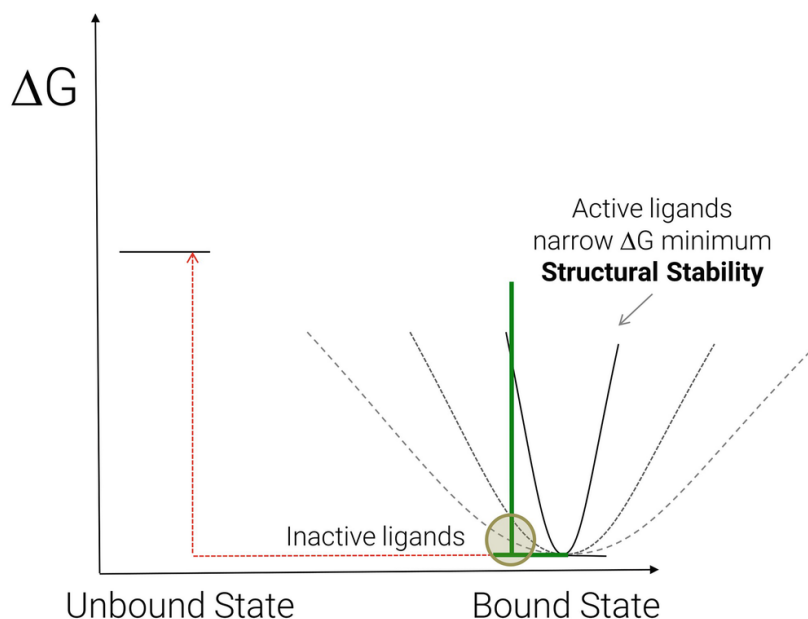


Figure 3.1: Structurally robust complex have a steep energy minimum, which can be quantified by DUck. Adapted from[149]

0.4 ns) at NVT and equilibration for 1 ns at NPT. Then the first steered molecular dynamics (SMD) simulation starts, during which the distance is increased by means of a harmonic potential (constant velocity and spring constant) from 2.5 Å to 5.0 Å during 0.5 ns. For computational efficiency, at this point, the protocol already measures the W_{QB} , so stops the process if the value is below a given threshold. Then the protocol performs non-steered MD also starting from the equilibrated trajectory, during which spontaneous rupture of the H-bond is prevented by applying gradual restraints for distances beyond 3 Å, after which W_{QB} is measured again. Finally, the protocol performs SMD at two temperatures (300K and 325K), and non-steered MD starting from the previous non-steered MD trajectory for as many replicas as specified. The W_{QB} is measured after both SMD runs, as well as after non-steered MD. During every simulation, the protocol applies harmonic restraints to all heavy atoms of the receptor to prevent structural changes. The DUck protocol uses MOE[150] to automatically prepare the scripts for the simulation, and to prepare the structure (AMBER force field 99SB[151]) and ligand (Parm@Frosst[152]) parameters.

The protocol described above uses relatively short simulation times, suitable for the time required to break an H-bond. Furthermore, the protocol only uses part of the protein, referred to as the chunk, which significantly reduces simulation times. The chunk is meant to preserve the local environment such that a W_{QB} value would be obtained, as if the full protein was used. When selecting residues for the chunk, the following should be considered: I) simulating as little residues as possible to reduce computational time, II) not selecting

residues that would block the ligand from exiting the pocket, III) not removing residues that may cause solvent entering the pocket from areas other than where the ligand is exiting.

3.5 Automated Pipeline Protocol

3.5.1 Structure Preparation

The structure must be prepared with or without conserved waters or cofactors. In every case here, MOE V.2016[152] was used to prepare the structures. The default settings were used in the “Structure Preparation” module, which caps (acetylates/methylates) C- and N- termini residues to avoid artificial charges, adds missing residues through homology modeling, and protonates the atoms at pH 7.

3.5.2 Interaction Analysis with MDMix

If the target is not well-known, an analysis of the complex to identify key H-bond features is required, as these will be used to evaluate structural robustness. In this thesis MDMix[153] aided in the identification of possible interaction sites. The MDMix protocol involved solvation of prepared structures in water, a mixture of 20% ethanol in water and a mixture of 20% acetamide in water in a truncated octahedral box constructed from replicas of a pre-equilibrated box of the same solvent mixture. These solvents were selected based on their ability to recapitulate features (polar and apolar) of drug-like compounds. Then, the MDMix protocol performs equilibration of the solvated system (temperature increase every 800ps to reach 300K NPT and 1ns NVT at the final temperature). Next, three simulations of 20ns NPT are carried out for each solvent combination (water, ethanol-water, acetamide-water), storing atomic coordinates every picosecond. During the simulation heavy atoms of the protein are restrained with soft harmonic potentials. Analysis of the trajectories consists of comparing the observed density with the expected density of polar and apolar probe atoms in small grids. The inverse Boltzmann relationship then converts the differences to binding free energies of the probed atoms. The final output is a collection of possible polar and apolar hotspots at varying energetic values.

3.5.3 Pose Clustering

If no crystal structure of the fragment hit was available, all plausible binding modes were derived. The fragment hits were docked for 100 runs using key H-bonds as pharmacophoric restraints, with a radius of 1, and a tolerance of 0.5 Å. Hereafter, the generated poses were

clustered using the `sdrmsd` function of rDock, where the best scoring pose is used as the first pose to measure the RMSD. If the next best scoring pose had an RMSD above 2 Å, the pose would be added to the second cluster, and become the ligand for comparison for the next cluster. DUCK was performed on the defined pharmacophore H-bonds on fragment poses with the best rDock S^{total} scores.

3.5.4 Library Preparation

For the projects described here, two libraries were used as the search space, an “In-Stock” and “Clean” subset of ZINC15[13], comprising 10,775,814 ligands, as well as an “On-Demand” and “Clean” subset of ZINC15, comprising 389,472,892 ligands, both of which were downloaded on October 9, 2017. The downloaded smiles were then split according to the number of heavy atoms, and subsequently converted to the fingerprints described in Section 3.5.5.

3.5.5 Similarity Searching

Similarity searching provides an efficient strategy for searching large libraries, however, the definition of similarity between two molecules depends on how the molecules are described. Molecular descriptors can be 1D (whole molecule descriptors), or representations based on the 2D or 3D description of molecules. A single 1D descriptor is usually not sufficiently discriminating to allow meaningful comparisons of molecules, while 3D descriptors are more complex since they must take into account that molecules are conformationally flexible.[154] At the same time, 3D fingerprints do not always outperform simple fingerprint similarity search. Substructure-based descriptors characterize a molecule, either by its 2D chemical graph or by its fingerprint. Encoding chemical graphs significantly reduces search efficiency, thus we've opted for fingerprint searching. There are two main ways of encoding substructures in a fingerprint, namely, dictionary-based (i.e. substructure keys-based e.g. MACCS) and topological-based (path-based e.g. Daylight or circular e.g. MORGAN).[155] In dictionary-based fingerprints, each bit position is associated with a fragment and accounts for its presence or absence. This makes these fingerprints chemically intuitive. In contrast, hashed fingerprints work by analyzing all the fragments of the molecule following a path up to a certain number of bonds, and then hashing every one of these paths to create the fingerprint.[156]

There are several studies comparing fingerprints in their ability to discriminate between actives and inactives. The developed protocol uses permissive similarity searches to retrieve near and far neighbors in a reduced search space, where after structure-based methods are

used for discriminating actives from inactives. As such, any fingerprint with a low similarity measure would suffice. We've opted for MACCS fingerprints with 0.5 Tanimoto similarity, since it's chemically intuitive and widely used. Furthermore, we've implemented the MACCS FastSearch method of Openbabel[157] (an indexed file where the fingerprints are stored in binary format) for optimal computational efficiency. On the other hand, even if a low similarity measure is used, another fingerprint can lead to a slightly different search space. For this reason, MORGAN fingerprints (radii 1 and 2) were also implemented using RDKit, although a lower threshold than 0.5 should be used in order to be as permissive.[158] For the protocol, the parent compounds are automatically converted to the same fingerprint and format. The platform then performs the similarity search on the split library for ligands that have only 2 heavy atoms more (user defined) or less compared to the parent compounds.

3.5.6 Ligand Preparation

Proper treatment of ligands for docking is imperative to success.[159] In summary, our current ligand preparation steps include pretreatment (removal of salts, neutralization), ionization, tautomer and stereoisomer generation, ring conformer generation, and minimization. Of these, ring conformer generation and minimization are the least straightforward.[160][161][162][163] Minimization whilst maintaining stereochemistry is usually not included in open source software. Furthermore, open source software often use templates for ring conformer generation, therefore do not accurately generate conformers for which no similar templates are available. Lastly, only few programs unify all these ligand preparation steps. As Schrödinger's LigPrep[164] does all of the above, it is used to prepare the smiles selected after the similarity search. The parameters were so that ligands above 300 atoms would be ignored, at most eight stereoisomers would be generated, at most six tautomers would be generated, at most eight ring conformers would be generated and lastly probable ionization states within the pH range of six to eight would be generated. The platform performs LigPrep on-the-fly only for ligands that are retrieved after MCS extraction (Section 3.5.7).

3.5.7 MCS Extraction and Superposition

After ligand preparation, the automated protocol extracts the Maximum Common Substructure (MCS) between the fragment and the similarity hits with RDKit. It uses the settings for maximizing bond overlap and ensures that for matching bonds, the bond order should also be exactly equal. Several options can be defined by the user, namely, if ring atoms should match ring atoms and if only complete rings should be matched. Furthermore, the user can specify the minimum percentage of atoms that the analogues must have in common

with the parent, or with a previously defined core (“MCS cutoff”). The automated protocol then performs flexible or rigid superposition of the automatically derived MCS. Flexible superposition uses RDKit’s constrained embedding and rigid superposition uses OpenBabel’s SMARTS matching.

3.5.8 Tethered Docking

As the first step, the initial fragment uses the reference ligand method to determine the cavity. Then, after flexible or rigid MCS superposition, the protocol performs high throughput VS (HTVS) for optimal efficiency. The HTVS protocol consists of three stages, in which the number of docking cycles for each ligand increases, and the cutoff for the score becomes more strict, at every stage, in order to early on remove bad scoring ligands. Three HTVS protocol templates are available for use, shown in Figure 3.2. The HTVS protocol in the first iteration is user defined and may depend on the rDock score of the starting fragment. During tethered HTVS, the program samples the molecule’s poses whilst the maximum translational deviation of the MCS is set to 0.001 Å and the maximum rotational degrees of freedom of the MCS is set to 0.001 degree per cycle of the genetic algorithm in rDock.

The final deviation will typically be very small, however, the user can specify a cutoff for the RMSD of the superposed MCS. The default for this cutoff is $< 0.2\text{Å}$, so the user may desire a more permissive tethering. Furthermore, the MCS is ensured not to be more than 3 Å away from the defined pharmacophore used for docking. This is to prevent cases where the MCS is not involved in the main interaction, but a feature in another part of the molecule can still interact. So, rDock penalizes the molecules if the extracted superposed MCS does not adhere to this restraint, for which the penalty cutoff is defined by the user, while the pharmacophore radius is set to 1Å with a tolerance of 0.5 Å. At least one mandatory pharmacophore restraint always corresponded to a restraint for the H-bond interaction used to assess structural robustness by DUck. Lastly, compounds are filtered if the RMSD between the MCS in the final docked pose and the superposition is larger than a user specified value, using `fconv`[167]. The protocol does tethered docking for every protomer and stereoisomer generated by LigPrep which pass the MCS cutoff (described in Section 3.5.7). Ultimately, only the best scoring isomer is selected for the next step.

3.5.9 Diversity Selection

As stated in Section 3.1, the platform makes a diverse selection after tethered docking, MMGBSA and DUck. The diversity selection is done using RDKit’s MaxMin algorithm[165], which generate descriptors (in this case MORGAN, radius 3) for all the molecules, and then

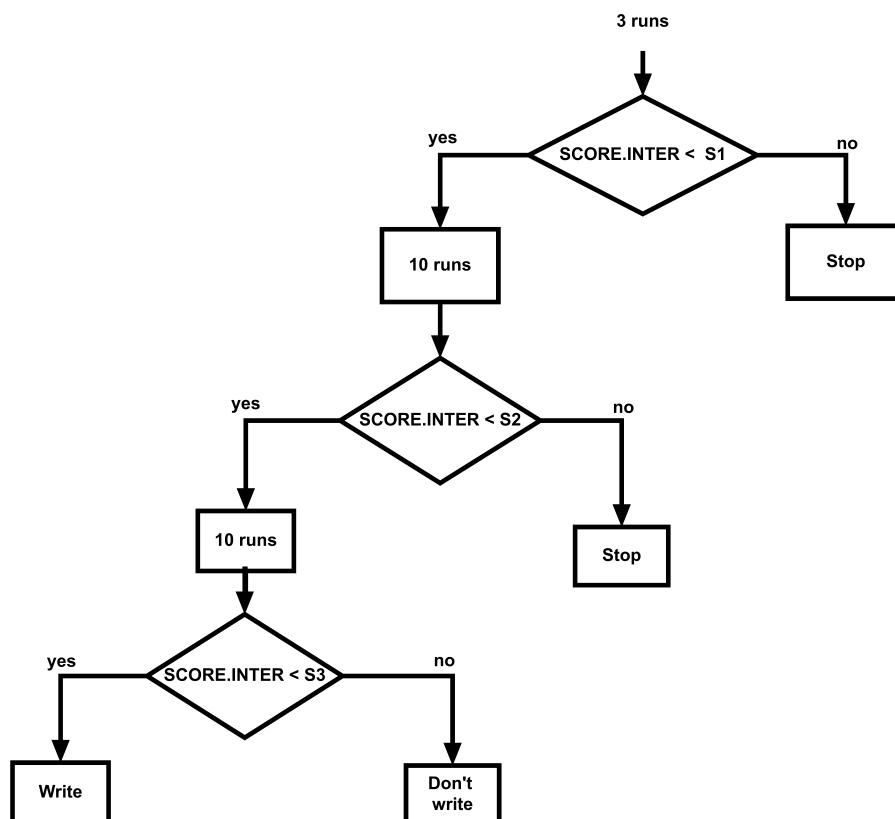


Figure 3.2: HTVS Protocols. Filter 1: S1 = -5, S2 = -8, S3 = -8. Filter 2: S1 = -8, S2 = -13, S3 = -15. Filter 3: S1 = -13, S2 = -15, S3 = -20

from a list of molecules finds the one that has the maximum value for its minimum distance to the picked set. The selected molecule is the most distant one to those already picked so is transferred to the picked set. The last step is iterated until the required number of molecules have been picked. This approach was taken because it is faster than Butina clustering, as only a subset of the matrix is generated. Additionally, in the first iteration, the picked molecule was defined as the molecule with the best S^{inter} .

3.5.10 MMGBSA

For ligands that pass the docking step, the protocol makes a diverse selection of 1000 molecules for the MMGBSA calculations. Schrödinger's Prime MMGBSA[166] with minimization is used, and only the ΔG_{sol} term in Equation 3.8 is used for filtering. If the score of the child is equal or lower than the score+10 of the initial fragment, for example, if the initial score was -39 kcal/mol, and that of the child, -29 kcal/mol, then the child can still pass to the next step. This is to encourage the scores to become better in each cycle,

while not excluding candidates that score slightly worse. Lastly, the RMSD between the cores of the docked pose and the minimized pose is calculated with `fconv[166]` for which the cutoff is user defined.

3.5.11 DUck

For DUck, a diversity selection of 500 compounds that pass the MMGBSA step is made. The platform performs 5 SMD and 5 MD replicas of DUck on the defined H-bond. The chunk residues were selected from the prepared structures. After selection of the residues for the chunk (described in Section 3.4) on the prepared structures, each section of residues was split into separate chains, where after each chain was acetylated and methylated, and lastly, checked for clashes possibly created during capping of the chains. The protocol performs DUck with a user defined threshold, which may depend on the W_{QB} value of the starting fragment. If more than 50 compounds pass the threshold, a final diversity selection of 50 compounds is made, which become the new parent molecules.

3.5.12 Subsequent Iterations

In subsequent iterations, the protocol performs the similarity search for each new parent molecule, however, the similarity hits are limited to the most representative set of the best scoring parents. Each parent is assigned a “parent score”, which consists of the ranking according to rDock, MMGBSA and DUck. The average of the ranking in the three methods is then summed for each parent that produced the child during the similarity search. The list of children is then sorted by the sum of the average rank of the parents and the top-ranked children scores and the top (100,000 for ZINC “In Stock” or 500,000 for ZINC “On-Demand”) are selected for MCS derivation. The parent that is used for superposition corresponds to the parent that has the largest MCS. After iteration one, the HTVS docking protocol depends on the rDock score of the parent, such that the docking score cutoffs in the HTVS protocol changes to a parent score + 10 in stage 1, a parent score + 5 in stage 2, and a parent score in stage 3. For MMGBSA, the filters are applied in the same way as for iteration one and for DUck, the threshold for iterations over one, is set to the truncated parent W_{QB} values.

Chapter 4

Results

Section 4.1 gives a summary of the methods described in Chapter 3, while Section 4.2, Section 4.3 and Section 4.4 describe its validation, including its application to various targets in different scenarios. It should be noted that the platform has also been applied to several other projects that were not be described in this thesis. Additional work on side projects (submitted and published) are also not be described here, but have been added to the Appendix.

4.1 The Automated Pipeline

The platform can be viewed as a focused virtual screening (VS) of the chemical space surrounding a given fragment, conceptually depicted in Figure 4.1. A permissive similarity search performs the first stage of narrowing down the chemical space. Then, the platform performs the VS iteratively, where in every iteration analogues of increasing size are virtually screened.

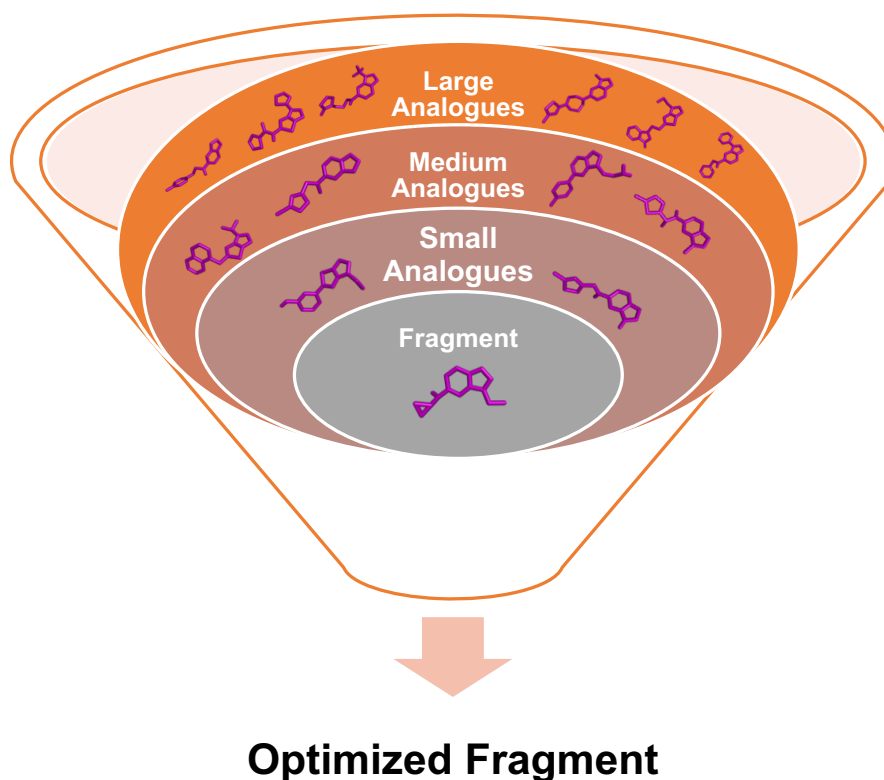


Figure 4.1: Conceptual depiction of fragment focused chemical space virtual screening.

The best scoring analogues are retrieved, which determine the chemical space for the next round. In this way, the information (features, chemotypes) of the best scoring analogues are harnessed, thereby guiding the optimization towards promising areas of chemical space that are also related to the starting fragment. Furthermore, this iterative procedure is in line with the criteria of ligand efficiency (LE) maintenance in FBDD. LE-guided optimization aims to emphasize the most important features for molecular recognition, avoiding increases in potency that are a mere result of an increase in size.[8,9] Thus, by limiting the platform to screen only sets of analogues categorized by size, we can effectively control LE. The platform does this through encouraging the analogues to score better in every iteration, while screening analogues in small size increments.

Since the platform is iterative, it is essential to maintain the most likely binders, and to ensure the removal of highly unlikely binders. Otherwise, there is a risk of error propagation. Ideally, the virtual hits should be experimentally validated after every iteration, in order to guide the optimization towards the right chemical space. But this would make the exercise slow and unpractical. Instead, in this thesis we challenged the computational methods to distinguish binders from non-binders. As such, the platform uses three orthogonal structure-based approaches as depicted in Figure 4.2, namely, docking with rDock[144], where pharmacophoric restraints (known or predicted) are applied and only slight deviations of the maximum common substructure (MCS) between the fragment and analogues are permitted, II) Schrödinger's MMGBSA[166] (Molecular Mechanics energies combined with Generalized Born and Surface Area continuum solvation free energy) to account for solvation effects and III) Dynamic Undocking (DUck)[150], to assess structural robustness through the defined key hydrogen bonds (H-bonds).

A combined rDock and DUck approach has been shown previously to reduce false positives and retrieve higher experimental hit rates[150]. And, although MMGBSA uses some crude approximations, it has been shown to improve docking results.[147] Moreover, here it is used to explicitly take into account the change in solvation free energy between the bound and unbound states, which can be an important source of error in docking (unadequately accounted for by the scoring function) and DUck (the unbound state is not sampled). Figure 4.3 describes the platform in full. Starting from permissive similarity searching for analogues two heavy atoms more or less compared to the fragment, the maximum common substructure (MCS) is automatically derived, the 3D geometries of the analogues are prepared on-the-fly, and are subsequently subjected to the VS protocol, which is meant to retrieve the most likely binders. Then the platform performs the similarity search anew on the new parents. To ensure scalability, the analogues that are most representative of the best scoring parents continue to the subsequent steps. The platform does this by selecting analogues with the most parents,

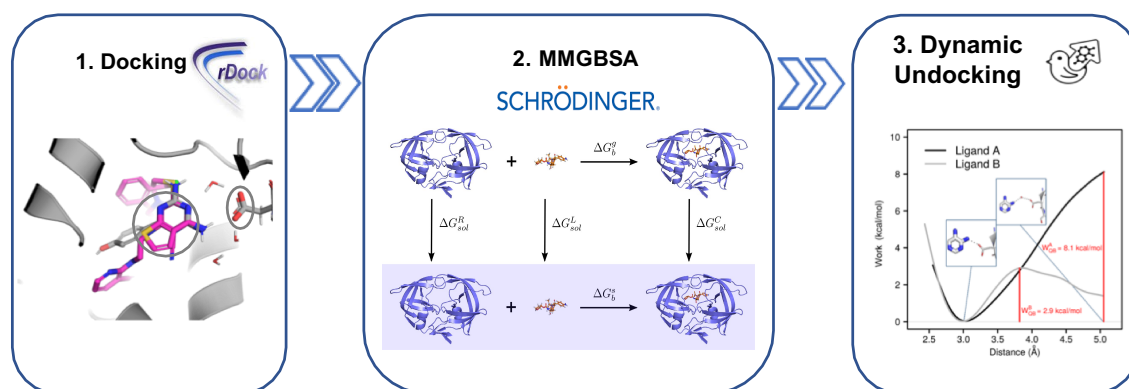


Figure 4.2: VS filtering with orthogonal approach. 1) Docking maintains the core or maximum common substructure, and key H-bonds. 2) MMGBSA accounts for solvation effects, and 3) DUck ensures structural robustness through the evaluation of the stability of key H-bonds.

taking into consideration their ranking during docking, MMGBSA, and DUck (Figure 4.3 most left box).

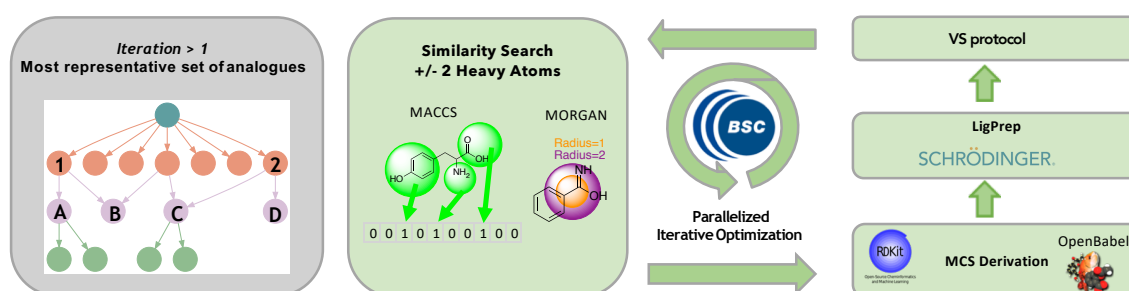


Figure 4.3: The developed pipeline. An example when selecting the most representative set of the best scoring parents: the top 3 analogues would be: child A, B and C if parent 1 scores better than parent 2, or child B, C and D if parent 2 scores better than parent 1.

Further limitations include diversity selections after docking, MMGBSA, and DUck, if necessary. Lastly, each step is performed in parallel to ensure scalability. Each iteration takes approximately 24 hours using the Barcelona Supercomputing Center based on a search space of 107, but performance may vary depending on the neighbor density of the starting fragment. Nonetheless, multiple fragment optimization projects can be run simultaneously.

4.2 The Platform Applied to NUDT21

In this study, an X-ray fragment screening campaign at XChem led to fragment hits for the NUDT21 protein, which has been found to be associated with both oncogenic and tumor suppressing roles.[167] CFIm25 (Cleavage Factor 25kDa) encoded by NUDT21 contains a Nudix domain, typically found to hydrolyze ATP and Ap4A.[168] Instead, CFIm25 has been found to bind exclusively to larger CFIm homodimers (CFIm68 and CFIm59), and RNA (sequence specific recognition of the UGUA motif by the Nudix domain).[169] The tetrameric complex CFIm is known to be essential for pre-mRNA processing.[171] We aimed to optimize the fragment hits (with almost unmeasurable potency) in collaboration with XChem in order to confirm the protein's ligandability.[14][170] We've done this using two distinct strategies, namely using the poised library concept (described in Section 1.3.1.2)[41], as well as, using the developed pipeline. The results, as well as the implications and benefits of using one approach over the other will be discussed.

4.2.1 The Putative Binding Site

The X-ray fragment hits were used to evaluate interaction sites, which can be seen in Figure 4.4. As stated in Section 3.4 the platform uses H-bonds to assess the structural stability of the possible binders, therefore a thorough interaction analysis is required especially if the target is unknown. In the structure, only one common interaction point was observed, namely, between fragment hit X-0401 and X-0404 which both make a water-mediated interaction with Tyr191.

Other interaction points with X-0404 involve Lys105 and the backbone oxygen of Gly108, while fragment hit X-0401, is seen to interact with Lys56 and Glu110. Fragment hit X-0581 is only seen to interact with the backbone of Leu106 via a dual H-bond, however, when the structure is minimized, an additional interaction with Lys105 is possible. Furthermore, a network of interactions comprising two waters and the four residues, Arg77, Thr79, Gln157, Glu170 and Lys172 is observed. These two waters, one of which is able to interact with fragment X-0581, were conserved during the VS protocols.

Mixed solvent MD (MDmix) also aided in the identification of interactions hotspots. The results of the simulation with structure X-0581 (ligand excluded from the simulation) is shown in Figure 4.5.

MDmix identified an acceptor hotspot near interacting residue Gly108 and a donor hotspot and water site near interacting residue Leu106. Waters are present in the structures of fragment hits X-0401 and X-0404 in the position of the identified water site. This water is displaced in the structure of fragment hit X-0581.

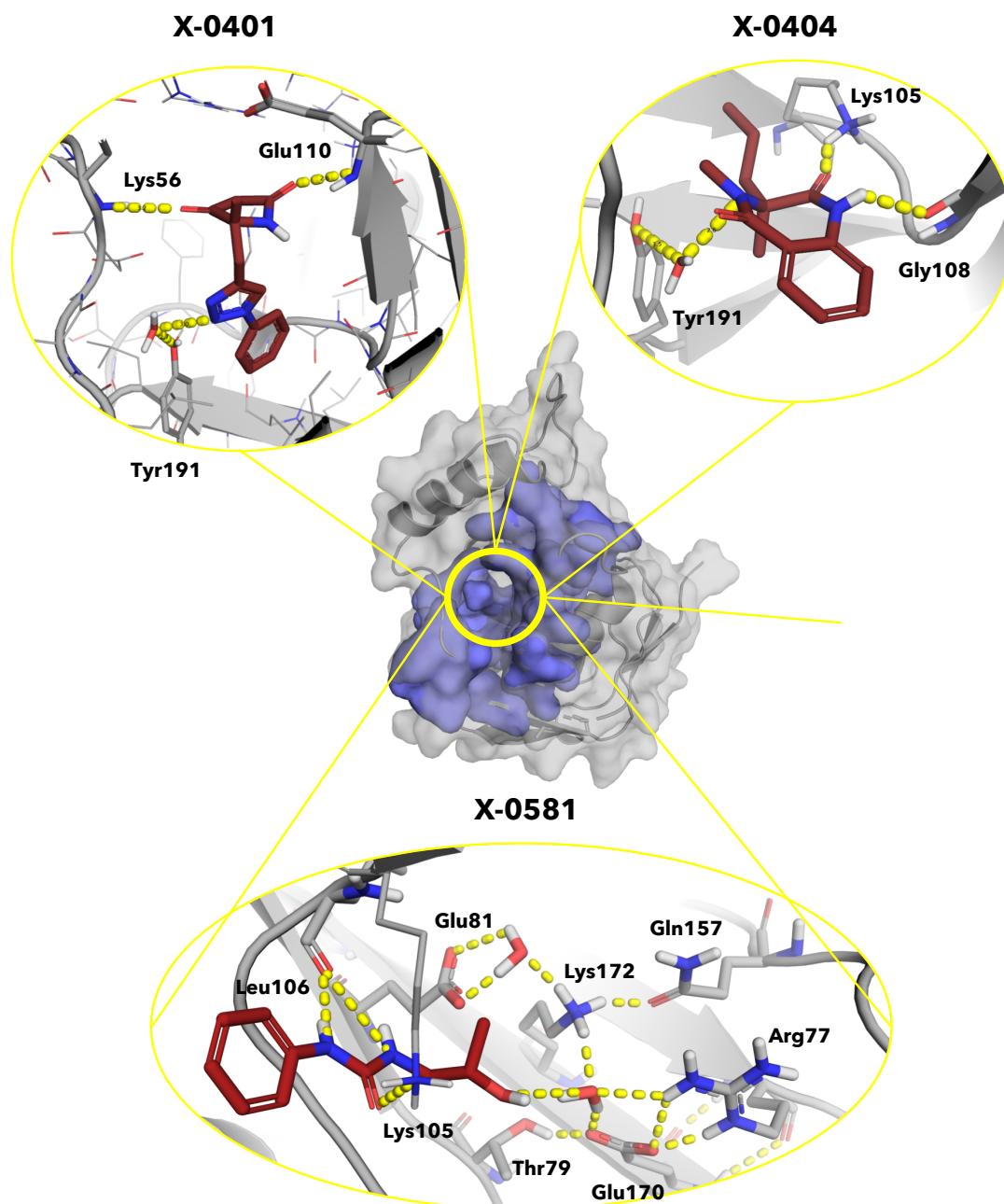


Figure 4.4: NUDT21 Fragment Hits: X-0401 (Chain A), B) X-0404 (Chain B), C) X-0581 (Chain A). The minimized protein structure of X-0581 is shown to interact with the ligand through Lys105.

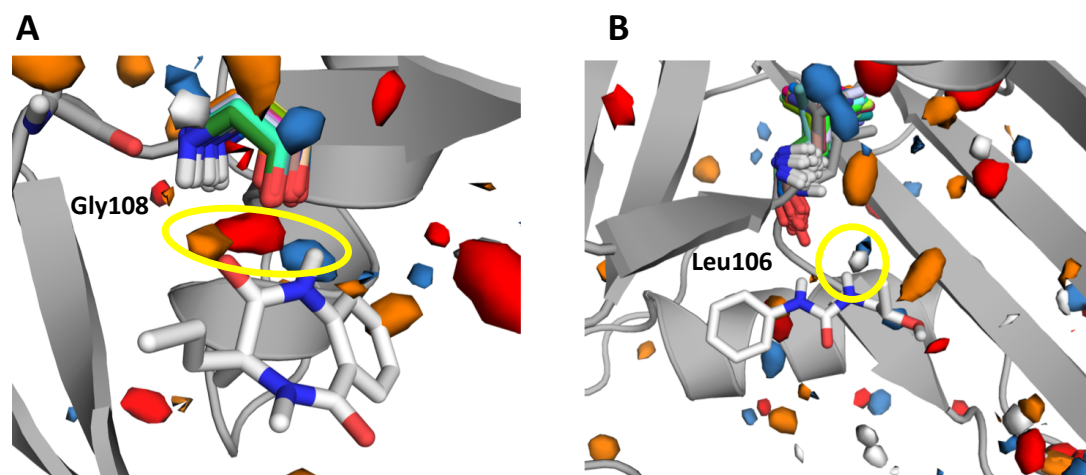


Figure 4.5: Consensus interaction hotspots derived from MDmix. Red=acceptor probe from acetamide, Blue=donor probe from acetamide, Orange=donor/acceptor (hydroxyl) probe from ethanol, White=water probe. It should be noted that, during MDmix simulations, residues were treated as flexible, therefore hotspots may appear shifted.

Duck on the interaction with Gly108 resulted in a W_{QB} value of 0.28 kcal/mol, and for Leu106, a W_{QB} value of 2.44 kcal/mol. These results indicate low structural robustness, however, the observation of these interactions in the experimental structures, and the consensus with the MDmix results gave confidence to use them as the key interactions for optimization, while the W_{QB} threshold was set to 2 kcal/mol for the first iteration. At the same time, as no consensus hotspots were identified for fragment hit X-0401, it was not further considered.

Besides the difference in interactions between the fragments, the structures also differed in conformation of the putative binding site. For a fragment-based approach, exploiting structural information is highly beneficial, however, structure-based design that does not account for flexibility will be biased towards selection of the structure. In Figure 4.6, a previously published x-ray structure (3Q2T)[173] containing the tetrameric complex of CFIm68 and CFIm25 bound to RNA is superposed with the structures retrieved in this study, along with a previously published structure (3BHO), containing CFIm25 bound to bis-adenosine-5'-tetrphosphate (Ap4A). A low RMSD of 0.69 Å was found for the CFIm25 domains. The fragment hits are found deep within the same pocket RNA and Ap4A are found to bind superficially (only fragment hit X-0404 is shown in Figure 4.6), however, no common interaction points are observed.

Interestingly, the fragment co-crystal structures display distinct conformations of the pocket, compared to conformations observed in previously published structures. The most

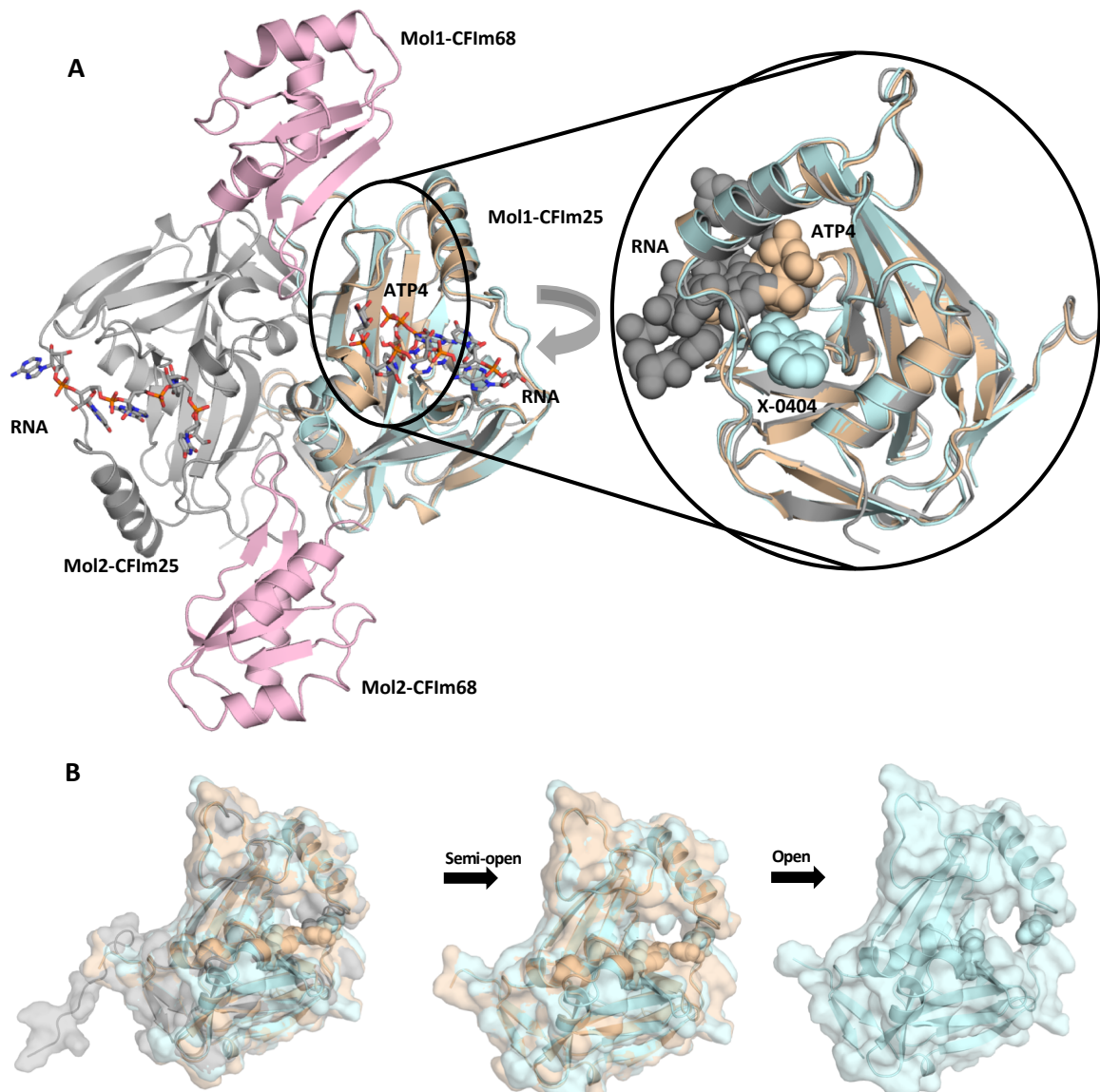


Figure 4.6: A) The tetrameric complex (3Q2T) of one homodimer of CFIm68 and one homodimer of CFIm25 bound to RNA aligned to a monomer of CFIm25 bound to Ap4A and the co-crystal structure of fragment hit X-0404. Zoomed in, fragment hit X-0404 is seen to bind deep in the pocket RNA and Ap4A superficially bind (ligands shown in spheres). B) From left to right, RNA in 3Q2T (closed conformation), X-0401/X-0581 (semi-open conformation) and, the surface of the structure bound to fragment hits X-0404 (open conformation), with residues Lys105 and Glu55 shown in spheres.

divergent structure corresponds to the co-crystal of fragment X-0404, where the pocket displays the most open conformation (“open”). The co-crystal structures of fragment X-0401 and fragment X-0581 share near identical conformations and have a slightly less open pocket (“semi-open”), whereas the pocket in previously described structures bound to RNA and

Ap4A have a closed conformation (“closed”). Specifically, residues Lys105 and Glu55 largely contribute to the pocket having a closed or open conformation (highlighted in Figure 4.6B as spheres). Although MDMix indicated no hotspots to interact with Lys105, it is the only interaction fragments X-0581 and X-0404 could have in common. Furthermore, DUck indicated the interaction to contribute just as much to structural robustness (W_{QB} of 2.1 kcal/mol for X-0581 in its minimized structure and 1.75 kcal/mol for X-0404) as Leu106 and Gly108. Its role in the conformation of the protein and the reasons mentioned above gave incentive to incorporate an interaction with Lys105 as a secondary pharmacophoric restraint.

4.2.2 Comparing VS Strategies

Not uncommon for crystallography-based fragment screening, fragments X-0581 and X-0404 had almost unmeasurable potency for NUDT21 from biophysical assays (SPR). They were optimized with two distinct design strategies: the iterative workflow (as described in Section 4.2), and the poised synthesis approach (Figure 4.7) combined with one iteration of the VS protocol described in Figure 4.2. The comparison between protocols is particularly interesting, because both strategies can be fully integrated into streamlined fragment optimization workflows. At the same time, while both strategies exploit structural information through structure-based VS, they differ significantly in the design strategy and in their intrinsic capabilities of chemical space exploration. With the second approach, the size of the library will scale linearly with the time it takes to run the protocol and will always retain the central chemical moiety defining the ‘poised’ library. On the other hand, one of the main benefits of the developed iterative platform is its ability to search very large chemical spaces and to perform scaffold hops.

The poised synthesis approach is depicted in Figure 4.7 and led to 1100 compounds for X-0581 and 351 compounds for X-0404, which were prepared with Schrödinger’s LigPrep[172] as described in Section 3.5.4. The VS protocol described in Figure 4.2 led to 283 compounds for X-0581 and 38 compounds for X-0404. The protocol involved the same parameters as described for the automated platform. It also applied the same pharmacophoric features (for docking and DUck), and used the same chunk (residues 52-54, 77-82, 96-99, 130-111, 124-125, 127-129, 172-174 and 187-188). After visual inspection based on diversity (meta vs. para position and the possibility of additional interactions), 19 compounds were selected for X-0581, and 4 compounds were selected for X-0401. Out of these compounds, only 7 could be synthesized (shown in the Supplementary Information), stemming from the X-0581 series.

The iterative approach used the ZINC “In-Stock” collection comprising roughly 11M compounds as the search space. Three rounds of iterative screening led to the selection

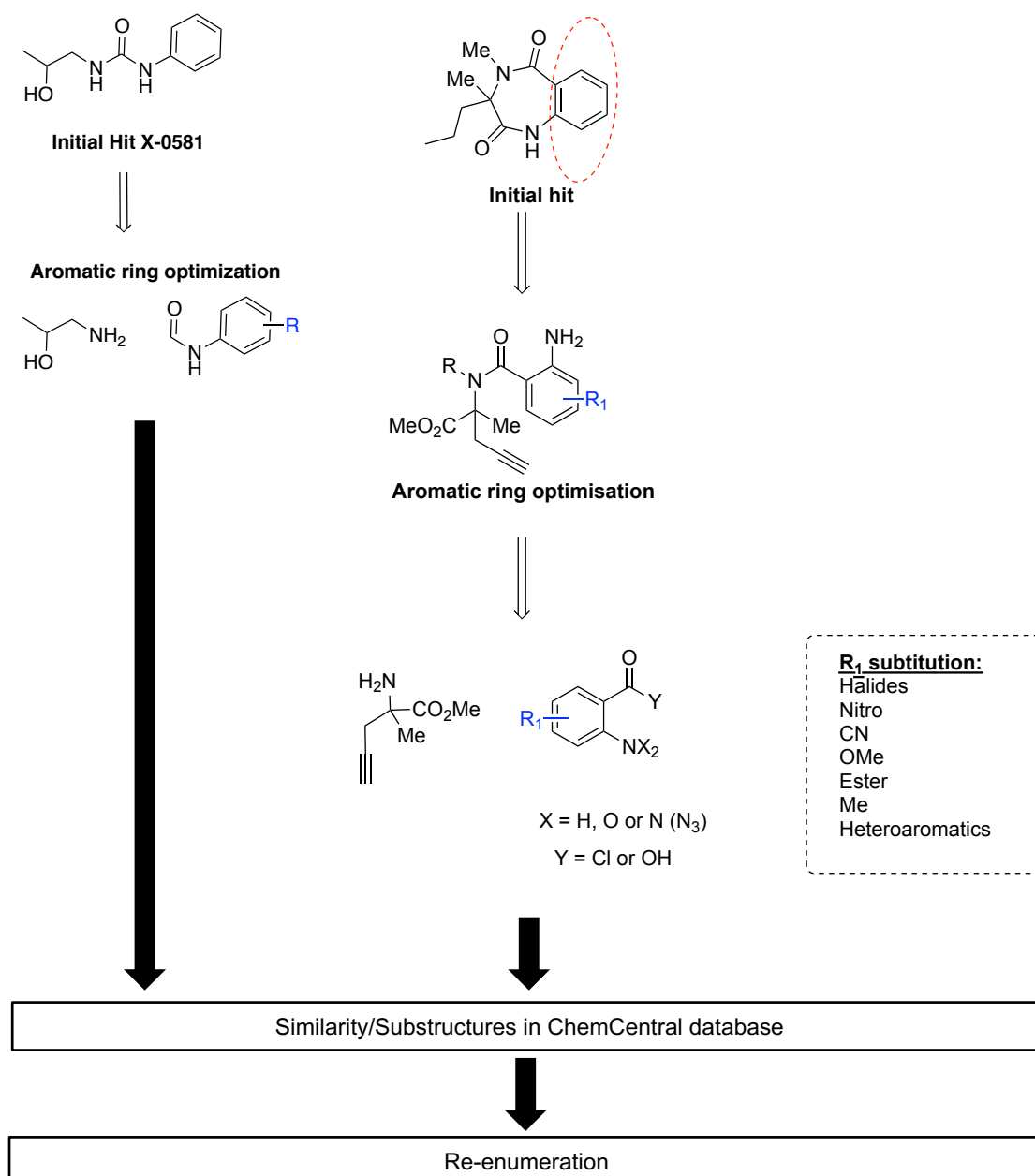


Figure 4.7: Poised fragment deconstruction based on chemical reaction. Substructure and similarity searches were performed in the ChemCentral database[171] (provided on the Squonk platform as searchable libraries).

of 41 compounds of which 37 were directly available for X-0581, and 8 compounds of which 8 were directly available for X-0404. Based on these results, the advantage of using commercially available chemical spaces is emphasized, even when comparing it to the use of explicit synthetic feasibility for library generation. Furthermore, no human intervention was necessary using the automated approach, therefore preventing any bias from influencing the results.

To help explain the difference in virtual hits found between X-0581 and X-0404, an overview of the attrition after each step of the automated protocol is shown in Table 4.1. It can be observed that the superpose step generally gives a larger attrition for X-0404 when compared to that of X-0581. This is mainly due to the fact that the platform removed compounds that had an RMSD larger than 0.2 Å between the MCS atoms. As many of the analogues actually contained a planar urea moiety, and the initial fragment contained a flexible ring scaffold (an extreme scaffold hop), the large attrition due to RMSD is expected. A significantly larger proportion is removed at the docking stage. This was then mainly due to the fact that the analogues did not score better than their parents.

Table 4.1: Attrition Rates VS Comparison

Iteration	1	2	3
X-0581			
1. Similarity	5,962	253,546	865,180
2. Superposed	3,257	34,050	32,368
3. Docked	2,315	12,536	12,550
4. MMGBSA	559	411	509
5. DUck	346	81	41
X-0404			
1. Similarity	112,126	905,843	1,880,142
2. Superposed	4,963	9,502	14,458
3. Docked	63	459	713
4. MMGBSA	11	87	39
5. DUck	10	47	8

In Figure 4.8a and Figure 4.8b, a comparison of rDock S^{inter} score distributions is shown. A tendency of the selected analogues to improve in every round of optimization can be observed for both fragments. However, starting fragment X-0581 had an rDock S^{inter} score of -11 kJ/mol, while X-0404 had an rDock S^{inter} score of -8 kJ/mol. The rDock HTVS protocols both applied “Filter 1” (depicted in Figure 3.2). Thus, the analogues of X-0404 had to score relatively higher than their parents, than the analogues of X-0581. For comparison

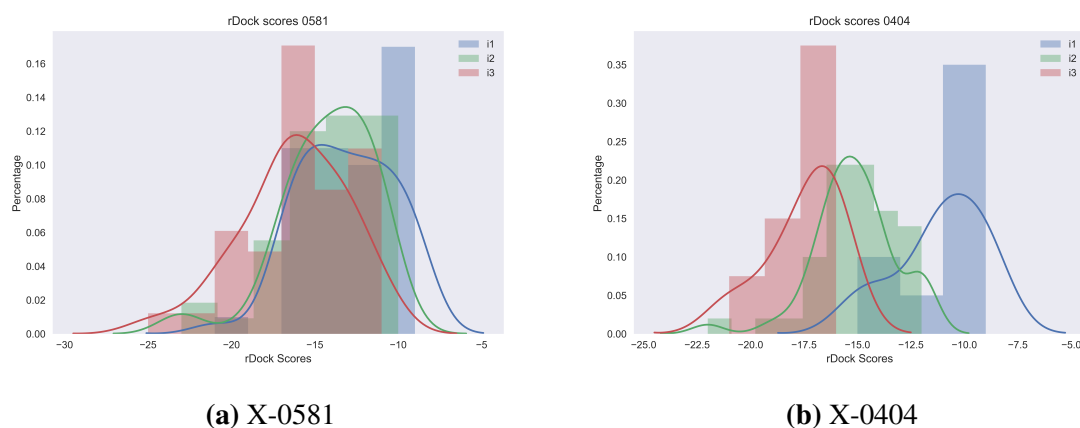


Figure 4.8: rDock Score Distributions of the Iterative Virtual Screening Approach.

to the poised approach, the structures of the 7 synthesized compounds are shown in the Supplementary Information along with rDock S^{inter} scores.

4.2.3 Experimental Results

A total of 52 compounds (37 from the iterative approach for X-0581, 8 from the iterative approach for X-0404, and 7 from the poised library approach) were selected for experimental testing by Surface Plasmon Resonance (SPR). Since these experiments were mainly carried out by another person, the methods will not be thoroughly described here, but SPR, in general terms is a biophysical technique that detects changes to the composition of a surface on which the protein is immobilized. Ligand binding effects are measured through differences in light refraction.[173] Each experiment was performed in duplicate in three separate channels, two with varying levels of protein, and one empty channel to control that ligands were not binding to the surface.

An efficient screening cascade was used by first screening the compounds at a relatively high concentration of 500 μM , where after compounds showing a positive response were subjected to dose-response experiments at 31.25 μM , 62.5 μM , 125 μM and 500 μM . Out of the 52 compounds that went through the cascade, 20 gave a positive response at 500 μM , of which 15 showed dose dependency. The dose response curves of the 15 compounds and the reference fragments (X-0581, X-0404 and X-0401) are shown in the Supplementary Information (empty channels not shown), Table 4.2 gives a summary of the results, and Figure 4.9 and Figure 4.10 give an overview of the compound structures and binding modes. It should be noted that many compounds suffered from aggregation at higher concentrations, thus saturation of the signal could not be met. So, K_D values could only be estimated, based on the theoretical

Table 4.2: NUDT21 SPR Result Summary VS Comparison

Binder	<500 μ M	500 μ M - 1mM	>1mM
Initial X-ray Binders			
X-0401			
X-0581			
X-0404			
Automated X0581 evolution			
B9			
C7			
C8			
D2			
D8			
A5			
B11			
B12			
C11			
C12			
Automated X0404 evolution			
B1			
B2			
C10			
A2			
A8			

maximum response (calculated from protein and ligand molecular weight). Furthermore, one-to-one binding was assumed i.e. that the ligands should not show unspecific binding to other putative binding sites.[<http://www.bio-rad.com/webroot/web/pdf/lsr/literature/6476.pdf>]

Surprisingly, all 15 compounds that passed the screening stem from the automated iterative VS approach, which comprises a hit rate of 27% for fragment X-0581, 63% for fragment X-0404, or a combined hit rate of 29%. Considering the much lower diversity of the compounds retrieved by the poised approach, which only introduced aromatic ring substitutions, it was reasonable to expect that some would bind. Nevertheless, this strategy may not be well-suited to evolve fragments that have almost unmeasurable potency. By contrast, the developed iterative VS approach successfully evolved the initial binders into hits that are still not very potent, but decent ligand efficient owing to their relatively small size. Moreover, the hits are quite diverse and provide a better starting point for further optimization. Interestingly, the hits evolved from fragment X-0404 resemble those from X-0581, although generally slightly larger, which would likely cause clashes with the used protein conformation.

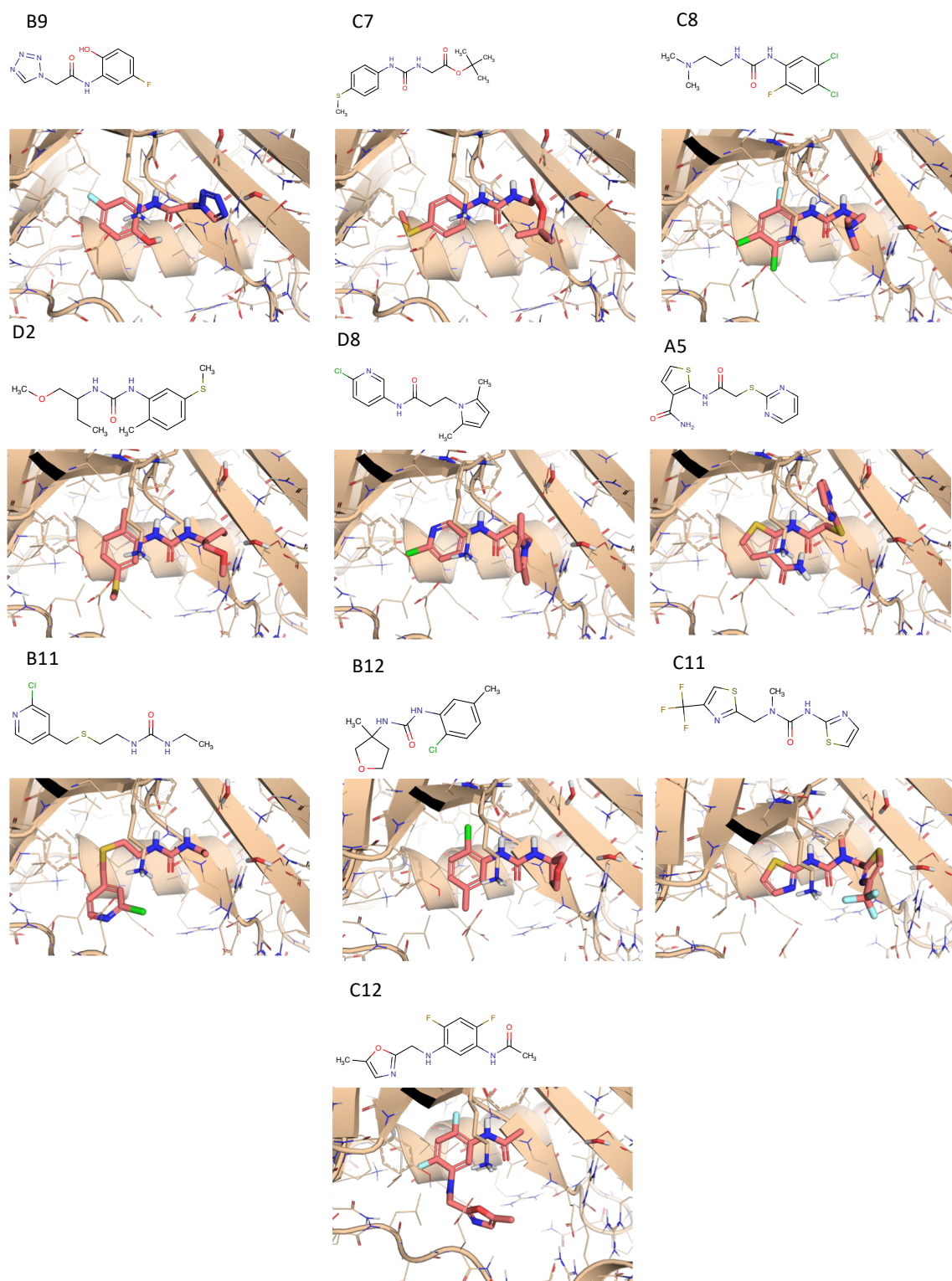


Figure 4.9: 2D structures and predicted binding modes of the SPR hits evolved from fragment X-0581.

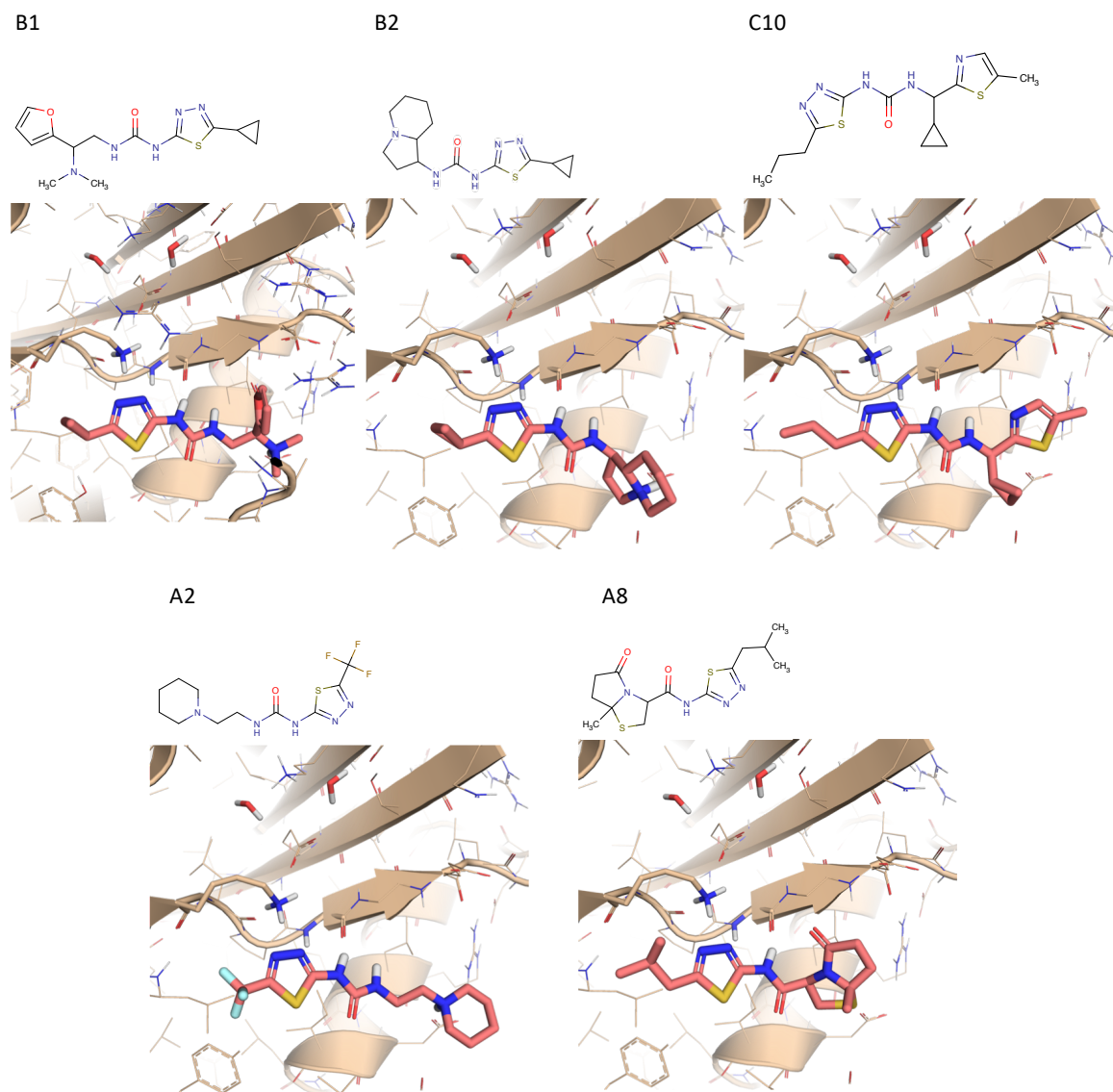


Figure 4.10: 2D structures and predicted binding modes of the SPR hits evolved from fragment X-0404.

In Section 4.3, however, we will discuss the adaptability of the platform to perform less extreme scaffold hopping. Besides the successful validation of the virtual screening platform, these results provide the most potent NUDT21 ligand described to date (D8; $K_D=200\mu\text{M}$; $\text{HAC}=19$; $\text{LE}=0.26$), are important to confirm the ligandability of this protein and for the future development of more potent chemical probes.

4.2.4 Future Perspectives

The next steps include crystallography of the compounds to detect binding with an orthogonal method and to confirm the predicted binding mode. While multiple crystallography attempts will be made for the compounds that were found to be active by SPR, it is not expected that there will be significant overlap between the SPR and x-ray hits. Rather, the combination of both screening methods will likely provide orthogonal hits[174]. However, we also aim to shed light on the biological effects these compounds could have. The first step includes RNA displacement assays. Since the fragment hits bind to the pocket that RNA is found to bind, there is reason to expect that they will affect RNA binding. Further design of compounds will be focused on RNA displacement through two strategies.

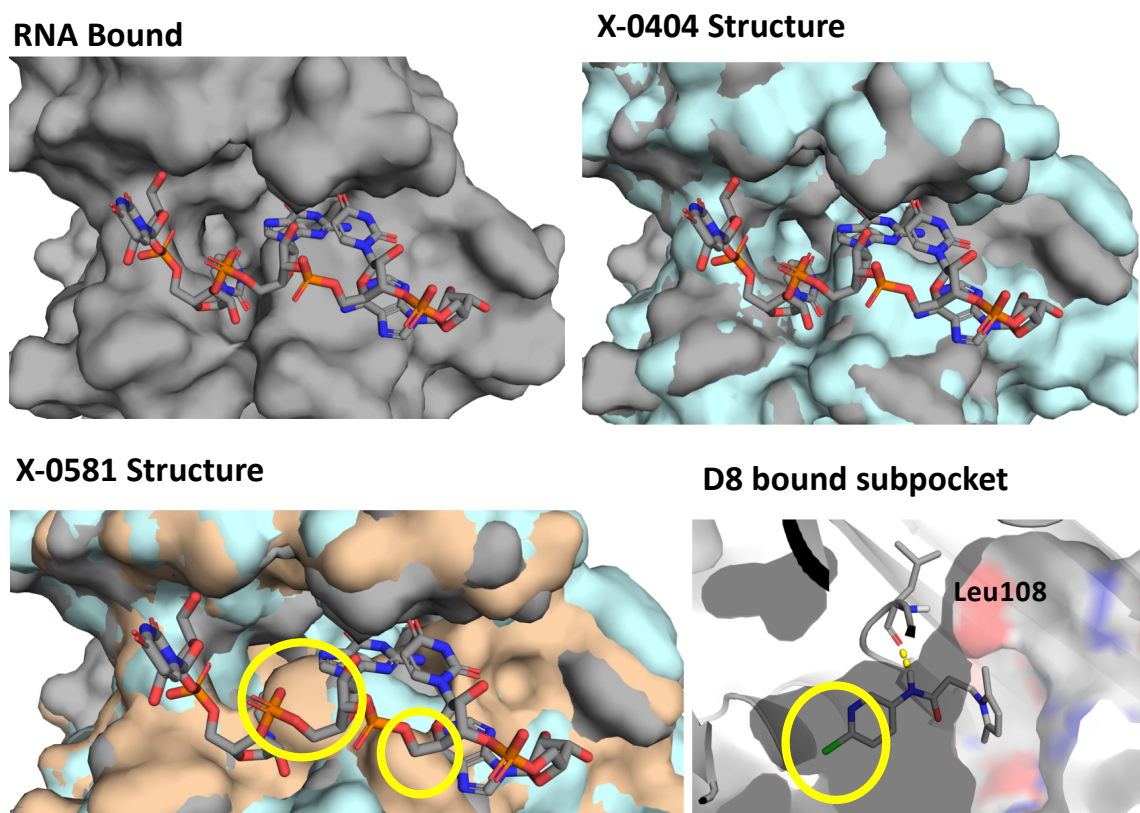


Figure 4.11: Conformational trapping strategy. Structure bound to X-0581 with a conformation that clashes with RNA site. Substituted aromatic rings in compounds derived from X-0581 hypothesized to induce conformational change.

The first study involves trapping the conformation in a state where RNA is unable to bind. As stated in Section 4.2, the structure of X-0404 has an open conformation and the structure

of X-0581 has a semi-open conformation. The latter would cause clashes with RNA binding (Figure 4.11). We hypothesize that the aromatic ring of X-0581 induces this conformational change. Thus, further design will focus on exploring the subpocket further.

The second strategy includes the design of ligands that would directly compete with RNA binding. As stated in Section 4.2.1, the fragment hits are found to bind deep within the pocket, and do not have overlapping interactions with RNA. However, there are other fragment hits, not described here, that do. The aim is to optimize the fragment that would directly compete.

4.3 Exploratory Strategies

In this section we describe the developed platform's adaptability to the user's desired design strategy. In Chapter 3 the methods were thoroughly described. Table 4.3 gives an overview of user defined parameters that can influence the design strategy. These settings can determine the degree of diversity or novelty that can be expected. At the same time, these settings can be tuned to be less stringent thereby allowing more compounds to be processed. The next section will give some examples of how settings can be changed for specific requirements.

4.3.1 Scaffold Hopping Strategies

Section 4.2.2 described the optimization of two NUDT21 X-ray fragments hits, X-0581 and X-0404. Optimization with the developed platform led to experimental hits for both fragments, however, for fragment X-0404 an extreme scaffold hop was observed (from a tetrahydro diazepinone scaffold to a urea scaffold). Figure 4.12 explains how this scaffold hop occurred. Figure 4.12A displays the child parent connections throughout three iterations of X-0404 optimization, and Figure 4.12B those of X-0581. The networks show that in the X-0581 optimization more diverse parents have led to the final selection, while the final selection from the X-0404 optimization stems from just two parents. Figure 4.12C shows the parent that has led to the most hits, containing a urea moiety, leading to the final selection of primarily urea scaffolds. The compound containing the tetrahydro diazepinone scaffold was not selected because the children with the most parents are prioritized. If the aim is to maintain the tetrahydro diazepinone scaffold, then the optimization can be biased by specifying that MCS extraction should consider that ring atoms must match other ring atoms, and that only complete rings must be maintained ("MCS ring match ring" and "MCS complete rings" in Table 4.3), which would translate to iterative substructure searching. This, however, led to most of the analogues being removed in the second iteration. Therefore,

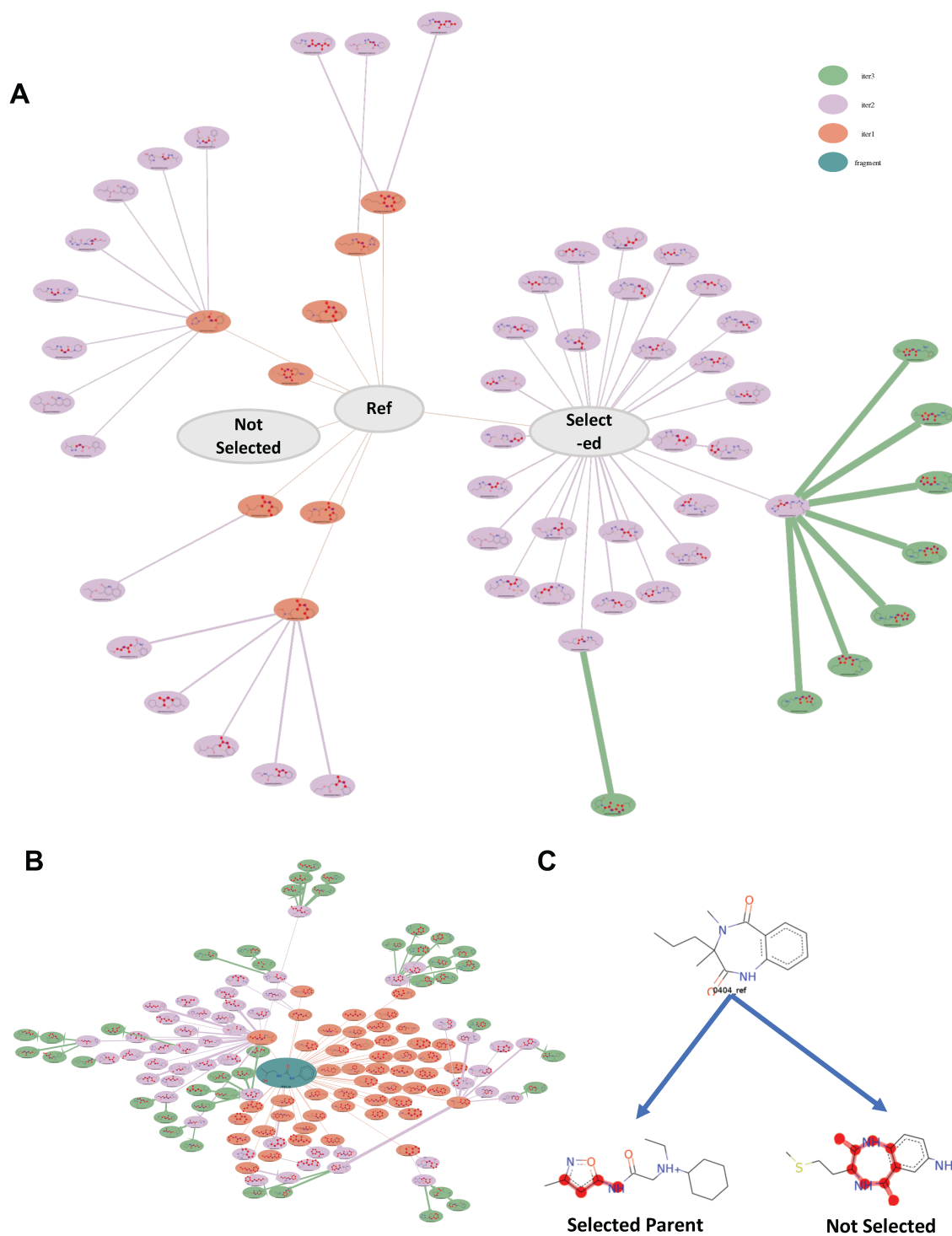


Figure 4.12: A) Parent child relationships for X-0404 evolution. B) Parent child relationships for X-0581 evolution. C) Scaffold hop to selected parent.

Table 4.3: The platform’s user defined parameters.

Settings	Default	Description
Library	In-Stock	Pre-prepared libraries (In-Stock or On-Demand).
Fingerprint	MACCS	Pre-prepared fingerprints (MACCS or MORGAN).
Similarity	0.5	Similarity search (smaller cutoff recommended for MORGAN).
Superposition	FLEX	RIGID recommended if the fragment is planar and FLEX recommended if the fragment is non-planar.
HA growth	2	The number of heavy atoms that the fragment size may increase per iteration. The maximum decrease is 2 HA.
MCS cutoff	0.5	The cutoff for the size of the MCS (percentage) compared to the children molecules.
MCS core	True	MCS extraction from previous MCS core or from parent.
Tethered RMSD cutoff	0.2	The RMSD cutoff for MCS superposition.
MCS ph4	0.5	The cutoff for which the MCS may deviate from the area defined as necessary.
Docked RMSD	1.5	The RMSD between the superposed MCS and docked MCS.
Complete rings	False	That the MCS must include complete rings.
Match rings	True	That for the MCS ring atoms must match ring atoms.
HTVS protocol	1	HTVS protocol for the first generation (see Figure 3.2).
DUck score	2	DUck score (W_{QB}) filter for the first iteration.

another strategy was tried, namely “MCS complete rings” was turned off, to allow for more permissive scaffold hopping, with the hopes to retrieve more hits that could pass all filters. And indeed, these settings retrieved virtual hits containing less extreme scaffold hops, however, still not as many virtual hits as X-0581. We hypothesized that the reason for this was that the screening library was not large or diverse enough to provide analogues of X-0404 that could pass all the filters incorporated into the protocol. As such, we tried a larger chemical space.

4.3.2 Chemical Space Exploration

The choice to use commercial chemical space for our iterative VS was also based on the fact that these libraries are continuously increasing in size. Commercial chemical space has experienced continuous growth over the last decades. The emergence of carefully designed virtual chemical libraries representing compounds that can be synthesized rapidly and at low cost has expanded the purchasable chemical space into the tens of billions. These number can increase by going beyond the commercial chemical space, but may involve synthetic routes that afford lower synthetic success rates. Besides that, larger numbers do not necessarily translates into better virtual screening outcomes, as they may create challenges in regards to soft and hard memory constraints.[175] Nonetheless, when using the On-Demand space consisting of 40 times more chemical matter than the In-Stock space, we already find many more virtual hits (depicted in Table 4.4).

Table 4.4: Attrition Rates In-Stock and On-Demand Space.

Iteration	1	2	3	4
In-Stock				
1. Similarity	112,192	1,216,791	2,326,173	3,018,340
2. Superposed	2,263	21,852	42,826	3,308
3. Docked	149	2,476	4,956	295
4. MMGBSA	35	169	193	113
5. DUck	25	69	66	24
On-Demand				
1. Similarity	550,528	6,599,934	29,099,709	100,181,069
2. Superposed	12,450	181,889	306,759	236,747
3. Docked	971	10,602	12,428	4,403
4. MMGBSA	139	1,349	1,396	586
5. DUck	114	125	143	236

Important questions that we aim to answer are whether we find novel chemical matter, if the use of larger collections increases the chance of becoming trapped in false positive space, as suggested[16] and understand the scalability of the process (i.e. what is the increase in computational cost and what library sizes could we potentially screen). From this exercise, screening the On-Demand set took twice as long, while the library is 40 fold larger. On a more practical aspect, we also want to compare costs and procurement rates.

From the In-Stock selection, 13 compounds were purchased based on visual inspection, diversity and availability, and from the On-Demand selection, 12 compounds were purchased. In principle, the In-Stock set should contain ligands that are directly purchasable, where as the

On-Demand set contains ligands that can be synthesized within 2 weeks, with 80% certainty. From this experience, the In-Stock collection indeed benefits from a better procurement rate (In-Stock: 26 out of 43 selected after visual inspection, On-Demand: 21 out of 63 selected after visual inspection). However, it did not cost that much more to purchase compounds from the On-Demand space (In-Stock: EUR 2,145, On-Demand: EUR 2,250), although delivery took approximately twice as long.

Figure 4.13a and Figure 4.13 show the rDock score distribution between the selected compounds from the In-Stock (Figure 4.13a) and from the On-Demand (Figure 4.13) space. These results show the compounds from the OnDemand space to score slightly better.

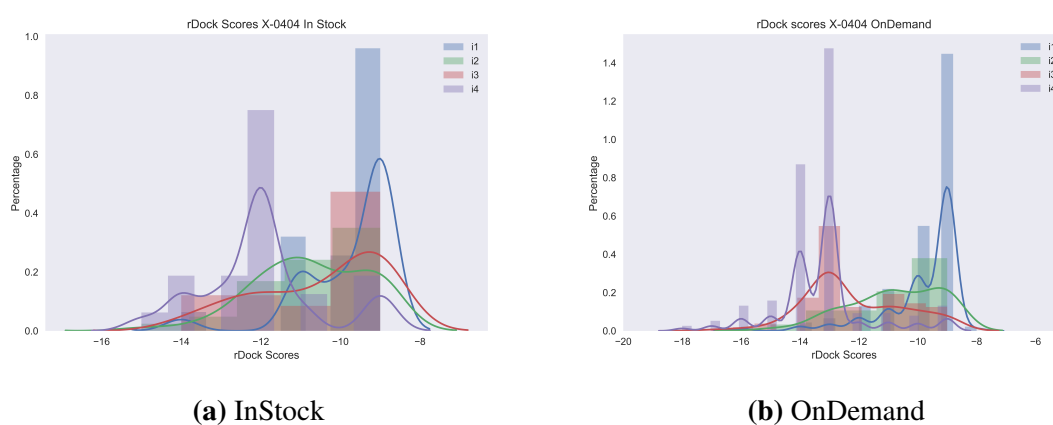


Figure 4.13: rDock Score Distribution Exploratory Strategies

4.3.2.1 Experimental Results

Table 4.5 shows the experimental results when less extreme scaffold hopping settings was applied (compared to Section 4.2.2) using the In-Stock space and the On-Demand space.

The same screening cascade as described in Section 4.2.3 resulted in a 54% hit rate using the In-Stock space, and a 33% hit rate using the On-Demand space. Both hit rates are lower than previously found for this fragment (63%), however, when considering the potency of the hits, the comparison changes. With extreme scaffold hopping for X-0404 optimization, 3 out of the 5 hits (60%) were less than 1 mM in potency and provided a clear signal by SPR. Here, with less extreme scaffold hopping the rates are 71% (5 out of 7) and 75% (3 out of 4) for In-Stock and On-Demand, respectively. Importantly, the last case afforded the most potent hit to date (**MR12**; $KD=10\mu M$; $HAC=25$; $LE=0.27$), a 20-fold potency improvement over the best hit obtained previously (**D8**; $KD=200\mu M$; $HAC=19$; $LE=0.26$). The raw SPR data is given in the Supplementary Information. Figure 4.14 and Figure 4.15 show the structures

Table 4.5: SPR Results Summary In-Stock and On-Demand Space

Binder	<500 μ M	500 μ M - 1mM	>1mM
X-0404 evolution In-Stock			
MR2			
MR7			
MR14			
MR1			
MR10			
MR11			
MR8			
X0404 evolution On-Demand			
MRE3			
MR12			
MRE6			
MRE4			

and binding modes of the experimentally validated hits. It should be noted that at the point the experiments had been done, the protein seemed to have lowered response levels, thus, the theoretical maximum response has also been reduced according to the baseline for affinity estimations.

4.3.2.2 Scaffold Analysis

Here we discuss how using a larger chemical space could be more advantageous, since using the On-Demand space led to slightly better scoring compounds, and our most potent hit so far for NUDT21. For this, an analysis of the scaffolds of all selected compounds was done. Figure 4.16A shows the clustering of manually defined scaffolds.

Scaffolds were defined as ring systems including attachment points with maximum one heavy atom dispersed. K-means clustering using RDK fingerprints (inspired by Daylight fingerprints) resulted in 15 clusters at the level of branching shown. It found three clusters with scaffolds only found in On-Demand hits, four clusters with scaffolds only found in In-Stock hits, and eight clusters with scaffolds that compounds from both spaces had in common were found.

Inspection of the interactions of the scaffolds unique to both search spaces (Figure 4.16B) found that some of the In-Stock scaffolds made less ideal interaction (clusters 14 and 6 in Figure 4.16C). Then, the observation that the On-Demand scaffolds were less diverse, and the fact that In-Stock cluster 9 contained the tri-one pyrimidine moiety also contained in our best hit (MR12 retrieved from the On-Demand space) led us to believe that the compounds in

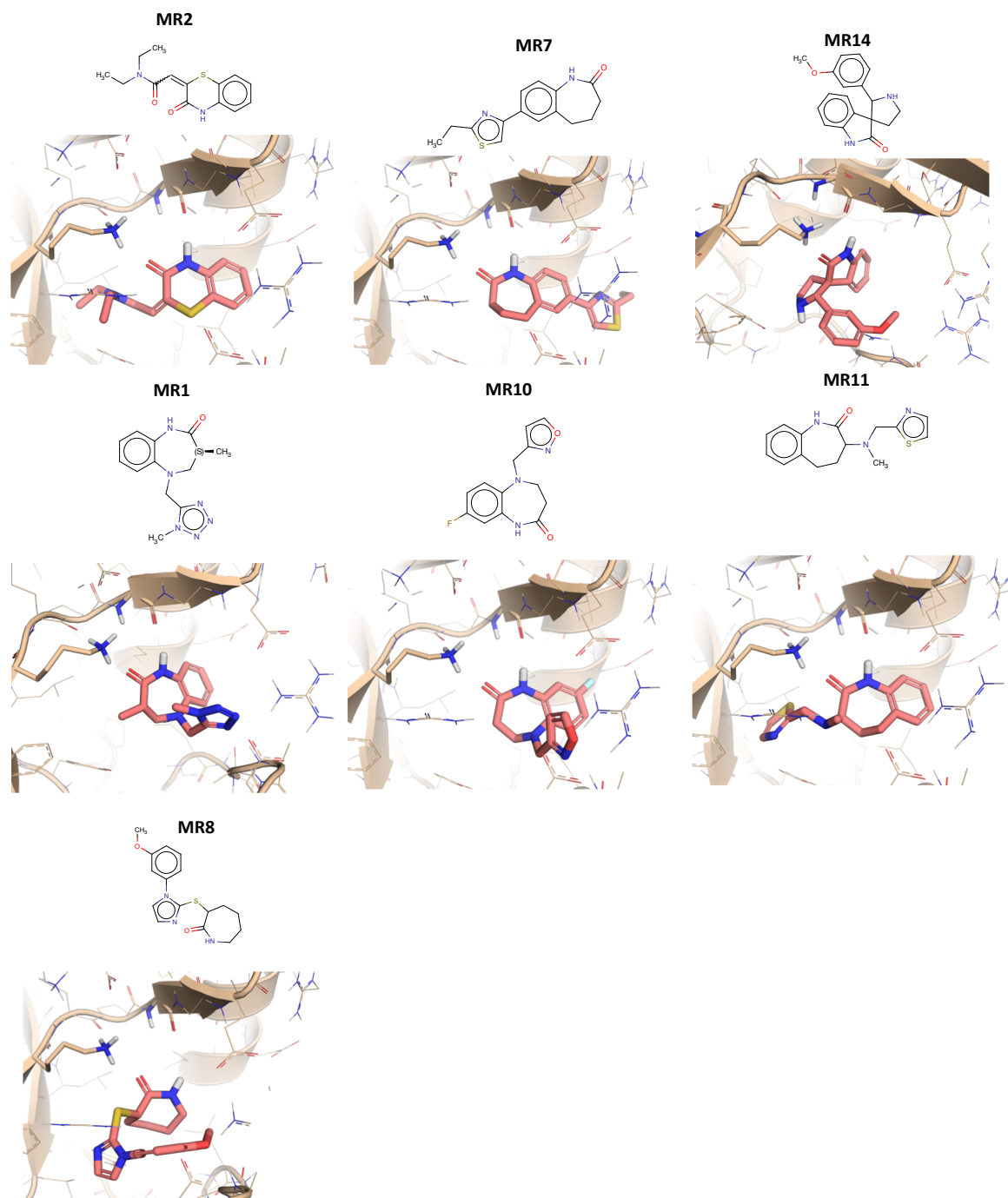


Figure 4.14: X-0404 In-Stock SPR Hits Binding Modes

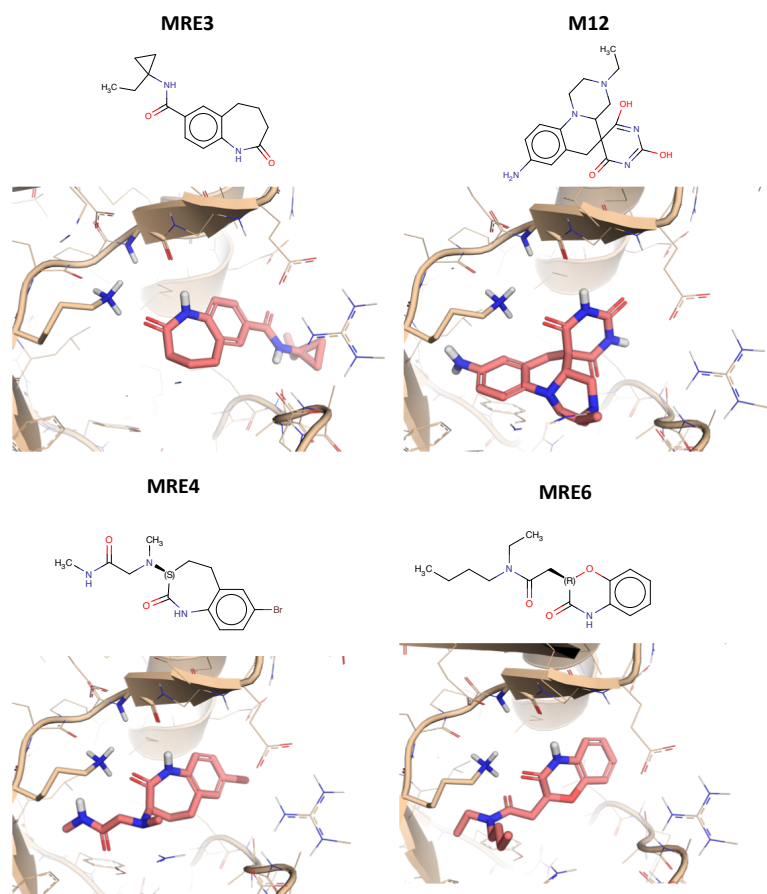


Figure 4.15: X-0404 On-Demand SPR Hits Binding Modes

the On-Demand space are not necessarily more diverse, but that in our case, the compounds making the best interactions are more elaborated in the On-Demand space, making them more probable to be selected by the platform. Indeed, the tri-one pyrimidine scaffold is found 10 times more often in the On-Demand space, than the In-Stock space.

4.4 Blinded Fragment Optimization

Fragment optimization without the availability of an X-ray is a very realistic scenario, but it can be quite challenging. Since we have developed a platform that is automated, and has been shown to lead to binders (described in Section 4.2 and Section 4.3), we decided to challenge to platform to blinded fragment optimization starting from all viable binding modes of fragment hits for which the binding mode have not been confirmed.

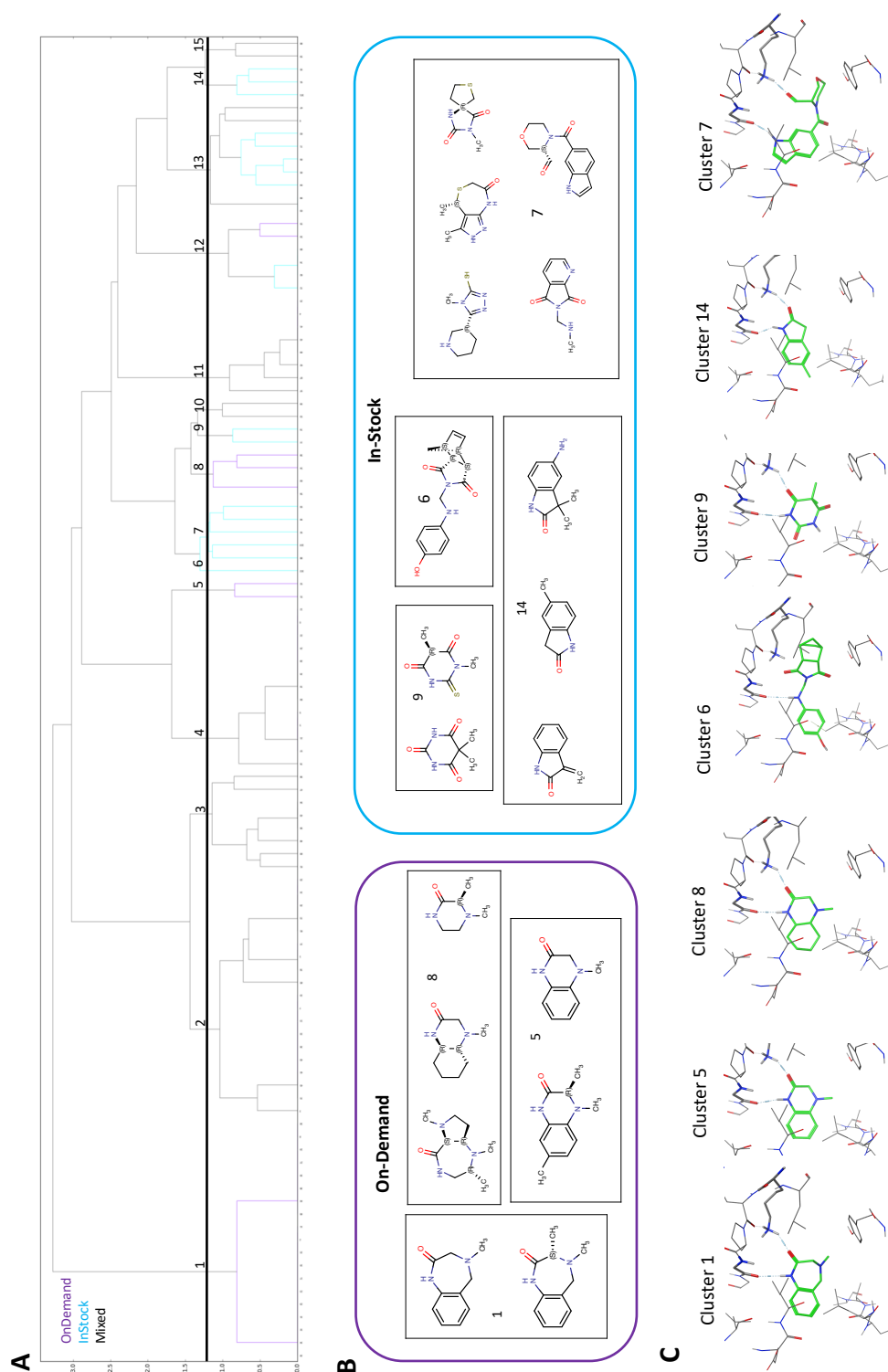


Figure 4.16: A) Clustering of manually defined scaffolds of all selected molecules. B) Scaffolds for clusters unique to molecules retrieved from On-Demand and In-Stock spaces. C) Binding modes of example scaffolds from unique clusters.

4.4.1 Target Selection

This study was divided into four case studies on three systems, namely, for HSP90, BRD4 and DYRK1A. The choice for these systems was based on fragments derived from literature, and received from our collaborators (Vernalis), which were found to be hits according to various screening assays, but could not be crystallized after at least one attempt. At the same time, these systems form robust test systems with an abundance of known ligands. Furthermore, their therapeutic relevance cannot be neglected. HSP90 (heat shock protein) is a molecular chaperone for hundreds of protein substrates or clients and not only plays a key role in protein folding under physiological and stress conditions, but is also involved in many cellular processes such as DNA repair, development, the immune response and neurodegenerative disease. This makes HSP90 an attractive drug target for which a number of HSP90 inhibitors are being evaluated in clinical trials.[176] BRD4 is an epigenetic regulator that has been reported as a potential therapeutic target for cancers.[177][178] The first BRD4 inhibitor JQ1 was developed in 2010[179], and since then, several BRD4 inhibitors have entered clinical trials. Additionally, a number of BRD4 degraders have been reported, which show more efficient anticancer activities, through protein specific degradation by small bifunctional molecules (PROTACs).[180] DYRK1A (dual-specificity tyrosine phosphorylation-regulated kinase 1A) that belongs to the group of CMGC kinases has attracted attention, as DYRK1A inhibition has been shown to weaken cognitive dysfunction in animal models for Down Syndrome and Alzheimer's disease. In addition, DYRK1A has been studied as a potential cancer therapeutic target. The current status is that inhibitors are being developed to determine the roles of this target.[181]

4.4.2 Pose Selection

Figure 4.17 shows an overview of the fragments and the selected binding modes. The BRD4 fragment was extracted from the literature[182], for which two viable poses were derived. For HSP90 two fragments were received from Vernalis, and for each, two viable poses were derived. For DYRK1A, two fragments were received from Vernalis, both of which contain the same core, however, bind in two slightly different binding modes. For pose selection and optimization, a structure from the PDB containing a similar ligand to the initial fragments was used (structures shown in the Supplementary Information). Similar ligands were retrieved with MACCS fingerprints, based on all downloaded PDB ligands for each target. For the structures, all waters were removed, except for those that are known to be conserved, namely for BRD4, four conserved water molecules[183], and for HSP90, three conserved water molecules[184]. The protocol used the H-bond pharmacophoric points shown in Table 4.6

for docking and DUck. For the DYRK1A fragments, a pharmacophoric feature that interacts with LEU241, N was implemented in addition to LEU241, O, since dual interactions with the hinge region in kinases are pronounced.[185] Clustering was performed as described in Section 3.5.3, after which the best scoring pose in each cluster was selected if the rDock score S^{total} was lower than -5 kJ/mol. Finally DUck[150] was performed on the selected fragment poses. A pose was considered, if the W_{QB} values was higher than 1.5 kcal/mol. The chunk residues for each PDB structure are given in the Supplementary Information.

The aim of these case studies is to answer two questions. Will the platform be able to distinguish if one starting pose is preferred? And will the platform evolve active ligands for which both binding modes are possible?

Table 4.6: Attrition Rates VS Comparison.

*Additional for docking: LEU241, N

Pose ID	rDock S^{total}	HTVS Protocol	W_{QB}	W_{QB} Cutoff	Constraint
BRD4 – Case Study 1					
B1	-20	2	2.4	2	ASN140,N
B2	-16		1.2		
HSP90 – Case Study 2					
H1-1	-17	1	6.4	4	ASP93,O
H1-2	-10		5.0		
HSP90 – Case Study 3					
H2-1	-20	2	7.4	4	ASP93,O
H2-2	-19		6.6		
DYRK1A– Case Study 4					
D1	-10	1	2.1	2	LEU241,O*
D2	-8		2.8		

4.4.3 Case Study 1 – BRD4

Before analysis of the results for this case study, it's important to mention that pose B2 is assumedly more probable than pose B1. According to several studies, the conserved water network contained in the binding pocket, is seen to accommodate hydrophobic groups, for example in 5M39[186] . Although, a flipped binding mode, in which the hydrophobic group is solvent exposed is also possible, for example in 5M3A (with a resolution of 1.65Å).

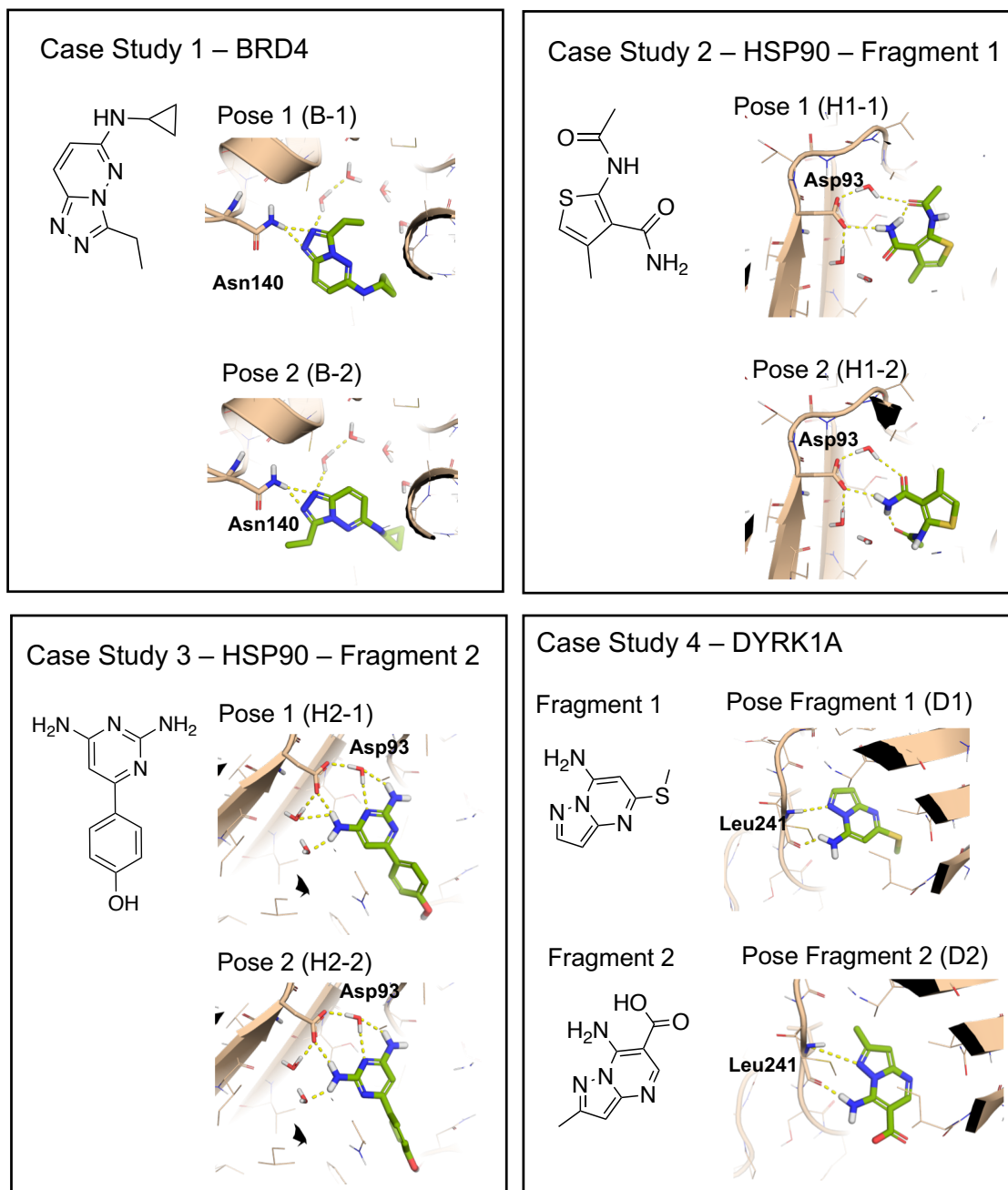


Figure 4.17: Overview of Case Studies and Predicted Binding Modes.

For this fragment, four iterations were performed starting from both binding modes. The initial fragment contained 15 heavy atoms, thus ligands up to 23 heavy atoms were explored by the fourth iteration. For B1, compounds for experimental validation were selected from the fourth iteration based on diversity, and availability. The nine compounds that were selected can be seen in Figure 4.18A. The compounds selected for testing, evolved from B2, stem from the third iteration, as those from the fourth iteration did not contain hydrophobic groups facing the conserved water network. The three selected compounds selected are shown in Figure 4.18B. The ligands evolved from both poses, maintain the same binding mode as the initial pose, but the scaffold is not preserved in all cases.

The first experiment for binding assessment was DSF (differential scanning fluorimetry), a thermal shift assay that measures changes in protein stability, which is changed upon binding.[187] The compounds were measured at 10 μ M, for which the results are shown in Figure 4.18. Based on these results, all compounds seem to bind. Binding assessment with an orthogonal biophysical assay based on time-resolved fluorescence resonance energy transfer (TR-FRET) is currently being executed. All 11 compounds were also subjected to co-crystallization experiments. Two structures could be obtained with proper fitting, one for compound **4**, and another for compound **11**. Compound **11** had a shifted binding mode compared with that observed in the crystal structure, while compound **4** had a flipped binding mode. Figure 4.19A shows the overlaid binding modes. Additionally, ligands crystallized in 5M39 and 5M3A, are also shown overlaid with the predicted binding modes (Figure 4.19B). Here it can be seen that the confirmed pose of compound **11** has shifted likely due to the change from methyl to ethyl. Our predicted binding mode for compound **4** on the other hand, does not resemble the binding mode observed in 5M3A, but has flipped, and resembles the confirmed binding mode of compound **11**, to accommodate the hydrophobic hotspot, observed in many other BRD4 ligands.

Tethered docking of compound **4** and **11** with the core constrained according to that observed, led to significantly worse rDock scores (compound **4**: -20 kJ/mol predicted vs. -12 kJ/mol observed, compound **11**: -21 kJ/mol predicted vs. -13 kJ/mol observed). This indicates that the protocol with the same settings would not identify these compounds even if we had started with the confirmed binding mode. Tethered docking with the initial fragment constrained according to the observed binding mode gives a score of -12 kJ/mol, and a W_{QB} of 0.9 kcal/mol. As such, this binding mode was not selected.

4.4.4 Case Study 2 – HSP90 – Fragment 1

For H2-1, four iterations were done, starting from 13 heavy atoms, thereby exploring ligands up to 21 heavy atoms. Three compounds for pose H1-1 and four compounds for H1-2 were

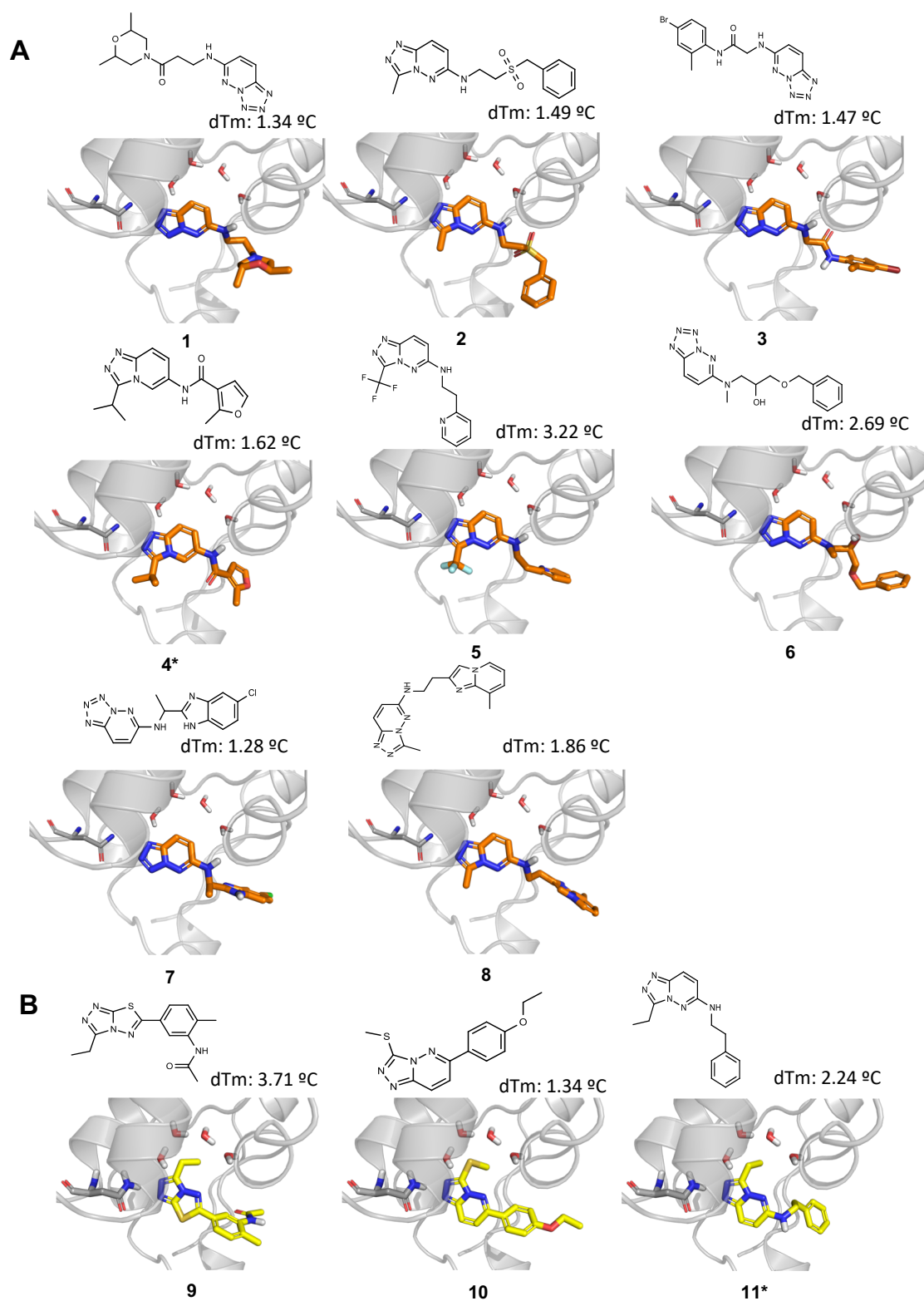


Figure 4.18: BRD4 Hits. A) Starting from pose B1. B) Starting from pose B2.

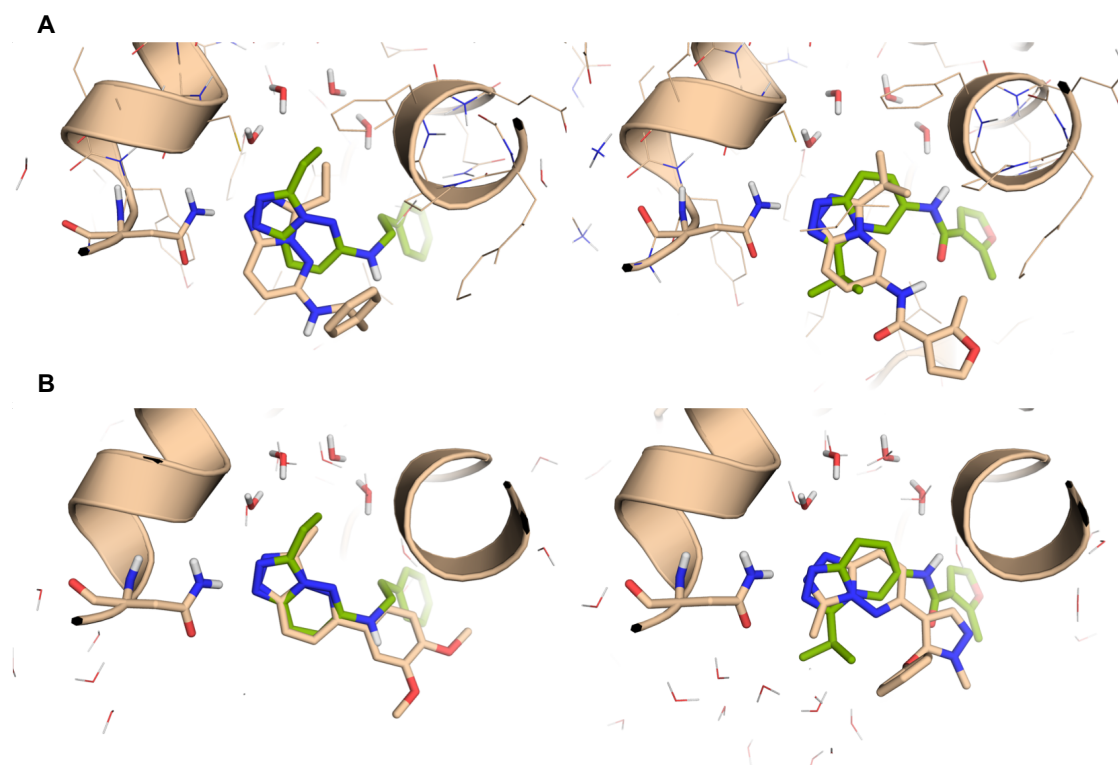
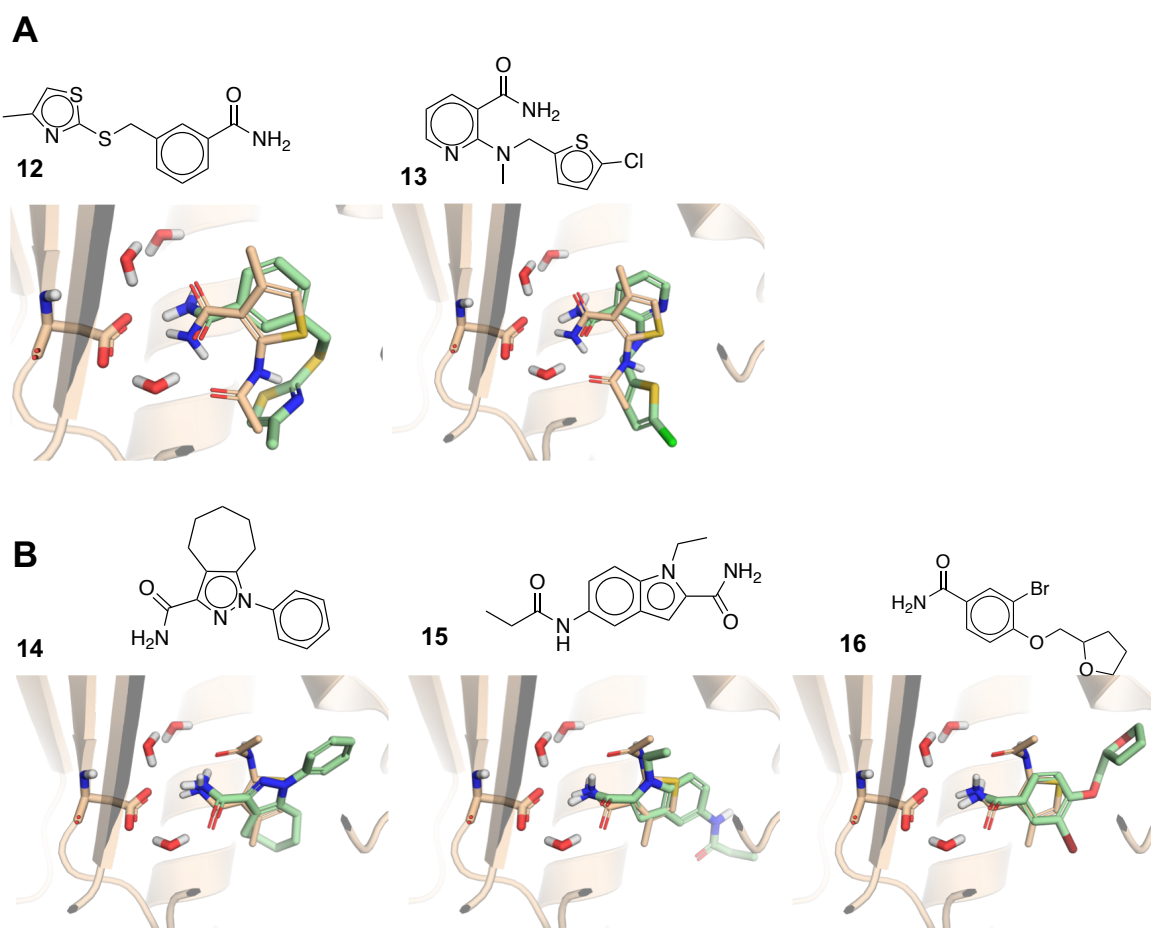


Figure 4.19: BRD4 X-ray (brown) vs. Predicted (green) binding modes. A) Overlay of obtained X-ray structures with Compounds 4 (right) and 11 (left). B) Overlay of 5M39 with compound 11 (left) and 5M3A with compound 4 (right).

selected for testing based on diversity and availability. Experimental validation was done in collaboration with Vernalis. The first step was quality control (LC-MS and NMR) and checking solubility. Two out of the seven compounds were insoluble, so no further data could be obtained. The next step was NMR (STD, LOGSY and CPMG). Table 4.7 shows the results. For the STD and LOGSY experiments, the initial signal should be positive and the delta should be negative if the compound is displaced by the competitor (VER-00082160). For the CPMG, the initial signal should be low, and the delta should be positive if displaced. Based only on CPMG results, all the compounds tested compounds seem to be binding. However, an empirical guide maintained by Vernalis to estimate the confidence level of true binding is that if a compound binds according to just one of these NMR techniques it's considered class 3, if it binds according to two of the techniques, it's considered class 2, and if it binds according to three of the techniques, it's considered class 1. With easily crystallisable systems class 1 gives ca. 80% success in crystallography, class 2 ca. 60% and class 3 ca. 40%.[188]

Table 4.7: NMR and SPR Results HSP90 Fragment 1.

Cpd.	STD	LOGSY		CPMG	dCPMG	Class	KD (μM)
		dSTD	dLOGSY				
H1-1							
12	0	0	0	2	2	3	183.0
13	0	0	0	1	2	3	399.0
H1-2							
14	0	0	0	2	1	3	3840
15	0	0	-1	3	2	3	378.0
16	0	0	0	0	3	2	841.0

**Figure 4.20:** HSP90 Fragment 1 Hits. A) Starting from pose H1-1. B) Starting from pose H1-2.

The experiments gave no class 1 compounds, and after soaking experiments, no crystal structures could be obtained. Nonetheless, SPR experiments for binding affinity could be obtained, for which the obtained KD values are also shown in Table 4.7 (raw SPR data is shown in the Supplementary Information). In Figure 4.20A, the experimentally validated compounds evolved from H1-1 are shown, and those of H1-2 are shown in Figure 4.20B. The IC50 value (fluorescent polarization/FP assay) for the initial fragment was 1.9mM. Even though we cannot directly compare the values because they were obtained with different biophysical techniques, it appears that at least compounds **12**, **13**, **15** and **16** bind with a better affinity, with compound **12** displaying significant potency and excellent ligand efficiency (KD=183 μ M; HAC=17; LE=0.30).

Notably, all the compounds from H1-1 have evolved in such a way that the amide is flipped according to how it's oriented in H1-2. How the binding mode of H1-1 flipped to that observed H1-2 is shown in the Supplementary Information.

4.4.5 Case Study 2 – HSP90 – Fragment 2

For H2-2, three iterations were performed, leading to the exploration of ligands up to 21 heavy atoms, starting from the 15 heavy atom fragment. Three compounds were selected for each binding mode evolution from iteration 3 based on diversity and availability. NMR and SPR results are shown in Table 4.8 and raw SPR data is shown in the Supplementary Information. Two out of the six compounds were insoluble so no further data is shown. The structures of the experimentally validated compounds evolved from H2-1 are shown in Figure 4.21A, and those evolved from H2-2 are shown in Figure 4.21B.

Table 4.8: NMR and SPR Results HSP90 Fragment 2. NB = not binding.

Cpd.	STD	dSTD	LOGSY	dLOGSY	CPMG	dCPMG	Class	KD (μ M)
H2-1								
17	0	0	0	0	3	0	NB	1080
18	0	0	0	0	0	3	3	3350
H2-2								
19	0	0	0	0	0	3	3	59.70
20	1	0	0	0	0	3	3	162.0

The starting fragment H2 had no measurable activity in FP. Compounds **17** and **18** show a moderate improvement (KD in the millimolar range), while compounds **19** and **20** show significant improvement and very good ligand efficiencies (0.29 and 0.26, respectively).

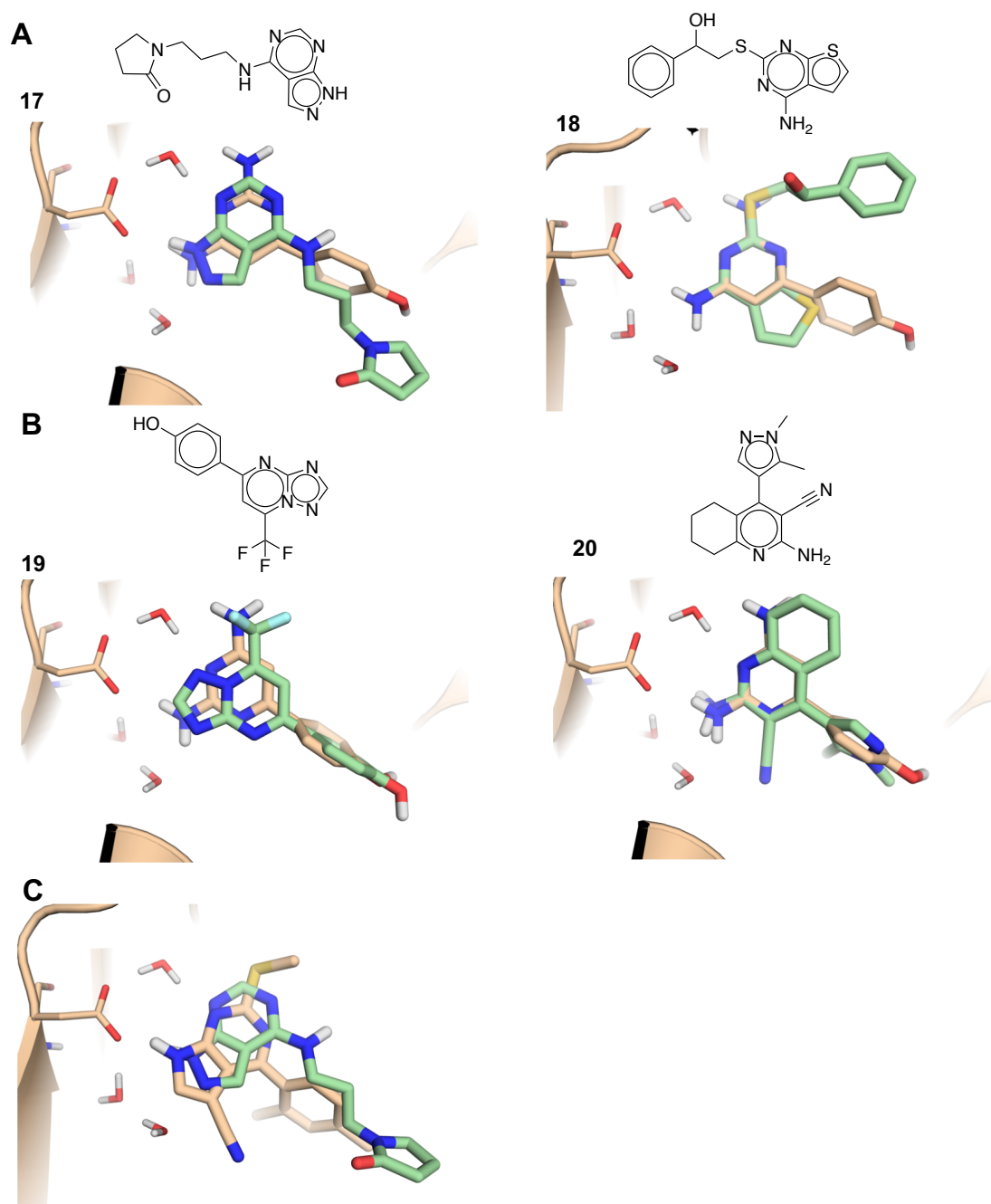


Figure 4.21: HSP90 Fragment 2 Hits. A) Compounds evolved from H2-1. B) Compounds evolved from H2-2. C) 4FCQ with compound 19.

Interestingly, compound **19** is the most potent, and was selected based on the fact that a C-H bond with Asp93 has never been reported for HSP90. Both compounds **17** and **18** stem from the second binding mode, indicating this pose to be the preferred starting point. However, compound **19**, contains a very similar scaffold as the co-crystallized ligand in 4FCQ[196], while none of the compounds evolved from H2-2, evolved to this scaffold. The overlay of the ligand in 4FCQ with compound **19** is shown in Figure 4.21C, for which a shift of the pyrimidine ring is observed. The evolution of this ring shift is shown in the Supplementary Information. Compound **19** could not be produced by H1-2, simply due to the fact that the flipped orientation, causes the MCS to be too small.

4.4.6 Case Study 4 – DYRK1A

For D1 (12 heavy atoms) five iterations were performed, while for D2 (14 heavy atoms), four iterations were performed, thereby sampling compounds up to 22 heavy atoms by the end of the evolution of both fragments. Four compounds were selected from pose D1 and five compounds were selected from pose D2, based on diversity and availability. One out of the 9 compounds was insoluble. NMR and SPR data (raw data in Supplementary Information) are shown in Table 4.9.

Table 4.9: NMR and SPR Results DYRK1A. NB = not binding. ND = No Data.

Cpd.	STD	dSTD	LOGSY	dLOGSY	CPMG	dCPMG	Class	KD (μ M)
D1								
21	1	-1	0	0	1	2	2	163.00
22	1	-1	0	0	0	2	2	329.00
23	3	-3	1	-1	1	3	1	73.50
D2								
24	0	0	0	0	1	3	3	ND
25	1	0	1	0	2	0	NB	444.00
26	0	0	0	0	0	1	3	31.60
27	2	-2	1	-1	1	3	1	436.00
28	1	-1	0	0	1	2	2	518.00

Although some class 1 and 2 compounds were retrieved, unfortunately, no crystal of the protein could be obtained from soaking experiments. Figure 4.22A shows the compounds evolved from D1, and Figure 4.22B shows the compounds evolved from D2 that were experimentally validated. The initial fragment D1 had an IC₅₀ of 115 μ M (inhibition of peptide phosphorylation), while D2 showed no measurable affinity. Only one compound

evolved from D1 had a better affinity (**23**; $KD=73\mu M$; $HAC=19$; $LE=0.30$) on which the platform performed a scaffold hop, while compound **26** had a significantly improved affinity over D2 ($KD=32\mu M$; $HAC=20$; $LE=0.31$), and maintained the same binding mode.

4.4.7 Summary

So, coming back to our two questions: will the platform be able to distinguish if one starting pose is preferred and will the platform evolve active ligands for which both binding modes are possible? We conclude that the platform is capable of deriving active compounds from fragments with an unknown binding mode, with success rates and jumps in potency similar to those obtained in other systems when the binding mode is known (Section 4.2 and Section 4.3). However, the platform is not suitable to identify the correct binding mode because it can find active molecules either by changing the scaffold or by changing the binding mode. Finally, we cannot rule out that in some cases active compounds derive from incorrect binding mode predictions.

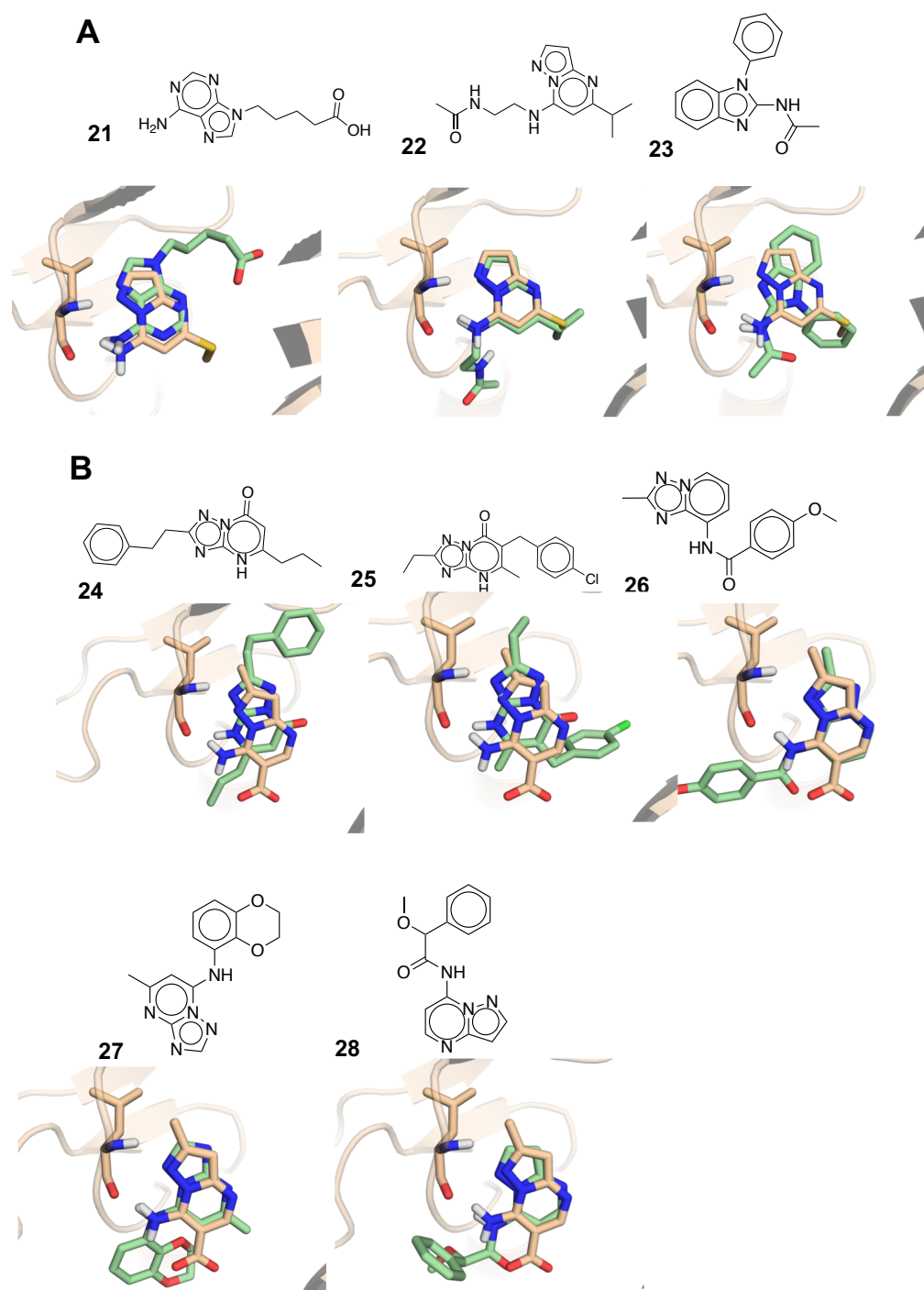


Figure 4.22: DYRK1A Hits. A) Compounds evolved from D1. B) Compounds evolved from D2.

Chapter 5

Discussion

Fragment-based drug discovery (FBDD) emerged as a disruptive technology and became established during the last two decades. Its rationality and low entry costs make it appealing, and the numerous examples of approved drugs discovered through FBDD validate the approach. However, FBDD still faces numerous challenges. Perhaps the most important one is the transformation of the initial fragment hits into viable leads. Fragment-to-lead (F2L) optimization is resource-intensive and is therefore limited in the possibilities that can be actively pursued. *In silico* strategies play an important role in F2L, as they can perform a deeper exploration of chemical space, prioritize molecules with high probabilities of being active and generate non-obvious ideas. This thesis describes the development of an automated pipeline for fragment-to-lead optimization. Its iterative nature allows seamless integration into fragment-based drug discovery pipelines (as shown in Figure 1.5) and its scalable feature can exploit continuously growing commercially available libraries.

Section 4.1 gave an overview of the developed platform. It uses three complementary structure-based methods that retrieve the most likely to bind analogues. Subsequent iterations harness information on chemotypes and features to determine the chemical space on which the next round of structure-based virtual screening takes place. Furthermore, it is scalable to screen ultra-large virtual libraries, which may lead to better or more hits. It searches through these very large chemical spaces, enabling diversity (e.g. scaffold hopping) but focusing on the most productive areas of chemical space by means of knowledge based (similarity, pharmacophores) and structure-based (protein-ligand interactions) constraints. All combined results indicate the platform's capabilities for being applied to future projects. Future prospects include making the platform available.

The ambitious aim to find a probe for every protein has been recently proposed through the "Target 2035" initiative.[2] Although daunting, any improvement over the mere 4% of the human proteome that has roughly been estimated to have been paired with a chemical probe, would be considered as progress. To this end, fragment-based approaches have been suggested as a promising strategy in the expansion of the number of proteins that can be deemed as druggable. Fragment screening campaigns very often will lead to higher hit rates, even for challenging targets, as fragment-sized molecules have a much better chance of binding than drug-sized molecules, albeit with lower affinity.[8]

Crystallographic fragment screening is sensitive enough to detect such weak binders, while also providing a clear mode of binding, so that fragment optimization, which is essential to increase potency, becomes less challenging. The automated crystal density analysis called PanDDA created by SGC-Diamond and XChem has proven extremely effective, where significant increase in fragment screening hit rates could be obtained when compared to manual inspection of the density.[138] An automated set-up that includes fragment X-ray

screening and iterative optimization could comprise an effective platform for expanding the druggable proteome through accelerating fragment-derived drugs.[189] A prime example introduced by the SGC-Diamond XChem partnership allows rapid follow-up synthesis using the concept of poised libraries (described in Section 1.3.1.2).[41] The resulting compound hits could then be synthesized via parallel, solution-phase synthesis. This in combination with crystallographic fragment screening, allowed Cox et al. to identify the first reported inhibitors of the atypical bromodomain, PHIP(2).[40]

Section 4.2 described the optimization of fragments that were found to bind by x-ray with almost unmeasurable potency to NUDT21. The lack of understanding of the mechanism and regulation of NUDT21 emphasizes the need for a chemical probe. We used two distinct approaches, compared the use of the poised approach and the iterative approach. Due to the employment of the “poised synthesis” strategy, scaffold hopping or scaffold optimization was not made possible. On the other hand, the iterative VS approach imposes just two main requirements (besides a guided improvement of scores) namely, that I) similar analogues contain known or predicted ligand binding features and II) that these features belong to a common core with the initial fragment. In this way, scaffold hopping is incorporated. The scaffold hopping approach led to improved versions of the starting fragments. This is particularly useful as fragment progression requires a verified high quality starting point with measurable potency, which may require tweaking of (the core in) the fragment. The poised approach did not lead to any hits. The search space of both strategies differed significantly due to the efficiency of both methods, ultimately, resulting in one giving results, and the other not. The hits retrieved from the automated and iterative protocol were very diverse, and if they can be crystallized, would make for better starting points for further optimization to probe NUDT21’s function.

Section 4.3 described the adaptability of the platform to perform no scaffold hopping to extreme scaffold hopping. Its scalability was also demonstrated. The platform can switch in between extreme, less extreme, an no scaffold hopping modes. Both the platforms ability to exploit large chemical spaces, and its ability to switch in between these different modes, allow for the opportunity to identify chemical entities related to an initial fragment with different levels of novelty and diversity.

Experimental results showed that less extreme scaffold hopping in the On-Demand set, led to better quality hits. The statistics would be different if more compounds would be tested, but at least we can say that even when adapting the settings to retrieve results that are more or less diverse or novel, we can still obtain binders. Analysis of the scaffolds did show that the On-Demand space led to better interacting scaffolds, due to their prevalence in the library. Furthermore, studies have shown that screening larger libraries can lead to

better hits.[14] In any case, the financial and computational investment was almost equal when using the On-Demand versus the In-Stock space. What we gained was ultimately more hits, one of which is the most potent reported to date for NUDT21 and is approaching the potency needed for a good chemical probe.

Practitioners agree that structural information on fragment-target complexes is essential for the success of FBLD.[190][191][192] More often than not, fragments that have not been crystallized will be disregarded[193][194][195], even though they may very well be good starting points for an FBDD campaign.[196] Ideally fragment hits can be visualized by X-ray, however, there are enough examples of fragment optimization without an X-ray.[90,108][197] In fact, the initial fragment of one of the first fragment derived drugs on the market was not crystallized.[186] Instead, the evolution of the fragment was guided by NMR. Furthermore, with well-known proteins, a confirmed binding mode might not even be a pre-requisite, as important interactions are known, which can be used to guide the optimization. Various other factors may influence the decision to optimize a fragment hit without the availability of the crystallized fragment, including binding affinity, ligand efficiency, and physical parameters such as solubility. Lastly, X-ray crystallography is used less often as the primary screening tool, mainly because its throughput is thought to be low, even though it has been shown that using X-ray as a secondary or tertiary screen will lead to less expected hits by X-ray[178]. These reasons make fragment optimization without the availability of an X-ray a very realistic scenario, so we decided to challenge to platform to blinded fragment optimization as described in Section 4.4 .

Four case studies were derived for fragment hits without a binding mode to address if the platform can distinguish if one starting pose is preferred of the system and if the platform will evolve active ligands for which both binding modes are possible. The initial fragment of BRD4 had a clear preference for one binding mode. Nonetheless, the pose selection protocol described above selected virtual hits from both the preferred and the less ideal binding mode. Moreover, both binding modes have been previously crystallized, confirming their plausibility. Extraordinarily, all the virtual screening hits actually tested were confirmed as active by DSF (TR-FRET experiments on-going), but for the only two compounds that produced crystals, their binding mode turned out to be different from the predicted ones. Not only that, but the crystallographic structures score poorly in our scoring methods and would not have been selected by our protocol. Unless the compounds actually bind in more than one way (which cannot be completely ruled out based on the binding mode of similar compounds in the PDB; Figure 4.19B), it would appear that the right compounds have been selected, but for the wrong reason. As we are starting from active compounds, the virtual screening protocol explores a privileged area of chemical space and it may be possible to find

hits even when the structure-based predictions fail. Indeed, the protocols are far from perfect, and inadequate accounting of solvation effects may be behind the results obtained in Case Study 1. But it is highly improbable that all hits stem from incorrect binding modes. For Case Study 2 the second pose is preferred due to an internal strain in pose 1. The platform actually corrected this by evolving the compounds to the flipped orientation seen in pose 2. For Case Study 3, there was no clear reason why one would be preferred over the next. The hypothesis was that the platform would lead to binders in both cases, and so it did. For Case Study 4 both fragments contain the same core, however, due to its substituents, Fragment D2 has less ideal interactions. Nonetheless, the pose selection protocol selected both the preferred and the less ideal binding mode. The hypothesis was that the platform will lead to actives for the preferred pose, and inactives for the less ideal pose. However, the compounds evolved from both poses showed good binding, indicating that the deviation in pose D2 is tolerable. The results revealed two different scenarios, namely, I) an unchanged binding mode and/or scaffold, and, II) if the scaffold cannot be evolved, a scaffold hop takes place, or a change in binding mode occurs.

The platform is capable of deriving active compounds from fragments with an unknown binding mode, with success rates and jumps in potency similar to those obtained in other systems when the binding mode is known (Section 4.2 and Section 4.3). However, the platform is not suitable to identify the correct binding mode. The reason being that it can detect if an alteration in the binding mode is necessary, by either changing the binding mode, or performing scaffold hopping. This addresses another challenge with fragment optimization in that it assumes that the binding mode is preserved.[198] Malhotra and colleagues showed that 14% of bound fragments alter their binding mode. In some cases, the added substituent may have resulted in a clash with the protein, while in other cases an alternate binding mode was the result of stronger interactions, only possibly in a different pose.[199] Lastly, we cannot rule out that in some cases active compounds derive from incorrect binding mode predictions due to the intrinsic limitations of the methods used. However, we also cannot completely rule out that the compounds may bind in more than one way.

Chapter 6

Conclusions

Global Conclusions

This thesis described the development and validation of an automated and iterative fragment-to-lead optimization, including its application to diverse scenarios.

Specific conclusions

1. Development

- The developed pipeline has been shown to be extremely efficient by evolving fragments to likely to bind compounds in under 24 hours.
- All the steps have been automated, and are run in parallel. Furthermore, carefully thought-out filters reduce the search space to promising areas. For these reasons, it is scalable.
- Because it is scalable, multiple automated evolutions can be run simultaneously.
- Because it is scalable, very large screening libraries can be used.

2. Validation: application to diverse scenarios.

- The developed pipeline has been shown to be superior to VS strategies that are not iterative and do not consider scaffold hopping.
- The platform has been shown to be adaptable to various discovery settings, from extreme to no scaffold hopping.
- The developed pipeline evolved active binders even when the starting fragment hit did not have a confirmed binding mode.
- The developed pipeline has been applied by external parties although future plans include making it available to all.

Bibliography

- [1] J. P. Hughes, S. S. Rees, S. B. Kalindjian, and K. L. Philpott. Principles of early drug discovery. *British Journal of Pharmacology*, 162(6):1239–1249, 2011.
- [2] Asher Mullard. A probe for every protein. *Nature Reviews Drug Discovery*, 18(10):733–736, 2019.
- [3] BioMedTracker 2016. Clinical development success rates 2006-2015. <https://www.bio.org/sites/default/files/Clinical%20Development%20Success%20Rates%202006-2015%20-%20BIO,%20Biomedtracker,%20Amplion%202016.pdf>, 2016.
- [4] P. G. Polishchuk, T. I. Madzhidov, and A. Varnek. Estimation of the size of drug-like chemical space based on GDB-17 data. *Journal of Computer-Aided Molecular Design*, 27(8):675–679, 2013.
- [5] Lars Ruddigkeit, Ruud Van Deursen, Lorenz C. Blum, and Jean Louis Reymond. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *Journal of Chemical Information and Modeling*, 52(11):2864–2875, 2012.
- [6] Duncan E. Scott, Anthony G. Coyne, Sean A. Hudson, and Chris Abell. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry*, 51(25):4990–5003, 2012.
- [7] Benjamin D. Cons Amanda J. Price, Steven Howard. Fragment-based drug discovery and its application to challenging drug targets. *Essays in Biochemistry*, 61(5):475–484, 2017.
- [8] Michael M. Hann, Andrew R. Leach, and Gavin Harper. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *Journal of Chemical Information and Computer Sciences*, 41(3):856–864, 2001.
- [9] Andrew L. Hopkins, Colin R. Groom, and Alexander Alex. Ligand efficiency: A useful metric for lead selection. *Drug Discovery Today*, 9(10):430–431, 2004.
- [10] Gisbert Schneider and Uli Fechner. Computer-based de novo design of drug-like molecules. *Nature Reviews Drug Discovery*, 4(8):649–663, 2005.
- [11] Peter S. Kutchukian and Eugene I. Shakhnovich. De novo design: Balancing novelty and confined chemical space. *Expert Opinion on Drug Discovery*, 5(8):789–812, 2010.
- [12] John J. Irwin and Brian K. Shoichet. ZINC - A free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*, 45(1):177–182, 2005.
- [13] Teague Sterling and John J. Irwin. ZINC 15 - Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, 55(11):2324–2337, 2015.
- [14] Jiankun Lyu, Sheng Wang, Trent E Balius, Isha Singh, Anat Levit, Yurii S Moroz, Matthew J O Meara, Tao Che, Enkhjargal Algaa, Kateryna Tolmacheva, Andrey A. Tolmachev, Brian K. Shoichet, Bryan L. Roth, and John J. Irwin. Ultra-large library docking for discovering new chemotypes. *Nature*, 566:224–229, 2019.

- [15] Christoph Gorgulla, Andras Boeszoermyeni, Zi-Fu Wang, Patrick D. Fischer, Paul Coote, Krishna M. Padmanabha Das, Yehor S. Malets, Dmytro S. Radchenko, Yurii S. Moroz, David A. Scott, Konstantin Fackeldey, Moritz Hoffmann, Iryna Iavniuk, Gerhard Wagner, and Haribabu Arthanari. An open-source drug discovery platform enables ultra-large virtual screens. *Nature*, 580(March 2019), 2020.
- [16] Xavier Barril. Computer-aided drug design: time to play with novel chemical matter. *Expert Opinion on Drug Discovery*, 12(10):977–980, 2017.
- [17] W. Patrick Walters. Virtual screening libraries. *Journal of Medicinal Chemistry*, 62:1116–1124, 2018.
- [18] Gisbert Schneider, Man Ling Lee, Martin Stahl, and Petra Schneider. De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. *Journal of Computer-Aided Molecular Design*, 14(5):487–494, 2000.
- [19] Dominique Douguet, Helene Munier-Lehmann, Gilles Labesse, and Sylvie Pochet. LEA3D: A computer-aided ligand design for structure-based drug design. *Journal of Medicinal Chemistry*, 48(7):2457–2468, 2005.
- [20] Xiao Qing Lewell, Duncan B. Judd, Stephen P. Watson, and Michael M. Hann. RECAP - Retrosynthetic Combinatorial Analysis Procedure: A powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *Journal of Chemical Information and Computer Sciences*, 38(3):511–522, 1998.
- [21] H Bohm. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *Journal of Computer-Aided Molecular Design volume*, 6:61–78, 1992.
- [22] Valerie J. Gillet, William Newell, Paulina Mata, Glenn Myatt, Sandor Sike, Zsolt Zsoldos, and A. Peter Johnson. SPROUT: Recent Developments in the de Novo Design of Molecules. *Journal of Chemical Information and Computer Sciences*, 34(1):207–217, 1994.
- [23] Renxiao Wang, Ying Gao, and Luhua Lai. LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design. *Journal of Molecular Modeling*, 6(7-8):498–516, 2000.
- [24] Yoshihiko Nishibata and Akiko Itai. Confirmation of usefulness of a structure construction program based on three-dimensional receptor structure for rational lead generation. *Journal of Medicinal Chemistry*, 36(20):2921–2928, oct 1993.
- [25] Peter S. Kutchukian, David Lou, and Eugene I. Shakhnovich. FOG: Fragment optimized growth algorithm for the de novo generation of molecule: Occupying druglike chemical Space. *Journal of Chemical Information and Modeling*, 49(7):1630–1642, 2009.
- [26] Albert C. Pierce, Govinda Rao, and Guy W. Bemis. BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, P38, and HIV protease. *Journal of Medicinal Chemistry*, 47(11):2768–2775, 2004.
- [27] David A. Pearlman Mark A. Murcko. CONCEPTS: New dynamic algorithm for de novo drug suggestion. *Journal of Computational Chemistry*, 14:1184–1193, 1993.
- [28] Haitao Ji, Wannian Zhang, Min Zhang, Makiko Kudo, Yuri Aoyama, Yuzo Yoshida, Chunquan Sheng, Yunlong Song, Song Yang, Youjun Zhou, Jianguo Lü, and Jü Zhu. Structure-based de novo design, synthesis, and biological evaluation of non-azole inhibitors specific for lanosterol 14 α -demethylase of fungi. *Journal of Medicinal Chemistry*, 46(4):474–485, 2003.
- [29] Chun Feng Chang, Wen Hsing Lin, Yi Yu Ke, Yih Shyan Lin, Wen Chieh Wang, Chun Hwa Chen, Po Chu Kuo, John T.A. Hsu, Biing Jiun Uang, and Hsing Pang Hsieh. Discovery of novel inhibitors of Aurora kinases with indazole scaffold: In silico fragment-based and knowledge-based drug design. *European Journal of Medicinal Chemistry*, 124:186–199, 2016.

- [30] Steven L. Warner, Sridevi Bashyam, Hariprasad Vankayalapati, David J. Bearss, Haiyong Han, Daniel D. Von Hoff, and Laurence H. Hurley. Identification of a lead small-molecule inhibitor of the Aurora kinases using a structure-assisted, fragment-based approach. *Molecular Cancer Therapeutics*, 5(7):1764–1773, 2006.
- [31] Hans Joachim Böhm, David W. Banner, and Lutz Weber. Combinatorial docking and combinatorial chemistry: Design of potent non-peptide thrombin inhibitors. *Journal of Computer-Aided Molecular Design*, 13(1):51–56, 1999.
- [32] Hans Joachim Boehm, Markus Boehringer, Daniel Bur, Hans Gmuender, Walter Huber, Werner Klaus, Dirk Kostrewa, Holger Kuehne, Thomas Luebbbers, Nathalie Meunier-Keller, and Francis Mueller. Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *Journal of Medicinal Chemistry*, 43(14):2664–2674, 2000.
- [33] Honma et al. Structure-based generation of a new class of potent Cdk4 inhibitors: New de novo design strategy and library design. *Journal of Medicinal Chemistry*, 44(26):4615–4627, 2001.
- [34] Kandil et al. Discovery of a novel HCV helicase inhibitor by a de novo drug design approach. *Bioorganic and Medicinal Chemistry Letters*, 19(11):2935–2937, 2009.
- [35] Shuaishuai Ni, Yaxia Yuan, Jin Huang, Xiaona Mao, Maosheng Lv, Jin Zhu, Xu Shen, Jianfeng Pei, Luhua Lai, Hualiang Jiang, and Jian Li. Discovering potent small molecule inhibitors of cyclophilin A using de novo drug design approach. *Journal of Medicinal Chemistry*, 52(17):5295–5298, 2009.
- [36] René Barone and Michel Chanon. A new and simple approach to chemical complexity. Application to the synthesis of natural products. *Journal of Chemical Information and Computer Sciences*, 41(2):269–272, 2001.
- [37] T. Honma, K. Hayashi, T. Aoyama, N. Hashimoto, T. Machida, K. Fukasawa, T. Iwama, C. Ikeura, M. Ikuta, I. Suzuki-Takahashi, Y. Iwasawa, T. Hayama, S. Nishimura, and H. Morishima. Structure-based generation of a new class of potent Cdk4 inhibitors: New de novo design strategy and library design. *Journal of Medicinal Chemistry*, 44(26):4615–4627, 2001.
- [38] Ewgenij Proschak, Kerstin Sander, Heiko Zettl, Yusuf Tanrikulu, Oliver Rau, Petra Schneider, Manfred Schubert-Zsilavec, Holger Stark, and Gisbert Schneider. From molecular shape to potent bioactive agents II: Fragment-based de novo design. *ChemMedChem*, 4(1):45–48, 2009.
- [39] Markus Hartenfeller, Martin Eberle, Peter Meier, Cristina Nieto-Oberhuber, Karl Heinz Altmann, Gisbert Schneider, Edgar Jacoby, and Steffen Renner. A collection of robust organic synthesis reactions for in silico molecule design. *Journal of Chemical Information and Modeling*, 51(12):3093–3098, 2011.
- [40] Markus Hartenfeller, Martin Eberle, Peter Meier, Cristina Nieto-Oberhuber, Karl Heinz Altmann, Gisbert Schneider, Edgar Jacoby, and Steffen Renner. A collection of robust organic synthesis reactions for in silico molecule design. *Journal of Chemical Information and Modeling*, 51(12):3093–3098, 2011.
- [41] Oakley B. Cox, Tobias Krojer, Patrick Collins, Octovia Monteiro, Romain Talon, Anthony Bradley, Oleg Fedorov, Jahangir Amin, Brian D. Marsden, John Spencer, Frank Von Delft, and Paul E. Brennan. A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chemical Science*, 7(3):2322–2330, 2016.
- [42] Stephen D Roughley and Allan M Jordan. The Medicinal Chemist’s Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates †. *J Med Chem*, 54:3451–3479, 2011.
- [43] Mark Whittaker. Picking up the pieces with FBDD or FADD: invest early for future success. *Drug Discovery Today*, 14(13-14):623–624, 2009.
- [44] Zhe Wang, Huiyong Sun, Xiaojun Yao, Dan Li, Lei Xu, Youyong Li, Sheng Tian, and Tingjun Hou. Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.*, 18(18):12964–12975, 2016.

- [45] Hwangseo Park, Yongje Shin, Jinhee Kim, and Sungwoo Hong. Application of Fragment-Based de Novo Design to the Discovery of Selective Picomolar Inhibitors of Glycogen Synthase Kinase-3 Beta. *Journal of Medicinal Chemistry*, 59(19):9018–9034, 2016.
- [46] Qiuping Xiang, Chao Wang, Yan Zhang, Xiaoqian Xue, Ming Song, Cheng Zhang, Chenchang Li, Chun Wu, Kuai Li, Xiaoyan Hui, Yulai Zhou, Jeff B. Smaill, Adam V. Patterson, Donghai Wu, Ke Ding, and Yong Xu. Discovery and optimization of 1-(1H-indol-1-yl)ethanone derivatives as CBP/EP300 bromodomain inhibitors for the treatment of castration-resistant prostate cancer. *European Journal of Medicinal Chemistry*, 147:238–252, 2018.
- [47] Martina et al. Structure-Based Design of an in Vivo Active Selective BRD9 Inhibitor. *Journal of Medicinal Chemistry*, 59(10):4462–4475, 2016.
- [48] Hale et al. From fragments to leads: Novel bacterial NAD⁺-dependent DNA ligase inhibitors. *Tetrahedron Letters*, 56(23):3108–3112, 2015.
- [49] Ahmed et al. Fragment-based discovery of a new family of non-peptidic small-molecule cyclophilin inhibitors with potent antiviral activities. *Nature Communications*, 7, 2016.
- [50] Bennet et al. Design, synthesis and biological evaluation of novel 4-phenylisoquinolinone BET bromodomain inhibitors. *Bioorganic and Medicinal Chemistry Letters*, 28(10):1811–1816, 2018.
- [51] Sarah M. Bronner, Jeremy Murray, F. Anthony Romero, Kwong Wah Lai, Vickie Tsui, Patrick Cyr, Maureen H. Beresini, Gladys De Leon Boenig, Zhongguo Chen, Edna F. Choo, Kevin R. Clark, Terry D. Crawford, Hariharan Jayaram, Susan Kaufman, Ruina Li, Yingjie Li, Jiangpeng Liao, Xiaorong Liang, Wenfeng Liu, Justin Ly, Jonathan Maher, John Wai, Fei Wang, Aijun Zheng, Xiaoyu Zhu, and Steven Magnuson. A Unique Approach to Design Potent and Selective Cyclic Adenosine Monophosphate Response Element Binding Protein, Binding Protein (CBP) Inhibitors. *Journal of Medicinal Chemistry*, 60(24):10151–10171, 2017.
- [52] Rongjun He, Zhi-hong Yu, Ruo-yu Zhang, Li Wu, Andrea M Gunawan, and Zhong-yin Zhang. Cefsulodin Inspired Potent and Selective Inhibitors of mPTPB, a Virulent Phosphatase from *Mycobacterium tuberculosis*. -, 6:1231–1235, 2015.
- [53] Maciej Majewski and Xavier Barril. Structural stability predicts the binding mode of protein-ligand complexes. *Journal of Chemical Information and Modeling*, 6:1644–1651, 2020.
- [54] Moira M. Rachman, Xavier Barril, and Roderick E. Hubbard. Predicting how drug molecules bind to their protein targets. *Current Opinion in Pharmacology*, 42:34–39, 2018.
- [55] Edwige Lorthiois, Karen Anderson, Anna Vulpetti, Olivier Rogel, Frederic Cumin, Nils Ostermann, Stefan Steinbacher, Aengus Mac Sweeney, Omar Delgado, Sha Mei Liao, Stefan Randl, Simon Rüdissler, Solene Dussauge, Kamal Fettis, Laurence Kieffer, Andrea De Erkenez, Louis Yang, Constanze Hartwig, Upendra A. Argikar, Laura R. La Bonte, Ronald Newton, Viral Kansara, Stefanie Flohr, Ulrich Hommel, Bruce Jaffee, and Jürgen Maibaum. Discovery of Highly Potent and Selective Small-Molecule Reversible Factor D Inhibitors Demonstrating Alternative Complement Pathway Inhibition in Vivo. *Journal of Medicinal Chemistry*, 60(13):5717–5735, 2017.
- [56] Anna Vulpetti, Stefan Randl, Simon Rüdissler, Nils Ostermann, Paul Erbel, Aengus Mac Sweeney, Thomas Zoller, Bahaa Salem, Bernd Gerhartz, Frederic Cumin, Ulrich Hommel, Claudio Dalvit, Edwige Lorthiois, and Jürgen Maibaum. Structure-Based Library Design and Fragment Screening for the Identification of Reversible Complement Factor D Protease Inhibitors. *Journal of Medicinal Chemistry*, 60(5):1946–1958, 2017.
- [57] Erik Gawehn, Jan A. Hiss, J. B. Brown, and Gisbert Schneider. Advancing drug discovery via GPU-based deep learning. *Expert Opinion on Drug Discovery*, 13(7):579–582, 2018.
- [58] Arkadiusz Dudek, Tomasz Arodz, and Jorge Galvez. Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review. *Combinatorial Chemistry & High Throughput Screening*, 9(3):213–228, 2006.

- [59] Sheng Yong Yang. Pharmacophore modeling and applications in drug discovery: Challenges and recent advances. *Drug Discovery Today*, 15(11-12):444–450, 2010.
- [60] Marcus Olivecrona, Thomas Blaschke, Ola Engkvist, and Hongming Chen. Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics*, 9(1):1–14, 2017.
- [61] Miha Skalic, José Jiménez, Davide Sabbadin, and Gianni De Fabritiis. Shape-Based Generative Modeling for de Novo Drug Design. *Journal of Chemical Information and Modeling*, 59(3):1205–1214, 2019.
- [62] Marwin H.S. Segler, Thierry Kogej, Christian Tyrchan, and Mark P. Waller. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Central Science*, 4(1):120–131, 2018.
- [63] Wengong Jin, Regina Barzilay, and Tbmimi Jaakkola. Junction tree variational autoencoder for molecular graph generation. *35th International Conference on Machine Learning, ICML 2018*, 5:3632–3648, 2018.
- [64] Rafael Gómez-Bombarelli, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science*, 4(2):268–276, 2018.
- [65] Connor W. Coley, William H. Green, and Klavs F. Jensen. Machine Learning in Computer-Aided Synthesis Planning. *Accounts of Chemical Research*, 51(5):1281–1289, 2018.
- [66] Marwin H.S. Segler, Mike Preuss, and Mark P. Waller. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*, 555(7698):604–610, 2018.
- [67] Connor W. Coley, Dale A. Thomas, Justin A.M. Lummiss, Jonathan N. Jaworski, Christopher P. Breen, Victor Schultz, Travis Hart, Joshua S. Fishman, Luke Rogers, Hanyu Gao, Robert W. Hicklin, Pieter P. Plehiers, Joshua Byington, John S. Piotti, William H. Green, A. John Hart, Timothy F. Jamison, and Klavs F. Jensen. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science*, 365:1–9, 2019.
- [68] Sara Szymkuć, Ewa P. Gajewska, Tomasz Klucznik, Karol Molga, Piotr Dittwald, Michał Startek, Michał Bajczyk, and Bartosz A. Grzybowski. Computer-Assisted Synthetic Planning: The End of the Beginning. *Angewandte Chemie - International Edition*, 55(20):5904–5937, 2016.
- [69] Sai Krishna Gottipati, Boris Sattarov, Sufeng Niu, Yashaswi Pathak, Haoran Wei, Shengchao Liu, Karam M. J. Thomas, Simon Blackburn, Connor W. Coley, Jian Tang, Sarath Chandar, and Yoshua Bengio. Learning To Navigate The Synthetically Accessible Chemical Space Using Reinforcement Learning. *arXiv*, pages 1–23, 2020.
- [70] Peter Ertl and Ansgar Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics*, 1(1):1–11, 2009.
- [71] Fergus Imrie, Anthony R. Bradley, Mihaela van der Schaar, and Charlotte M. Deane. Deep Generative Models for 3D Linker Design. *Journal of Chemical Information and Modeling*, 6:1983–1995, 2020.
- [72] Christopher A. Lipinski, Franco Lombardo, Beryl W. Dominy, and Paul J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 64(SUPPL.):4–17, 2012.
- [73] Han van de Waterbeemd and Eric Gifford. ADMET in silico modelling: Towards prediction paradise? *Nature Reviews Drug Discovery*, 2(3):192–204, 2003.
- [74] Dean R. Artis, Jack J. Lin, Chao Zhang, Weiru Wang, Upasana Mehra, Mylene Perreault, David Erbe, Heike I. Krupka, Bruce P. England, James Arnold, Alexander N. Plotnikov, Adhirai Marimuthu, Hoa Nguyen, Sarah Will, Maxime Signaevsky, John Kral, John Cantwell, Calvin Settachatgull, Douglas S. Yan, Daniel Fong, Angela Oh, Shenghua Shi, Patrick Womack, Benjamin Powell, Gaston Habets, Brian L. West, Kam Y.J. Zhang, Michael V. Milburn, George P. Vlasuk, K. Peter Hirth, Keith Nolop,

- Gideon Bollag, Prabha N. Ibrahim, and James F. Tobin. Scaffold-based discovery of indeglitazar, a PPAR pan-active anti-diabetic agent. *Proceedings of the National Academy of Sciences of the United States of America*, 106(1):262–267, 2009.
- [75] Xin Yang, Yifei Wang, Ryan Byrne, Gisbert Schneider, and Shengyong Yang. Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. *Chemical Reviews*, 119(18):10520–10594, 2019.
- [76] Jérémy Besnard, Gian Filippo Ruda, Vincent Setola, Keren Abecassis, Ramona M. Rodriguiz, Xi Ping Huang, Suzanne Norval, Maria F. Sassano, Antony I. Shin, Lauren A. Webster, Frederick R.C. Simeons, Laste Stojanovski, Annik Prat, Nabil G. Seidah, Daniel B. Constam, G. Richard Bickerton, Kevin D. Read, William C. Wetsel, Ian H. Gilbert, Bryan L. Roth, and Andrew L. Hopkins. Automated design of ligands to polypharmacological profiles. *Nature*, 492(7428):215–220, 2012.
- [77] Anvita Gupta, Alex T. Müller, Berend J.H. Huisman, Jens A. Fuchs, Petra Schneider, and Gisbert Schneider. Generative Recurrent Networks for De Novo Drug Design. *Molecular Informatics*, 37(1), 2018.
- [78] Laurent Batiste, Andrea Unzue, Aymeric Dolbois, Fabrice Hassler, Xuan Wang, Nicholas Deerain, Jian Zhu, Dimitrios Spiliotopoulos, Cristina Nevado, and Amedeo Caffisch. Chemical Space Expansion of Bromodomain Ligands Guided by in Silico Virtual Couplings (AutoCouple). *ACS Central Science*, 4(2):180–188, 2018.
- [79] Florent Chevillard, Helena Rimmer, Cecilia Betti, Els Pardon, Steven Ballet, Niek Van Hilten, Jan Steyaert, Wibke E. Diederich, and Peter Kolb. Binding-Site Compatible Fragment Growing Applied to the Design of β 2 -Adrenergic Receptor Ligands. *Journal of Medicinal Chemistry*, 61(3):1118–1129, 2018.
- [80] Laurent Hoffer, Yuliia V. Voitovich, Brigitt Raux, Kendall Carrasco, Christophe Muller, Aleksey Y. Fedorov, Carine Derviaux, Agnès Amouric, Stéphane Betzi, Dragos Horvath, Alexandre Varnek, Yves Collette, Sébastien Combes, Philippe Roche, and Xavier Morelli. Integrated Strategy for Lead Optimization Based on Fragment Growing: The Diversity-Oriented-Target-Focused-Synthesis Approach. *Journal of Medicinal Chemistry*, 61(13):5719–5732, 2018.
- [81] Yan Li, Yaping Sun, Yunpeng Song, Dongcheng Dai, Zhixiong Zhao, Qing Zhang, Wenge Zhong, Liaoyuan A. Hu, Yingli Ma, Xun Li, and Renxiao Wang. Fragment-Based Computational Method for Designing GPCR Ligands. *Journal of Chemical Information and Modeling*, 2019.
- [82] Peter Kolb, Catherine Berset Kipouros, Danzhi Huang, and Amedeo Caffisch. Structure-based tailoring of compound libraries for high-throughput screening: Discovery of novel EphB4 kinase inhibitors. *Proteins: Structure, Function and Genetics*, 73(1):11–18, 2008.
- [83] Mohamad Sabbah, Vitor Mendes, Robert G. Vistal, David M. G. Dias, Monika Záhorszská, Katarína Mikušová, Jana Korduláková, Anthony G. Coyne, Tom L. Blundell, and Chris Abell. Fragment-Based Design of Mycobacterium tuberculosis InhA Inhibitors. *Journal of Medicinal Chemistry*, 63:4749–4761, 2020.
- [84] Jacek Kwiatkowski, Boping Liu, Doris Hui Ying Tee, Guoying Chen, Nur Huda Binte Ahmad, Yun Xuan Wong, Zhi Ying Poh, Shi Hua Ang, Eldwin Sum Wai Tan, Esther Hq Ong, Dinie Nurul, Anders Poulsen, Vishal Pendharkar, Kanda Sangthongpitag, May Ann Lee, Sugunavathi Sepramaniam, Soo Yei Ho, Joseph Cherian, Jeffrey Hill, Thomas H. Keller, and Alvin W. Hung. Fragment-Based Drug Discovery of Potent Protein Kinase C Iota Inhibitors. *Journal of Medicinal Chemistry*, 61(10):4386–4396, 2018.
- [85] Tom D. Heightman, Valerio Berdini, Hannah Braithwaite, Ildiko M. Buck, Megan Cassidy, Juan Castro, Aurélie Courtin, James E.H. Day, Charlotte East, Lynsey Fazal, Brent Graham, Charlotte M. Griffiths-Jones, John F. Lyons, Vanessa Martins, Sandra Muench, Joanne M. Munck, David Norton, Marc O'Reilly, Nick Palmer, Puja Pathuri, Michael Reader, David C. Rees, Sharna J. Rich, Caroline Richardson, Harpreet Saini, Neil T. Thompson, Nicola G. Wallis, Hugh Walton, Nicola E. Wilsher, Alison J.A. Woolford, Michael Cooke, David Cousin, Stuart Onions, Jonathan Shannon, John Watts, and Christopher W. Murray. Fragment-Based Discovery of a Potent, Orally Bioavailable Inhibitor That Modulates the Phosphorylation and Catalytic Activity of ERK1/2. *Journal of Medicinal Chemistry*, 61(11):4978–4992, 2018.

- [86] Feng Wang, Kyu Ok Jeon, James M. Salovich, Jonathan D. Macdonald, Joseph Alvarado, Rocco D. Gogliotti, Jason Phan, Edward T. Olejniczak, Qi Sun, Shidong Wang, Demarco Camper, Joannes P. Yuh, J. Grace Shaw, Jiqing Sai, Olivia W. Rossanese, William P. Tansey, Shaun R. Stauffer, and Stephen W. Fesik. Discovery of Potent 2-Aryl-6,7-dihydro-5 H-pyrrolo[1,2- a]imidazoles as WDR5-WIN-Site Inhibitors Using Fragment-Based Methods and Structure-Based Design. *Journal of Medicinal Chemistry*, 61(13):5623–5642, 2018.
- [87] Thomas Drapier, Pierre Geubelle, Charlotte Bouckaert, Lise Nielsen, Saara Laulumaa, Eric Goffin, Sébastien Dilly, Pierre Francotte, Julien Hanson, Lionel Pochet, Jette Sandholm Kastrop, and Bernard Pirotte. Enhancing Action of Positive Allosteric Modulators through the Design of Dimeric Compounds. *Journal of Medicinal Chemistry*, 61(12):5279–5291, 2018.
- [88] Alla Korepanova, Kenton L. Longenecker, Steve D. Pratt, Sanjay C. Panchal, Richard F. Clark, Marc Lake, Sujatha M. Gopalakrishnan, Diana Raich, Chaohong Sun, and Andrew M. Petros. Fragment-based discovery of a potent NAMPT inhibitor. *Bioorganic and Medicinal Chemistry Letters*, 28(3):437–440, 2018.
- [89] Ting Chen, Venkataswamy Sorna, Susie Choi, Lee Call, Jared Bearss, Kent Carpenter, Steven L. Warner, Sunil Sharma, David J. Bearss, and Hariprasad Vankayalapati. Fragment-based design, synthesis, biological evaluation, and SAR of 1H-benzo[d]imidazol-2-yl)-1H-indazol derivatives as potent PDK1 inhibitors. *Bioorganic and Medicinal Chemistry Letters*, 27(24):5473–5480, 2017.
- [90] Fatiha Benmansour, Iuni Trist, Bruno Coutard, Etienne Decroly, Gilles Querat, Andrea Brancale, and Karine Barral. Discovery of novel dengue virus NS5 methyltransferase non-nucleoside inhibitors by fragment-based drug design. *European Journal of Medicinal Chemistry*, 125:865–880, 2017.
- [91] Michael F. Mesleh, Jason B. Cross, Jing Zhang, Jan Kahmann, Ole A. Andersen, John Barker, Robert K. Cheng, Brunella Felicetti, Michael Wood, Andrea T. Hadfield, Christoph Scheich, Terence I. Moy, Qingyi Yang, Joseph Shotwell, Kien Nguyen, Blaise Lippa, Roland Dolle, and M. Dominic Ryan. Fragment-based discovery of DNA gyrase inhibitors targeting the ATPase subunit of GyrB. *Bioorganic and Medicinal Chemistry Letters*, 26(4):1314–1318, 2016.
- [92] John B. Jordan, Douglas A. Whittington, Michael D. Bartberger, E. Allen Sickmier, Kui Chen, Yuan Cheng, and Ted Judd. Fragment-Linking Approach Using 19F NMR Spectroscopy to Obtain Highly Potent and Selective Inhibitors of β -Secretase. *Journal of Medicinal Chemistry*, 59(8):3732–3749, 2016.
- [93] Madeline E. Kavanagh, Anthony G. Coyne, Kirsty J. McLean, Guy G. James, Colin W. Levy, Leonardo B. Marino, Luiz Pedro S. De Carvalho, Daniel S.H. Chan, Sean A. Hudson, Sachin Surade, David Leys, Andrew W. Munro, and Chris Abell. Fragment-Based Approaches to the Development of Mycobacterium tuberculosis CYP121 Inhibitors. *Journal of Medicinal Chemistry*, 59(7):3272–3302, 2016.
- [94] Thomas G. Davies, William E. Wixted, Joseph E. Coyle, Charlotte Griffiths-Jones, Keisha Hearn, Rachel McMenamin, David Norton, Sharna J. Rich, Caroline Richardson, Gordon Saxty, Henriëtte M.G. Willems, Alison J.A. Woolford, Joshua E. Cottom, Jen Pyng Kou, John G. Yonchuk, Heidi G. Feldser, Yolanda Sanchez, Joseph P. Foley, Brian J. Bolognese, Gregory Logan, Patricia L. Podolin, Hongxing Yan, James F. Callahan, Tom D. Heightman, and Jeffrey K. Kerns. Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein-Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery. *Journal of Medicinal Chemistry*, 59(8):3991–4006, 2016.
- [95] William D. Shipe, Steven S. Sharik, James C. Barrow, Georgia B. McGaughey, Cory R. Theberge, Jason M. Uslaner, Youwei Yan, John J. Renger, Sean M. Smith, Paul J. Coleman, and Christopher D. Cox. 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*, 58(19):7888–7894, 2015.
- [96] Andreas Ritzén, Morten D. Sørensen, Kevin N. Dack, Daniel R. Greve, Anders Jerre, Martin A. Carnerup, Klaus A. Rytved, and Jesper Bagger-Bahnsen. Fragment-Based Discovery of 6-Arylindazole JAK Inhibitors. *ACS Medicinal Chemistry Letters*, 7(6):641–646, 2016.

- [97] Dace Rasina, Martins Otikovs, Janis Leitans, Rosario Recacha, Oleksandr V. Borysov, Iveta Kanepelapsa, Ilona Domranceva, Teodors Pantelejevs, Kaspars Tars, Michael J. Blackman, Kristaps Jaudzems, and Aigars Jirgensons. Fragment-Based Discovery of 2-Aminoquinazolin-4(3H)-ones As Novel Class Nonpeptidomimetic Inhibitors of the Plasmepsins I, II, and IV. *Journal of Medicinal Chemistry*, 59(1):374–387, 2016.
- [98] Daniel J. Burdick, Shumei Wang, Christopher Heise, Borlan Pan, Jake Drummond, Jianping Yin, Lauren Goeser, Steven Magnuson, Jeff Blaney, John Moffat, Weiru Wang, and Huifen Chen. Fragment-based discovery of potent ERK2 pyrrolopyrazine inhibitors. *Bioorganic and Medicinal Chemistry Letters*, 25(21):4728–4732, 2015.
- [99] Dawn M. George, Eric C. Breinlinger, Michael Friedman, Yang Zhang, Jianfei Wang, Maria Argiriadi, Pratima Bansal-Pakala, Martine Barth, David B. Duignan, Prisca Honore, Qingyu Lang, Scott Mittelstadt, Dominique Potin, Lian Rundell, and Jeremy J. Edmunds. Discovery of selective and orally bioavailable protein kinase C θ (PKC θ) inhibitors from a fragment hit. *Journal of Medicinal Chemistry*, 58(1):222–236, 2015.
- [100] Ola Fjellström, Sibel Akkaya, Hans Georg Beisel, Per Olof Eriksson, Karl Erixon, David Gustafsson, Ulrik Jurva, Daiwu Kang, David Karis, Wolfgang Knecht, Viveca Nerme, Ingemar Nilsson, Thomas Olsson, Alma Redzic, Robert Roth, Jenny Sandmark, Anna Tigerström, and Linda Öster. Creating novel activated factor XI inhibitors through fragment based lead generation and structure aided drug design. *PLoS ONE*, 10(1), 2015.
- [101] Sophie M. Bertrand, Nicolas Ancellin, Benjamin Beaufile, Ryan P. Bingham, Jennifer A. Borthwick, Anne Bénédicte Boullay, Eric Boursier, Paul S. Carter, Chun Wa Chung, Ian Churcher, Nerina Dodic, Marie Hélène Fouchet, Charlene Fournier, Peter L. Francis, Laura A. Gummer, Kenny Herry, Andrew Hobbs, Clare I. Hobbs, Paul Homes, Craig Jamieson, Edwige Nicodeme, Stephen D. Pickett, Iain H. Reid, Graham L. Simpson, Lisa A. Sloan, Sarah E. Smith, Donald O.N. Somers, Claus Spitzfaden, Colin J. Suckling, Klara Valko, Yoshiaki Washio, and Robert J. Young. The Discovery of in Vivo Active Mitochondrial Branched-Chain Aminotransferase (BCATm) Inhibitors by Hybridizing Fragment and HTS Hits. *Journal of Medicinal Chemistry*, 58(18):7140–7163, 2015.
- [102] Jason P. Burke, Zhiguo Bian, Subrata Shaw, Bin Zhao, Craig M. Goodwin, Johannes Belmar, Carrie F. Browning, Dominico Vigil, Anders Friberg, De Marco V. Camper, Olivia W. Rossanese, Taekyu Lee, Edward T. Olejniczak, and Stephen W. Fesik. Discovery of tricyclic indoles that potently inhibit Mcl-1 using fragment-based methods and structure-based design. *Journal of Medicinal Chemistry*, 58(9):3794–3805, 2015.
- [103] John A. Christopher, Sarah J. Aves, Kirstie A. Bennett, Andrew S. Doré, James C. Errey, Ali Jazayeri, Fiona H. Marshall, Krzysztof Okrasa, Maria J. Serrano-Vega, Benjamin G. Tehan, Giselle R. Wiggin, and Miles Congreve. Fragment and Structure-Based Drug Discovery for a Class C GPCR: Discovery of the mGlu5 Negative Allosteric Modulator HTL14242 (3-Chloro-5-[6-(5-fluoropyridin-2-yl)pyrimidin-4-yl]benzonitrile). *Journal of Medicinal Chemistry*, 58(16):6653–6664, 2015.
- [104] Sarah Picaud, Maria Strocchia, Stefania Terracciano, Gianluigi Lauro, Jacqui Mendez, Danette L. Daniels, Raffaele Riccio, Giuseppe Bifulco, Ines Bruno, and Panagis Filippakopoulos. 9 H -purine scaffold reveals induced-fit pocket plasticity of the brd9 bromodomain. *Journal of Medicinal Chemistry*, 58(6):2718–2736, 2015.
- [105] Daniel L. Cheney, Jeffrey M. Bozarth, William J. Metzler, Paul E. Morin, Luciano Mueller, John A. Newitt, Alexandra H. Nirschl, Alan R. Rendina, James K. Tamura, Anzhi Wei, Xiao Wen, Nicholas R. Wurtz, Dietmar A. Seiffert, Ruth R. Wexler, and E. Scott Priestley. Discovery of novel P1 groups for coagulation factor VIIa inhibition using fragment-based screening. *Journal of Medicinal Chemistry*, 58(6):2799–2808, 2015.
- [106] Thomas A. Halgren. Identifying and characterizing binding sites and assessing druggability. *Journal of Chemical Information and Modeling*, 49(2):377–389, 2009.
- [107] Andriy Kovalenko and Fumio Hirata. Three-dimensional density profiles of water in contact with a solute of arbitrary shape: A RISM approach. *Chemical Physics Letters*, 290(1-3):237–244, 1998.

- [108] David Carcache, Anna Vulpetti, Joerg Kallen, Henri Mattes, David Orain, Rowan Stringer, Eric Van-grevelinghe, Romain M. Wolf, Klemens Kaupmann, Johannes Ottl, Janet Dawson, Nigel G. Cooke, Klemens Hoegenauer, Andreas Billich, Juergen Wagner, Christine Guntermann, and Samuel Hintermann. Optimizing a Weakly Binding Fragment into a Potent ROR γ t Inverse Agonist with Efficacy in an in Vivo Inflammation Model. -, 61:6724–6735, 2018.
- [109] Jason B. Cross, Jing Zhang, Qingyi Yang, Michael F. Mesleh, Jan Antoinette C. Romero, Bin Wang, Doug Bevan, Katherine M. Poutsiaka, Felix Epie, Terence Moy, Anu Daniel, Joseph Shotwell, Brian Chamberlain, Nicole Carter, Ole Andersen, John Barker, M. Dominic Ryan, Chester A. Metcalf, Jared Silverman, Kien Nguyen, Blaise Lippa, and Roland E. Dolle. Discovery of Pyrazolopyridones as a Novel Class of Gyrase B Inhibitors Using Structure Guided Design. *ACS Medicinal Chemistry Letters*, 7(4):374–378, 2016.
- [110] Robert Abel, Tom Young, Ramy Farid, Bruce J. Berne, and Richard A. Friesner. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *Journal of the American Chemical Society*, 130(9):2817–2831, 2008.
- [111] Yumi Matsui, Isao Yasumatsu, Takashi Asahi, Takahiro Kitamura, Kazuo Kanai, Osamu Ubukata, Hitoshi Hayasaka, Sachiko Takaishi, Hiroyuki Hanzawa, and Shinichi Katakura. Discovery and structure-guided fragment-linking of 4-(2,3-dichlorobenzoyl)-1-methyl-pyrrole-2-carboxamide as a pyruvate kinase M2 activator. *Bioorganic and Medicinal Chemistry*, 25(13):3540–3546, 2017.
- [112] Jark Böttcher, David Dilworth, Ulrich Reiser, Ralph A. Neumüller, Michael Schleicher, Mark Petronczki, Markus Zeeb, Nikolai Mischerikow, Abdellah Allali-Hassani, Magdalena M. Szewczyk, Fengling Li, Steven Kennedy, Masoud Vedadi, Dalia Barsyte-Lovejoy, Peter J. Brown, Kilian V.M. Huber, Catherine M. Rogers, Carrow I. Wells, Oleg Fedorov, Klaus Rumpel, Andreas Zoepfel, Moriz Mayer, Tobias Wunberg, Dietrich Böse, Stephan Zahn, Heribert Arnhof, Helmut Berger, Christoph Reiser, Alexandra Hörmann, Teresa Krammer, Maja Corcokovic, Bernadette Sharps, Sandra Winkler, Daniela Häring, Xiao Ling Cockcroft, Julian E. Fuchs, Barbara Müllauer, Alexander Weiss-Puxbaum, Thomas Gerstberger, Guido Boehmelt, Christopher R. Vakoc, Cheryl H. Arrowsmith, Mark Pearson, and Darryl B. McConnell. Fragment-based discovery of a chemical probe for the PWWP1 domain of NSD3. *Nature Chemical Biology*, 15(8):822–829, 2019.
- [113] Rosanna Meine, Walter Becker, Hannes Falke, Lutz Preu, Nadège Loaëc, Laurent Meijer, and Conrad Kunick. Indole-3-carbonitriles as DYRK1A inhibitors by fragment-based drug design. *Molecules*, 23(2):1–23, 2018.
- [114] Robert Schulz, Amira Atef, Daniel Becker, Franziska Gottschalk, Carolin Tauber, Stefan Wagner, Christoph Arkona, Atef A. Abdel-Hafez, Hassan H. Farag, Jörg Rademann, and Gerhard Wolber. Phenylthiomethyl Ketone-Based Fragments Show Selective and Irreversible Inhibition of Enteroviral 3C Proteases. *Journal of Medicinal Chemistry*, 61(3):1218–1230, 2018.
- [115] M. Dawidowski, L. Emmanouilidis, V. C. Kalel, K. Tripsianes, K. Schorpp, K. Hadian, M. Kaiser, P. Mäser, M. Kolonko, S. Tanghe, A. Rodriguez, W. Schliebs, R. Erdmann, M. Sattler, and G. M. Popowicz. Inhibitors of PEX14 disrupt protein import into glycosomes and kill Trypanosoma parasites. *Science*, 355(6332):1416–1420, 2017.
- [116] William McCoull, Roman D. Abrams, Erica Anderson, Kevin Blades, Peter Barton, Matthew Box, Jonathan Burgess, Kate Byth, Qing Cao, Claudio Chuaqui, Rodrigo J. Carbajo, Tony Cheung, Erin Code, Andrew D. Ferguson, Shaun Fillery, Nathan O. Fuller, Eric Gangl, Ning Gao, Matthew Grist, David Hargreaves, Martin R. Howard, Jun Hu, Paul D. Kemmitt, Jennifer E. Nelson, Nichole O’Connell, D. Bryan Prince, Piotr Raubo, Philip B. Rawlins, Graeme R. Robb, Junjie Shi, Michael J. Waring, David Whittaker, Marta Wylot, and Xiahui Zhu. Discovery of Pyrazolo[1,5-a]pyrimidine B-Cell Lymphoma 6 (BCL6) Binders and Optimization to High Affinity Macrocytic Inhibitors. *Journal of Medicinal Chemistry*, 60(10):4386–4402, 2017.
- [117] Frederick Cohen Paola Di Lello*,† Richard Pastor*,† Jeremy M. Murray, Robert A. Blake, Till Maurer Terry D. Crawford, Joy Drobnick, Jason Drummond, Lorna Kategaya, Tracy Kleinheinz, Lionel Rougé, Xianrui Zhao, Ingrid Wertz, Chudi Ndubaku, and Vickie Tsui*. Discovery of Small-Molecule Inhibitors

- of Ubiquitin Specific Protease 7 (USP7) Using Integrated NMR and in Silico Techniques. *Journal of Medicinal Chemistry*, 60:10056–10070, 2017.
- [118] Qiufeng Liu, Fubao Huang, Xiaojing Yuan, Kai Wang, Yi Zou, Jianhua Shen, and Yechun Xu. Structure-Guided Discovery of Novel, Potent, and Orally Bioavailable Inhibitors of Lipoprotein-Associated Phospholipase A2. *Journal of Medicinal Chemistry*, 60(24):10231–10244, 2017.
- [119] Axel Rudling, Robert Gustafsson, Ingrid Almlöf, Evert Homan, Martin Scobie, Ulrika Warpman Berglund, Thomas Helleday, Pål Stenmark, and Jens Carlsson. Fragment-Based Discovery and Optimization of Enzyme Inhibitors by Docking of Commercial Chemical Space. *Journal of Medicinal Chemistry*, 60(19):8160–8169, 2017.
- [120] Mark Adams, Toshitake Kobayashi, J. David Lawson, Morihisa Saitoh, Kenichiro Shimokawa, Simone V. Bigi, Mark S. Hixon, Christopher R. Smith, Takayuki Tatamiya, Masayuki Goto, Joseph Russo, Charles E. Grimshaw, and Steven Swann. Fragment-based drug discovery of potent and selective MKK3/6 inhibitors. *Bioorganic and Medicinal Chemistry Letters*, 26(3):1086–1089, 2016.
- [121] David C. McKinney, Charles J. Eyermann, Rong Fang Gu, Jun Hu, Steven L. Kazmirski, Sushmita D. Lahiri, Andrew R. McKenzie, Adam B. Shapiro, and Gloria Breault. Antibacterial FabH Inhibitors with Mode of Action Validated in *Haemophilus influenzae* by in Vitro Resistance Mutation Mapping. *ACS Infectious Diseases*, 2(7):456–464, 2016.
- [122] Alison J.A. Woolford, Philip J. Day, Véronique Bénétou, Valerio Berdini, Joseph E. Coyle, Yann Dudit, Pascal Grondin, Pascal Huet, Lydia Y.W. Lee, Eric S. Manas, Rachel L. McMennamin, Christopher W. Murray, Lee W. Page, Vipulkumar K. Patel, Florent Potvain, Sharna J. Rich, Yingxia Sang, Don O. Somers, Lionel Trotter, Zehong Wan, and Xiaomin Zhang. Fragment-Based Approach to the Development of an Orally Bioavailable Lactam Inhibitor of Lipoprotein-Associated Phospholipase A2 (Lp-PLA2). *Journal of Medicinal Chemistry*, 59(23):10738–10749, 2016.
- [123] Jan Lanz and Rainer Ried. Merging allosteric and active site binding motifs: De novo generation of target selectivity and potency via natural-product-derived fragments. *ChemMedChem*, 10(3):451–454, 2015.
- [124] Gianni Chessari, Ildiko M. Buck, James E.H. Day, Philip J. Day, Aman Iqbal, Christopher N. Johnson, Edward J. Lewis, Vanessa Martins, Darcey Miller, Michael Reader, David C. Rees, Sharna J. Rich, Emiliano Tamanini, Marc Vitorino, George A. Ward, Pamela A. Williams, Glyn Williams, Nicola E. Wilsher, and Alison J.A. Woolford. Fragment-Based Drug Discovery Targeting Inhibitor of Apoptosis Proteins: Discovery of a Non-Alanine Lead Series with Dual Activity Against cIAP1 and XIAP. *Journal of Medicinal Chemistry*, 58(16):6574–6588, 2015.
- [125] Stephan G. Zech, Anna Kohlmann, Tianjun Zhou, Feng Li, Rachel M. Squillace, Lois E. Parillon, Matthew T. Greenfield, David P. Miller, Jiwei Qi, R. Mathew Thomas, Yihan Wang, Yongjin Xu, Juan J. Miret, William C. Shakespeare, Xiaotian Zhu, and David C. Dalgarno. Novel Small Molecule Inhibitors of Choline Kinase Identified by Fragment-Based Drug Discovery. *Journal of Medicinal Chemistry*, 59(2):671–686, 2016.
- [126] Jean Rémy Marchand, Andrea Dalle Vedove, Graziano Lolli, and Amedeo Caffisch. Discovery of Inhibitors of Four Bromodomains by Fragment-Anchored Ligand Docking. *Journal of Chemical Information and Modeling*, 57(10):2584–2597, 2017.
- [127] Hongtao Zhao, Jing Dong, Karine Lafleur, Cristina Nevado, and Amedeo Caffisch. Discovery of a novel chemotype of tyrosine kinase inhibitors by fragment-based docking and molecular dynamics. *ACS Medicinal Chemistry Letters*, 3(10):834–838, 2012.
- [128] Min Xu, Andrea Unzue, Jing Dong, Dimitrios Spiliotopoulos, Cristina Nevado, and Amedeo Caffisch. Discovery of CREBBP Bromodomain Inhibitors by High-Throughput Docking and Hit Optimization Guided by Molecular Dynamics. *Journal of Medicinal Chemistry*, 59(4):1340–1349, 2016.

- [129] Andrea Unzue, Min Xu, Jing Dong, Lars Wiedmer, Dimitrios Spiliotopoulos, Amedeo Cafisch, and Cristina Nevado. Fragment-Based Design of Selective Nanomolar Ligands of the CREBBP Bromodomain. *Journal of Medicinal Chemistry*, 59(4):1350–1356, 2016.
- [130] Christopher W. Murray, Valerio Berdini, Ildiko M. Buck, Maria E. Carr, Anne Cleasby, Joseph E. Coyle, Jayne E. Curry, James E.H. Day, Phillip J. Day, Keisha Hearn, Aman Iqbal, Lydia Y.W. Lee, Vanessa Martins, Paul N. Mortenson, Joanne M. Munck, Lee W. Page, Sahil Patel, Susan Roomans, Kirsten Smith, Emiliano Tamanini, and Gordon Saxty. Fragment-Based Discovery of Potent and Selective DDR1/2 Inhibitors. *ACS Medicinal Chemistry Letters*, 6(7):798–803, 2015.
- [131] Timo Heikkilä, Srinath Thirumalairajan, Matthew Davies, Mark R. Parsons, A. Glenn McConkey, Colin W.G. Fishwick, and A. Peter Johnson. The first de novo designed inhibitors of Plasmodium falciparum dihydroorotate dehydrogenase. *Bioorganic and Medicinal Chemistry Letters*, 16(1):88–92, 2006.
- [132] Matthew Davies, Timo Heikkilä, Glenn A. McConkey, Colin W.G. Fishwick, Mark R. Parsons, and A. Peter Johnson. Structure-based design, synthesis, and characterization of inhibitors of human and Plasmodium falciparum dihydroorotate dehydrogenases. *Journal of Medicinal Chemistry*, 52(9):2683–2693, 2009.
- [133] N. Yi Mok, James Chadwick, Katherine A.B. Kellett, Eva Casas-Arce, Nigel M. Hooper, A. Peter Johnson, and Colin W.G. Fishwick. Discovery of biphenylacetamide-derived inhibitors of BACE1 using de novo structure-based molecular design. *Journal of Medicinal Chemistry*, 56(5):1843–1852, 2013.
- [134] Mark Rogers-Evans, Alexander I. Alanine, Konrad H. Bleicher, Dagmar Kube, and Gisbert Schneider. Identification of novel cannabinoid receptor ligands via evolutionary de novo design and rapid parallel synthesis. *QSAR and Combinatorial Science*, 23(6):426–430, 2004.
- [135] Gerhard Wolber and Thierry Langer. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *Journal of Chemical Information and Modeling*, 45(1):160–169, 2005.
- [136] Yaxia Yuan, Jianfeng Pei, and Luhua Lai. LigBuilder V3: A Multi-Target de novo Drug Design Approach. *Frontiers in Chemistry*, 8(February):1–18, 2020.
- [137] Nicholas M. Pearce, Tobias Krojer, Anthony R. Bradley, Patrick Collins, Radosław P. Nowak, Romain Talon, Brian D. Marsden, Sebastian Kelm, Jiye Shi, Charlotte M. Deane, and Frank Von Delft. A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nature Communications*, 8:24–29, 2017.
- [138] Alex Zhavoronkov. Medicinal Chemists versus Machines Challenge: What Will It Take to Adopt and Advance Artificial Intelligence for Drug Discovery? -, pages 0–2, 2021.
- [139] Olivia Doppelt-Azeroual, Fabien Mareuil, Eric Deveaud, Matúš Kalaš, Nicola Soranzo, Marius van den Beek, Björn Grüning, Jon Ison, and Hervé Ménager. ReGaTE: Registration of galaxy tools in Elixir. *GigaScience*, 6(6):1–4, 2017.
- [140] Carles Perez, Daniel Soler, Robert Soliva, and Victor Guallar. FragPELE: Dynamic Ligand Growing within a Binding Site. A Novel Tool for Hit-To-Lead Drug Design. *Journal of Chemical Information and Modeling*, 60:1728–1736, 2020.
- [141] William L. Jorgensen. Efficient drug lead discovery and optimization. *Accounts of Chemical Research*, 42(6):724–733, 2009.
- [142] Sergio Ruiz-Carmona, Daniel Alvarez-Garcia, Nicolas Foloppe, A. Beatriz Garmendia-Doval, Szilveszter Juhos, Peter Schmidtke, Xavier Barril, Roderick E. Hubbard, and S. David Morley. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Computational Biology*, 10(4):1–7, 2014.

- [143] Ruth Huey, Garrett M Morris, Arthur J Olson, and David S Goodsell. A semiempirical free energy force field with charge-based desolvation. *Journal of computational chemistry*, 28(6):1145–1152, apr 2007.
- [144] Samuel Genheden and Ulf Ryde. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, 10(5):449–461, 2015.
- [145] William L Jorgensen and Julian Tirado-Rives. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, mar 1988.
- [146] Author Manuscript, Jianing Li, Robert Abel, Kai Zhu, Yixiang Cao, Suwen Zhao, Richard A. Friesner, Author Manuscript, Jianing Li, Robert Abel, Kai Zhu, Yixiang Cao, Suwen Zhao, and Richard A. Friesner. The VSGB 2.0 Model: A Next Generation Energy Model for High Resolution Protein Structure Modeling. *Proteins: Structure, Function, and Bioinformatics*, 79(10):2794–2812, 2012.
- [147] Sergio Ruiz-Carmona, Peter Schmidtke, F. Javier Luque, Lisa Baker, Natalia Matassova, Ben Davis, Stephen Roughley, James Murray, Rod Hubbard, and Xavier Barril. Dynamic undocking and the quasi-bound state as tools for drug discovery. *Nature Chemistry*, 9(3):201–206, 2017.
- [148] Maciej Majewski, Sergio Ruiz-Carmona, and Xavier Barril. An investigation of structural stability in protein-ligand complexes reveals the balance between order and disorder. *Communications Chemistry*, 2(1), 2019.
- [149] Ingo Muegge. PMF Scoring Revisited. *Journal of Medicinal Chemistry*, 49(20):5895–5902, oct 2006.
- [150] CCG. Molecular Operating Environment (MOE) CCG ULC 2016.08. 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2018.
- [151] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8):1950–1958, jun 2010.
- [152] JF Truchon CI Bayly, D McKay. An Informal AMBER Small Molecule Force Field : parm@Frosst. -, 2010.
- [153] Daniel Alvarez-Garcia and Xavier Barril. Molecular simulations with solvent competition quantify water displaceability and provide accurate interaction maps of protein binding sites. *Journal of Medicinal Chemistry*, 57(20):8530–8539, 2014.
- [154] Peter Willett. Similarity Searching Using 2D Structural Fingerprints. -, pages 133–158, 2011.
- [155] Dagmar Stumpfe and Jürgen Bajorath. Similarity searching. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(2):260–282, 2011.
- [156] Adrià Cereto-Massagué, María José Ojeda, Cristina Valls, Miquel Mulero, Santiago Garcia-Vallvé, and Gerard Pujadas. Molecular fingerprint similarity search in virtual screening. *Methods*, 71(C):58–63, 2015.
- [157] Noel M O Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. Open Babel: An open chemical toolbox - 1758-2946-3-33.pdf. -, pages 1–14, 2011.
- [158] Greg Landrum. RDKit Documentation. *Reading and Writing*, 2011.
- [159] G Madhavi Sastry, Matvey Adzhigirey, Tyler Day, Ramakrishna Annabhimoju, and Woody Sherman. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design*, 27(3):221–34, mar 2013.
- [160] Nils Ole Friedrich, Christina De Bruyn Kops, Florian Flachsenberg, Kai Sommer, Matthias Rarey, and Johannes Kirchmair. Benchmarking Commercial Conformer Ensemble Generators. *Journal of Chemical Information and Modeling*, 57(11):2719–2728, 2017.

- [161] Nils Ole Friedrich, Florian Flachsenberg, Agnes Meyder, Kai Sommer, Johannes Kirchmair, and Matthias Rarey. Conformer: A Novel Method for the Generation of Conformer Ensembles. *Journal of Chemical Information and Modeling*, 59(2):731–742, 2019.
- [162] Jean Paul Ebejer, Garrett M. Morris, and Charlotte M. Deane. Freely available conformer generation methods: How good are they? *Journal of Chemical Information and Modeling*, 52(5):1146–1158, 2012.
- [163] I. Jen Chen and Nicolas Foloppe. Tackling the conformational sampling of larger flexible compounds and macrocycles in pharmacology and drug discovery. *Bioorganic and Medicinal Chemistry*, 21(24):7898–7920, 2013.
- [164] Schrödinger. Schrödinger Release 2016-3: Schrödinger, LLC, New York, NY, 2016.
- [165] Mark Ashton, John Barnard, Florence Casset, Michael Charlton, Geoffrey Downs, Dominique Gorse, John Holliday, Roger Lahana, and Peter Willett. Identification of diverse database subsets using property-based and fragment-based molecular descriptions. *Quantitative Structure-Activity Relationships*, 21(6):598–604, 2002.
- [166] Gerd Neudert and Gerhard Klebe. fconv: format conversion, manipulation and feature computation of molecular data. *Bioinformatics*, 27(7):1021–1022, feb 2011.
- [167] Mohammad Hassan Jafari Najaf Abadi, Rana Shafabakhsh, Zatollah Asemi, Hamid Reza Mirzaei, Roxana Sahebnaasagh, Hamed Mirzei, and Michael R. Hamblin. CFIm25 and alternative polyadenylation: Conflicting roles in cancer. *Cancer Letters*, 459(April):112–121, 2019.
- [168] Molly Coseno, Georges Martin, Christopher Berger, Gregory Gilmartin, Walter Keller, and Sylvie Doublé. Crystal structure of the 25 kDa subunit of human cleavage factor I m. *Nucleic Acids Research*, 36(10):3474–3483, 2008.
- [169] Qin Yang, Gregory M. Gilmartin, and Sylvie Doublé. Structural basis of UGUA recognition by the Nudix protein CFIm25 and implications for a regulatory role in mRNA 3 processing. *Proceedings of the National Academy of Sciences of the United States of America*, 107(22):10062–10067, 2010.
- [170] Christoph Grebner, Erik Malmerberg, Andrew Shewmaker, Jose Batista, Anthony Nicholls, and Jens Sadowski. Virtual Screening in the Cloud: How Big Is Big Enough? *Journal of Chemical Information and Modeling*, page acs.jcim.9b00779, 2019.
- [171] Tim Dungeon. <https://sqonk.it/>.
- [172] Schrödinger. LigPrep 2.3 User Manual. *Schrödinger Press*, page 1, 2009.
- [173] Shuwen Zeng, Dominique Baillargeat, Ho-Pui Ho, and Ken-Tye Yong. Nanomaterials enhanced surface plasmon resonance for biological and chemical sensing applications. *Chem. Soc. Rev.*, 43(10):3426–3452, 2014.
- [174] Johannes Schiebel, Nedyalka Radeva, Stefan G. Krimmer, Xiaojie Wang, Martin Stieler, Frederik R. Ehrmann, Kan Fu, Alexander Metz, Franziska U. Huschmann, Manfred S. Weiss, Uwe Mueller, Andreas Heine, and Gerhard Klebe. Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study. *ACS Chemical Biology*, 11(6):1693–1701, 2016.
- [175] Torsten Hoffmann and Marcus Gastreich. The next level in chemical space navigation: going far beyond enumerable compound libraries. *Drug Discovery Today*, 24(5):1148–1156, 2019.
- [176] Florian H. Schopf, Maximilian M. Biebl, and Johannes Buchner. The HSP90 chaperone machinery. *Nature Reviews Molecular Cell Biology*, 18(6):345–360, 2017.

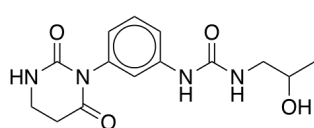
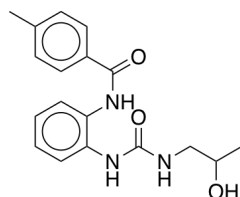
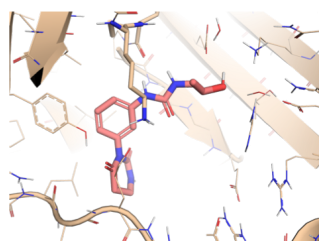
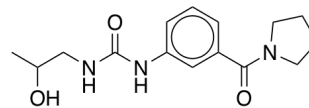
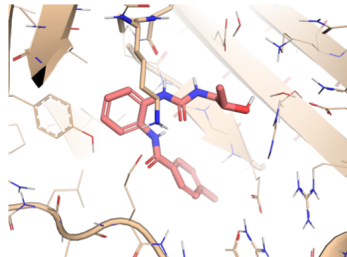
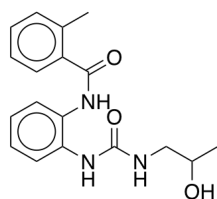
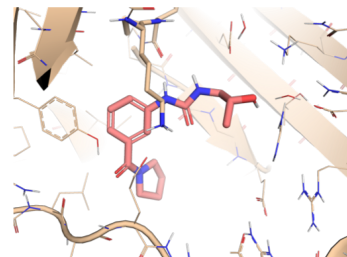
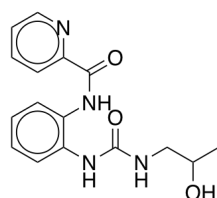
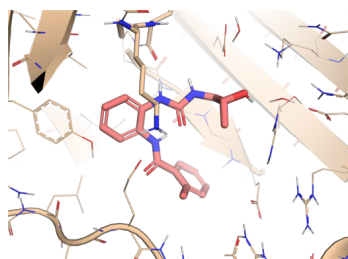
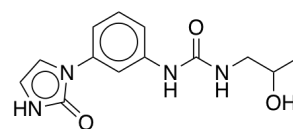
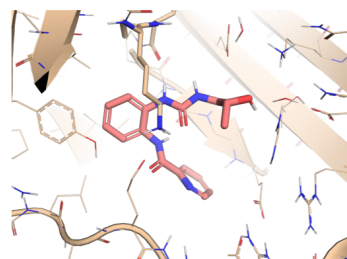
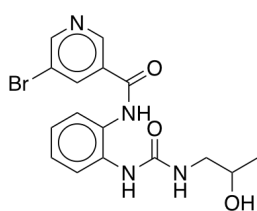
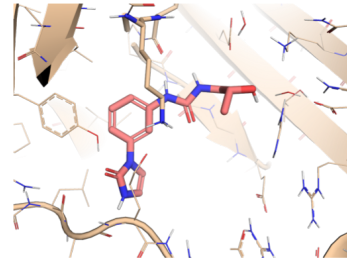
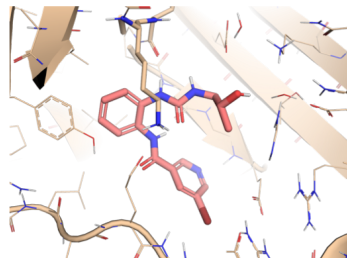
- [177] Jake E. Delmore, Ghayas C. Issa, Madeleine E. Lemieux, Peter B. Rahl, Junwei Shi, Hannah M. Jacobs, Efsthios Kastiris, Timothy Gilpatrick, Ronald M. Paranal, Jun Qi, Marta Chesi, Anna C. Schinzel, Michael R. McKeown, Timothy P. Heffernan, Christopher R. Vakoc, P. Leif Bergsagel, Irene M. Ghobrial, Paul G. Richardson, Richard A. Young, William C. Hahn, Kenneth C. Anderson, Andrew L. Kung, James E. Bradner, and Constantine S. Mitsiades. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell*, 146(6):904–917, 2011.
- [178] Jennifer A. Mertz, Andrew R. Conery, Barbara M. Bryant, Peter Sandy, Srividya Balasubramanian, Deanna A. Mele, Louise Bergeron, and Robert J. Sims. Targeting MYC dependence in cancer by inhibiting BET bromodomains. *PNAS*, 108(40):16669–16674, 2011.
- [179] Panagis Filippakopoulos, Jun Qi, Sarah Picaud, Yao Shen, William B. Smith, Oleg Fedorov, Elizabeth M. Morse, Tracey Keates, Tyler T. Hickman, Ildiko Felletar, Martin Philpott, Shonagh Munro, Michael R. McKeown, Yuchuan Wang, Amanda L. Christie, Nathan West, Michael J. Cameron, Brian Schwartz, Tom D. Heightman, Nicholas La Thangue, Christopher A. French, Olaf Wiest, Andrew L. Kung, Stefan Knapp, and James E. Bradner. Selective inhibition of BET bromodomains. *Nature*, 468(7327):1067–1073, 2010.
- [180] Yingchao Duan, Yuanyuan Guan, Wenping Qin, Xiaoyu Zhai, Bin Yu, and Hongmin Liu. Targeting Brd4 for cancer therapy: inhibitors and degraders. *MedChemComm*, pages 1779–1802, 2018.
- [181] Dnyande B. Jarhad, Karishma K. Mashelkar, Hong Rae Kim, Minsoo Noh, and Lak Shin Jeong. Dual Specificity Tyrosine-Phosphorylation-Regulated Kinase 1a (DYRK1a) Kinase Inhibitors as Potential Therapeutics. *Journal of Medicinal Chemistry*, 61:9791–9810, 2018.
- [182] Jun Lan Yu, Tian Tian Chen, Chen Zhou, Fu Lin Lian, Xu Long Tang, Yi Wen, Jing Kang Shen, Ye Chun Xu, Bing Xiong, and Nai Xia Zhang. NMR-based platform for fragment-based lead discovery used in screening BRD4-targeted compounds. *Acta Pharmacologica Sinica*, 37(7):984–993, 2016.
- [183] Matteo Aldeghi, Gregory A. Ross, Michael J. Bodkin, Jonathan W. Essex, Stefan Knapp, and Philip C. Biggin. Large-scale analysis of water stability in bromodomain binding pockets with grand canonical Monte Carlo. *Communications Chemistry*, 1(1):19, 2018.
- [184] Nicholas G.M. Davies, Helen Browne, Ben Davis, Martin J. Drysdale, Nicolas Foloppe, Stephanie Geoffrey, Ben Gibbons, Terance Hart, Roderick Hubbard, Michael Rugaard Jensen, Howard Mansell, Andrew Massey, Natalia Matassova, Jonathan D. Moore, James Murray, Robert Pratt, Stuart Ray, Alan Robertson, Stephen D. Roughley, Joseph Schoepfer, Kirsten Scriven, Heather Simmonite, Stephen Stokes, Allan Surgenor, Paul Webb, Mike Wood, Lisa Wright, and Paul Brough. Targeting conserved water molecules: Design of 4-aryl-5-cyanopyrrolo[2,3-d] pyrimidine Hsp90 inhibitors using fragment-based screening and structure-based optimization. *Bioorganic and Medicinal Chemistry*, 20(22):6770–6789, 2012.
- [185] Li Xing, Jacquelyn Klug-Mcleod, Brajesh Rai, and Elizabeth A. Lunney. Kinase hinge binding scaffolds and their hydrogen bond patterns. *Bioorganic and Medicinal Chemistry*, 23(19):6520–6527, 2015.
- [186] Leonhard Geist, Moriz Mayer, Xiao Ling Cockcroft, Bernhard Wolkerstorfer, Dirk Kessler, Harald Engelhardt, Darryl B. McConnell, and Robert Konrat. Direct NMR Probing of Hydration Shells of Protein Ligand Interfaces and Its Application to Drug Design. *Journal of Medicinal Chemistry*, 60(21):8708–8715, 2017.
- [187] Melanie L Dart, Thomas Machleidt, Emily Jost, Marie K Schwinn, Matthew B Robers, Ce Shi, Thomas A Kirkland, Michael P Killoran, Jennifer M Wilkinson, James R Hartnett, Kristopher Zimmerman, and Keith V Wood. Homogeneous Assay for Target Engagement Utilizing Bioluminescent Thermal Shift. *ACS medicinal chemistry letters*, 9(6):546–551, apr 2018.
- [188] Roderick E. Hubbard and James B. Murray. *Experiences in fragment-based lead discovery*, volume 493. Elsevier Inc., 1 edition, 2011.

- [189] Anthony R. Bradley, Aude Echaliier, Michael Fairhead, Claire Strain-Damerell, Paul Brennan, Alex N. Bullock, Nicola A. Burgess-Brown, Elisabeth P. Carpenter, Opher Gileadi, Brian D. Marsden, Wen Hwa Lee, Wyatt Yue, Chas Bountra, and Frank Von Delft. The SGC beyond structural genomics: Redefining the role of 3D structures by coupling genomic stratification with fragment-based discovery. *Essays in Biochemistry*, 61(5):495–503, 2017.
- [190] Ellene Mashalidis, Paweł Sledz, Steffen Lang, and Chris Abell. A three-stage biophysical screening cascade for fragment-based drug discovery. *Nature Protocols*, 8(11):2309–2324, 2013.
- [191] Thomas G. Davies and Ian J. Tickle. Fragment screening using X-ray crystallography. In *Fragment-Based Drug Discovery and X-Ray Crystallography.*, volume 317, pages 33–59. -, 2011.
- [192] Rob L M van Montfort and Paul Workman. Structure-based drug design: aiming for a perfect fit. *Essays in biochemistry*, 61(5):431–437, 2017.
- [193] Harren Jhoti, Glyn Williams, David C. Rees, and Christopher W. Murray. The 'rule of three' for fragment-based drug discovery: where are we now? *Nature Reviews Drug Discovery*, 12(8):644–644, 2013.
- [194] Monya Baker. Fragment-based lead discovery grows up, NRDD 2013. -, 12(January):5–7, 2013.
- [195] Diane Joseph-McCarthy, Arthur J. Campbell, Gunther Kern, and Demetri Moustakas. Fragment-based lead discovery and design. *Journal of Chemical Information and Modeling*, 54(3):693–704, 2014.
- [196] Daniel A Erlanson, Ben J Davis, Wolfgang Jahnke, and George Box. Perspective Fragment-Based Drug Discovery: Advancing Fragments in the Absence of Crystal Structures. *Cell Chemical Biology*, -(2018):1–7, 2019.
- [197] Magdalena Korczynska, Daniel D. Le, Noah Younger, Elisabet Gregori-Puigjané, Anthony Tumber, Tobias Krojer, Srikanthasan Velupillai, Carina Gileadi, Radosław P. Nowak, Eriko Iwasa, Samuel B. Pollock, Idelisse Ortiz Torres, Udo Oppermann, Brian K. Shoichet, and Danica Galonić Fujimori. Docking and Linking of Fragments to Discover Jumonji Histone Demethylase Inhibitors. *Journal of Medicinal Chemistry*, 59(4):1580–1598, 2016.
- [198] Malgorzata N Drwal, Guillaume Bret, Carlos Perez, Célien Jacquemard, Jérémy Desaphy, and Esther Kellenberger. Structural insights on fragment binding mode conservation. *Journal of Medicinal Chemistry*, page acs.jmedchem.8b00256, 2018.
- [199] Shipra Malhotra and John Karanicolas. When Does Chemical Elaboration Induce a Ligand To Change Its Binding Mode? *Journal of Medicinal Chemistry*, 60(1):128–145, 2017.

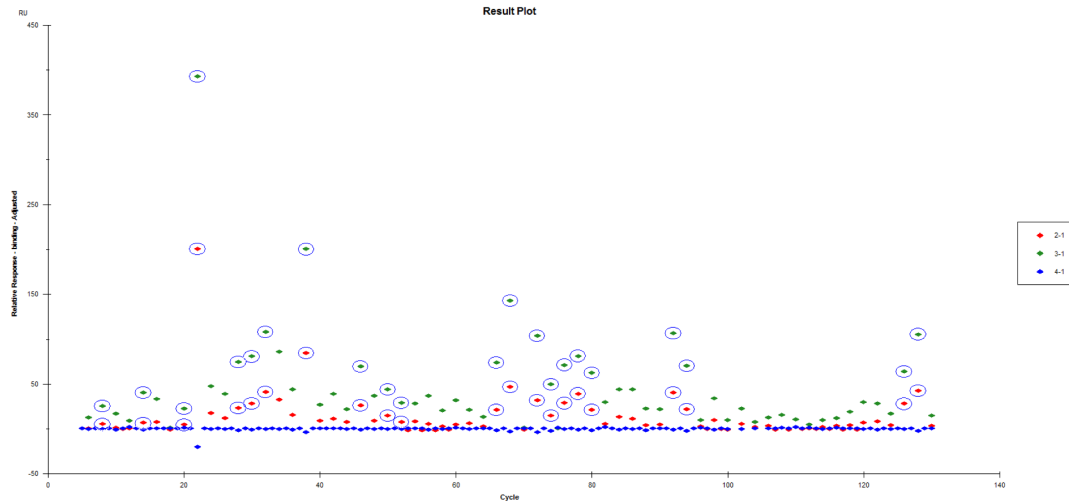
Supplementary Information

The Platform Applied to NUDT21

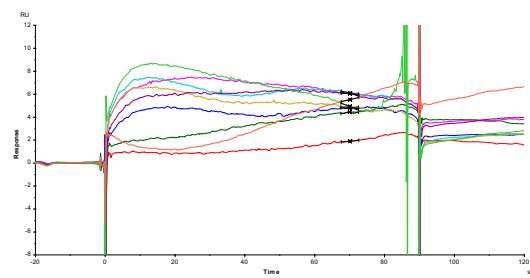
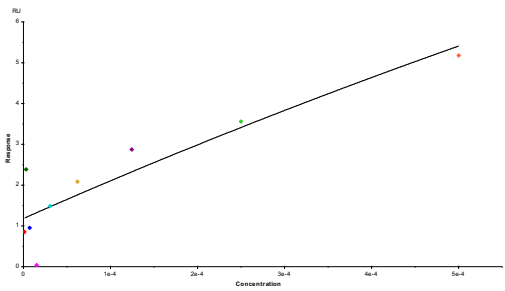
Selected compounds from poised strategy.

rDock S^{inter} Score: -25.016rDock S^{inter} Score: -22.014rDock S^{inter} Score: -24.135rDock S^{inter} Score: -22.776rDock S^{inter} Score: -21.654rDock S^{inter} Score: -25.373rDock S^{inter} Score: -28.498

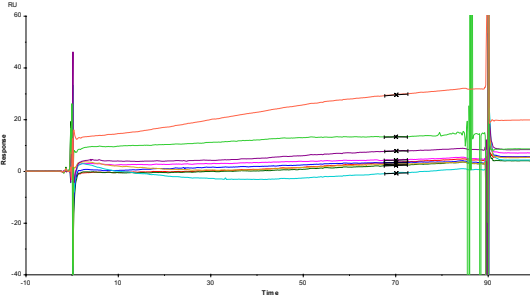
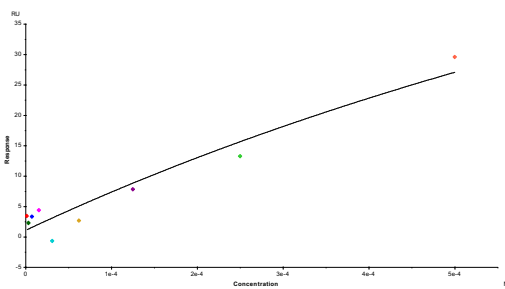
SPR NUDT21



X-0401



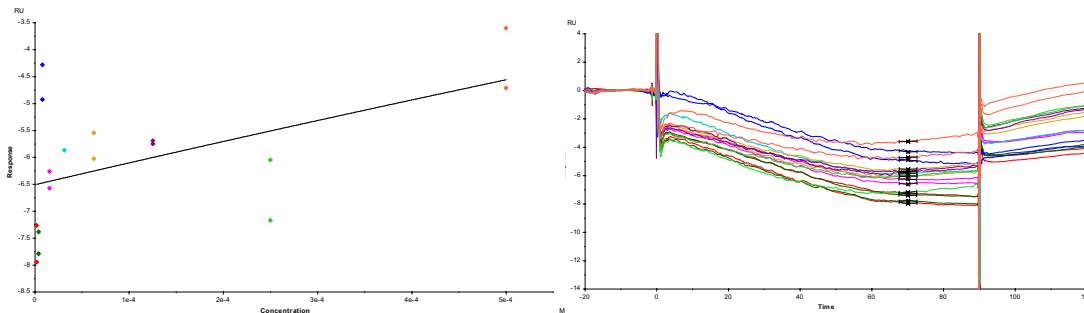
KD (M) 0.004229 Rmax (RU) 40.00 offset (RU) 1.181 Chi² (RU²) 0.527



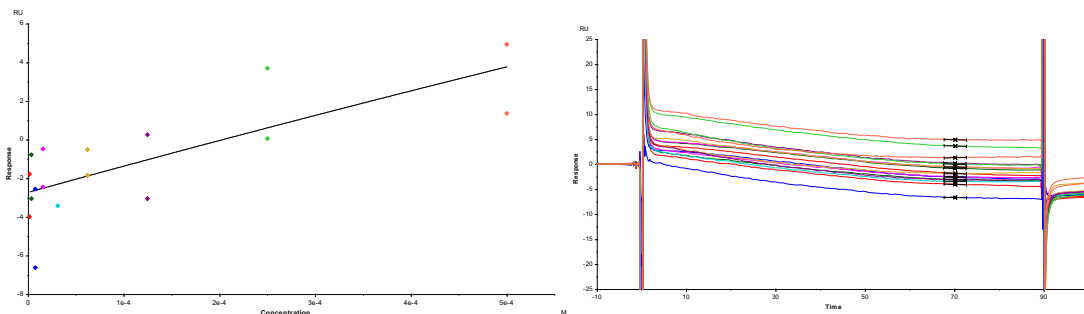
KD (M) 0.001807 Rmax (RU) 120.0 offset (RU) 1.104 Chi² (RU²) 6.75

SPR NUDT21

X-0581

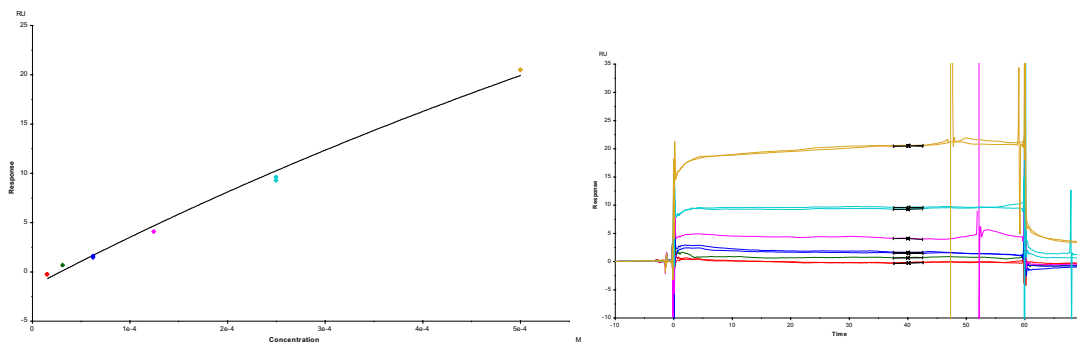
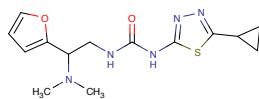


KD (M) 0.009764 Rmax (RU) 40.00 offset (RU) -6.506 Chi² (RU²) 1.16



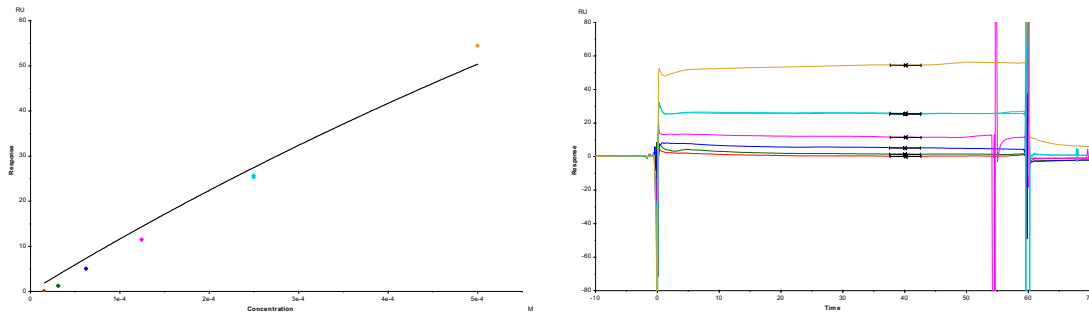
KD (M) 0.008737 Rmax (RU) 120.0 offset (RU) -2.705 Chi² (RU²) 3.48

B1



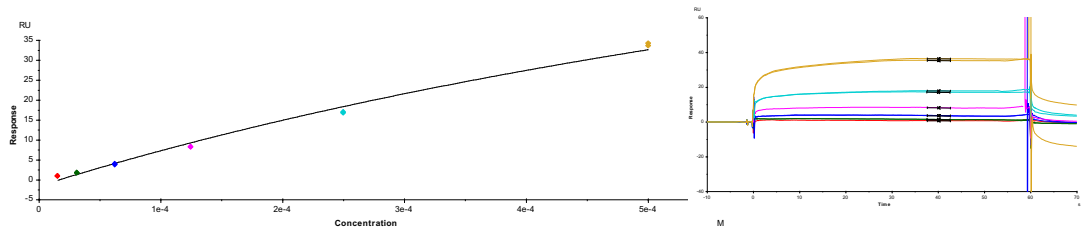
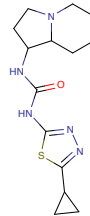
KD (M) 0.002303 Rmax (RU) 120.0 offset (RU) -1.475 Chi² (RU²) 0.408

SPR NUDT21

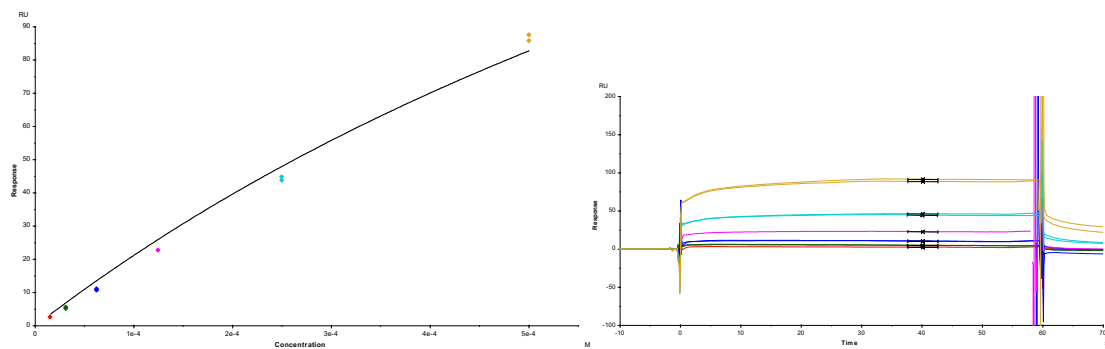


KD (M) 0.002437 Rmax (RU) 300.0 offset (RU) 0.000 Chi² (RU²) 8.30

B2



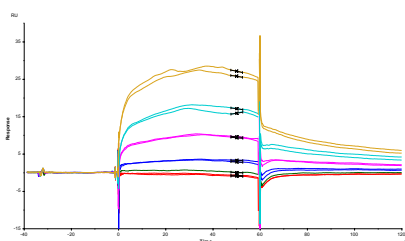
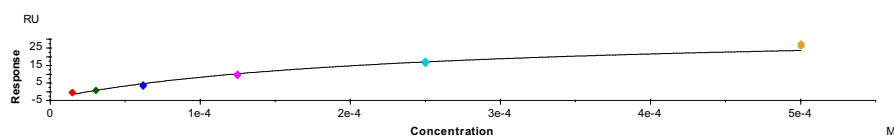
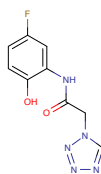
KD (M) 0.001292 Rmax (RU) 120.0 offset (RU) 0.000 Chi² (RU²) 35.0



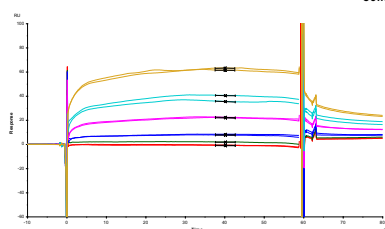
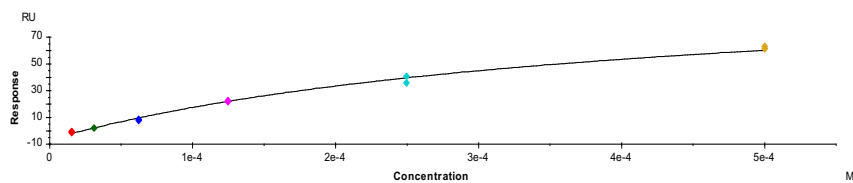
KD (M) 0.001313 Rmax (RU) 300.0 offset (RU) 0.000 Chi² (RU²) 10.0

SPR NUDT21

B9



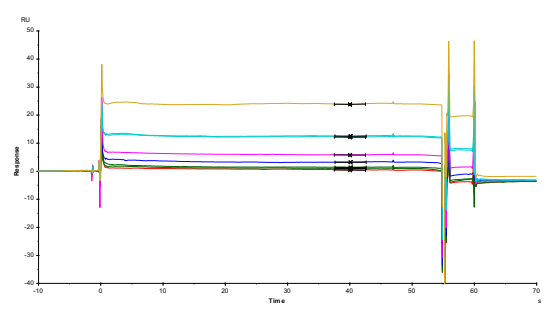
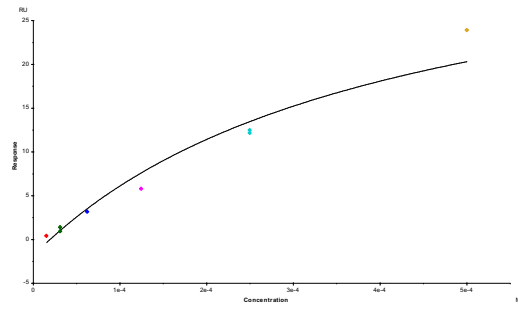
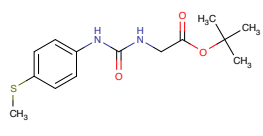
KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
2.252E-4	40.00	-3.978	2.98



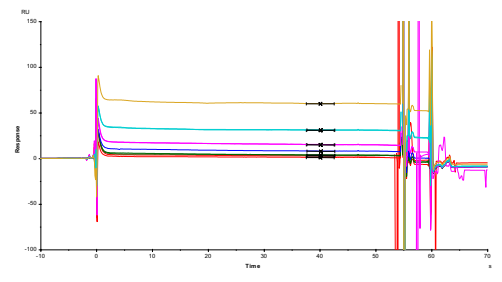
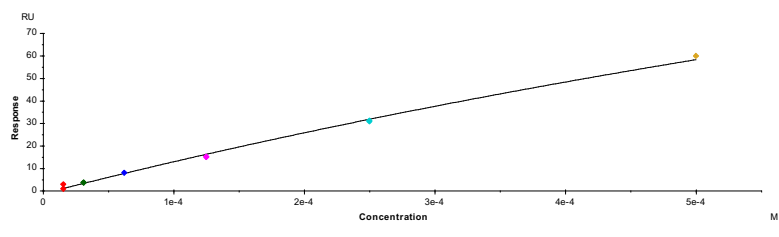
KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
3.997E-4	120.0	-6.508	4.19

SPR NUDT21

C7



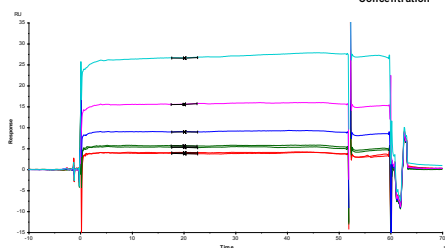
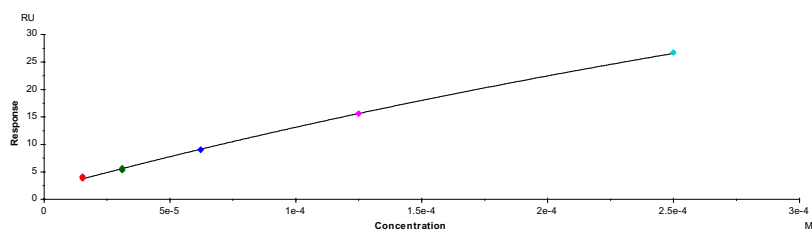
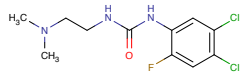
KD (M)	Rmax (RU)	offset (RU)	Chi ²
8.096E-6	120.0	-85.88	178



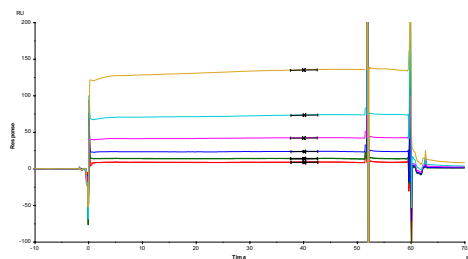
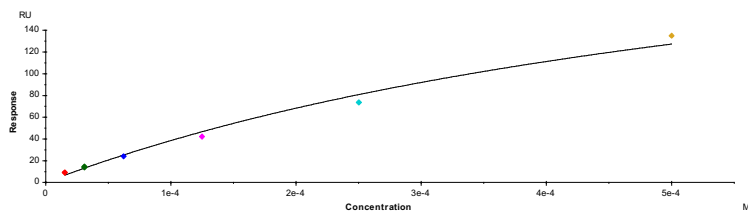
KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
0.002019	300.0	-1.126	1.22

SPR NUDT21

C8



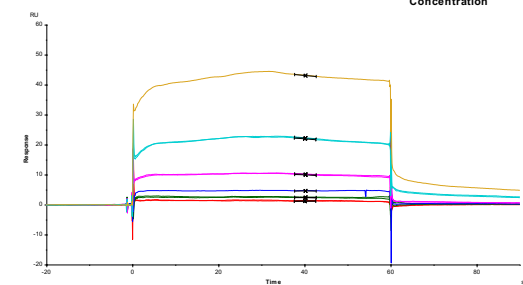
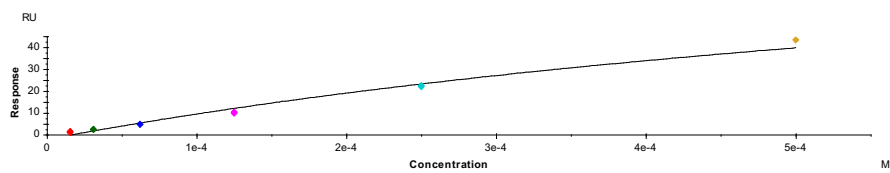
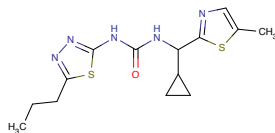
KD (M)	Rmax (RU)	offset (RU)	Chi ²
9.632E-4	120.0	1.856	0.0429



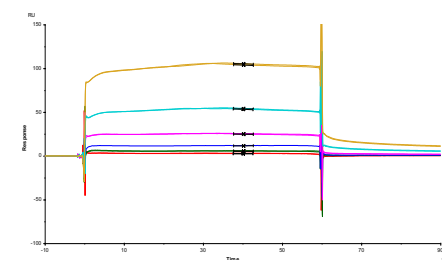
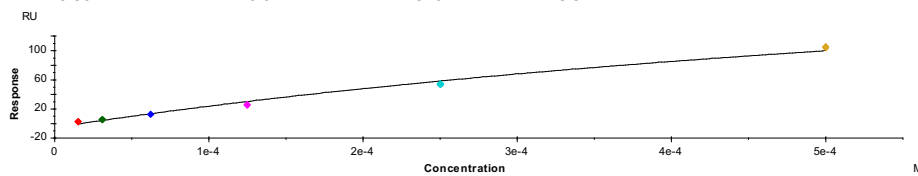
KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
6.778E-4	300.0	0.000	21.2

SPR NUDT21

C10



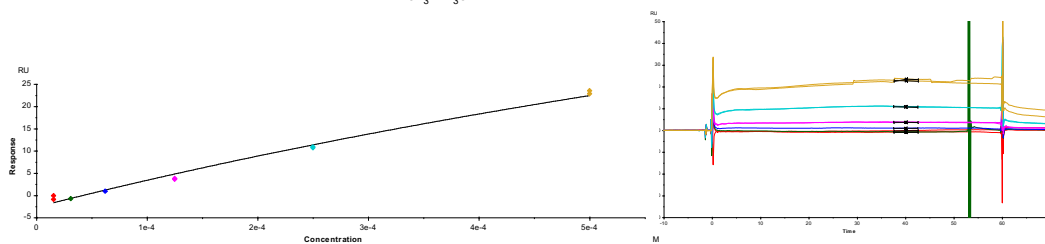
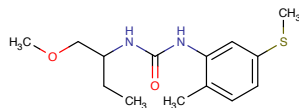
KD (M) 9.364E-4 Rmax (RU) 120.0 offset (RU) -1.946 Chi² (RU²) 3.37



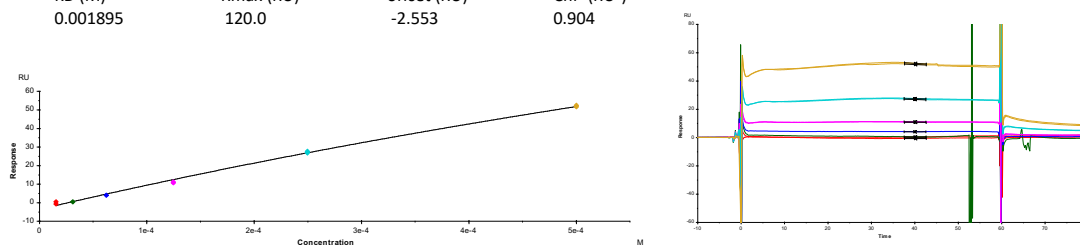
KD (M) 9.273E-4 Rmax (RU) 300.0 offset (RU) -5.486 Chi² (RU²) 18.8

SPR NUDT21

D2

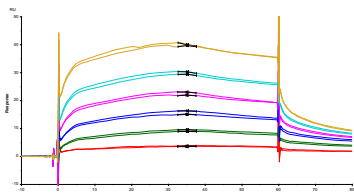
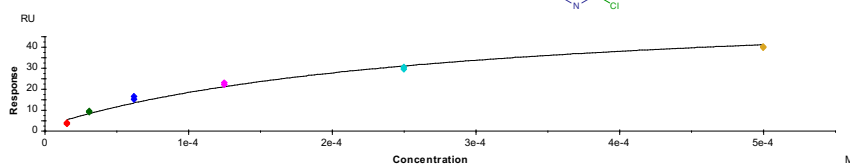
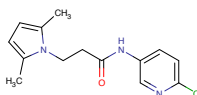


KD (M) Rmax (RU) offset (RU) Chi² (RU²)
 0.001895 120.0 -2.553 0.904



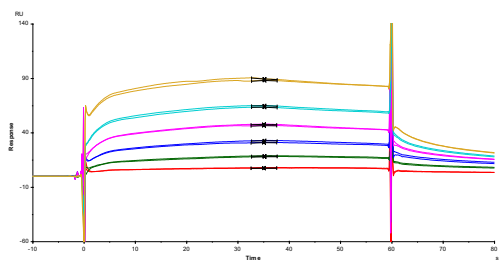
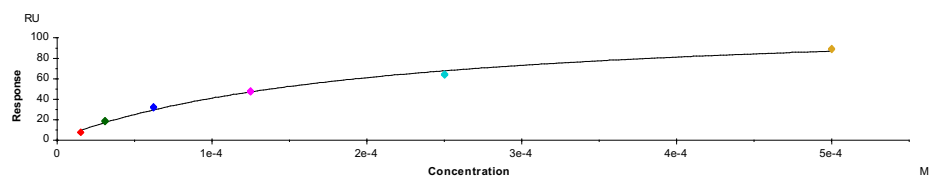
KD (M) Rmax (RU) offset (RU) Chi² (RU²)
 0.002213 300.0 -3.564 1.22

D8



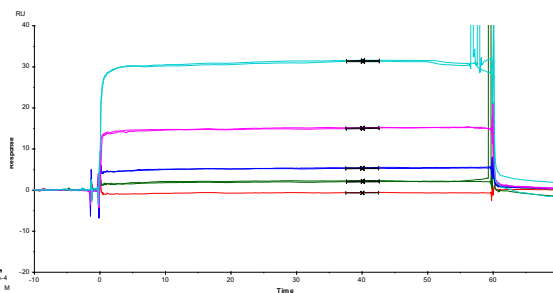
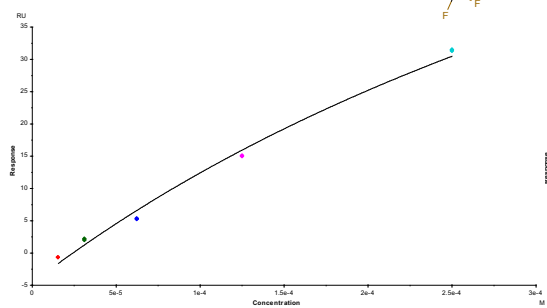
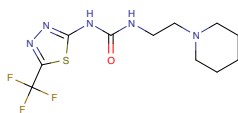
KD (M) Rmax (RU) offset (RU) Chi² (RU²)
 2.714E-4 60.00 2.229 2.96

SPR NUDT21

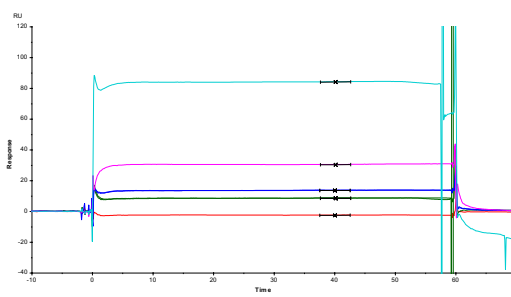
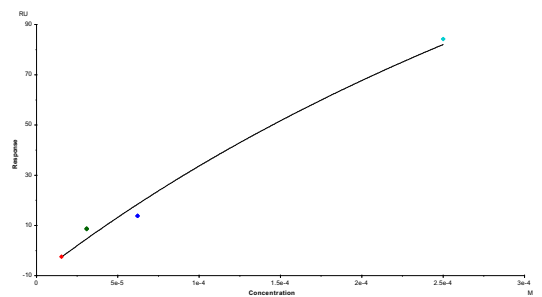


KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
2.015E-4	120.0	1.157	5.95

A2



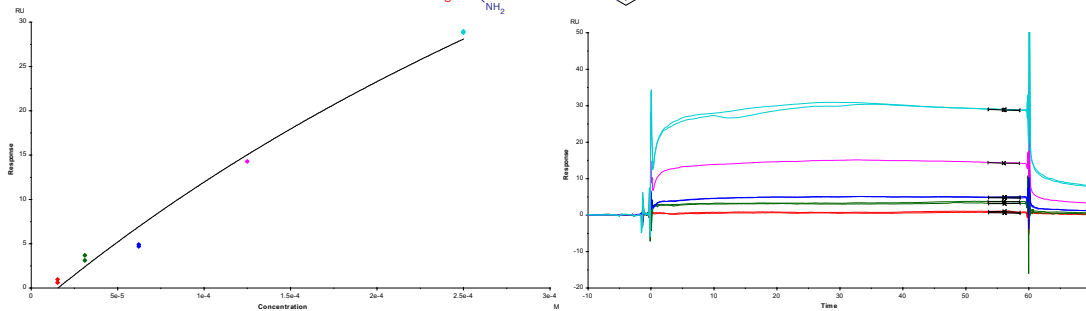
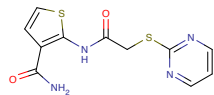
KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
6.044E-4	120.0	-4.617	1.29



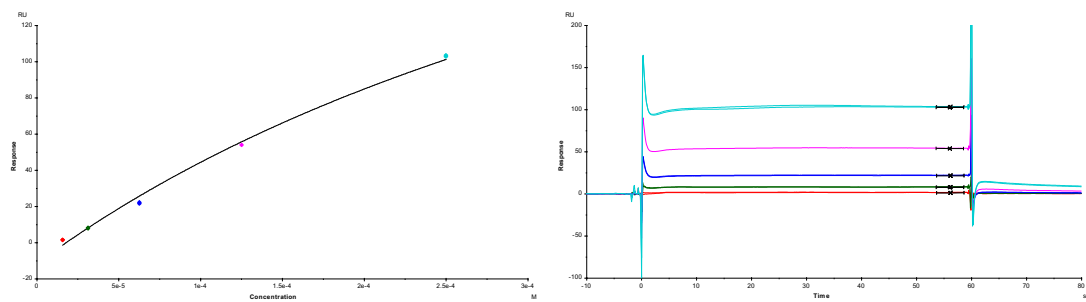
KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
7.000E-4	350.0	-10.11	20.3

SPR NUDT21

A5

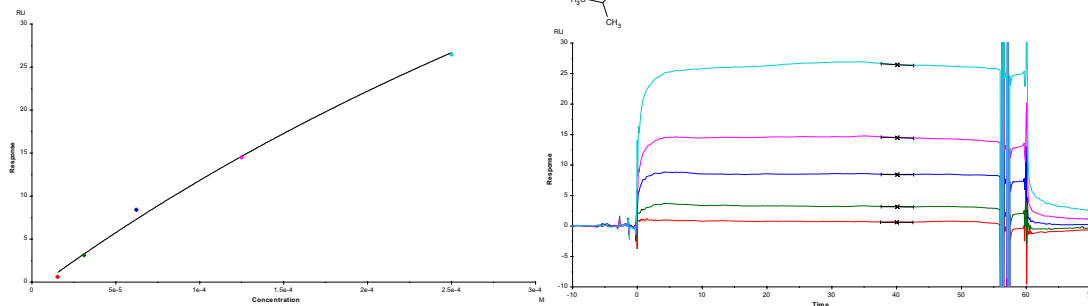
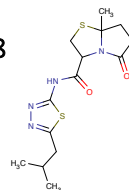


KD (M) 7.307E-4 Rmax (RU) 120.0 offset (RU) -2.485 Chi² (RU²) 2.05



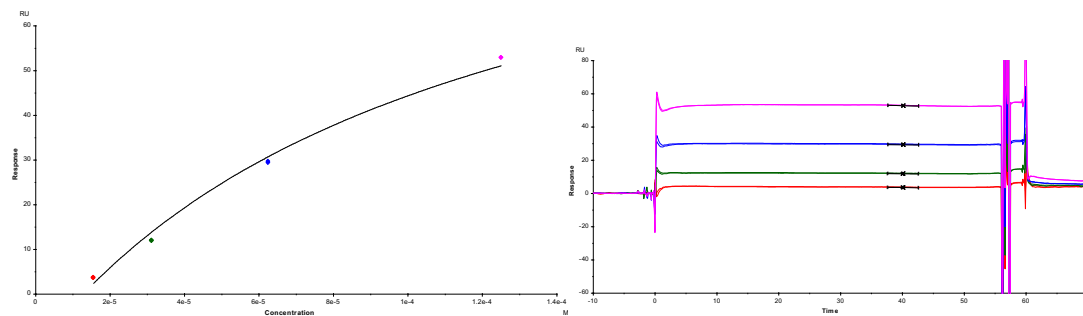
KD (M) 5.267E-4 Rmax (RU) 350.0 offset (RU) -11.42 Chi² (RU²) 8.09

A8



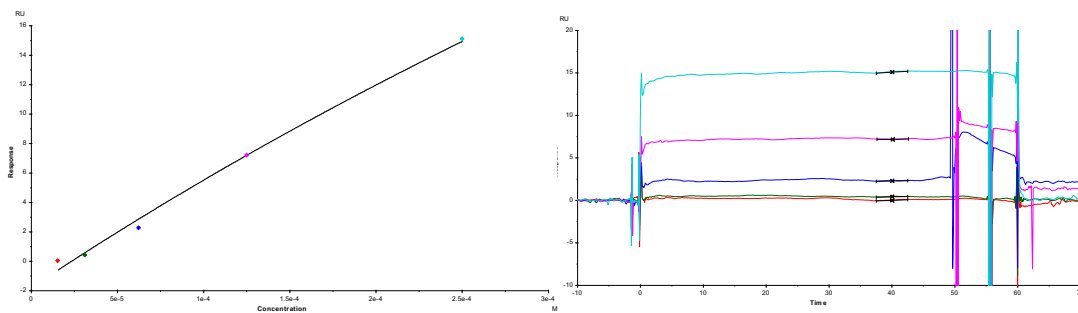
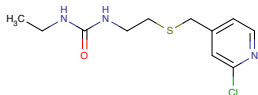
KD (M) 8.319E-4 Rmax (RU) 120.0 offset (RU) -1.043 Chi² (RU²) 0.526

SPR NUDT21

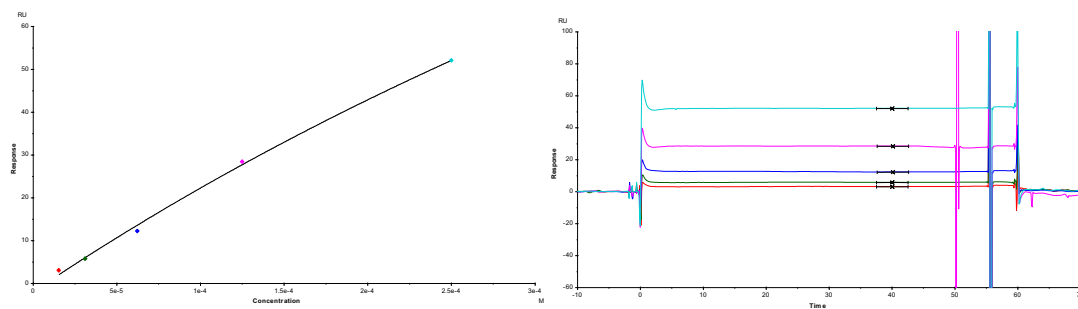


KD (M) 5.974E-4 Rmax (RU) 350.0 offset (RU) -13.41 Chi² (RU²) 2.99

B11



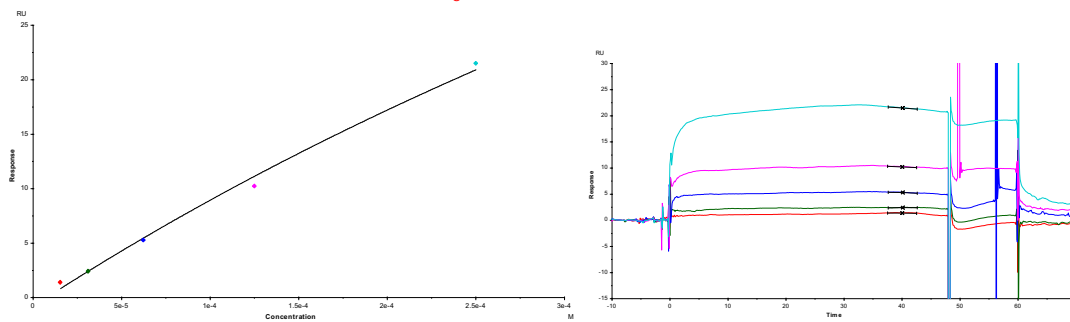
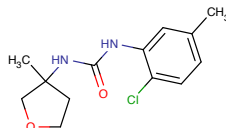
KD (M) 0.001544 Rmax (RU) 120.0 offset (RU) -1.784 Chi² (RU²) 0.260



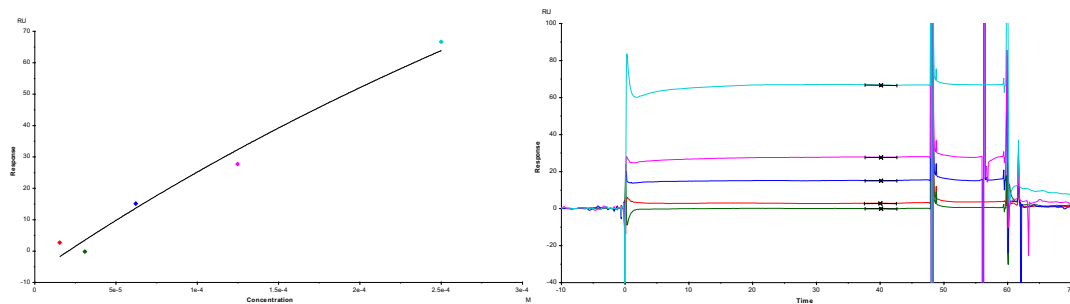
KD (M) 0.001139 Rmax (RU) 300.0 offset (RU) -1.926 Chi² (RU²) 1.21

SPR NUDT21

B12

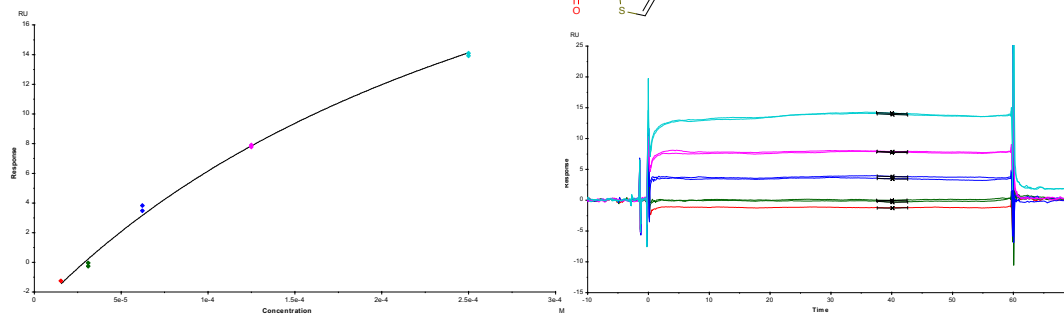
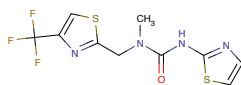


KD (M) 0.001132 Rmax (RU) 120.0 offset (RU) -0.7944 Chi² (RU²) 0.510



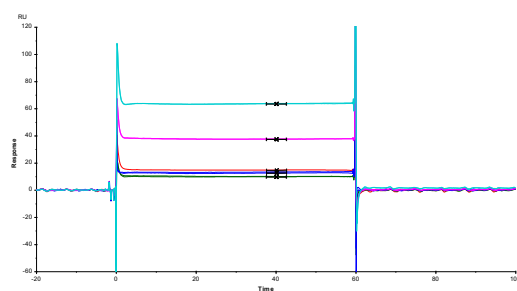
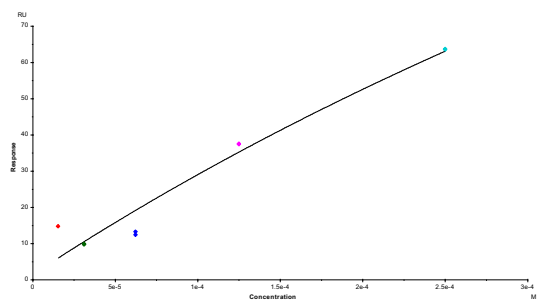
KD (M) 9.814E-4 Rmax (RU) 350.0 offset (RU) -7.175 Chi² (RU²) 21.7

C11



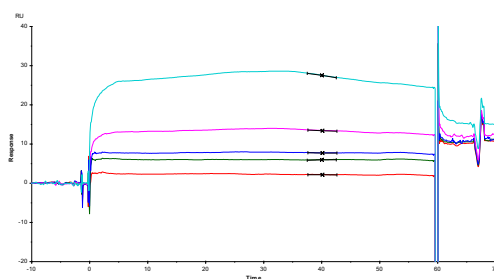
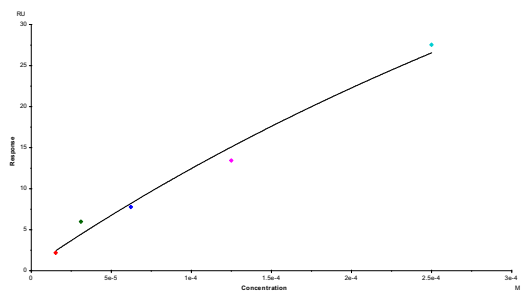
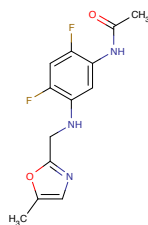
KD (M) 3.253E-4 Rmax (RU) 40.00 offset (RU) -3.252 Chi² (RU²) 0.134

SPR NUDT21

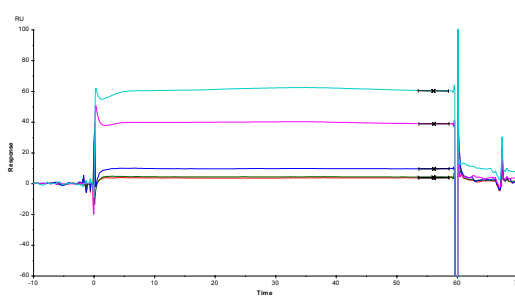
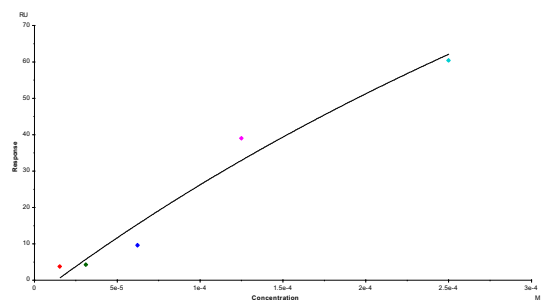


KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
0.001169	350.0	1.484	23.7

C12



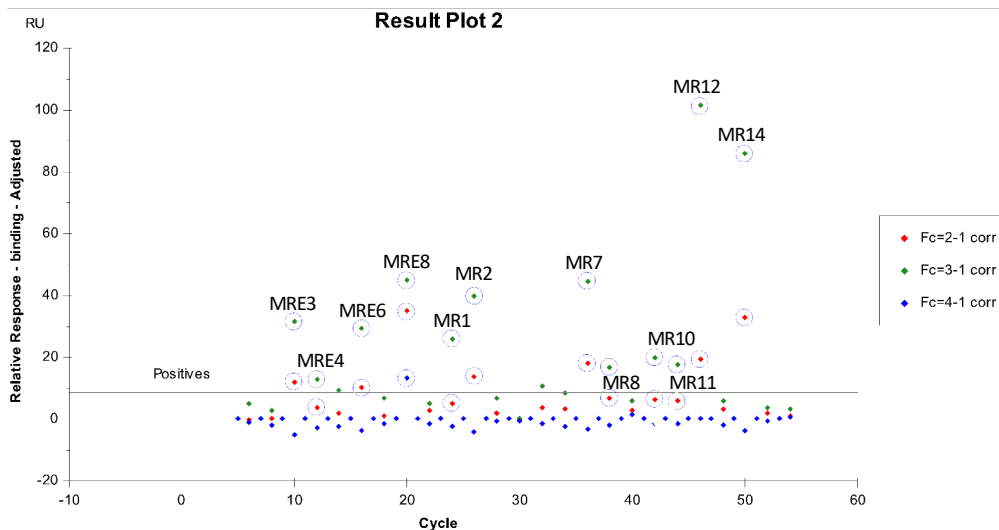
KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
8.983E-4	120.0	0.4304	2.09



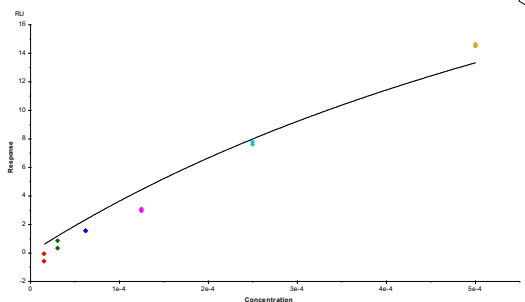
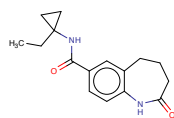
KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
8.755E-4	300.0	-4.530	28.0

Exploratory Strategies

SPR Exploratory Strategies



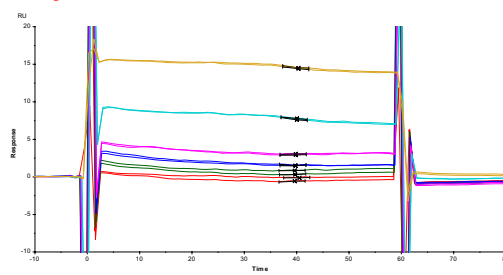
MRE3



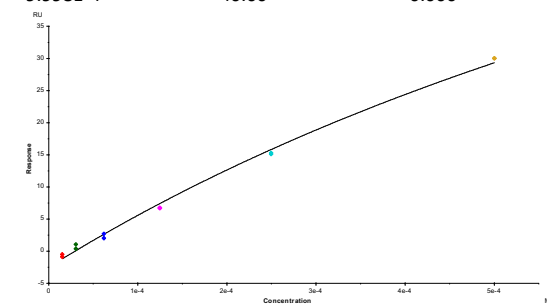
KD (M)
9.998E-4

Rmax (RU)
40.00

offset (RU)
0.000



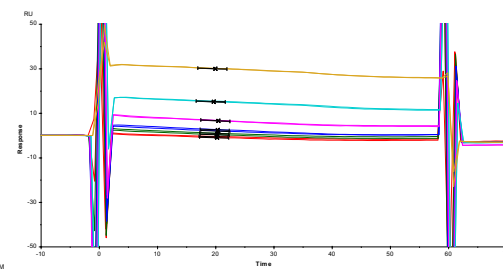
Chi² (RU²)
1.04



KD (M)
0.001386

Rmax (RU)
120.0

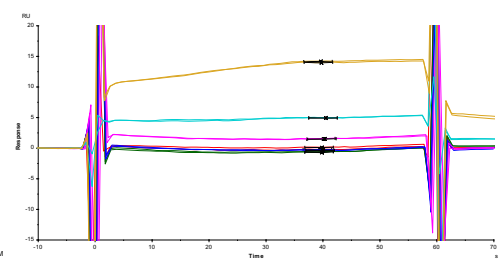
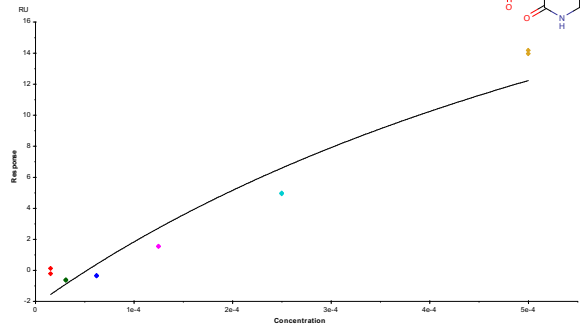
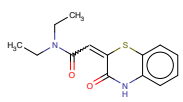
offset (RU)
-2.495



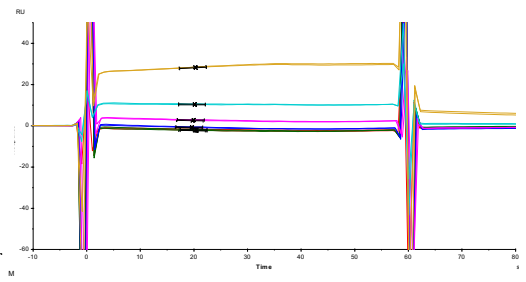
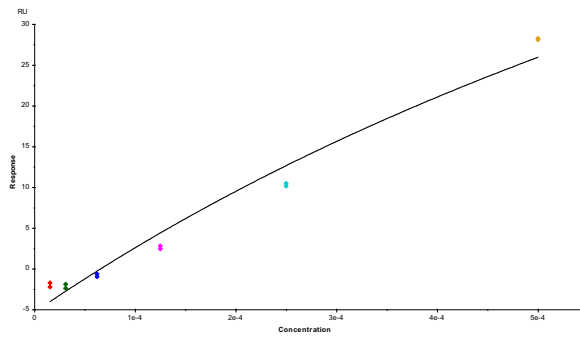
Chi² (RU²)
0.457

SPR Exploratory Strategies

MR2

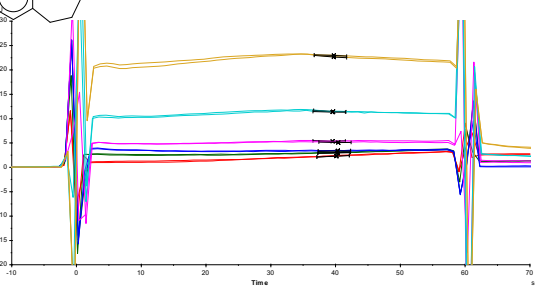
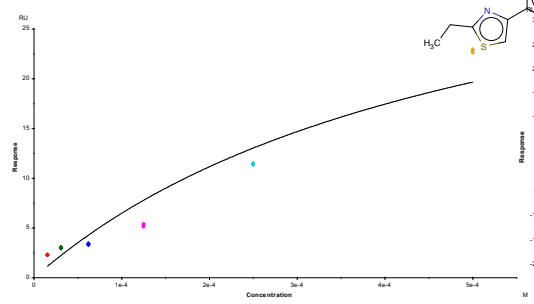
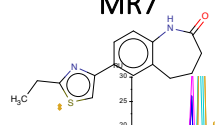


KD (M) 8.819E-4 Rmax (RU) 40.00 offset (RU) -2.234 Chi² (RU²) 2.09



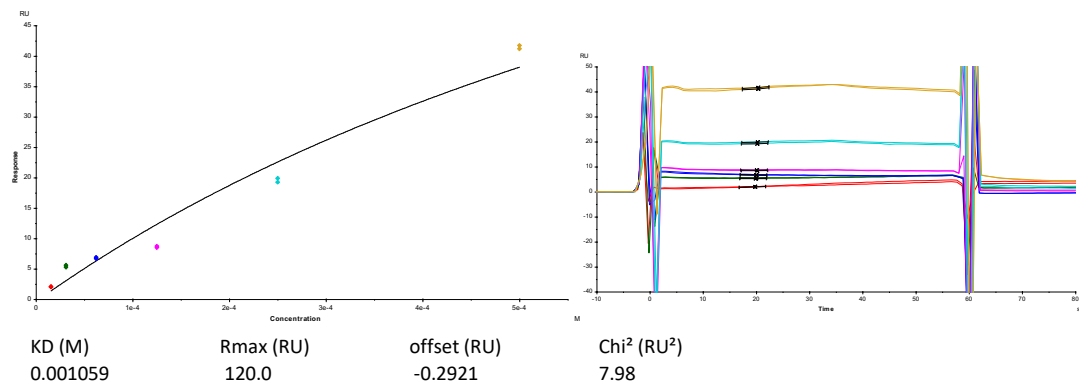
KD (M) 0.001419 Rmax (RU) 120.0 offset (RU) -5.272 Chi² (RU²) 3.68

MR7

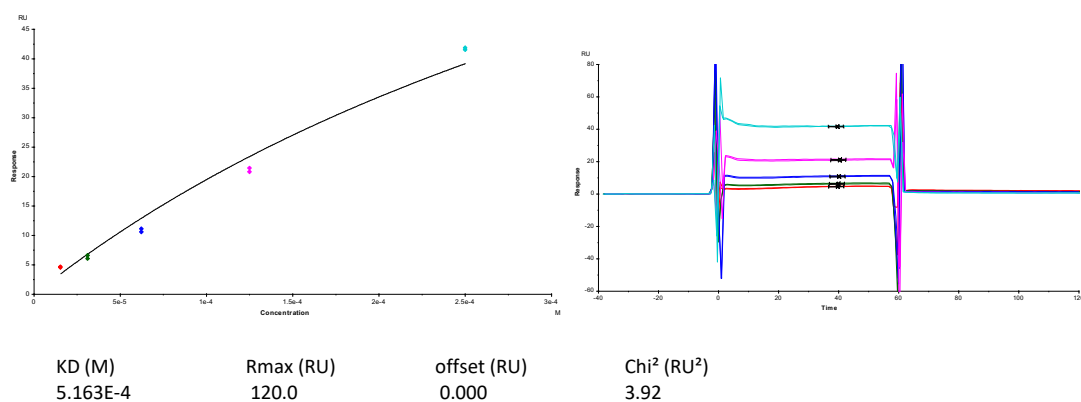
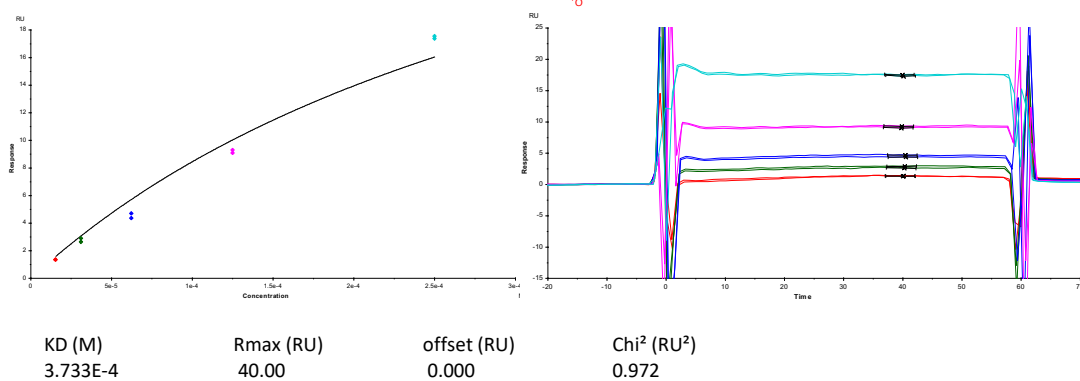
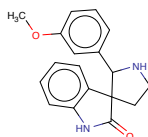


KD (M) 5.163E-4 Rmax (RU) 40.00 offset (RU) 0.000 Chi² (RU²) 3.88

SPR Exploratory Strategies

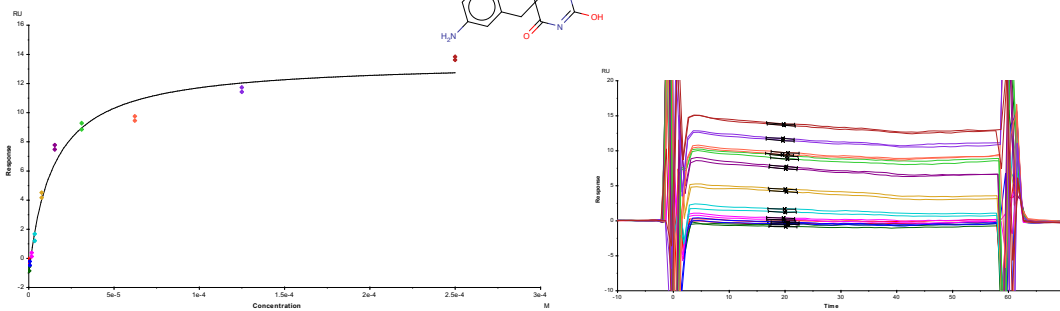
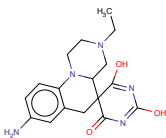


MR14

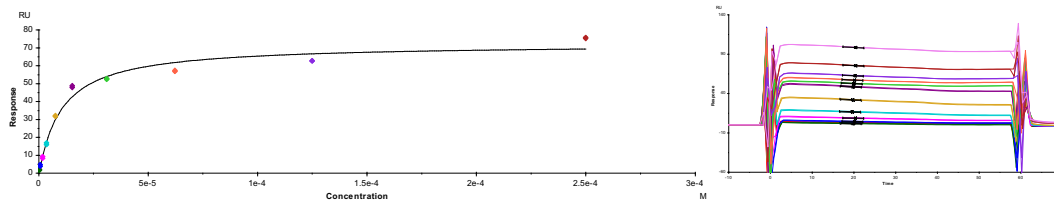


SPR Exploratory Strategies

MR12

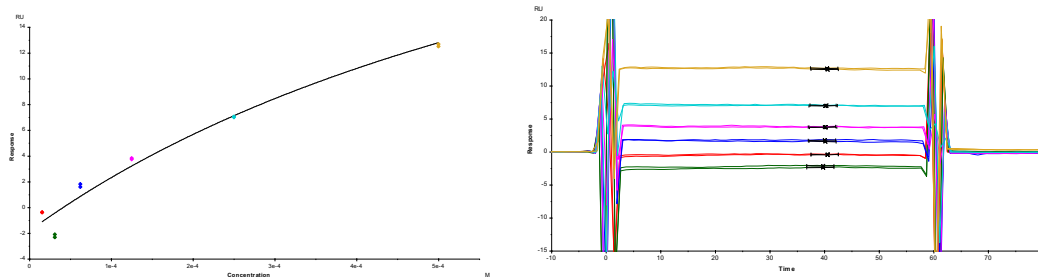
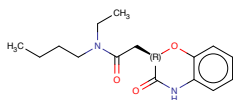


KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
1.409E-5	14.63	-1.114	0.539



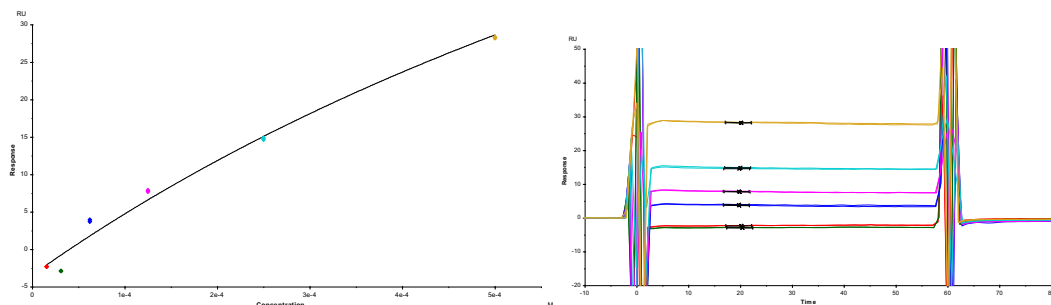
KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
1.031E-5	73.53	-1.240	13.2

MRE6



KD (M)	Rmax (RU)	offset (RU)	Chi ² (RU ²)
8.690E-4	40.00	-1.791	0.961

SPR Exploratory Strategies



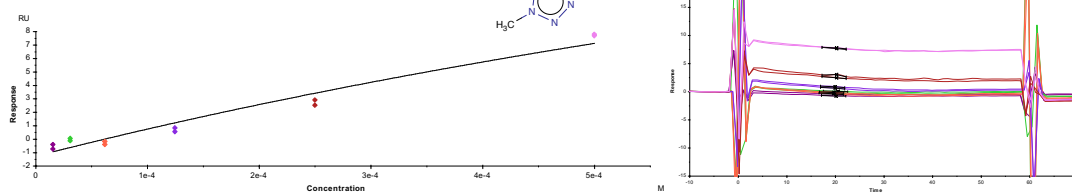
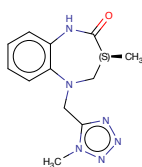
KD (M)
0.001373

Rmax (RU)
120.0

offset (RU)
-3.360

Chi² (RU²)
2.08

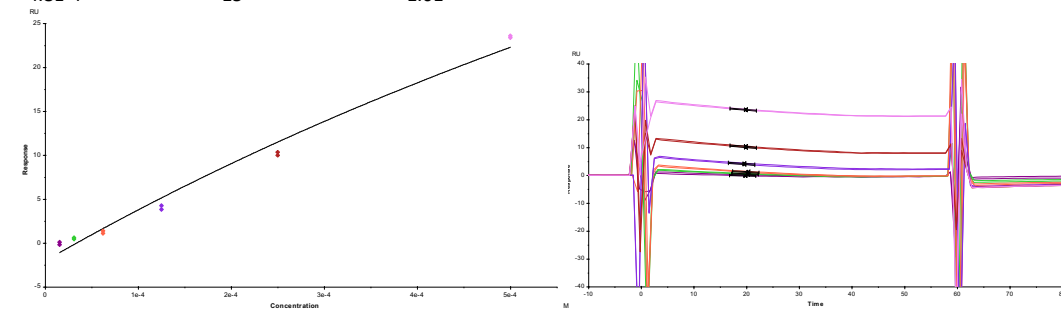
MR1



KD (M)
4.8E-4

Rmax (RU)
15

Chi² (RU²)
1.01



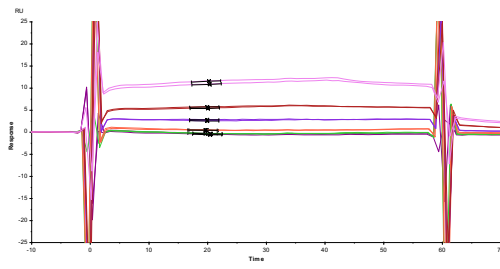
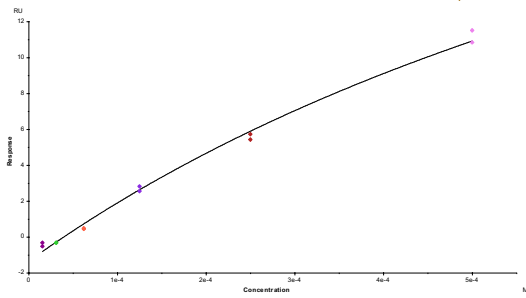
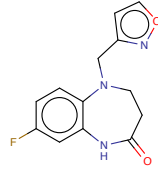
KD (M)
0.001163

Rmax (RU)
80

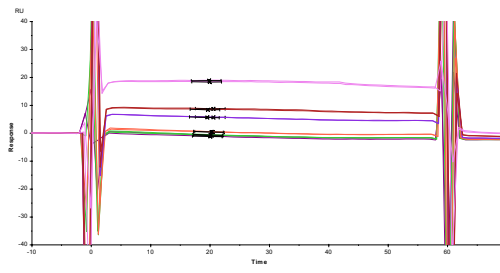
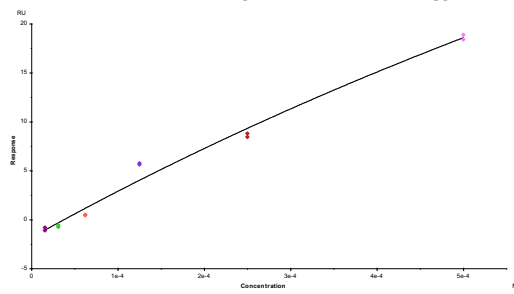
Chi² (RU²)
2.04

SPR Exploratory Strategies

MR10

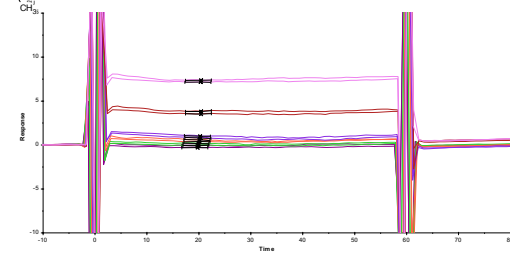
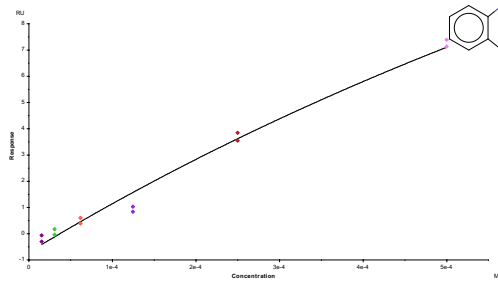
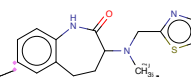


KD (M) 2.24E-4 Rmax (RU) 15 Chi² 1.89



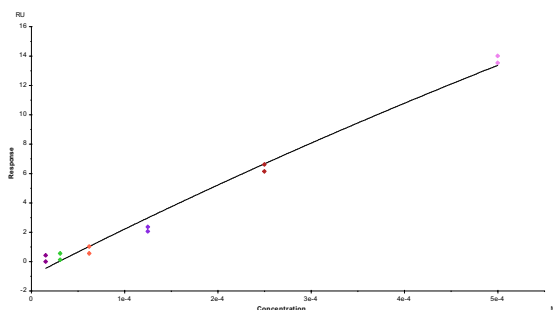
KD (M) 0.001456 Rmax (RU) 80 Chi² (RU²) 0.811

MR11



KD (M) 0.002058 Rmax (RU) 40.00 offset (RU) -0.7094 Chi² (RU²) 0.123

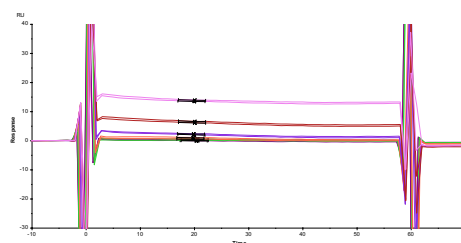
SPR Exploratory Strategies



KD (M)
0.003687

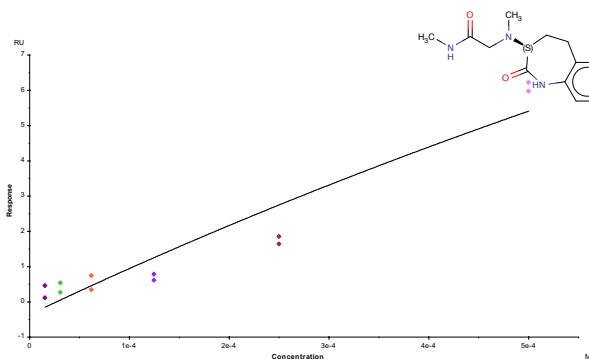
Rmax (RU)
120.0

offset (RU)
-0.9521



Chi² (RU²)
0.337

MRE4

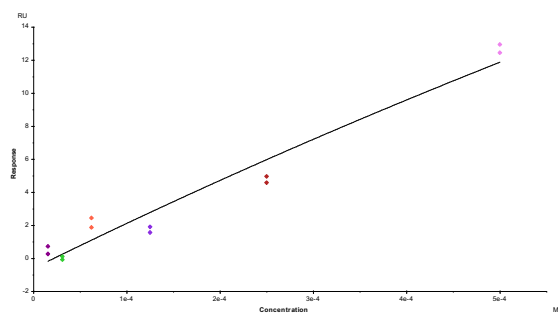
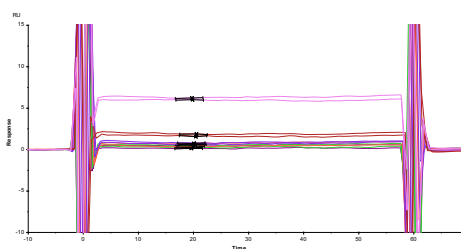


KD (M)
0.002966

Rmax (RU)
40.00

offset (RU)
-0.3587

Chi² (RU²)
0.449

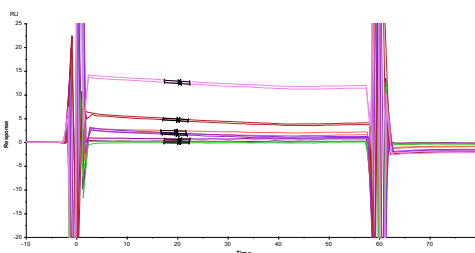


KD (M)
0.004307

Rmax (RU)
120.0

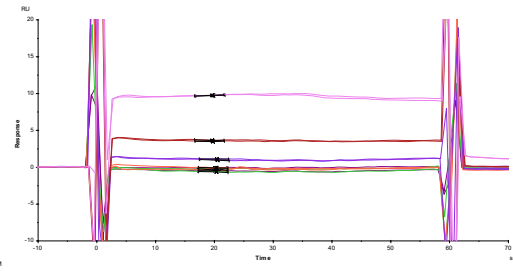
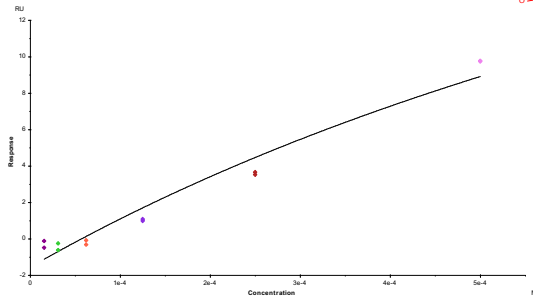
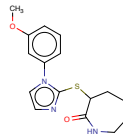
offset (RU)
-0.5938

Chi² (RU²)
1.00

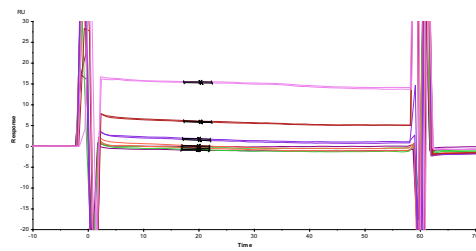
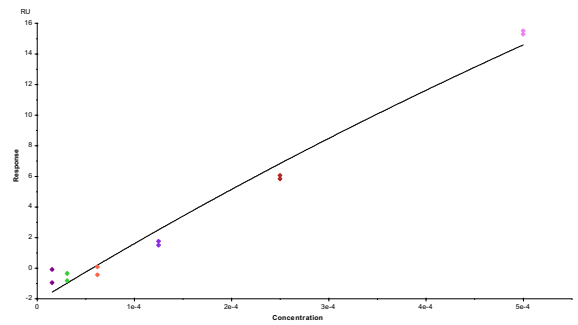


SPR Exploratory Strategies

MR8



KD (M) 0.001411 Rmax (RU) 40.00 offset (RU) -1.543 Chi² 0.577



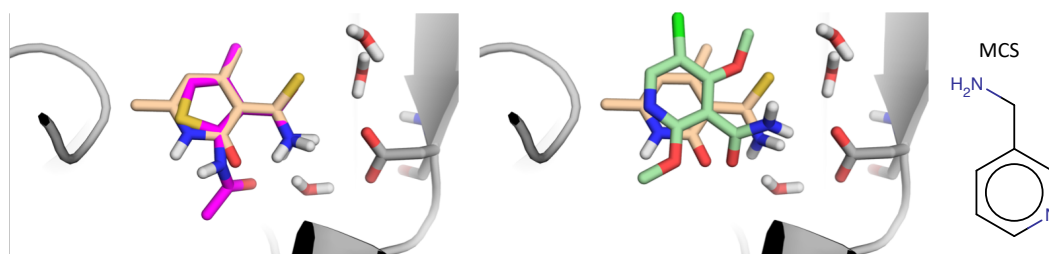
KD (M) 0.003078 Rmax (RU) 120.0 offset (RU) -2.167 Chi² (RU²) 0.787

Blinded Optimization

Structures and chunk residues.

Target	Case	PDB	Resolution	Chunk Residues
BRD4	1	3M49	2.00Å	81-89, 91-94, 97, 101, 104-105, 131, 135-137, 139-140, 144-146, 149, and the four conserved waters
HSP90	2	3HHU	1.56Å	47-58, 78, 91-98, 137-138, 150-152, 183-186, and the three conserved waters
HSP90	3	4FCQ	15Å	47-58, 78, 91-98, 137-138, 150-152, 183-186 and the three conserved waters
DYRK1A	4 (D1)	5A4T	2.15Å	165, 173-175, 184-187, 219-225, 237-247, 290-308
DYRK1A	4 (D2)	2WO6	2.5Å	166, 173-175, 184-188, 219-225, 237-247, 290-308

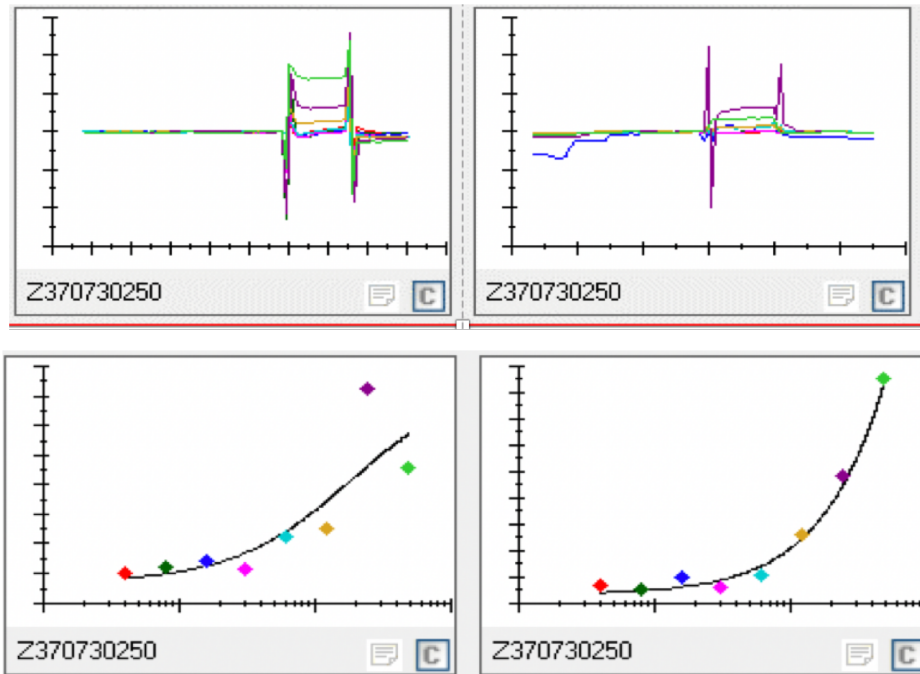
Binding mode flip Case Study 2



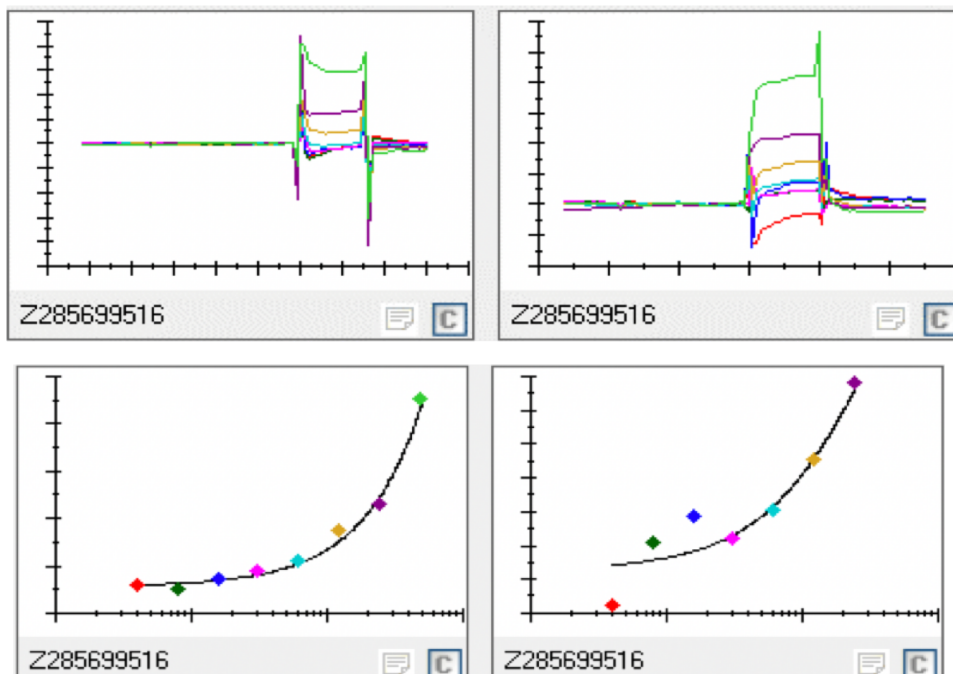
H1-1/great grandparent (magenta) superposed with grandparent (light pink) on the left. Grandparent (light pink) superposed with parent that has a flipped amide on the right. The corresponding MCS is shown, for which the RMSD was 0.025. The tethered RMSD cutoff was increased otherwise no hits were found.

SPR HSP90 Fragment 1

12

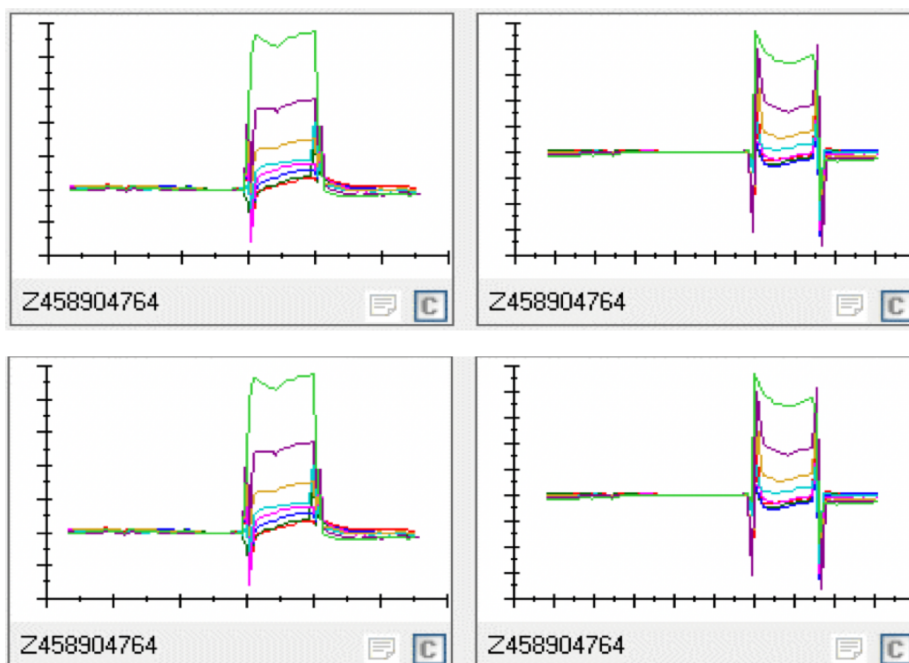


13

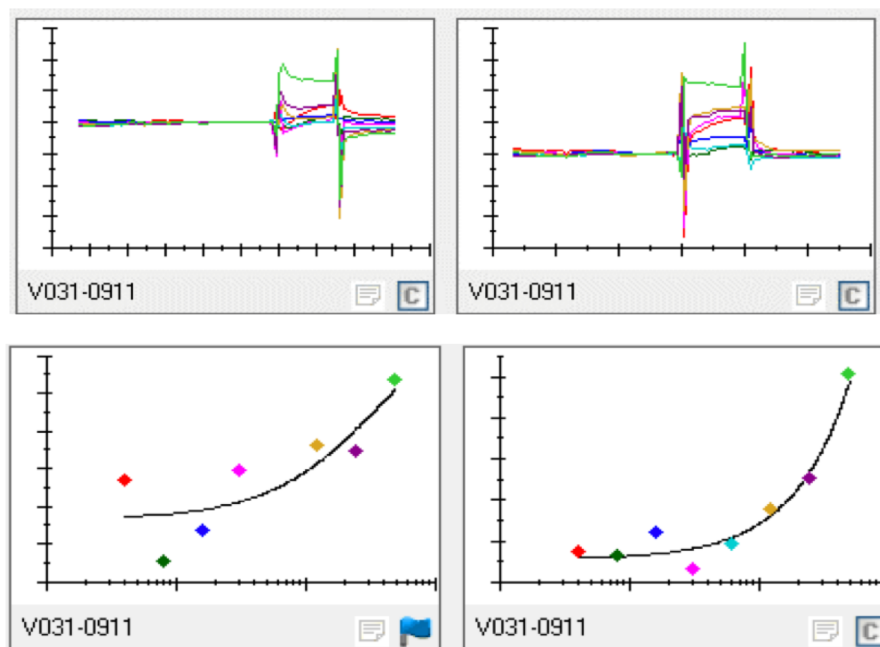


SPR HSP90 Fragment 1

14

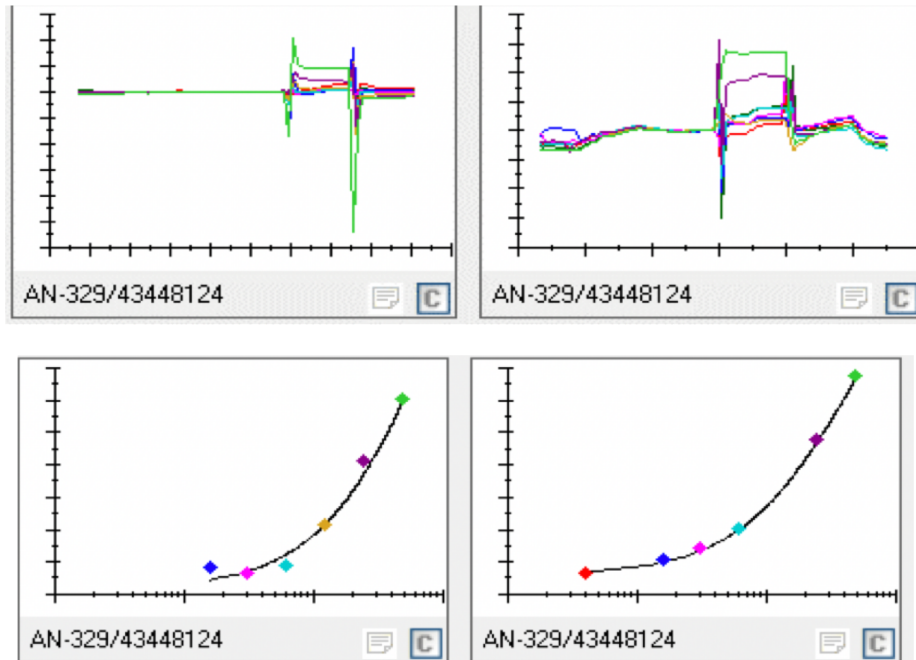


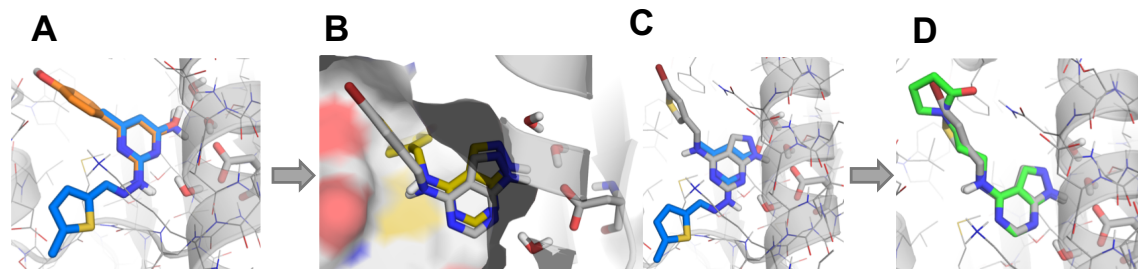
15



SPR HSP90 Fragment 1

16

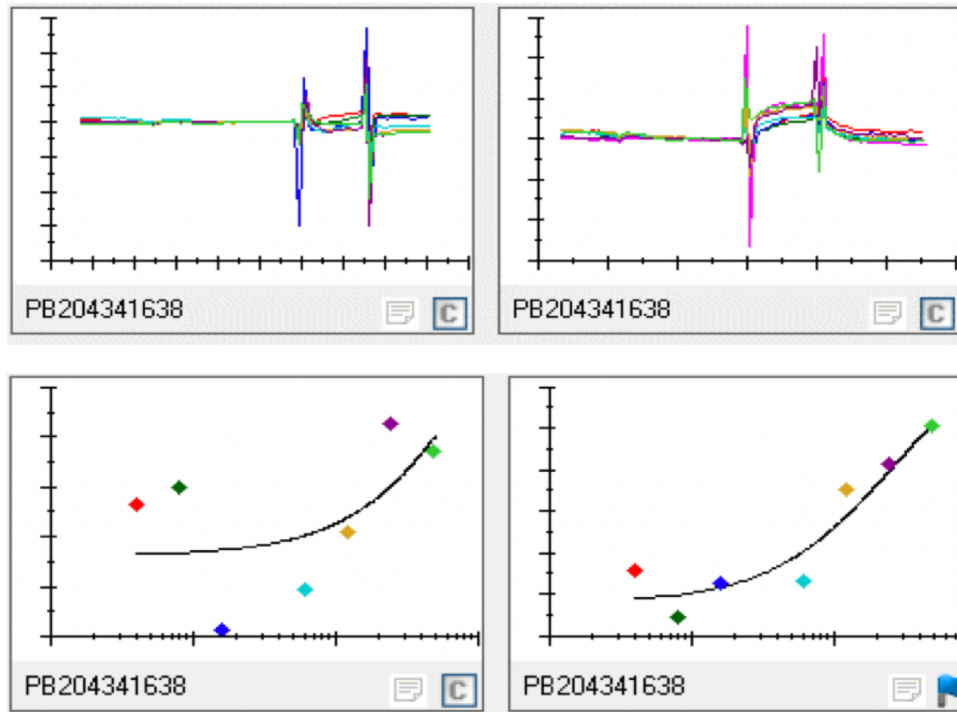


Binding Mode Shift Case Study 3 – HSP90 – Fragment 2

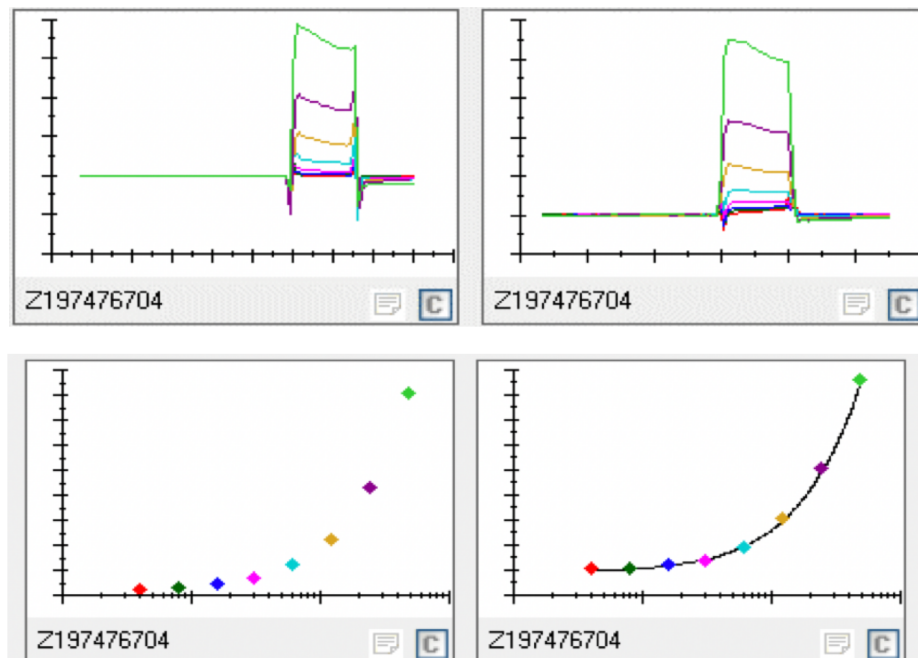
A) H2-1/great-grandparent (orange), with the grandparent ZINC0002452 (blue). B) The parent after docking (grey) and before docking (olive-green). C) The grandparent (blue) with the parent (grey) in which a slight shift is observed. D) The parent (grey) with compound 21 (green).

SPR HSP90 Fragment 2

17

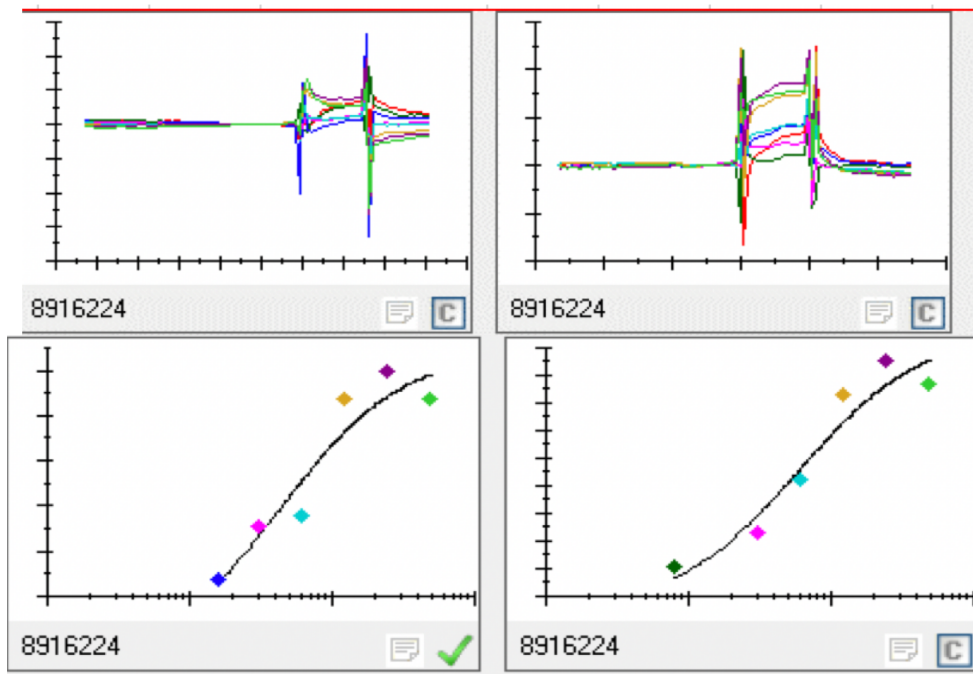


18

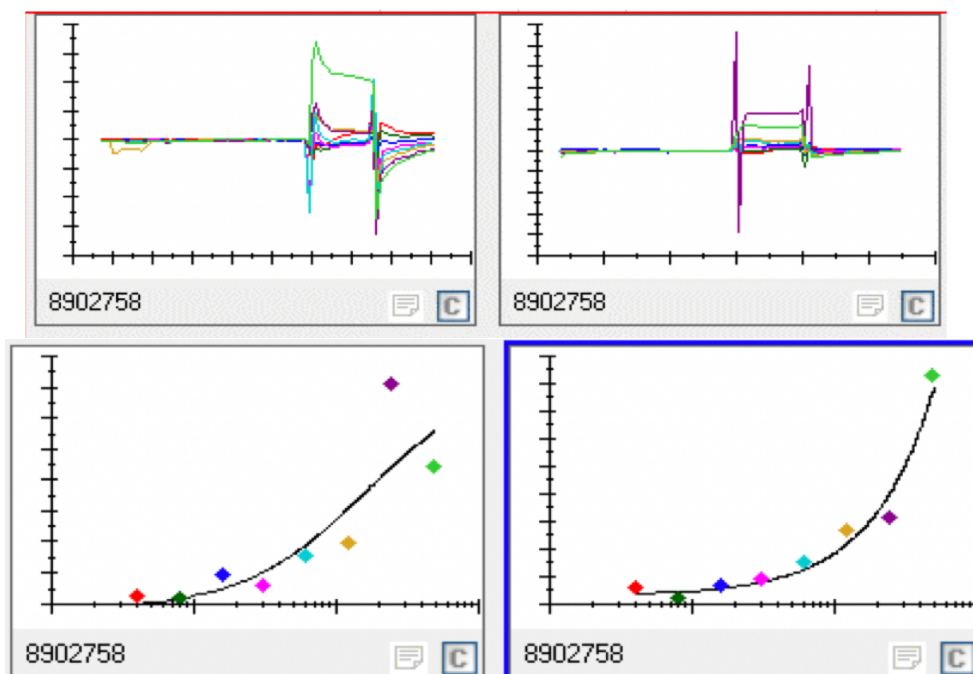


SPR HSP90 Fragment 2

19



20



Additional Publications

Predicting how drug molecules bind to their protein targets

Moira Rachman¹, Xavier Barril^{1,2} and Roderick E Hubbard^{3,4}

In *Current Opinion in Pharmacology* (2018), 42, 34-9

1. Institut de Biomedicina de la Universitat de Barcelona (IBUB) and Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII 27-31, 08028 Barcelona, Spain.
2. Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.
3. YSBL, University of York, Heslington, York YO10 5DD, UK
4. Vernalis (R&D) Ltd, GrantaPark, Abington, Cambridge CB21 6GB, UK



Predicting how drug molecules bind to their protein targets

Moira M Rachman¹, Xavier Barril^{1,2} and Roderick E Hubbard^{3,4}



There have been substantial advances in the application of molecular modelling and simulation to drug discovery in recent years, as massive increases in computer power are coupled with continued development in the underlying methods and understanding of how to apply them. Here, we survey recent advances in one particular area — predicting how a known ligand binds to a particular protein. We focus on the four contributing classes of calculation: predicting where a binding site is on a protein; characterizing where chemical functional groups will bind to that site; molecular docking to generate a binding mode for a ligand and dynamics simulations to refine that pose and allow for protein conformation change. Examples of successful application are provided for each class.

Addresses

¹ Facultat de Farmàcia and Institut de Biomedicina, Universitat de Barcelona, Av. Joan XXIII, 27-31, 08028 Barcelona, Spain

² Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

³ YSBL, University of York, Heslington, York YO10 5DD, UK

⁴ Vernalis (R&D) Ltd, Granta Park, Abingdon, Cambridge CB21 6GB, UK

Corresponding author: Hubbard, Roderick E (roderick.hubbard@york.ac.uk)

Current Opinion in Pharmacology 2018, 42:34–39

This review comes from a themed issue on **New technologies**

Edited by **Antoine Bril, Oliver Nosjean and Patrick Genissel**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 21st July 2018

<https://doi.org/10.1016/j.coph.2018.07.001>

1471-4892/© 2018 Published by Elsevier Ltd.

Introduction

The majority of drug discovery projects begin with identification of a small molecule compound which binds to a defined site on a specific biological molecule (usually a protein), affecting the function of that target protein. This initial hit is then optimized to incorporate adequate drug-like properties (affinity, selectivity, efficacy, ADME, etc.) into a candidate compound that generates the desired therapeutic effect and is suitable for clinical trials.

Over the past thirty years, there has been a steady increase in the use of structure-based methods in this drug discovery process where models of how compounds

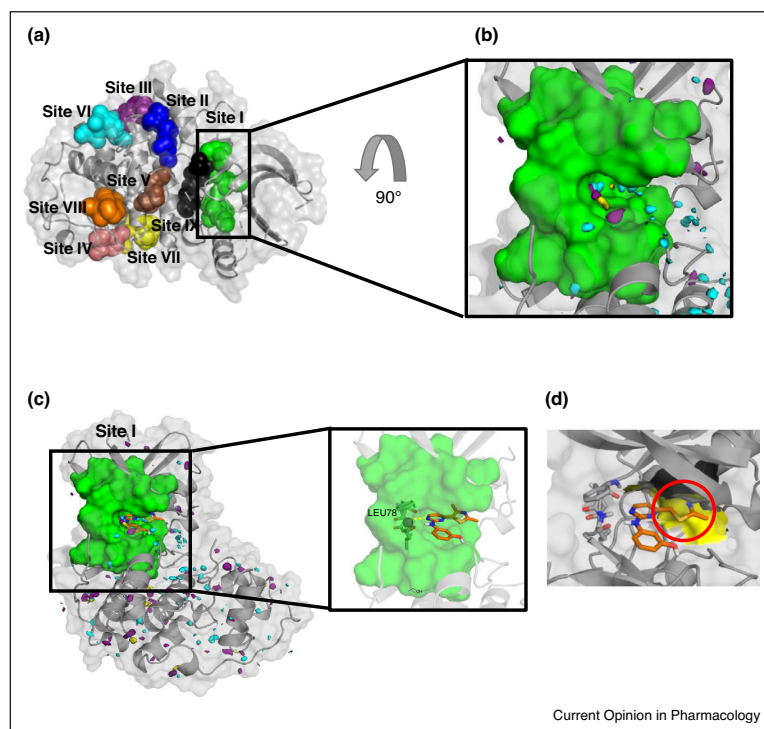
bind to the target can allow rational design of the required improvements in the compounds. For some targets, experimental methods can provide structural information with sufficient throughput and speed to interactively guide the structure-based design. For example, X-ray crystallography provides an atomic level picture of how compounds bind and NMR spectroscopy can provide varying levels of information on interactions between the compound and the protein, such as whether a compound binds, where it is binding to and (in some limited cases) a structure of the compound binding to the target. However, it is often not possible to generate such structures with sufficient speed to inform decisions about compound optimization.

In this review, we survey recent developments in the computational methods that predict how compounds bind to their protein target using either an experimentally determined structure of the target or a model based on sequence homology. Some of the methods can be used to screen compound libraries (real or virtual) for initial hits; in addition, the methods can help to guide optimization of compounds in structure-based design. These applications are not discussed in detail here. What we focus on are the methods that, once a compound is demonstrated to bind, can be used to predict the position and orientation or ‘pose’ of the compound binding. As summarized in [Figure 1](#), we have loosely divided these methods into four categories: (1) identifying binding sites; (2) characterizing the potential of a binding site to bind chemical matter; (3) predicting the position and orientation (or pose) of compound binding and (4) dynamic docking to explore both the energetics of binding and conformational change to refine the pose. Before summarizing these in turn, we first survey some history and the issue that underpins all molecular modelling — the ability to estimate energy of interaction.

Origins of the methods

A more detailed description of the origins of structure-based design methods is provided elsewhere [1] but there are three influential developments that should be highlighted — CHARMM [2], GRID [3] and DOCK [4]. The Karplus group developed molecular dynamics (MD) simulations of a protein in 1977 [2], which led to the development of the CHARMM [5] (Chemistry at Harvard Molecular Mechanics) program which became a central platform for many molecular simulation methods over the following decades. One of the most influential developments for structure-based drug discovery in the 1980s was the program GRID from Goodford [3]. This introduced the idea of

Figure 1



Methods for predicting ligand binding modes illustrated through an example of calculations for the kinase, CDK2 (protein structure taken from the PDB code: 1CKP). **(a)** Binding site prediction by fpocket [15] (default settings) by clustering solvent inaccessible spheres and disregarding solvent exposed spheres. A 'druggability' score is assigned to each predicted pocket. In this example, Site I, obtained the highest score (0.8), while the remaining eight pockets score very low (<0.1). **(b)** Polar hot spots identified through mixed solvent MD simulations using MDMix [20]. Ethanol and water were used to probe the binding pocket, from which high and low energy areas are identified. The low energy areas probed by ethanol (deep purple), help to identify donor or acceptor features that could be exploited by ligand binding. Water (cyan) and hydrophobic (yellow) sites are also probed. **(c)** These hot spots were then used to guide docking of the ligand from PDB structure 1PXM. Docking was performed with rDock [31], using a donor as a pharmacophoric restraint (sphere) to interact with the backbone of LEU78 (yellow dashed line) in the CDK2 structure. **(d)** This was followed by pose refinement using MD [54] to explore the flexibility of the pocket; for example, the yellow surface indicates possibility for a clash between ligand and protein.

characterizing what types of chemical functionality would bind to a binding site by calculating the energy of interaction between the protein and a functional group at each point on a grid. Finally, there is the DOCK program from the Kuntz group [4] which was the first widely used program for computationally docking compounds into the structure of a protein. Although some of the ideas within these programs were built on the work of others, the programs (and their authors) became major promoters of the ideas of using computational methods to characterize and predict how compounds can bind to proteins and formed the foundation of the current generation of methods.

Predicting the energy of interaction between protein and ligand – scoring functions

All structure-based design methods critically rely on an estimate of the energy of interaction between a ligand (or

probe) and the protein. Most approaches still rely on the rather simplistic treatment established in the early methods [2–4] where the non-covalent interactions are treated with simple Coulombic (for electrostatics) or Lennard-Jones (for van der Waals) interaction potentials but there is increasing use of more sophisticated treatments. The theoretical bases for these more advanced calculations were established a long time ago. What has changed in recent years is the relentless increase in computer power allowing these methods to be applied within a realistic timeframe. There are three main areas to highlight.

The first are the perturbation methods [6] which calculate changes in free energy by performing extensive MD while transforming (in this case the compound) from one chemical structure to another. The second is a number of approaches for more extensive treatment of

electrostatics: for example with Poisson-Boltzman or Generalised Born models within a molecular mechanics calculation (called MM-PBSA and MM-GBSA) — this can account for bulk solvent effects [7]; or using explicit quantum mechanics for part of the model within a molecular mechanics model to account more fully for both electrostatics [8] and metal ion charges [9]. Finally, there has been a growing realization that water molecules, both their stability and their interaction networks, can make important contributions to the thermodynamics of ligand binding [10,11^{*}]. However, even with these advances, our ability to predict the thermodynamics from a static model of a protein–ligand structure is still far from predictive. Nonetheless, the methods have been central to recent successes in structure-based design, such as potent non-nucleoside reverse transcriptase inhibitors [12] and other examples described below.

Pocket detection

The first challenge of SBDD is identifying where the ligand binds, for which geometry-based and energy-based programs have been successfully developed [13]. Geometry-based programs such as CAVITY [14] or fpocket [15] identify the largest and deepest cavities within a static structure; examples of successful use are identification of two new allosteric sites in the crystal structure of PHGDH in which virtual screening (VS) identified compounds with anti-tumor activity [16^{**}] and cavity detection in models of the intrinsically disordered protein Myc for which VS identified cell active compounds [17]. Another method uses hidden Markov models to identify cryptic pockets which can then be exploited as allosteric sites [18^{*}]. The more difficult to identify pockets are where a flexible model of the protein is needed to identify previously unidentified cavities. In these cases, energy-based methods similar to those used to identify protein–ligand hot-spots [19,20] can be used to map the protein surface repeatedly while exploring low energy conformers of side chains. Such calculations complement some of the experimental methods which explore transient sites and can be incorporated into strategies for ligand design [21,22]. Finally, a comprehensive review [23^{*}] summarises the large number of methods developed to compare binding sites on the basis of protein structural information, with many examples of how this has been used in ligand design.

Solvent mapping/probe mapping

Many approaches have evolved to characterize what will bind to a site since the original GRID (which uses a point probe [3]) and MCSS (which was the first example of using functional groups [24]). A number of groups [19,20] have developed variations on mixed solvent dynamics (reviewed in [25^{**}]), where simulations use as many as 16 different chemical probes to explore what could bind to (a usually flexible) binding site. In one ambitious study [26], the methods identified key interaction hot spots on

an ensemble of structures derived from MD simulations of X-ray structures and homology models from which VS identified an FGF-23 antagonist. The optimized compound was subsequently demonstrated to have activity in mouse models.

A more specific example is the characterization of water molecules in protein structures. Programs like WaterMap [11^{*}] and MDMix [20] can identify which water molecules are energetically favorable to displace in compound optimization. One striking example is the identification of two high energy waters in the binding site of the FGFR kinase in the presence of a lead compound, which could be displaced by modification of the scaffold leading to Rogaratinib (BAY 1163877), a pan FGFR inhibitor [27^{*}]; another example is identification of Acetyl-CoA carboxylase inhibitors which are active in animal models of obesity and diabetes [28].

A variant on solvent mapping is hybrid structure-based and ligand-based methods. One example is the ALTA-VS strategy [29], where pharmacophoric features are derived from the ligands. A virtual library of rigid fragments is constructed cutting all the rotatable bonds of the compounds in an available library and then computationally docking each fragment into the binding site, then evaluating energy of interaction (in this case using an MM-GBSA force field). The best scoring fragments can then be used as pharmacophoric points to direct VS. A more extreme example of this type of approach is where compounds validated as binding after VS against a homology model is used to refine the homology model for more comprehensive modelling [30].

Pose prediction and refinement

Once a binding pocket has been identified and characterized, the next step is to predict the binding pose of a ligand in that site — molecular docking. Assessing how well different programs can make this prediction has been a continuous industry for the molecular modelling community (see our own work with the program, rDock [31]). A recent paper [32] provides one of the more comprehensive studies of the past 5 years, assessing how well the experimentally observed pose can be predicted by ten different programs for a test set of 2002 protein–ligand crystal structures. There are two main criteria on which programs can be assessed: can the program generate the correct binding pose?, and does the scoring function successfully identify this pose as the most favored? In general, where receptor flexibility is not important, the correct pose can be generated and although for some protein families, there was reasonable correlation between docking scores and experimental binding affinities, the ranking of the binding affinity was not well predicted across the whole dataset.

Success can be improved by combining docking with alternate scoring schemes. MM-GBSA can be used to rank order docking hits with some exciting examples [33–35]. A more recent development is dynamic approaches such as Duck [36**] (Figure 1). DUCK uses steered MD where a virtual ‘force’ is applied to pull the ligand out of the binding site, while performing MD simulations to calculate the energetic cost of breaking a key hydrogen bond between a ligand and the protein. The theoretical relationship is tenuous, but this could be taken as related to the activation energy for the interaction, which will affect the binding energy. Another approach is free energy perturbation (introduced in [36**]) with a number of examples [37–39].

Dynamic docking

The evaluation of docking programs [32] usually tests for re-docking a ligand into the protein conformation obtained in the protein ligand crystal structure. What is more challenging (but the more realistic application scenario) is to predict the binding pose where there is some adjustment in the detailed conformation of the protein binding site. There are a number of approaches to address this problem. The first is to use different conformations of the protein as the target for docking. Ensembles of conformations for the protein can be generated by MD [40,41] or different experimental crystal structures can be used [42] as a project proceeds and more crystal structures obtained. One recent method development is a fast algorithm for sampling flexible protein–ligand conformations (known as PELE [43]) that was able to reproduce ligand induced side chain rearrangements and small main chain protein movement in a set of protein–ligand complexes, and performed better than MD simulations and induced fit docking.

A second approach is to account for receptor flexibility while docking and there have been a number of recent reviews [44–46] which survey progress. The methods rely on substantial computer power, so at present can only be applied to a few compounds and are more applicable for hit to lead optimization rather than hit identification. The methods fall into two main categories — those requiring an *a priori* definition of the pathway for conformational change and those that do not. In a recent publication [47] an induced fit docking protocol is used to generate possible conformational changes that enable ligand binding. These possible conformations are then assessed using metadynamics, a computational device to encourage a molecular simulation to explore across the possible conformations available, by introducing energy terms which discourage the simulation returning to conformations that have already been visited. Significant improvement in the quality of docking is reported across 42 test systems.

There are a number of recent publications where the simulation is unsteered and explores conformational

change unsupervised. In one case [48], a technique called potential scaled MD was used to predict the binding pose in two test cases with conformational change. The method works by lowering the barrier between conformational states, in some ways emulating use of a high temperature. An alternate unsupervised method is adaptive electrostatic bias [49] where the electrostatic interactions are modulated depending on proximity of the ligand to the binding site — this also reduces the barrier to conformational change. These new dynamics methods complement more conventional MD, as in the FGF example discussed above [26].

This is a promising new concept for which several retrospective examples have emerged — for example adaptive sampling allowed high throughput MD of a small fragment library for CXCL2 [50], while there are even examples of long MD simulation used for *in silico* fragment screening [51], which also emphasizes the power of GPU-accelerated computational power in structure-based drug design.

Concluding remarks

The October 1st, 1981 edition of Fortune magazine heralded a ‘New Industrial Revolution’ in which drugs can be designed by computer. The pharmaceutical industry has encountered waves of new technology (e.g. combinatorial chemistry, genomics). In most cases, it takes many years (decades) for the methods but also the expertise to develop so that the methods can make effective contributions to the drug discovery process. Structure-based, computational methods have suffered more than most new technologies in achieving this routine, productive phase. However, the techniques have now made recognizable contributions to the design of more than 50 compounds in clinical trials [52] and to several drugs on the market [53].

In this mini-review, we have focused on new developments in methods for just one aspect of SBDD — predicting how compounds bind to their target. What we have not discussed in detail is how these methods (combined with advances in predicting strength of protein–ligand binding) are now contributing to an increasing number of success stories where potent compounds are being identified for a range of targets. Looking back over the past forty years, perhaps it has not been an ‘industrial revolution’, but more a continuous scientific evolution. Steady improvements in quality of the methods and understanding of how they can be used has led to increased acceptance and confidence so that the medicinal chemistry community now can appreciate how SBDD methods can contribute to the drug design process.

Acknowledgements

This work was supported by the European Union’s Horizon2020 MSCA Programme under grant agreement 675899 (FRAGNET).

References

- Hubbard RE: **3D structure and the drug-discovery process.** *Mol BioSyst* 2005, **1**:391-406.
 - McCammon JA, Gelin BR, Karplus M: **Dynamics of folded proteins.** *Nature* 1977, **267**:585-590.
 - Goodford PJ: **A computational procedure for determining energetically favorable binding sites on biologically important macromolecules.** *J Med Chem* 1985, **28**:849-857.
 - Kuntz ID, Blaney JM, Oatley SJ, Ferrin RE, Thomas L: **A geometric approach to macromolecule–ligand interactions.** *J Mol Biol* 1982, **161**:269-288.
 - Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S *et al.*: **CHARMM: the biomolecular simulation program.** *J Comput Chem* 2009, **30**:1545-1614.
 - Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J *et al.*: **Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field.** *J Am Chem Soc* 2015, **137**:2695-2703.
 - Sun H, Li Y, Shen M, Tian S, Xu L, Pan P, Guan Y, Hou T: **Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring.** *Phys Chem Chem Phys* 2014, **16**:22035-22045.
 - Kolár MH, Hobza P: **Computer modeling of halogen bonds and other σ -hole interactions.** *Chem Rev* 2016, **116**:5155-5187.
 - Chaskar P, Zoete V, Rohrig UF: **On-the-fly QM/MM docking with attracting cavities.** *J Chem Inf Model* 2017, **57**:73-84.
 - Spyrakakis F, Ahmed MH, Bayden AS, Cozzini P, Mozzarelli A, Kellogg GE: **The roles of water in the protein matrix: a largely untapped resource for drug discovery.** *J Med Chem* 2017, **60**:6781-6827.
 - Cappel D, Sherman W, Beuming T: **Calculating water thermodynamics in the binding site of proteins – applications of water map to drug discovery.** *Curr Top Med Chem* 2017, **17**:2586-2598.
- A collection of examples of how simulations can identify water molecules that can readily be displaced by modifications to a ligand in a binding site.
- Kudalkar SN, Beloor J, Quijano E, Spasov KA, Lee WG, Cisneros JA, Saltzman WM, Kumar P, Jorgensen WL, Anderson KS: **From in silico hit to long-acting late-stage preclinical candidate to combat HIV-1 infection.** *Proc Natl Acad Sci U S A* 2018, **115**:E802-E811.
 - Xie ZR, Hwang MJ: **Methods for predicting protein–ligand binding sites.** *Methods Mol Biol* 2015, **1215**:383-398.
 - Yuan Y, Pei J, Lai L: **Binding site detection and druggability prediction of protein targets for structure-based drug design.** *Curr Pharm Des* 2013, **19**:2326-2333.
 - Le Guilloux V, Schmidtke P, Tuffery P: **Fpocket: an open source platform for ligand pocket detection.** *BMC Bioinform* 2009, **10**:168.
 - Wang Q, Liberti MV, Liu P, Deng X, Liu Y, Locasale JW, Lai L: **Rational design of selective allosteric inhibitors of PHGDH and serine synthesis with anti-tumor activity.** *Cell Chem Biol* 2017, **24**:55-65.
- Two previously unexploited binding pockets were identified computationally in the structure of the enzyme PHGDH; virtual screening identified compounds which showed activity and models of how these compounds bound used to rationally design potent, cell-active inhibitors.
- Yu C, Niu X, Jin F, Liu Z, Jin C, Lai L: **Structure-based inhibitor design for the intrinsically disordered protein c-Myc.** *Sci Rep* 2016, **6**:1-11.
 - Hart KM, Moeder KE, Ho CMW, Zimmerman MI, Frederick TE, Bowman GR: **Designing small molecules to target cryptic pockets yields both positive and negative allosteric modulators.** *PLoS ONE* 2017, **12**:1-13.
- Another example of where computational methods identify additional binding sites on a model enzyme for which virtual screening identified both activators and inhibitors.
- Kozakov D, Grove LE, Hall DR, Bohndud T, Mottarella SE, Luo L, Xia B, Beglov D, Vajda S: **The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins.** *Nat Protoc* 2015, **10**:733-755.
 - Alvarez-Garcia D, Barril X: **Molecular simulations with solvent competition quantify water displaceability and provide accurate interaction maps of protein binding sites.** *J Med Chem* 2014, **57**:8530-8539.
 - Fischer M, Shoichet BK, Fraser JS: **One crystal, two temperatures: cryocooling penalties alter ligand binding to transient protein sites.** *ChemBioChem* 2015, **16**:1560-1564.
 - Fischer M, Fraser JS, Stoichet BK: **The incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery.** *Nat Chem* 2015, **6**:575-583.
 - Ehrt C, Brinkjost T, Koch O: **Impact of binding site comparisons on medicinal chemistry and rational molecular design.** *J Med Chem* 2016, **59**:4121-4151.
- A comprehensive review of methods for characterising binding sites with many examples of the use of computational methods for drug design.
- Miranker A, Karplus M: **Functionality maps of binding sites: a multiple copy simultaneous search method.** *Proteins* 1991, **11**:29-34.
 - Ghanakota P, Carlson HA: **Driving structure-based drug discovery through cosolvent molecular dynamics.** *J Med Chem* 2016, **59**:10383-10399.
- A comprehensive and well written survey of the new generation of cosolvent molecular dynamics simulations which characterise what functional groups could bind to a site.
- Xiao Z, Riccardi D, Velazquez HA, Chin AL, Yates CR, Carrick JD, Smith JC, Baudry J, Quarles LD: **A computationally identified compound antagonizes excess FGF-23 signaling in renal tubules and a mouse model of hypophosphatemia.** *Sci Signal* 2016, **9**:1-14.
 - M-P Collin, Lobell M, Hübsch W, Brohm D, Schirok H: **Discovery of rogaratinib (BAY 1163877): a pan-FGFR inhibitor.** *ChemMedChem* 2018:1-10.
- A good example of using *in silico* methods to identify a clinical candidate molecule.
- Harriman G, Greenwood J, Bhat S, Huang X, Wang R, Paul D, Tong L, Saha AK, Westlin WF, Kapeller R *et al.*: **Acetyl-CoA carboxylase inhibition by ND-630 reduces hepatic steatosis, improves insulin sensitivity, and modulates dyslipidemia in rats.** *Proc Natl Acad Sci U S A* 2016, **113**:E1796-E1805.
 - Marchand JR, Dalle Vedove A, Lolli G, Caflich A: **Discovery of inhibitors of four bromodomains by fragment-anchored ligand docking.** *J Chem Inf Model* 2017, **57**:2584-2597.
 - Levoine N, Labeeuw O, Billot X, Calmels T, Danvy D, Krief S, Berrebi-Bertrand I, Lecomte JM, Schwartz JC, Capet M: **Discovery of nanomolar ligands with novel scaffolds for the histamine H4 receptor by virtual screening.** *Eur J Med Chem* 2017, **125**:565-572.
 - Ruiz-Carmona S, Alvarez-Garcia D, Follope N, Beatriz Garmendia-Doval A, Juhos S, Schmidtke P, Barril X, Hubbard RE, Morley SD: **rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids.** *PLoS Comput Biol* 2014, **10**.
 - Wang Z, Sun H, Yao X, Li D, Xu L, Li Y, Tian S, Hou T: **Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power.** *Phys Chem Chem Phys* 2016, **18**:12964-12975.
 - Manglik A, Lin H, Aryal DK, McCorvy JD, Dengler D, Corder G, Levit A, Kling RC, Bernat V, Hübner H *et al.*: **Structure-based discovery of opioid analgesics with reduced side effects.** *Nature* 2016, **537**:185-190.

34. McCorvy JD, Butler KV, Kelly B, Rechsteiner K, Karpiak J, Betz RM, Kormos BL, Shoichet BK, Dror RO, Jin J *et al.*: **Structure-inspired design of β -arrestin-biased ligands for aminergic GPCRs.** *Nat Chem Biol* 2018, **14**:126-134.
35. Wang S, Wacker D, Levit A, Che T, Betz RM, McCorvy JD, Venkatakrisnan AJ, Huang XP, Dror RO, Shoichet BK *et al.*: **D4 dopamine receptor high-resolution structures enable the discovery of selective agonists.** *Science* 2017, **358**:381-386.
36. Ruiz-Carmona S, Schmidtke P, Luque FJ, 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.
- Demonstration of a method that uses steered molecular dynamics to estimate the energy required to 'undock' a compound which can be taken a surrogate for the energetics of binding.
37. Lenselink EB, Louvel J, Forti AF, Van Veldhoven JPD, De Vries H, Mulder-Krieger T, McRobb FM, Negri A, Goose J, Abel R *et al.*: **Predicting binding affinities for GPCR ligands using free-energy perturbation.** *ACS Omega* 2016, **1**:293-304.
38. Wang L, Deng Y, Knight JL, Wu Y, Kim B, Sherman W, Shelley JC, Lin T, Abel R: **Modeling local structural rearrangements using FEP/REST: application to relative binding affinity predictions of CDK2 inhibitors.** *J Chem Theory Comput* 2013, **9**:1282-1293.
39. Ciordia M, Pérez-Benito L, Delgado F, Trabanco AA, Tresadern G: **Application of free energy perturbation for the design of BACE1 inhibitors.** *J Chem Inf Model* 2016, **56**:1856-1871.
40. Salvatella X: **Understanding protein dynamics using conformational ensembles.** *Adv Exp Med Biol* 2014, **805**:67-85.
41. Miao Y, Goldfeld DA, Moo EV, Sexton PM, Christopoulos A, McCammon JA, Valant C: **Accelerated structure-based design of chemically diverse allosteric modulators of a muscarinic G protein-coupled receptor.** *Proc Natl Acad Sci U S A* 2016, **113**:E5675-E5684.
42. Ahmed-Belkacem A, Colliandre L, Ahnou N, Nevers Q, Gelin M, Bessin Y, Brillet R, Cala O, Douguet D, Bourguet W *et al.*: **Fragment-based discovery of a new family of non-peptidic small-molecule cyclophilin inhibitors with potent antiviral activities.** *Nat Commun* 2016, **7**.
43. Grebner C, Iegre J, Ulander J, Edman K, Hogner A, Tyrchan C: **Binding mode and induced fit predictions for prospective computational drug design.** *J Chem Inf Model* 2016, **56**:774-787.
44. Gioia D, Bertazzo M, Recanatini M, Masetti M, Cavalli A: **Dynamic docking: a paradigm shift in computational drug discovery.** *Molecules* 2017, **22**.
45. Sledz P, Caffisch A: **Protein structure-based drug design: from docking to molecular dynamics.** *Curr Opin Struct Biol* 2017, **48**:93-102.
46. Ganesan A, Coote ML, Barakat K: **Molecular dynamics-driven drug discovery: leaping forward with confidence.** *Drug Discov Today* 2017, **22**:249-269.
47. Clark AJ, Tiwary P, Borrelli K, Feng S, Miller EB, Abel R, Friesner RA, Berne BJ: **Prediction of protein-ligand binding poses via a combination of induced fit docking and metadynamics simulations.** *J Chem Theory Comput* 2016, **12**:2990-2998.
48. Bertazzo M, Bernetti M, Recanatini M, Masetti M, Cavalli A: **Fully flexible docking via reaction-coordinate-independent molecular dynamics simulations.** *J Chem Inf Model* 2018.
49. Spitaleri A, Decherchi S, Cavalli A, Rocchia W: **Fast dynamic docking guided by adaptive electrostatic bias: The MD-binding approach.** *J Chem Theory Comput* 2018.
50. Martinez-Rosell G, Harvey MJ, De Fabritiis G: **Molecular-simulation-driven fragment screening for the discovery of new CXCL12 inhibitors.** *J Chem Inf Model* 2018 <http://dx.doi.org/10.1021/acs.jcim.7b00625>.
51. Pan AC, Xu H, Palpant T, Shaw DE: **Quantitative characterization of the binding and unbinding of millimolar drug fragments with molecular dynamics simulations.** *J Chem Theory Comput* 2017, **13**:3372-3377.
52. Ban F, Dalal K, Li H, LeBlanc E, Rennie PS, Cherkasov A: **Best practices of computer-aided drug discovery: lessons learned from the development of a preclinical candidate for prostate cancer with a new mechanism of action.** *J Chem Inf Model* 2017, **57**:1018-1028.
53. Rognan D: **The impact of in silico screening in the discovery of novel and safer drug candidates.** *Pharmacol Ther* 2017, **175**:47-66.
54. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Zhang W, Merz KM, Roberts B, Hayik S, Roitberg A, Seabra G, Swails J, Götz AW, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wolf RM, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Salomon-Ferrer R, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA: **Amber12.** San Francisco: University of California; 2012.

DUckCov: a Dynamic Undocking-Based Virtual Screening Protocol for Covalent Binders

Moira Rachman^{1,2}, Andrea Scarpino², David Bajusz², Gyula Palfy³, Istvan Vida³, Andras Perczel³, Xavier Barril,^{1,4} and Gyorgy M. Keseru²

“

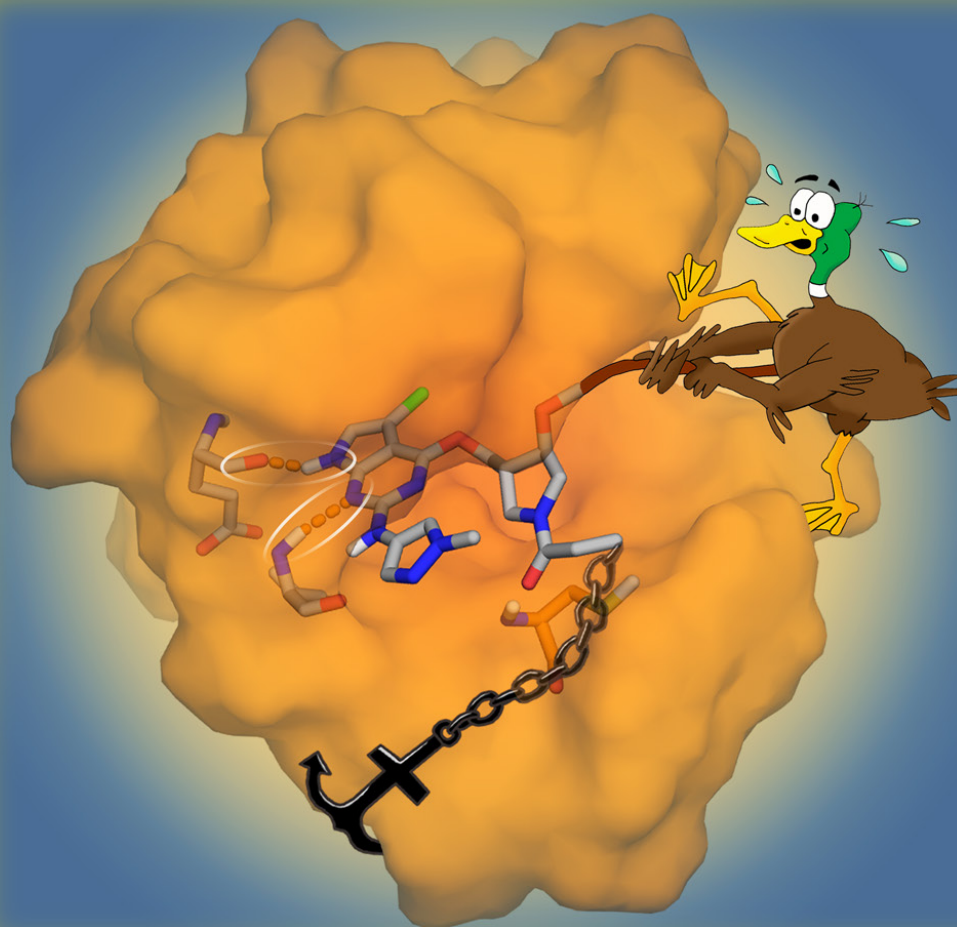
In *ChemMedChem* (2019), 14, 1011-21

1. Institut de Biomedicina de la Universitat de Barcelona (IBUB) and Facultat de Farmacia, Universitat de Barcelona, Av. Joan XXIII 27-31, 08028 Barcelona, Spain.
2. Medicinal Chemistry Research Group, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest (Hungary).
3. Laboratory of Structural Chemistry and Biology and MTA-ELTE Protein Modelling Research Group, Budapest (Hungary)
4. Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.



CHEM MED CHEM

CHEMISTRY ENABLING DRUG DISCOVERY



10/2019

Front Cover:

György M. Keserű et al.

DUckCov: a Dynamic Undocking-Based Virtual Screening Protocol
for Covalent Binders

A Journal of



WILEY-VCH

www.chemmedchem.org



DUckCov: a Dynamic Undocking-Based Virtual Screening Protocol for Covalent Binders

Moira Rachman,^[a, b] Andrea Scarpino,^[b] Dávid Bajusz,^[b] Gyula Pálffy,^[c] István Vida,^[c] András Perczel,^[c] Xavier Barril,^[a, d] and György M. Keserü^{*,[b]}

Thanks to recent guidelines, the design of safe and effective covalent drugs has gained significant interest. Other than targeting non-conserved nucleophilic residues, optimizing the noncovalent binding framework is important to improve potency and selectivity of covalent binders toward the desired target. Significant efforts have been made in extending the computational toolkits to include a covalent mechanism of protein targeting, like in the development of covalent docking methods for binding mode prediction. To highlight the value of the noncovalent complex in the covalent binding process,

here we describe a new protocol using tethered and constrained docking in combination with Dynamic Undocking (DUck) as a tool to privilege strong protein binders for the identification of novel covalent inhibitors. At the end of the protocol, dedicated covalent docking methods were used to rank and select the virtual hits based on the predicted binding mode. By validating the method on JAK3 and KRas, we demonstrate how this fast iterative protocol can be applied to explore a wide chemical space and identify potent targeted covalent inhibitors.

Introduction

Despite the activity of a large number of drugs approved by the US Food and Drug Administration (FDA) that depend on a covalent mode of action,^[1] classical drug discovery screening cascades typically eliminate electrophilic compounds, mainly due to the toxicity risks associated with their mechanism. Indeed, a majority of these drugs was discovered by serendipity in biological assays, and their mechanism was elucidated later on, typically after approval. The reluctance to use reactive ligands, and more specifically promiscuous “suicide inhibitors”, is related to increased risks of carcinogenicity, hepatotoxicity, and potential idiosyncratic effects caused by protein haptization.^[2,3]

More recently, the reputation of covalent binders has changed thanks to the guidelines introduced for the rational design of targeted covalent inhibitors (TCIs). According to these guidelines, the ligand’s selectivity toward its protein target is still to be achieved by optimizing the noncovalent interactions (hydrogen bonding, van der Waals, electrostatic, etc.) at the binding site, as in the case of traditional approaches. Furthermore, increased specificity can be obtained by targeting a poorly conserved reactive residue within the protein family.^[3] To this effect, the development of methods to identify poorly conserved reactive residues have aided the acceleration of TCI design. For example, activity-based protein profiling techniques (ABPP, isoTOP-ABPP^[4,5]) can be used to both investigate the activity at the proteomic level and quantify the intrinsic reactivity of functional cysteines. Also, Liu and colleagues have coined the term “kinase cysteinome” to refer to the collection of targetable cysteine residues in the human kinome^[6] and published a computational methodology to identify such cysteines.^[7]

Ligands that bind through a covalent mechanism are not subject to classical equilibrium kinetics, as their residence time in the binding pocket can last up to days. As a consequence, the potency of these drugs is capable of surpassing the theoretical limits of potency/ligand efficiency.^[2] Another advantage is the prolonged duration of action, which can persist even when the ligand has already been cleared from the body. This can be beneficial for alleviating the drug burden of a patient due to less frequent drug dosing (depending on the turnover rate of the protein) and therefore a possibly lower risk of idiosyncratic toxicity, which has been linked to daily drug dosage.^[8]

[a] M. Rachman, Prof. X. Barril
Facultat de Farmàcia and Institut de Biomedicina, Universitat de Barcelona,
Av. Joan XXIII 27–31, 08028 Barcelona (Spain)

[b] M. Rachman, A. Scarpino, Dr. D. Bajusz, Prof. G. M. Keserü
Medicinal Chemistry Research Group, Research Centre for Natural Sciences,
Hungarian Academy of Sciences, Magyar tudósok körútja 2, 1117 Budapest
(Hungary)
E-mail: keseru.gyorgy@ttk.mta.hu

[c] G. Pálffy, I. Vida, Prof. A. Perczel
Laboratory of Structural Chemistry and Biology & MTA-ELTE Protein Model-
ling Research Group, Eötvös Loránd University, Pázmány Péter sétány 1/A,
1117 Budapest (Hungary)

[d] Prof. X. Barril
Catalan Institution for Research and Advanced Studies (ICREA), Passeig
Lluís Companys 23, 08010 Barcelona (Spain)

Supporting information and the ORCID identification number(s) for the
author(s) of this article can be found under:
<https://doi.org/10.1002/cmdc.201900078>.

© 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA.
This is an open access article under the terms of the Creative Commons
Attribution License, which permits use, distribution and reproduction in
any medium, provided the original work is properly cited.

In addition, the TCI approach has proven to be a valuable tool in targeting protein binding sites, which were previously considered as undruggable, as well as to combat drug resistance by targeting poorly conserved non-catalytic residues. Overall, all of these aspects have contributed to a resurgence of covalent drug discovery programs, which has already led to an increase of clinical candidates acting via a covalent mechanism.^[9]

In general, a covalent binder first requires the formation of an initial noncovalent complex with its target, followed by the chemical reaction between the ligand's electrophilic warhead and the nucleophilic residue. As such, the most straightforward covalent drug design approach is based on the modification of a known noncovalent binder to introduce an electrophilic warhead. This could indeed allow to reach and covalently modify the targeted nucleophile on the protein by maintaining the overall binding mode in the rest of the pocket. Additionally, an important strategy is to fine-tune the warhead reactivity based on the target nucleophilicity in order to limit possible side effects arising from off-target modifications.^[9–11]

From a computational perspective, once an appropriate nucleophile and warhead are identified, a structure-based approach can be used to screen or optimize ligands to fit the binding site, while also being able to place the warhead in the vicinity of the targeted residue to form the covalent bond. Several covalent docking methods have recently been developed to model the structural changes occurring when covalent ligands bind to their target. However, other than the inherent limitations of traditional docking methods (i.e., scoring, protein flexibility, solvation, and nonclassical effects),^[12] these tools also have to address additional challenges in the simulation. Predicting the optimal geometry of the reacting groups upon covalent bond formation is of key importance for accurate simulations. Furthermore, covalent docking programs face the inability to evaluate the energy of bond formation, which would require QM-based simulations of the reaction. Depending on the method of choice, modeling all the different and key aspects characterizing the binding of covalent ligands is often reflected in higher computational costs than for traditional noncovalent docking.

Among the first developed tools are GOLD^[13] and AutoDock:^[14] the former enforces the covalent reaction through the definition of a link atom in both the ligand and receptor before initiating the genetic algorithm search, while the latter offers the opportunity to choose between the *two-point attractor* approach and the better performing *flexible side-chain* method, in which the ligand is sampled as part of the protein. In addition to a lack of the energetic contribution of covalent binding, the manual definition of the atoms involved in the reaction hinders the applicability of covalent docking programs to large libraries. A recent approach taken by CovalentDock^[15] automatically detects reactive atoms for linking and rewards the energy contribution of the binding event as an additional MM-based term. The authors retrospectively validated their method on 76 covalently bound ligands in the Protein Data Bank (PDB), for which CovalentDock showed better performance than GOLD and AutoDock. However, CovalentDock is

limited in reaction types (only Michael addition and β -lactam opening are supported) and does not account for the flexibility of the reacted residue. Furthermore, the cloud web server developed for its usage appears to no longer be available (access attempted on October 16, 2018). More recently, other web-based servers such as DOCKoValent,^[16] or proprietary software such as ICM-Pro,^[17] FITTED,^[18] and DOCKTITE^[19] (an SVL-based workflow for the modeling software MOE^[20]) enabled covalent docking-based virtual screening applications by using predefined and customizable reactions to identify reacting groups.

Schrödinger's CovDock^[21] takes it one step further and mimics the full binding process of covalent ligands (as opposed to only taking into account the covalently attached ligand–protein complex). With this, CovDock highlights the importance of the noncovalent interactions formed prior to covalent binding. The multistep algorithm provides two alternative solutions by means of a “pose prediction” module and a virtual screening module (CovDock-VS). The former includes an extensive protocol for the prediction of the covalently bound pose, namely: I) ligand conformation generation; II) positioning the pre-reaction form of the ligand warhead close to the receptor reactive residue (mutated to Ala) using a constrained docking; III) resetting the mutation to the original residue, sampling its rotameric states, and generating the covalent attachment; IV) clustering and minimization of the poses (including the reacted residue); and V) scoring by means of the Prime energy model. An additional affinity score, which averages GlideScore on both the pre- and post-reaction forms of the ligand, is provided to compare different compounds equipped with the same or similar reactive warheads. While it shows good binding mode prediction accuracy, this protocol takes roughly 1–2 CPU hours per ligand, so it is not suited for high-throughput screenings. Toledo Warshaviak and colleagues addressed this issue by developing CovDock-VS,^[22] which I) skips the ConfGen step, II) limits the number of resulting pose clusters to three, III) excludes minimization by Prime, and IV) scores and ranks protein–ligand complexes based only on the initial GlideScore. Ultimately, this led to significantly improved speeds (\approx 15 minutes per structure on a single CPU according to the info on CovDock's latest release) over the pose prediction module, but also yielded less accurate binding mode predictions, unless known interaction patterns were incorporated.

In general, the performance gap in terms of binding mode prediction among the different covalent docking programs was shown to vary significantly depending on various factors (i.e., protein target, accessibility of the nucleophilic residue, amount of noncovalent interactions occurring in the complex).^[23] On the other hand, the speed of the simulation remains one of the main bottlenecks that can drastically affect the size and diversity of the covalent libraries used for screening applications. To this end, herein we present DUCkCov, a time-efficient multistep VS protocol for the identification of novel covalent binders. It was devised to emphasize the role of the interactions mediating the initial noncovalent complex, whose optimization can, therefore, result in both an increase of the selectivity for the target and in an opportunity to decrease the reactivity of the electrophile. As depicted in

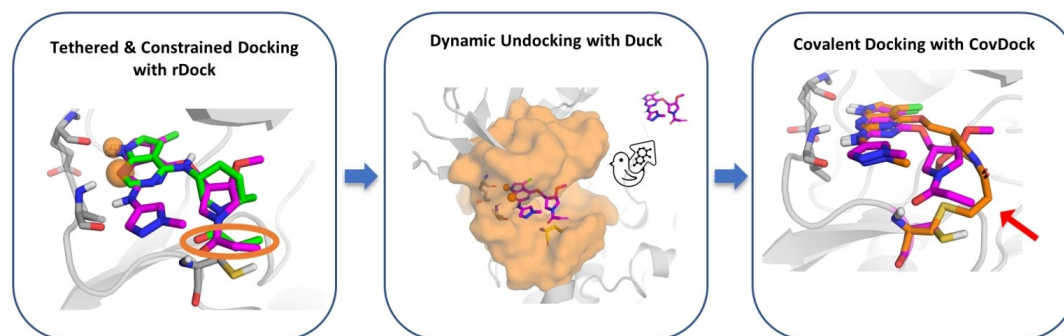


Figure 1. Starting from a library of covalent ligands, the general workflow is as follows: 1) docking with rDock with pharmacophoric constraints (orange spheres) and positional restraints for the warhead (encircled by orange ellipse), 2) dynamic undocking to test the strength of the H-bond interaction that was enforced during docking, and 3) covalent docking of ligands that display the best noncovalent interactions to account for warhead flexibility (red arrow).

Figure 1, rDock^[24] is first used to constrain the reactive warhead close to the targeted residue. During docking, pharmacophoric restraints are applied to known H-bond interaction points, if any. Dynamic Undocking (DUck)^[25] is then used to assess the strength of these H-bonds. DUck evaluates structural stability, rather than thermodynamic stability, and has been shown to be orthogonal to methods that attempt to estimate the binding energy. H-bonds are suggested to be the main determinants of structural stability based on their sharp distance and angular dependencies, and their role in structure-kinetic relationships.^[25,26] Finally, CovDock is used to evaluate the binding mode of those ligands that optimally bind through noncovalent interactions, and to check if the same interaction pattern is maintained in the predicted covalent docking pose.

The protocol was prospectively validated in two case studies: a target with highly conserved noncovalent interactions (JAK3) and another one where the noncovalent interactions are not conserved across known inhibitors (KRas^{G12C}).

Results and Discussion

Case study 1: JAK3

JAK3 is one of the four Janus kinases (a subfamily of tyrosine kinases), the only of which is primarily restricted to leukocytes. Its functional modulation has been associated with a phenotype of severe combined immunodeficiency.^[27] Although significant effort has been put into the discovery of JAK inhibitors, the search for JAK-specific ligands is still on-going. JAK3 specificity over other family members can be achieved by targeting C909 with ligands that are able to covalently bind this residue. H-bond interactions are known to play a prominent role in building up affinity toward kinase targets. Because the majority of kinase inhibitors bind to the highly conserved hinge motif, DUckCov application on JAK3 was focused on the identification of covalent ligands displaying strong interactions at this region.

The DUckCov workflow for JAK3 is described in Figure 2B. Based on the selected JAK3 structure, tethered and constrain-

ed docking filtered the acrylamide dataset from roughly 50000 compounds to 249 compounds that satisfied the H-bond interactions with E903 (backbone C=O) and L905 (backbone NH) at the hinge region (depicted in cyan in Figure 2A), while conserving the acrylamide group close to the reactive C909 as observed in the prepared reference ligand (depicted in grey in Figure 2A). Next, DUck was performed, using the H-bond established with E903-O as the simulation coordinate, leading to 92 remaining hits. From those, a second round of DUck on the H-bond formed with L905-NH resulted in 66 compounds. In both cases, a W_{OB} (work necessary to pull the ligand from 2.5 to 5.0 Å relative to the defined H-bond interaction) threshold of 6 kcal mol⁻¹ was maintained. The consensus of both interactions was used to filter the rDock docking poses using DUck, as both these interactions are made by the reference ligand (with W_{OB} = 13 kcal mol⁻¹ and 11 kcal mol⁻¹, respectively). Then, in order to get both a quantitative ranking and more accurate binding mode predictions, CovDock in the "Pose prediction" module was used for the covalent docking simulations on the 66 DUck hits. Finally, we have selected the top 10 ligands according to their CovDock affinity scores. Out of these, five compounds (1, 3, 5, 8 and 9) were available for immediate purchase (Figure 2C–D); they were experimentally validated in an enzyme-based activity assay (see Methods section). For the rest of the top 10 ligands, see Supporting Information Figure S6.

Compound 1, the top-ranked ligand, has been originally reported as a potent double mutant EGFR^{L858R/T790M} inhibitor (PF-06459988, IC₅₀ = 7 nM),^[28] and based on the DUckCov prediction and subsequent *in vitro* testing, we found that the compound inhibits JAK3 with similar potency (5 nM). Interestingly, JAK3 activity of this compound was not reported previously; however, profiling against 54 human kinases at 1 μM revealed its moderate JAK1 and JAK2 inhibitory activity.^[28] Compounds 5, 6, 8 and 9 are characterized by a higher rigidity of the linker between the hinge binding region and the warhead. Compound 5 displayed an IC₅₀ value of 389 nM on JAK3, and 18% and 62% inhibition on JAK1 and JAK2 at 10 μM, respectively, suggesting a covalent bond driven improvement of the inhibi-

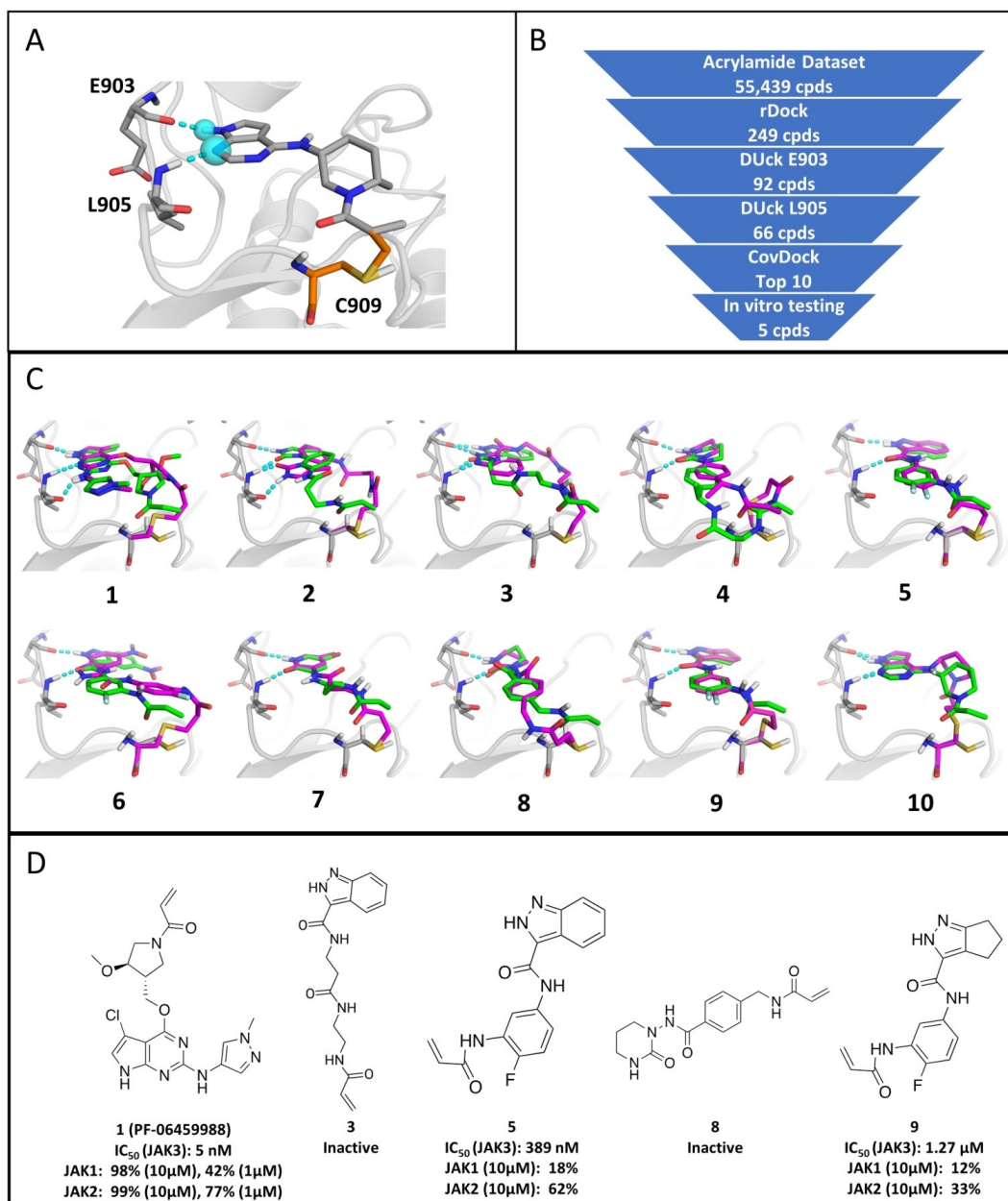


Figure 2. A) Pre-reaction reference ligand in the structure 5TOZ in grey, and covalent attachment in orange in the post-reaction form, with defined features as cyan spheres, and interactions in cyan dashes. B) DUckCov protocol for JAK3. C) Top-ten compounds, ranked by CovDock affinity scores (green: rDock binding modes, magenta: CovDock binding modes; interactions with the hinge residues E903 and L905 are displayed in cyan). D) Experimentally tested compounds for JAK3; the three confirmed compounds correspond to a hit rate of 60%.

tion due to the presence of the JAK3 unique reactive cysteine. Compound **9**, a partially saturated analogue of **5**, was found to

be less potent (1.27 μM IC₅₀), in line with the observed vast majority of aromatic hinge binding moieties known in the liter-

ature. In the same line, compound **8** showed no activity in the biochemical assay, further highlighting the preference for planar hinge-binding cores. Finally, compound **3** was also experimentally tested (as an analogue of **5** with a flexible linker), but has shown no activity.

In Figure 2C the poses generated by rDock and CovDock for the top 10 ligands are shown in green and magenta, respectively. In nine out of 10 cases the interactions used for pulling with DUck were reproduced in the best scoring pose generated by CovDock. (RMSD values, CovDock affinity, rDock scores, and DUck W_{OB} values, along with the ZINC codes of the compounds are given in Supporting Information Table S7). Compounds **1**, **4**, **5**, **7**, **8**, **9** and **10** retain the same orientation of the hinge binding region, while compounds **2** and **3** contain the hinge binding scaffold in a flipped orientation, due to the flexibility of the linker. However, compound **3** maintained both interactions at the hinge, while compound **2** showed a dual interaction with L905. In addition to the higher deviations (RMSD) between the binding modes predicted by rDock and CovDock, the linker flexibility of compounds **1–4** results in higher strain energies as well (relative to minimum energies of the free ligands, see Supporting Information Table S8). However, this does not necessarily prevent the compounds to be potent inhibitors of JAK3 (as exemplified by compound **1**), in accordance with the general notion that the linker rarely has a profound effect on activity. Compound **6** is the only case where neither an interaction with E903, nor L905 is observed. (The ten lowest ranked ligands are included as counter-examples in Supporting Information Figure S9, most of them lacking any kind of interaction with the hinge).

The 60% hit rate observed with DUckCov against JAK3 demonstrates the importance of noncovalent interactions established by covalent ligands. However, these findings also clearly show the need for dedicated covalent docking programs sampling multiple rotameric states of the reactive residue. This would increase the chance to identify the most optimal geometry of the covalent attachment, which could consequently reflect in a rearrangement of the overall binding mode that would be generated by tethered docking.

It is also important to note that running the DUckCov workflow took a total 1200 CPU/GPU hours, while running CovDock

Virtual screening on the whole dataset would have taken about 13750 CPU hours (15 minutes per ligand according to the software manual). The roughly 11-fold speedup can be mostly attributed to the quick tethered docking step, leaving only a fraction of the ligands to be evaluated by the more expensive Dynamic Undocking. If we account for parallelization as well, running this specific workflow in parallel on 24 GPUs of the Barcelona Supercomputing Center has required a total 50 hours of runtime, while our license token limit would have allowed us to run CovDock Virtual screening on three parallel threads (three ligands), resulting in about 4600 hours of total runtime, translating to a roughly 92-fold decrease in speed compared to the DUckCov workflow. The reported speedups can be considered typical for academic groups (based on the accessible resources), but in a more general sense, CPU/GPU time is more accessible (cheaper) than state-of-the-art software licenses (such as Schrödinger) for industrial researchers as well.

Case study 2: KRas^{G12C}

To challenge the method's applicability domain, it was also applied to another oncological target, the catalytic domain of KRas^{G12C}. For KRas^{G12C}, even the best irreversible binders show low potency if their covalent warhead is removed.^[29] KRas is a small G protein, which is rendered constitutively active by the G12C mutation, leading to abnormal cell growth. The mutation has been shown to be implicated in 40% of KRas-driven lung cancers.^[30] Known covalent ligands bind to a highly flexible allosteric pocket, which traps KRas^{G12C} in the inactive GDP-bound state (thereby confirming its druggability).^[31] Additionally, covalent ligands can specifically target the mutated KRas^{G12C}, sparing the wildtype protein and offering the opportunity for oncogene-specific inhibition.^[30]

In Table 1, the various H-bond interactions are displayed for 10 KRas^{G12C} structures containing a covalently bound acrylamide ligand, as well as the W_{OB} values obtained for each interaction on the reference ligand. For the remaining two of the 12 selected structures (PDB IDs 4M21 and 6ARK), no H-bonds could be reliably identified. An interaction was used in DUckCov if the work required to break the H-bond was higher than 6 kcal mol⁻¹. Thus, structures 5F2E (pulling from atoms R68-

Table 1. Interaction patterns for the selected KRas^{G12C} structures.^[a]

PDB ID	LIG	Chain	R68 NH2 [Don]	K16 NZ [Don]	E63 O [Acc]	E63 OE2 [Acc]	D69 OD1 [Acc]	H95 NE2 [Don]
4M22	22C	B		6.5				
5F2E	5UT	A	10	X	10		13	
5V6S	8YD	A		X			9.2	X
5V71	8ZG	A		X				X
5V9L	91D	A		X				X
5V9O	91G	A		10		12	21	
5V9U	91S	A						X
5YXZ	94C	A		X				X
5YY1	94F	A		X				X
6B0V	C8G	A	X			X		

[a] X indicates that the calculated DUck W_{OB} value was < 6 kcal mol⁻¹, otherwise the value corresponds to the work necessary to break the H-bond during the DUck simulation in kcal mol⁻¹.

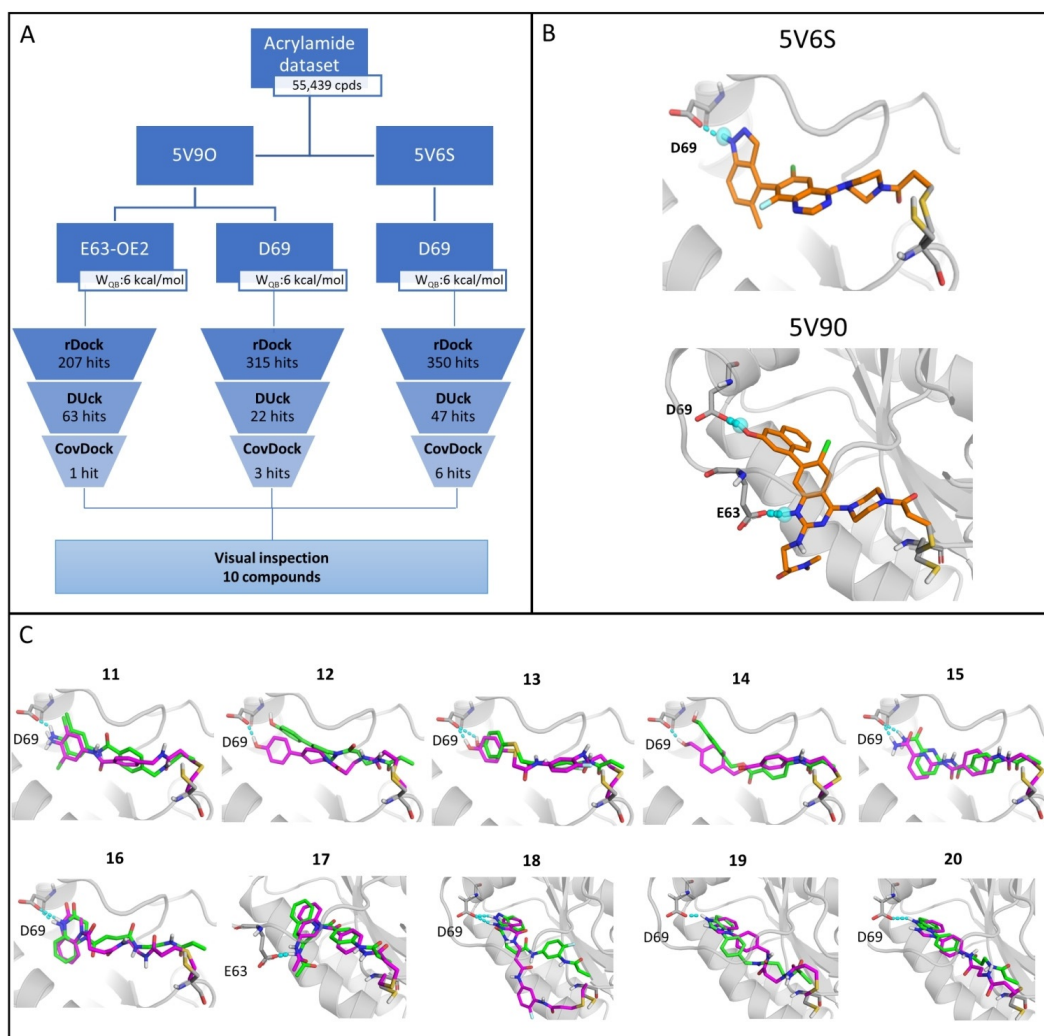


Figure 3. A) DUckCov workflow for KRas^{G12C} for the structures and interactions that eventually led to virtual hits. B) Pre-reaction reference ligands in 5V6S and 5V90 in orange (warhead used for tethering), with covalent attachment in grey, and pharmacophoric features for docking/interaction for DUck in cyan. C) Compounds resulting from DUckCov workflow against 5V6S (11–16) and 5V90 (17–20). In the latter case, compound 17 was retrieved considering the feature/interaction with E63, while compounds 18–20 were retrieved considering the feature/interaction with D69. rDock poses and CovDock poses are shown in green and magenta, respectively.

NH2, E63-O and D69-OD1), 5V90 (pulling from atoms K16-NZ, E63-OE2 and D69-OD1), and 5V6S (pulling from atom D69-OD1) were used to validate the DUckCov protocol on KRas^{G12C}.

In Figure 3A, the workflow is summarized for the two structures and three interactions that led to virtual hits, namely 5V90, E63-OE2 and D69-OD1, and 5V6S, D69-OD1. Finally, 63, 22 and 47 hits were generated by DUckCov from these interaction points, respectively. Based on the results of JAK3, the compounds that were selected for experimental testing were en-

sured to have maintained the inspected H-bond and/or displayed similar binding modes according to rDock and CovDock. The RMSD between rDock and CovDock poses and diversity of the hits were also used to support the final selection (Figure 3C).

In Figure 3C, the compounds retrieved using the stepwise workflow are shown. Compounds 11, 13, 15 and 16, maintain the defined interaction with D69 of the KRas structure 5V6S, in both rDock and CovDock poses. It should be noted that the

pharmacophoric restriction in rDock has a tolerance of 1 Å relative to the reference coordinates. As a result, in some cases the input geometry for DUck does not form a hydrogen bond. Yet, the initial step in the DUck protocol involves a minimization that can repair the H-bond. For this reason, as shown in compounds **12** and **14**, some interactions present high W_{OB} values even though they were not recapitulated by rDock. These interactions are also present in the CovDock pose.

The overall conservation of the binding modes in all ten compounds is reasonably good, in accordance with the RMSD values between the rDock and CovDock poses. RMSD values, CovDock affinity and rDock SCORE.INTER (interaction score energy) scores, as well as W_{OB} values, for the final ten compounds are reported in Supporting Information Table S10, along with their ZINC codes. These compounds were purchased and tested in HSQC NMR measurements (except for compound **14** that was not available for immediate purchase at the time of this study). The 2D structures are included in Figure 4B for the confirmed hits, and Supporting Information Figure S11 for the rest. Strain energies of the resulting binding modes are reported in Supporting Information Table S12.

The site specific binding of compounds **11**, **13**, **17** and **19** was confirmed by $^1\text{H},^{15}\text{N}$ -HSQC (2D NMR) measurements. After the appropriate incubation time, changes in the HSQC spectrum were detected based on chemical shift perturbation confirming the binding of the mentioned small molecules (see Figure 4A, and Supporting Information Figure S13 for the full spectra). The perturbed chemical shifts are located mostly in the well-known Ras functional regions: the P-loop (G10-S17), Switch I (D30-D38), and Switch II (L56-G77). Overall, the findings suggest that the compounds can bind to KRas^{G12C} covalently at the C12 residue, and are located in the allosteric binding pocket of KRas, similarly to known inhibitors such as ARS-853^[31] and ARS-1620.^[32] It is worth to note that from the three protein structure/H-bond combinations that were applied for the DUckCov workflow (Figure 3A), all of them have produced at least one confirmed hit compound.

The 44% hit rate retrieved for KRas using the DUckCov protocol is exceptional, considering the lower druggability of this target. Moreover, the correlation between affinity and activity remains elusive.^[33] Considering this, the protocol was successful in identifying four novel KRas covalent binders in an efficient manner by first focusing on the noncovalent interactions, even though these are known to be non-conserved. Furthermore, a dedicated covalent docking program is imperative for the evaluation of possible rearrangements of the overall binding mode generated by rDock. By sampling multiple rotameric states of the reactive residue, forming the covalent bond between the reactive atoms and performing structural optimization of the covalently attached ligand, the binding mode prediction module in CovDock could increase the chance to find an optimal geometry for the ligands. Additionally, a comparison with Schrödinger's CovDock Virtual screening module clearly highlights the advantage of the DUckCov workflow, as three out of the four confirmed hits were not included in the top ten virtual hits by CovDock (Supporting Information Table S14).

Conclusions

DUckCov is presented here as a novel protocol for the identification of covalent binders that models every stage of the multistep binding mechanism of covalent ligands in an efficient hierarchical manner. In this protocol, only molecules that can form a stable and productive pre-reactive state are evaluated before assessing the post-reactive state, thereby allowing to explore large chemical spaces. Dynamic Undocking (DUck) is the main feature of the workflow, as it is used to analyze the pre-reactive state of the ligands by evaluating the strength of H-bond interactions driving the formation of the initial noncovalent protein–ligand complex. Furthermore, DUck calculations are performed on focused protein chunks, thus enabling fast simulations by decreasing the size of the system. Therefore, DUckCov relies on DUck as a stringent and efficient filter for the selection of molecules to be subjected to the following steps. Next, the post-reactive state is analyzed by performing covalent docking with CovDock in the most accurate pose prediction module. This step is used to generate bound conformations, thus allowing to compare binding modes in the pre- and post-reaction states, and to assess if the key H-bonds are maintained when the ligand is covalently bound to the targeted nucleophilic residue.

Our protocol was successfully validated in two case studies. For JAK3, we reported a hit rate of 60% (three actives out of five molecules tested), identifying two novel, low micromolar and high nanomolar ligands, as well as a low nanomolar inhibitor, originally developed for another kinase target (EGFR). For the more challenging KRas^{G12C} protein target, four novel covalent ligands were experimentally confirmed out of nine tested. Due to the highly flexible nature of the KRas^{G12C} allosteric binding pocket, the resulting 44% hit rate can be considered exceptionally good. The two case studies display the broad applicability of DUckCov, in identifying novel chemical matter for structurally better characterized, as well as more challenging targets. It is also important to highlight that, depending on the available resources, the presented workflow can provide a roughly ten- to hundred-fold speedup, as compared to a commercially available virtual screening tool for covalent binders (Schrödinger CovDock).

Experimental Section

Target structure selection: For JAK3, the PDB structure 5TOZ (chain A) was used as the template, using the co-crystallized inhibitor PF-06651600 as the reference ligand. At the moment of selection, eight structures were available containing covalently bound ligands having a terminal acrylamide as warhead (4QPS, 4V0G, 4Z16, 5TOZ, 5TTS, 5TTU, and 5TTV). Alignment and superposition of these structures in MOE^[20] led to an average RMSD of 0.78 Å. Given the structural conservation of the JAK3 kinase, the choice for structure 5TOZ was based on its co-crystallized ligand having the best inhibitory potency (0.4 nM).^[34]

For the second case study, a structure ensemble approach was used, due to the pronounced flexibility of the KRas allosteric binding site. At the end of May 2018, 23 KRas structures containing covalently bound ligands had been deposited in the PDB. The majori-

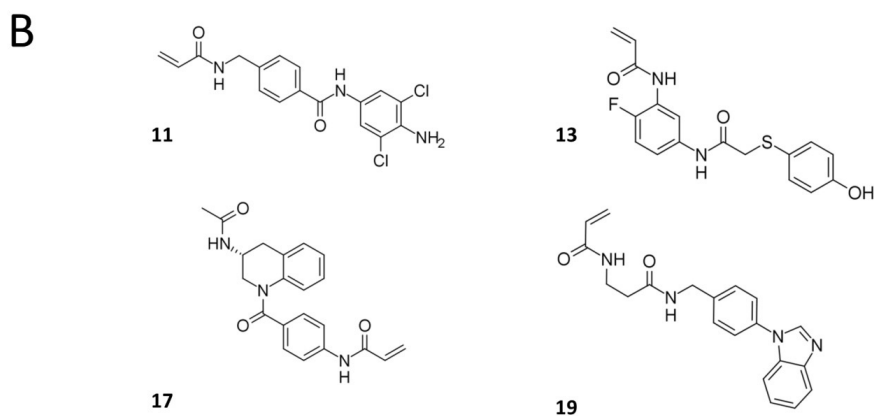
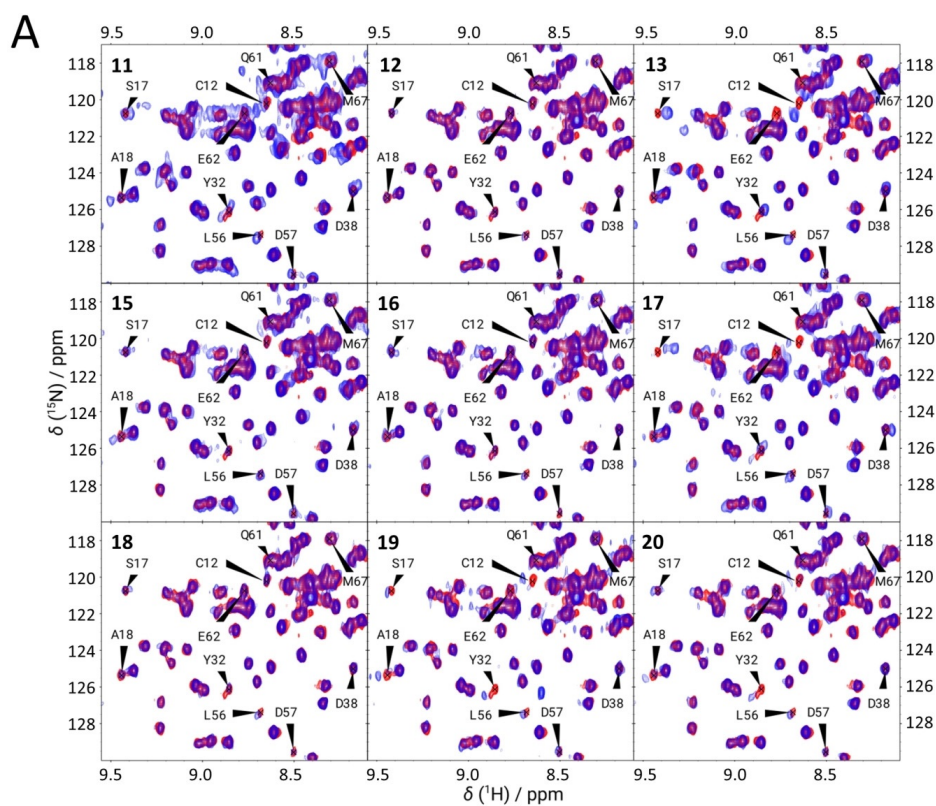


Figure 4. A) ^1H , ^{15}N -HSQC spectra of the tested nine molecules and KRas $^{\text{G12C}}$, showing the spectral region of C12 and other binding site residues (blue), overlaid on the reference spectrum of the free protein (red), after incubation. For compounds 11, 13, 17 and 19, most of the highlighted residues from the P-loop (C12, S17, A18), Switch I (Y32, D38) and Switch II (L56, D57, Q61, E62, M67) regions are significantly perturbed, while almost no changes are detected for 12, 16 and 18. Small (inconclusive) changes are detected for compounds 15 and 20. B) 2D structures of the experimentally confirmed hit compounds against KRas $^{\text{G12C}}$.

ty of unique ligands formed a covalent bond via Michael addition (20/23). Of these, 12 were acrylamide-based covalent ligands, while eight contained a vinylsulfonamide moiety as the electrophile. As for the remaining three complexes, one was formed via a ring opening reaction (5V6V) and the other two were formed through disulfide formation (4LUC, 4LV6). We limited the set to the 12 acrylamide-based complexes due to the limited commercial availability of screening compounds able to react via ring opening or disulfide formation, and the narrower chemical space of available vinylsulfonamides (roughly 2000 purchasable compounds in ZINC, versus 50 000 purchasable acrylamides^[16]). The average RMSD of these 12 structures in the flexible switch II loop was 2.41 Å (Supporting Information Table S1). From these 12, those structures in which at least one H-bond with the co-crystallized ligand was stronger than 6 kcal mol⁻¹ (as evaluated by DUck) were selected for the ensemble approach in DUckCov.

Protein structure preparation (in silico): All of the selected PDB structures were prepared in MOE as follows: I) the structure was corrected (termini were capped, gaps were capped or a homology of the sequence of a similar structure was built, alternate conformations were chosen if more than one was present, correct tautomeric states for the residues were assigned); II) the structure (including the covalently bound ligand) was protonated at pH 7; III) the covalent bond between the residue and the ligand was manually broken; IV) the ligand warhead in its pre-reaction form was built with MOE builder by making the acrylamide's C α -C β bond double, and finally; V) the cysteine was rebuilt, then minimized in the presence of the pre-reaction ligand.

The structures for the CovDock simulations were prepared with the Protein Preparation Wizard provided in the Schrödinger Suite,^[35,36] in order to further refine the protein's H-bond network and to perform a restrained minimization of hydrogen atoms. The receptor grid box required for docking calculations was centered on the corresponding co-crystallized ligand.

Datasets for VS: As one of the main features of the protocol is its efficiency, it is most beneficial when a large collection of electrophilic ligands is available. In general, if the protein has already been targeted by covalent inhibitors, the library can be compiled by collecting commercially available ligands and/or by enumerating synthetically accessible compounds bearing the same warhead type as the crystallized inhibitor. For JAK3, the vast majority of known covalent inhibitors bind through an acrylamide warhead, while for KRas^{G12C}, most of the known potent covalent inhibitors bind through either an acrylamide or a vinylsulfonamide warhead. Further on, due to the limited commercial availability of screening compounds containing other types of warheads, we used the acrylamide dataset (roughly 50 000 compounds) collected by London et al. (for testing their recently published covalent docking program, DOCKovalent), to validate our protocol.^[16] Prior to docking simulations, LigPrep by Schrödinger was used to prepare 3D conformations from SMILES codes and to generate tautomeric and ionization states at pH 6–8 while retaining specified chiralities.^[35,36]

General workflow description: An overview of the protocol is shown in Figure 1. The collected library (here: ZINC acrylamide collection) is first docked with rDock against the target of interest, while simultaneously tethering the covalent warhead to its reference coordinates and using pharmacophoric constraints to enforce the main noncovalent interaction. Because the protein structure is derived from a crystallized covalent complex, a distance cutoff is set to avoid large deviations of the electrophilic warhead from the position defined in the reference ligand. H-bond pharmacophoric

constraints are applied in the docking simulation to keep only those ligands that can establish the H-bond interactions defined as important for binding. DUck is then used to evaluate the strength of the H-bond. For DUck, only H-bonds are assessed, as they are known to be key contributors to affinity in many targets.^[37,38] In a DUck simulation, the ligands are pulled from 2.5 to 5.0 Å relative to the defined H-bond interaction point in the protein, during a user-defined number of MD and SMD replicas. The force necessary to pull out the ligand is then used to calculate a work value (W_{08}), which corresponds to the strength of the H-bond. What makes DUck exceptionally fast, is that only the local environment of the residue involved in the interaction is required for the simulation. Lastly, only those ligands that display the best noncovalent interactions (according to rDock and DUck) are covalently docked with CovDock using the most accurate pose prediction module.

Tethered and constrained docking: Using rDock, the cavity was prepared with the reference ligand method, using the respective co-crystallized ligand. Tethered docking was used to restrain the electrophile close to the reactive residue. Tethered docking consists of two steps, namely, I) superposing atoms according to the defined SMARTS pattern, and II) docking, during which the superposed atoms can only deviate from the original position by a user-defined cutoff. The warhead was defined by the SMARTS pattern "[#6]=[#6]-[#6]=[#8]" for the acrylamide motif. During docking, the tethered part of the ligand could move freely in terms of the dihedral degrees of freedom, while the translational and rotational degrees of freedom could deviate by max. 0.1 Å per docking run. This is meant to allow some flexibility in the sampling, also taking into consideration that the targeted cysteine residue could display a significant degree of flexibility.

Furthermore, a pose is penalized if the defined pharmacophoric constraints are not met. Here, a 1 Å deviation was permitted to increase sampling, considering that the strength of the H-bond would still be assessed by DUck later on. For JAK3, the pharmacophoric constraints were defined as an acceptor (E903 backbone C=O) and a donor (L905 backbone NH), both of which interact with the reference ligand in 5TOZ. For KRas^{G12C}, the pharmacophoric constraints were defined based on the H-bond interactions observed between the reference ligands and the protein in the selected structures: these interactions (along with their W_{08} values evaluated by DUck) are summarized in Table 1.

Next, the high-throughput VS protocol (HTVS) of rDock was implemented, which consisted of three docking stages for each ligand. In each stage, the number of docking runs increases (for better sampling), and the threshold for the docking score decreases (better scores). The ligand only proceeds to the next stage if its docking score is better than the defined threshold within the specified number of runs. This is done to increase the efficiency of the simulation by progressively decreasing the number of ligands moved forward. The docking score filters were selected based on the score of the reference ligands, while being stricter for JAK3 (as the defined noncovalent interactions necessary for binding are well known) and less strict for KRas (as the defined noncovalent interactions necessary for binding are not known). For the same reason, the total number of docking runs for JAK3 was significantly lower than for KRas. The exact HTVS protocols are given in Supporting Information Schemes S2 (for JAK3) and S3 (for KRas) along with the in-place rDock SCORE.INTER scores in Supporting Information Table S4.

Dynamic undocking: The first step for a DUck simulation is the definition of the chunk (a part of the protein structure) that repre-

sents the local environment surrounding the residue interacting with the ligand. Thus, for every interaction point, a separate chunk is created. When selecting residues for the chunk, the following guidelines were considered: I) selecting as little residues as possible to reduce computational time; II) residues were not selected if they would block the ligand from exiting the pocket during the simulations based on the directionality of the H-bond; III) residues were not removed if this would lead to the possibility of solvent entering the pocket from areas other than where the ligand is exiting; and lastly IV) preserving the local environment. This was done from the already prepared structures. The gaps created during the process of selecting the chunk residues were capped. For this, each section of residues was split into separate chains, and the termini of each chain were acetylated or methylated. Lastly, the chunk was checked for clashes possibly created during the capping of the chains. The corresponding chunk definitions for JAK3 and KRas are included in Supporting Information Table S5.

After production of the chunk, DUCK automatically does the following: I) automatic ligand parameterization in MOE, II) minimization, III) equilibration, and IV) a series of SMD (at two different temperatures), then MD simulations, in which the ligand is pulled from 2.5 to 5.0 Å relative to the defined H-bond interaction. Steps II) to IV) were performed with GPU-based pmemd.cuda in AMBER.^[39] Five replicas of step IV) were performed, during which a W_{QB} threshold of 6 kcal mol⁻¹ (force necessary to pull the ligand) was maintained, so that the simulation was stopped if the measured W_{QB} value of the H-bond was smaller. Additionally, for KRas, the inclusion of co-crystallized GDP in the chunk was necessary, as its absence would have led to the surface being more exposed to bulk solvent. For this, GDP was parameterized using MOE's PFROST forcefield, and the generated parameters were automatically included in the DUCK protocol.

Covalent docking with CovDock: Because Schrödinger's CovDock pose prediction module outperformed most of the other covalent docking tools, we used this approach to rank and predict the binding mode of the virtual hits identified to have strong noncovalent interactions by the previous workflow steps.^[23] CovDock ranks the compounds according to an "Affinity Score", which is calculated as the average of the pre-reaction Glide score and the post-reaction in-place docking score. By deeming the energy of bond formation as constant across a set of compounds having the same warhead involved in the chemical reaction (as in the case of DUCKov), the affinity score can be used to compare and rank ligands in a set. CovDock affinity scores for the reference ligands of the structures that led to hits are given in Supporting Information Table S4. Contrary to the first docking step in the workflow, no additional restraints were applied in the covalent docking simulation other than those used by default in CovDock. Binding site residues were defined by centering the receptor grid on the ligand co-crystallized in the structure under investigation. When setting up the simulation, the acrylamide warhead was automatically recognized in each ligand structure through the SMARTS-based definition of the "Michael addition" reaction type. Ultimately, this step was incorporated to evaluate if a change in binding mode would take place upon covalent bond formation, which would prevent the ligand from establishing the interactions defined as necessary by previous workflow steps. To that end, root-mean-squared deviation (RMSD) values between the rDock and CovDock conformations were calculated by means of a Python script provided by Schrödinger (rmsd.py). A small RMSD, typically lower than 2.0 Å, was considered as favorable. Furthermore, the defined interaction patterns were also visually inspected for consensus.

Biochemical and structural characterization of the identified virtual hits: Compounds **1**, **3**, **5**, **8** and **9** were tested at 10 μM in duplicate with the Z'-LYTE kinase inhibition assay (Life Technologies). The assay uses a fluorescence-based format and is based on the different sensitivity of phosphorylated and non-phosphorylated peptides to proteolytic cleavage. A suitable peptide substrate is labeled with two fluorophores (coumarin and fluorescein), forming a FRET pair. After incubating the kinase + peptide + test compound mixture for an hour, a development reaction is carried out. Any peptide that was not phosphorylated by the kinase is cleaved, disrupting the resonance energy transfer between the FRET pair. The reaction progress is quantified based on the ratio of the detected emission at 445 nm (coumarin) and 520 nm (fluorescein), that is, the ratio of cleaved versus intact peptide. A more detailed description of the assay is available on the website of Life Technologies.^[40] IC₅₀ values were determined from 10 points titration measurements using the same assay.

Binding of compounds **11–13** and **15–20** to KRas^{G12C} was tested and structurally characterized by NMR measurements, performed on a Bruker Avance III 700 MHz spectrometer equipped with a 5-mm Prodigy TCI H&F-C/N-D, z-gradient probe head operating at 700.05 MHz for ¹H and 70.94 MHz for ¹⁵N nuclei. ¹H, ¹⁵N-HSQC spectra were recorded at 298 K to obtain the protein ¹H and ¹⁵N resonances in both free and small-molecule-bound state and the changes in chemical shifts were followed upon complex formation. NMR samples contained ¹⁵N-labeled KRas^{G12C} (catalytic domain, residues 1–169) in 150 and 50 μM concentration in free protein measurement (as a reference) and binding test, respectively, 5 mM GDP, 10 mM EDTA, 15 mM MgCl₂ in PBS buffer, 5% DMSO and 10% D₂O at pH 7.4 and 150–500 μM ligand. Because some of the ligands were not fully dissolved, we used a longer incubation time (96 h), and a high number of scans for every HSQC spectrum (NS = 128). To avoid false-positive results, the free protein was incubated for four days as well, and the spectra of the samples were compared with the spectrum of the incubated free protein. All ¹H chemical shifts were referenced to the DMSO peaks (which were calibrated to DSS resonance before in free protein measurements) as DSS was not added to avoid any side reactions with the ligand. ¹⁵N chemical shift values were referenced indirectly using the corresponding gyromagnetic ratios according to the IUPAC convention. Sequence-specific assignments of H^N and N in the bound KRas^{G12C} spectra were transferred from our results to be published elsewhere (BMRB entry code: 27646). There were ambiguities in a number of resonances in crowded spectral regions; however, this fact did not influence the final outcome. All spectra were processed with Bruker TOPSPIN and analyzed using NMRFAM-SPARKY software.^[41]

Acknowledgements

This study was supported by the MSCA ITN FRAGNET (project 6758993) grant to M.R., A.S, X.B., and G.M.K., and by the National Research, Development and Innovation Office of Hungary (OTKA, grant number K116904; and NVKP_16, grant number NVKP_16-1-2016-0020). Additionally, X.B. is supported by grants SAF2015-68749-R (Spanish Ministerio de Economía) and 2014 SGR 1189 (Catalan government). This work was completed in the ELTE Institutional Excellence Program (783-3/2018/FEKUTSRAT) supported by the Hungarian Ministry of Human Capacities and co-financed by the European Regional Development Fund (VEKOP-2.3.2-16-2017-00014 and VEKOP-2.3.3-15-2017-00018).

Conflict of interest

The authors declare no conflict of interest.

Keywords: covalent docking · dynamic undocking · targeted covalent inhibitors · virtual screening

- [1] J. G. Robertson, *Biochemistry* **2005**, *44*, 5561–5571.
- [2] R. A. Bauer, *Drug Discovery Today* **2015**, *20*, 1061–1073.
- [3] J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, *Nat. Rev. Drug Discovery* **2011**, *10*, 307–317.
- [4] B. F. Cravatt, A. T. Wright, J. W. Kozarich, *Annu. Rev. Biochem.* **2008**, *77*, 383–414.
- [5] E. Weerapana, C. Wang, G. M. Simon, F. Richter, S. Khare, M. B. D. Dillon, D. A. Bachovchin, K. Mowen, D. Baker, B. F. Cravatt, *Nature* **2010**, *468*, 790–797.
- [6] Q. Liu, Y. Sabnis, Z. Zhao, T. Zhang, S. J. Buhrlage, L. H. Jones, N. S. Gray, *Chem. Biol.* **2013**, *20*, 146–159.
- [7] Z. Zhao, Q. Liu, S. Bliven, L. Xie, P. E. Bourne, *J. Med. Chem.* **2017**, *60*, 2879–2889.
- [8] D. S. Johnson, E. Weerapana, B. F. Cravatt, *Future Med. Chem.* **2010**, *2*, 949–964.
- [9] R. Mah, J. R. Thomas, C. M. Shafer, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 33–39.
- [10] C. Jöst, C. Nitsche, T. Scholz, L. Roux, C. D. Klein, *J. Med. Chem.* **2014**, *57*, 7590–7599.
- [11] R. Lonsdale, J. Burgess, N. Colclough, N. L. Davies, E. M. Lenz, A. L. Orton, R. A. Ward, *J. Chem. Inf. Model.* **2017**, *57*, 3124–3137.
- [12] H. M. Kumalo, S. Bhakat, M. E. S. Soliman, *Molecules* **2015**, *20*, 1984–2000.
- [13] G. Jones, P. Willett, R. Glen, A. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727–748.
- [14] G. Bianco, S. Forli, D. S. Goodsell, A. J. Olson, *Protein Sci.* **2016**, *25*, 295–301.
- [15] X. Ouyang, S. Zhou, C. T. T. Su, Z. Ge, R. Li, C. K. Kwok, *J. Comput. Chem.* **2013**, *34*, 326–336.
- [16] N. London, R. M. Miller, S. Krishnan, K. Uchida, J. J. Irwin, O. Eidam, L. Gibold, P. Cimermančić, R. Bonnet, B. K. Shoichet, et al., *Nat. Chem. Biol.* **2014**, *10*, 1066–1072.
- [17] R. Abagyan, M. Totrov, K. Dmitry, *J. Comput. Chem.* **1994**, *15*, 488–506.
- [18] C. R. Corbeil, P. Englebienne, N. Moitessier, *J. Chem. Inf. Model.* **2007**, *47*, 435–449.
- [19] C. Scholz, S. Knorr, K. Hamacher, B. Schmidt, *J. Chem. Inf. Model.* **2015**, *55*, 398–406.
- [20] Molecular Operating Environment (MOE) 2016.08, Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, H3A 2R7 (Canada), **2018**.
- [21] K. Zhu, K. W. Borrelli, J. R. Greenwood, T. Day, R. Abel, R. S. Farid, E. Harder, *J. Chem. Inf. Model.* **2014**, *54*, 1932–1940.
- [22] D. Toledo Warshaviak, G. Golan, K. W. Borrelli, K. Zhu, O. Kalid, *J. Chem. Inf. Model.* **2014**, *54*, 1941–1950.
- [23] A. Scarpino, G. G. Ferenczy, G. M. Keserü, *J. Chem. Inf. Model.* **2018**, *58*, 1441–1458.
- [24] S. Ruiz-Carmona, D. Alvarez-Garcia, N. Foloppe, A. B. Garmendia-Doval, S. Juhos, P. Schmidtke, X. Barril, R. E. Hubbard, S. D. Morley, *PLoS Comput. Biol.* **2014**, *10*, e1003571.
- [25] S. Ruiz-Carmona, P. Schmidtke, F. J. Luque, L. Baker, N. Matassova, B. Davis, S. Roughley, J. Murray, R. Hubbard, X. Barril, *Nat. Chem.* **2017**, *9*, 201–206.
- [26] P. Schmidtke, F. Javier Luque, J. B. Murray, X. Barril, *J. Am. Chem. Soc.* **2011**, *133*, 18903–18910.
- [27] Y. Rochman, R. Spolski, W. J. Leonard, *Nat. Rev. Immunol.* **2009**, *9*, 480–490.
- [28] H. Cheng, S. K. Nair, B. W. Murray, C. Almaden, S. Bailey, S. Baxi, D. Behenna, S. Cho-Schultz, D. Dalvie, D. M. Dinh, et al., *J. Med. Chem.* **2016**, *59*, 2005–2024.
- [29] J. M. Ostrem, U. Peters, M. L. Sos, J. A. Wells, K. M. Shokat, *Nature* **2013**, *503*, 548–551.
- [30] C. I. Nnadi, M. L. Jenkins, D. R. Gentile, L. A. Bateman, D. Zaidman, T. E. Balius, D. K. Nomura, J. E. Burke, K. M. Shokat, N. London, *J. Chem. Inf. Model.* **2018**, *58*, 464–471.
- [31] M. P. Patricelli, M. R. Janes, L. S. Li, R. Hansen, U. Peters, L. V. Kessler, Y. Chen, J. M. Kucharski, J. Feng, T. Ely, et al., *Cancer Discovery* **2016**, *6*, 316–329.
- [32] M. R. Janes, J. Zhang, L. S. Li, R. Hansen, U. Peters, X. Guo, Y. Chen, A. Babbar, S. J. Firdaus, L. Darjania, et al., *Cell* **2018**, *172*, 578–589.e17.
- [33] J. M. Jansen, C. Wartchow, W. Jahnke, S. Fong, T. Tsang, K. Pfister, T. Zavorotinskaya, Di. Bussiere, J. M. Cheng, K. Crawford, et al., *PLoS One* **2017**, *12*, e0174706.
- [34] J. B. Telliez, M. E. Dowty, L. Wang, J. Jussif, T. Lin, L. Li, E. Moy, P. Balbo, W. Li, Y. Zhao, et al., *ACS Chem. Biol.* **2016**, *11*, 3442–3451.
- [35] *Small-Molecule Drug Discovery Suite* 2018–2013, Schrödinger LLC, New York, NY (USA), **2018**.
- [36] G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *J. Comput. Aided. Mol. Des.* **2013**, *27*, 221–234.
- [37] E. Nittinger, T. Inhester, S. Bietz, A. Meyder, K. T. Schomburg, G. Lange, R. Klein, M. Rarey, *J. Med. Chem.* **2017**, *60*, 4245–4257.
- [38] S. Raschka, A. J. Wolf, J. Bemister-Buffington, L. A. Kuhn, *J. Comput. Aided. Mol. Des.* **2018**, *32*, 511–528.
- [39] D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York, P. A. Kollman (2018), AMBER 2018, University of California, San Francisco.
- [40] Z'-LYTE Kinase Assay Kits: <http://www.lifetechnologies.com/hu/en/home/life-science/drug-discovery/target-and-lead-identification-and-validation/kinasebiology/kinase-activity-assays/z-lyte.html> (accessed October 25, 2018).
- [41] W. Lee, M. Tonelli, J. L. Markley, *Bioinformatics* **2015**, *31*, 1325–1327.

Manuscript received: February 5, 2019

Accepted manuscript online: February 20, 2019

Version of record online: March 8, 2019

Discovery of a novel kinase hinge binder fragment by dynamic undocking

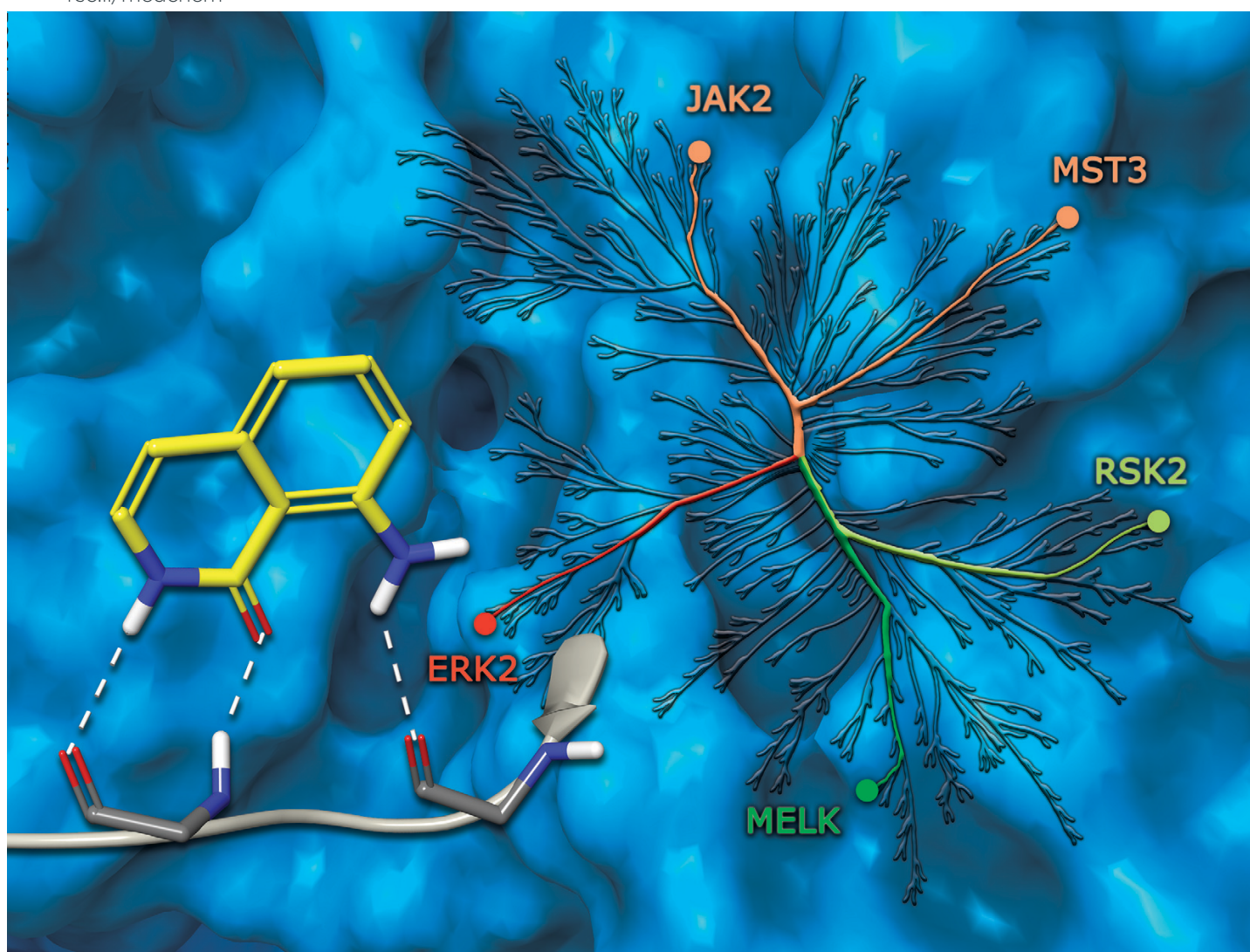
Moira Rachman^{1,2}, Dávid Bajusz², Anasztázia Hetényi³, Andrea Scarpino², Balázs Merő⁴,
Attila Egyed², László Buday⁴, Xavier Barril^{1,5} and György M. Keserű²

In *RCS Medicinal Chemistry* (2020), 11, 552-8

1. Institut de Biomedicina de la Universitat de Barcelona (IBUB) and Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII 27-31, 08028 Barcelona, Spain.
2. Medicinal Chemistry Research Group, Research Centre for Natural Sciences, Magyar Tudósok Körútja 2, Budapest 1117, Hungary.
3. Department of Medical Chemistry, University of Hungary.
4. Signal Transduction and Functional Genomics Research Group, Research Centre for Natural Sciences, Budapest, Hungary
5. Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.

RSC Medicinal Chemistry

rsc.li/medchem



ISSN 2632-8682

RESEARCH ARTICLE

View Article Online

View Journal | View Issue

Cite this: *RSC Med. Chem.*, 2020, **11**, 552

Discovery of a novel kinase hinge binder fragment by dynamic undocking

Maira Rachman,^{a,b} Dávid Bajusz,^b Anasztázia Hetényi,^c Andrea Scarpino,^b Balázs Merő,^d Attila Egyed,^b László Buday,^d Xavier Barril^{a,e} and György M. Keserő^{b,*}

One of the key motifs of type I kinase inhibitors is their interactions with the hinge region of ATP binding sites. These interactions contribute significantly to the potency of the inhibitors; however, only a tiny fraction of the available chemical space has been explored with kinase inhibitors reported in the last twenty years. This paper describes a workflow utilizing docking with rDock and dynamic undocking (DUck) for the virtual screening of fragment libraries in order to identify fragments that bind to the kinase hinge region. We have identified 8-amino-2*H*-isoquinolin-1-one (**MR1**), a novel and potent hinge binding fragment, which was experimentally tested on a diverse set of kinases, and is hereby suggested for future fragment growing or merging efforts against various kinases, particularly MELK. Direct binding of **MR1** to MELK was confirmed by STD-NMR, and its binding to the ATP-pocket was confirmed by a new competitive binding assay based on microscale thermophoresis.

Received 6th November 2019,
Accepted 11th February 2020

DOI: 10.1039/c9md00519f

rsc.li/medchem

Introduction

Protein kinases are a major class of drug targets, with over 30 marketed drugs and 250 drug candidates undergoing clinical studies, mainly but not only in various oncology indications.¹ The majority of kinase inhibitors target ATP binding sites (type I), which contain a conserved hinge region. For this reason, one of the main challenges in the development of kinase inhibitors is obtaining selectivity. Another major challenge is avoiding an already congested IP space, as the diversity of hinge binder scaffolds used in type I kinase inhibitors is relatively limited.² Consequently, these chemotypes have barely been sampled, which prompts the thorough exploitation of their chemical space.^{3,4} For example, it has been shown recently that only 1% of potential hinge binders are present in known kinase inhibitors.⁵

Fragment-based drug discovery (FBDD) has already been used in many kinase programs⁴ and has the potential to

identify novel fragment-sized hinge binders and specifically “evolve” the fragment for the targeted kinase.^{6,7} In fact, the first FBDD derived drug on the market, vemurafenib, targets the oncogenic V600E mutant of the B-Raf kinase.⁸ Structure-based virtual screening approaches have been used effectively for both the identification of novel fragment-sized hinge binders and their optimization. In one case study, Kolb and colleagues screened 730 000 compounds and discovered two ligands with different hinge binder moieties. The initial set was first filtered by kinase hinge binding pharmacophore restraints and the resulting 21 418 compounds were docked to the ATP site.⁹ In another study, Ulrich and colleagues extracted core fragments from 2.3M commercially available compounds. The resulting unique fragments were filtered for kinase hinge pharmacophores, and were subsequently docked into a panel of protein kinases. This strategy identified a number of hinge binder fragments with no previously reported activity against the investigated kinases.¹⁰

In this work, we used a novel screening strategy to find potent hinge binding fragments. Our dynamic undocking¹¹ based approach identified **MR1** (Fig. 1A) that has been experimentally validated against five kinase targets (Table 1). Future fragment growing or merging efforts toward these targets could avoid a congested druggable chemical space (e.g. for JAK2) and could provide a suitable chemical starting point to unmet medical needs (e.g. for MELK). JAK2 belongs to the non-receptor tyrosine kinase family, for which the V617F mutation is known to be implicated in myeloproliferative disorders.^{12,13} MST3 belongs to the Ste20

^a Facultat de Farmàcia and Institut de Biomedicina, Universitat de Barcelona, Av. Joan XXIII 27-31, 08028 Barcelona, Spain

^b Medicinal Chemistry Research Group, Research Centre for Natural Sciences, Magyar Tudósok Körútja 2, Budapest 1117, Hungary. E-mail: keseru.gyorgy@ttk.hu

^c Department of Medical Chemistry, University of Szeged, Dóm tér 8, H-6720 Szeged, Hungary

^d Signal Transduction and Functional Genomics Research Group, Research Centre for Natural Sciences, Magyar Tudósok Körútja 2, Budapest 1117, Hungary

^e Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluis Companys 23, 08010 Barcelona, Spain



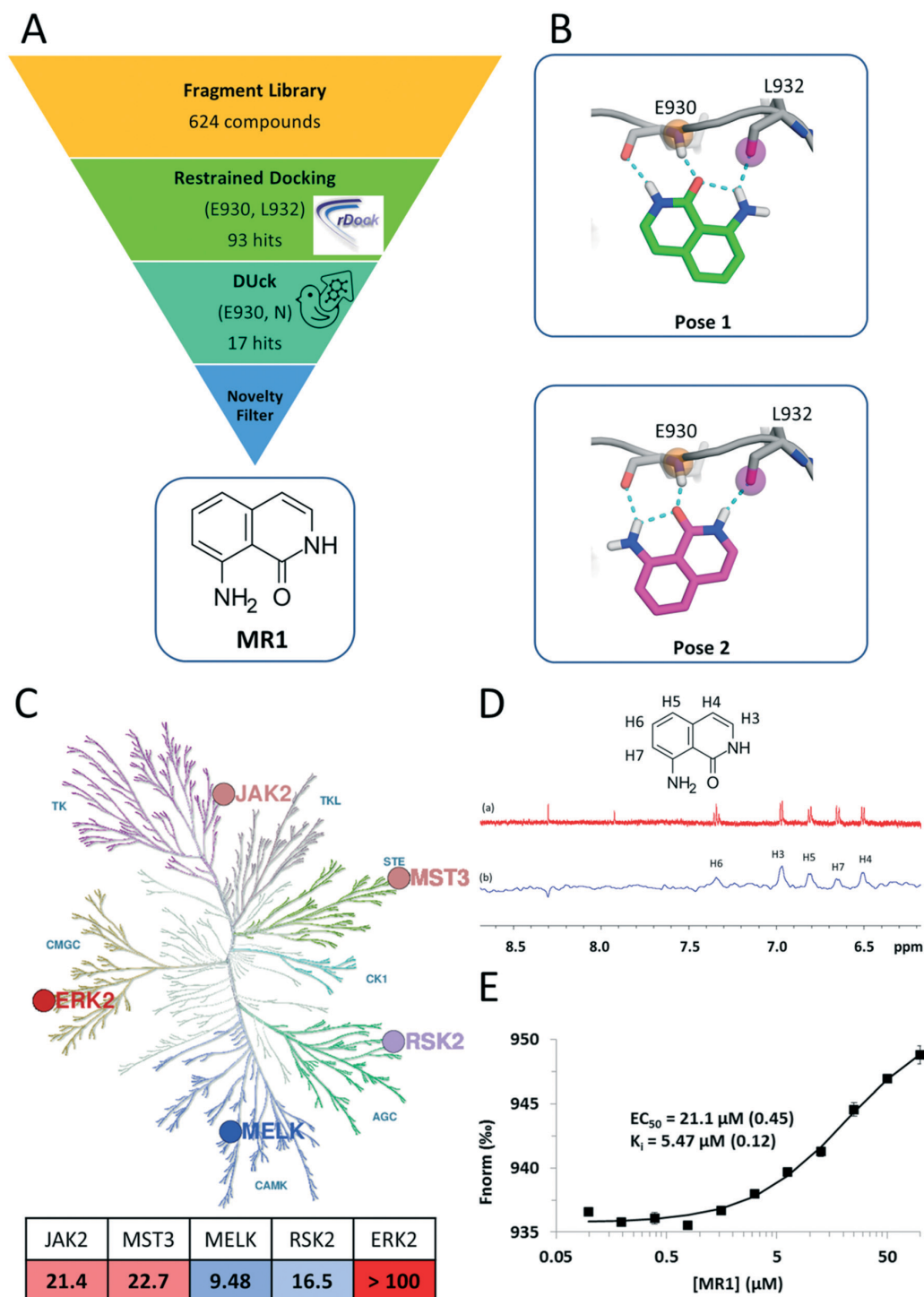


Fig. 1 A) Summary of the virtual screening workflow. B) Predicted binding modes of MR1 in the JAK2 ATP pocket. Both binding modes utilize tridentate H-bond interactions with the backbone carbonyl and amide groups of the hinge residues E930 and L932. C) Kinase inhibitory profile of MR1 overlaid on the kinase phylogenetic tree²⁸ (coloured blue to red, from most active to least active). The table summarizes the respective IC_{50} values in μM units, with the same colouring scheme (illustration reproduced courtesy of Cell Signaling Technology, Inc., www.cellsignal.com). D) ^1H NMR spectrum of MR1 (a) and saturation transfer difference NMR spectrum of MR1 in the presence of MELK (b). E) In the competitive, MST-based binding assay, titration of MELK with MR1 reveals concentration-dependent binding with a K_i value of $5.47 \mu\text{M}$ and confirms the ATP-pocket as the location of binding.



Table 1 Summary of the kinase targets investigated in this study

Branch	Kinase	Uniprot ID	Clinical indications
AGC	RSK2	P51812	Oncogenesis and leukemia ¹⁵
CAMK	MELK	Q14680	Oncogenesis and cancer treatment resistance ¹⁶
CMGC	ERK2	P28482	Cervical ¹⁷ and colorectal ¹⁸ cancer
STE	MST3	Q9Y6E0	Breast cancer ¹⁴
TK	JAK2	O60674	Myeloproliferative neoplasms (MPNs) ^{12,13}

serine/threonine protein kinase family, which has been shown to promote proliferation and tumorigenicity.¹⁴ Maternal embryonic leucine zipper kinase (MELK) is a serine/threonine kinase belonging to the CAMK family, and was initially found to be expressed in a wide range of early embryonic cellular stages.

More recently, MELK has been identified to be present in several human cancers and stem cell populations with a unique spatial and temporal pattern, which suggests a prominent role in cell cycle control, cell proliferation, apoptosis, cell migration, cell renewal, embryogenesis, oncogenesis, and cancer treatment resistance and recurrence.¹⁶ ERK2 is part of the Ras/Raf/MEK/ERK pathway, which is often overactivated in a very wide range of cancers, while RSK2 is a downstream effector of this pathway and phosphorylates substrates involved in transcription, translation, cell cycle regulation, and cell survival.¹⁵

Results and discussion

Here, we present a novel and potent hinge binding fragment, **MR1** (Fig. 1A), that was discovered through virtual screening (VS) of our in-house fragment library¹⁹ using kinase hinge pharmacophore information in a hierarchical VS strategy (Fig. 1A). Our initial screening has been performed on JAK2 as a prototypic protein kinase with a large number of type I inhibitors. It is well known that the kinase hinge region exhibits a pronounced H-bond pattern with ATP or ATP-site competitors and that these interactions are imperative for type I kinase inhibitors to bind. As such, a typical VS strategy would incorporate such information by ensuring that the retained ligands maintain these features through the use of *e.g.* pharmacophoric restraints. Here, we employ an additional filter that emphasizes the importance of the defined H-bonds through dynamic undocking (DUck¹¹), which estimates the robustness of H-bonds. DUck is an orthogonal method to approaches meant to predict binding free energies. Instead, it assesses structural stability through evaluation of the robustness of H-bonds. The VS strategy consisted of docking with rDock²⁰ using pharmacophoric restraints to ensure that the features necessary for H-bonding with the hinge were present, then using DUck to estimate the resistance to rupture of these H-bonds. This strategy led to the discovery of **MR1**, which was tested on five kinases from the major branches of the kinase phylogenetic tree (Fig. 1C), each with several relevant indications (Table 1). MST3 and MELK, in particular, have only few known potent inhibitors.

MR1 was found to inhibit four of the target kinases with IC₅₀ values in the low to mid-micromolar range (corresponding to ligand efficiencies between 0.54–0.59),²¹ which is in line with its small size and other hinge binders of this complexity.⁷ ERK2 was the only kinase that was not inhibited; one possible reason behind this is the higher ATP concentration (100 μM) in the ERK2 inhibition assay, as compared to the rest of the kinases (10–50 μM), resulting in a stronger competition of ATP toward the binding site. The assessment of the binding mode of **MR1** was based on the docking and DUck scores. Two plausible solutions were generated (Fig. 1B), both of which adopt a tridentate interaction with the hinge. This was verified also for MELK: here, pose 2 is clearly preferred (*W*_{QB} value of 7.7 kcal mol⁻¹, *vs.* 1.6 kcal mol⁻¹ for pose 1). The ability of fragments to bind simultaneously in multiple orientations has been reported in numerous occasions, and can even be favourable from an entropic perspective.²² Depending on how the fragment is then elaborated, it freezes into one binding mode or the other,^{23,24} which will likely stay conserved during the elaboration process²⁵ (this also means that the mentioned entropic gain will be lost, but ideally, this is compensated by a larger favourable enthalpy decrease due to additional secondary interactions between the protein and the elaborated ligand). The tridentate interaction is not often observed within kinases (a bidentate interaction being the most common), but – as supported by the inhibition data – a third interaction allows for high versatility.²⁶ This can be constrained at the optimization phase for introducing family-specific substituents. In fact, it is not the hinge binder but rather its decoration that determines kinase selectivity. Consequently, selective kinase inhibitors can stem from an unselective fragment.^{7,27}

In addition to identifying **MR1**, we have observed a stepwise enrichment of known hinge binders along the workflow: while the whole fragment library contained known hinge binders in 14.4% of the compounds (90 out of 624), this ratio was 64.7% for the virtual hits after the DUck calculations (11 out of 17). This can be considered as a retrospective validation of the presented virtual screening workflow.

Next, we aimed to confirm the direct and ATP-competitive binding of **MR1** experimentally. For this purpose, we have selected MELK, as **MR1** displayed the strongest inhibitory activity against this kinase. Direct binding to MELK was confirmed by an STD-NMR measurement (Fig. 1D), while the binding site was validated with a competitive MST-based (microscale thermophoresis) assay developed in our lab (Fig. 1E).



The assay is based on the displacement of a fluorescently labelled type I reference ligand (see the Experimental section) and it has additionally enabled us to quantify the dissociation constant of **MR1** from MELK as 5.47 μM . Its high ligand efficiency ($\text{LE} = 0.67$) nominates **MR1** as a viable starting point for optimizing ATP-site kinase inhibitors.

Experimental

Virtual screening

The structures were prepared using MOE 2016,²⁹ by removing water and cofactors, capping the termini and gaps, and for protonation with default settings. For the virtual screening, the JAK2 PDB structure, 3E64 (chain A) was used.³⁰ For docking, the cavity was defined in the prepared structure by the reference ligand method, using the crystallized ligand as reference. The fragment library consisted of 624 compounds and was prepared with LigPrep,³¹ so that ligands above 300 Da would be ignored, at most eight stereoisomers, six tautomers and eight ring conformers would be generated and lastly, probable ionization states within the pH range of six to eight would be generated.

The prepared library was docked with pharmacophoric restraints at the hinge region, namely, Glu930, O and Leu932, N. The pharmacophore was defined as a 2 Ångström radius around the mentioned receptor atoms. If the feature did not adhere to the positional constraints, rDock would assign a positive (unfavourable) pharmacophore restraint score, for which the cutoff was set to 0.5. Furthermore, a high-throughput VS (HTVS) protocol was implemented, which consisted of three stages, for which at every stage the number of docking runs increases, and the rDock "SCORE.INTER" filter becomes stricter. The filter was adapted to scores expected for fragment-sized molecules which resulted in 93 compounds in total. DUCK was performed on this set of compounds, pulling from Leu932, N (pulling from Glu930, O did not result in any additional filtering). The first step for a DUCK simulation is the definition of the chunk (a part of the protein structure) that represents the local environment surrounding the residue interacting with the ligand. When selecting residues for the chunk, the following guidelines were considered: I) selecting as few residues as possible to reduce computational time, II) residues were not selected if they would block the ligand from exiting the pocket during the simulations based on the directionality of the H-bond, III) residues were not removed if this would lead to the possibility of solvent entering the pocket from areas other than where the ligand is exiting, and lastly IV) preserving the local environment of the interacting atoms in the already prepared structures. The sequence gaps created during the process of selecting the chunk residues were capped. For this, each section of residues was split into separate chains, and the termini of each chain were acetylated or methylated. Lastly, the chunk was checked for clashes possibly created during the capping of the chains. The chunk included the

following residues: 853–859, 861–865, 879–882, 898, 902, 911, 912, 927, 929–941, 976, 978, 980–984 and 993–996. After production of the chunk, DUCK performs I) automatic ligand parameterization in MOE, II) minimization, III) equilibration, and IV) two SMD simulations (at two different temperatures, 300 K and 325 K), in which the distance between the interacting atoms in the ligand and protein is increased from 2.5 to 5.0 Å, and V) if the W_{QB} value (work necessary to break the H-bond) in the previous step reaches a pre-defined threshold, then the system is sampled by a short unbiased MD simulation, after which the resulting new structures are fed into steps IV) and V); the last two steps being repeated in a finite number of cycles (replicas).¹¹ Steps II) to V) were performed with GPU-based pmemd.cuda in AMBER. Here, up to five replicas of steps IV) and V) were performed, during which a W_{QB} threshold of 6 kcal mol⁻¹ (work necessary to break the H-bond) was used, so that the simulations were discontinued if the measured W_{QB} in any replica was below the threshold. If five runs were completed, the lowest obtained W_{QB} value was used, which resulted in 17 compounds that surpassed the threshold. Finally, we checked the novelty of these fragments against known kinase inhibitors by substructure searches, which ultimately led to the selection of compound **MR1** for experimental validation.

To suggest a binding pose for MELK, **MR1** was docked in all liganded MELK structures available. Both poses in Fig. 1B were identified and submitted to a DUCK simulation, pulling from the NH atom in the hinge (Cys 89 N). Chunks were derived by selecting residues within 10 Å from Cys89 N.

To check for known hinge binders in the fragment library, all kinase inhibitors found in the ATP binding site were downloaded from the KLIFS database.³² The cores (5 Ångstrom from the hinge NH group) were extracted from known binders from KLIFS and from the compounds in the fragment library. A substructure search was performed to find which of the fragment cores in the library were present in the known hinge binder cores.

Kinase inhibition assay

MR1 (purity 97% by LC/MS) was tested against the target kinases with the Z'-LYTE kinase inhibition assay (Life Technologies). The assay employs a fluorescence-based format and is based on the different sensitivities of phosphorylated and nonphosphorylated peptides to proteolytic cleavage. A suitable peptide substrate is labelled with two fluorophores, forming a FRET pair. After incubating the kinase + peptide + test compound mixture for an hour, a development reaction is carried out. Any peptide that was not phosphorylated by the kinase is cleaved, disrupting the resonance energy transfer between the FRET pair. The reaction progress is quantified based on the ratio of the detected emission at 445 nm (coumarin) and 520 nm (fluorescein), *i.e.* the ratio of cleaved *vs.* intact peptides. A more detailed description of the assay is available on the website of Life Technologies.³³ IC_{50} values were determined



Research Article

RSC Medicinal Chemistry

from 10-pt titration measurements (with duplicate datapoints) using the SelectScreen™ Biochemical Kinase Profiling Service available at Life Technologies.

NMR measurements

The human MELK protein (Q14680) DNA was obtained from OriGene and the kinase-uba domain (1–337) sequence was cloned into a modified pET vector encoding an N-terminal His-tag. The recombinant protein was expressed in *E. coli* BL21 pLysS cells, harvested by centrifugation and purified by immobilized metal affinity chromatography and anion exchange chromatography. ¹H and STD-NMR measurements were performed using a 600 MHz Bruker Avance III spectrometer equipped with a 5 mm cryo-TXI (¹H, ¹³C, ¹⁵N) probe with z-gradient at 298 K. The MELK protein and **MR1** were dissolved in a 10% (v/v) D₂O and H₂O mixture containing 20 mM Tris buffer (pH 8.0), 260 mM NaCl, 1 mM TCEP and 5% glycerol. Spectra were acquired with water suppression using excitation sculpting with the pulsed gradient scheme. For the ¹H and STD measurements, the MELK and **MR1** concentrations were 2.14 and 50 μM, respectively. As a reference, STD experiments were also performed without the target, containing the ligand species alone.

STD-NMR spectra were acquired using a series of 40 equally spaced 50 ms Gaussian-shaped pulses for selective saturation of the protein, with a total saturation time of 2 s and a 50 ms spinlock to suppress protein signals. The frequency of the on-resonance saturation was set at 1.0 ppm and the off-resonance saturation frequency was set at 40.0 ppm. A total of 2000 scans were collected for each pseudo-2D experiment.

Microscale thermophoresis binding assay

Microscale thermophoresis measurements were conducted on a NanoTemper Monolith NT.115 device, using the red fluorescence channel. MELK was stored and applied in a buffer consisting of 20 mM Tris, 260 mM NaCl, 1 mM TCEP and 5% glycerol, at a pH of 8.0. The **MR1** stock was prepared in DMSO and diluted into the protein buffer, with the final DMSO concentrations not exceeding 1%. We have applied the ligand displacement assay principle to detect the thermophoresis of the fluorescently labeled MELK reference ligand Kinase Tracer 236 (ThermoFisher, cat. no. PV5592) at increasing concentrations of **MR1**. A titration curve was acquired with serial 1:1 dilutions starting from 100 μM ligand concentration (with 0.56 μM protein, 50 nM Kinase Tracer 236, 60% LED power, 20% MST power), with 11 datapoints. Each datapoint was acquired in triplicate. The primary result was a relative EC₅₀ value of **MR1** against MELK (21.1 μM), which was converted to a K_i value of 5.47 μM, using the following formula:

$$K_i = \frac{EC_{50}}{\frac{[T]_{50}}{K_d} + \frac{[P]_0}{K_d} + 1} \quad (1)$$

In formula (1), [P]₀ is the protein concentration at 0% binding, [T]₅₀ is the concentration of free tracer at 50% and K_d is the dissociation constant of the tracer and protein, which was evaluated to be 208 nM under the assay conditions in a direct binding measurement.

Conclusions

Here, we report the discovery of a novel hinge binder using a new computational screening protocol based on dynamic undocking. 8-Amino-2*H*-isoquinolin-1-one (**MR1**) was found to inhibit kinases from four major branches of the kinase phylogenetic tree, with low to mid-micromolar potencies (translating to ligand efficiencies between 0.54–0.59). All of these kinases are relevant clinical targets, mostly for oncological indications. In particular, MST3 and MELK have few known potent inhibitors, presenting the opportunity to create novel chemical matter for these kinases with a less congested IP space. In addition to its versatile inhibitory profile, direct binding of **MR1** to MELK was confirmed by STD-NMR, and its binding to the ATP site was validated by an MST-based (microscale thermophoresis) ligand displacement assay developed as part of this work. Together, these findings nominate **MR1** as a viable fragment starting point for a range of type I kinase inhibitor programs.

Abbreviations

AGC	PKA, PKG and PKC containing families
ATP	Adenosine triphosphate
CAMK	Ca ²⁺ /calmodulin-dependent protein kinases
CMGC	CDK, MAPK, GSK3 and CLK containing families
DMSO	Dimethyl sulfoxide
DUck	Dynamic undocking
ERK2	Common alternative name for mitogen-activated protein kinase 1 (MAPK1)
JAK2	Janus kinase 2
LC/MS	Liquid chromatography/mass spectrometry
MD	Molecular dynamics
MELK	Maternal embryonic leucine zipper kinase
MST3	Common alternative name for serine/threonine-protein kinase 24 (STK24)
NMR	Nuclear magnetic resonance
RSK2	Ribosomal protein S6 kinase alpha-3
SMD	Steered molecular dynamics
STD	Saturation transfer difference
STE	Yeast sterile 7-, 11- and 20-homologue kinases
TK	Tyrosine kinases

Conflicts of interest

There are no conflicts to declare.



Acknowledgements

The authors thank Kitti Koprivanacz and Tamás Szimler for their help in protein expression and purification. This study was supported by the MSCA ITN FRAGNET (project 6758993) grant to M. R., A. S., X. B. and G. M. K., by the National Research, Development and Innovation Office of Hungary (OTKA K116904), the MCIU/AEI/FEDER, UE (SAF2015-68749-R and RTI2018-096429-B-I00), the Catalan government (2014 SGR 1189), and the National Research, Development and Innovation Fund of Hungary (HunProEx 2018-1.2.1-NKP-2018-00005).

Notes and references

- 1 D. Fabbro, S. W. Cowan-Jacob and H. Moebitz, *Br. J. Pharmacol.*, 2015, **172**, 2675–2700.
- 2 D. Bajusz, G. G. Ferenczy and G. M. Keserü, *Curr. Top. Med. Chem.*, 2017, **17**, 2235–2259.
- 3 S. Klaeger, S. Heinzlmeir, M. Wilhelm, H. Polzer, B. Vick, P. A. Koenig, M. Reinecke, B. Ruprecht, S. Petzoldt, C. Meng, J. Zecha, K. Reiter, H. Qiao, D. Helm, H. Koch, M. Schoof, G. Canevari, E. Casale, S. Re Depaolini, A. Feuchtinger, Z. Wu, T. Schmidt, L. Rueckert, W. Becker, J. Huenges, A. K. Garz, B. O. Gohlke, D. P. Zolg, G. Kayser, T. Vooder, R. Preissner, H. Hahne, N. Tönisson, K. Kramer, K. Götze, F. Bassermann, J. Schlegl, H. C. Ehrlich, S. Aiche, A. Walch, P. A. Greif, S. Schneider, E. R. Felder, J. Ruland, G. Médard, I. Jeremias, K. Spiekermann and B. Kuster, *Science*, 2017, **358**, eaan4368.
- 4 P. N. Mortenson, V. Berdini and M. O'Reilly, *Methods Enzymol.*, 2014, **548**, 69–92.
- 5 H. Zhao and A. Caflisch, *Bioorg. Med. Chem. Lett.*, 2015, **25**, 2372–2376.
- 6 D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke and H. Jhoti, *Nat. Rev. Drug Discovery*, 2016, **15**, 605–619.
- 7 P. Bamborough, M. J. Brown, J. A. Christopher, C. Chung and G. W. Mellor, *J. Med. Chem.*, 2011, **54**, 5131–5143.
- 8 G. Bollag, P. Hirth, J. Tsai, J. Zhang, P. N. Ibrahim, H. Cho, W. Spevak, C. Zhang, Y. Zhang, G. Habets, E. A. Burton, B. Wong, G. Tsang, B. L. West, B. Powell, R. Shellooe, A. Marimuthu, H. Nguyen, K. Y. J. Zhang, D. R. Artis, J. Schlessinger, F. Su, B. Higgins, R. Iyer, K. Dandrea, A. Koehler, M. Stumm, P. S. Lin, R. J. Lee, J. Grippo, I. Puzanov, K. B. Kim, A. Ribas, G. A. McArthur, J. A. Sosman, P. B. Chapman, K. T. Flaherty, X. Xu, K. L. Nathanson and K. Nolop, *Nature*, 2010, **467**, 596–599.
- 9 P. Kolb, C. B. Kipouros, D. Huang and A. Caflisch, *Proteins: Struct., Funct., Bioinf.*, 2008, **73**, 11–18.
- 10 R. Ulrich, G. Wishart, M. Kiczun, A. Richters, N. Tidtenlusch, D. Rauh, B. Sherborne, P. G. Wyatt and R. Brenk, *Chem. Biol.*, 2013, **8**, 1044–1052.
- 11 S. Ruiz-Carmona, P. Schmidtke, F. J. Luque, L. Baker, N. Matassova, B. Davis, S. Roughley, J. Murray, R. Hubbard and X. Barril, *Nat. Chem.*, 2017, **9**, 201–206.
- 12 E. J. Baxter, L. M. Scott, P. J. Campbell, C. East, N. Fourouclas, S. Swanton, G. S. Vassiliou, A. J. Bench, E. M. Boyd, N. Curtin, M. A. Scott, W. N. Erber, T. Avis, A. Barthorpe, G. Bignell, M. Blow, L. Brackenbury, G. Buck, S. Clegg, J. Clements, J. Cole, H. Davies, S. Edkins, K. Gray, M. Gorton, S. O'Meara, K. Halliday, R. Harrison, W. Haynes, K. Hills, C. Hunter, D. Jones, V. Kosmidou, R. Laman, R. Lugg, A. Parker, J. Perry, R. Petty, A. Small, H. Solomon, P. Stephens, Y. Stephens, C. Stevens, R. Smith, P. Tarpey, C. Tofts, J. Varian, S. West, S. Widaa, S. Bamford, A. Butler, E. Dawson, E. Dicks, K. Edwards, S. Forbes, C. Greenman, J. Hinton, A. Menzies, K. Raine, R. Shepherd, J. Teague, A. Yates, R. Wooster, A. Futreal, M. Stratton and A. R. Green, *Lancet*, 2005, **365**, 1054–1061.
- 13 R. L. Levine, M. Wadleigh, J. Cools, B. L. Ebert, G. Wernig, B. J. P. Huntly, T. J. Boggon, I. Wlodarska, J. J. Clark, S. Moore, J. Adelsperger, S. Koo, J. C. Lee, S. Gabriel, T. Mercher, A. D'Andrea, S. Fröhling, K. Döhner, P. Marynen, P. Vandenberghe, R. A. Mesa, A. Tefferi, J. D. Griffin, M. J. Eck, W. R. Sellers, M. Meyerson, T. R. Golub, S. J. Lee and D. G. Gilliland, *Cancer Cell*, 2005, **7**, 387–397.
- 14 C. Cho, K. Lee, W. Chen and C. Wang, *Oncotarget*, 2016, **7**, 14586–14604.
- 15 K. A. Casalvieri, C. J. Matheson, D. S. Backos and P. Reigan, *Trends Cancer*, 2017, **3**, 302–312.
- 16 R. Ganguly, A. Mohyeldin, J. Thiel, H. I. Kornblum, M. Beullens and I. Nakano, *Clin. Transl. Med.*, 2015, **4**, 11.
- 17 A. I. Ojesina, L. Lichtenstein, S. S. Freeman, C. S. Pedamallu, I. Imaz-Rosshandler, T. J. Pugh, A. D. Cherniack, L. Ambrogio, K. Cibulskis, B. Bertelsen, S. Romero-Cordoba, V. Treviño, K. Vazquez-Santillan, A. S. Guadarrama, A. A. Wright, M. W. Rosenberg, F. Duke, B. Kaplan, R. Wang, E. Nickerson, H. M. Walline, M. S. Lawrence, C. Stewart, S. L. Carter, A. McKenna, I. P. Rodriguez-Sanchez, M. Espinosa-Castilla, K. Woie, L. Bjorge, E. Wik, M. K. Halle, E. A. Hoivik, C. Krakstad, N. B. Gabiño, G. S. Gómez-Macias, L. D. Valdez-Chapa, M. L. Garza-Rodríguez, G. Maytorena, J. Vazquez, C. Rodea, A. Cravioto, M. L. Cortes, H. Greulich, C. P. Crum, D. S. Neuberger, A. Hidalgo-Miranda, C. R. Escareno, L. A. Akslen, T. E. Carey, O. K. Vintermyr, S. B. Gabriel, H. A. Barrera-Saldaña, J. Melendez-Zajgla, G. Getz, H. B. Salvesen and M. Meyerson, *Nature*, 2014, **506**, 371–375.
- 18 L. Fang, W. Lu, H. H. Choi, S.-C. J. Yeung, J.-Y. Tung, C.-D. Hsiao, E. Fuentes-Mattei, D. Menter, C. Chen, L. Wang, J. Wang and M.-H. Lee, *Cancer Cell*, 2015, **28**, 183–197.
- 19 M. Congreve, R. Carr, C. Murray and H. Jhoti, *Drug Discovery Today*, 2003, **8**, 876–877.
- 20 S. Ruiz-Carmona, D. Alvarez-Garcia, N. Foloppe, A. B. Garmendia-Doval, S. Juhos, P. Schmidtke, X. Barril, R. E. Hubbard and S. D. Morley, *PLoS Comput. Biol.*, 2014, **10**, e1003571.
- 21 A. L. Hopkins, G. M. Keserü, P. D. Leeson, D. C. Rees and C. H. Reynolds, *Nat. Rev. Drug Discovery*, 2014, **13**, 105–121.
- 22 G. C. P. van Zundert, B. M. Hudson, S. H. P. de Oliveira, D. A. Keedy, R. Fonseca, A. Heliou, P. Suresh, K. Borrelli, T.



Research Article

RSC Medicinal Chemistry

- Day, J. S. Fraser and H. van den Bedem, *J. Med. Chem.*, 2018, **61**, 11183–11198.
- 23 S. Malhotra and J. Karanicolas, *J. Med. Chem.*, 2017, **60**, 128–145.
- 24 D. Erlanson, *Pract. Fragm. blog*, 2017.
- 25 M. N. Drwal, G. Bret, C. Perez, C. Jacquemard, J. Desaphy and E. Kellenberger, *J. Med. Chem.*, 2018, **61**, 5963–5973.
- 26 L. Xing, J. Klug-Mcleod, B. Rai and E. A. Lunney, *Bioorg. Med. Chem.*, 2015, **23**, 6520–6527.
- 27 S. L. Posy, M. A. Hermsmeier, W. Vaccaro, K. H. Ott, G. Todderud, J. S. Lippy, G. L. Trainor, D. A. Loughney and S. R. Johnson, *J. Med. Chem.*, 2011, **54**, 54–66.
- 28 M. Chartier, T. Chénard, J. Barker and R. Najmanovich, *PeerJ*, 2013, **1**, e126.
- 29 Mol. Oper. Environ. CCG ULC 2016.08, 1010 Sherbooke St. West, Suite #910, Montr. QC, Canada, H3A 2R7, 2018.
- 30 S. Antonysamy, G. Hirst, F. Park, P. Sprengeler, F. Stappenbeck, R. Steensma, M. Wilson and M. Wong, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 279–282.
- 31 Schrödinger Release 2016-4, LigPrep, Schrödinger, LLC, New York, NY, 2016.
- 32 A. J. Kooistra, G. K. Kanev, O. P. J. van Linden, R. Leurs, I. J. P. de Esch and C. de Graaf, *Nucleic Acids Res.*, 2016, **44**, D365–D371.
- 33 Z'-LYTE Kinase Assay Kits. <http://www.lifetechnologies.com/hu/en/home/life-science/drug-discovery/target-and-lead-identification-and-validation/kinasebiology/kinase-activity-assays/z-lyte.html> (Accessed 14 Oct 2019).



