Biophysical and structural approaches to discover small-molecule modulators of challenging targets in cancer

Míriam Martínez Cartró

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

BIOPHYSICAL AND STRUCTURAL APPROACHES TO DISCOVER SMALL-MOLECULE MODULATORS OF CHALLENGING TARGETS IN CANCER

MÍRIAM MARTÍNEZ CARTRÓ

2021

BIOPHYSICAL AND STRUCTURAL APPROACHES TO DISCOVER SMALL-MOLECULE MODULATORS OF CHALLENGING TARGETS IN CANCER

Aquesta tesi ha estat realitzada per Míriam Martínez Cartró sota la direcció del Dr Carles Galdeano Cantador, Professor Lector Serra Hunter i Dr Xavier Barril Alonso, Professor d'Investigació ICREA, tots dos del Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica 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.

Carles Galdeano Cantador   Xavier Barril Alonso   Míriam Martínez Cartró

Director de tesi   Director de tesi   Doctoranda

MÍRIAM MARTÍNEZ CARTRÓ
JULY 2021

ABSTRACT

Advances in structural and molecular biology have increased the knowledge of relevant and *undrugged* targets and have favoured the rational development of novel drugs through structure-based drug design (SBDD). Computational techniques are widely employed in SBDD, resulting as inexpensive, rapid and efficient tools for hit discovery and optimization. Hits from computational approaches have then to be assessed experimentally. Biophysical techniques are ideal candidates for hit validation, since they can directly study compound binding to a particular target. Biophysical techniques can provide deep knowledge of target-compound interactions ranging from binding assessment to binding site determination or information about the atomic structure of the target-compound complex. Overall, the combination of computational and biophysical techniques is a strategy that can enhance our ability to modulate challenging and *undruggable* targets in early-stage drug discovery.

E3 ligases have been described as relevant targets in cancer. Besides, the irruption of the targeted protein degradation technology has situated this target family in the forefront. In the present thesis we have applied a structure-based approach in order to study E3 ligases ligandability. This study has provided valuable information of the binding preferences of the studied proteins, while illustrating the possibility to increase the number of binders of these challenging family. Being FBW7 E3 ligase one of the most mutated proteins in cancer, we have used the previous information to identify and characterize small molecules that bind to this E3 ligase by combining computational and biophysical techniques. These small molecules could be a point of departure to develop drugs able to modulate this E3 ligase.

TET2 is a tumour suppressor that losses its function by mutations or gene repression in different types of cancer, particularly hematologic. Besides, inhibition of TET2 has been described to have a therapeutic interest due to its implication to cancer relapse. Potential TET2 modulators have been developed following a structure-based approach. In the present thesis we have developed and characterized TET2 modulators applying biophysical techniques.

Bromodomains have been recently described for their interest in cancer. Specifically, BRD4 has also been used as a test system for computational techniques due to its ease of production and constant behaviour. Computer-aided drug design faces several challenges, being the prediction of solvation preferences and fragment evolution two of them. On one hand, we have applied computational tools to study the solvation preferences of BRD4 BD1. With that information, we have developed and characterized novel chemical entities. On the other hand, fragments interacting with BRD4 BD1 have been identified by an automated fragment evolution platform developed in our group. In the present thesis, interactions of the evolved fragments with BRD4 BD1 have been characterized. The resulting information has helped to validate the applied computational tools and could be used to develop novel BRD4-based therapies.

TABLE OF CONTENTS

ABBREVIATIONS

α-KG	α-ketoglutarate		CDC20	Cell-division cycle 20
ΔG	Changes in free energy		CETSA	Cellular thermal shift assay
ΔH	Changes in enthalpy		C-JUN	Transcription factor AP-1
ΔS	Changes in entropy		C-MYC	Myc proto-oncogene protein
ΔT	Temperature difference		COP1	Constitutive photo-morphogenesis protein 1
ΔTm	Difference of melting temperature		CPD	Conserved phosphorylation motif
2-HG	L-α-Hydroxyglutaric acid disodium salt		CRBN	Cereblon
5-caC	5-carboxylcytosine		CRLs	Cullin-RING ligases
5-fC	5-formylcytosine		CTLH	C-terminal LisH motif
5-hmC	5-hydroxymethylcytosine		DCAF1	DDB1 and CUL4 associated factor 1
5-mC	5-methylcytosine		DCAF15	DDB1 and CUL4 associated factor 15
ADME	Absorption, distribution, metabolism, excretion		DISC1	Disrupted in schizophrenia 1 protein
ALA, A	Alanine		DMSO	Dimethyl sulfoxide
ASB9	Ankyrin repeat and SOCS box protein 9		DNA	Deoxyribonucleic acid
ASN, N	Asparragine		DOL	Degree of labelling
BET	Bromodomain and extra-terminal domain		dsDNA	Double-stranded deoxyribonucleic acid
BRD	Bromodomain		DSF	Differential scanning fluorimetry
BRD4 BD1	Bromdomain 1 of bromodomain-containing protein 4		DTT	Dithiothreitol
BTB	Bric-a-brac, tramtrack, and broad complex		DUck	Dynamic undocking
BTRC	F-box/WD repeat-containing protein 1A		*E. coli*	*Escherichia coli*
CADD	Computer-aided drug design		EDC	1-ethyl-3-(3-(dimethylamino)propyl) carbodiimidehydrochloride

ESI	Electrospray ionization		LED	Light emitting diode
FBA	F-box associated		LR	Leucine-rich
FBO44	F-box only protein 44		LYTACs	Lysosomal targeting chimeras
FBW7	F-box/WD repeat-containing protein 7		MADTACs	Macroautophagy degradation targeting chimeras
Fe	Iron		MALDI	Matrix-assisted laser desorption/ionization
FP	Fluorescence polarization		MATH	Meprin and TRAF homology
Fw	Forward		MET, M	Methionine
GID4	Glucose-induced degradation protein 4 homolog		MQW	Milli-Q water
GST	Glutathione S-transferase		MST	Microscale thermophoresis
HECT	Homologous to E6AP C-terminus		NaCl	Sodium chloride
HIS	Histidine		NHS	N-hydroxysuccinimide
ILE, I	Isoleucine		NMR	Nuclear magnetic resonance
IMAC	Immobilized metal ion affinity chromatography		NOG	N-oxalylglycine
IPTG	Isopropyl-β-D-thiogalactopyranoside (IPTG)		NOTCH1	Neurogenic locus notch homolog protein 1
IR	Infrared		OD	Optical density
ITC	Isothermal titration calorimetry		PAGE	Polyacrylamide gel electrophoresis
K_a	Equilibrium association constant		PBS	Phosphate-buffered saline
KCTD5	Potassium channel tetramerization domain containing 5		PCR	Polymerase chain reaction
K_d	Equilibrium dissociation constant		PDB	Protein data bank
KEAP1	Kelch-like ECH-associated protein 1		PEG	Polyethylene glycol
KLHDC2	Kelch domain-containing protein 2		PHD	Plant homeodomain
k_{off}	Dissociation rate constant		PHE, F	Phenylalanine
k_{on}	Association rate constant		PPI	Protein-protein interactions
LB	Luria broth		PRKN	Parkin

X

PROTAC	Proteolysis targeting chimera		SPRY	Protein sprouty homolog
R	Gas constant		SPSB1	SPRY domain-containing SOCS box protein 1
RBCC	N-terminal RING finger/B-box/coiled coil		S_T	Soret coefiicient
RBR	Ring-between-ring		SUMO	Small ubiquitin-like modifier
RING	Really interesting new gene		T	Temperature
Rmax	Maximum response (in SPR)		TET2	Ten-eleven translocation-2
RNF4	RING finger protein 4		TEV	Tobacco etch virus
RNF43	Ring finger protein 43		Tm	Melting temperature
ROS	Rosetta		TRIM21	Tripartite motif-containing protein 21
Rv	Reverse		TRIM28	Transcription intermediary factor 1-beta
SAR	Structure-activity relationships		Tris	Tris(hydroxymethyl)aminomethane
SCF	S-phase kinase-associated protein 1 (Skp1)–Cul1–F-box protein		TRP, W	Tryptophan
SH2	Src Homology 2		TYR, Y	Tyrosine
SKP1	S-phase kinase-associated protein 1		ULP1	Ubiquitin-like-specific protease 1
SKP2	S-phase kinase-associated protein 2		UPS	Ubiquitin proteasome system
SMD	Steered molecular dynamics		VS	Virtual screening
SOCS2	Suppressor of cytokine signalling 2		WDR	Tryptophan-aspartic acid (WD) repeat
SPOP	Speckle-type POZ protein		WT	Wild type
SPR	Surface plasmon resonance			

LIST OF FIGURES

CHAPTER 5: RESULTS – Pharmacological modulation of TET2 epigenetic enzyme

CHAPTER 6: RESULTS – Development of new chemical entities that bind to the bromodomain BRD4 BD1

LIST OF BOXES

LIST OF TABLES

CHAPTER 5: RESULTS – Pharmacological modulation of TET2 epigenetic enzyme

CHAPTER 6: RESULTS – Development of new chemical entities that bind to the bromodomain BRD4 BD1

CHAPTER 1: INTRODUCTION

INTRODUCTION

1.1. The drug discovery process

The recent worldwide emergency caused by coronavirus disease has emphasised the importance to tackle unmet medical needs[1]. Drug discovery is a multifaceted process, which involves the identification of a drug chemical therapeutically useful in the treatment and management of a disease condition[2]. The development of a completely novel drug is an extremely arduous, long, expensive and risky process. Before launching the drug to the market, different stages need to be fulfilled in order to develop a safe and effective drug with all regulatory requirements approved (Figure 1.1). In general, drug discovery process has been described to last around 15 years costing more than a billion dollars[2,3].

THE DRUG DISCOVERY PROCESS

RESEARCH & DEVELOPMENT	PRECLINICAL STUDIES	CLINICAL TRIALS	REVIEW, APPROVAL & PHASE IV
3-6 YEARS	1-2 YEARS	4-7 YEARS	1 - ∞ YEARS
- Unmet medical need identification - Hit identification - Hit to lead - Lead optimization	- Safety studies *in vitro* and *in vivo* - Efficacy studies *in vitro* and *in vivo*	- Phase I trials (safety) - Phase II trials (efficacy) - Phase III trials (safety and efficacy in larger population)	- Evidence from trials - Approval and manufacture - Phase IV (Post-release monitoring of side effects)

Figure 1.1. Scheme of the drug discovery process. Different stages and the corresponding times to undertake the development of a novel drug are shown.

The initial step in the drug discovery process is the research and development, also referred as early drug discovery. Here, a hypothesis that the modulation of a target or pathway would result in a therapeutic effect has to be generated. This step usually involves the participation of academia. In the case of target-based drug discovery, this phase involves the identification and validation of a target, which has to be related to a particular disease while triggering a phenotypic response[4]. Chemical entities are subsequently screened and if binding or activity is presented

they are considered hits. These last need to be assessed by different experiments and assays to become validated hits. Chemical series are then prioritized for hit-to-lead follow up. This selection not only involves hit potency, but also binding mechanism, chemical tractability, pharmacokinetic properties or patentability. A multidisciplinary team modifies chemical features of the selected hits to develop a structure-activity-relationship in order to optimize the hit[5].

Leads undergo preclinical phase, in which their efficacy and safety is assessed by *in vitro* and *in vivo* studies. Besides, pharmacokinetics are also exhaustively evaluated at this stage. Pharmacokinetics involve the study of distribution of the drug over the body, the so-called ADME (absorption, distribution, metabolism, excretion). The range of potentially safe and tolerable doses of leads must also be assessed to pass the preclinical phase. In fact, at least two non-human species are usually used to provide information about pharmacokinetics and toxicity of leads[6].

Compounds that passed the preclinical phase enter to the clinical trials. This stage is comprised of four different phases (I-IV). Phase I studies are generally performed in healthy volunteers, or sometimes in patients who failed to improve on existing therapies. The objectives of these studies are to determine tolerability and characterize pharmacokinetics and a preliminary assessment of drug activity[7]. Phase II studies are conducted on larger groups of patients (few hundreds) to evaluate the efficacy of the drug and to have additional safety data. Phase III clinical trials consist on studying the safety and efficacy for a larger number of patients than in phase II and for a longer period. These allow to detect long-term or uncommon side effects[2].

Once passed clinical trials phase I-III, the drug has to be submitted to regulatory agencies approval to be launched to the market. Phase IV of clinical trials is conducted after the drug is approved by the regulatory agency. This experiments are performed in a much larger number of patients for longer periods of time (unless the drug is retired from market, phase IV persists). Therefore, rare or long-term side-effects can be assessed in phase IV and the balance between benefit and harm becomes clearer[7].

1.2. The (un)druggable proteome and unmet clinical needs

The effect of a drug can interfere with any of the four types of macromolecules in human body: proteins, polysaccharides, lipids and nucleic acids[8]. Nevertheless, toxicity and low specificity are more related to the latter three types. This fact explains why almost all approved drugs on the market are directed against protein targets. Remarkably, drugged proteins need to be suitable for drug interactions, while appropriate drug targets. Druggability term appeared 15 years ago. Nowadays is defined as the ability of a protein to bind a small molecule or antibody with required affinity, adequate chemical properties, and at the same time be a potential drug target (for example being linked to a disease)[9]. This leads to the term "druggable proteome", which can be defined as the fraction of proteins that accomplish all the aforementioned characteristics. Mainly, the druggable proteome is constituted by enzymes, ion channels, receptors and transporters, being their function classification illustrated in Figure 1.2 A.

It has been reported that 10% of proteins have a described small-molecule binder. Nevertheless, some of these proteins have been described as druggable and other as ligandable[10]. Ligandability is a pre-requisite for druggability, since it just implies binding of a small molecule to the target, without the need of accomplishing the pharmacodynamic and pharmacokinetic mechanisms[11]. Considering that, 3% of proteins are linked to at least one approved drug, while 7% are known to bind to small molecules with high potency lacking the druggability requirements. Proteins with confirmed bioactivity and clinical phenotype but not small-molecules developed represent the 55% of the proteome. Finally, the 35% of the remaining proteins constitute to the so-called dark proteome (Figure 1.2 B).

Cancer is nowadays an unmet medical need that causes millions of deaths every year, entailing a need to develop novel treatments to combat it. Modulating disease-associated proteins with no described small-molecule binders could be crucial for that. Over the last 40 years, improvements in prevention, detection and treatment have revolutionised cancer medicine supposing an increase of the survival. Nevertheless, progress has not advanced equally for all forms of the disease. Lung, pancreatic, oesophageal cancers and brain tumours share

poor five-year survival with limited improvement in the past decade. Besides, cancer continues being the most common cause of death in children and the main cause of death by disease among teenagers and young adults. An additional unmet medical need within the field of oncology is the treatment of rare cancers, being 5% of all cancers diagnoses. The expansion of the druggable proteome for the development of novel treatments would imply a huge benefit for this disease, especially in the aforementioned forms of cancer[12].
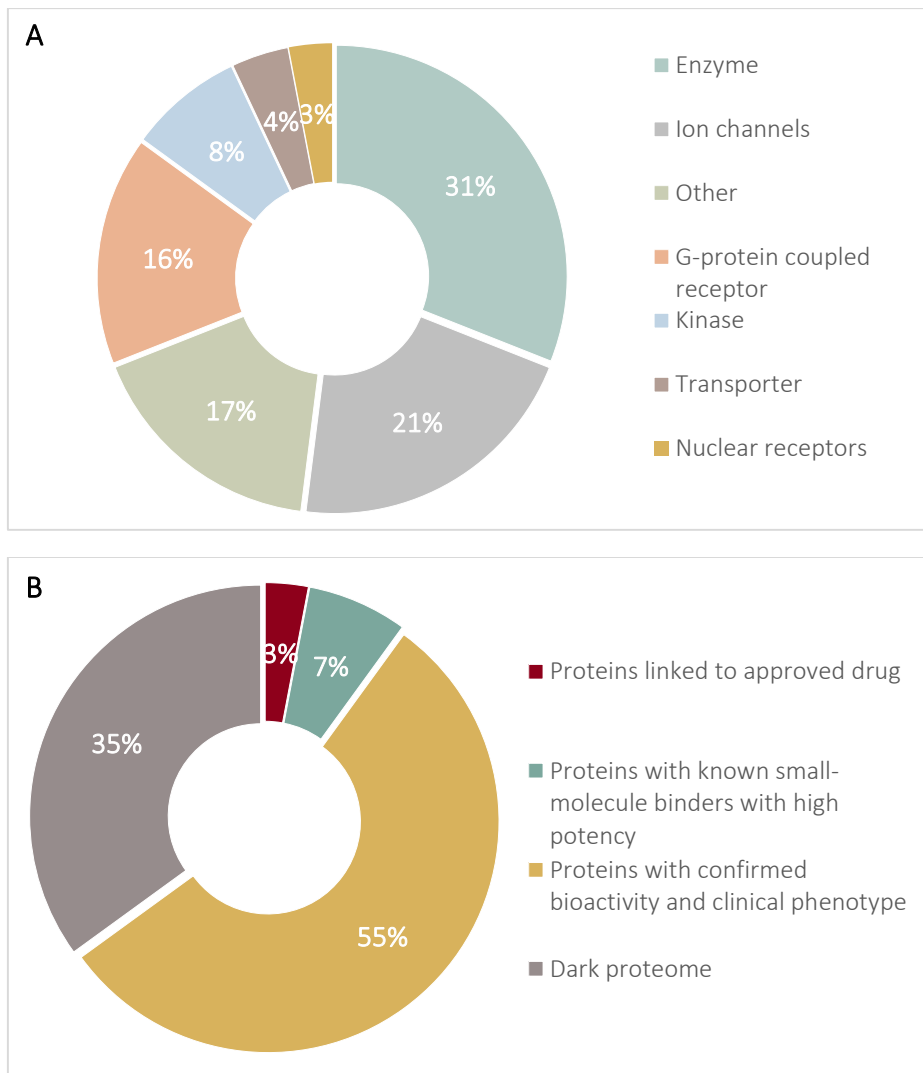
Figure 1.2. The (un)druggable proteome. A. Major protein families with a described drug. Data obtained from [10]. **B.** Percentages of the whole proteome regarding described knowledge. Data obtained from from [10].

1.3. Exploiting novel mechanisms of action

Drugs can be mainly classified in two categories: small molecules and biologics. On one hand, biologics are medicines derived from living cells or through biological processes. In fact, they are complex molecules consisting of proteins, carbohydrates, nucleic acids, cells or tissues for transplantation. Examples of biologics would be vaccines, monoclonal antibodies, recombinant proteins, gene therapies, blood products, among others. On the other hand, small molecules are made by chemical synthesis. In contrast to biologics, they possess a relative simple structure, with molecular weights lower than 1 kDa. Besides, they do not trigger the typical immune response produced by biologics[13]. On 2020, FDA approved 57 new drugs: 40 small molecules and 17 biologics[14]. Regarding the focus on small molecules of the present thesis, in this section we will center on the novel mechanisms of action of small-molecule drugs.

Traditionally, small-molecule drugs modulated protein activity by acting at the orthosteric site, the pocket were endogenous ligands and substrates bind. Therefore, these drugs directly competed or enhanced endogenous substrates[15]. The concept of allosterism was originally introduced by Monod et al. 1963[16]. In contrast to orthosteric drugs, allosteric drugs interact with other regions of the protein modulating its activity indirectly. Consequently, these drugs are resistant to orthosteric competition and affect protein activity by changing the conformation at the allosteric binding site[17]. The advantages of allosteric drugs include more specificity. This can achieved because differently to allosteric sites, orthosteric sites tend to be highly conserved among different family members, making challenging the development of selective orthosteric drugs. Besides, less toxicity is also related to allosterism due to the indirect mechanism they perform, modulating protein activity rather than blocking it completely. Allosteric drugs are also an alternative when disease-related proteins are resistant to orthosteric activation or inhibition[18]. Finally, in combination with orthosteric ligands they have been described to be able to overcome drug-resistant mutations[15].

A relevant type of novel small-molecule drugs are interfacial modulators, which encompass all small molecules that mediate binding between two or more macromolecules as the multimeric complex. Therefore, this

supposes a powerful mechanism to modulate cellular events for drug discovery purposes. Examples of interfacial modulators include small molecules affecting protein-nucleic acid interfaces[19], as well as those that condition protein interactions. The interactome refers to the network of interactions between proteins. It has been described to be essential in both, biological systems and disease development. Studies of protein-protein interactions (PPI) have led to the realization that rather to be equally distributed across the protein surfaces, PPI are focused in certain residues or regions, largely responsible for driving binding[20]. Targeting PPI has been considered challenging for two main reasons. One corresponds to the huge surface area that needs to be targeted, unless an interaction hotspot is identified. The second is that PPI are usually flat, lacking of cavities to fit a small molecule. Small molecules can affect PPI through orthosteric or allosteric modulation[21]. On one hand, PPI inhibition can involve direct competition against the interacting partners or rather be produced by an alternative mechanism. The first requires the binding of the small molecule to the interacting areas of the binding protein, directly inhibiting the complex formation. The second implies binding of the small molecule in cavities different from the macromolecular interface. This interaction induces changes in either the static conformation or the dynamic properties of the target protein(s) and hinders the macromolecular interaction in an allosteric manner[21]. Initially, PPI inhibitors, especially the orthosteric, tended to be larger and more hydrophobic than typical orally available drugs. However, recently PPI inhibitors have reached these "drug-like" values, especially in the cases where PPI contain small, high-affinity interface including a hot segment that can recapitulate the binding of the partner protein[22]. On the other hand, small molecules enhancing PPI have also been developed, known as molecular glues. Originally, molecular glues were related to natural products and there was a belief that they could only be designed from these compounds. However, this was not the case and several chemical entities have been developed and have acted as molecular glues. This mechanism of action is a non-covalent extension of the small modifications already encompassed in the cell that mediate protein-protein associations, such as post-translational modifications. Recently, molecular glues have gained popularity for their relevance in the target protein degradation field[23]. For instance, molecular glues

provide novel interactions between a substrate receptor and an E3 ubiquitin ligase, which hijacks the ubiquitin proteasome system, leading the degradation of the substrate[24].

BOX 1.1. Targeted protein degradation through PROTACs
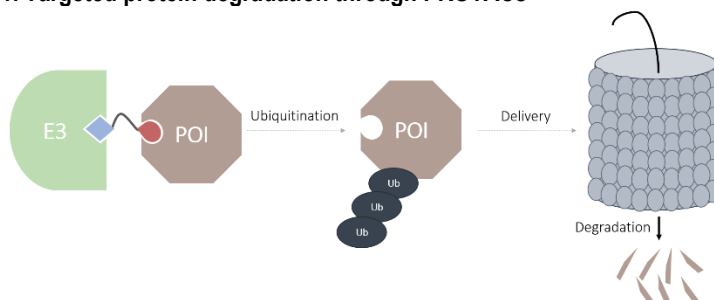
PROTACs are bifunctional small molecules that hijack E3 ligases to cause ubiquitination and degradation of a particular protein of interest (POI). The main characteristics of PROTAC technology are:

- **They work through a validated sub-stoichiometrically catalytic mechanism** based on the dual interaction, allowing lower drug doses. One warhead of the PROTAC molecule binds to the POI, while the other interacts with a specific E3 ligase. After dissociation, PROTAC molecule can continue mediating ubiquitination and subsequent degradation of an additional POI.
- Rather than mainly depending on the affinity of the wareheads, **PROTACs activity strongly depends on cooperativity**. Cooperativity is the ability to form the ternary complex: E3 ligase – PROTAC – POI. This phenomenon allows weak binders to be potential wareheads for successfully mediating degradation, since the resulting ternary complex may be highly compatible and even selective. Consequently, this technology could permit the expansion of the druggable proteome.
- **PROTACs have been described to overcome resistance from small-molecule inhibitors**, since they mediate degradation through a completely different mechanism, acting as a chemical knockdown.
- **Resistance can also be developed in PROTAC molecules** affecting the involved proteins of the ubiquitin proteasome system.

Despite having around 600 E3 ligases in the human body, few of them have described ligands, which are mainly employed for PROTACs construction. Expanding the toolbox of E3 ligases would help to disentangle the remaining unknowns of PROTAC technology. Besides, it would allow the overcoming of PROTAC resistance, and also a more selective targeted protein degradation (e.g., tissue or organ specific)[25,26].

Targeted protein degradation is a novel field in drug discovery that has boosted in recent years. Several types of small molecules involved in the targeted protein degradation have been described. These drugs can be

differentiated regarding the hijacked mechanism of degradation. On one hand, protein degradation can be mediated through the ubiquitin proteasome system[27]. This is the case of PROteolysis TArgeting Chimera (PROTAC) molecules, which are bifunctional small molecules that hijack E3 ligases to cause the ubiquitination and degradation of a particular protein of interest. Unlike traditional occupancy-driven inhibition carried out by typical drugs, PROTACs behave catalytically, promoting target degradation at lower drug concentrations[28]. Indeed, several PROTAC molecules have recently achieved clinical trials, proving to be a successful drug discovery strategy[25]. Further detailed information about PROTACs can be found in Box 1.1. The aforementioned molecular glues act similarly to PROTACs. These drugs mediate the interaction of an E3 ligase and the protein of interest. While PROTACs are able of binding to the E3 ligase and the degraded protein separately, molecular glues are specific binders of the complex. On the other hand, protein degradation can be mediated through lysosome employing LYsosomal TArgeting Chimeras (LYTACs) or MAcroautophagy Degradation TArgeting Chimeras (MADTACs). The first are small molecules that mediate protein degradation of extracellular proteins hijacking the lysosomal degradation pathway. The latter are drugs that mediate protein degradation through macroautophagy, a degradation pathway that promotes degradation of cytosolic substrates through lysosomes. A part from proteins, this substrates can be larger objects such as dysfunctional organelles and intracellular pathogens. Overall, targeted protein degradation field has expanded the possibility of how protein function can be modulated by small molecules[27].

Another significant mechanism of action that can be driven by small-molecule drugs is the one encompassed by pharmacological chaperones. Most disease mutations destabilize the affected protein without affecting the binding site and the body's quality control system eliminates them. Pharmacological chaperones are small molecules designed to selectively interact with the incorrectly folded target protein, while stabilizing it and preventing their degradation, resulting in an enhancement of its activity[29]. Pharmacological chaperones have entered in clinical practice for some rare diseases[30]. Nonetheless, they have some limitations as being mutation specific or the cases where

large deletions are introduced[29]. Moreover, most pharmacological chaperones are described inhibitors of their targets. Therefore, they can rescue the protein for degradation but continue inhibiting its activity[29]. Allosteric ligands stabilizing the protein while not affecting the binding site would be ideal candidates for pharmacological chaperones[30].

Imposing our will with small molecules to orchestrate protein proximity by remote control is a powerful capability. The interaction of a drug to a protein can affect its interactomes, modifying its behavior. Novel mechanisms of action can enlarge the opportunities of targeting disease-associated proteins for a therapeutic benefit[23].

1.4. Drug discovery strategies: Target-based drug discovery versus phenotypic screening

Currently drug discovery strategies can be based on molecular or empirical approaches. The latter is referred as phenotypic, since it relies on a physiological response provided by a chemical entity, which specific mechanism of action could be disentangled a posteriori. In contrast, molecular approaches, also referred as target-based, are hypothesis-driven and are focused on a specific molecular target that has an important role in a disease[31]. The typical pathways in drug discovery for both approaches are illustrated in Figure 1.3.
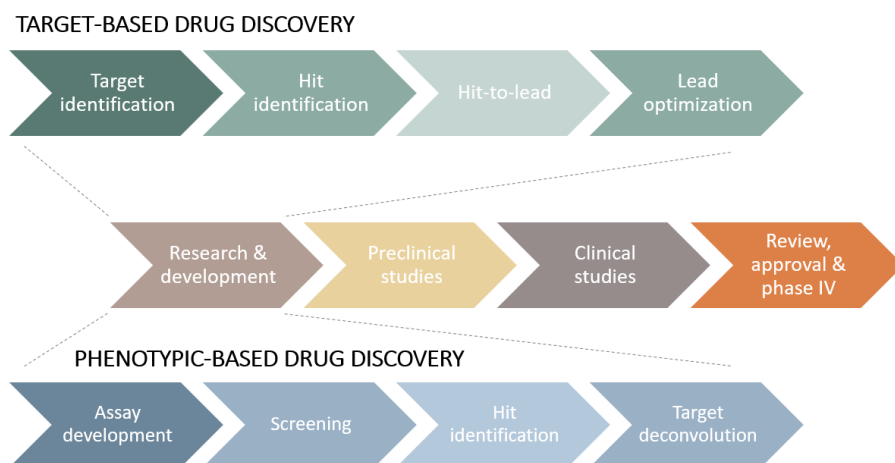
Figure 1.3. Target-based versus phenotypic-based drug discovery. Scheme of the drug discovery process including the typical steps that would follow the target-based approach and the phenotype-based approach.

Target-based strategies are considered to be more robust and cost-effective than phenotype-based. Besides, the first approach has also been considered to reduce the median time for drug discovery and development. Challenges are faced in target-based drug discovery in target validation step, which needs to be accurately developed[32]. Indeed, this is one of the reasons why hits found through this strategy do not always translate to a complex clinical phenotype[32]. In contrast, phenotype-based assays use intact biological system that closely mimic clinical responses[33] and can unearth new biology and pathways. Besides, selected hits usually have more drug-like properties, including cellular permeability and hydrophilicity[34].Phenotype-base strategy also faces fundamental challenges, being hit selection one of them. Initial phenotype assays typically contain thousands of hit compounds, being a considerable amount of them unselective or toxic, substances with unwanted mechanisms of action or false positives[35]. Furthermore, difficulties arise in defining mechanisms of action through target deconvolution. This is an essential step in the whole drug discovery pipeline, being often a time- and resource- intensive process[33].

Edner et al. performed an analysis comparing the strategy of drug discovery employed for the FDA approved first-in-class drugs from 1999 to 2013 (Figure 1.4 A). First-in-class drugs were the ones modulating an — until then — unprecedented target or biological pathway. This study showed that the majority of drugs were discovered by a target-based approach, considering both small molecules and biologics. In fact, they presented an alternative categorization of drug discovery approaches, which a part from the aforementioned included the chemocentric approach. This strategy is based on the use of compounds with known pharmacology as starting points. They proposed to include it in the system-based approach, together with the phenotypic drug discovery (Figure 1.4 B).

Despite the classical comparative between target-based and phenotype-based drug discovery, the choice should not be between them. They are both complementary approaches. The actual goal is to use an appropriate combination of both strategies to enable promising drug discovery ideas to successfully move forward[31,35].
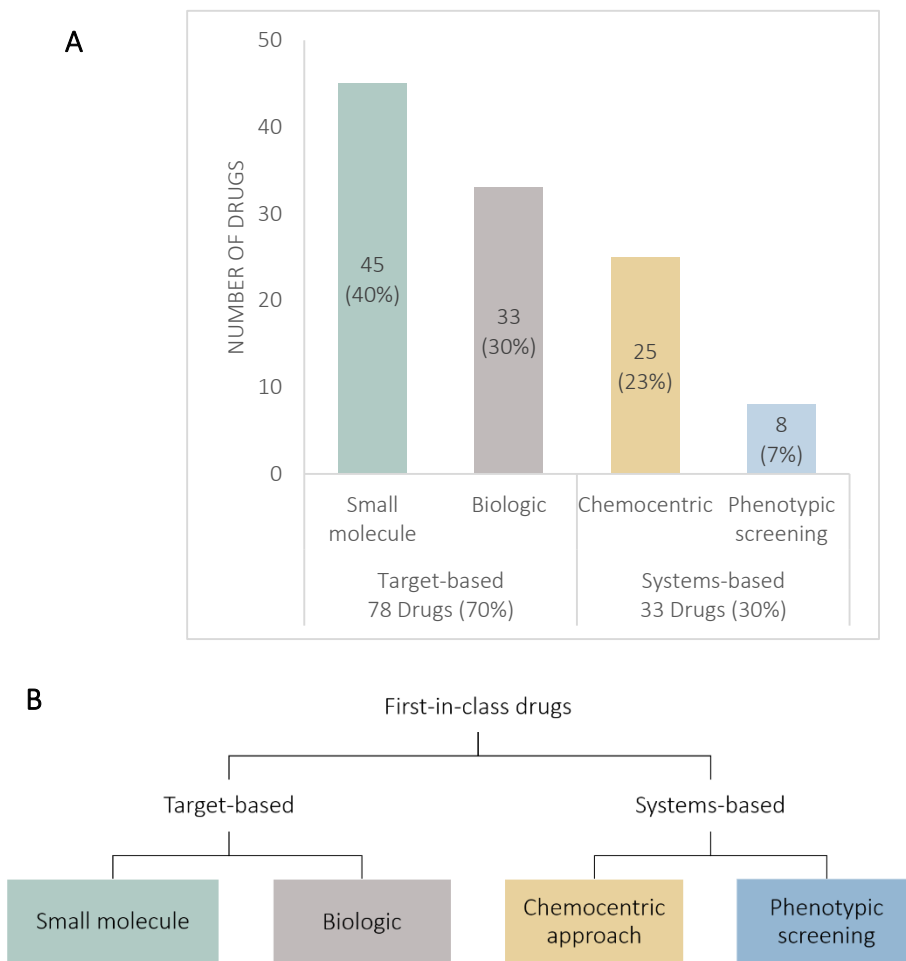
Figure 1.4. First-in-class drugs. A. Number of first-in-class drugs approved by the FDA from 1999 to 2013 according to their discovery strategy. **B.** Types of drug discovery strategies according to [35]. Both figures are modified from [35].

1.5. Target-based small-molecule hit identification strategies

1.5.1. Chemical probes for target validation

A successfully early drug discovery program strongly depends on target identification and validation. In order to interrogate the function of a protein, several mechanisms have been employed, being the use of small-molecule chemical probes one of them. A useful chemical probe interacts with the target of interest with high specificity, providing a better understanding of the target function in a biological context[36]. The development of new chemical probes is described to be particularly

challenging, since the exploration a new mechanism of action for target validation implies a lack in the previous knowledge[26].

From a technical point of view, the discovery of chemical probes does not differ much from hit identification in early stage drug discovery. The main differences between small molecule drugs and chemical probes include the need of a defined mechanism and selectivity of the latter. In contrast, chemical probes are not necessarily required to have drug-like properties, such as bioavailability[37]. Indeed, in order to avoid bad quality chemical probes, several criteria have been proposed, being potency, selectivity, demonstrated target engagement and availability of a negative control the most important ones[26].

1.5.2. Computational tools for hit identification

Computer-aided drug design (CADD) is a target-based methodology that encompass all the different range of computational techniques that can be involved in drug discovery[38]. CADD techniques have resulted to speed up and reduce the cost of the drug discovery process in a number of ways[39]. Indeed, they have been involved in three major purposes: (i) filter large compound libraries into smaller sets of potential active compounds that can be tested experimentally; (ii) guide hit and lead optimization, whether with the aim of increasing its affinity or optimizing drug-like properties for ADME; (iii) design novel compounds, either by "growing" starting molecules or by piecing together fragments into novel chemotypes. CADD is typically classified in two categories: structure-based and ligand-based. The first, relies on the knowledge of the target protein structure[38]. Actually, advances in structural and molecular biology have increased the knowledge of relevant and undrugged targets and have favored the rational development of novel drugs through structure-based CADD techniques[40]. In contrast, ligand-based CADD exploits the knowledge of known active and inactive molecules through chemical similarity searches or construction of predictive, quantitative structure-activity relationship[38]. Figure 1.5 illustrates the position of CADD in the drug discovery pipeline. Due to the scope of the present thesis, the computational methods detailed a posteriori will be based on structure-based drug design for hit identification, unless mentioned otherwise.
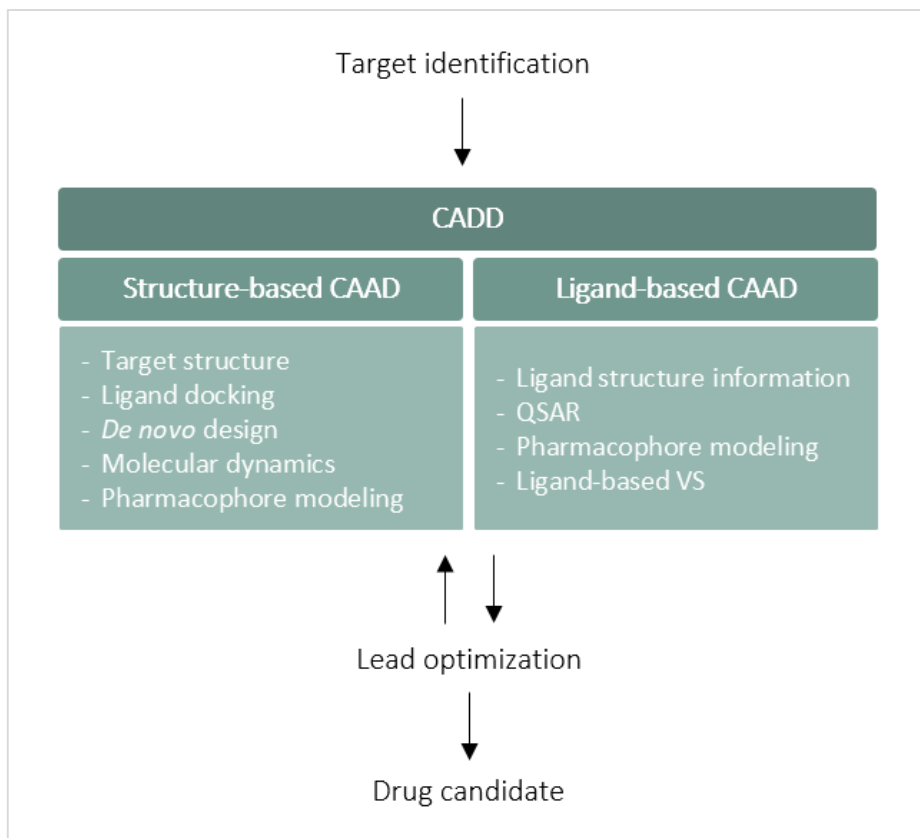
Figure 1.5. CADD methodologies applied in the early stage drug discovery procedure. Abbreviations: CADD, computer-aided drug design; QSAR, quantitative structure-activity relationship; VS, virtual screening. Modified from[38].

1.5.2.1. Tools for druggability assessment and binding site prediction

Plenty of computational methods have been developed to automatically detect protein binding sites and subsequently assess their druggability. These predictions appear to be really valuable when no ligands are described for a particular target or to discovery novel allosteric pockets. Besides, druggability predictions allow the avoidance of proteins not suitable for drug interactions[41]. Binding site prediction methods were classified by Carl et al. as template-based methods (also referred as genomic-based methods), geometry-based methods, energy-based methods and others. Template-based methods are algorithms guided by the already known protein information. They assume that the predicted binding sites of the protein of reference will correlate with the ones of its structural homologues[42]. Some examples of this methods include:

FINDSITE[43], IntFOLD[44], and 3DLigandSite[45]. Geometry-based methods characterize protein surface using different parameters such as Van der Waals radii in order to locate pockets or clefts. Besides, they can also define the pocket by calculation of molecular distance, solvent accessible surface area, and cavity volume. The advantage of geometric measurements is that they do not require previous knowledge of the protein[42]. Examples of these technologies are fpocket[46], SiteMap[47] and Depth[47]. Energy-based methods calculate hydrogen bonding or π-stacking to locate the most druggable regions over the protein. This is achieved by using probe molecules or chemical moieties[42]. Some of their examples include AutoSite[48], PocketFinder[49] and FTSite[49]. Other methods include machine-learning methods, which have been the primary focus of recent developments[42].

The aforementioned methods for rational design of novel cavities can undergo different challenges, as the not being able to detect conformational changes[50]. Methods based on co-solvent molecular dynamics simulation can help to overcome this fact. This methods involve molecular dynamic simulations of a protein in the presence of explicit water molecules mixed with co-solvent molecules to perform hotspot detection, binding site identification, and binding energy estimation[51]. Indeed, they rely of the demonstrated the ability of proteins to unspecifically bind to small organic co-solvents in regions over the surface, which correlate well with binding sites and important interactions for the drug binding affinity[52]. Examples of co-solvent molecular dynamics involve MDmix[53], SILCS[50] and EXPLORER[51].

1.5.2.2. Virtual screening methodologies

Popular virtual screening (VS) techniques were originated in the 1980s and the first related publication appeared in 1997 (Horvath, 1997). In recent times, VS is considered an excellent alternative to high throughput screening, especially for its cost-effectiveness. In fact, the application of this methodology has already contributed to the development compounds that can be now found in the market, such as captopril (antihypertensive), tirofiban (fibrinogen antagonist), several antivirals (saquanivir, ritonavir, indinavir, etc.), among others[54].

In short, VS refers to the use of computational tools to select compounds with potential to bind a specific target, that will subsequently be screened in, for example, biochemical assays. Indeed, the term VS usually is used to refer to docking of large compound databases[39]. Docking is a technique able to predict the conformation and binding of small molecules to the protein of interest. It can be applied to a virtual compound library and it predicts binding energies to rank the resulting ligands according different scoring functions[40]. To achieve that, docking programs use two different procedures: the search algorithms and the scoring functions. The firsts generate the possible conformation of each ligands, while the latter permit the ranking of the different molecules. Furthermore, since it is a structure-based technique it requires the availability of 3D structure of the protein on interest[54]. If no structure is available, homology models can be developed using structures of proteins with a similar sequence[55]. A virtual database of small molecules is also necessary for docking-based virtual screening. Several compound databases have been published[54] and, with the aim of expanding the chemical space, ultra-large libraries have recently been employed[56].

Consensus docking has been applied in some VS studies, which consist in the combination of two different docking programs to improve the reliability of the predictions. With the same purpose, machine learning-based algorithms have also been developed[54]. Furthermore, covalent docking has been applied in order to assess additional mechanisms of action. Most of this last methods focus on modelling the conformation of covalent inhibitors in the bound covalent complex[57]. Finally, additional methods completely orthogonal to docking have flourished to increase VS success rate. This is the case of Dynamic Undocking (DUck), a particular implementation of steered molecular dynamics that measures the work (W_{QB}) needed to break the main hydrogen bond between a small molecule and a protein. This technique has been proved to be useful to detect true ligands and increase the success of VS campaigns[58].

Structure-based VS presents several disadvantages, such as the challenge to simulate protein flexibility, which is translated in not accurate prediction of binding mode determination. This often leads to false positives and false negatives. Nevertheless, VS allows the preselection of a smaller subset of compounds to test experimentally

with potential to bind in the predicted site. This supposes a decrease in the time and cost of hit identification[54].

1.5.3. Target-based screening of chemical libraries

1.5.3.1. High-throughput screening

As an alternative of applying computational tools for initial hit selection, libraries of compounds can be directly tested upon targets of interest. These libraries can be comprised of different types of molecules[4]. High-throughput screening using drug-like molecules (accomplishing Lipinski rule of 5) is one of the most widely used approaches for drug discovery in pharmaceutical research[59]. Large amounts of compounds (can be millions) are tested in these types of assays, requiring the use of complex laboratory automation. Initial hits from these screenings have affinities in the low µM range. However, the resulting small molecules usually do not perform high quality interactions and further information needs to be provided to improve their ligand efficiency[60]. As an advantage, no previous knowledge of the chemotype likely to interact with the protein needs to be assumed[4].

1.5.3.2. Fragment-based drug discovery

Fragment-based drug discovery is a different strategy widely used in recent years. It is based on the screening of smaller libraries of smaller compounds. Indeed, fragments usually follow the rule of three, what supposes having a molecular weight below 300 Da, fewer than three hydrogen-bond donors and acceptors, fewer than three rotatable bonds and a cLogP of three or below. Fragments tend to be less complex than drug-like molecules, while providing a higher coverage of the chemical space. Subsequently, fragments are able to interact with a greater number of sites and proteins, which is especially useful for challenging targets. In counterpart, they are prone to be unspecific. Despite fragment hits are weak binders, they need to make high-quality interactions with the target to bind with sufficient affinity detection. As a consequence, ligand efficiency of fragments is usually higher than drug-like initial hits[60]. The main challenge of fragment-based drug discovery is fragment optimization, which requires structural information. For that purpose, three different strategies are used:

fragment growing (adding atoms to a fragment), fragment merging (combining binding features of fragments and other hits) and fragment linking (linking two fragments that bind to distinct sites)[61,62].

1.5.3.3 Screening of natural products

Natural products have long tradition of being valuable starting points for drug discovery. This interest is related to their structural diversity and various bioactivities. However, technical barriers found in their screening, isolation, characterization and optimization contributed to reduce the use of natural products in drug discovery programs. In recent years, several technological and scientific developments have overcome these challenges reviving the interest in them. This has allowed the recent enrichment of screening libraries based on natural products[63,64].

1.6. Biophysical techniques for hit characterization

Biophysical techniques are ideal candidates for hit validation and characterization, since they can directly study compound binding to a particular target. They can be applied for the screening of chemical libraries or in combination with computational tools, using a smaller number of compounds preselected as potential binders. A large number of biophysical techniques can be employed, which not only differ in the physical principle that underlies the detection of binding, but also in the throughput, information content, sample requirements, sensitivity and robustness of the resulting data. Biophysical techniques can provide deep knowledge of target-compound interactions ranging from binding assessment to binding site determination or information about the atomic structure of the target-compound complex[65]. A summary of some of the most used biophysical techniques and its applicability is shown in Table 1.1. The biophysical techniques employed in the present thesis are further detailed a posteriori.

Table 1.1. Summary of the main biophysical techniques employed in drug discovery

Technique or method	Strengths	Limitations	Affinity range	Throughput
X-ray crystallography				
Protein-compound co-crystallization or apo protein crystal soaking	- Identification of binding site and binding mode - Direct visualization at atomic resolution of target-ligand interactions to optimize compound structure	- Need for diffraction quality crystals suitable for ligand soaking or co-crystallization - Often requires access to large research of infrastructures (synchrotrons) - No quantitative affinity information	No lower limit to 1M	Medium
Ligand-observed NMR				
Shift change in magnetic state of ligand due to binding	- Confirmation of ligand binding to unlabeled proteins of any size - Integrity of ligand and protein in every experiment	- Considerably large amounts of protein required (typically tens of milligrams for screening) - Limited derived structural information	100 nM to 10 mM	Medium
Protein-observed NMR				
Protein NMR peak shift induced by binding	- Monitor protein integrity upon ligand binding - Titration can reliably determine K_d - Binding epitope resolved from pattern of perturbations - Can determine structure if NMR spectrum is assigned	- A large amount of isotopically labelled protein is required - Screening of large libraries is only possible in multiplexing mode - Only suitable for small proteins (MW < 40 kDa)	100 nM to 1 mM	Medium
SPR				
Refractive index change due to ligand binding to immobilized target on sensor	- Allows obtaining K_d and other kinetic data - High sensitivity allows analysis of fragments - Use of low amounts of protein	- Requires immobilization of target with high stability over time - Signals affected by solvent effects	1nM to 500 µM	Medium

Table 1.1 (cont.). Summary of the main biophysical techniques employed in drug discovery

Technique or method	Strengths	Limitations	Affinity range	Throughput
DSF				
Monitoring of protein thermal unfolding with a fluorescent reporter ligand. Thermal stability is increased upon ligand binding.	- Fast and robust assay development - Functional knowledge of target not necessary - Cheap materials - Equipment widely available	- Requires a fluorescent dye - Artefacts occur owing to fluorescence quenching or aggregation - Not suitable for disordered or hydrophobic proteins	1 nM to 100 µM	High
ITC				
Heat of the system changes upon binding event	- Direct determination of thermodynamic parameters for a binary system	- Only useful for binding events with enthalpic component - Very high protein consumption - Requires high solubility of titrated component	1 nM to 100 µM	Low
DSC				
Amount of heat required to increase temperature of sample changes upon binding	- Determination of the effect of a ligand on the thermal stability of a protein - Can be used to study ultra-tight binding	- Very high protein consumption	Dissociation constants down to 10^{-20} M	Low
NC-MS (also called native MS; usually ESI-MS)				
Mass detection of protein-ligand complex in gas phase	- Direct visualization of complex formation - High sensitivity - Very low protein consumption - Accurate mass measurements - High-content information - No labelling required	- Requires sample desalting - Protein has to be stable in an ESI-MS suitable buffer - Detergents are not tolerated - Unspecific binding makes the detection of low-affinity binders more difficult	No lower limit to 500 µM	Medium

Table 1.1 (cont.). Summary of the main biophysical techniques employed in drug discovery

Technique or method	Strengths	Limitations	Affinity range	Throughput
AS-MS (also called SEC-MS)				
Incubation of protein in fragment mixture then separation of bound from unbound molecules by SEC, followed by MS detection	- Ultra-high throughput when compound cocktails used - Can be applied to solubilized membrane proteins - Ligand mass detection enables verification of compound structure	- Low-affinity binders (including fragments) are hard to detect because they tend to dissociate from the protein during the SEC step owing to high off-rates	No lower limit to 10 µM	Ultra-high
HDX-MS				
Ligand binding affects deuteration rate of protein residues, which is detectable by mass	- Direct detection of protein binding site and/or changes in protein conformation or protein dynamics upon binding	- Spatial resolution is limited and depends on peptide lengths and coverage of target sequence after protease cleavage	No lower limit to 20 µM	Low
MST				
Change in the molecular motion of the target in a temperature gradient due to ligand binding	- In-solution measurements - Applicable to solubilized membrane proteins	- Requires labelling or strong intrinsic fluorescence	1 pM to 1 mM	Medium

Data extracted from [65] and [66]. Abbreviations: AS-MS, affinity selection mass spectrometry; DSC, differential scanning calorimetry; DSF, differential scanning fluorimetry; ESI-MS, electrospray ionization mass spectrometry; HDX-MS, hydrogen–deuterium exchange mass spectrometry; ITC, isothermal titration calorimetry; MS, mass spectrometry; MST, microscale thermophoresis; NC-MS, non-covalent mass spectrometry; NMR, nuclear magnetic resonance; SEC-MS, size-exclusion chromatography mass spectrometry; SPR, surface plasmon resonance.

1.6.1. Differential scanning fluorimetry

Proteins exist in thermodynamic equilibrium between multiple conformational states. In the simplest case we can consider two different states of a protein: the folded or native and the unfolded or denatured. In most cases, the population of the unfolded state increases

upon an increase of temperature. Binding of a small molecule to the protein usually alters the population of these states[67]. Differential scanning fluorimetry (DSF), also referred as thermal shift assay, is a biophysical method based on this principle. Indeed, it consists of the controlled increase of temperature of the protein solution in the presence of an environmentally sensitive fluorophore (in most cases is SYPRO orange). This supposes a gradual increase of the population of the unfolded state, making the protein expose its solvent hydrophobic residues normally buried inside its core. This hydrophobic environment, makes the fluorescent signal increase significantly. Thus, it makes possible to calculate the denaturing process of the protein over temperature[68]. The melting temperature (Tm) of the protein is calculated as the midpoint of the resulting fluorescence versus the temperature (Figure 1.6 A). The Tm is usually represented by plotting the derivative of the fluorescence signal against the temperature (Figure 1.6 B). In order to assess small molecule binding, Tm of the protein is compared with the Tm of the protein in solution with the small molecule[69] (Figure 1.6). Binding of a small molecule to a protein supposes in most cases an increase of its stability, despite it has also been possible to observe protein destabilization in known binders[70].

DSF is widely used in high-throughput assays for hit identification for the ease of implementation in microplate formats using a real-time thermal cycler instrument[67]. A part form its applicability as a primary screening, DSF can also be used for buffer optimization to identify optimal storage conditions, for assay screenings or crystallization experiments. DSF has also been implemented to measure thermal stability of proteins in cells and tissues, what is known as cellular thermal shift assay (CETSA)[66]. Notwithstanding, DSF also presents some limitations such as artefacts, related to fluorescent quenching or aggregation. Sometimes the stabilization of the small molecule can be too low to be measured. These facts make DSF a source of false-negative and false-positive hits. Besides, it is not suitable for all types of proteins or complexes[65]. Taking everything into consideration, DSF is a cost-effective, parallelizable, practical, and accessible biophysical technique, which make it be a widely used method in primary screenings of small molecules[66].
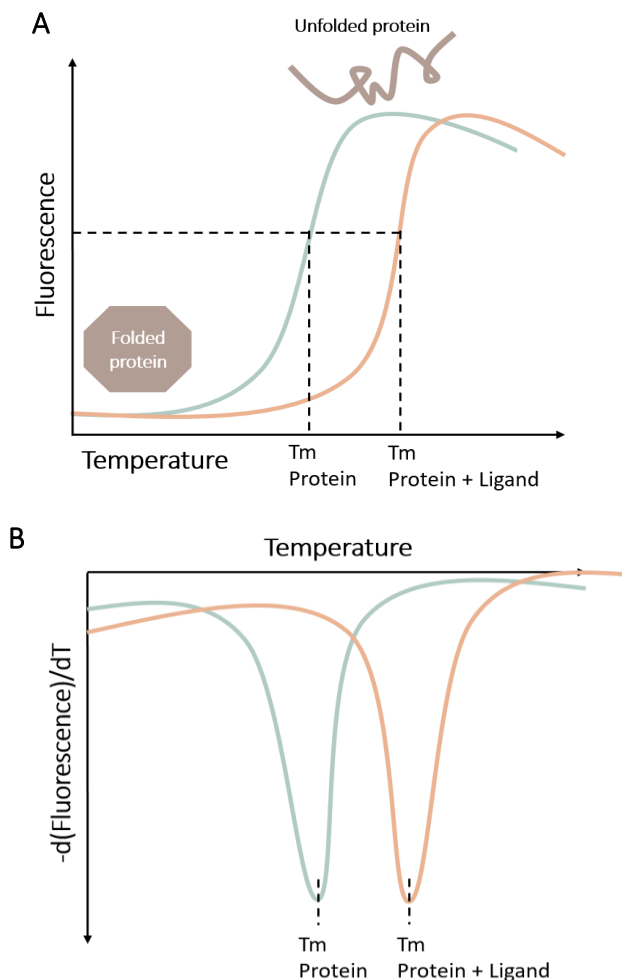
Figure 1.6. Typical DSF curves. A. Typical recording of fluorescence intensity versus temperature for the unfolding of protein in the presence (orange) and absence (green) of small-molecule ligand. Melting temperature (Tm) corresponds to the point of inflection of the curve. **B.** Tm can be easily assessed by plotting the derivative of the fluorescence signal against the temperature.

1.6.2. Surface plasmon resonance

Surface plasmon resonance (SPR) is an optical technique that allows real-time, label-free detection of biomolecular interactions. In SPR, protein is usually covalently attached to a chip, by means of a dextran matrix linked in the solution side of a gold film of the chip. In a standard SPR set up, a beam of polarized monochromatic light is shone through a prism at the

thin-layer of gold coating one surface of the prism[67]. This generates an evanescence wave that extends up to 300 nm in solution. The light will be reflected and recorded at a particular angle of incidence (Figure 1.7 A and B). The binding of the small molecule (analyte) to the protein immobilized at the surface of the fold layer, will suppose a shift in the resonance angle that can be monitored in real time in a so-called sensogram (Figure 1.7 C). In the sensogram, the resonance or response units (RU), corresponding to changes in the refractive index, are plotted over time in seconds[71]. Since small molecules interactions are measured in real time, SPR sensograms can give information about the rate of interaction (association (k_{on}), dissociation (k_{off}) or both). Besides the binding level achieved in the steady-state can be measured at different concentrations to extract the affinity of binding (K_d)[72].

Low sample consumption and wide range of binding detection (from 1 nM to 500 µM) are two of the strengths of SPR. Besides, it allows direct time-resolved determination of interaction. However, the SPR signals can also be affected by solvent effects. Covalent immobilization of the protein is one of the major limitations of the technique[65]. This can be overcome by performing different immobilization protocols, involving the immobilization of protein tags, for example. These protocols are widely established and special chips and immobilization protocols are available.

SPR can be used in different stages of the early drug discovery phase. It can be used for hit identification not only of drug-like molecules, but also fragment-like. Binding level obtained at a single concentration can be used to perform an initial screen of potential binders. The selected subset can be further studied at a dose-response manner to select best candidates considering stoichiometry, affinity and kinetics. Besides, competition assays can be performed to study binding site specificity. SPR can also be useful in hit-to-lead or lead optimization stages, mainly for the kinetic information provided. SPR allows the measurement of drug residence time ($1/k_{off}$), considered a relevant parameter to predict *in vivo* behavior of drugs[73].
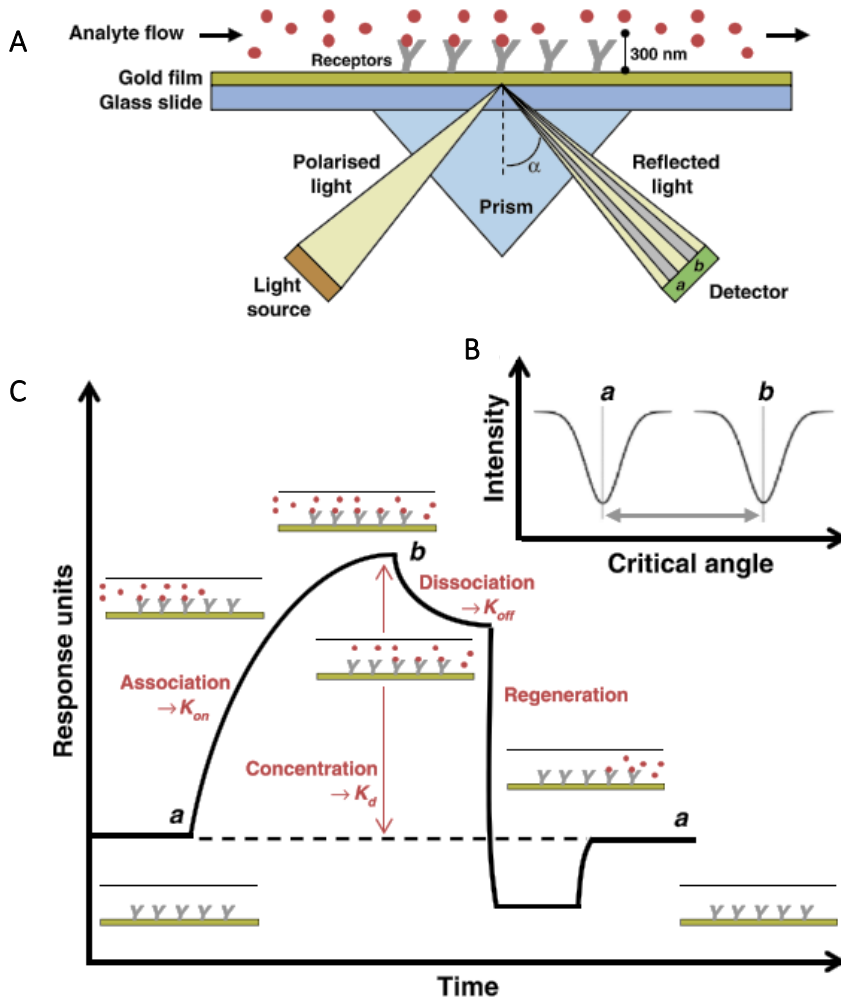
Figure 1.7. SPR principles. A. Schematic illustration of the basic SPR experiment representing binding measurement of a small molecule (red) to the immobilized protein (grey). Polarized light is directed through a prism to the under surface of the gold film, generating an evanescence wave that extends up to 300 nm in solution. The light is then reflected at a particular angle of incidence. Once small molecule (analyte in red) binds to the immobilized protein (receptor in grey) in the gold film refractive index changes (from a to b). **B.** Change in the critical angle of incident light from angle a (protein) to angle b (protein-small molecule interaction). **C.** Response of the SPR experiment in the form of a sensogram. Initially, protein (grey) has not been exposed to small molecules (analyte in red) and the critical angle is a. Once small molecules are injected the association can be observed (k_{on}). When a steady-state is achieved (all binding sites occupied in this example) the RU correspond to the changed final critical angle b. This is considered the observed Rmax and can be used to extrapolate the binding affinity (K_d). Finally, small molecules are removed and the dissociation can be measured (k_{off}). The surface can then be regenerated and returned to the critical angle a to start the experiment again. Figure modified from [71].

1.6.3. Microscale thermophoresis

Microscale thermophoresis (MST) is a fluorescent-based technique that permits the quantification of biomolecular interactions. As it name implies, it is based on thermophoresis, which relies on the fact that a temperature difference (ΔT) leads to a local change in molecule concentration (depletion or enrichment), quantified by the Soret coefficient (S_T): $c_{hot}/c_{cold}=\exp(-S_T\Delta T)$[74]. Thermophoresis strongly depends on a variety of molecular properties such as size, charge, hydration shell or conformation. Therefore, this technique is highly sensitive to any change of molecular properties and can be employed for the measurement of small-molecule interactions[75]. In short, MST assay consists on monitoring the movement of fluorescent molecules (using covalently attached or intrinsic fluorophores[76]) through a microscopic temperature gradient produced by an IR-Laser[77].

Before the IR-Laser is switched on, a homogenous molecule distribution is observed inside the capillary (Figure 1.8 B-I). Next, IR-laser turns on heating specific focal zones in the sample solution and leads to changes in fluorescence intensity known as T-jump (Figure 1.8 B-II). Then, thermophoretic movements of the molecules start and fluorescence intensity will decrease until it reaches the steady state, depending on molecular depletion out of the heated zone according to the typical thermophoresis, described as the movement of molecules from hot to cold zone (Figure 1.8 B-III). Thereafter, IR-laser switches off to induce mass diffusion of molecules, depending on concentration gradient, called back-diffusion state (Figure 1.8 B-IV). Binding quantifications are performed by analysing the change in fluorescence intensity, which is estimated as relative fluorescence (normalized fluorescence, F_{norm} = F_{hot}/F_{cold})[77]. Binding quantification is depicted in Figure 1.8 C.

MST is a very sensitive technique suited to measure a wide range of interactions, including binding of small molecule to proteins[75]. Indeed, it is able to measure a wide range of affinities, ranging from 1 pM to 1 mM[65]. MST present several advantages such as allowing in-solution measurements with no need of immobilization, simplicity and sensitivity[77]. Its main limitation is the requirement of labelling or strong intrinsic fluorescence. Overall, MST has been described to be a suitable biophysical technique for hit confirmation and validation[65].
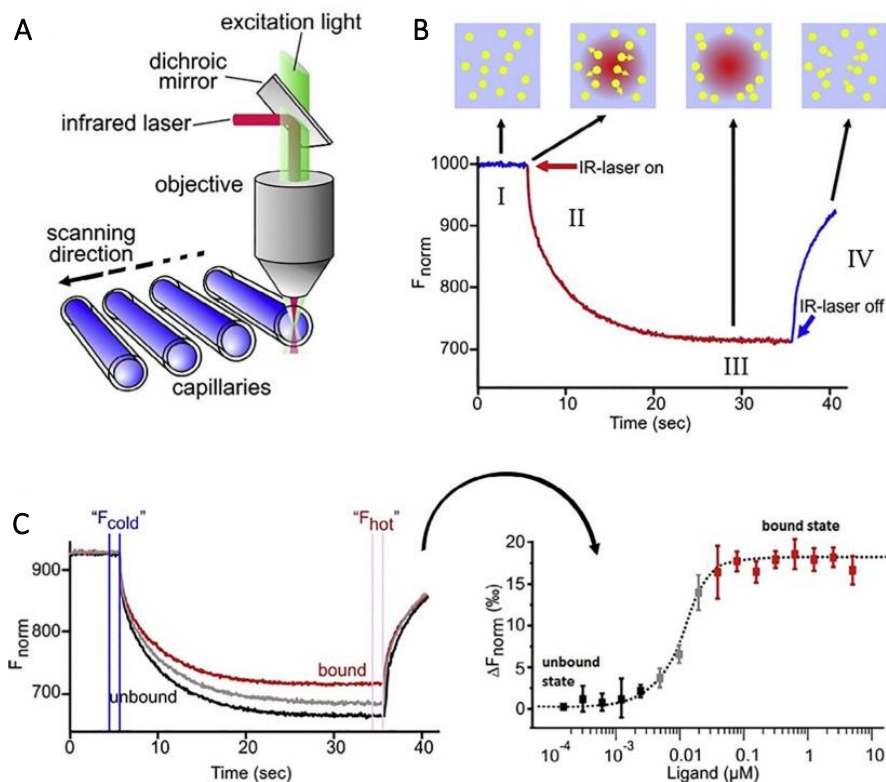
Figure 1.8. MST principles. A. Schematic setup of MST instrument. **B.** Thermophoresis signal in every step of the MST experiment. **C.** Thermophoretic signals for bound/unbound molecules translated into a binding curve. Adapted from [77].

1.6.4. Isothermal titration calorimetry

Isothermal titration calorimetry is a biophysical technique that measures the heat of a reaction to obtain its thermodynamic profile. Indeed, ITC is a label-free technique that provides direct measurement of heat (q) (either released or absorbed) in molecular binding during gradual titration. ITC experiments are performed on an instrument that contains two different cells, the reference cell and the sample cell. In the sample cell a syringe can be inserted (Figure 1.9 A). Usually the sample cell contains the protein, while the syringe is filled with the small molecule. An ITC experiment proceeds by injection of the small-molecule ligand into a temperature controlled stirred-cell containing the other component[67]. Different injections of small molecule will be titrated into the cell, supposing a release or absorption of heat. Indeed, what is directly measured in each injection is the differential power (DP), which

is the power applied in the reference cell to maintain zero temperature difference between reference and sample cell. An example of a raw data obtained from an ITC experiment is depicted in Figure 1.9 B. Peak areas from the raw data are integrated and plotted against the molar ratio of molecules in the syringe versus the cell in a Wiseman plot[78] (Figure 1.9 B). The Wiseman plot or isotherm provides de direct measurement of the difference of enthalpy (ΔH), the stoichiometry (n) and the binding constants (K_a and K_d). From those values, the free energy change (ΔG) and the entropy of binding (ΔS) can be calculated applying the following equation: $\Delta G = \Delta H - T \times \Delta S = R \times T \times \ln K_d$, where R is the gas constant and T the temperature[67].

A critical parameter that determines the shape of the binding isotherm is the so-called c-value, (c value = n x [protein] x K_d^{-1}, where n is stoichiometry and [Protein] the concentration of protein). It has been calculated that c value of minimum 10 is desired to obtain a sigmoidal shape[78]. For low affinity binders, the c value is also low. This could be compensated by increasing protein concentration, what would suppose increasing concentrations of the small molecule. Subsequently, the amount of protein needed and solubility of small molecule would suppose a limitation of the technique. Figure 1.9 C shows how the shape of the isotherm varies with the c value.

c
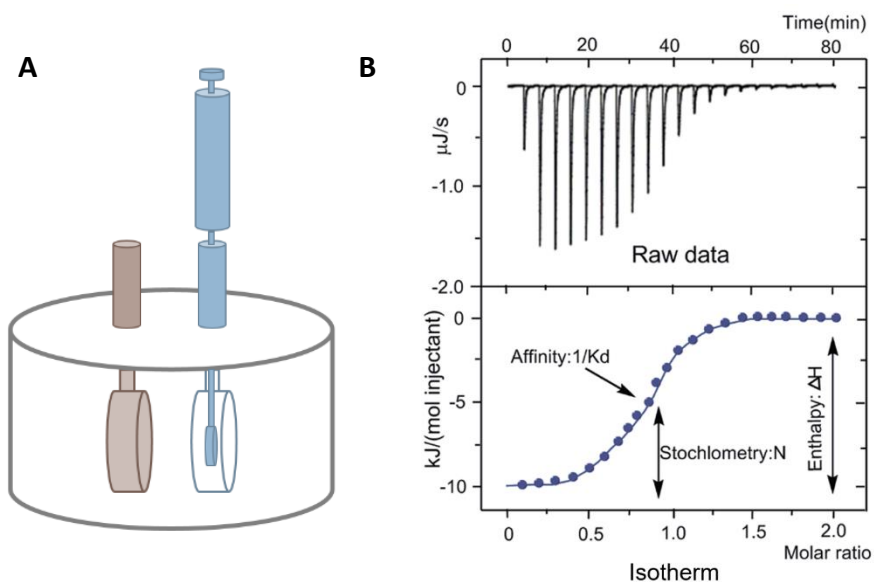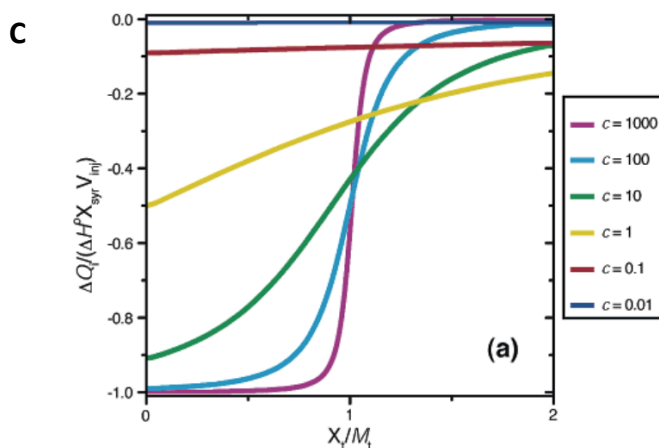
Figure 1.9. Basis of ITC. A. Schematic representation of ITC instrument with the reference cell in brown, the sample cell in white and the syringe in blue. **B.** On the top, raw data characteristic of an ITC experiment. On the bottom, binding isotherm showing the thermodynamic parameters that can be directly extracted. Adapted from [79]. **C.** Variation of the shape of the binding isotherm regarding c value. Adapted from [78].

The principal strength of ITC is that it allows the measurement of the full thermodynamic profile of a reaction in a single experiment. Therefore, it is not only possible to assess the affinity of the interaction, but also its binding mechanisms. For instance, if the reaction is primarily mediated by enthalpic contributions it can be related to hydrogen bonds or van der Waals interactions. In contrast, entropic contributions are related to hydrophobic effect of water release (favourable) or conformational changes and reduction in degrees of freedom (unfavourable)[80]. Additional advantages of the technique are high precision, reproducibility and sensitivity. Besides, the measurement is performed in solution without the need of labelling. In contrast, the main weakness of ITC is not being able to properly assess reactions with low enthalpic contributions. Besides, it is considered to be low throughput with high sample consumption, which is specially challenging for macromolecules (proteins) that are difficult to prepare in large quantities[80]. Finally, ITC cannot properly assess weak binders with K_d higher than 100 µM[65]. ITC is the gold standard for characterizing biological interactions. Despite its throughput, it is a suitable technique for hit-to-lead or lead optimization steps of the drug discovery process[81].

1.6.5. Fluorescence polarization

Fluorescence polarization (FP) is a fluorescence-based technique in which the fluorophore is irradiated with linearly polarized light. The resultant fluorescence intensity is measured through a polarization filter placed in front of the detector and oriented either parallel or perpendicular to the incident polarized light[82]. Measurements of polarization are related to the molecular mobility of the fluorescent probe, which is dependent of size and shape. In particular, they depend on the rotational correlation time of the fluorophore in solution. Usually, the fluorophore is a ligand of the protein you want to study. This ligand presents a high rate of rotation, which is altered when binding to the protein. As a consequence, the rate of motion decreases allowing a higher detection of polarization (Figure 1.10). Since not all small molecules are fluorescent, usually FP assays are performed in competition mode with a fluorescent (or fluorescent-labelled) known binder. The assay needs to be carefully optimized to ensure the proper evaluation of the small molecules to be screened[67].

FP is a biophysical technique performed in solution with selective but inexpensive reagents and equipment[83]. As a ratiomeric method, it is less sensitive to the absorptive interferences or inner filter effects, as well as, to the environmental interferences, such as pH changes. In contrast, other effects, such as autofluorescence and light scattering can be presented. Additional disadvantages include the requirement of a fluorescent ligand, either intrinsically or labelled, and the need of exhaustive assay optimization. Furthermore, with FP assays only compounds that affect interaction of the fluorescent probe can be assessed, which could lead to false negatives. In a whole, FP is a biophysical technique that has been widely used for hit identification, confirmation and validation[80, 67].
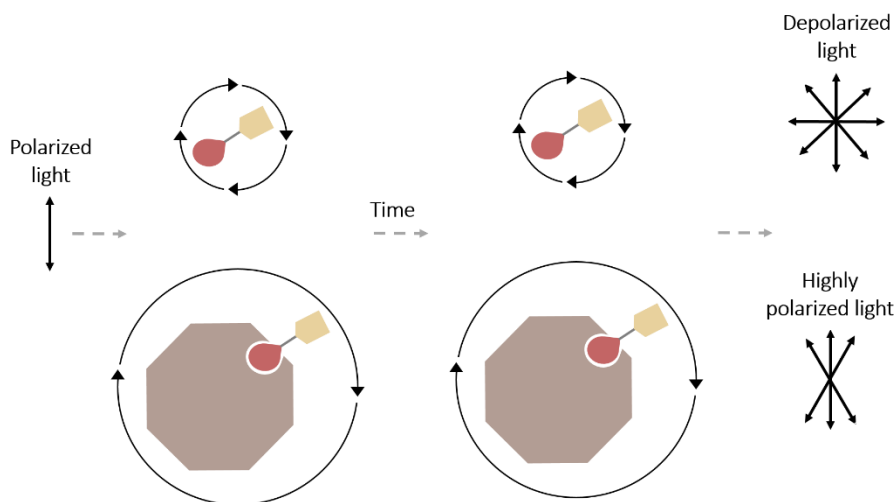
Figure 1.10. Schematic description of fluorescence polarization principle. When a fluorescent-labelled (yellow) ligand (red) is excited by polarized light, it rapidly rotates resulting in low polarization. However, when it is bound to a protein (brown), the rotation of the fluorescent probe is more slowly. This is translated into a highly polarized signal.

1.6.6. X-ray crystallography

X-ray crystallography has been the main methodology to determine the three-dimensional structure of macromolecules, including protein-small molecule complexes. Consequently, this technique allows the binding site and binding mode characterization of small-molecule drugs[84]. Nevertheless, obtaining this valuable information is not usually a simple pathway, and several steps need to be successfully accomplished (Figure 1.11). Beforehand, the protein of interest needs to be produced and purified in relevant amounts for crystals production[85]. In some cases, optimization of this step will be required by modifying protein construct or protein production protocol[86]. Indeed, before studying protein-small molecule complexes, the protein target must be crystallized and its structure solved[67]. Then, in order to obtain crystals of the complex, two different techniques can be applied: soaking and co-crystallization. Soaking consists on the introduction of the ligands into pre-existing apo crystals, whereas co-crystallization involves adding the ligand to the protein to form a complex that is subsequently crystallized[87]. Obtained crystals are exposed to an X-ray source and diffraction patterns are recorded. This information is used to construct the electron density map of the crystallized entity[67]. The resolution of these maps is dependent of

the quality of the crystal[86]. To determine if a small-molecule is bound to a protein, electron density map of the protein alone is compared with the one of the complex. Further studies of these differences and fitting of molecular models permit the identification of small molecule binding mode[67].

PROTEIN PRODUCTION
- Construct optimization
- Suitable purification protocol

CRYSTAL GROWTH
- Crystallization screening
- Soaking or co-crystallization for complexes

X-RAY DIFFRACTION
- Crystal mounting
- Data collection of diffraction patterns

DATA PROCESSING
- Electron density maps

3D STRUCTURE DETERMINATION
- Structure elucidation
- Refinement
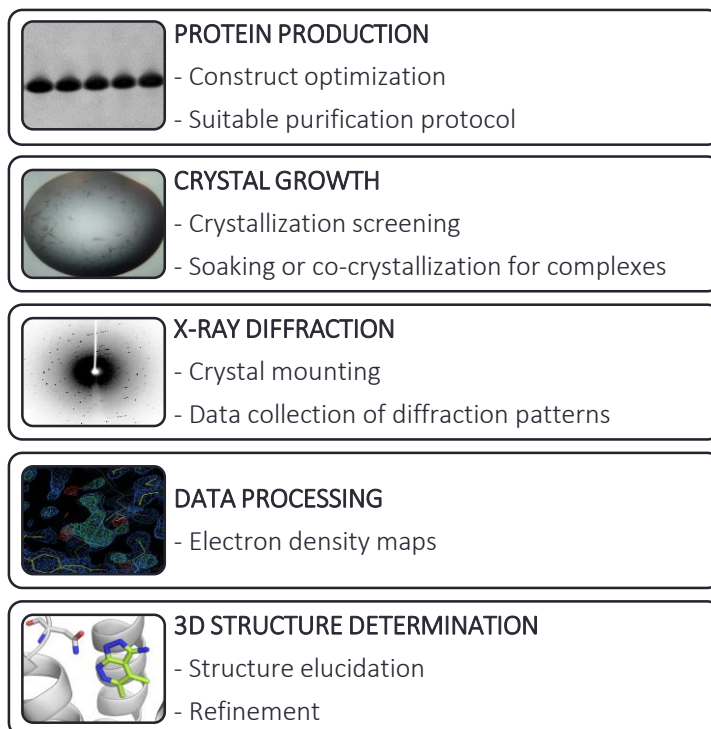
Figure 1.11. Schematic representation of steps needed to obtain a 3D structure of protein-small molecule complex.

The initial challenge in X-ray crystallography is obtaining a protein crystal suitable for structure elucidation. This step usually requires exhaustive optimization in the protein production steps. Indeed, even if this is achieved, obtaining crystals of the complex is also challenging. Soaking the apo protein crystal is the most straightforward option. Nevertheless, it does not always work and co-crystallization screenings need to be performed to determine the best conditions. Besides, the electron density map developed needs to be of enough quality and the density of the ligands needs to be high enough to enable fitting the ligand in the structre[86].

Despite not providing information about ligand affinity, X-ray crystallography is a technique widely used in early drug discovery, especially in hit-to-lead or lead optimization steps[67]. More recently, it has also been used as a primary screening, particularly for fragment-based screening campaigns, since the information obtained allows the direct optimization of the hits[86].

CHAPTER 2: OBJECTIVES

OBJECTIVES

2.1. General objective

Tackling unmet medical needs is the main purpose of drug discovery. Cancer is nowadays a disease that causes millions of deaths every year, entailing a need to develop novel treatments to combat it by modulating until now considered undruggable targets. In this scenario, the general objective of the present thesis is the implementation of structure-based approaches in combination with biophysical techniques for the identification and characterization of small molecules able to modulate challenging targets in cancer. These small molecules could be employed to elucidate which is the more convenient strategy to manipulate the targets described in this thesis.

2.2. Specific objectives

OBJECTIVE 1. E3 ligases have been described as relevant targets in cancer. Besides, the irruption of the targeted protein degradation technology has situated this target family in the forefront. In the present thesis we aim to **apply a structure-based approach in order to study E3 ligases ligandability**. Being FBW7 E3 ligase one of the most mutated proteins in cancer, we also aim to **use the previous structural information to identify and characterize small molecules that bind (binders) to this E3 ligase** by combining computational and biophysical techniques.

OBJECTIVE 2: Both inhibition and activation of the undrugged epigenetic enzyme TET2 would suppose a benefit for different cancer treatment scenarios. Potential TET2 modulators were previously identified following a structure-based approach in our Lab. In the present thesis we aim to **develop and characterize novel series of TET2 modulators applying biophysical techniques**.

OBJECTIVE 3: Bromodomains have been recently described for their interest in cancer. Specifically, BRD4 has also been used as a test system for computational techniques due to its ease of production and constant behaviour. Computer-aided drug design faces several challenges, being the prediction of solvation preferences and fragment evolution two of them. In the present thesis, we aim to **apply computational tools to study**

the solvation preferences of BRD4 BD1 and to develop and characterize novel chemical entities according to that. Besides, we also aim to characterize evolved fragments binding to BRD4 BD1 identified by an automated fragment evolution platform. The resulting information will help to validate the applied computational tools and to develop novel BRD4-based therapies.

CHAPTER 3: MATERIALS AND METHODS

MATERIALS AND METHODS

3.1. General materials and methods

3.1.1. General materials

3.1.1.1. Reagents

General reagents and consumed materials were purchased from suppliers Sigma-Aldrich, VWR and Fisher Scientific. For SD-PAGE, Mini-PROTEAN TGX precast gels, Precision Plus Protein™ Dual Colour Standards and Coomassie Brilliant Blue were purchased from Bio-Rad. Tobacco etch virus (TEV) protease was purchased from Sigma-Aldrich, while thrombin protease was purchased from Fisher Scientific. ULP1 protease was a kind donation of Dr Miquel Coll (IRB-CSIC, Barcelona).

3.1.1.2. Bacterial strains and growth media

XL1 blue competent cells (Agilent) and MC061 cells were used for DNA preparation, while *Escherichia coli* (*E. coli*) ROS(DE3) were used for protein production. MC061 and ROS(DE3) were a kind donation of Dr Raimon Sabaté (Universitat de Barcelona). Luria-Bertani (LB) broth and LB agar were purchased from VWR.

3.1.2. Preparation of competent *E. coli* cells

A small amount (around 10 μL) of frozen competent *E. coli* cells was plated in a LB agar plate without antibiotic and incubated at 37ºC over-night (O/N). A single colony was picked and added to a sterile 50 mL falcon tube with 15 mL of LB broth for further incubation at 37 ºC and 200 rpm. Once the bacterial culture reached an Optical Density of 600 nm (OD_{600}) value of 0.6, the bacterial cells were cooled on ice for 10 minutes. Cells were then harvested by centrifugation at 6000 rpm and 4 °C for 3 minutes. The pellet was gently resuspended in 10 mL of pre-cooled 0.1 M $CaCl_2$ and incubated on ice for 20 minutes. Another cycle of centrifugation (6000 rpm, 4 °C and 3 minutes) was carried out to resuspend the pellet in 5 mL of pre-cooled 0.1 M $CaCl_2$ containing 15% (v/v) glycerol. The resulting cells were dispensed as 200 μL aliquots into micro-centrifugation tubes and stored at -80 °C until further use.

3.1.3. Determination of DNA concentration, protein concentration and cell growth

UV-light absorption of DNA has a maximum at 260 nm as a consequence of the aromatic rings of its bases. For that reason, DNA concentration was determined by measuring UV-light absorption at 260 nm using a NanoDrop spectrophotometer (Thermo-Scientific). In a similar way, protein concentration was determined by measuring UV-light absorption at 280 nm, since the presence of tryptophan allows proteins to their maximum of UV-light absorption at 280 nm.

Cell growth was also assessed using a NanoDrop spectrophotometer (Thermo-Scientific) by measuring the amount of light scattering by microbial cells (Optical Density) at 600 nm (OD_{600}).

3.2. Materials and methods of Chapter 4: Disentangling E3 ligases ligandability: application to FBW7

3.2.1. Computational tools

3.2.1.1. Solvent-mixed molecular dynamics simulations with MDMix to study E3 ligases ligandability

Prior to perform MDMix[53], E3 ligase structures were directly downloaded from Protein Data Bank (PDB)[88]. Protein preparation was performed with MOE 2016[89], consisting on cleaning crystallography artefacts, correcting missing side chains, building-up loops up to 6 missing amino acids, capping gaps bigger than 6 amino acids, selecting correct protonation states, etc.

Three replicas of 50 ns of solvent-mixed molecular dynamics simulations were carried out for each E3 ligase. The mixed solvent used was ethanol-water (1:4). The simulations were performed using an octahedral solvation box, at 300K, 1 atmosphere and saving trajectories each 1500 steps (3 picoseconds). Soft constraints on the protein heavy atoms (0.01 kcal·mol^{-1}·Å^{-2}) were applied, allowing certain mobility. Simulations were performed using ff14SB amber force field. Resulting trajectories were aligned. The ethanol co-solvent was divided into two chemical probes: one polar (corresponding to the ethanol's hydroxyl group) and the other hydrophobic (corresponding to the terminal methyl), followed by the

calculation of a density map for each probe. The observed density distributions were converted into free energy applying the Boltzmann relationship. From the high energetic regions in the energy grids (the 0.02 percentile with the best free energy) a hierarchical clustering was used to extract the binding hotspots.

3.2.1.2. Virtual screening to determine potential binders for the E3 ligase FBW7

3.2.1.2.1. Docking-based virtual screening

Docking-based virtual screening with pharmacophoric restraints was performed with rDock[90]. A snapshot taken from MDMix of FBW7-SKP1 structure was selected in order to carry out the virtual screening. The structure was prepared using MOE[91], by removing water and solvent molecules. The cavity used for docking was defined in the prepared structure considering 8 Å radius from MDMix hotspots of FBW7-Pocket G.

A collection of 6 million purchasable compounds were docked into FBW7-SKP1. Initially, three pharmacophoric restraints, considering MDMix hotspots, were applied consisting of: a hydrogen bond acceptor with FBW7 N635 and two hydrophobic interactions close to FBW7 W365, F636 and I361. Pharmacophore tolerance was defined to 0.5 Å radius around the pharmacophoric points that related to hydrogen bond and to 1 Å radius around for hydrophobic interactions. If the compound conformation did not have features accomplishing the positional constraints, rDock would assign a positive restraint score (unfavourable). The high-throughput virtual screening (HTVS) protocol consisted in three different stages in order to increase the efficiency of the simulation. In each stage the compounds were filtered depending on the "SCORE.INTER" (general docking score) and the "SCORE.RESTR.PHARMA" (pharmacophoric restraint score). First, compounds passed 2 rounds of docking, discarding the ones with "SCORE.INTER" higher than -18 kJ/mol and "SCORE.RESTR.PHARMA" higher than 5 kJ/mol. In the second round, the "SCORE.INTER" and the "SCORE.RESTR.PHARMA" should be less than -23 kJ/mol and 3 kJ/mol, respectively, in order to complete the 50 docking runs (step three). Compounds with "SCORE.INTER" and "SCORE.RESTR.PHARMA" less than

-23 kJ/mol and 0.4 kJ/mol, respectively, were selected (3 restraints docking).

On one hand, an additional HTSV protocol was run with the resulting compounds but adding an extra pharmacophoric restraint. It consisted of a hydrogen bond donor with FBW7 S678. The same HTSV restraints and filters were applied for the following set of compounds, (4 restraints docking). These molecules were then filtered by dynamic undocking (DUck[92]). On the other hand, the compounds from the 3 restraints docking were filtered in order to select a representative subset that allowed DUck simulations.

3.2.1.2.2. Compound clustering of hits from 3 restraints docking

Compounds from the 3 restraints docking were clustered in order to perform DUck. For that, the 10000 compounds with the best docking score ("SCORE.INTER") and the 1000 compounds with the best ligand efficiency ("SCORE.INTER.NORM") were selected. The first set of compounds were clustered with Reynolds MOE clustering script[93] taking into account 85% similarity. The top 1500 clustered compounds according to "SCORE.INTER" were merged with the selected from "SCORE.INTER.NORM" and clustered using MACCS fingerprint[94] by taking into account 85% similarity.

3.2.1.2.3. Applying DUck to filter docking hits

Compounds from both HTVS campaigns were filtered by DUck[92] using the hydrogen bond produced by an acceptor group of the compound and FBW7 N635. Moreover, the compounds resulting from the 4 pharmacophoric restrains docking, were also filtered considering the hydrogen bond produced by a donor group of the compound and FBW7 S678. For both interactions studied, the same protein chunk was prepared, which is the part of the protein structure involved to create the necessary environment for the ligand interaction. When selecting the residues of the chunk we considered the following guidelines: (1) Selecting the minimum number of residues necessary in order to reduce computational time; (2) Not selecting residues that would block the ligand from exiting the pocket; (3) Not removing residues if this would lead to the possibility of solvent entering the pocket from areas other

than where the ligand is exiting; (4) Preserving the local environment of the interacting atoms. The sequence gaps produced when selecting the chunk were capped with acetyl and N-methyl groups in order to prevent charged ends and unnatural electrostatic forces. Finally, possible clashes produced when capping were checked. The selected chunk included the following residues: GLU328, GLY329, ILE 330, GLN358, HID359, ARG360, ILE361, ASP362, THR363, ASN364, TRP365, ARG366, CYS390, LYS590, GLN631, PHE632, ASN633, LYS634, ASN635, PHE636, VAL637, ILE638, LEU648, TRP649, ASP650, LEU651, LYS652, LEU660, ILE675, ARG676, ALA677, SER678, ASN679, THR680, LYS681, LEU682, VAL683, LEU700, ASP701, PHE702. Ligands were parameterized using a MOE[91] svl script. The steered-molecular dynamics (SMD) consisted of four different steps: I) minimization, II) equilibration, III) SMD simulations at two different temperatures (300 K and 325 K), in which the hydrogen bond of the binder is pulled to go from 2.5 Å to 5.0 Å, IV) Calculation of the W_{QB} value (work necessary to break the hydrogen bond) and application of a threshold (W_{QB} = 4 kcal/mol) to repeat steps III and IV during 6 cycles (6 replicas). If the threshold value was not achieved in one of those cycles, SMD was discontinued. The lowest W_{QB} value of the 6 replicas was the one considered. SMD steps were performed with GPU-based pmemd.cuda in AMBER.

3.2.2. Molecular cloning

3.2.2.1. Protein construct and expression plasmids

FBW7-SKP1 encoding construct was a kind donation of Dr Bing Hao (UconnHealth, USA), being designed to express the glutathione S-transferase (GST)-tagged human FBW7 (residue 263-707) and truncated SKP1. The plasmid also contains a thrombin protease cleavage site and confers ampicillin resistance. Plasmids containing $FBW7_{N635A}$-SKP1, $FBW7_{N635I}$-SKP1, $FBW7_{A677I}$-SKP1, $FBW7_{A677F}$-SKP1 and $FBW7_{N679W}$-SKP1 were generated in house using QuickChange II site-directed mutagenesis kit (Agilent) to make single-point mutations to FBW7-SKP1 plasmid.

3.2.2.2. Kits: Site-directed mutagenesis of Fbw7-Skp1 mutants

To perform site-directed mutagenesis of FBW7-SKP1, QuickChange II site-directed mutagenesis kit (Agilent) was employed. In order to carry

out DNA amplification, GeneJet plasmid miniprep kit (Fisher Scientific) was used.

Primers design was performed following QuickChange manual guidelines and purchased in biomers.net. The polymerase chain reactions (PCR) were performed on 9700 GeneAmp® PCR System (Applied Biosystems®). In order to digest parental methylated and hemimethylated DNA, Dpn I was added to the mixture. The resulting product was transformed to XL1-Blue competent cells. A single colony of transformed cells was picked and cultured over-night at 37ºC and 150 rpm in 10 mL of LB medium supplied with 100 µg/L of ampicillin. DNA was purified using GeneJet plasmid miniprep kit (Fisher Scientific) and sequenced by the Genomics Service of the *Centres Tecnològics de la Universitat de Barcelona* (CCTiUB) to confirm the presence of mutations.

3.2.3. Transformation of *E. coli* cells

Competent *E. coli* cells were thawed on ice for 30 minutes. Afterwards, 50 µL of cells were mixed with 5 ng of vector-DNA and kept on ice for other 30 minutes. Heat shock was carried out for 1 minute at 42 ºC followed by 5 minutes incubation on ice. 500 µL of LB without antibiotic were added to transformed *E. coli* cells and the mixture was incubated at 37 ºC and 400 rpm for 1 hour. Then, 100 µL of bacterial cells were spread on LB-agar plates that contained 100 µg/L of ampicillin and incubated over-night at 37 ºC.

3.2.4. Protein expression: Fbw7-Skp1, Fbw7$_{N635A}$-Skp1, Fbw7 $_{N635I}$-Skp1, Fbw7$_{A677I}$-Skp1, Fbw7$_{N679W}$-Skp1 and Fbw7$_{A677F}$-Skp1 complexes

A single colony of transformed cells was picked and cultured over-night at 37ºC and 150 rpm in 10 mL of LB medium supplied with 100 µg/L of ampicillin. The following day, the 10 mL of grown cells were transferred to a flask containing 1 L of LB media also supplied with 100 µg/l of ampicillin. The flask was incubated at 37 ºC and 180 rpm until reaching an OD$_{600}$ between 0.6 and 0.8. Temperature was then decreased at 18 ºC and cells were induced with 1 mM IPTG. *E. coli* cell culture was incubated at 18 ºC and 180 rpm for 12 to 18 hours. Cells were harvested by centrifugation at 3700 rpm and 6 ºC during 30 minutes. Cell pellets were stored at -20 ºC before use.

3.2.5. Cell lysis and lysate clarification

While cell lysis consists on cell disruption in order to release intracellular materials, lysate clarification implies the separation of these materials to have the proteins soluble into the lysis buffer.

Cell pellets were defrosted at room temperature followed by a resuspension with 15 mL of protein buffer A (Table 3.1) supplemented with Pierce protease inhibitor cocktail (ThermoFisher Scientific). Cell lysis was carried out by sonication during 2 minutes in intervals of 10 seconds of sonication followed by 30 seconds of break at 19 ºC. Lysates were then clarified by double centrifugation at 4 ºC and 8800 rpm during 30 minutes. Supernatant was collected and filtered using 0.8 µm syringe filters (Sigma-Aldrich) in order to be used for protein purification.

Table 3.1. Protein purification buffers for FBW7-SKP1 purification

Name	Buffer	Protein purified
A	50 mM Hepes pH 8.0, 200 mM NaCl, 5 mM DTT	FBW7-SKP1
B	50 mM Hepes pH 8.0, 200 mM NaCl, 20 mM Glutathione, 5 mM DTT	FBW7-SKP1
C	50 mM Hepes pH 8.0, 50 mM NaCl	FBW7-SKP1
D	50 mM Hepes pH 8.0, 1 M NaCl	FBW7-SKP1

3.2.6. Protein purification: Fbw7-Skp1, Fbw7$_{N635A}$-Skp1, Fbw7 $_{N635I}$-Skp1, Fbw7$_{A677I}$-Skp1, Fbw7$_{N679W}$-Skp1 and Fbw7$_{A677F}$-Skp1 complexes

FBW7-SKP1 complexes were all purified following the same three-step purification procedure by using an ÄKTA start system (Cytiva). The protein obtained after lysate clarification was applied to a 5 mL GSTrap 4B column (Cytiva), washed with buffer A (Table 3.1). Bound protein to the column was eluted with buffer B (Table 3.1). Thrombin protease was added to the resulting protein fraction, for the GST cleavage. The mixture was dialyzed overnight at 4 °C against buffer A (Table 3.1) using a 3.5 kDa dialysis membrane (Spectra/Por). The solution with the uncleaved protein was introduced to a second GST purification step,

performed as described above. The flow through was collected and dialyzed against Buffer C (Table 3.1) to be loaded onto 5 mL anion exchange Hitrap Heparin column (Cytiva). The protein was eluted at the flow-through and its mass and purity were subsequently verified by SDS–PAGE and mass spectrometry.

3.2.7. Biophysical techniques

3.2.7.1. Differential scanning fluorimetry to determine binding to FBW7-SKP1

Differential scanning fluorimetry (DSF) experiments were performed using a Light Cycler 480 II (Roche Applied Science) at the genomics service (CCTiUB).

Prior to interaction assessment, the setup of the experiment was performed. For that, different concentrations of protein were tested, ranging from 1 µM to 5 µM, as well as different concentrations of the SYPRO Orange dye (2.5x and 5x). Besides, buffer screening was also performed testing the different buffers shown in Table 3.2.

Table 3.2. Buffers employed to perform DSF optimization

Name	Buffer
DSF-B1	50 mM sodium acetate pH 4.6, 200 mM NaCl
DSF-B2	50 mM potassium phosphate pH 6.0, 200 mM NaCl
DSF-B3	100 mM Hepes pH 7.5, 1.12 M lithium sulphate
DSF-B4	50 mM TRIS pH 8.0, 50 mM NaCl
DSF-B5	50 mM TRIS pH 8.0, 50 mM NaCl
DSF-B6	50 mM Hepes pH 8.0, 50 mM NaCl
DSF-B7	50 mM Hepes pH 8.0, 50 mM NaCl, 0.5 M urea
DSF-B8	50 mM Hepes pH 8.0, 50 mM NaCl, 1 M urea

The buffer selected to perform compound screening was DSF-B6 (Table 3.2). FBW7-SKP1 was tested at a final concentration of 5 µM, while compounds were diluted at final concentrations of 500 µM or 250 µM at 1% DMSO, depending on solubility. A 26-mer peptide of the natural substrate CYCLIN-E (sequence: KAMLSEQNRASPLPSGLL[pT]PPQ[pS]GKK;

K_d 70 nM[95]) was used as a positive control. The peptide was tested at three different concentrations: 500 µM, 250 µM and 100 µM, all of them at 1% DMSO. SYPRO Orange dye (Sigma-Aldrich) was added as a fluorescence probe at a dilution of 1:1000 (5x). Excitation and emission filters for the SYPRO-Orange dye were set to 465 nm and 580 nm, respectively. The screening was carried out in a 96-well plate at a final volume of 25 µL, raising the temperature from 20 ºC to 85 ºC in 0.6 ºC per minute steps. Data analysis was performed using the Light Cycler 480 software. Control experiments were carried out with protein and dye mixture, and buffer and dye mixture. When necessary, controls with compound and dye mixture were also performed. All experiments were performed in triplicates. To consider compound binding, thermal shifts (ΔTm) of the protein-ligand complexes (ΔTm=Tm protein-ligand complex – Tm protein) had to be at least twice the standard deviation of the Tm replicas of the protein. The melting curves and thermal shifts were visualized using Microsoft Excel spreadsheet.

3.2.7.2. FBW7-SKP1 surface plasmon resonance screening for identification and characterization of binders

All surface plasmon resonance (SPR) experiments were carried out at 25ºC using Biacore T200 SPR biosensor instrument (Cytiva) at the molecular interaction analysis service (CCTiUB).

Prior to the start of each new experiment, the whole system was cleaned using Biacore maintenance kit. Then, CM7 sensor chip (Cytiva) was inserted and preconditioned and normalized following the protocol proposed by the supplier (Cytiva). In all cases, immobilization was carried out using standard amine coupling procedure, which started with the activation of the carboxymethyl dextran matrix of the sensor chip with 0.1 M N-hydroxysuccinimide (NHS) and 0.4 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) at a flow rate of 15 µL/min for 7 minutes. The protein immobilization was then performed at 5 µL/min, using a 1:50 protein mixture diluted with 10 mM sodium acetate at pH 5.0. To determine the amount of protein immobilized, formula from Figure 3.1 was applied. Once the protein was immobilized, 1 M ethanolamine hydrochloride was injected for 7 minutes at 15 µL/min to block activated groups of the dextran matrix.

Phosphate buffered saline (PBS, 10 mM phosphate, pH 7.4, 150 mM NaCl) was used as immobilization running buffer.

SPR PROTEIN IMMOBILIZATION

$$R_L = \frac{MW_L}{MW_A} \times \frac{R_{max}}{s}$$

R_L: Response level (RU) of immobilized ligand

MW_L: Molecular weight of ligand (protein)

MW_A: Molecular weight of analyte (compound)

R_{max}: Maximum binding capacity

s: Stoichiometry (number of binding sites per analyte)

Amount of protein to immobilize in order to obtain a R_{max} value of approximately 50 RU

Figure 3.1. SPR protein immobilization. Formula applied to calculate the amount of protein immobilized on the surface of the SPR chip.

All the screenings were performed at a flow rate of 60 µL/min and the compound association and dissociation times were set to 60 seconds and 120 seconds, respectively. In all screenings, solvent correction was employed.

The Biacore T200 evaluation software 2.0 was used for data analysis. Nonspecific binding to the chip surface and the baseline drift were corrected subtracting the signals of the reference surface (where the immobilization procedure was carried out without proteins) to the signals obtained on the surfaces with protein. Artefacts derived from DMSO interferences were corrected using series of solvent standards (solvent correction). Background signals were corrected subtracting blank injections (blank subtraction) to the injected compound signals.

3.2.7.2.1. FBW7-SKP1 screening for identification of binders

Potential binders from virtual screening cascade were screened to determine FBW7-SKP1 binding. For this initial screening, the four channels in the chip were used: one with no protein immobilized but treated to block the dextran (reference) and the others with low and high density of protein to determine the optimal R_{max} considering signal/noise ratio. The forth channel was used to discard unspecific binding with the dextran matrix. To determine the densities of channel

2 and 3, formula from Figure 3.1 was applied to have the high-density channel with an expected R_{max} of 50 RU and the low-density channel with an expected R_{max} of 25 RU.

The 41 hits from virtual screening were initially tested at 3 different concentrations: 100 µM, 10 µM, 1 µM at 5% DMSO. For that, 20 mM stock solutions were prepared by dissolving compound powder in 100% DMSO. A bank of dilutions was prepared in 100% of DMSO from which desired samples concentrations were obtained by diluting them with 1.05x PBS, 0.05% tween-20. The running buffer employed consisted of 1x PBS, 0.05% tween-20, 5% DMSO. For the solvent correction, 8 dilutions that ranged from 3% to 8% DMSO in 1x PBS and 0.05% tween-20 were prepared.

3.2.7.2.2. FBW7-SKP1 characterization of binders

Compounds that showed a dose-response sensogram in FBW7-SKP1 were screened in a wider range of concentrations in duplicates. While protein immobilization procedure was the same as Section 3.2.7.2.1, running buffer had to be adjusted in order to discard interactions with the dextran surface matrix. The optimized running buffer consisted of 1.05x PBS, 0.05% tween-20, 0.5 mg/mL dextran, 5% DMSO. Initially, compounds were tested at an initial 2-fold dilution series starting from 250 µM in 5% DMSO. For that, a bank of dilutions was prepared in 100% of DMSO from which desired samples concentrations were obtained by diluting them with 1.05x PBS, 0.05% tween-20 and 0.5 mg/mL dextran. Once there was an estimation of the K_d of the compounds, concentration range was adjusted for each compound. For the solvent correction, 8 dilutions that ranged from 3% to 8% DMSO in 1.05x PBS, 0.05% tween-20 and 0.5 mg/mL dextran were prepared. To estimate binding affinity, SPR data was fitted to a single interaction model, where steady state values were extracted from the sensograms recorded and plotted against the different concentrations assayed. If necessary R_{max} was fixed considering the R_{max} expected regarding the amount of protein immobilized in the chip surface (extrapolating it from Figure 3.1).

3.2.7.2.3. FBW7-SKP1 competitive SPR of best binders

SPR competitive assay was set up to determine if compounds interacted in the same binding site by following a protocol already published[96]. Fixed concentrations of compounds at saturation conditions were tested in triplicates to establish a SPR response and a standard deviation for compound. Mixtures of two compounds at the same concentration were also screened (in triplicates). In particular, compounds MMC2, MMC21, MMC35, MMC42, A1_MMC11, A2_MMC11, A3_MMC21 and A6_MMC17 were tested at 125 µM. MMC4, MMC11, MMC37, MMC40, A7_MMC2 and A8_MMC37 were tested at 250 µM. A5_MMC17 was tested at 200 µM, whereas A9_MMC37 and A10_MMC40 were tested at 50 µM. For compound preparation, 20 mM stock solutions were prepared by dissolving compound powder in 100% DMSO. Dilutions were prepared in 100% of DMSO from which desired samples concentrations were obtained by diluting them with 1.05x PBS, 0.05% tween-20 and 0.5 mg/mL dextran.

To determine competition, we used the following formulas:

$$\text{Degree of Competition (DoC)} = 1 - \frac{R_{A\text{-}B} - R_A}{R_B} \quad \textbf{(Eq. 3.1)}$$

$$R_{A\text{-}B} + SD_{A\text{-}B} \; < \; RTh_{A\text{-}B} - (SD_A + SD_B) \; \textbf{(Eq. 3.2)}$$

$R_{A\text{-}B}$ is the SPR response given by the mixture of compound A + compound B, R_A is the response given by compound A, R_B is the response given by compound B, $RTh_{A\text{-}B}$ is the theoretical SPR response if no competition (calculated by the sum $R_A + R_B$), $SD_{A\text{-}B}$ is the standard deviation of the triplicates of the mixture A-B, SD_A is the standard deviation of the triplicates of compound A, SD_B is the standard deviation of the triplicates of compound B.

For these experiments protein immobilization procedure was the same as Section 3.2.7.2.1, while buffers employed were the same as 3.2.7.2.2.

3.2.7.2.4. FBW7-SKP1 and mutants: binding site characterisation

Three different experiments were carried out comparing the interaction of the best binders among mutants. Channels contents for each experiment are listed in Table 3.3. For compound screening, the same

procedure as Section 3.2.7.2.2 was followed and the same buffers used for sample preparation and experiment running.

Table 3.3. Channel contents of SPR experiments carried out with FBW7-SKP1 and mutants

Experiment	Channel	Channel Content	R_{max} expected
FBW7-SKP1 and mutants – SPR experiment 1	Channel 1	Reference	-
	Channel 2	FBW7-SKP1	48
	Channel 3	FBW7$_{N635A}$-SKP1	45
	Channel 4	FBW7$_{N635I}$-SKP1	45
FBW7-SKP1 and mutants – SPR experiment 2	Channel 1	Reference	-
	Channel 2	FBW7-SKP1	45
	Channel 3	FBW7$_{N635I}$-SKP1	45
	Channel 4	FBW7$_{N679W}$-SKP1	64
FBW7-SKP1 and mutants – SPR experiment 3	Channel 1	Reference	-
	Channel 2	FBW7-SKP1	45
	Channel 3	FBW7$_{N635I}$-SKP1	55
	Channel 4	FBW7$_{A677F}$-SKP1	45

3.2.7.2.5. FBW7-SKP1 SPR mass transport control

Mass transfer is understood in SPR as the transfer of analyte (in our case drug-like small-molecules) from the bulk solution to the surface of the chip. Mass transfer can affect evaluation of binding, when the transfer is slow in relation to the association rate[97]. Therefore, if binding rate would depend on experiment flow rate, the SPR results would be limited by mass transfer. In order to verify that the results were not limited by this phenomenon, we analysed the interaction of a control compound at two different flow rates: 60 µL/min and 90 µL/min.

3.2.7.3. Microscale thermophoresis

Microscale thermophoresis (MST) experiments were performed on Monolith NT.115 (Nanotemper) at the laboratory of Dr Xavier Salvatella (IRB, Barcelona). Prior to performing MST experiments, protein labelling had to be carried out. For that, Monolith Protein Labelling Kit RED-NHS 2nd Generation (Nanotemper) was purchased, which dye carries a

reactive NHS-ester group that reacts with primary amines (lysine residues) to form a covalent bond. FBW7-SKP1 and FBW7$_{A677I}$-SKP1 were labelled following the corresponding protocol provided by Nanotemper. Both were diluted to obtain a concentration of approximately 10 µM with the Labelling Buffer NHS (130 mM NaHCO$_3$, 50 mM NaCl, pH 8.2-8.3). Dye RED-NHS 2nd Generation was added with a 2.5-fold excess of dye for labelling FBW7-SKP1 and with a 1-fold excess of dye for labelling FBW7$_{A677I}$-SKP1 (3-fold recommended). Protein concentration and its degree-of-labelling (DOL) were calculated following equation 3 and equation 4, respectively:

$$c = \frac{A_{280} - (A_{650} \times 0.04)}{\epsilon_{Protein} \times d} \quad \textbf{(Eq. 3.3)}$$

$$DOL = \frac{A_{650}}{195000 \times c} \quad \textbf{(Eq. 3.4)}$$

Where c is the protein concentration in molar units, ε Protein is the extinction coefficient of the protein in M^{-1}cm^{-1}, d is the path length d of the spectrophotometer in cm, A$_{280}$ the measured absorption at 280 nm, A$_{650}$ the measured absorption at 650 nm, 0.04 the correction factor at 280 nm and 195000 the molar absorbance of dye in M^{-1}cm^{-1}. To measure the absorptions, NanoDrop spectrophotometer (Thermo-Scientific) was employed, which normalizes the results for a path length of 1 cm. The DOL parameter indicates the amount of dye molecules are bound to the protein, meaning that a DOL of 1 refers to a dye:protein ratio of 1:1. DOL values greater than 1 were avoided to avoid adverse effects on protein function, whereas DOL values lower than 0.5 were avoided since it could lead to a reduced signal-to-noise ratio.

MST experiments were carried out at 25 °C, using Monolith NT.115 Premium Capillaries (Nanotemper). LED (light emitting diode) power was set to 40% or 60%, while MST power was set at 20% and 40%. All the experiments were carried out in a buffer that contained 50 mM Hepes (pH 8.0), 50 mM NaCl, 0.05% tween-20 and 1 mM DTT (MST buffer). Samples were prepared at a final volume of 20 µL and labelled FBW7-SKP1 final concentration was 50 nM. Compound stock solutions were prepared at 20 mM by dissolving its powder in 100% DMSO. In all cases, 2:1 dilution was carried out, starting at the maximum concentration possible regarding compound solubility at 5% DMSO. For that, a bank of

dilutions was prepared at 2x the desired compound concentration at 10% DMSO using MST buffer. The solutions were then diluted by 2x labelled protein concentration (100 nM).

Results were analysed using MO.Affinity Analysis v2.3 (Nanotemper). Data was evaluated using initial fluorescence mode, where raw fluorescence counts was plotted against compound concentration to obtain the K_d. In order to discriminate if initial fluorescence was caused by the compound interaction with the protein or caused by sample loss due to aggregation or surface adsorption, SD-test was carried out with the three highest and the three lowest concentrations of compound-protein mixture. It consisted on first centrifuging the samples at 15000 g during 10 minutes, to remove precipitates, followed by protein denaturation by performing 1:1 dilution of samples of 4% SDS and 40 mM DTT and incubation at 95 ºC during 5 minutes. Samples were loaded into capillaries and initial fluorescence was assessed. If no differences in fluorescence were determined between low and high compound concentrations after SD-test, it was concluded that the observed fluorescence change was induced by a binding event.

3.2.7.4. Isothermal titration calorimetry to characterize binding and determine binding site of A5_MMC17 in FBW7-SKP1

Isothermal titration calorimetry (ITC) assays were performed using Nano ITC Low Volume (TA Instruments) at the laboratory of Dr Maria Macias (IRB, Barcelona). All titrations were carried out at a temperature of 25 ºC. DMSO concentrations were the same in the sample cell with the protein solution and in the syringe with the compound, which were filled at final volumes of 320 µL and 50 µL, respectively. Reference cell was filled with 320 µL of milli-Q water (MQW). FBW7-SKP1 and FBW7$_{A677I}$-SKP1 were buffer exchanged in 50 mM HEPES pH 8.0, 50 mM NaCl. In all experiments an initial injection of 0.48 µL was performed followed by 20 identical injections of 2.5 µL. 30 µM FBW7-SKP1 solution was titrated against 600 µM of A5_MMC17 at 5% DMSO. In order to compare their binding affinities, 30 µM FBW7$_{A677I}$-SKP1 solution was titrated against 600 µM of A5_MMC17 at 5% DMSO. The heat of dilution was determined by carrying out the same experiment with buffer instead of

protein in the sample cell and was subtracted from the experimental data.

Data was analysed using the NanoAnalyze™ software (TA Instruments) to directly obtain the enthalpy of binding (ΔH) and the binding constants (K_a and K_d). From that, thermodynamic parameters were calculated using the following equation:

$$\Delta G = \Delta H - T \times \Delta S = R \times T \times \ln K_d \textbf{ (Eq. 3.5)}$$

Where ΔG, ΔH and ΔS are the changes in free energy, enthalpy and entropy of binding respectively, R is the gas constant (1.987 cal x K^{-1} x mol^{-1}) and T the temperature (in Kelvin). In all cases single binding site model was applied and the first data point was excluded from the analysis.

3.2.7.5. Fluorescence polarization

Fluorescence polarization (FP) experiments were performed on a ClarioStar reader (BMG LabTech) in NUNC 384-well plates using the excitation filter of 482-16, the emission filter of 530-40 and the dichroic LP504 filter. The fluorophore probe to be followed was a FICT labelled 16-mer diphosporilated peptide of DISC1, an endogenous substrate of FBW7, which K_d has been described in literature[98]. The fluorescent labelled peptide (PEVPPpTPPGpSHSAFTK(FICT)) and its unlabelled form (PEVPPpTPPGpSHSAFTK) were purchased on demand (ChinaPeptides). Initially the experiment was set up by optimizing buffer conditions, incubation times and the concentrations of the protein, FICT-DISC1 and the compounds. The quality of the optimized conditions was assessed by calculating Z-factor (Z'):

$$Z' = \frac{3 \times (\sigma_{free} + \sigma_{bound})}{mP_{bound} - mP_{free}} \textbf{ (Eq. 3.6)}$$

Where σ_{free} and σ_{bound} are the standard deviations of the unbound and bound peptide replicas respectively, and mP_{bound} and $mP_{unbound}$ the means of the polarization values of unbound and bound peptide respectively. All assays were carried out in duplicates at a final volume of 20 µL and the optimized buffer consisted of 50 mM Hepes (pH 8.0), 50 mM NaCl, 0.01% Tween-20, 1 mM DTT.

Two types of competitive assays were carried out to assess compound activity on substrate recognition. On one hand, the effect of the compounds on the K_d of FICT-DISC1 was followed (FP competitive assay 1). For that, FBW7-SKP1 was screened at 8 different concentrations (2-fold dilution series starting from 25 µM), while fixing the concentrations of FICT-DISC1 (50 nM) and compound (100 µM, 1% DMSO) or unlabelled DISC1 peptide (1 µM, 1% DMSO (control)). On the other hand, the effect of increasing compound concentrations was determined (FP competitive assay 2). 2-fold dilution series of compound starting from 800 µM (8% DMSO) were screened with fixed concentrations of FBW7-SKP1 (10 µM) and labelled-DISC1 (50 nM). For the positive control, unlabelled DISC1 peptide was tested at 2-fold dilution series of compound starting from 3 µM (8% DMSO). All prepared samples were centrifuged at 0.2 rpm for 30 seconds and incubated for 30 minutes at room temperature in the dark. Data was obtained in duplicates and evaluated with Mars software (BMG LabTech). Means of polarization values from replicas were plotted against concentrations to get the corresponding dose-response curves. For FP competitive assay 1, the "apparent" dissociation constant (K_d^{app}) of FICT-DISC1 was calculated, which allowed us to determine the K_d of the competitive entity tested (unlabelled peptide or compound, K_d^U) by using the following equation:

$$K_d^U = \frac{K_d^L}{K_d^{app} - K_d^L} \times [U]_{tot} \quad \textbf{(Eq. 3.7)}$$

Where $[U]_{tot}$ is the total concentration of the unlabelled entity, K_d^L is the dissociation constant of the labelled peptide and K_d^{app} the apparent dissociation constant of the labelled peptide in the competition assay. For FP competitive assay 2, EC_{50} values were calculated.

3.3. Materials and methods of Chapter 5: Pharmacological modulation of TET2 epigenetic enzyme

3.3.1. Molecular cloning

3.3.1.1. Protein construct and expression plasmids

TET2 construct 1 is a kind donation of Professor Xu (Fundan University, China) and it consist on TET2 (1129–2002 with residues 1471–1843

replaced by a 15-residue GS-linker) expressed in a pET-28a(+)psumo vector with kanamycin resistance and an N-terminal His6-Sumo tag, with expression in *E. coli*. TET2 construct 2 is a modification of TET2 construct 1 done by the collaborator Dr Isabel Puig (VHIO, Barcelona). It expresses TET2 residues 1129–1936 (with residues 1471–1843 replaced by a 15-residue GS-linker).

3.3.2. Transformation of *E. coli* cells

Competent *E. coli* cells were thawed on ice for 30 minutes. Afterwards, 50 µL of cells were mixed with 5 ng of vector-DNA and kept on ice for other 30 minutes. Heat shock was carried out for 1 minute at 42 ºC followed by 5 minutes incubation on ice. 500 µL of LB without antibiotic were added to transformed *E. coli* cells and the mixture was incubated at 37 ºC and 400 rpm for 1 hour. Then, 100 µL of bacterial cells were spread on LB-agar plates that contained 50 µg/L of kanamycin and incubated over-night at 37 ºC.

3.3.3. Protein expression: TET2 construct 1 and TET2 construct 2

A single colony of transformed cells was picked and cultured over-night at 37ºC and 150 rpm in 10 mL of LB medium supplied with 50 µg/L of kanamycin. The following day, the 10 mL of grown cells were transferred to a flask containing 1 L of LB media also supplied with 50 µg/L of kanamycin. The flask was incubated at 37 ºC and 180 rpm until reaching an OD_{600} between 0.6 and 0.8. Temperature was then decreased at 18 ºC and cells were induced with 0.4 mM IPTG. *E. coli* cell culture was incubated at 18 ºC and 180 rpm for 12 to 18 hours. Cells were harvested by centrifugation at 3700 rpm and 6 ºC during 30 minutes. Cell pellets were stored at -20 ºC before use.

3.3.4. Cell lysis and lysate clarification

Cell pellets were defrosted at room temperature followed by a resuspension with 15 mL of protein buffer E (Table 3.4) supplemented with Pierce protease inhibitor cocktail (ThermoFisher Scientific). Cell lysis was carried out by sonication during 2 minutes in intervals of 10 seconds of sonication followed by 30 seconds of break at 19 ºC. Lysates were then clarified by double centrifugation at 4 ºC and 8800 rpm during 30

minutes. Supernatant was collected and filtered using 0.8 µm syringe filters (Sigma-Aldrich) in order to be used for protein purification.

Table 3.4. Protein purification buffers for TET2 purifications

Name	Buffer	Protein purified
E	50 mM Hepes pH 8.0, 150 mM NaCl, 30 mM imidazole, 2 mM ß-mercaptoethanol	TET2
F	50 mM Hepes pH 8.0, 150 mM NaCl, 250 mM imidazole, 2 mM ß-mercaptoethanol	TET2
G	50 mM Hepes pH 8.0, 150 mM NaCl, 2 mM ß-mercaptoethanol	TET2
H	50 mM Hepes pH 8.0, 50 mM NaCl, 2 mM ß-mercaptoethanol	TET2
I	50 mM Hepes pH 8.0, 1 M NaCl, 2 mM ß-mercaptoethanol	TET2

3.3.5. Protein purification

3.3.5.1. TET2 for biophysical assays

To perform biophysical assays, TET2 construct 1 was employed for protein production, which consisted on a two-step procedure by using an ÄKTA start system (Cytiva). The protein obtained after lysate clarification was applied to a 5 mL HisTrap HP column (Cytiva), washed with buffer E (Table 3.4). Bound protein to the column was eluted with a linear gradient of buffer F (Table 3.4). Fractions containing protein from the elution were pooled and concentrated up to 4 mL to perform size exclusion chromatography with a HiPrep 16/60 Sephacryl S-100 HR (Cytiva) using buffer G (Table 3.4). The mass and purity of the protein were verified by SDS–PAGE and mass spectrometry.

3.3.5.2. TET2 for crystallization assays

For crystallization trials, TET2 construct 1 and TET2 construct 2 were employed for protein production following the same purification procedure and using an ÄKTA start system (Cytiva). The protein obtained after lysate clarification was applied to a 5 mL HisTrap HP column (Cytiva) to perform an immobilized metal ion affinity chromatography (IMAC),

washed with buffer E (Table 3.4). Bound protein to the column was eluted with a linear gradient of buffer F (Table 3.4). Fractions containing protein from the elution were pooled and concentrated. ULP1 protease was added to protein fraction in order to cleave the His-SUMO tag. The mixture was dialyzed overnight at 4 °C against buffer E (Table 3.4) using a 3.5 kDa dialysis membrane (Spectra/Por). Another IMAC was performed as described above and the flow through was collected and dialyzed against Buffer H (Table 3.4). Protein solution was loaded to a 5 mL anion exchange Hitrap Heparin column (Cytiva) and eluted with a linear gradient of buffer I (Table 3.4). The mass and purity of the protein were verified by SDS–PAGE and mass spectrometry.

3.3.6. Biophysical techniques

3.3.6.1. Differential scanning fluorimetry to determine binding to TET2

DSF experiments were performed using a Light Cycler 480 II (Roche Applied Science) at the genomics service (CCTiUB). TET2 screening was carried out with TET2 construct 1 protein at a final concentration of 2.5 µM. Experiment buffer consisted on 50 mM HEPES pH 8.0, 150 mM NaCl. Screening compounds were tested at a final concentration of 100 µM or 50 µM, while priority compounds were screened at 200 µM, 100 µM, 50 µM and 10 µM, all of them at 1% DMSO. SYPRO Orange dye (Sigma-Aldrich) was added as a fluorescence probe at a dilution of 1:1000 (5x). Excitation and emission filters for the SYPRO-Orange dye were set to 465 nm and 580 nm, respectively. The screening was carried out in a 96-well plate in a final volume of 25 µL, raising the temperature from 20 ºC to 85 ºC in 0.6 ºC per minute steps. Data analysis was performed using the Light Cycler 480 software. Control experiments were carried out with protein and dye mixture, and buffer and dye mixture. When necessary, controls with compound and dye mixture were also performed. All experiments were performed in triplicates. To consider compound binding, thermal shifts (ΔTm) of the protein-ligand complexes (ΔTm = Tm protein-ligand complex–Tm protein) had to be at least twice the standard deviation of the Tm replicas of the protein. The melting curves and thermal shifts were visualized using Microsoft Excel spreadsheet.

3.3.6.2. TET2 surface plasmon resonance screening for identification and characterization of binders

All SPR experiments were carried out at 25ºC using Biacore T200 SPR biosensor instrument (Cytiva) at the molecular interaction analysis service (CCTiUB). Prior to the start of each new experiment, the whole system was cleaned using Biacore maintenance kit.

3.3.6.2.1. Surface plasmon resonance for hit identification and characterization

For SPR experiments, TET2 protein from Section 3.3.5.1 was employed. A CM5 sensor chips (Cytiva) was inserted, preconditioned and normalized following the protocol proposed by the supplier (Cytiva). For the experiments two channels in the chip were used: one with no protein immobilized but treated to block the dextran (reference) and the other with immobilized TET2 protein. Immobilization was carried out using standard amine coupling procedure, which started with the activation of the carboxymethyl dextran matrix of the sensor chip with 0.1 M N-hydroxysuccinimide (NHS) and 0.4 M 1-ethyl-3-(3-(dimethylamino)propyl) carbodiimide hydrochloride (EDC) at a flow rate of 15 µL/min for 7 minutes. The protein immobilization was then performed at 5 µL/min, using a protein mixture diluted 1:100 with 10 mM sodium acetate at pH 4.0. To determine the amount of protein immobilized, formula from Figure 3.1 was applied to have an expected R_{max} of 40 RU. Once the protein was immobilized, 1 M ethanolamine hydrochloride was injected for 7 minutes at 15 µL/min to block activated groups of the dextran matrix. Phosphate buffered saline (PBS, 10 mM Phosphate, pH 7.4, 150 mM NaCl) was used as immobilization running buffer.

All the screenings were performed at a flow rate of 60 µL/min and the compound association and dissociation times set were 60 seconds and 180 seconds, respectively. In all screenings, running buffer consisted on 1x PBS and 5% DMSO and solvent correction was applied, which consisted on 8 dilutions that ranged from 3% to 8% DMSO in 1.05x PBS. Compound stock solutions were prepared at 20 mM by dissolving its powder in 100% DMSO. Initially, single concentration assays were performed. For that, several doses were tried to establish the optimal

ones for the screening, being: 10 µM, 100 µM, 50 µM or 25 µM at 5% DMSO. The hits obtained for this primary screening were then characterized to determine the K_d, by carrying out 2-fold dilution series starting from 250 µM in 5% DMSO. For that, a bank of dilutions was prepared in 100% of DMSO from which desired samples concentrations were obtained by diluting them with 1.05x PBS. Once there was an estimation of the K_d of the compounds, concentration range was adjusted if necessary.

The Biacore T200 evaluation software 2.0 was used for data analysis. Nonspecific binding to the chip surface and the baseline drift were corrected subtracting the signals of a reference surface (where the immobilization procedure was carried out without proteins) to the signals obtained on the surfaces with protein. Artefacts derived from DMSO interferences were corrected using series of solvent standards (solvent correction). Background signals were corrected subtracting blank injections (blank subtraction) to the injected compound signals. To estimate binding affinity, SPR data was fitted to a single interaction model, where steady state values were extracted from the sensograms recorded and plotted against the different concentrations assayed. When necessary, R_{max} was fixed considering the R_{max} expected regarding the amount of protein immobilized in the chip surface.

3.3.6.2.2. Surface plasmon resonance to study mechanism of action of TET2 binders

SPR experiments were performed to study how TET2 binders affected TET2 interaction with biotinylated DNA. For those experiments, TET2 protein of Section 3.3.5.1 and double-stranded DNA (Forward (5'-3'): Biotin-GTATGCCTCATGCCGGACTTAACTGCAGTG. Reverse (5'-3'): CACTG CAGTTTAGTCCGGCATGAGGCAAAC) were employed. Sensor Chip SA (Cytiva) was inserted, preconditioned and normalized following the protocol proposed by the supplier (Cytiva). For the experiments two channels in the chip were used: one empty and the other with immobilized DNA. Immobilization of the biotinylated DNA was carried out using the streptavidin matrix of the chip, which has high affinity for biotin (K_d: 10 $^{-15}$ M). A solution of 200 nM of DNA was directly immobilized at 5 µL/min. To determine the amount of DNA immobilized,

formula from Figure 3.1 was applied to have an expected R_{max} of 100 RU and considering as analyte TET2. HBSN (10 mM Hepes pH 8.0, 150 mM NaCl) buffer with 0.005% Tween-20 was used as immobilization running buffer. All the screenings were performed at a flow rate of 30 µL/min and the compound association and dissociation times set were 60 seconds and 90 seconds, respectively. In all screenings, running buffer consisted on 50 mM Hepes pH 8.0, 150 mM NaCl, 2 mM β-mercaptoethanol, 0.005% Tween20 and 5% DMSO. Solvent correction was applied, which consisted on 8 dilutions that ranged from 4% to 6% DMSO in the same running buffer without DMSO. Compound stock solutions were prepared at 20 mM by dissolving its powder in 100% DMSO. Protein stock was at 2.43 µM buffered in 50 mM Hepes pH 8.0, 150 mM NaCl and 2 mM β-mercaptoethanol. Different protein concentrations were tested with and without compounds in order to determine the effect of known binders in DNA recognition. TET2 was screened at different concentrations that ranged from 0.1 µM to 2 µM. Compounds were tested at a single concentration of 50 µM in the case of ONR-7B and 100 µM for 6H and 3D. Controls with compound mixtures without protein were performed. Blanks consisted on the buffer employed to dilute samples adding 5% DMSO.

The Biacore T200 evaluation software 2.0 was used for data analysis. Nonspecific binding to the chip surface and the baseline drift were corrected subtracting the signals of a reference surface (where the immobilization procedure was carried out without DNA) to the signals obtained on the surface with DNA. Artefacts derived from DMSO interferences were corrected using series of solvent standards (solvent correction). Background signals were corrected subtracting blank injections (blank subtraction) to the injected compound signals. To estimate binding affinity, SPR data was fitted to a single interaction model, where steady state values were extracted from the sensograms recorded and plotted against the different concentrations assayed.

3.3.6.3. X-ray crystallography to characterize TET2 – ligand complexes

3.3.6.3.1. Crystallization strategy for characterization of TET2 binders

X-ray crystallography is a technique used for determining the atomic and molecular structure of a protein. In order to apply to this methodology, the protein target must be first be crystallized[67]. Crystallization experiments were performed at the laboratory of Dr Miquel Coll (IRB-CSIC, Barcelona).

TET2 structure has been previously elucidated, but only in complex with a double stranded DNA and N-oxalylglycine (NOG, an α-ketoglutarate (α-KG) analogue)[99,100]. For that reason, three types of complex were screened for both constructs: TET2 in complex with DNA and an α-KG analogue, TET2 in complex with DNA, an α-KG analogue and one of the best binders assessed and TET2 in complex with one of the best binders assessed.

Two different constructs were tested for crystallization experiments: TET2 construct 1 and TET2 construct 2. Prior to crystallization TET2 construct 1 and TET2 construct 2 where buffered to 50 mM Hepes pH 8.0, 150 mM NaCl and 2 mM ß-mercaptoethanol. Trials with TET2 construct 2 buffered in 10 mM Hepes pH 7.0, 100 mM NaCl and 10 mM ß-mercaptoethanol were also performed. TET2 construct 1 protein was tried at concentrations that ranged from 450 μM to 300 μM. Its uncleaved form (His-SUMO TET2 construct 1) was also tested at concentrations around 450 μM. A wide range of concentrations were tested for TET2 construct 2 crystallization trials, from 170 μM to 1 mM. A palindromic 12-bp double stranded DNA (dsDNA) was used for crystallization trials with the following sequence: 5′-ACCACXGGTGGT-3′, where X is 5-hydroxymethyldeoxycytosine. Dry DNA was purchased in biomers.net and was resuspended in sterile MQW to have a final concentration of 2 mM after annealing. The annealing was performed by heating the DNA solution at 80 ºC during 30 minutes and a gradual cool over-night. DNA was tested in 1:1 ratio with protein. Two different α-KG analogue were tested in complex with TET2 and DNA: NOG and 2-HG (L-α-hydroxyglutaric acid disodium salt). Their stock solutions were prepared at 250 mM by dissolving its powder in 100% DMSO. In all experiments they were tested in a concentration of 2 mM. Best binders

were tested at final concentrations ranging from 0.2 mM to 2 mM. Their stock solutions were prepared at 40 mM by dissolving its powder in 100% DMSO. DMSO final concentration was of 0.8% when no binder was in the complex (from α-KG analogue), and ranged from 4.4% to 10% when the complex included one of the best binders.

Crystallization assays were performed using vapour diffusion. For the screening conditions and to perform a seeding screening, sitting drop method was used, whereas for the optimization plates hanging drop method was performed. All crystal growth was carried out at 4 ºC, except for the screenings with His-Sumo-TET2, in which 20 ºC was also tried since the protein should be more stable. Crystallization and seeding screens were performed at the Automated Crystallography Platform (PAC, from IBMB-IRB, Barcelona) using Phoenix robot (Art Robbins). For that, 2 Lens Crystallization Microplates (Swissci) were used, which consist on 96-well plates with two concave sub wells per well, allowing to have two droplets per condition. Generally, one droplet contained the protein complex and the other was used as a control with the same conditions of the complex but without the protein. 1:1 drop ratio between sample solution and reservoir were performed in all crystallization screenings in a final volume of 200 nL and 300 nL for TET2 construct 1 and TET2 construct 2 crystallization tests, respectively.

Different screening conditions available in the Automated Crystallography Platform (PAC, from IBMB-IRB) were used for crystallization referred as: PAC1, PAC2, PAC9, PAC10, PAC21 and BCS, being PAC10 specific for protein-DNA complexes. To crystallize complexes with TET2 construct 1 PAC1, PAC9 and PAC21 screens were performed with the following complexes: TET2, DNA, 2HG and binder; TET2 and binder; His-Sumo-TET2, DNA, 2HG and binder; His-Sumo-TET2 and binder. TET2, DNA and 2HG concentrations were similar in all screens and binder concentrations and DMSO concentrations varied. For complexes with TET2 construct 2 PAC1, PAC2, PAC9, PAC10, PAC21 and BCS were performed with the following complexes: TET2, DNA, 2HG and binder; TET2, DNA, NOG and binder; TET2, DNA and 2HG; TET2 and binder. In these experiments, different TET2 concentrations were screened (maintaining 1:1 DNA ratio and 2mM of 2HG or NOG), as well as different binder concentrations.

Crystal seeding consists on adding crystal seeds to a crystallizing solution in order to optimize the crystallization process[101]. For seeding screening, a 96-well optimization plate was prepared in the Automated Crystallography Platform (PAC, from IBMB-IRB) facilities, combining an optimization screening of PAC10-C11 and PAC10-E12 conditions. Seeds were prepared by mixing 30 µL of reservoir solution with 2 µL of crystals in a centrifuge tube containing a Seed Bead (Hampton Research), followed by vortexing the sample solution. The plate was prepared with Phoenix robot (Art Robbins) and drops contained 200 nL of sample solution, 40 nL of seeds and 160 nL of reservoir. Two droplets were prepared for each condition, one that contained the complex formed by TET2, DNA and 2HG and the other with TET2, DNA, 2HG and binder complex.

In order to optimize hit conditions from crystal screenings, optimization screens were prepared in 24-well plates (Crystalgen), being the reservoir final volume of 1 mL. Optimization plates with the conditions already published[99,100] were tried for TET2 construct 1 and TET2 construct 2. For TET2 construct 1, two optimization plates were prepared considering mixing conditions from PAC1-F11 and PAC9-D9. 1:1 drop ratio between sample solution and reservoir were performed with a final volume of 2µL per drop. Sample mixture contained TET2 construct 1, DNA 2HG and a different binder per plate. In the case of TET2 construct 2, optimization plates containing TET2, DNA and 2HG complex were prepared from PAC10-E12 conditions. These plates were further optimized varying crystallization conditions, protein concentration and complex-reservoir drop ratios, being 1:1 or 2:1 with final volumes of 2 µL and 3 µL, respectively. PAC10-G3 condition was also optimized in 2 µL drops (1:1 ratio) for TET2, DNA and 2HG complex and TET2, DNA, 2HG and binder complex. PAC10-C11 condition was optimized in 2 µL drops (1:1 ratio) for TET2, DNA, 2HG and binder complex. Finally, PAC1-H3 condition was also optimized in 2 µL drops (1:1 ratio) for TET2, DNA, NOG and binder.

3.3.6.3.2. TET2 – DNA Complex formation for crystallization assays

In order to confirm complex formation and discard potential excess of DNA, size exclusion chromatography was performed. For that, a Superdex 75 increase 10/300 GL (Cytiva) was used. First, in order to

establish the volumes of elution, a small amount of DNA sample and protein sample were injected separately to the column. Then, complex was formed mixing 20 mg/mL TET2 in 1:1 ratio with DNA including NOG, as for crystallization assays. Complex was then loaded to the column and the fractions related to the complex were used for crystallization.

3.4. Materials and methods of Chapter 6: Development of new chemical entities that bind to the bromodomain BRD4 BD1

3.4.1. Computational tools

3.4.1.1. Study of solvation preferences of bromodomain BRD4 BD1

The most conserved water molecules in BRD4 BD1 were established by carrying out a crystal structural analysis of all of the structures in the PDB. All BRD4 BD1 structures found in PDB were overlaid with the software MOE[91] and the positions of water molecules in the active site of the protein were compared. To achieve this, first the waters of the active site were established as the ones of reference. These waters were the ones found 6 Å far from the reference inhibitor XD14[102]. Subsequently, the coordinates of these hydration sites were compared with all PDB files that contained BRD4 BD1 in holo- conformation, in order to find out which structure was found lying within 1.0 Å in these positions. From this data, we obtained the fractional conservation (F) and the displacement fraction (D):

$$F = \frac{N_{Hydrated}}{N_{Total}} \textbf{ (Eq. 3.8)} \quad D = 1 - F \textbf{ (Eq. 3.9)}$$

$N_{Hydrated}$ corresponded to the total number of PDB in which a molecule of water has been found in the place of the waters of reference, and N_{Total} was the total number of PDB.

3.4.1.2. Virtual screening to determine BRD4 BD1 potential binders that displace water molecules

3.4.1.2.1. Tethered docking with rDock to select acetamides

In order to enforce acetamide binding mode, tethered docking was performed in a collection of approximately 6 million purchasable compounds. The *sdtether* script (part of rDock[90]) was employed using

the following inputs: acetamide moiety as a reference bound ligand structure and the SMARTS code of a primary acetamide (O=C-[NH2]).

3.4.1.2.2. Docking-based virtual screening with pharmacophoric restraints for tethered docking hits

Hits from tethered docking were subsequently screened by docking-based virtual screening with pharmacophoric restraints. For that, rDock[90] was also used. Protein cavity was defined by applying *rbcavity* using the "reference ligand method". For that, a mol2 file from PDB file 4O77[103] was used, which contained holo BRD4 BD1. Prior to that, ligand and waters were removed, a part from the three waters of the cavity that should not be displaced by acetamide. The cavity was formed considering the space at 6 Å radius from the reference ligand from PDB file 4O77[103]. The pharmacophoric restraint applied consisted on an hydrogen bond acceptor 2 Å from ASN140, considered a key interaction of BRD4 BD1 described ligands[104]. Pharmacophore tolerance was defined to 1 Å radius around the pharmacophoric point. If the compound conformation did not have features accomplishing the positional constraints, rDock would assign a positive restraint score (unfavourable).

The high-throughput virtual screening (HTVS) protocol consisted in three different stages in order to crease the efficiency of the simulation. In the first stage, the compounds were filtered depending on the "SCORE.INTER" (higher than -5 kJ/mol were discarded) and the "SCORE.RESTR.PHARMA" (higher than 2 kJ/mol were discarded). In the second round, the "SCORE.INTER" and the "SCORE.RESTR.PHARMA" should be less than -10 kJ/mol and 1 kJ/mol, respectively, in order to complete the 20 docking runs (step three). Finally compounds with "SCORE.INTER" and "SCORE.RESTR.PHARMA" less than -17 kJ/mol and 0.4 kJ/mol, respectively, were selected.

3.4.1.2.3. Compound clustering of docking hits

MACCS fingerprints[94] were used to cluster docking hits that had 95% selectivity. For that, MOE[91] software was employed.

3.4.1.2.4. Applying DUck in docking hits that displace water molecules of BRD4 BD1

After clustering, selected compounds were filtered with DUck[58] using the hydrogen bond produced by an acceptor group of the compound and BRD4 BD1 ASN140. For DUck calculations only, a subset of the protein (referred as "chunk") is employed. Protein chunk was prepared manually by selecting the residues within 6 Å from N140. Besides, when selecting the residues of the chunk we considered the following guidelines: (1) Selecting the minimum number of residues necessary in order to reduce computational time; (2) Not selecting residues that would block the ligand from exiting the pocket; (3) Not removing residues if this would lead to the possibility of solvent entering the pocket from areas other than where the ligand is exiting; (4) Preserving the local environment of the interacting atoms. The sequence gaps produced when selecting the chunk were capped with acetyl and N-methyl groups in order to prevent charged ends and unnatural electrostatic forces. Finally, possible clashes produced when capping were checked. The selected chunk of BRD4 BD1 included the following residues: TRP81, PRO82, PHE83, GLN84, GLN85, PRO86, VAL87, ASP88, ALA89, LYS91, LEU92, ASN93, LEU94, TYR97, ILE101, PRO104, MET105, THR131, ASN135, CYS136, TYR137, TYR139, ASN140, ASP144, ASP145, ILE146, MET149. Ligands were parameterized using a MOE[91] svl script. The steered-molecular dynamics (SMD) consisted in four different steps: I) minimization, III) equilibration, III) SMD simulations at two different temperatures (300K and 325K), in which the hydrogen bond of the binder is pulled to go from 2.5 Å to 5.0 Å, IV) Calculation of the W_{QB} value (work necessary to break the hydrogen bond) and application of a threshold (W_{QB} = 6 kcal/mol) to repeat steps III and IV during 6 cycles (6 replicas). If the threshold value was not achieved in one of those cycles, SMD was discontinued. The lowest W_{QB} value of the 6 replicas was the one considered. SMD steps were performed with GPU-based pmemd.cuda in AMBER.

3.4.1.3. Docking-based virtual screening including conserved waters in the cavity

The resulting molecules from the virtual screening that were purchased to test biophysically were redocked using rDock[90]. The main difference was that in the BRD4 BD1 structure the network of preserved water

molecules was considered. A part from that, the structure preparation and the pharmacophoric restraints were the same as in Section 3.4.1.2.2. The virtual screening protocol consisted on performing 50 docking runs per compound.

3.4.2. Molecular cloning

3.4.2.1. Protein construct and expression plasmids

BRD4 BD1 construct is a kind donation of Professor Alessio Ciulli (University of Dundee, UK) and it encodes residues 44-168 of BRD4 BD1 protein with a His6-tag, TEV protease cleavage site and kanamycin resistance.

3.4.3. Transformation of *E. coli* cells

Competent *E. coli* cells were thawed on ice for 30 minutes. Afterwards, 50 µL of cells were mixed with 5 ng of vector-DNA and kept on ice for other 30 minutes. Heat shock was carried out for 1 minute at 42 ºC followed by 5 minutes incubation on ice. 500 µL of LB without antibiotic were added to transformed *E. coli* cells and the mixture was incubated at 37 ºC and 400 rpm for 1 hour. Then, 100 µL of bacterial cells were spread on LB-agar plates that contained 50 µg/L of kanamycin and incubated over-night at 37 ºC.

3.4.4. Protein expression: BRD4 BD1

A single colony of transformed cells was picked and cultured over-night at 37ºC and 150 rpm in 10 mL of LB medium supplied with 50 µg/L of kanamycin. The following day, the 10 mL of grown cells were transferred to a flask containing 1 L of LB media also supplied with 50 µg/L of kanamycin. The flask was incubated at 37 ºC and 180 rpm until reaching an OD_{600} of 2.5. Temperature was then decreased at 18 ºC and cells were induced with 0.4 mM IPTG. *E. coli* cell culture was incubated at 18 ºC and 180 rpm for 12 to 18 hours. Cells were harvested by centrifugation at 3700 rpm and 6 ºC during 30 minutes. Cell pellets were stored at -20 ºC before use.

3.4.5. Cell lysis and lysate clarification

When purified for biophysical experiments, BRD4 BD1 cell pellets were defrosted at room temperature followed by a resuspension with 15 mL of protein Buffer J (Table 3.5) supplemented with Pierce protease inhibitor cocktail (ThermoFisher Scientific). Cell lysis was carried out by sonication during 2 minutes in intervals of 10 seconds of sonication followed by 30 seconds of break at 19 ºC. Lysates were then clarified by double centrifugation at 4 ºC and 8800 rpm during 30 minutes. Supernatant was collected and filtered using 0.8 μm syringe filters (Sigma-Aldrich) in order to be used for protein purification.

Cell lysis and lysate clarification when purifying BRD4 BD1 for crystallization assays followed a different procedure. Defrosted pellets were resuspended with buffer L (Table 3.5) complemented with protease inhibitors. Cells were lysed using the French Press at 25000 psi. 0.15% (v/v) polyethylenimine was then added to the lysate to help DNA precipitate and the mixture was incubated on ice for 30 min. Cell debris was centrifuged at 21000 rpm and 4ºC for 45 min. Supernatant was collected for protein purification.

Table 3.5. Protein purification buffers for BRD4 BD1 purifications

Name	Buffer	Protein purified
J	50 mM Hepes pH 7.4, 150 mM NaCl, 30 mM imidazole, 2 mM ß-mercaptoethanol	BRD4 BD1
K	50 mM Hepes pH 7.4, 150 mM NaCl, 250 mM imidazole, 2 mM ß-mercaptoethanol	BRD4 BD1
L	50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol, 1mM TCEP	BRD4 BD1
M	50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol, 10 mM imidazole	BRD4 BD1
N	50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol, 30 mM imidazole	BRD4 BD1
O	50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol, 50 mM imidazole	BRD4 BD1
P	50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol, 100 mM Imidazole	BRD4 BD1

Table 3.5 (cont.). Protein purification buffers for BRD4 BD1 purifications

Name	Buffer	Protein purified
Q	50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol, 150 mM imidazole	BRD4 BD1
R	50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol, 250 mM imidazole	BRD4 BD1
S	50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol, 500 mM imidazole	BRD4 BD1
T	10 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol	BRD4 BD1

3.4.6. Protein purification

3.4.6.1. BRD4 BD1 for biophysical assays

To perform biophysical assays, BRD4 BD1 protein was purified using an ÄKTA start system (Cytiva). The protein obtained after lysate clarification was applied to a 5 mL HisTrap HP column (Cytiva), washed with buffer J (Table 3.5). Bound protein to the column was eluted with a linear gradient of buffer K (Table 3.5). Fractions containing protein from the elution were pooled and concentrated. TEV protease was added to protein fraction in order to cleave the His-tag. The mixture was dialyzed overnight at 4 °C against buffer J (Table 3.5) using a 3.5 kDa dialysis membrane (Spectra/Por). Another IMAC was performed as described above and the flow through with the protein was collected. The mass and purity of the protein were verified by SDS–PAGE and mass spectrometry.

3.4.6.2. BRD4 BD1 for crystallization assays

Purification of BRD4 BD1 for crystallization assays was performed at Structural Genomics Consortium (SGC, Oxford) in the laboratories of Professor Frank von Delft and Professor Panagis Filippakopoulos. The protein obtained after lysate clarification was applied to a 7 mL cobalt resin, washed with buffer M and N (Table 3.5), subsequently. Bound protein was eluted with buffers O, P, Q, R and S (Table 3.5) and each elution was separated to run SDS-PAGE. Purer samples were pooled and

TEV protease was added to cleave the His-tag over-night. ÄKTA prime system (Cytiva) was used to perform size exclusion chromatography with a Superdex75 16/60 Prep Grade (Cytiva) using buffer T. The mass and purity of the protein were verified by SDS–PAGE and mass spectrometry.

3.4.7. Biophysical techniques

3.4.7.1. Differential scanning fluorimetry to determine binding to BRD4 BD1

DSF experiments were performed using a Light Cycler 480 II (Roche Applied Science) at the Genomics Service (CCTiUB). BRD4 BD1 DSF screening was carried out with BRD4 BD1 at a final concentration of 2.5 µM. Experiment buffer consisted on 50 mM HEPES pH 7.5, 150 mM NaCl. Compounds were tested at different final concentrations, depending on the project they were involved. In general, final concentrations of 200 µM, 100 µM, 50 µM and 10 µM were screened, all in 1% DMSO. In each experiment the already described binder (+)-JQ1[105] was tested at 10 µM (1% DMSO) as a positive control. Besides (+)-JQ1 was also tested at 200 µM (1% DMSO). SYPRO Orange dye (Sigma-Aldrich) was added as a fluorescence probe at a dilution of 1:1000 (5x). Excitation and emission filters for the SYPRO-Orange dye were set to 465 nm and 580 nm, respectively. The screening was carried out in a 96-well plate in a final volume of 25 µL, raising the temperature from 20 ºC to 85 ºC in 0.6 ºC per minute steps.

Data analysis was performed using the Light Cycler 480 software. Control experiments were carried out with protein and dye mixture, and buffer and dye mixture. When necessary, controls with compound and dye mixture were also performed. All experiments were performed in triplicates. To consider compound binding, thermal shifts (ΔTm) of the protein-ligand complexes ($\Delta Tm = Tm$ protein-ligand complex $- Tm$ protein) had to be at least twice the standard deviation of the Tm replicas of the protein. The melting curves and thermal shifts were visualized using Microsoft Excel spreadsheet.

3.4.7.2. Isothermal titration calorimetry to characterize potential binders of BRD4 BD1

ITC experiments were performed using VP-ITC microcalorimeter (MicroCal™). All titrations were carried out at 25 ºC. DMSO concentrations were the same in the sample cell with the protein solution and in the syringe with the compound, which were filled at final volumes of 1.2 mL and 250 µL, respectively. Reference cell was filled with 1.2 mL of MQW. BRD4 BD1 was buffer exchanged in 50 mM HEPES pH 7.5, 150 mM NaCl. An initial injection of 2 µL was performed followed by 23 identical injections of 12 µL (in the case of (+)-JQ1, fragment of reference (SPF-REF), SPF5 and SFP22), 29 identical injections of 8 µL (in the case of SPF14 and SSR4) or 34 identical injections of 8 µL (in the case of SPF23). 50 µM BRD4 BD1 solutions were titrated against SPF14 (500 µM, 1% DMSO). 40 µM BRD4 BD1 solutions were titrated against SPF22 (400 µM, 5% DMSO) and SSR4 (400 µM, 1% DMSO). 30 µM BRD4 BD1 solutions were titrated against fragment of reference (SPF-REF, 300 µM, 1% DMSO), SPF5 (300 µM, 3% DMSO) and SPF23 (1 mM, 3% DMSO). 20 µM BRD4 BD1 solutions were titrated against (+)-JQ1 (200 µM, 1% DMSO).

Data was analysed using the MicroCal ORIGIN software to directly obtain the enthalpy of binding (ΔH) and the binding constants (K_a and K_d). From that, thermodynamic parameters were calculated using the already mentioned equation 3.5:

$$\Delta G = \Delta H - T \times \Delta S = R \times T \times \ln K_d$$

Where ΔG, ΔH and ΔS are the changes in free energy, enthalpy and entropy of binding respectively, R is the gas constant (1.987 cal x K^{-1} x mol^{-1}) and T the temperature (in Kelvin). In all cases single binding site model was applied and the first data point was excluded from the analysis.

3.4.7.3. X-Ray crystallography to characterize BRD4 BD1 – ligand complexes

X-ray crystallography is a technique used for determining the atomic and molecular structure of a protein. In order to apply to this methodology, the protein target must be first be crystallized[67]. Two different techniques to obtain crystals of the complex have been used: soaking

and co-crystallization. Soaking consists on the introduction of the ligands into pre-existing apo crystals, whereas co-crystallization involves adding the ligand to the protein to form a complex that is subsequently crystallized[87].

3.4.7.3.1. Crystallization strategy for characterization of BRD4 BD1 binders

Crystallization experiments were performed at Structural Genomics Consortium (SGC, Oxford) in the laboratories of Professor Frank von Delft and Professor Panagis Filippakopoulos.

Prior to crystallization experiments, protein (from Section 3.4.6.2) was buffer exchanged in 10 mM Hepes pH 7.5, 150 mM NaCl, 2% glycerol using a Superdex75 16/60 Prep Grade (Cytiva) in an ÄKTA prime system (Cytiva). All crystallization assays were performed using the sitting drop vapour diffusion method at 4 °C. For that, 3 Lens Crystallization Microplates (Swissci) were used, which consist on 96-well plates with three concave sub wells per well, allowing to have three droplets per condition. Final volume for reservoir sub well was of 20 µL and for drops was of 300 nL, performing three different drop ratios of reservoir and protein mixture (200 nL + 100 nL, 150 nL + 150 nL, 100 nL + 200 nL). Plates were prepared using a mosquito® crystallization robot (TTP Labtech).

BRD4 BD1 was crystallized in an apo state for soaking experiments or in solution with a compound for co-crystallization experiments. In all cases, it was tested at final concentrations that ranged from 333 µM to 450 µM. Compound stock solutions were prepared at 100 mM by dissolving its powder in 100% ethylene glycol for soaking experiments, while prepared at 100 mM in 100% DMSO for soaking and co-crystallization experiments. Different ranges of concentration and percentages of DMSO were assayed in co-crystallization trials. RVX-208[106] and (+)-JQ1[105] were tested as control compounds in soaking and in co-crystallization experiments, being their stock concentration 50 mM 100% DMSO.

Different screening conditions already available in the protein crystallography facilities at SGC were used for crystallization referred as: LFS6, BCS, HIN3, JCSG7 and MIDAS. LFS6 stands for Ligand Friendly Screen 6 and is an in-house screen, which samples different combinations of PEG, salts and at near-physiological pH to obtain crystals

of protein-compound complexes. Basic ChemSpace (BCS) is another in-house screen that consolidates a wide diversity of PEG precipitants into a single screen[107]. JCSG7 and Modern Intelligent Dynamic Alternative Screen (MIDAS) can be purchased from Molecular Dimensions, whereas Index (HIN) can be purchased from Hampton Research. Three optimization plates from LFS6 screen were also prepared with Phoenix robot (Art Robbins Instruments). LFS6-G1 and LFS6-H7 were designed from conditions G1 and H7 of the screen, respectively. The other optimization screen consisted on varying the reservoir solutions per row in order to carry out a co-crystallization screen with 8 compounds per plate. The selected conditions were the ones of LFS6 screen that had given co-crystals of the complex formed by BRD4 BD1 and (+)-JQ1.

CHAPTER 4: RESULTS

DISENTANGLING E3 LIGASES LIGANDABILITY: APPLICATION TO FBW7

RESULTS: DISENTANGLING E3 LIGASES LIGANDABILITY: APPLICATION TO FBW7

4.1. Background

4.1.1. The ubiquitin proteasome system

During normal cellular homeostasis, proteins are constantly synthesized and destroyed. The most common degradation pathway for proteins is the ubiquitin proteasome system (UPS) pathway, a highly regulated signalling cascade that is ultimately responsible for the controlled degradation of a large number of proteins[108]. Protein ubiquitination is a post-translational modification, which involves the covalent attachment of ubiquitin proteins (8.6 kDa) to a lysine residue. Polyubiquitinated proteins can be recognized by the proteasome and destroyed[109]. This procedure occurs through a cascade of enzymatic reactions. Initially, ubiquitin is activated by an E1 enzyme, which consumes an ATP to produce a thioester linkage with the catalytic cysteine of ubiquitin. The ubiquitin is then transferred to an E2 conjugating enzyme. Finally, E3 ubiquitin ligases are in charge of substrate recognition and mediate the mechanism to transfer the ubiquitin to the substrate protein to be degraded. Repeated iterations of this ubiquitination process, result in long chains of ubiquitin on a given substrate and posterior degradation by the proteasome[110,111]. This procedure is illustrated in Figure 4.1.

The therapeutic potential of intervention in the UPS has been demonstrated by proteasome inhibitors such as peptide boronic acids (bortezomib®) and epoxyketones (carfilzomib®) that have had success in the clinic for the treatment of multiple myeloma amongst other diseases[108]. Despite these early successes, proteasome inhibitors have several limitations including offering no selectivity for the large number of proteins targeted by the UPS. Other members of the UPS have been targeted for drug discovery purposes, such as E1 ubiquitin activating enzymes. UBA1 and UBA6 are the two E1 enzymes identified in humans. Even though many E1 inhibitors have been developed, little success has been achieved due to lack of specificity and poor drug-like properties. In contrast to proteasome inhibitors, E1 do not induce the accumulation of ubiquitinated substrates[112]. E2 enzymes have also been potently

modulated by small-molecules. Around 40 E2 ligases have been described in humans. Targeting E2 ligases corresponds to an increase on selectivity compared to E1 enzymes[112,113]. Key to protein degradation by the UPS is the recruitment of the substrate protein by the E3 ubiquitin ligase. E3 ubiquitin ligases (of which >600 are known in humans) confer substrate specificity for ubiquitination and would represent more attractive targets for therapeutic intervention than current proteasome inhibitors. Consequently, this unconventional enzyme class has resulted to be more appealing for drug discovery efforts[114].
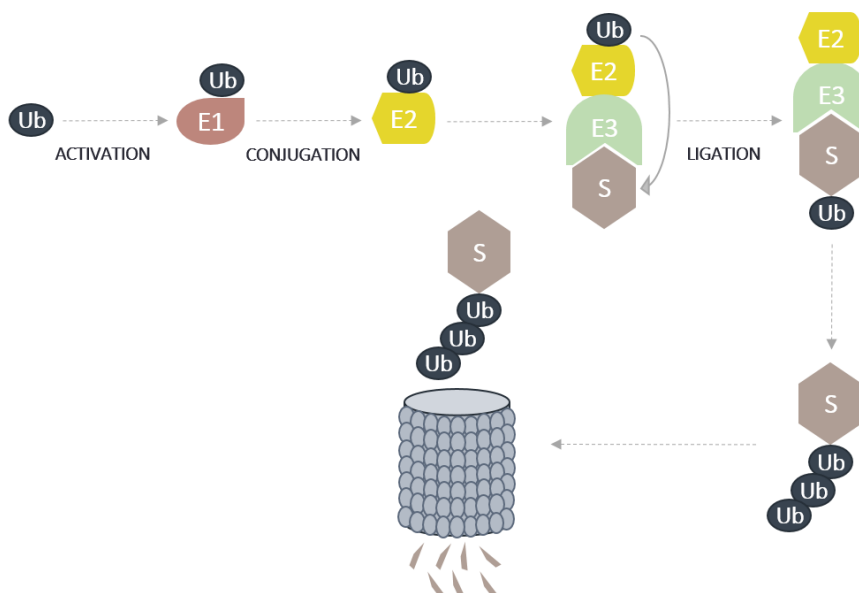
Figure 4.1. Natural pathway of the ubiquitin proteasome system (UPS) involving the enzymes E1, E2 and E3 to lead to substrate degradation by the proteasome. Abbreviations: S, substrate; Ub, ubiquitin.

4.1.2. E3 ligases: relevance, state-of-the-art and therapeutic modulation

Based on the mechanism of ubiquitin transfer to the substrate protein and the presence of unique domains, E3 ligases can be classified in three different types: really interesting new gene (RING), homologous to E6AP C-terminus (HECT), and ring-between-ring (RBR). Being the largest family of ubiquitin ligases, RING E3 ligases bind to a substrate protein while catalysing the direct transfer of ubiquitin from E2. HECT and RBR E3 ligases transfer ubiquitin to the substrate through a two-step reaction, forming an E3-ubiquitin intermediate mediated by a thioester bond[111].

E3 ligases specificity makes this protein family extremely appealing for drug discovery therapies, not only for decreasing side effects, but also for being potential targets in personalized medicine[114]. Besides, this protein family has also gained popularity in recent years for their implication in PROteolysis TArgeting Chimeras (PROTAC) molecules (Section 1.3, Box 1.1). Despite the tremendous significance of this protein family, only for few E3 ligases binders have been reported and only small molecules targeting the E3 ligase CRBN are currently approved. This limited success is partly caused for the need of modulating protein-protein interactions (PPI). The main challenges for small molecules targeting PPI are interacting with large contact surfaces (1500-2000 Å^2), which tend to be almost flat, lacking suitable pockets for small molecules to bind[115]. Moreover, targeting E3 ligases in the substrate binding site (degron site), often requires competing with strong binding post-translational modifications. Consequently, difficulties to develop drug-like compounds arise, when it is needed to disrupt highly polar interactions done by the substrate (like phosphorylations).

Several strategies have been developed to successfully find small molecules targeting E3 ligases. For instance, the development of nanomolar ligands for VHL E3 ligase was triggered by a peptidomimetic approach[116]. Fragment screening with X-ray crystallography as primary screen allowed the development of nanomolar inhibitors of KEAP1[117]. Structure-based computational tools have also been employed for the identification of small-molecules binding to SKP2[118] and SPOP[119]. SKP2 inhibitors were designed to be allosteric inhibitors that prevented SKP2 complex formation to cause substrate ubiquitination[118].In contrast, SPOP inhibitors were designed to bind in the substrate binding site. Despite no structure of SPOP with its inhibitors could be disentangled, binding site was elucidated through mutations[119]. The development of covalent ligands has been a tendency that has allowed targeting different E3 ligases such as RNF114[120] and RNF4[121]. In the case of RNF114 covalent-ligand screening was applied to find small-molecules that mimic the covalent interaction of the described anti-cancer natural product nimbolide[120]. RNF4 covalent binders were discovered using chemoproteomics-enabled covalent ligand screening platform. The

resulting hits were employed for PROTAC construction, successfully degrading the protein of interest[121]. On the whole, the development of novel E3 ligases recruiters would help to disentangle the remaining unknowns of PROTAC technology while allowing a more selective targeted protein degradation (e.g., tissue or organ specific). Besides, it would undoubtedly imply a benefit in therapeutics[122,123].

4.2. Objectives of Chapter 4

Despite the tremendous significance of E3 ligases, only for few members of this large family of proteins have developed binders. In this scenario, the first objective of Chapter 4 is to study E3 ligases ligandability, using a computational platform that involves MDMix technology[53] and allows the identification of regions (both allosteric and orthoesteric) in the selected E3 ligases that are more prompt to be targeted. Furthermore, we aim to develop and apply a computational workflow that enables the straightforward discovery of novel and therapeutic relevant small molecules able to bind to E3 ubiquitin ligases. The workflow consists on performing a high throuput virtual screening, followed by the implementation of biophysical techniques to characterize binding of potential hits (binders). Besides, binding site and mechanism of action of positive binders will be disentangled in the present chapter.

4.3. Study of E3 ligases ligandability

4.3.1. Selection of representative subset of E3 ligases

A comprehensive database of all available E3 ligases with X-ray structure with their structural and functional data was built. Ubihub[124] was used to obtain the full list of E3 ligases and to classify them based on their involvement in protein degradation (UPS score). The E3 ligases with available crystal structure were downloaded from the PDB database and relevant data about the structure was annotated (crystal quality, coverage, ligands, and etcetera). More precise data such as domains and their function was obtained from Uniprot[125] and PROSITE[126]. With all this information a representative subset of 23 E3 ligases was selected based on protein subfamilies and structural motifs (Figure 4.2). Selected E3 ligases are listed in Table 4.1. In order to make a retrospective validation,

some of the E3 ligases selected had already a ligand described. Compilation of E3 ligases dataset was performed together with Álvaro Serrano, a PhD student in the group. We reviewed the resulting database to make the E3 ligase selection.

Table 4.1. List of selected E3 ligases for MDMix studies

E3 ligase	MDMix structural motif	PDB code	Covered sequence	E3 Ligase subfamily	Ubihub class[124]
ASB9	ASB + SOCS-BOX	3ZKJ	88%	BC-BOX (SOCS-BOX)	Complex
BTRC	WDR + F-BOX	6M90	66%	F-BOX	Complex
CDC20	WDR	4GGC	63%	APC	Complex
CRBN		5FQD	85%	DCAF	Complex
COP1	WDR	5HQG	44%	Miscellaneous	Simple
DCAF1	WDR	4CC9	20%	DCAF	Complex
DCAF15	WDR	6UD7	71%	DCAF	Complex
FBO44	FBA + F-BOX	3WSO	99%	F-BOX	Complex
FBW7	WDR + F-BOX	2OVP	62%	F-BOX	Complex
GID4	GID/CTLH	6CCU	55%	Miscellaneous	Simple
KCTD5	BTB Domain	3DRX	76%	BTB	Complex
KEAP1	KELCH	6TYM	46%	BTB	Complex
KLHDC2	KELCH	6DO3	78%	BC-BOX (VHL-BOX)	Complex
PRKN	RING1-IBR-RING2	5C1Z	83%	RBR	Simple
RNF4	Zinc finger RING-type	4PPE	35%	Miscellaneous	Simple
RNF43	Extracellular domain	4KNG	19%	Miscellaneous	Simple
SKP2	LR + F-BOX	2AST	77%	F-BOX	Complex
SOCS2	SH2 + SOCS-BOX	6I5N	81%	BC-BOX (SOCS-BOX)	Complex
SPOP	MATH + BTB	3HQI	78%	BTB	Complex
SPSB1	SPRY	2JK9	74%	BC-BOX (SOCS-BOX)	Complex
TRIM21	SPRY	2IWG	38%	TRIM	Simple
TRIM28	PHD-BRD	2RO1	23%	TRIM	Simple
	RBCC	6QAJ	33%	TRIM	Simple
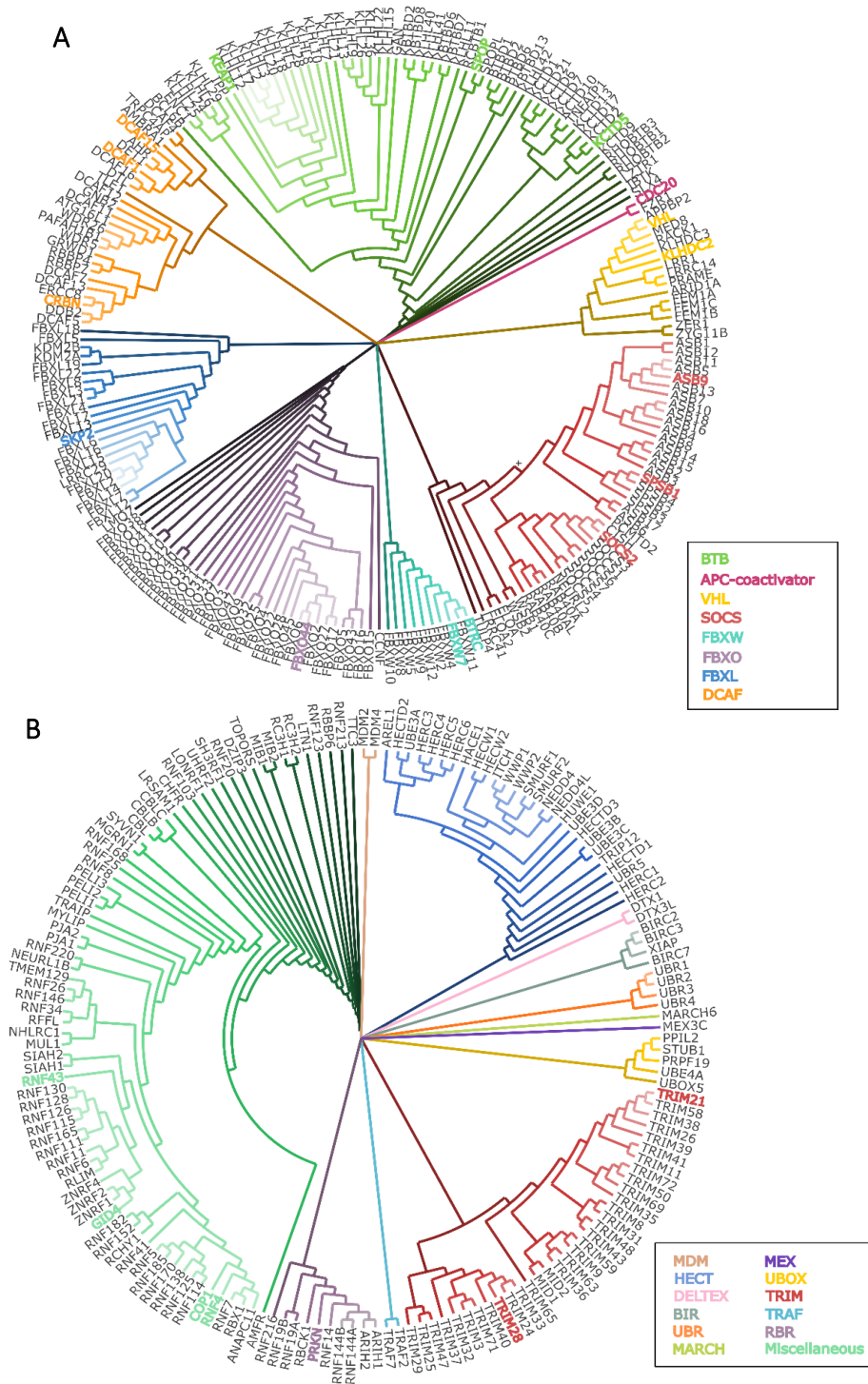
Figure 4.2. Dendrograms of complex (A) and simple (B) E3 ligases. Selected E3 ligases for MDMix experiments are highlighted. Modified from Ubihub[124].

4.3.2. Mixed solvent molecular dynamics simulations

Multiple experimental studies have demonstrated the ability of proteins to unspecifically bind small organic co-solvents in regions over the surface, and these regions correlate well with binding sites and important interactions for the drug binding affinity[52]. MDMix exploits this idea of promiscuity to identify hotspots using small co-solvent molecules. MDMix was carried out for all 23 E3 ligases by performing 50 ns molecular dynamics simulations with ethanol-water (1:4) solvent mixtures. Specifically, the adaptor proteins present in ASB9, CRBN, DCAF15, FBO44, FBW7, SKP2, SOCS2 and VHL crystal structures were conserved and included in the simulations. MDMix provided binding hotspots over the proteins surface, which were clustered by surface distance. Clusters were built 6 Å around the most probable (energetic) hotspots. To ensure they were good starting points to build protein-ligand anchors, a minimum density of hotspots was required ($\geq$ 4 hotspots in a maximum of 500 Å^3). The most potential ligandable pockets were selected manually considering the surface geometry of the protein and the drug-likeness of the cluster.

In the next pages, results of this study are developed, but in abstract, all the studied E3 ligases presented at least one potential ligandable pocket, except RNF4. We hypothesized that this is related to the fact that the structural information provided consisted of 35% of a total sequence of 190 amino acids[127]. Overall, 2.3 ligandable pockets per E3 ligase were described by MDMix.

4.3.2.1. MDMix results compared with already crystallized ligands

To assess MDMix capacity retrospectively, E3 ligases with already crystallized ligands were studied (6 out of 23). These E3 ligases were: BTRC, CDC20, CRBN, DCAF15, KEAP1 and VHL. From a drug discovery point of view, VHL is one of the most exploited E3 ligases, not only for its biological interest, but also for PROTAC construction. Interestingly, one of the most potent cluster of hotspots (VHL-Pocket E) identified by MDMix was a polar hotspot that overlapped with the hydroxyl group of the hydroxiproline post translational modification, which is also characteristic of orthosteric VHL inhibitors. Indeed, an hydrophobic hotspot was placed coinciding with the typical tert-butyl group of VHL

ligands, which was determined to provide an enhance of efficiency[116] (Figure 4.3 A). MDMix was also able to predict the binding site of the VHL allosteric ligand described, corresponding to VHL-Pocket C[128] (Figure 4.3 B).

CRBN is well-known for interacting with immunomodulatory drugs (IMiDs), which promote the degradation of non-native protein substrates. CRBN-Pocket G predicted by MDMix coincided with the binding site of these molecular glues. Despite the main hotspot corresponded to the hydrophobic site of glutarimide moiety of IMiDs[129], polar interactions were also overlapping. For the phthalimide moiety, just the pyrrolidine group was coinciding, since the aniline one was found to be solvent exposed (Figure 4.3 C).

In the same manner as IMiDs, indisulam was described as a molecular glue between DCAF15 and RBM39[130], causing the degradation of this last. Despite it was not selected as ligandable pocket for not accomplishing drug-like features, DCAF15-Pocket L was found in indisulam binding site. However, the interactions were not overlapping perfectly (Figure 4.3 D).

KEAP1 described ligands bind in the degron recognition site, inhibiting the interaction with its natural substrate NRF2[131,117,132]. MDMix hotspots overlapped with the interactions of compound 7 (PDB code 5NFU, K_d 1.3 nM)[117] with SER602 and TYR334 of KEAP1 (KEAP1-Pocket B). However, the rest of interactions that compound 7 shares with other inhibitors[131] were not predicted. This fact is reasonable, since these compounds have a carboxylic acid moiety that mimics the substrate interaction with ARG483 placed at the other site of the pocket (Figure 4.3 E). Similarly to KEAP1, an energetic cluster of hotspots (BTRC-Pocket C) appeared in the ligand and substrate binding site of BTRC. Although hotspots did not match completely with the substrate and ligand interactions, the most energetic hotspot (hydrophobic) coincided with the trifluoromethyl moiety from the ligands (Figure 4.3 F).

The CDC20 inhibitor Apcin[133] was corresponding with CDC20-Pocket E predicted by MDMix as well. Despite being mainly a hydrophobic cluster with the region of the Apcin trifluoromethyl overlapping, the hydrogen bond with CDC20 Asp177 was also predicted (Figure 4.3 G). Taken

together, these data showed consistency in MDMix results for ligand design.

A. VHL- Pocket E

B. VHL- Pocket C

C. CRBN-Pocket G

D. DCAF15-Pocket L
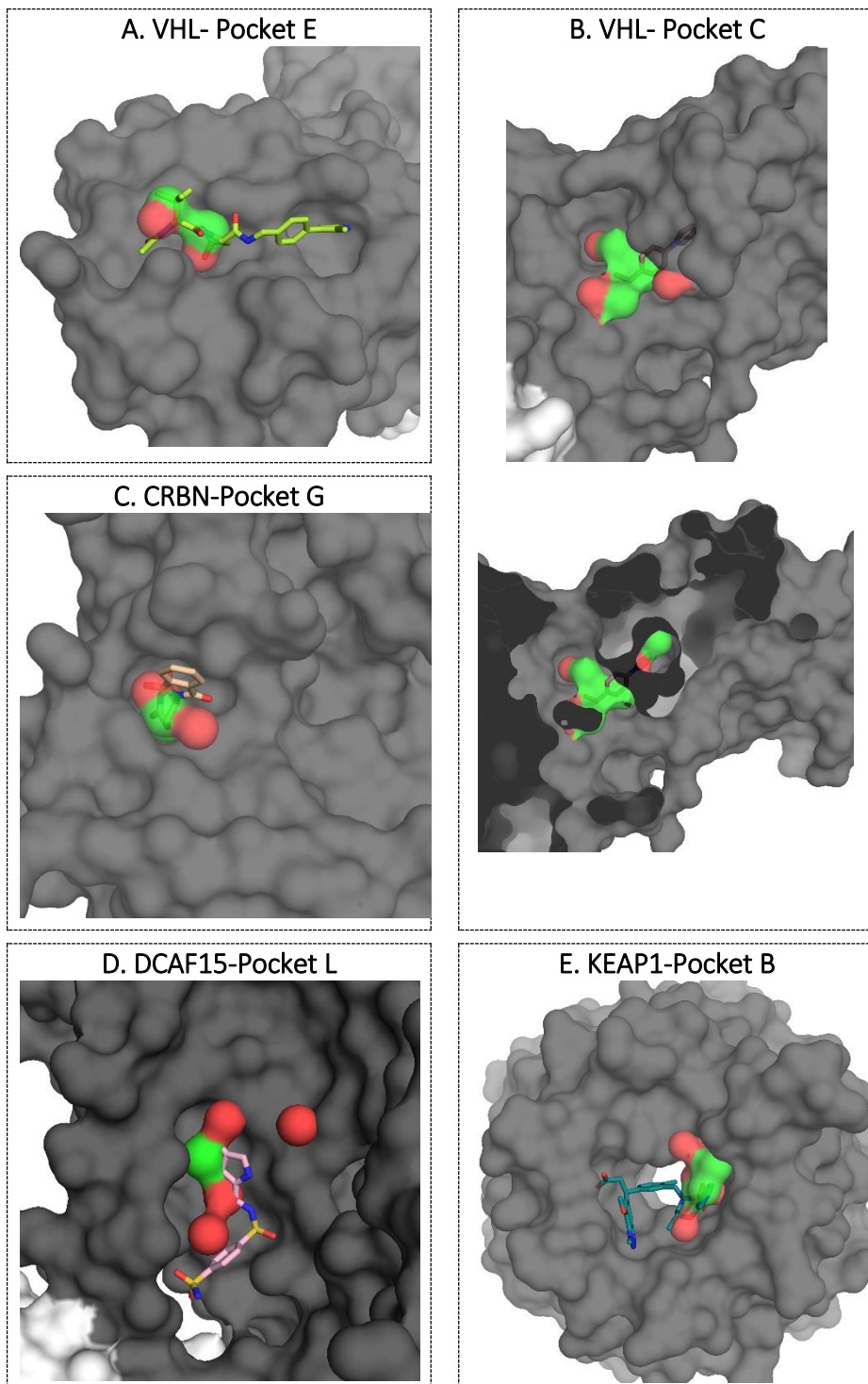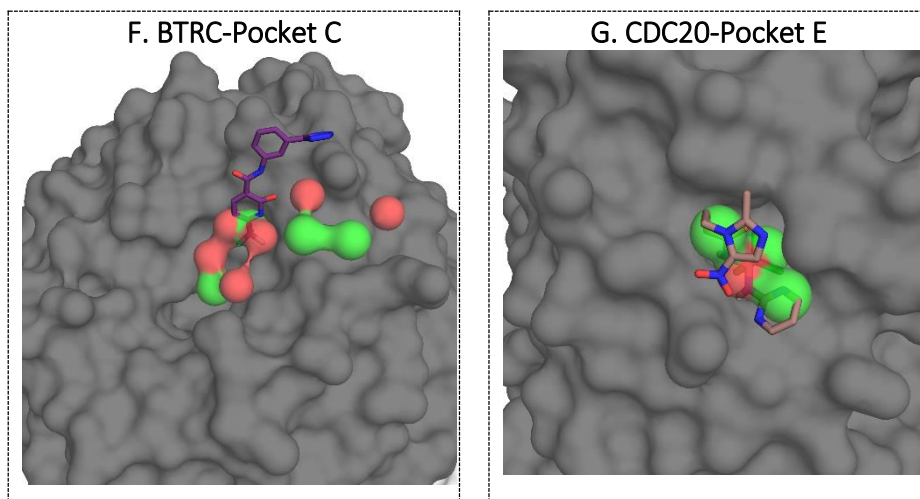
E. KEAP1-Pocket B

Figure 4.3. Structure images of E3 ligases with crystallized ligands overlapped with MDMix hotspots: VHL-Pocket E (A), VHL-Pocket C (B), CRBN-Pocket G (C), DCAF15-Pocket L (D), KEAP1-Pocket B (E), BTRC-Pocket C (F) and CDC20-Pocket E (G). In all cases, protein structure is shown in grey (surface mode), crystallized ligands are shown in different colours (stick mode), and polar and hydrophobic hotspots are shown in red and green surface, respectively**.**

4.3.2.2. MDMix results compared with degron recognition sites

As previously mentioned in the cases of VHL, KEAP1 and BTRC, several E3 ligase ligands have been developed for binding to the degron recognition site, modulating E3 ligases activity. MDMix was able to predict hotspots placed on protein pockets in the degron recognition site. CRBN is one of the most exploited E3 ligases, especially for PROTAC construction. However, its most studied functions are related to drug-dependent degradation of neo-substrates. In fact, few is known about CRBN native substrates[134], which are predicted to be involved in the Lon N-terminal domain[135]. Three different pockets (CRBN-Pocket B, CRBN-Pocket J and CRBN-Pocket E) were reported by MDMix in this domain, which hotspots accomplished the characteristics of energy, density and drug-likeness to become ligandable pockets. Interestingly, all of them were found near the interface of CRBN adaptor DDB1.

Another member of the DCAF family studied, was DCAF1 E3 Ligase. In this particular case, the protein structure available consisted on the WDR domain, which was involved in lentiviral infections. DCAF1 recognizes the

viral accession protein (VPX), which recruits SAMHD1 for proteasomal degradation[136]. An energetic cluster of hotspots (DCAF1-Pocket A) was determined by MDMix corresponding to VPX binding site. TRIM21 is another E3 ligase involved in viral infections, since it recruits for degradation antibody-bound virus[137]. Crystal structure disentangling the interactions between TRIM21 and IG gamma was elucidated by James et al.[138]. MDMix predicted part of this recognition site (TRIM21-Pocket C), in particular the interaction with TRIM21 ASN451 (Figure 4.4 A).

As a representative of RBR E3 ligases family, PRKN had two druggable pockets predicted with MDMix. PRKN-Pocket C was placed in the substrate binding region (described with its substrate SYT11[139]) and was also close to the described ubiquitin binding site[140]. PRKN-Pocket A appeared to be in the interface between this substrate binding region and the ubiquitin-like domain, being the last also related to PRKN modulation. Moreover, PRKN-pocket A was placed close to PRKN SER65 in ubiquitin-like domain, which phosphorylation by PINK1 has been related to PRKN activation[141].

GID4 E3 ligase structure was recently released recognizing several synthetic peptides emulating the degron[142,143]. MDMix determined GID4-Pocket D, predicting proline interactions of the initial series of peptides with GLU237 and TYR258[142]. Besides, MDMix hydrophobic hotspots were placed in GID4 hydrophobic pocket (TYR273, PHE254, LEU240, LEU171, LEU159, VAL141, ILE161 and LEU164) (Figure 4.4 B).

On 2018, Rusnac *et al.* presented a study of degron recognition by KLHDC2[144], resulting in different structures of the E3 ligase bound to degron peptides. KLHDC2-Pocket B assessed by MDMix coincided with part of those peptides, particularly the ones forming hydrogen bonds with KHLDC2 ARG189 and LYS147 (Figure 4.4 C).

SKP2 degrades its substrate p27 using CSK1B as a protein anchor[145], being the modulation of CSK1B-SKP2 interface able to regulate p27. In fact, MDMix SKP2-Pocket D hotspots overlapped with SER41 of CSK1B (Figure. 4.7 D). COP1 is another protein that uses a protein anchor to degrade substrates, i.e. TRIB1[146]. As MDMix COP1-Pocket D hotspots, TRIB1 peptides have been described to interact in the donut hole of COP1. Indeed, they overlapped when performing interactions with LYS472, PHE645 and TRP517 (Figure 4.4 E).

Considering we were studying protein-protein interactions, it was consistent to find some degron sites coinciding with MDMix "non-ligandable" pockets. This was the case of the hotspots presented in the members of the SOCS family SPSB1 and SOCS2. MDMix predicted SPSB1-Pocket G, a quite polar cluster of hotspots placed in a flat area of the degron recognition site of the E3 ligase. Interestingly, it overlapped of part of the already crystallized substrate peptides[147]. SOCS2 structure interacting with substrate peptides was recently described by Kung et al.[148] MDMix SOCS2-Pocket B and part of SOCS2-Pocket D were placed in the degron recognition site, even the interaction with the phosphotyrosines characteristic of SOCS substrates[149] was not predicted. These pockets were not selected because they were placed in a flat area of SOCS2 (Figure 4.4 F).

A summary of these results is listed in Table 4.2. All these data provided insights to illuminate the different possibilities of ligandability in degron recognition sites. Besides, it remarks the importance of protein surface geometry in ligand design and the challenges faced when targeting protein-protein interactions.

A. TRIM21-Pocket C
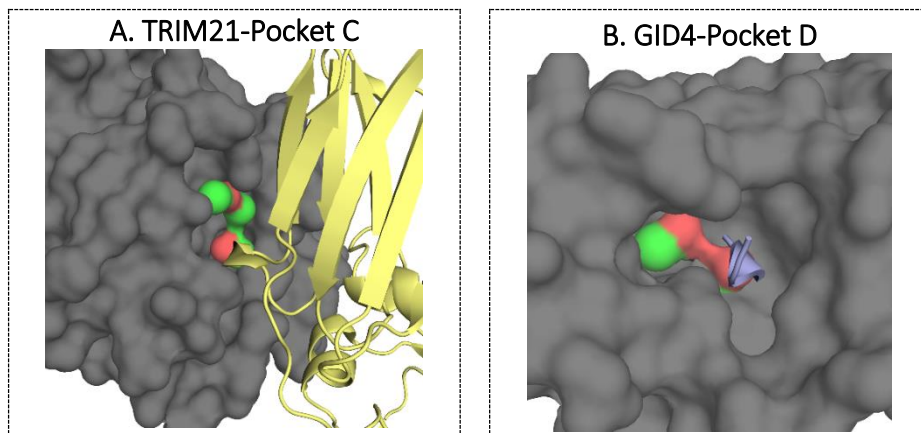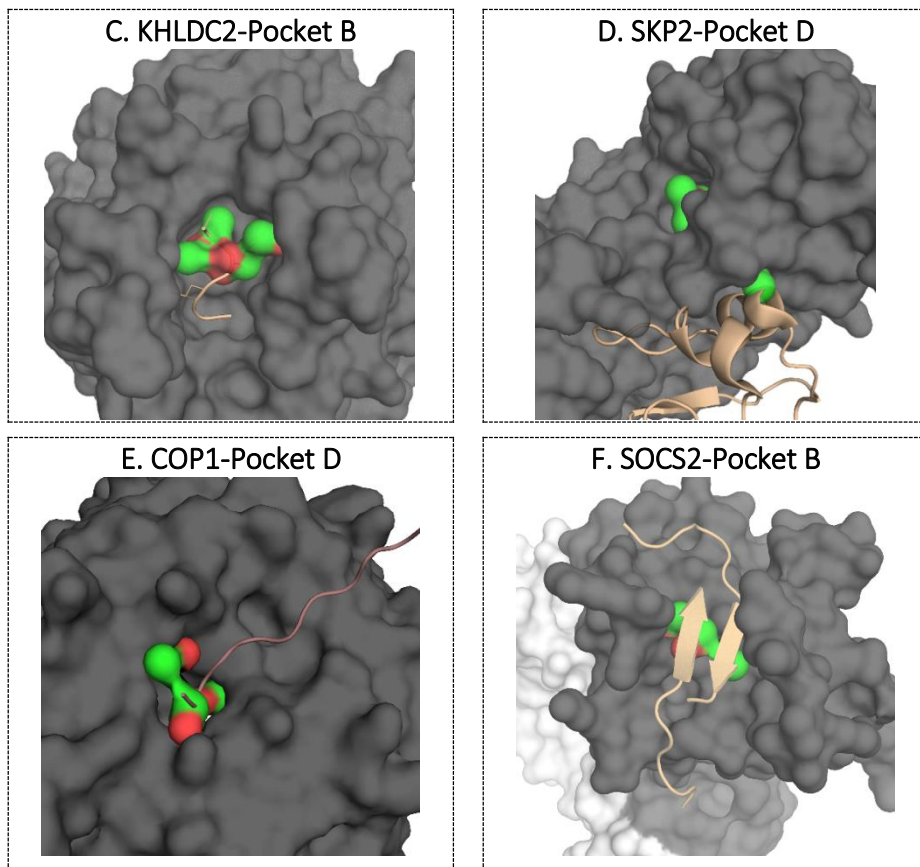
B. GID4-Pocket D

Figure 4.4. Structure images of E3 ligases with crystallized substrates overlapped with MDMix hotspots: TRIM21-Pocket C (A), GID4-Pocket D (B), KHLDC2-Pocket B (C), SKP2-Pocket D (D), COP1-Pocket D (E), and SOCS2-Pocket B (F). E3 ligase structure is shown in grey (surface mode), crystallized substrates are shown in different colours (cartoon mode), and polar and hydrophobic hotspots are shown in red and green surface, respectively.

Table 4.2. Summary of MDMix results regarding substrate recognition site

E3 ligase	Known degron site	Crystallized with substrate or substrate adaptor	Pockets found in degron site or close	Degron site is in ligandable pocket
ASB9	Yes	Yes	B, C, M	No
BTRC	Yes	Yes	C	Yes
CDC20	Yes	Yes	No	No
COP1	Yes	Yes	D	Yes
CRBN	Predicted	No	B, J, E	Yes
DCAF1	Yes	Yes	A	Yes
DCAF15	No	-	-	-

Table 4.2 (cont.). Summary of MDMix results regarding substrate recognition site

E3 ligase	Known degron site	Crystallized with substrate or substrate adaptor	Pockets found in degron site or close	Degron site is in ligandable pocket
FBO44	No	-	-	-
FBW7	Yes	Yes	B	Yes
GID4	Yes	Yes*	D	No
KCTD5	No	-	-	-
KEAP1	Yes	Yes	B	Yes
KLHDC2	Yes	Yes	B	Yes
PRKN	Yes	No	D, F	Yes
RNF4	Yes	No	-	-
RNF43	No	Yes	No	No
SKP2	Yes	Yes	D	Yes
SOCS2	Yes	Yes	B, D	No
SPOP	Yes	Yes	-	No
SPSB1	Yes	Yes	G	No
TRIM21	Yes	Yes	C	Yes
TRIM28	Yes	No	J, O, I	No
VHL	Yes	Yes	E	Yes

*GID4 is not crystallized with natural substrates but rather with synthetic peptides emulating the substrate.

4.3.2.3. MDMix predicting novel and relevant allosteric hotspots

MDMix can be a useful tool to discover new protein cavities that have never been explored previously. Indeed, 66% of all predicted pockets were novel and allosteric, not being described before. A summary of MDMix results regarding allosteric pockets is listed in Table 4.3. Despite using minimum restraints, MDMix was able to predict some potential cryptic pockets such in the case of DCAF1-Pocket J (Figure 4.5 A), which clearly opened during simulation to allow the interaction with ethanol molecules. Similarly, RNF43-Pocket F also opened during simulation giving place to a deep cavity in the protein surface. Other interesting pockets (RNF43-Pocket C, RNF43-Pocket D and RNF43-Pocket E) were discovered for this E3 ligase placed in the binding site of RSPO1, a modulator of RNF43 levels[150].

SOCS-box family has no ligands described nowadays[123] and despite no druggable pockets were found in the degron binding site using MDMix, all family members studied presented potential allosteric pockets. SOCS2-Pocket A was slightly opened during simulation and could be an interesting SOCS2 modulator for being close to Elognin C binding site. SPSB1 presented very energetic hotspots, especially in SPSB1-Pocket B (Figure 4.5 B). Although SPSB1-Pocket C was placed in a flatter protein surface, it had favourable energy values and possibilities for growing through the cavity using the hotspots not automatically selected. ASB9 also presented two potential pockets: ASB9-Pocket H, which was small but had a high energetic hotspot, and ASB9-Pocket G, which had growing possibilities and was quite close to CKB binding site[151].

Similarly, to SOCS-box family, E3 ligases with KCTD domains in the BTB family are also orphan of small-molecule modulators. KCTD5 was the member of this subfamily studied and presented three potential pockets, being KCTD5-Pocket E and KCTD5-Pocket A1 quite close in proximity in the protein surface and more drug-sized pockets. Besides, KCTD5-Pocket H was a fragment-sized pocket with energetic hotspots (Figure 4.5 C). With a different structural motif, SPOP was another member of the BTB family studied by MDMix, which presented a really promising allosteric pocket (SPOP-Pocket C) with high energetic hotspots (Figure 4.5 D).

TRIM E3 Ligases are a structurally diverse family of proteins, although most of them contain a RING-finger domain, a B1-box and/or a B2-box domain, and a coiled-coil region[152]. Allosteric regions with potential to be targeted were found in TRIM21. While TRIM21-Pocket A had possibilities of growing considering the non-automatically selected hotspots, TRIM21-Pocket D was fragment-sized but very energetic. Two different simulations were carried out with TRIM28, studying RBCC domain and PHD-BRD domain separately, which present ubiquitin E3 ligase activity and SUMO E3 ligase activity, respectively. Having substrates with different characteristics than other family members, TRIM28 PHD-BRD function has been considered atypical[153]. Two interesting allosteric pockets were found in this particular domain, being TRIM28 PHD-BRD-Pocket C the most ligandable, since it presented a deep pocket with good ligand efficiency (Figure 4.5 E). TRIM28 RBCC

allosteric pockets were located in the central region of the dimer in the coiled-coil domain. Whereas TRIM28 RBCC-Pocket B and TRIM28 RBCC-Pocket K presented nice cavities with favourable energy values, TRIM28 RBCC-Pocket A was found to be close to M312 of both assemblies of the dimer, resulting an interesting pocket for the development of covalent binders.

Regarding the atypical E3 ligases, two different cavities were assessed in GID4 outside the degron recognition site, GID4-Pocket A and GID4-Pocket B. However, the most energetic hotspots in both of them were hydrophobic. In CDC20 WDR domain, it was predicted CDC20-Pocket A, which despite being flat, was predicted to have high efficiency and probabilities of growing.

Potential allosteric regions were also found in members of the F-BOX family. For instance, FBO44-Pocket B was predicted, even though it was mainly hydrophobic. In both members of the FBXW subfamily tested (FBW7 and BTRC) this allosteric pockets were also determined. BTRC-Pocket A was a polar cavity with a hydrophobic hotspot as the most energetic one (Figure 4.5 F). In a similar region of the protein was found FBW7-Pocket G (Figure 4.9 C), despite in this case the anchor hotspot predicted by MDMix was a polar one.

Table 4.3. Summary of MDMix results regarding predicted novel and allosteric pockets

E3 ligase	Known allosteric pocket with ligand	Known allosteric pocket predicted by MDMix	Novel/allosteric pockets by MDMix
ASB9	No	-	Yes
BTRC	No	-	Yes
CDC20	Yes	Yes	Yes
COP1	No	-	No
CRBN	Yes	Yes	Yes
DCAF1	No	-	Yes
DCAF15	Yes	Yes	Yes
FBO44	No	-	Yes
FBW7	No	-	Yes
GID4	No	-	Yes
KCTD5	-	-	Yes
KEAP1	No	-	Yes
KLHDC2	No	-	Yes
PRKN	No	-	No

Table 4.3 (cont.). Summary of MDMix results regarding predicted novel and allosteric pockets

E3 ligase	Known allosteric pocket with ligand	Known allosteric pocket predicted by MDMix	Novel/allosteric pockets by MDMix
RNF4	Yes (covalent)	No	No
RNF43	No	-	Yes
SKP2	Yes	No	No
SOCS2	No	-	Yes
SPOP	No	-	Yes
SPSB1	No	-	Yes
TRIM21	No	-	Yes
TRIM28	No	-	Yes
VHL	Yes	Yes	Yes

A. DCAF1-Pocket J

B. SPSB1-Pocket B

C. KCTD5-Pocket H
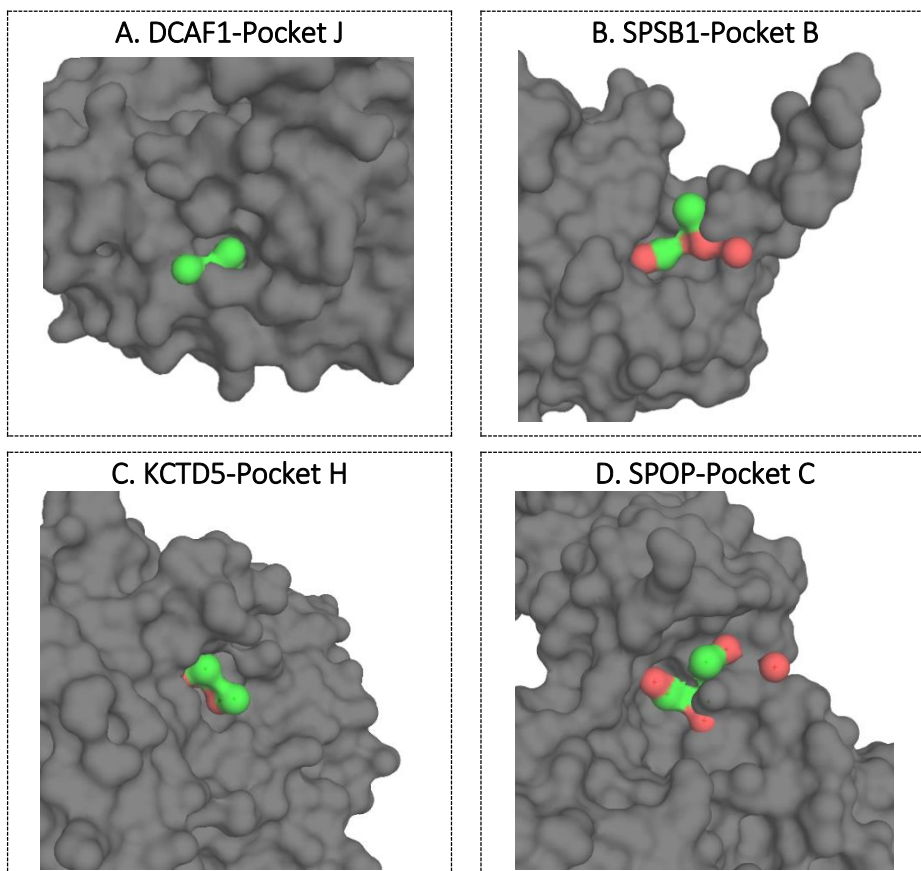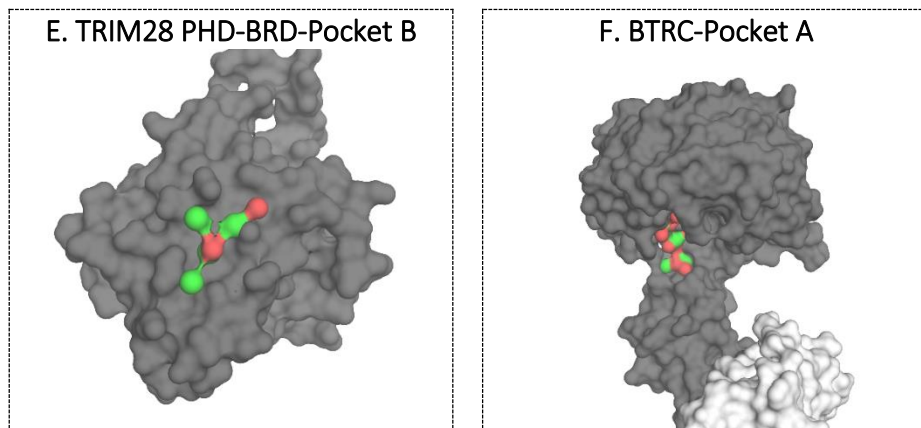
D. SPOP-Pocket C

Figure 4.5. MDMix novel allosteric pockets in the surface of E3 ligases. Structure images of DCAF1-Pocket J (A), SPSB1-Pocket B (B), KCTD5-Pocket H (C), SPOP-Pocket C (D), TRIM28 PHD-BRD-Pocket B (E), and BTRC-Pocket A (F). In all cases, E3 ligase structure is shown in grey (surface mode), and polar and hydrophobic hotspots are shown in red and green surface, respectively.

4.4. FBW7 E3 ligase: an undruggable protein with relevance in cancer

4.4.1. FBW7 structure and complex formation

RING E3 ligases can be divided in two sub-categories: simple RING E3 and multi-module cullin-RING ligases (CRLs). This last subtype of E3 ligases consist of four protein modules: RING-Box protein (RBX), which contains the RING domain binding E2; cullin protein; an adaptor protein, connecting the cullin to the substrate binding protein; and the substrate binding protein. Substrate binding proteins include two different domains: one that recognises the substrate and another that interacts with the adaptor[154]. Depending on all these features, CRLs can be divided in different subgroups. From these last, FBW7 is an SCF E3 ligase, what stands for S-phase kinase-associated protein 1 (SKP1)–Cul1–F-box protein), since it has an F-box domain recognising the adaptor SKP1. A schematic illustration of the complex formed with FBW7 for ubiquitination is depicted in Figure 4.6 A.

Up to date there are 4 structures in the Protein Data Bank (PDB) that contain FBW7 E3 ligase. In all of them Dr Bing Hao has been involved. FBW7 has been crystallized in complex with SKP1, either alone (PDB code

2OVP) or in complex with peptides of its natural substrates: two different CYCLIN-E peptides (PDB codes 2OVR and 2OVP)[95] and a DISC1 peptide (PDB code 5V4B)[98]. In all structures truncated SKP1 protein was produced with a human FBW7 fragment (residues 263-707). FBW7 structure is comprised by three different domains: an F-box domain (helices H-1, H0, H1-3), an α-helical linker domain (helices H4 and H5 and H4-H5 loop) and a WD40-repeat domain (Figure 4.6 B). WD40 domain is the domain in charge of substrate recognition. It is formed by a canonical eight-bladed β-propeller structure, with each blade consisting of four antiparallel β-strands (strands A to D) (Figure 4.6 C).
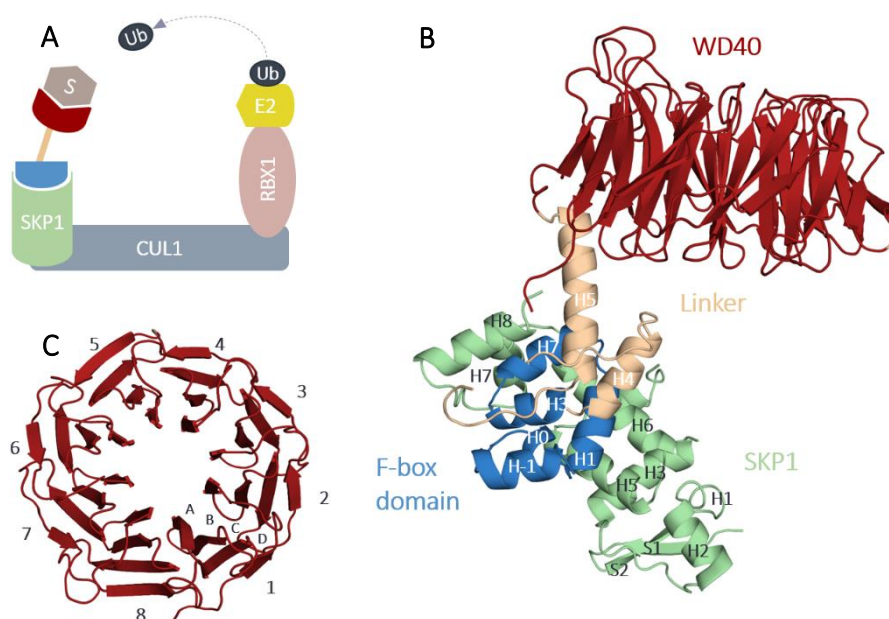
Figure 4.6. FBW7-SKP1 structure. A. Schematic illustration of the complex formed with FBW7 for ubiquitination procedure. FBW7 domains are coloured in blue (F-box), orange (linker) and red (WD40). **B.** Structure of FBW7 domains (F-box in blue, linker in orange and WD40 in red) interacting with SKP1 (green). The secondary structure elements of F-box, linker and SKP1 are labelled. **C.** WD40 domain with the eight β-propeller and the strands for one blade labelled. Abbreviations: S, substrate; Ub, ubiquitin.

Several computational studies were performed studying the mechanism of ubiquitination of the CLRs[154,155,156]. According to models that pieced together crystal structures of the CRL components, a gap of 50-60 Å is placed between the ubiquitin E2 binding site and the tip of the substrate binding protein. In order to disentangle the mechanism of

ubiquitination, Dr Ruth Nussinov performed molecular dynamics simulations of different CRLs, including FBW7. Substrate binding proteins were simulated in two forms: bound to the substrate and bound to both the substrate and the adaptor. In both cases the gap distance with the E2 was reduced, as a result of allosterically controlled linker domain motions, facilitating ubiquitin transfer. Consequently the linker domain of the substrate binding proteins, as FBW7, could be an allosteric mechanism to modulate protein function[155].

4.4.2. Therapeutic relevance of FBW7 E3 ligase

FBW7 is one of the most commonly deregulated UPS protein in human cancers, which targets a range of substrates for degradation including some key human oncoproteins such as CYCLIN-E, c-MYC, NOTCH and JUNK[157]. Diverse studies have proven its central role in tumorigenesis, since FBW7 is considered to be among the most mutated cancer genes and its tumour suppressor function has been demonstrated[158]. In fact, it has been reported that approximately 6% of all cancers have mutations in FBW7[159]. Furthermore, FBW7 has showed an important role in sensitizing cancer stem cells to chemotherapies[160]. More recently, FBW7 has been related to neuropsychiatric diseases, being in charge of the degradation of DISC1. This protein has been described to promote neural development and signally. Consequently, inhibition of FBW7 would stabilize DISC1 levels and could be beneficial to treat schizophrenia[98]. Given FBW7 relevance, finding modulators for this enzyme would be of utmost value from a therapeutic point of view while expanding the toolbox of E3 ligases.

4.4.3. FBW7 substrate recognition and mutations

As previously mentioned, FBW7 recognize its substrates in the WD40 domain in the so-called degron site. Most of FBW7 substrates contain a conserved phosphorylation motif (CPD), which sequence consist of: (L)-X-pT/pS-P-P-X-pS/pT/E/D (X represents any amino acid and p represents phosphorylation)[161]. Figure 4.7 illustrates the structural motif of relevant substrates and the final CPD. For FBW7 recognition is essential that substrates undergo post-translational modifications involving phosphorylation of serine or threonine residues. As a consequence,

phosphorylated residues interact with key arginine residues of the WD40 domain[158].

NOTCH1	**TP**TL**S**PPL
CYCLIN-E	**TPP**Q**S**GK
c-MYC	**TPP**L**S**PS
c-JUN	**TPP**L**S**PI
SREBP1	**TPP**P**S**DA
SV40LTA	**TPP**P**E**PE
DISC1	**TPP**G**S**HS
Consensus	**TPP**X**S**

Figure 4.7. Conserved phosphorylation motif of FBW7 substrates. Modified from [158].

According to COSMIC database, mutations in FBW7 are specially concentrated at the WD40 domain, particularly in three arginine residues (ARG465, ARG479 and ARG505) crucial for substrate recognition (Figure 4.8 A). In consequence of these mutations, FBW7 is not able to recognise its natural substrates leading to their accumulation. Since FBW7 is responsible for degradation of several oncoproteins, these mutations can induce cancer malignancies. Figure 4.8 B illustrate the FBW7 mutation frequency in different human cancers.
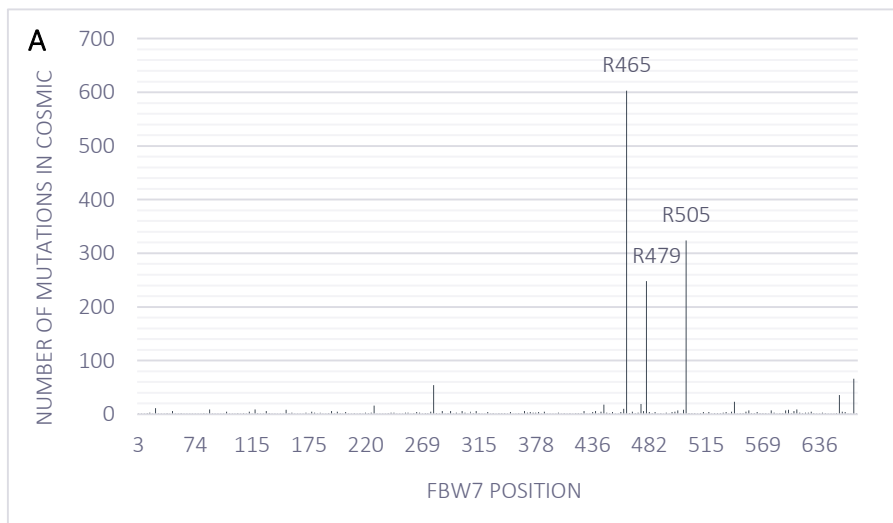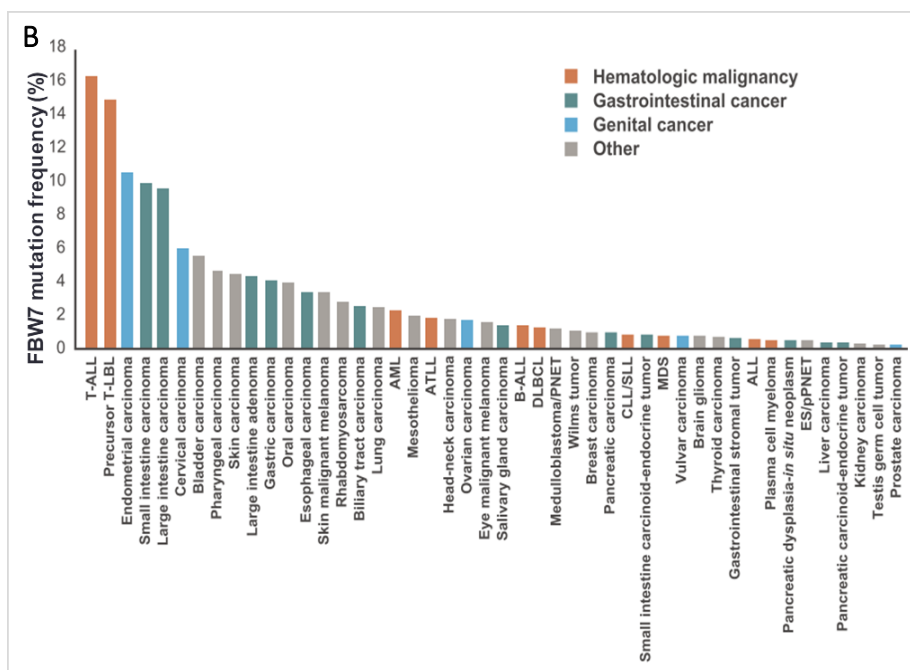
Figure 4.8. *FBW7* gene mutations. A. Distribution of mutations *FBW7* gene according to COSMIC database. The most persistent mutations are labelled. **B.** *FBW7* mutation frequency for different human cancer types in the COSMIC database. Cancer types with >100 patients registered are listed. Adapted from [158].

4.5. Identification of FBW7-SKP1 small-molecule binders

4.5.1. MDMix of FBW7: ligandable sites and pocket selection

Three regions in FBW7 E3 ligase more prompt to be targeted were identified using MDMmix: Pocket B, Pocket D and Pocket G (Figure 4.9). FBW7-Pocket B was found in the WD40 domain, close to the substrate binding site. It presented 8 different hotspots, being the sum of all of their energies -7.6 kcal/mol. FBW7-Pocket D was placed at the interface with SKP1, close to a flexible loop of FBW7. It was comprised of 5 hotspots with a total energy value of -5.31 kcal/mol. Finally, FBW7-Pocket G had 4 hotspots with a total energy value of -4.53 kcal/mol. Despite being the least energetic, FBW7-Pocket G presented the highest pocket efficiency (total energy/number of hotspots), being -1.13 kcal/mol. Besides, this particular pocket was found in the interface between the WDR and the linker domains of FBW7, being the flexibility

of this last described to play an important role in E3 ligases substrate ubiquitination[155] (Section 4.4.1). Indeed, one amino acid placed in the pocket, TRP365, which a hotspot for a hydrophobic interaction was determined, was found to be preserved among 70% of F-box/WD repeat proteins in all species. For these reasons, FBW7-Pocket G was selected to develop FBW7 binders from its hotspots.

FBW7-Pokcet G consisted on 4 different hotspots, two hydrophobic (ΔG_{bind} values of -1.29 kcal/mol and -0.99 kcal/mol) and two polar, one of which would interact as a hydrogen bond acceptor (ΔG_{bind} value: -1.28 kcal/mol) and the other as a hydrogen bond donor (ΔG_{bind} value: -0.97 kcal/mol). The hotspots assessed by MDMix were transformed into pharmacophoric restraints in order to perform a high-throughput virtual screening (HTVS).

A   B
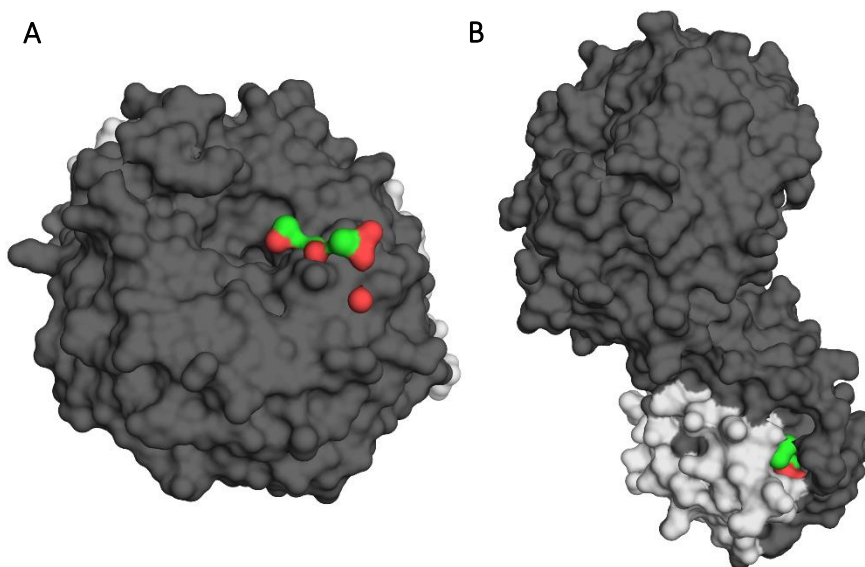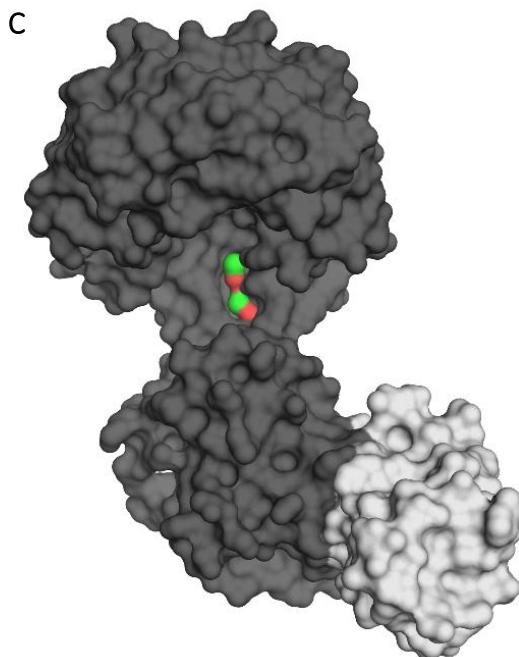
Figure 4.9. MDMix ligandable pockets from FBW7 (dark grey) in complex with SKP1 (white): FBW7-Pocket B (A), FBW7-Pocket D (B) and FBW7-Pocket G (C). Polar hotspots are shown in red and hydrophobic ones in green.

4.5.2. Virtual screening to find FBW7 binders in FBW7-Pocket G

The hotspots determined in FBW7-Pocket G were transformed into pharmacophoric restraints in order to perform HTVS, which consisted on the application of molecular docking, using rDock software[90], followed by dynamic undocking (DUck). DUck is an orthogonal methodology that consists on a simplified computational procedure that calculates the work needed to beak a key hydrogen bond and reach the quasi-bound state (W_{QB}). This technique has been proved to be useful to detect true ligands and increase the success of virtual screening campaigns[58].

In order to increase the possibilities of finding hits and due to the energy values, a HTVS campaign was initially performed considering only three out of the four hotspots transformed into pharmacophoric restraints. In particular, they consisted on the two hydrophobic interactions and the hydrogen bond acceptor. An in-house library of seven million compounds was screened with rDock, resulting in around 0.5 million compounds with favourable docking scores. To reduce the number of

molecules while preserving chemical diversity, similarity clustering was carried out ending up with around 2000 potential binders that were then filtered using DUck. For DUck experiments the key hydrogen selected was formed by an acceptor group for the compounds and FBW7 ASN635. A threshold of $W_{QB} > 4$ kcal·mol^{-1} was applied, concluding with 59 potential binders that passed all the filters from this first HTVS campaign. These compounds were manually filtered considering chemical diversity, drug-likeness and DUck scores. Finally, 38 compounds were selected in order to test FBW7 binding employing biophysical techniques.

Additionally, a second HTVS campaign was performed adding the forth hotspot that consisted on a hydrogen bond donor with SER678. Since the resulting compounds implicitly needed to accomplish the initial three pharmacophoric restraints, we rescreened the 0.5 million compounds from the first molecular docking. 453 compounds obtained favourable docking scores and were screened by DUck twice, taking separately into account the two different hydrogen bonds that they could form. On one hand, the hydrogen-bond selected was the same as for the 3 constraints docking (between an acceptor group from the compound and FBW7 ASN635), in which case just 2 molecules appeared to pass DUck filter ($W_{QB} > 4$ kcal·mol^{-1}). On the other hand, the hydrogen bond produced by a donor group of the compounds and SER678 was followed, with just 1 molecule passing the DUck filter. Interestingly, this last molecule, did not pass the initial DUck filter pulling from ASN635. These 3 molecules were also purchased to be tested, having a final amount of 41 molecules for the assessment of FBW7 binding. The overall virtual screening protocol is illustrated in Figure 4.10.
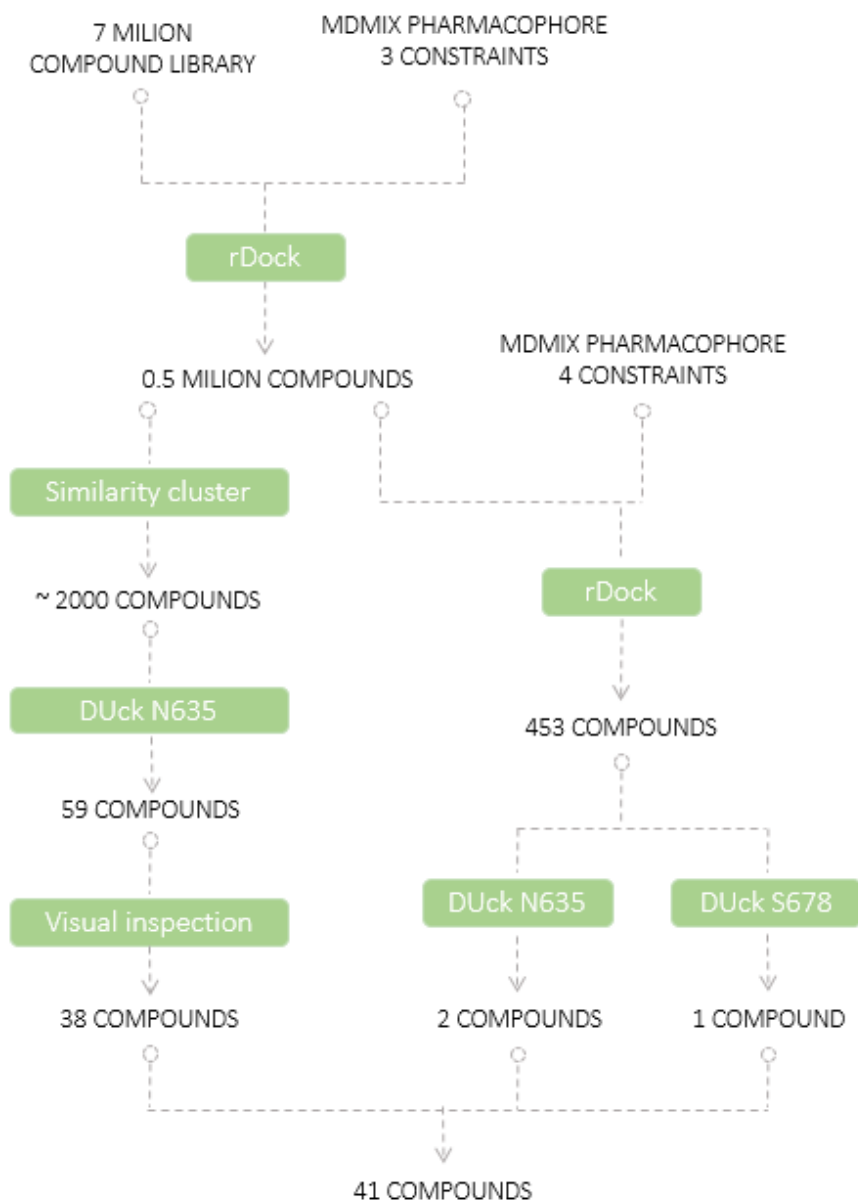
Figure 4.10. Virtual screening protocol to develop compounds able to interact with FBW7 E3 ligase from MDMix results.

4.5.3. Protein production of FBW7-SKP1 complexes

4.5.3.1. Site directed mutagenesis of FBW7-SKP1 complexes

Site directed mutagenesis of FBW7-SKP1 complexes were done with Roger Castaño, a master student that I supervised. Mutations performed are listed in Table 4.4. Site directed mutagenesis was carried out as described in Section 3.2.2.2. Forward and reverse primers were designed following the recommendations in QuickChange Manual (Agilent) and using a bioinformatic tool provided by the manufacturer[162]. Mutations were evaluated by Sanger sequencing. For that, primers were also designed using NCBI software tools. Primers had a melting temperature of more than 45 ºC and GC content between 30% and 80%. In all cases the mutation was placed at least 100bp before the edges of the amplicon. Mutations were verified by Sanger sequencing. All mutations were successfully produced as it is shown in Table 4.4.

Table 4.4. List of mutations produced in *FBW7* and results from Sanger sequencing

Mutations	Nucleotides change (Fw)		*FBW7* mutated nucleotides	Sanger sequencing
FBW7_{N635A}-SKP1	WT	AAC (ASN)	1903, 1904	
	M	GCC (ALA)		
FBW7_{N635I}-SKP1	WT	AAC (ASN)	1904	
	M	ATC (ILE)		
FBW7_{A677I}-SKP1	WT	GCC (ALA)	2029, 2030	
	M	ATC (ILE)		
FBW7_{A677F}-SKP1	WT	GCC (ALA)	2029, 2030	
	M	TTC (PHE)		

Table 4.4 (cont.). List of mutations produced in *FBW7* and results from Sanger sequencing

Mutations	Nucleotides change (Fw)		*FBW7* mutated nucleotides	Sanger sequencing
FBW7$_{N679W}$-SKP1	WT	AAC (ASN)	2035, 2036, 2037	
	M	TGG (TRP)		

Abbreviations: A (nucleotide), adenine; A and ALA (amino acid), alanine; C (nucleotide), cytosine; F and PHE (amino acid), phenylalanine; G (nucleotide), guanine; Fw, forward; I and ILE (amino acid), isoleucine; M, mutant; N and ASN (amino acid), asparagine; T (nucleotide), thymine; W and TRP (amino acid), tryptophan; WT, wild type.

4.5.3.2. Cloning and expression of FBW7-SKP1 complexes in *E. coli*

The cloned glutathione S-transferase (GST)-tagged human FBW7 (residue 263-707) and truncated SKP1 was co-expressed as a dicistronic message in *E. coli*. Both proteins were expressed together since this particular part of FBW7 cannot solubilize without SKP1. The co-expression was performed in ROS(DE3) after the induction with 1 mM of IPTG, as mentioned in Section 3.2.4.

The yield of expression was of 0.3 mg/mL, producing a low amount of protein after each purification. In order to improve this yield different bacterial strains were tried (BL21, BLD21(DE3) and BL21(DE3)pLysS) and different growth media cultures (Terrific Broth and Luria Broth). Moreover, induction conditions were also optimized changing IPTG concentration (0.4 mM or 1 mM), OD$_{600}$ (0.6-0.8 or 2.5) and induction time (10h, 18h or 24 h). None of these conditions helped to improve the yield, being the low yield of expression one of the bottlenecks of the project, especially for techniques as DSF, FP or ITC that require a considerable amount of protein in the mg range. As a consequence of that, 50L of FBW7-SKP1 WT were expressed. In the case of the mutants, which mainly were tested in SPR, 4L of culture were expressed for FBW7$_{N635A}$-SKP1, FBW7$_{N635I}$-SKP1 and FBW7$_{N679W}$-SKP1, 8L in the case of FBW7$_{A677F}$-SKP1 and 12 L for FBW7$_{A677I}$-SKP1. Mutants were expressed with Roger Castaño, a master student that I supervised.

4.5.3.3. Purification and characterization of FBW7-SKP1 complexes

Purification of all FBW7-SKP1 complexes followed the same protocol. Mutants were purified with Roger Castaño, a master student that I supervised. The purification procedure started with cell lysis followed by a clarification step. The filtered sample was purified by glutathione affinity chromatography. The GST-tag was removed by thrombin cleavage overnight. In order to remove the tag from the sample, a second glutathione affinity chromatography was performed. As a final purification step, an anion exchange chromatography was performed. Chromatograms of the purification steps are illustrated in Figure 4.11.

C

ANION EXCHANGE CHROMATOGRAPHY
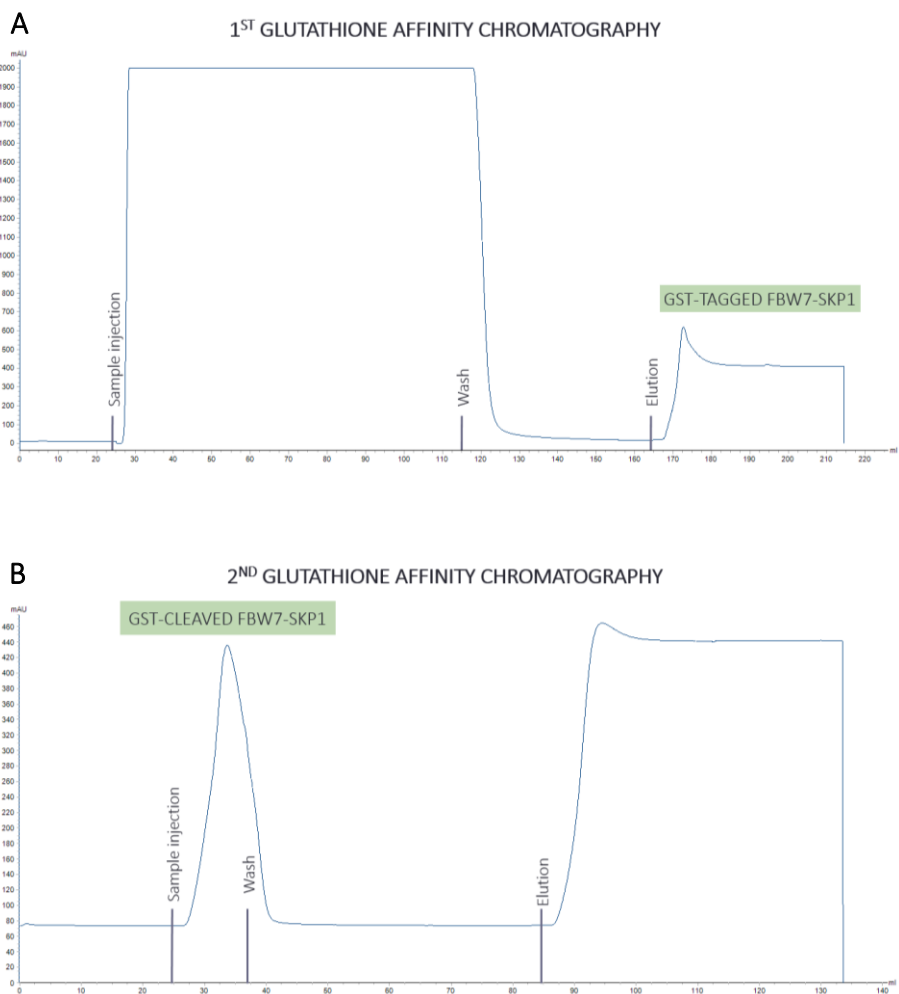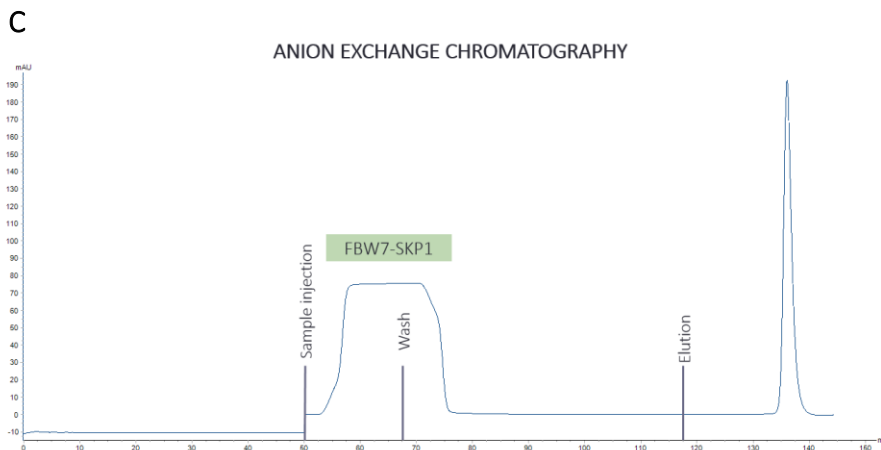
Figure 4.11. Chromatograms showing FBW7-SKP1 purification procedure. A. First glutathione affinity chromatography. **B.** Second glutathione affinity chromatography. **C.** Anion exchange chromatography.

After the purification procedure a yield of 0.3 mg/L was obtained with FBW7-SKP1 and mutants. The mass and purity of the protein were verified by SDS–PAGE and mass spectrometry (Figure 4.12).

A   B
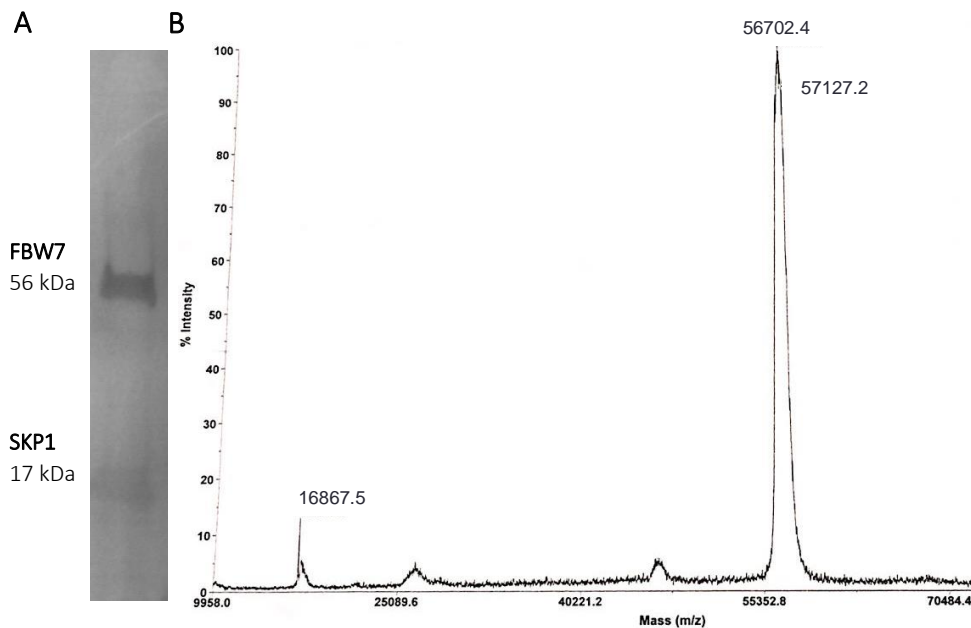
Figure 4.12. A. SDS polyacrylamide gel electrophoresis of Fbw7-Skp1 after purification. **B.** Matrix-assisted laser desorption/ionization (MALDI) of Fbw7 (56 kDa) and Skp1 (16 kDa).

In order to ensure that the protein produced was FBW7-SKP1, proteomic studies were carried out by Proteomics Unit of University of Barcelona

(CCTiUB). Results are shown in Table 4.5. Keratin proteins are contaminations of the sample by human skin or hair[163]. Giving the fact that FBW7 precipitates when is not in complex with SKP1, our conclusions lead to the fact that the protein produced is in complex.

Table 4.5. Proteomics results of protein solution

Description	Score	Coverage	AAs
F-box/WD repeat-containing protein 7 OS=Homo sapiens GN=FBXW7 PE=1 SV=1 - [FBXW7_HUMAN]	790.92	42.86	707
S-phase kinase-associated protein 1 OS=Homo sapiens GN=SKP1 PE=1 SV=2 - [SKP1_HUMAN]	143.97	46.01	163
Keratin, type II cytoskeletal 1 OS=Homo sapiens GN=KRT1 PE=1 SV=6 - [K2C1_HUMAN]	35.50	22.36	644
Keratin, type II cytoskeletal 2 epidermal OS=Homo sapiens GN=KRT2 PE=1 SV=2 - [K22E_HUMAN]	32.18	13.62	639
Keratin, type I cytoskeletal 10 OS=Homo sapiens GN=KRT10 PE=1 SV=6 - [K1C10_HUMAN]	29.56	16.10	584
Trypsin OS=Sus scrofa PE=1 SV=1 - [TRYP_PIG]	24.28	12.12	231
Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=3 - [K1C9_HUMAN]	15.18	9.31	623
Keratin, type II cytoskeletal 6B OS=Homo sapiens GN=KRT6B PE=1 SV=5 - [K2C6B_HUMAN]	6.33	3.37	564

Abbreviations: AAs, Amino acids.

Moreover, protein folding and stability was also assessed in the absence and in the presence of 5% of DMSO. Dr Salvatore Scaffidi did these experiments by carrying out 1D-NMR and recording a proton spectrum of FBW7-SKP1. The results showed how the immediate addition of DMSO did not cause an effect on protein stability. However, after 24h with DMSO a slight partial unfolding started to appear.

4.5.4. Differential scanning fluorimetry to determine binding to FBW7-SKP1

DSF was selected as a frontline primary screening to determine binding of virtual screening hits. Since DSF conditions using FBW7-SKP1 complex had never been described before, several optimization trials had to be performed. Different concentrations of protein were tested, ranging from 1 µM to 5 µM, as well as different concentrations of the SYPRO

Orange dye (2.5x and 5x), being 5 μM of FBW7-SKP1 and 5x SYPRO finally selected. In addition, buffer screening was performed by testing the stability of the protein using different tampon buffers, salt concentrations and a range of pH. 50 mM TRIS pH 8.0 and 50 mM NaCl buffer was considered the best one. Nevertheless, due to the high stability of the complex and the difficulties of being able to see a relevant thermal shift with the positive control, conditions changing the tampon buffer and adding denaturants were also tried (Section 3.2.7.1). Finally, 50 mM HEPES pH 8.0 and 50 mM NaCl was the buffer selected for the DSF screening. As a positive control, 26-mer peptide of the natural substrate CYCLIN-E (KAMLSEQNRASPLPSGLL[pT]PPQ[pS]GKK) was used (K_d 70 nM[95]). The peptide was tested at three different concentrations: 100 μM, 250 μM and 500 μM, giving thermal shifts in the 67ºC melting temperature of 0.65, 0.72 and 1.6 respectively (Figure 4.13). All 41 compounds from VS were tested at a final concentration of 500 μM or 250 μM, depending on their solubility. In order to be considered significant, thermal shifts (ΔTm) of the protein-ligand complexes (ΔTm = Tm protein-ligand complex − Tm protein) had to be at least twice the standard deviation of each Tm of FBW7-SKP1.

Unfortunately, no compound was able to stabilize the protein producing a thermal shift higher than 1ºC. Nevertheless, some of them were able to destabilize it with melting temperatures lower than -2 ºC, meaning that they could possible interact with the complex (Figure 4.14). ΔTm obtained are listed in Table 4.6. It is important to emphasize the fact that in FBW7-SKP1 DSF, the increasing of fluorescence has been observed twice, hence having two melting temperatures in the system, one at 55 ºC and the other at 67 ºC (Figure 4.13). Notwithstanding, we were not able to determine which protein or domain corresponds to each melting temperature. Establishing one for SKP1 and the other for FBW7 could make sense, but the fact that FBW7 has three different domains cannot be neglected. Moreover, thermal shift was randomly related to one melting temperature or the other, making the results hard to interpret. Due to these interpreting difficulties and the fact that little stabilization was observed in the 41 molecules from the virtual screening, no concluding results were able to be extracted from these experiments and SPR was performed as an alternative.
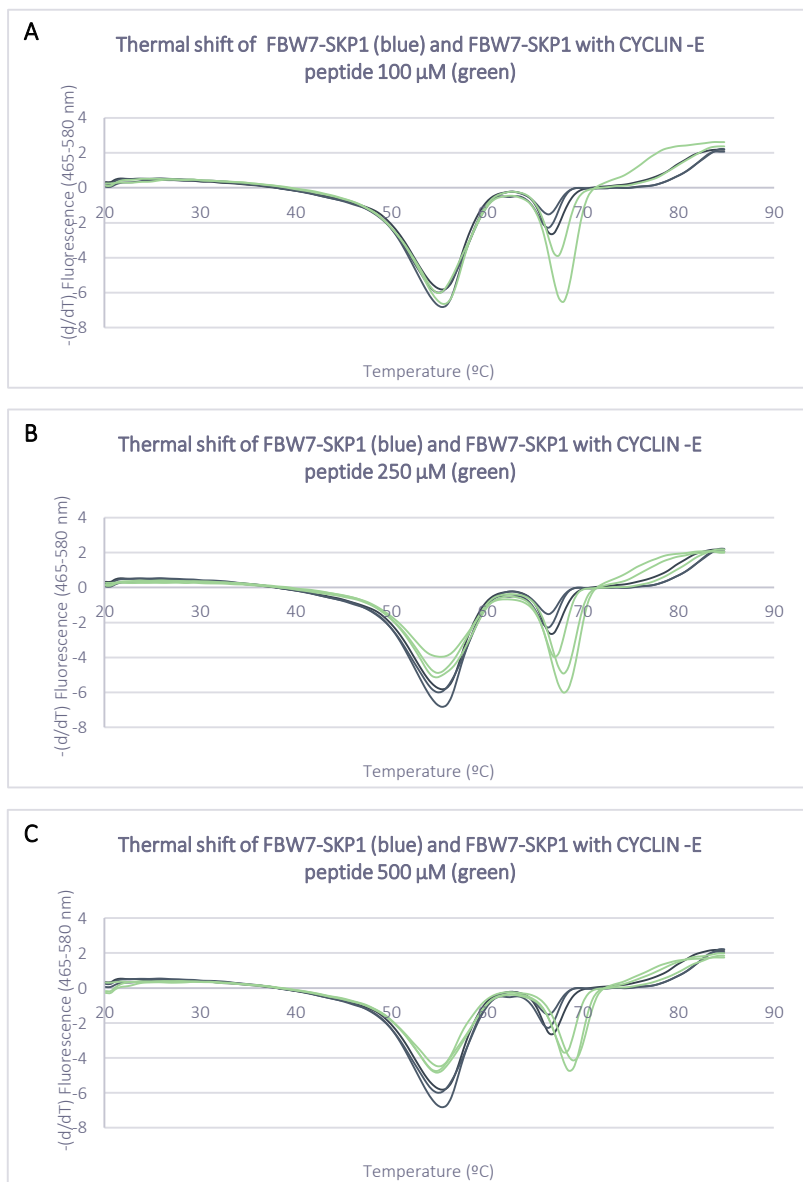
Figure 4.13. DSF graphics of FBW7-SKP1 and CYCLIN-E peptide. DSF graphics comparing melting temperatures of FBW7-SKP1 with FBW7-SKP1 in complex with 100 μM of CYCLIN-E peptide (A), 250 μM of CYCLIN-E peptide (B) and 500 μM of CYCLIN-E peptide (C).
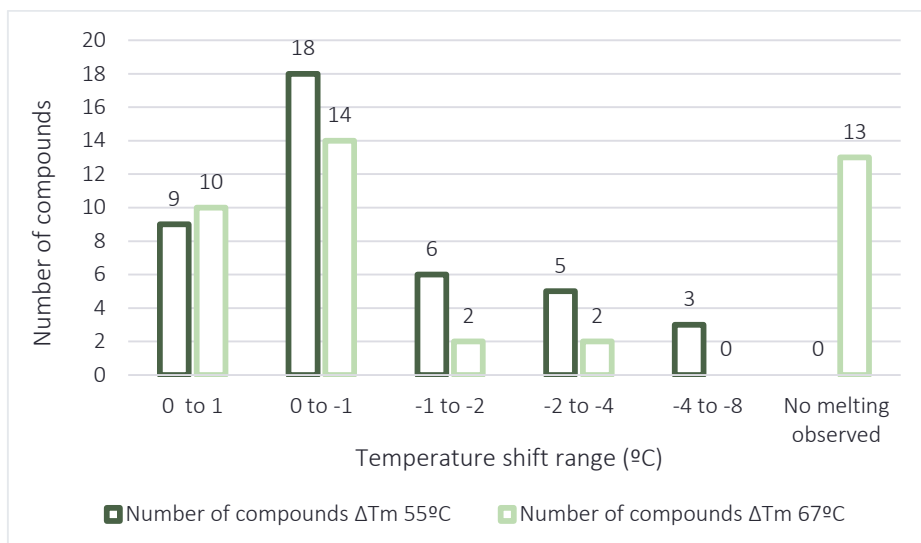
Figure 4.14. Summary of DSF results to evaluate FBW7-SKP1 interaction with VS compounds. Each column represents the number of compounds that produced a temperature shift in a particular range. For each range, temperature shift at 55ºC and at 67 ºC was evaluated. In some cases, no melting was observed at 67 ºC.

Table 4.6. DSF results of virtual screening compounds

Compound ID	[CPD]	ΔTm 55 ºC	SD Tm 55 ºC	ΔTm 67 ºC	SD Tm 67 ºC
MMC1	500 µM	-0,40	0,17	-0,04	0,70
MMC2	500 µM	-5,27	1,03		
MMC3	500 µM	-0,70	0,29	-0,56	0,23
MMC4	500 µM	-0,09	0,12	0,12	0,06
MMC5	500 µM	0,41	0,26	0,03	0,28
MMC6	250 µM	-1,95	0,16		
MMC7	500 µM	0,15	0,23	-0,02	0,25
MMC8	500 µM	-0,10	0,03	-0,53	0,05
MMC9	250 µM	-2,93	0,00		
MMC10	500 µM	-0,57	0,19	-0,29	0,03
MMC11	500 µM	-1,19	0,18	-0,02	0,03
MMC12	500 µM	-1,06	0,16	-1,82	0,21
MMC13	500 µM	-0,09	0,20	-3,07	0,21
MMC14	250 µM	0,50	0,17	-0,30	0,07
MMC15	500 µM	-0,09	0,20	0,12	0,09
MMC16	250 µM	0,13	0,17	-0,03	0,04
MMC17	500 µM	0,25	1,14		
MMC18	250 µM	-2,41	0,80		
MMC19	250 µM	-0,19	0,18	-1,46	0,20
MMC21	500 µM	-0,81	0,30	-0,13	0,23

Table 4.6 (cont.). DSF results of virtual screening compounds

Compound ID	[CPD]	ΔTm 55 ºC	SD Tm 55 ºC	ΔTm 67 ºC	SD Tm 67 ºC
MMC22	500 µM	-0,57	0,15	0,23	0,26
MMC23	500 µM	0,03	0,35	0,01	0,06
MMC24	250 µM	-1,99	0,10	-2,34	0,51
MMC25	250 µM	-0,36	0,55	-0,49	0,88
MMC26	250 µM	-6,94	0,37		
MMC27	500 µM	-0,90	0,47		
MMC29	500 µM	-3,67	0,65		
MMC31	500 µM	-1,10	0,25		
MMC32	500 µM	0,06	0,16	-0,38	0,01
MMC33	500 µM	-0,61	0,18	0,25	0,04
MMC34	500 µM	0,15	0,04	0,20	0,05
MMC35	500 µM	-5,20	0,17		
MMC36	250 µM	-1,66	0,31		
MMC37	250 µM	0,12	0,06	0,01	0,21
MMC38	250 µM	-2,84	0,50		
MMC39	500 µM	-0,43	0,25	0,14	0,03
MMC40	250 µM	-0,29	0,19	-0,02	0,00
MMC41	500 µM	-0,26	0,23		
MMC42	250 µM	-2,53	0,06	0,29	0,21
MMC43	500 µM	-0,34	0,41	-0,31	0,49
MMC44	500 µM	-0,31	0,07	-0,04	0,57

The two melting curves that FBW7-SKP1 presented have been considered. Abbreviations: ΔTm, thermal shift; CPD, compound; SD, standard deviation.

4.5.5. FBW7-SKP1 surface plasmon resonance screening for binders identification and characterization

FBW7-SKP1 complex was immobilized in a CM7 sensor chip to have immobilization levels of around 4500 RU and 9000 RU to achieve R_{max} of 25 RU and 50 RU, respectively. The immobilization protocol is described in Section 3.2.7.2. Initially, as a positive control, binding of the CYCLIN-E peptide used for DSF experiments (Section 4.5.4) was assessed. Nevertheless, CYCLIN-E peptide did not present any response. A potential explanation of this fact could be the limitation of accessibility to the degron site caused by the random immobilization of the protein to the sensor chip. Moreover, solubility problems faced when diluting the peptide could also contribute to these results. Without positive control, we decided to perform an initial screening of all 41 VS

compounds using 3 different concentrations (100 µM, 10 µM, 1 µM). 28 compounds that presented dose-response were selected to be screened in a wider range of concentrations. At this point, buffer was optimized to improve saturation of the compounds and dose-response fitting. 0.5 mg/mL of dextran was added to the running buffer (Section 3.2.7.2.2), thereby discarding interactions with the dextran surface matrix. 9 compounds from VS presented a dose-response sensogram with K_ds in the one to three-digit micromolar range. This supposed a hit rate of 22%, which is considered a desirable success rate taking into account the expected for other structure-based approaches[164]. Additionally, *SAR by catalogue* was performed selecting 10 additional compounds that passed docking-based VS filers. These new compounds had 70% or 80% of similarity to SPR positive VS compounds. They were purchased and screened in a dose-response manner. All of them presented K_ds in the one to three-digit micromolar range, supposing a small increase in potency. In some cases, saturation was not achieved, forcing to calculate the affinity by fixing the R_{max}. This R_{max} consisted on the theoretical maximum response regarding the amount of ligand immobilized. A summary of the affinity data obtained of the 19 SPR binders is illustrated in Table 4.7. Besides, dose-response curves and sensograms of representative binders are depicted in Figure 4.15.

Table 4.7. Summary of SPR affinity data obtained for binders ordered by K_d value

Compound ID	K_d (µM)	Chi²	R_{max} observed	R_{max} expected
A9_MMC37	2	0.37	5.30	23
MMC40	2	1.41	4.73	21
A8_MMC37	23	0.49	6.20	23
A3_MMC21	41	0.47	13.13	21
A5_MMC17	45	0.53	Fixed	23
MMC21	51	1.88	Fixed	23
A6_MMC17	60	0.283	Fixed	23
MMC4	63	0.09	7.65	23
MMC17	71	3.33	Fixed	55
MMC2	76	0.83	16.00	23
A2_MMC11	127	0.28	7.35	23
MMC11	135	22.40	Fixed	23
MMC42	135	21.80	Fixed	23

Table 4.7 (cont.). Summary of SPR affinity data obtained for binders ordered by K_d value

Compound ID	K_d (µM)	Chi^2	R_{max} observed	R_{max} expected
MMC35	136	1.29	17.84	23
A1_MMC11	139	0.59	Fixed	23
A4_MMC17	140	1.15	104.60	55
MMC37	234	0.33	16.79	23
A7_MMC2	280	0.23	Fixed	23
A10_MMC40	353	0.08	18.35	23

Abbreviations: Chi^2, chi square; K_d, dissociation constant; R_{max}, maximum response.

A. DOSE-RESPONSE CURVE: A9_MMC37
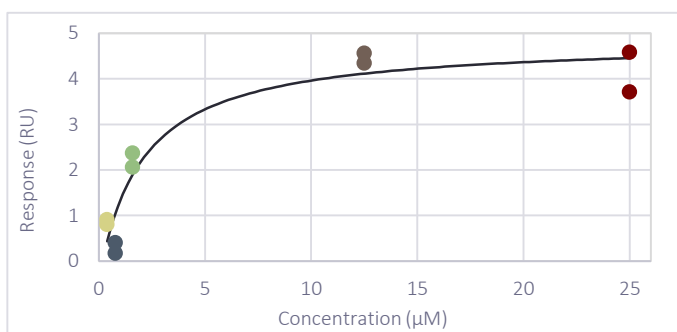
SENSOGRAM: A9_MMC37
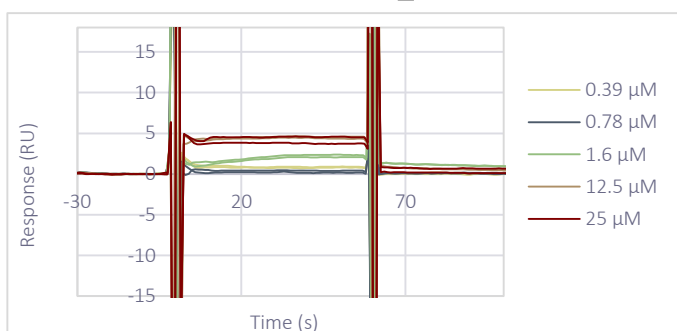
B. DOSE-RESPONSE CURVE: MMC40
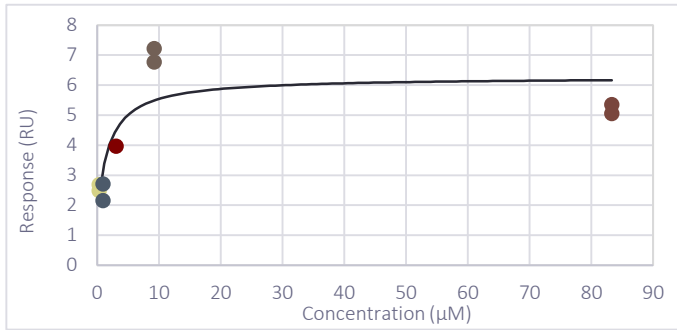
SENSOGRAM: MMC40
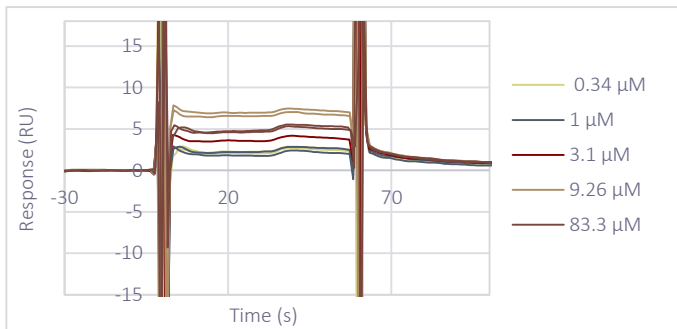
C. DOSE-RESPONSE CURVE: A8_MMC37
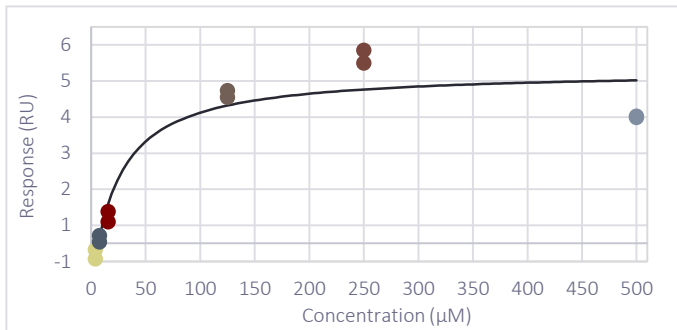
SENSOGRAM: A8_MMC37
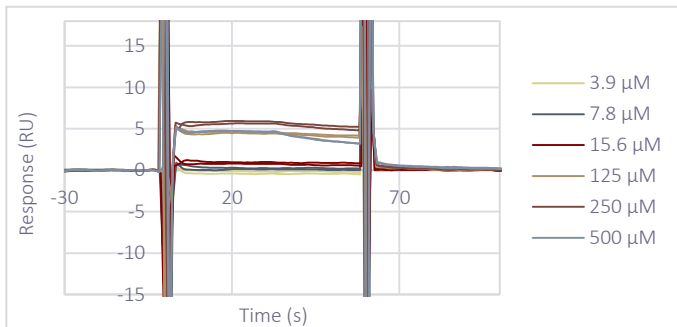
D. DOSE-RESPONSE CURVE: A3_MMC21
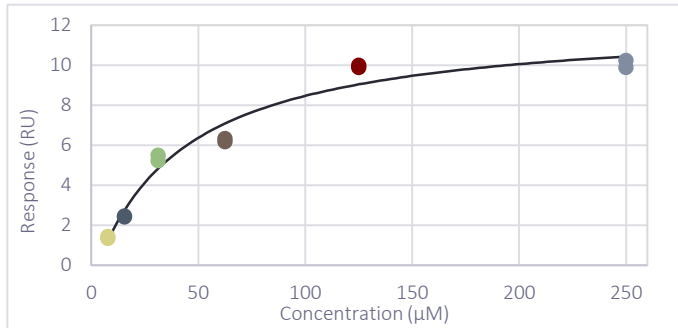
SENSOGRAM: A3_MMC21
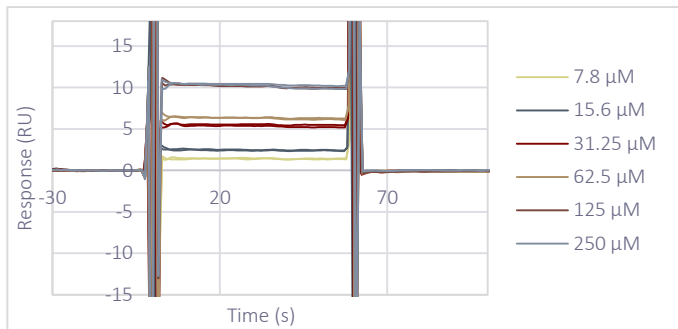
E. DOSE-RESPONSE CURVE: A5_MMC17
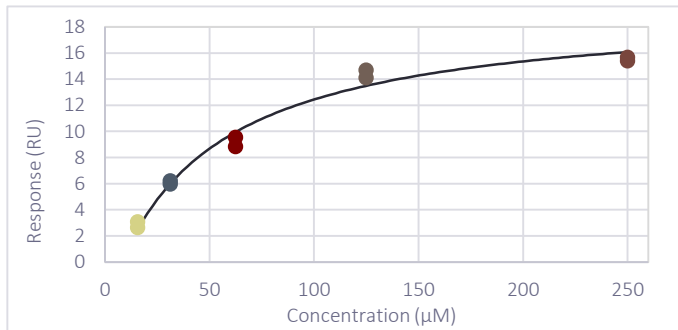
SENSOGRAM: A5_MMC17
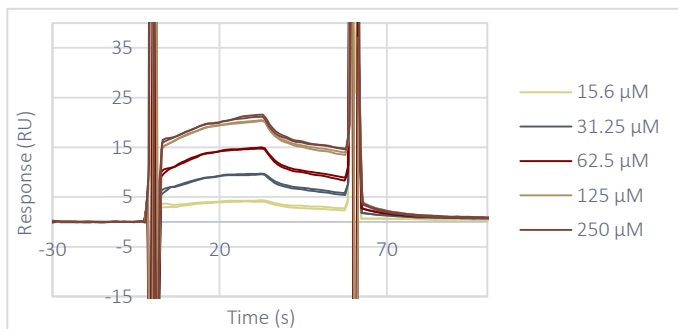
F. DOSE-RESPONSE CURVE: MMC21
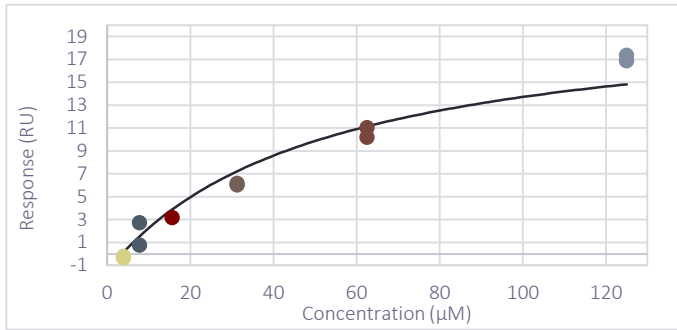
SENSOGRAM: MMC21
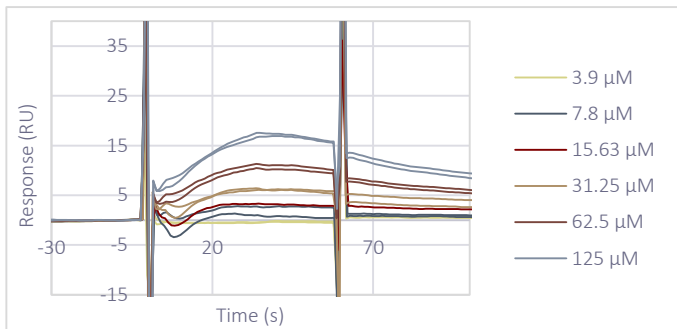
G. DOSE-RESPONSE CURVE: A6_MMC17
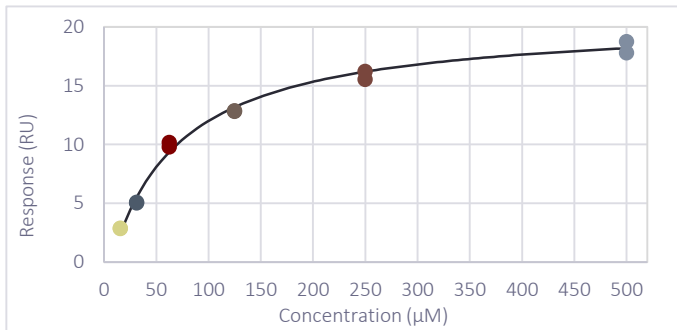
SENSOGRAM: A6_MMC17
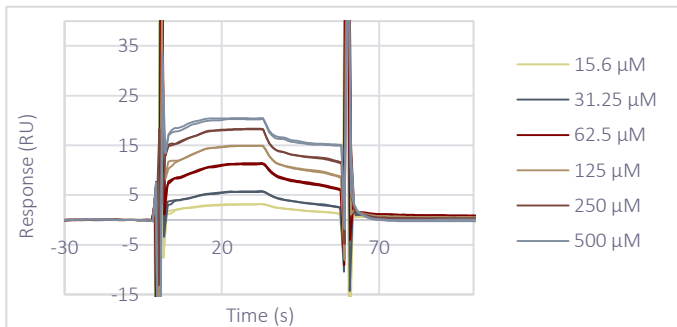
H. DOSE-RESPONSE CURVE: MMC4
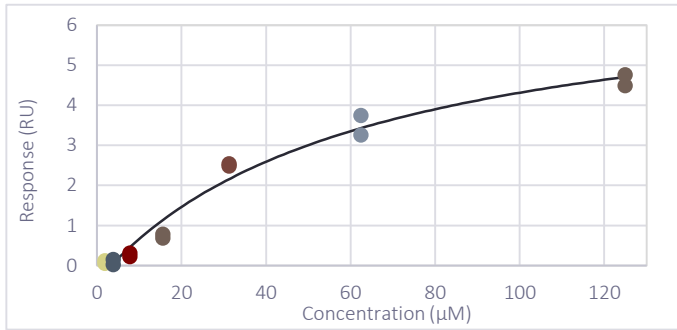
SENSOGRAM: MMC4
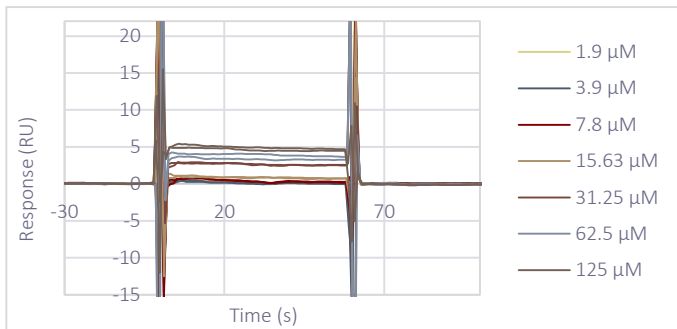
I. DOSE-RESPONSE CURVE: MMC17

SENSOGRAM: MMC17

J. DOSE-RESPONSE CURVE: A4_MMC17
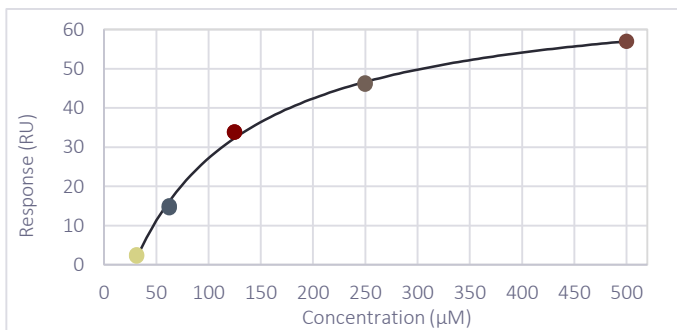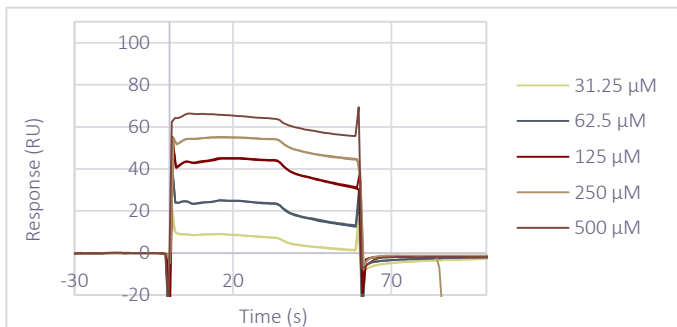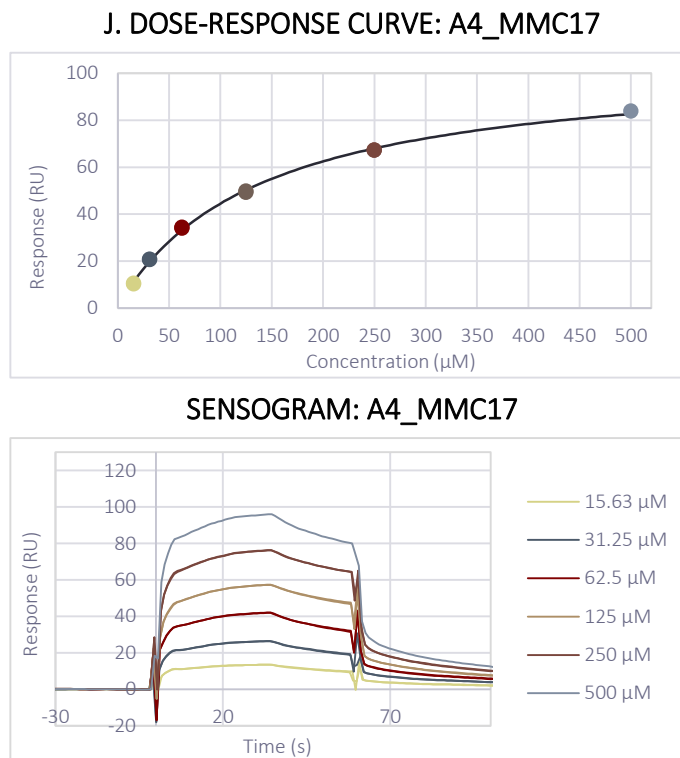
SENSOGRAM: A4_MMC17

Figure 4.15. Dose-response curves and sensograms of the representative FBW7-SKP1 binders: A9_MMC37 (A), MMC40 (B), A8_MMC37 (C), A3_MMC21 (D), A5_MMC17 (E), MMC21 (F), A6_MMC17 (G), MMC4 (H), MMC17 (I) and A4_MMC17 (J). In order to properly observe the sensogram, sensogram axis have been adjusted.

On one hand, A9_MMC37 and MMC40 were the compounds that presented lower K_d values. However, R_{max} obtained was really low, compared to other compounds that were able to reach the theoretical R_{max} (R_{max} expected). For that reason, these were not considered priority compounds. On the other hand, despite presenting lower affinities according to SPR, 5 compounds (MMC21, MMC17 and MMC17 analogues: A4_MMC17, A5_MMC17 and A6_MMC17) showed low association and dissociation behaviour, reaching theoretical R_{max} (Figure 4.15). To ensure that low kinetics were not dependent on flow rate and due to mass transport, A5_MMC17 was screened in a dose-response manner increasing the flow rate to 90 µL/min, while maintaining all the other parameters (Section 3.2.7.5). The same behaviour was observed in both flow rates, meaning that mass transport did not interfere in observed results. Association and dissociation constants were evaluated

for the 5 compounds with slow kinetics. However, proper data fitting could not be obtained. For that reason, K_d obtained considering steady state values was considered.

4.6. Elucidating the binding site of virtual screening binders

4.6.1. X-Ray crystallography of FBW7-SKP1 best binders

Determining the binding site of FBW7-SKP1 binders would be of utmost importance in order to validate our strategy. X-ray crystallography was the most direct approach to elucidate that. Therefore, we launched a collaboration with Dr Bing Hao (UconnHealth, USA), who had already been able to crystallize FBW7-SKP1[165,98]. Co-crystallization and soaking experiments were performed for compounds MMC2, MMC17 and MMC21. Even though no crystals were obtained with MMC21, 4 datasets could be collected for MMC2 and MMC17 crystals. In particular, one of them came from MMC2 co-crystal, and the others from FBW7-SKP1 soaked with MMC17 (2 datasets) and with MMC2 (1 dataset). Besides, datasets of crystals without compound (native crystals) were collected as negative controls. Unfortunately, no density relative to the compounds was found in the collected structures. In all crystals, even in native crystals, a density that seemed to correspond to a sulphate ion was observed in the virtual screening binding site (Figure 4.16 A). This could be due to the crystallization conditions, since it had Li_2SO_4. This density could be a hotspot used in ligand design. Moreover, when compared to FBW7-SKP1 published structure (PDB code 2OVP[165]), the binding site pocket is a bit opened, being similar to the pocket used for virtual screening (Figure 4.16 B). This confirms that this pocket is able to open in order to fit a compound. While no results were overcoming, indirect approaches as competitive assays or site-directed mutagenesis were attempted.
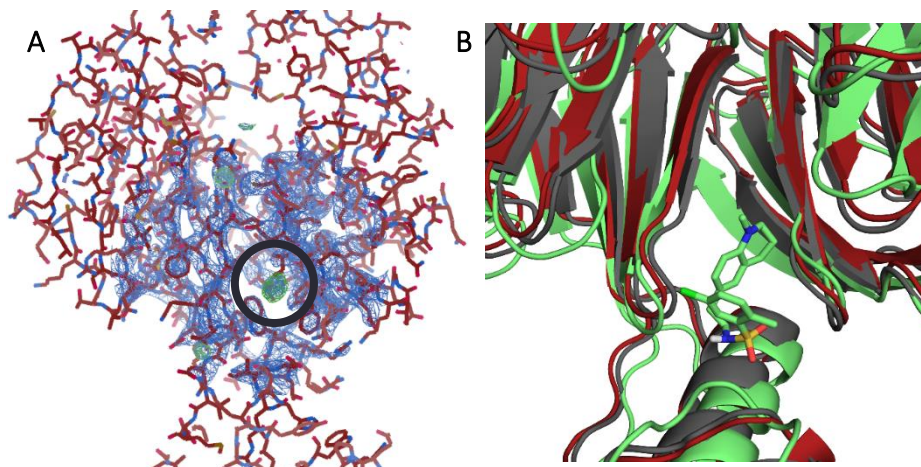
Figure 4.16. FBW7-SKP1 structure with MMC17. A. Density map of FBW7-SKP1 soaked with MMC17. Non-described density that seem to correspond to a sulphate ion is highlighted with a circle and can be observed in green, while density corresponding to FBW7 is seen in blue. **B.** Virtual screening pocket in FBW7 structure used for virtual screening with MMC17 (green), FBW7 in 2OVP structure[165] (grey) and FBW7 structure soaked with MMC17 (red).

4.6.2. Study of competition among FBW7-SKP1 binders performing surface plasmon resonance

SPR competitive assay was set up to determine if the compounds interacted in the same binding site by following a protocol already published[96]. Fixed concentrations of compounds at saturation conditions were tested in triplicates to establish a SPR response and a standard deviation for compound. Mixtures of two compounds at the same concentration were also screened (in triplicates). The strategy envisaged that if those compounds would interact in the same binding site, the compounds would be displacing each other and the response would be similar than the compounds on their own. However, if they would bind to different sites, their response would be similar to the theoretical sum of the responses given by compounds tested alone (Figure 4.17). To determine competition, we used equations 3.1 and 3.2 described in Section 3.2.7.2.3. Both equations presented different manners to assess if compounds would interact between each other. Concordance in both equations allowed us to determine if compounds would compete or not for the binding site. 17 of the 19 positive compounds from SPR were tested, since we had 4 analogues (MMC17, A4_MMC17, A5_MMC17 and A6_MMC17) and we just tested 2 of them as competitive controls

(A5_MMC17 and A6_MMC17). 4 representative compounds were selected to be prepared in mixtures with some or all the rest of the compounds. Results of competitive SPR are illustrated in Figure 4.18 and a summary is listed in Table 4.8. Almost all compounds were observed to compete between them.
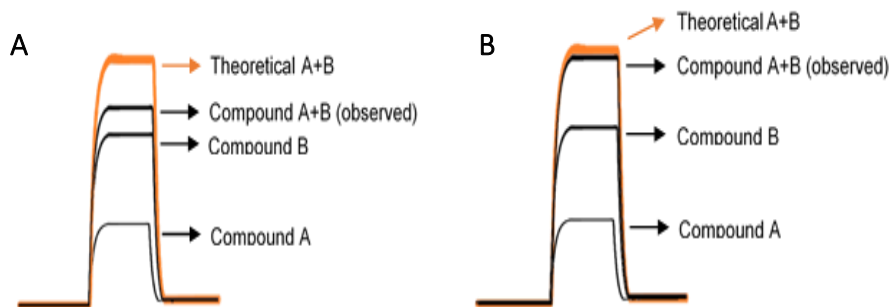
Figure 4.17. Schematic representation of competitive SPR. SPR sensograms when two compound compete **(A)** and when do not compete **(B).**
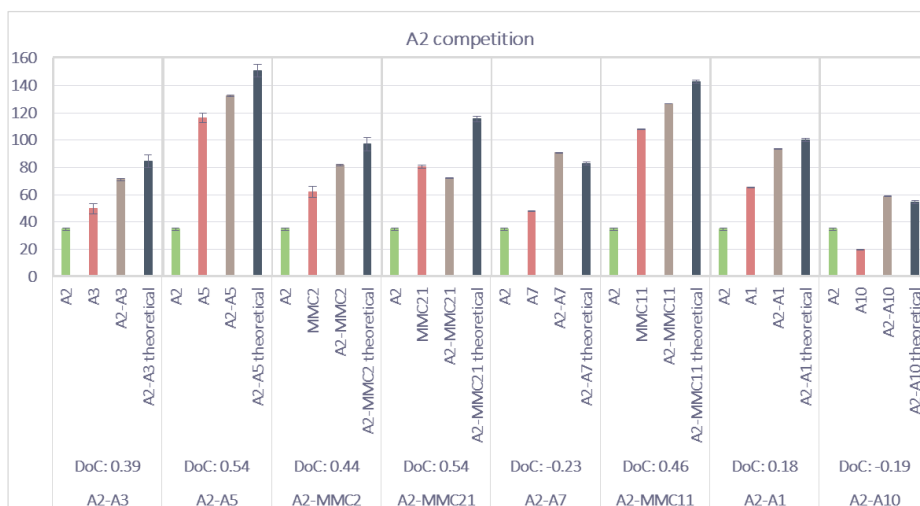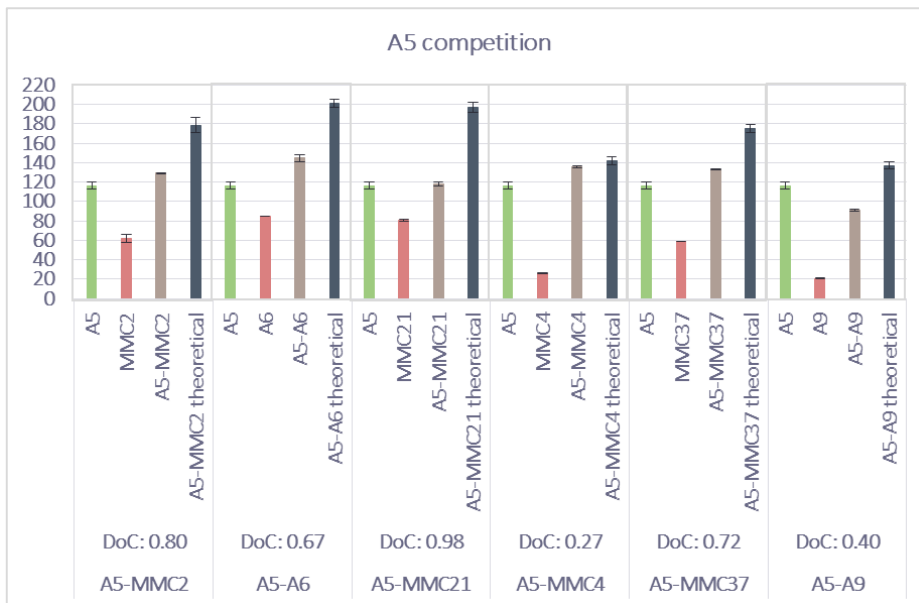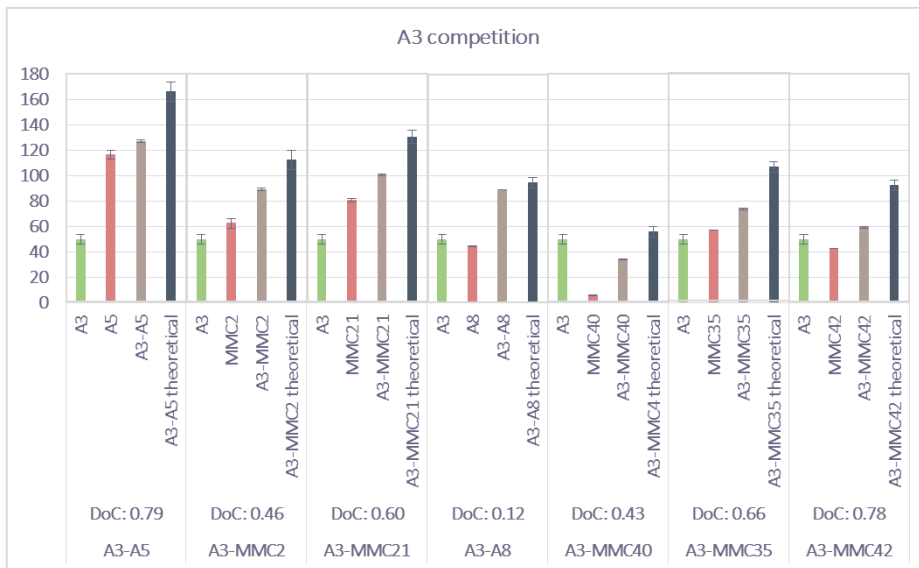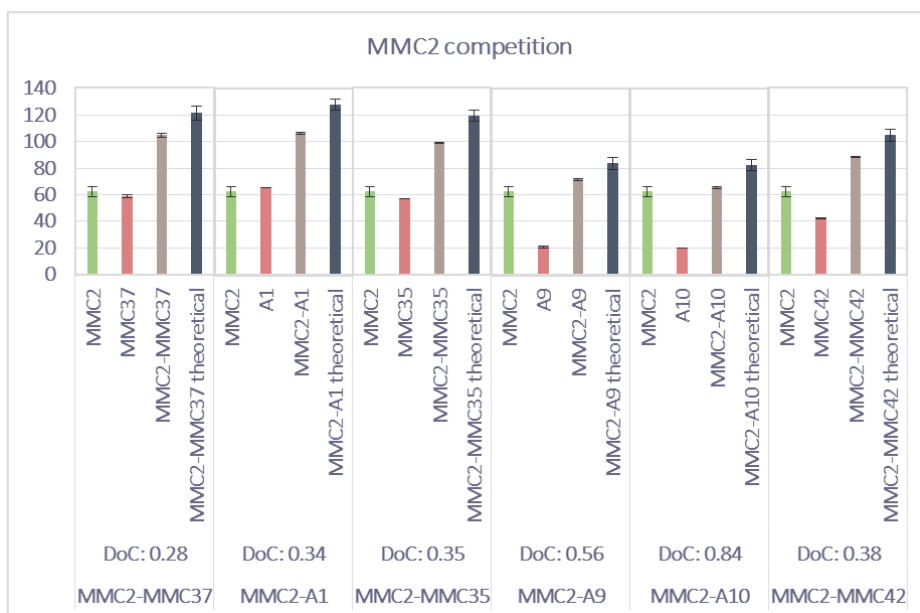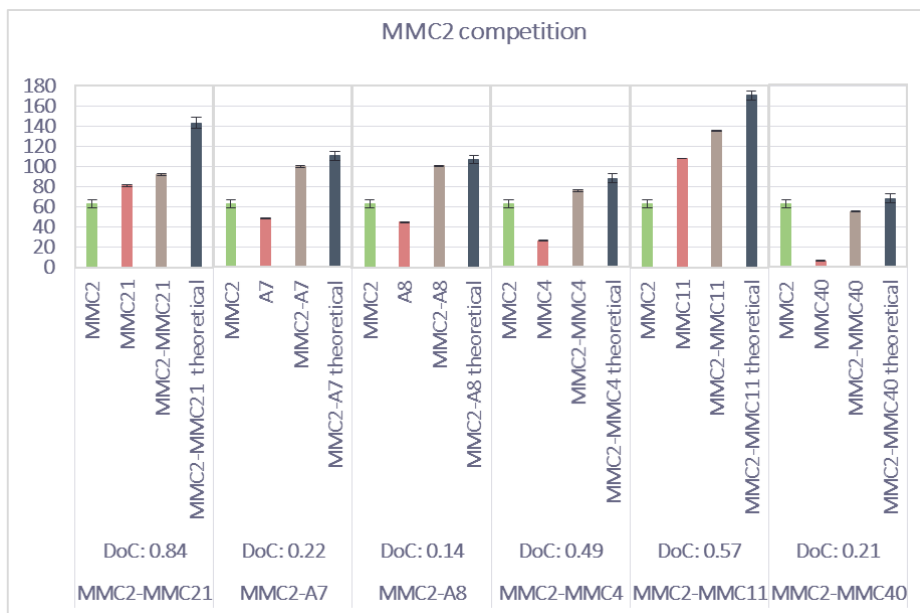
Figure 4.18. Bar chart of competition results. Comparison of the response given by the compounds alone or in mixture and the theoretical response given in case they would not compete (addition of compounds response alone). Besides, the degree of competition (DoC) calculated as equation 3.1 is also shown in each mixture.

Table 4.8. Competition table summarizing results from competitive SPR assay

COMPETES	A2_MMC11	A3_MMC21	A5_MMC17	MMC2
A2_MMC11		YES	YES	YES
A3_MMC21	YES		YES	YES
A5_MMC17	YES	YES		YES
A6_MMC17			YES	
MMC2	YES	YES	YES	
MMC21	YES	YES	YES	YES
A7_MMC2	NO			N.C.
A8_MMC37		N.C.		N.C.
MMC4			YES	YES
MMC11	YES			YES
MMC40		YES		YES
MMC37			YES	YES
A1_MMC11	YES			YES
MMC35		YES		YES
A9_MMC37			YES	YES
A10_MMC40	N.C			N.C.
MMC42		YES		YES

Abbreviations: N.C., not conclusive.

4.6.3. Binding site characterization by surface plasmon resonance: FBW7-SKP1 vs. single-point mutants

Despite competitive SPR results were promising, binding to the predicted pocket still needed to be validated. For that, site directed mutagenesis was carried out. Several single-point mutations of amino acids in the pocket were performed and the following proteins were produced: FBW7$_{N635A}$-SKP1, FBW7$_{N635I}$-SKP1, FBW7$_{A677I}$-SKP1, FBW7$_{A677F}$-SKP1 and FBW7$_{N679W}$-SKP1 (Section 4.5.3). Binding of best compounds was assessed by SPR by comparing WT and mutant forms. SPR experiments with mutants were performed with Roger Castaño, a master student that I supervised.

Initially, SPR experiments were performed with FBW7-SKP1, FBW7$_{N635A}$-SKP1 and FBW7$_{N635I}$-SKP1 (FBW7-SKP1 and mutants – SPR experiment 1). In this case, the amino acid mutated was in both cases ASN635, which had previously been used in DUck filter (Section 4.5.2). No differences were observed in compounds affinity between WT and mutants. A plausible explanation of this fact is that very close to ASN635, FBW7 has

ASN633 that could also perform the predicted hydrogen bond. Results from FBW7-SKP1 and mutants – SPR experiment 1 are listed in Table 4.9.

Table 4.9. Summary of SPR affinity data obtained in FBW7-SKP1 and mutants – SPR experiment 1

Compound ID	Protein	K_d (µM)	Chi²	R_{max} observed	R_{max} expected
A3_MMC21	FBW7-SKP1	58	1.62	19.7	48
	FBW7$_{N635A}$-SKP1	61	3.67	33.5	45
	FBW7$_{N635I}$-SKP1	60	3.19	22.7	45
A5_MMC17	FBW7-SKP1	61	13.3	Fixed	48
	FBW7$_{N635A}$-SKP1	44	216	Fixed	45
	FBW7$_{N635I}$-SKP1	46	47	Fixed	45
MMC21	FBW7-SKP1	68	8.36	Fixed	48
	FBW7$_{N635A}$-SKP1	37	78.2	Fixed	45
	FBW7$_{N635I}$-SKP1	52	9.12	Fixed	45
A6_MMC17	FBW7-SKP1	95	15.7	Fixed	48
	FBW7$_{N635A}$-SKP1	56	331	Fixed	45
	FBW7$_{N635I}$-SKP1	65	39.3	Fixed	45
MMC2	FBW7-SKP1	102	3.49	47.5	48
	FBW7$_{N635A}$-SKP1	104	8.97	73.3	45
	FBW7$_{N635I}$-SKP1	97	5.45	55.8	45
MMC42	FBW7-SKP1	81	111	Fixed	48
	FBW7$_{N635A}$-SKP1	69	289	Fixed	45
	FBW7$_{N635I}$-SKP1	80	126	Fixed	45
A1_MMC11	FBW7-SKP1	185	3.6	Fixed	48
	FBW7$_{N635A}$-SKP1	75	102	Fixed	45
	FBW7$_{N635I}$-SKP1	133	11.4	Fixed	45

Abbreviations: Chi2, chi square; K_d, dissociation constant; Rmax, maximum response.

Other single point mutants, FBW7$_{A677I}$-SKP1 and FBW7$_{N679W}$-SKP1 were tested in FBW7-SKP1 and mutants – SPR experiment 2. These mutations would suppose a smaller cavity in order to block, totally or partially, the entrance of binders. Results from FBW7-SKP1 and mutants – SPR experiment 2 are listed in Table 4.10. 3 compounds did not present differences in K_d between WT and mutants, suggesting they did not interact in the FBW7 predicted binding site. Nonetheless, the other 4 compounds screened appeared to be 3 to 6 fold more potent for FBW7-

SKP1 WT than for FBW7$_{A677I}$-SKP1 of FBW7$_{N679W}$-SKP1. In particular, A5_MMC17 and A6_MMC17 were both 4 and 6-fold more potent for FBW7-SKP1 WT than for FBW7$_{A677I}$-SKP1 of FBW7$_{N679W}$-SKP1, respectively. This can be explained for the fact they have a very similar structure (Figure 4.11).

Table 4.10. Summary of SPR affinity data obtained in FBW7-SKP1 and mutants – SPR experiment 2

Compound ID	Protein	K_d (µM)	Chi²	R_{max} observed	R_{max} expected
A3_MMC21	FBW7-SKP1	19	11.7	13.7	45
	FBW7$_{A677I}$-SKP1	19	9.31	13	45
	FBW7$_{N679W}$-SKP1	18.4	27.9	23.7	64
A5_MMC17	FBW7-SKP1	48	22.2	Fixed	45
	FBW7$_{A677I}$-SKP1	202	2.21	Fixed	45
	FBW7$_{N679W}$-SKP1	292	1.84	Fixed	64
MMC21	FBW7-SKP1	54	7.12	Fixed	45
	FBW7$_{A677I}$-SKP1	168	1.22	Fixed	45
	FBW7$_{N679W}$-SKP1	183	2.26	Fixed	64
A6_MMC17	FBW7-SKP1	62	6.56	Fixed	45
	FBW7$_{A677I}$-SKP1	268	1.49	Fixed	45
	FBW7$_{N679W}$-SKP1	367	524	Fixed	64
MMC2	FBW7-SKP1	54	1.72	25.9	45
	FBW7$_{A677I}$-SKP1	39	2.52	25.8	45
	FBW7$_{N679W}$-SKP1	40	6.07	43.1	64
MMC42	FBW7-SKP1	161	9.14	Fixed	45
	FBW7$_{A677I}$-SKP1	488	1.78	Fixed	45
	FBW7$_{N679W}$-SKP1	472	3.47	Fixed	64
A1_MMC11	FBW7-SKP1	147	19.2	Fixed	45
	FBW7$_{A677I}$-SKP1	239	5.05	239	45
	FBW7$_{N679W}$-SKP1	202	15.1	Fixed	64

Abbreviations: Chi2, chi square; K_d, dissociation constant; Rmax, maximum response.

To disrupt even more the interaction with FBW7-SKP1 binders while making the cavity smaller, FBW7$_{A677F}$-SKP1 was tested in SPR. In FBW7-SKP1 and mutants – SPR experiment 3, the mutants screened were FBW7$_{A677F}$-SKP1 and FBW7$_{A677I}$-SKP1. Results from FBW7-SKP1 and mutants – SPR experiment 2 are listed in Table 4.11. As in previous

experiments, no differences in K_d between WT and mutants were observed in A3_MMC21. Interestingly, K_d differences were more accentuated in FBW7$_{A677I}$-SKP1 than in FBW7$_{A677F}$-SKP1. We hypothesised that aromatics groups of the compounds could interact with phenyl ring of phenylalanine. Overall, FBW7$_{A677I}$-SKP1 was selected to be used in further experiments to validate the differences in K_d between WT and mutants, and to subsequently validate our predicted binding site. Dose-response curves of 4 representative compounds interacting with FBW7-SKP1 and all mutants are depicted in Figure 4.19.

Table 4.11. Summary of SPR affinity data obtained in FBW7-SKP1 and mutants – SPR experiment 3

Compound ID	Protein	K_d (µM)	Chi²	R_{max} observed	R_{max} expected
A3_MMC21	FBW7-SKP1	25	2.55	20.08	45
	FBW7$_{A677I}$-SKP1	22	1.88	8.812	55
	FBW7$_{A677F}$-SKP1	22	1.07	6.216	45
A5_MMC17	FBW7-SKP1	59	4.71	Fixed	45
	FBW7$_{A677I}$-SKP1	175	5.36	Fixed	55
	FBW7$_{A677F}$-SKP1	118	3.79	Fixed	45
MMC21	FBW7-SKP1	88	5.24	Fixed	45
	FBW7$_{A677I}$-SKP1	181	2.58	Fixed	55
	FBW7$_{A677F}$-SKP1	207	1.83	Fixed	45
A6_MMC17	FBW7-SKP1	84	4.05	Fixed	45
	FBW7$_{A677I}$-SKP1	252	1.15	Fixed	55
	FBW7$_{A677F}$-SKP1	155	2.69	Fixed	45
MMC42	FBW7-SKP1	200	25.6	Fixed	45
	FBW7$_{A677I}$-SKP1	263	17.3	Fixed	55
	FBW7$_{A677F}$-SKP1	236	13.8	Fixed	45

Abbreviations: Chi2, chi square; K_d, dissociation constant; Rmax, maximum response.

A. A3_MMC21

FBW7-SKP1 + A3_MMC21
(Experiment 3)

FBW7$_{N635A}$-SKP1 + A3_MMC21
(Experiment 1)

FBW7$_{N635I}$-SKP1 + A3_MMC21
(Experiment 1)

FBW7$_{A677I}$-SKP1 + A3_MMC17
(Experiment 3)

FBW7$_{A677F}$-SKP1 + A3_MMC21
(Experiment 3)

FBW7$_{N679W}$-SKP1 + A3_MMC21
(Experiment 2)
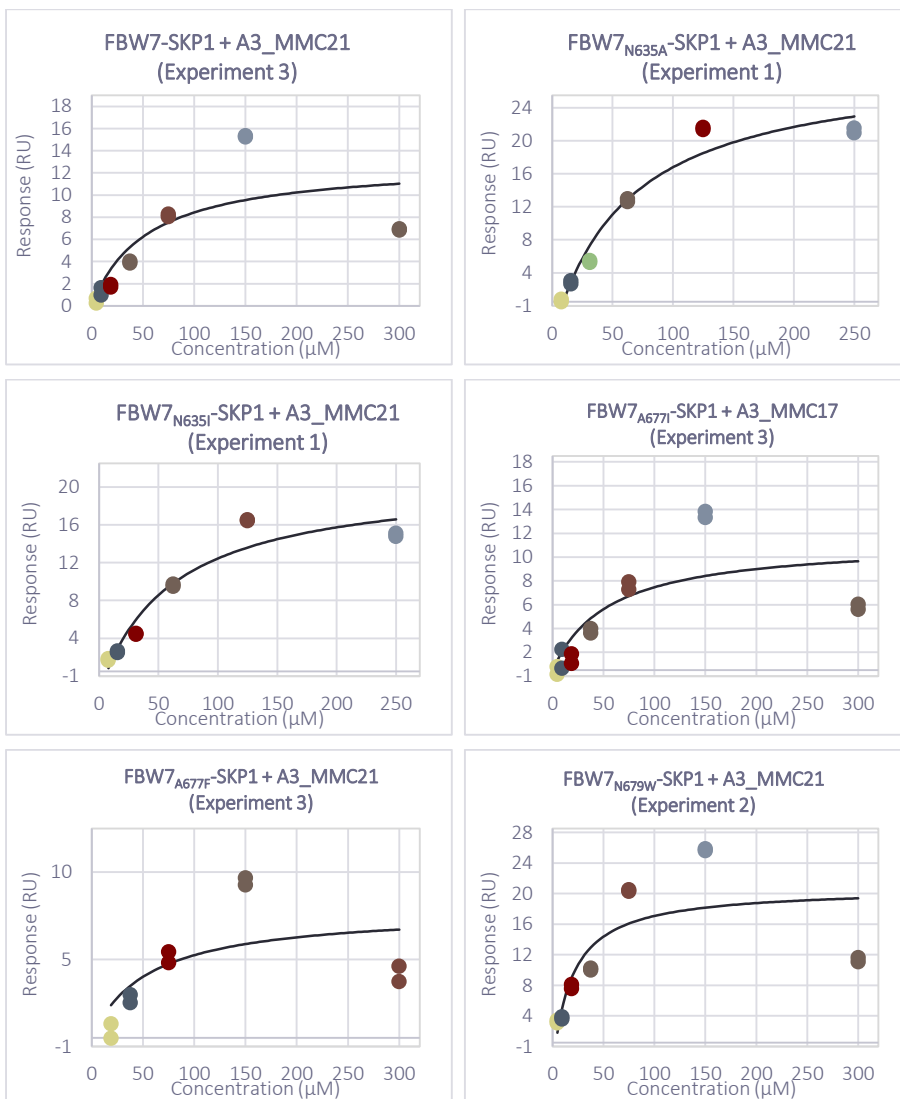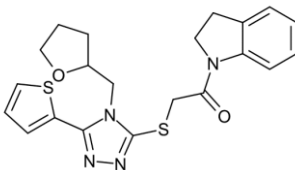
B. A5_MMC17
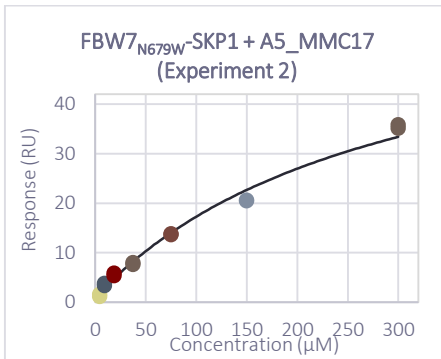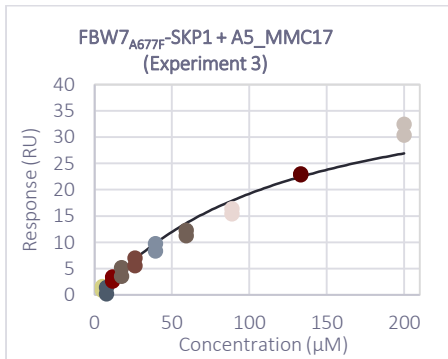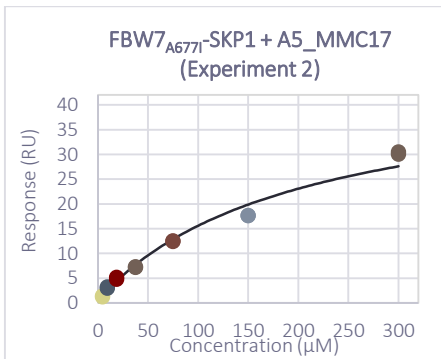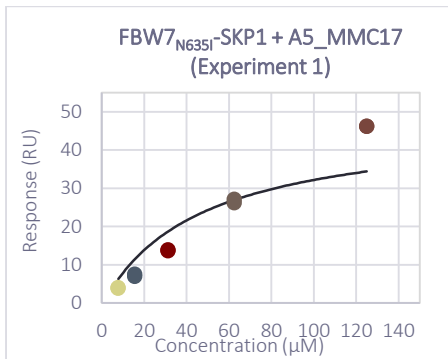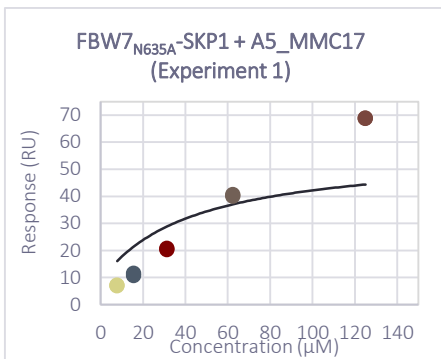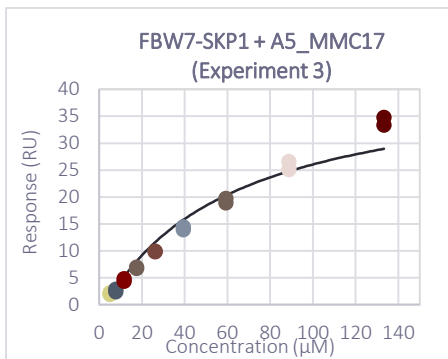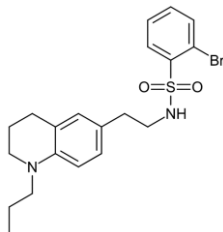
C. MMC21

D. A6_MMC17

Figure 4.19. SPR curves WT versus mutants. Structure and dose response curves of representative compounds interacting with FBW7-SKP1, FBW7$_{N635A}$-SKP1, FBW7$_{N635I}$-SKP1, FBW7$_{A677I}$-SKP1, FBW7$_{A677F}$-SKP1 and FBW7$_{N679W}$-SKP1. **A.** A3_MMC21. **B.** A5_MMC17. **C.** MMC21. **D.** A6_MMC17.

4.6.4. Binding site characterization by microscale thermophoresis: FBW7-SKP1 vs. FBW7$_{A677I}$-SKP1

Since the differences in affinity were not as disruptive as expected and in order to validate our previous results, MST was carried out. Prior to experiments FBW7-SKP1 and FBW7$_{A677I}$-SKP1 were labelled. Protein labelling was performed using Monolith Protein Labelling Kit RED-NHS 2nd Generation (Nanotemper) as described in Section 3.2.7.3. Final protein concentration (equation 3.3) and the degree of labelling (DOL; equation 3.4) were calculated for each protein, being the final concentration in both cases of 1.4 µM. Several attempts of labelling needed to be performed in order to achieve an appropriate DOL. In some cases DOL appeared to be higher than 1, which could lead to adverse effects on protein function. To solve that, the excess of dye added when labelling was decreased to 2.5 for FBW7-SKP1 and to 1 for FBW7$_{A677I}$-SKP1, being 3-fold excess the one recommended by the manufacturer. Finally, DOLs of 0.92 and 0.73 were obtained for FBW7-SKP1 and FBW7$_{A677I}$-SKP1, respectively.

After labelling, MST experiment was set up in order to select the concentration of the labelled protein, the LED power and optimal buffer. After optimization, MST experiments were performed in a buffer that consisted of 50 mM Hepes (pH 8.0), 50 mM NaCl, 0.05% Tween-20, 1 mM DTT. Labelled-protein was prepared at a final concentration of 50 nM and LED power was set to 40% or 60%. Affinity for FBW7-SKP1 and FBW7$_{A677I}$-SKP1 was determined for four of the best compounds: A5_MMC17, A6_MMC17, MMC21, and A3_MMC21. MST experiments were performed in triplicates, except for A3_MMC21 that was studied in duplicates. In all cases, changes on initial fluorescence were observed, making binding affinity to be assessed from these values instead of MST signal. However, a part from quenching from the fluorescence upon binding, changes in fluorescence could also be derived from nonspecific adsorption to capillaries or aggregation of the fluorescent molecule upon addition of the ligand. SD-test was performed to confirm that fluorescence changes were caused by compound interaction to the protein. It was based on comparing differences on fluorescence emission of the three highest and the three lowest concentrations of compounds in mixture with the protein. In the case of fluorescence caused by

compound interaction, differences in fluorescence should be observed between high and low concentrations with the folded protein, whereas when the protein was denatured they should present the same fluorescence, since compound binding was not causing any effect. SD-test procedure is described on Section 3.2.7.3. In all cases, SD-test proved that the change of fluorescence was caused by compound interaction. Each compound was screened at different concentrations depending on their solubility and K_D. A5_MMC17 was screened at 15 different concentrations ranging from 1 mM to 3.4 µM (2:1) dilutions. A6_MMC17 was screened at 15 different concentrations ranging from 444 µM to 1.5 µM (2:1) dilutions. MMC21 was screened at 16 different concentrations ranging from 800 µM to 1.8 µM (2:1) dilutions. A3_MMC21 was screened at 15 different concentrations ranging from 500 µM to 1.7 µM (2:1) dilutions. The K_D obtained with FBW7-SKP1 WT was in the same range as in SPR experiments. Differences in affinity between FBW7-SKP1 and FBW7$_{A677I}$-SKP1 were observed in A5_MMC17, A6_MMC17 and A3_MMC21. Surprisingly, no differences in affinity were observed in the case of MMC21. Summary of all results obtained is depicted in Figure 4.20.

A.   **FBW7-SKP1 with A5_MMC17**   **FBW7$_{A677I}$-SKP1 with A5_MMC17**
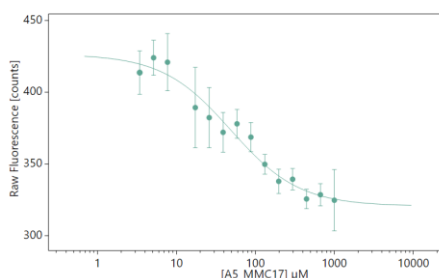
K_d: 49 ± 13 µM
Standard deviation: 7.51
Reduced Chi^2: 0.38
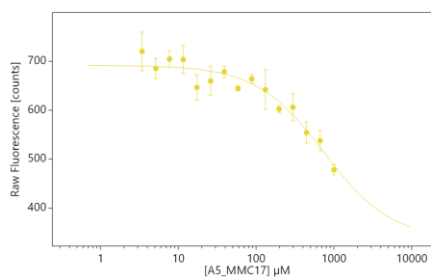
K_d: 745 ± 401 µM
Standard deviation: 19.70
Reduced Chi^2: 2.46

Comparison A5_MMC17 binding with FBW7-SKP1 (green) vs. FBW7$_{A677I}$-SKP1 (yellow)
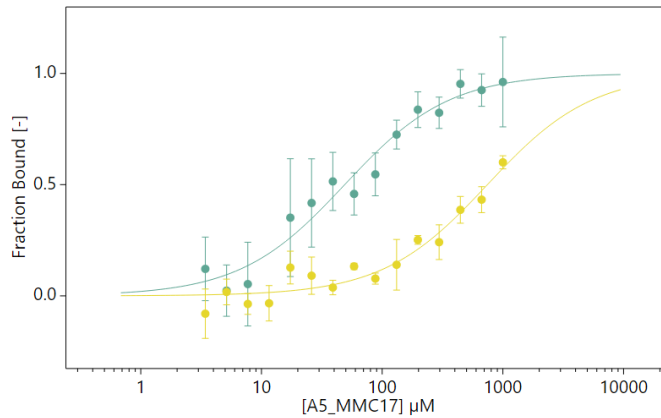
B. FBW7-SKP1 with A6_MMC17
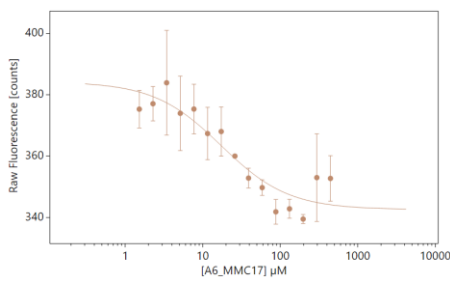
K_d: 17 ± 9 µM
Standard deviation: 5.69
Reduced Chi^2: 2.76

FBW7$_{A677I}$-SKP1 with A6_MMC17
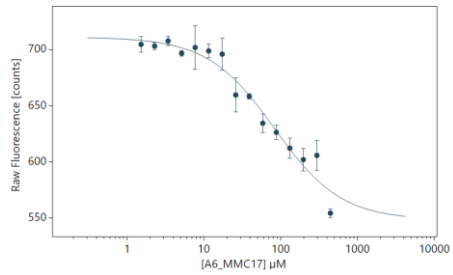
K_d: 83 ± 25 µM
Standard deviation: 11.00
Reduced Chi^2: 3.26

Comparison A6_MMC17 binding with FBW7-SKP1 (brown) vs. FBW7$_{A677I}$-SKP1 (blue)
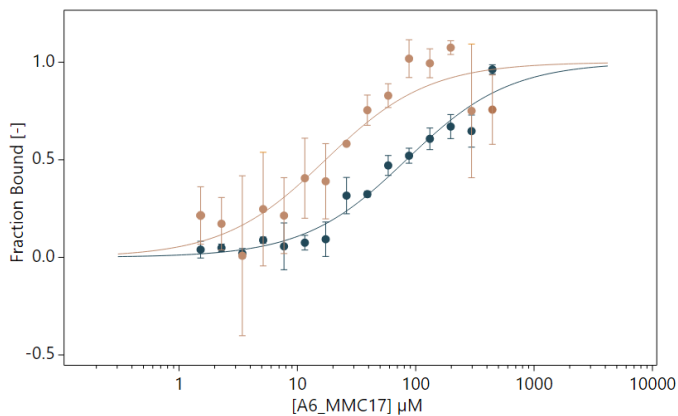
C. **FBW7-SKP1 with MMC21**

FBW7$_{A677I}$-SKP1 with MMC21
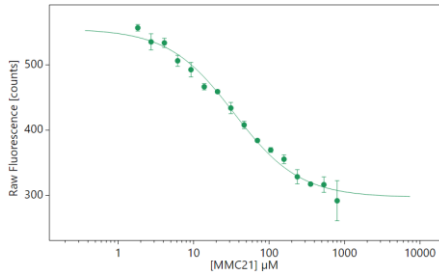
K_d: 34 ± 4 µM
Standard deviation: 9.35
Reduced Chi^2: 2.60
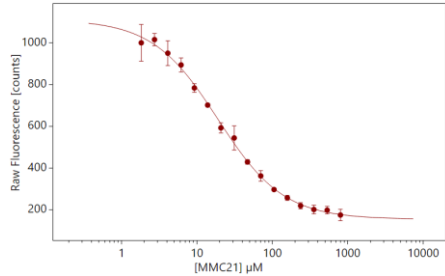
K_d: 19 ± 1 µM
Standard deviation: 15.86
Reduced Chi^2: 0.63

Comparison MMC21 binding with FBW7-SKP1 (green) vs. FBW7$_{A677I}$-SKP1 (red)
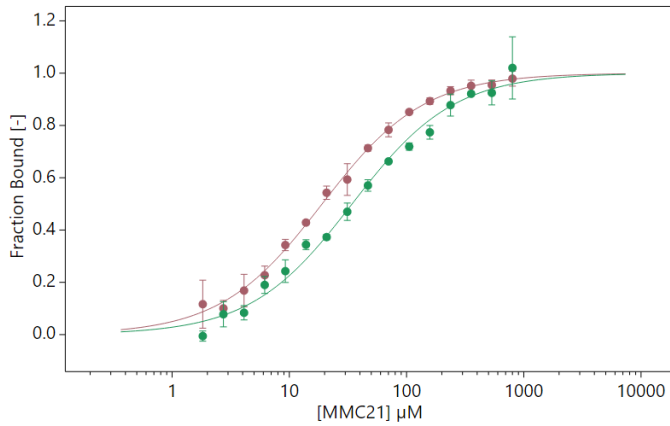
D. **FBW7-SKP1 with A3_MMC21**
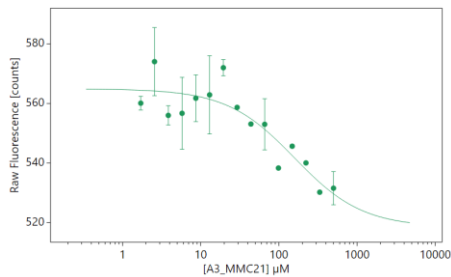
FBW7$_{A677I}$-SKP1 with A3_MMC21

K_d: 159 ± 119 µM
Standard deviation: 6.36
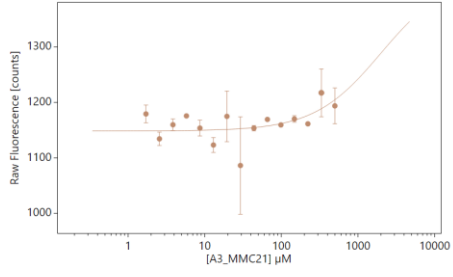
K_d: 2000 ± 15000 µM
Standard deviation: 9.35

Comparison A3_MMC21 binding with FBW7-SKP1 (green) vs. FBW7$_{A677I}$-SKP1 (brown)
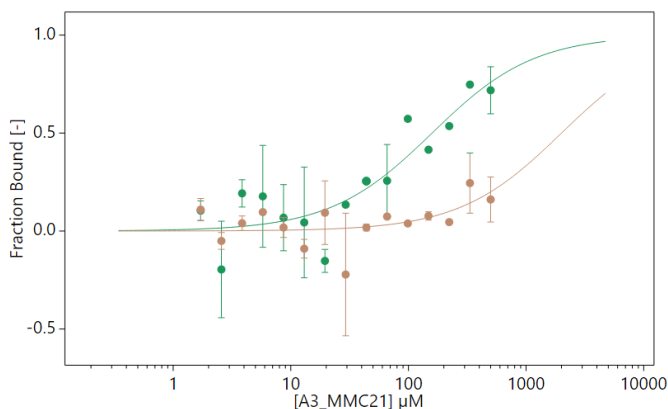
Figure 4.20. Results from MST experiments. MST curves of each compound screened with FBW7-SKP1 and FBW7$_{A677I}$-SKP1 and its comparison. **A.** A5_MMC17. **B.** A6_MMC17. **C.** MMC21. **D.** A3_MMC21.

4.6.5. Binding site characterization by isothermal titration calorimetry: FBW7-SKP1 vs. FBW7$_{A677I}$-SKP1

To finally confirm the differences observed between FBW7-SKP1 WT and FBW7$_{A677I}$-SKP1, ITC experiments were performed titrating the best compound (A5-MMC17) against both proteins. In both ITC experiments, the cell contained 30 µM of protein (FBW7-SKP1 or FBW7$_{A677I}$-SKP1) and series of a solution of A5-MMC17 at 600 µM were injected. All samples were prepared in a buffer that contained 50 mM Hepes (pH 8.0), 50 mM NaCl and the final DMSO concentration was 5%, as specified in Section 3.2.7.4. Differences in A5_MMC17 binding affinity were clearly observed between FBW7-SKP1 and FBW7$_{A677I}$-SKP1 and are illustrated in Figure 4.21 A. When binding to FBW7-SKP1, A5_MMC17 presented a ΔG of -5.85 kcal/mol and a K_d of 51 µM (Figure 4.21 B), being this last consistent with SPR and MST experiments. In contrast, K_d of A5_MMC17 interacting with FBW7$_{A677I}$-SKP1 could not be assessed.

K_d of A5_MMC17 binding to FBW7-SKP1 was determined by fixing the stoichiometry, since a proper sigmoidal shape could not be observed in ITC titration. This fact is related with what is called the c value[78] (c value = n x [FBW7-SKP1] x K_d^{-1}, where n is stoichiometry and [FBW7-SKP1] the concentration of protein). A c value of minimum 10 is desired to obtain a sigmoidal shape[78] and in our case it was of 0.6. To increase the c value,

we should have increased the concentration of protein, which was challenging to achieve due to low yields obtained in protein purification. Overall, ITC is a label-free technique that allows the quantification of protein-ligand interaction. Thereby, the differences observed in these experiments between FBW7-SKP1 WT and its mutant are significant and support our hypothesis regarding compounds binding site.
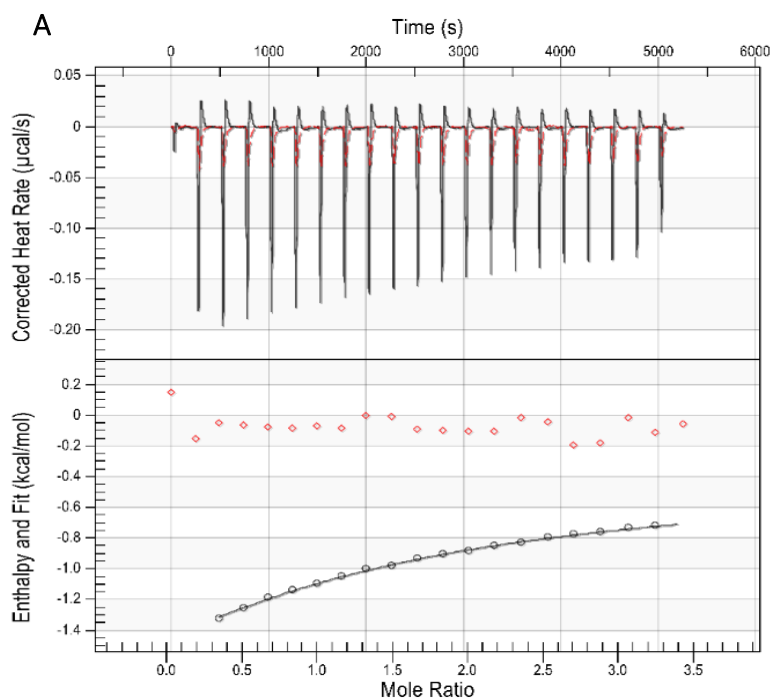
B. ITC results of A5_MMC17 binding to FBW7-SKP1

Model	Variable	Value	CI
Independent	K_d (µM)	50.900	6.007
	n	1	Fixed
	ΔH (kcal/mol)	-2.052	±0.140
	ΔS (cal/mol·K)	12.76	
	ΔG (kcal/mol)	-5.85	
Blank (linear)	Intercept (µcal)	-1	Fixed
	Slope	0.008	±0.001
	Confidence Level		90%

Figure 4.21. ITC results. A. Comparison between ITC experiments of A5-MMC17 interacting with FBW7-SKP1 (black) and FBW7$_{A677I}$-SKP1 (red). **B.** Thermodynamic profile of A5-MMC17 binding to FBW7-SKP1.

4.7. Elucidating the mechanism of action of virtual screening binders

4.7.1. Fluorescence polarization to determine the effect in FBW7 substrate recognition

In order to perform an initial evaluation of the mechanism of action of the positive molecules from the primary screening, a fluorescence polarization (FP) assay was carried out. This technique would allow us to determine the effect of our compounds in substrate recognition of FBW7-SKP1. As a fluorescent-labelled ligand, a labelled small peptide of DISC1, the endogenous substrate of FBW7[98], was used (FICT-DISC1).

FP assays were performed with Andrea Bertran, while supervising her master internship. Initially, FP assay had to be optimized, considering buffer conditions, incubation times and concentrations. For buffer optimization, we contemplated the use of PBS, Hepes, BSA, Tween-20, DTT and different pH. The final selected buffer consisted on 50 mM Hepes (pH 8.0), 50 mM NaCl, 0.01% Tween-20, 1 mM DTT. The last two conditions helped reducing aggregation and protein absorption onto the plate wells. The incubation time was established at 30 minutes as a compromise to let the components homogenize between them, but also to avoid fluorescence loss. Several concentrations of FICT-DISC1 were also tested ranging from 2 nM to 1000 nM, selecting 50 nM as the optimal one.

Initially, the K_d of the labelled peptide (FICT-DISC1) was extrapolated by screening FBW7-SKP1 at increasing concentrations ranging from 0.19 µM to 25 µM (2-fold dilutions), as it is shown in Figure 4.22. The resulting K_d obtained from different experiments performed at different days was between 1 µM and 4 µM, according to the one of the unlabelled peptide reported in the literature[98].

Compound competition with FICT-DISC1 was assessed by two different competition assays. FP competitive assay 1 consisted on following the effect of the compounds on the K_d of the FICT-DISC1 peptide. Initially, this competition was assessed with the unlabelled form of the peptide (DISC1). DISC1 was tested at 1 µM (in excess) to achieve a notable displacement of the labelled peptide. From this competitive assay, the "apparent" dissociation constant (K_d^{app}) of FICT-DISC1 was calculated,

which allowed us to determine the K_d of DISC1 using equation 3.7. Unlabelled DISC1 presented a K_d of 0.5 μM, being similar to the already described. Results of competition are depicted in Figure 4.23.
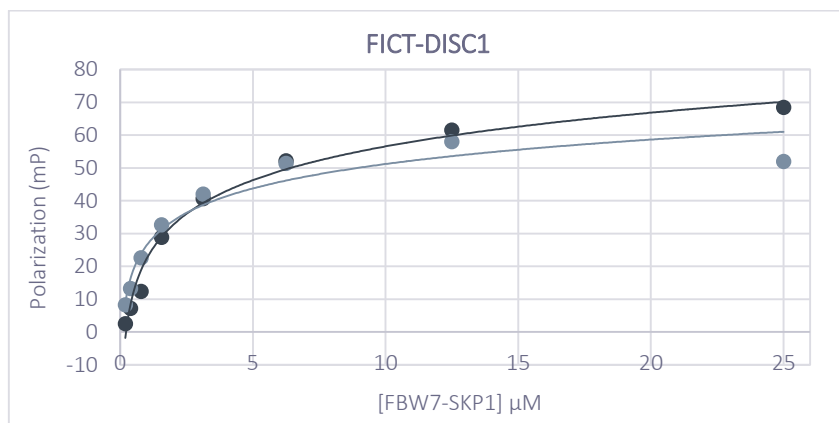
Figure 4.22. Curve replicates showing binding of the labelled peptide FICT-DISC1 to FBW7-SKP1.

Figure 4.23. Fluorescence polarization competitive assay 1 with DISC1. Curves of the average of the two replicates. In blue it is shown the curve of FICT-DISC1, whereas in green the competition of FICT-DISC1 and DISC1 can be observed.

Seven virtual screening compounds were tested by FP competitive assay 1: MMC2, MMC17, MMC21, MMC37, A2_MMC11, A5_MMC17 and A6_MMC17 at 100 μM. Slight differences were observed when comparing the K_d obtained by the control (FICT-DISC1 alone) in each experiment and the K_d^{app} observed in some competition assays as is listed in Table 4.12. Examples of curves obtained by this type of competition assay are shown in Figure 4.25.

141

Table 4.12. Results from fluorescence polarization competitive assay 1

Compound ID	K_d (µM) positive control	K_d^{app} (µM) in presence of compound
MMC2	2.5	1.26
MMC17	4.5	5.70
MMC21	2.5	0.28
MMC37	4.5	2.90
A2_MMC11	2.5	0.34
A5_MMC17	2.5	2.97
A6_MMC17	2.5	3.00

Table showing the comparison between the K_d^{app} of FICT-DISC1 in competition assays and the K_d of the positive control (FICT-DISC1 alone) in the same day compound was tested.

FP competitive assay 2 consisted on the direct observation of the effect of FICT-DISC1 binding when increasing compound concentration. Besides, these experiments allowed the elucidation of the significance of the slight differences in the K_d^{app} observed. For these experiments, a part of fixing FICT-DISC1 concentration (50 nM), FBW7-SKP1 concentration was also fixed. Calculated Z' factor (Equation 3.6) of dose-response of FBW7-SKP1 at 50 nM of FICT-DISC1 (Figure 4.22) allowed the calculation of the optimal FBW7-SKP1 concentration, which was 10 µM. Besides, to avoid compound aggregation and precipitation at the highest concentration (800 µM), 8% of DMSO was used. DISC1 unlabelled peptide was also used as positive control at concentrations ranging from 0.02 µM to 3 µM also at 8% DMSO. FP competitive assay 2 with DISC1 showed a clear decrease on FICT-DISC1 binding (Figure 4.24). These results also corroborated that the increase in DMSO does not affect either the stability of the protein or the binding of DISC1. A part from the 7 compounds tested with FP competitive assay 1, 12 additional compounds were tested. Therefore, all 19 virtual screening positive compounds were tested by FP competitive assay 2: MMC2, MMC4, MMC11, MMC17, MMC21, MMC35, MMC37, MMC40, MMC42, A1_MMC11, A2_MMC11, A3_MMC21, A4_MMC17, A5_MMC17, A6_MMC17, A7_MMC2, A8_MMC37, A9_MMC37 and A10_MMC40. None of those compounds seemed to affect FICT-DISC1 recognition by FBW7. Nevertheless, some of them emitted fluorescence at high concentrations and could not be properly evaluated (Figure 4.25 D).

Example of 4 representative curves of FP competitive assay 2 are illustrated in Figure 4.25.
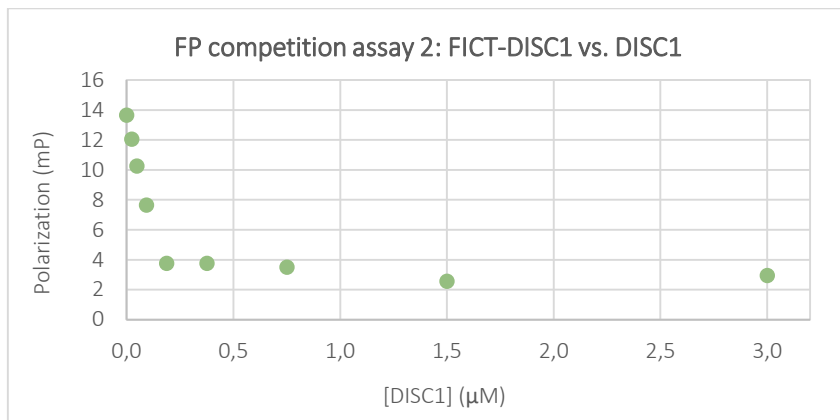
Figure 4.24. FP competition assay 2 at increasing concentrations of DISC1, from 0.02 μM to 3 μM at 8% DMSO.

A. FP COMPETITION ASSAY 1: MMC2   FP COMPETITION ASSAY 2: MMC2
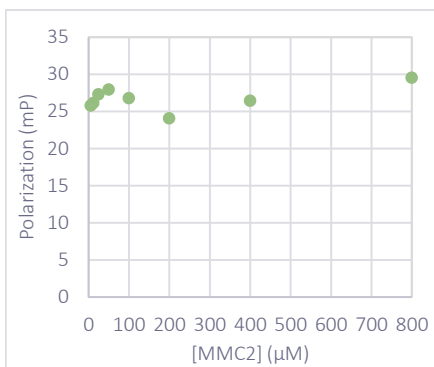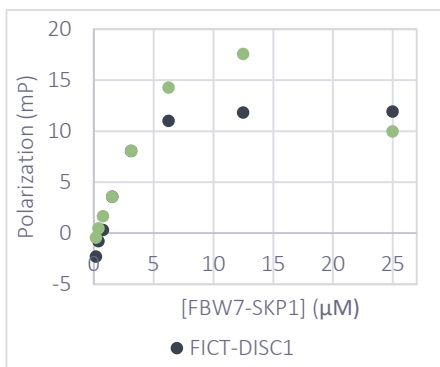
B. FP COMPETITION ASSAY 1: MMC37   FP COMPETITION ASSAY 2: MMC37

C.

D.

Figure 4.25. FP competition assay 1 and 2 of VS compounds: MMC2 (**A**), MMC37 (**B**), A2_MMC11 (**C**) and A5_MMC17 (**D**).

4.7.2. Cell-based assays

In order to disentangle the mechanism of action of A5_MMC17 and A6_MMC17, Andrea Bertran, a PhD student in the group, performed cell assays following the FBW7 substrate c-MYC by western-blot. HEK293 cells were treated for 6 hours with those compounds at three different concentrations: 10 µM, 100 µM, 250 µM. Besides, they were also tested at 250 µM after 1 hour treatment of MG-132[166] (10 µM), an inhibitor of the proteasome, to ensure proteasome-mediated degradation. Experiments were done in duplicates. Both, A5_MMC17 and A6_MMC17 appeared to decrease c-MYC levels in a proteasome-dependent manner

(Figure 4.26). Further experiments to study the mechanism of action of both compounds are going to be performed. CYCLIN-E, another FBW7 substrate, is being followed by western-blot and ubiquitination assays are being set up. Importantly, experiments that will allow us to relate the obtained results to FBW7 modulation will also be developed.
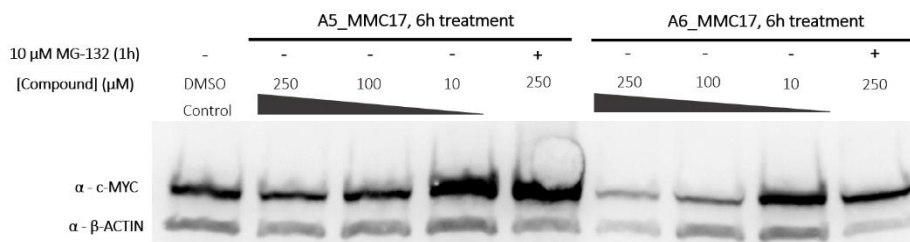
Figure 4.26. Results from cell-based assays. Western blot following FBW7 substrate c-MYC after A5_MMC17 and A6_MMC17 treatment, with and without addition of proteasome inhibitor MG-132. Abbreviations: α, anti.

4.8. Summary and future perspectives

Despite the tremendous significance of E3 ligases, only few of them have reported binders and just a handful number of small-molecules targeting them are in the market or in clinical trials. To contribute to the scientific need of increasing the toolbox of E3 ligases ligands, the initial objective of the present thesis was to study ligandability of this protein family. To achieve that, MDMix[53] methodology, developed in Dr Xavier Barril's group, was implemented. Multiple experimental studies have demonstrated the ability of proteins to unspecifically bind small organic co-solvents in regions over the surface, and these regions correlate well with binding sites and important interactions for the drug binding affinity[52]. MDMix exploits this concept of promiscuity performing molecular dynamic simulations with organic co-solvents to identify hotspots over the protein surfaces. On the whole, we aimed to find ligandable hotspots over protein surface with drug-like features. Encouraging results were obtained when validating the platform retrospectively with already crystallized ligands. MDMix ligandable pockets selected coincided with 6 out of 7 E3 ligases pockets with crystallized ligand. Indisulam interaction with DCAF15 E3 ligase was the one not coinciding. In fact, MDMix actually predicted this binding pocket, even though it was not selected after visual inspection for not

accomplishing drug-like features. Besides, indisulam has been reported to be potent when mediating DCAF15-RBM39 interaction acting as a molecular glue, rather than when binding to the two proteins separatedly[130]. Therefore, not selecting this pocket in MDMix studies was plausible. MDMix was also able to predict ligandable pockets in degron recognition sites, providing some starting anchors to drug these protein-protein interactions. Notwithstanding, some pockets predicted by MDMix in degron site were not considered as ligandable. These results emphasized the importance of protein surface geometry in ligand design and the challenges faced when targeting protein-protein interactions. Several E3 ligases have covalent described ligands. RNF4 was our example of this type of proteins and no ligandable pocket was found in this structure. Not having predicted a covalent pocket was expected, since these kinds of predictions would go beyond the scope of the methodology. Even the soft restraints applied when using MDMix do not allow the presence of big conformational changes, it did allow us to see the opening of pockets during simulations. This was the case of DCAF1-Pocket J, RNF43-Pocket F or VHL-Pocket C. As aforementioned, VHL-Pocket C coincides with the binding site of the already described VHL allosteric ligand[128]. Inspiringly, 66% of all predicted pockets were novel and allosteric, not being described before. All the information obtained could be point of departure to develop novel E3 ligase ligands. Besides, these results not only suppose valuable insights of the binding preferences of the selected E3 ligases in particular, but also presented a broad point of view on E3 ligases ligandability, illustrating the possibility to increase the number of binders of these challenging family.

FBW7 was one of the selected E3 ligases to perform MDMix, including at the structure its adaptor protein SKP1. FBW7 is one of the most commonly deregulated UPS protein in human cancers, which targets a range of substrates for degradation including some key human oncoproteins such as CYCLIN-E, c-MYC, NOTCH and JUNK[157]. Given its relevance, finding modulators for FBW7 would be of utmost value from a therapeutic point of view while expanding the toolbox of E3 ligases. Consequently, and in order to validate our platform, identification of small-molecules with the ability to bind FBW7 and their structural and functional characterization were pursued, using as starting point MDMix

results while applying computational tools. Virtual screening has demonstrated to be useful tool to target E3 ligases, since this approach has been employed for the identification of small-molecules binding to SKP2[118] and SPOP[119]. Starting from a library of 7 million compounds, FBW7-Pocket G resulting from MDMix experiments was used to develop binders of FBW7 E3 ligase starting from a computational approach that involved docking-based virtual screening and dynamic undocking, or DUck[92], a methodology developed in Dr Xavier Barril's group. While the first ensured that selected compounds accomplish the pharmacophore obtained from MDMix hotspots, the second has been proved to be useful to detect true ligands while increasing the success of virtual screening[92]. After visual inspection, 41 compounds were selected in order to be screened by biophysical techniques.

As a primary screening of virtual screening compounds, DSF technique was selected. DSF has been described as a versatile, fast and inexpensive methodology. These advantages make DSF an appropriate technique for high-throughput screenings for compound validation[65]. Nonetheless, DSF did not result to be the suitable technique to assess compound binding and results presented interpreting difficulties. We attributed these results to the complexity of the protein complex we were working with. As mentioned in Section 4.4.1, FBW7 is not only produced with SKP1, but also contains 3 differentiated domains (F-box, linker and WD40). Moreover, little increases in melting temperature were observed in the 41 molecules tested at 500 µM and 250 µM concentrations. For those reasons, SPR was decided to be performed instead.

SPR has been considered to be fast, involving low-sample consumption and quantitative, since it allows K_d determination[65]. Amine coupling reaction was selected as immobilization procedure, since MDMix binding pocket should not be affected. In order to better characterize compound binding, buffer had to be optimized by adding dextran. These discarded interactions with the dextran surface matrix and improved the quality of the data obtained. 9 compounds from VS presented a dose-response sensogram with K_ds in the one to three-digit micromolar range. This supposed a hit rate of 22%, which is considered a desirable success rate taking into account the expected for other structure-based

approaches[164]. Additionally, *SAR by catalogue* was performed selecting 10 additional compounds that passed docking-based VS filers. These new compounds had 70% or 80% of similarity to SPR VS positive compounds. This threshold is the reason why some parent compounds share few similarities with their analogues. All of them presented K_ds in the one to three-digit micromolar range, supposing a small increase in potency.

Crucial to this project was binding site characterization of SPR binders, what would confirm our hypothesis and the implementation of MDMix for ligand development. For that reason, we are collaborating with Dr Bing Hao to obtain crystal structures of the protein with our hits, although no results have overcome until the moment. In parallel, SPR competitive assay was performed and most binders seemed to clearly compete among each other. To ensure they bind to the predicted pocket, site directed mutagenesis experiments were performed to produce several single-point mutants in the predicted pocket. Site directed mutagenesis of the amino acid employed for DUck experiments did not affect compound binding in SPR experiments, probably because the interaction performed with this amino acid is easily exchanged by another residue. Differences in SPR binding were observed in the other single-point mutations performed. Orthogonal techniques were implemented to see differences in binding affinity between FBW7-SKP1 and mutants. MST was performed with 4 of the best SPR ligands. The resulting data had to be fitted by fluorescence intensity, rather than MST signal. This is a usual event that the provider (Nanotemper) contemplates within its manual. However, quenching of the fluorescence yielded upon our compounds binding was an indication that they were interacting or allosterically affecting fluorophore binding. Proteomic studies could be performed in order to study dye binding site. Binding differences with FBW7-SKP1 and mutants appeared to be inconsistent between MMC21 and A3_MMC21. To further study that and lead to an accurate conclusion, ITC of those compounds titrated against WT and mutants could be performed. Despite supposing a huge sample consumption, ITC is a sensitive, label-free and solution-based technique that provides unique data[67]. Encouragingly, compound A5_MMC17 presented differences in SPR, MST and ITC. These positive

results allowed us to initially conclude that it actually binds to the predicted binding site.

Initial studies of the effect of our compounds in FBW7 substrate recognition were carried out by fluorescence polarization. These binders were supposed to affect the ubiquitination process by constraining the system. Therefore, not modifying substrate recognition was not surprising. To disentangle the mechanism of action of these compounds, Andrea Bertran, a PhD student in the group, performed cell assays and followed c-MYC substrate of the FBW7 ligase by western-blot. Despite further confirmation assays need to be performed, A5_MMC17 and A6_MMC17 were observed to enhance the activity of FBW7 in a proteasome-dependent manner. Being both compounds analogues, their structure could be a point of departure to develop drugs able to modulate FBW7 E3 ligase, which has been considered undruggable until now, with no ligands described in thee literature.

CHAPTER 5: RESULTS

PHARMACOLOGICAL MODULATION OF TET2 EPIGENETIC ENZYME

RESULTS: PHARMACOLOGICAL MODULATION OF TET2 EPIGENETIC ENZYME

5.1. Background

5.1.1. Introduction to ten-eleven translocation (TET) protein family

Epigenetics involve the dynamic and reversible regulation of gene expression, contributing to normal cellular phenotype or to human diseases[167]. DNA methylation is a critical epigenetic modification, which regulates cell differentiation and proliferation, among other cellular processes. A key enzyme family for DNA demethylation is the ten-eleven translocation (TET) family, initially found as a chromosomal translocation partner in leukemia[168]. In humans, TET family is composed of three enzymes: TET1, TET2 and TET3, exhibiting different expression patterns. While TET1 and TET3 are ubiquitous, TET2 expression is predominant on several tissues, especially in hematopoietic and neuronal lineages[169].

TET proteins oxidize 5-methylcytosine (5-mC) in a α-ketoglutarate and Fe(II) dependent manner. Initially, TET enzyme converts 5-mC to 5-hydroxymethylcytosine (5-hmC), followed by oxidation to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC). This oxidative steps contribute to the demethylation of the cytosine, ultimately performed by the base excision repair mechanism (BER)[170] (Figure 5.1). Indeed, the preference of TET1 and TET2 for 5-mC rather than 5-hmC or 5-fC has already been described[100].
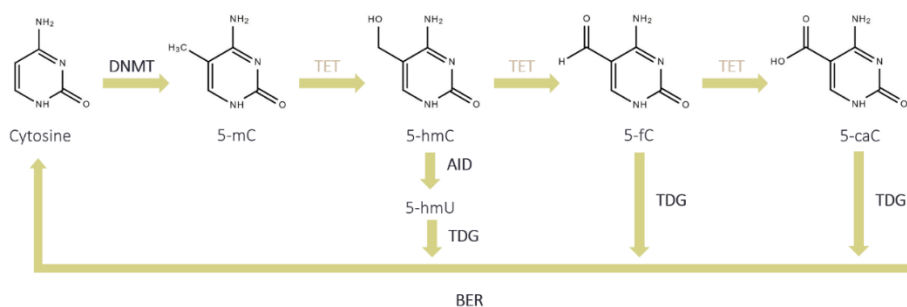
Figure 5.1. Pathway of cytosine demethylation mediated by TET family proteins. Abbreviations: AID, adenosine deaminase; BER, base excision repair; 5-caC, 5-carboxylcytosine; DNMT, DNA methyltransferase; 5-fC, 5-formylcytosine; 5-hmC, 5-hydroxymethylcytosine; 5-hmU, 5-hydroxymethyluracil; 5-mC, 5-methylcytosine; TDG, thymine DNA glycosylase.

Disruption of DNA methylation dynamics has been described to be a key feature in most cancers. Consequently, TET family is considered a critical regulator of cancer development together with the rest of DNA methylation and demethylation machineries. Furthermore, deregulation of TET enzymes can be observed in both haematological and solid cancers, being mutations more frequent in the first. As a consequence, tumour-suppressive functions of TET enzymes are impaired and carcinogenesis is promoted[171].

5.1.2. Therapeutic relevance of modulating TET2 epigenetic enzyme

DNA modifications performed by TET2 are essential for gen control in both cancer cells and immune-cell subtypes[172]. TET2 is specially mutated in hematopoietic cancers. Indeed, a loss of TET2 function has been reported in 10% of de *novo acute* myeloid leukaemia, 30% of myelodysplastic syndrome, and almost 50% of chronic myelomonocytic leukaemia patients[170]. Besides, several solid cancers also present mutations in TET2. In consequence, the activation of TET2 functional roles has been expected to be of utmost relevance in cancer treatment. The disclosure of a positive modulator of this protein would promote the disentangling of underlying molecular mechanisms of TET2 functions[172]. Vitamin C has been reported to be able to rescue TET2 function, blocking the progression of cancers such as acute myeloid leukaemia. This fact has made plausible that a pharmacological enhancement of TET2 wild type activity could actually block cancer[170].

In parallel, the research group of Dr Héctor G. Palmer at (VHIO, Barcelona) has also demonstrated the therapeutic interest of inhibiting TET2. TET2 has been found highly expressed in dormant tumour cells, which are known to be the cause of cancer relapses. Besides, TET2 has been identified as a key regulator of the number and survival of these cells, making it be applicable to predict prognosis and tumour recurrence. In short, negative modulation of TET2 could help to eliminate dormant tumour cells and to avoid cancer relapse[173]. On 2019, Bobcat399, an inhibitor of TET proteins was developed based on cytosine structure. Bobcat399 presented selectivity in TET proteins over DNA methyltransferase, specifically DNA methyltransferase 3A. The IC_{50}

for TET1 and TET2 were of 33 μM and 73 μM, respectively. Bobcat399 has been the first inhibitor of TET2 reported until now[174].

5.1.3. Preliminary results in our Lab

Our group has been collaborating with the group of Prof Héctor G. Palmer (VHIO, Barcelona) in order to perform a drug discovery project to develop a first-in-class TET2 modulator. TET2 had already been crystallized[99,100], allowing the application of a structure-based approach to find compounds interacting with this enzyme. Dr Sergi Ruiz a former member of the group, carried out MDMix[53] to determine TET2 druggable sites and an allosteric pocket in TET2 was identified. The hotspots of the selected pocket were subsequently used as pharmacophoric restraints to perform a virtual screening. Hence, small drug-like molecules with high-predicted binding capacity were developed. We proved that some of these molecules are active, increasing or decreasing TET2 enzymatic activity (Figure 5.2). During the present thesis, in order to improve the physicochemical properties of the initial active hits, potency and to define a composition of matter patent, analogues of best virtual screening hits were purchased and synthetized. Besides, in parallel activity of the positive binders was assessed *in vitro* and *in vivo* by our collaborators at VHIO. Structures of the compounds studied are not shown in the present thesis for confidentiality reasons.

A

ALLOSTERIC POCKET IDENTIFICATION

TET2-CD

IDENTIFICATION OF
POTETIAL TET2 BINDERS

Computational docking screening
of 6×10^6 diverse molecules with
drug-like chemical features

SELECTION OF 95
small moelcule
candidates
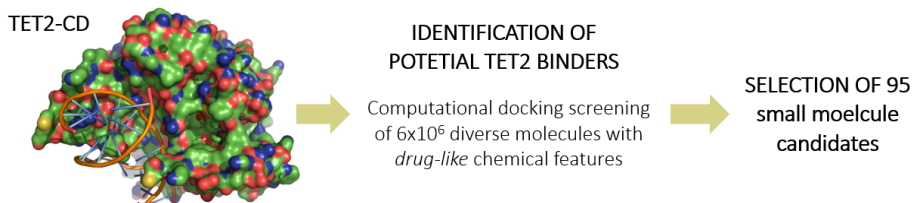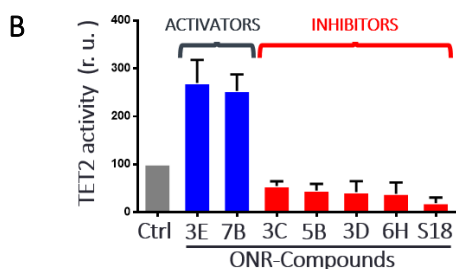
Figure 5.2. Preliminary results. A. Computational approach followed to determine the initial modulators of TET2 activity**. B.** Enzymatic assays of best compounds from virtual screening.

5.2. Objectives of Chapter 5

Given the relevant role of TET2 in cancer, the development of both negative and positive TET2 modulators would suppose an advance in the cancer drug discovery field. As previously mentioned, in collaboration with the group of Dr Héctor G. Palmer at VHIO, we identified an allosteric pocket in TET2 and potential modulators of this epigenetic enzyme. Enzymatic and functional assays were performed by our collaborators in VHIO.

The general objective of the present thesis is to perform biophysical techniques to study the affinity of TET2 potential binders and to structurally characterize the binding of TET2 with the best hits. In particular, the first specific objective is to determine optimal technique to study virtual screening hits and analogues of best binders from enzymatic assays (i), and to subsequently asses their binding and affinity (ii). Besides, another specific objective is to perform X-ray crystallography experiments to characterize binding of the best TET2 modulators (iii).

5.3. Protein production of TET2: cloning, expression and purification

5.3.1. Cloning, expression and purification of TET2 construct 1 for biophysical assays

To perform biophysical assays TET2 construct 1 was employed, performing its expression in ROS(DE3) cells after the induction of 0.4 mM

IPTG, as mentioned in Section 3.3.3.3. TET2 construct 1 purification started with cell lysis and clarification. The filtered sample was purified by immobilized metal ion affinity chromatography (IMAC) followed by size exclusion chromatography. Chromatograms of the purification steps are shown in Figure 5.3.
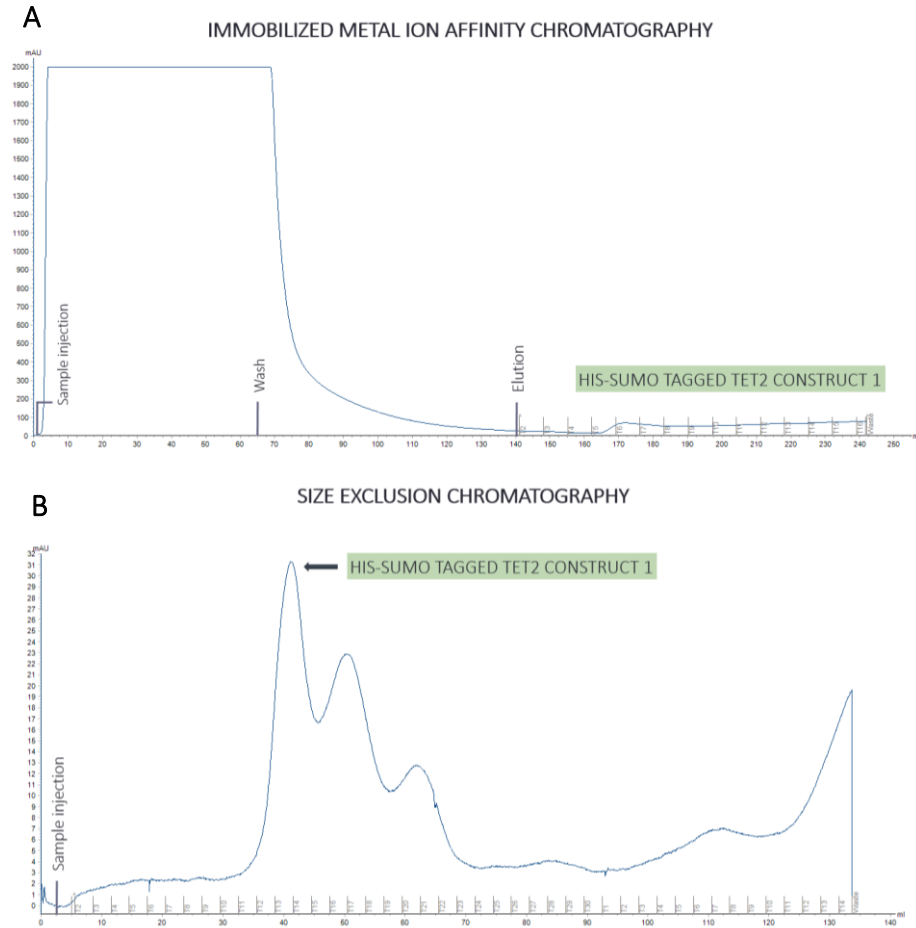
Figure 5.3. Chromatograms showing TET2 construct 1 purification procedure to perform biophysical assays. A. Immobilized metal ion affinity chromatography. **B.** Size exclusion chromatography.

After the purification procedure a yield of 0.4 mg/L was obtained. The mass and purity of the protein were verified by SDS–PAGE and mass spectrometry (Figure 5.4).
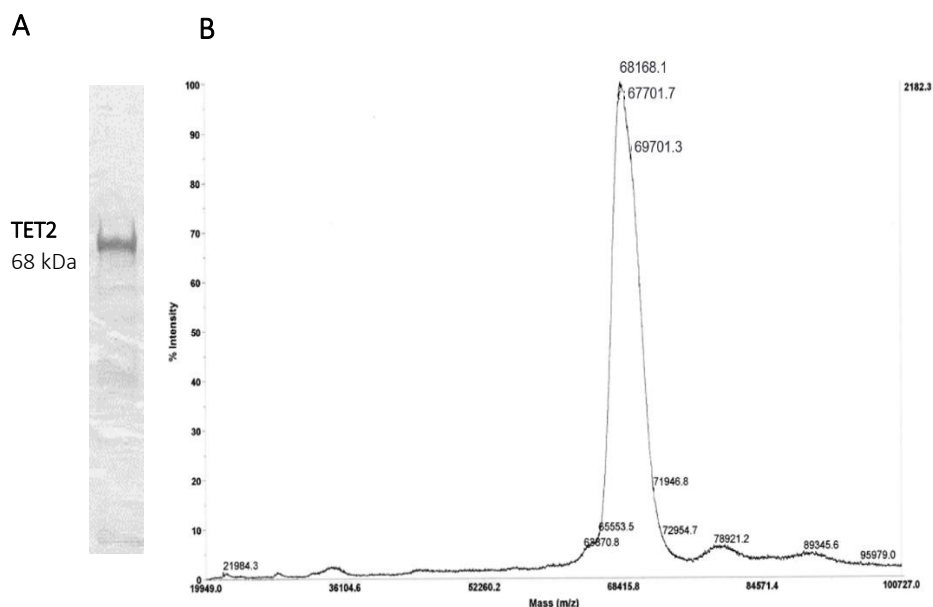
Figure 5.4. A. SDS polyacrylamide gel electrophoresis of His-SUMO-TET2 construct 1 after purification for biophysical assays. **B.** Matrix-assisted laser desorption/ionization (MALDI) of His-SUMO-TET2 construct 1 (68 kDa).

5.3.2. Cloning, expression and purification of TET2 construct 1 and TET2 construct 2 for crystallization experiments

The same cloning, expression and purification procedure was followed for TET2 construct 1 and TET2 construct 2 when performing crystallization experiments. They were both expressed in ROS(DE3) cells after the induction of 0.4 mM IPTG, as mentioned in Section 3.3.3.3. The purification procedure started with cell lysis followed by a clarification step. The filtered sample was purified by His-affinity chromatography (IMAC). The His-SUMO-tag was removed by ULP1 cleavage overnight. In order to remove the tag from the sample, a second His-affinity chromatography was performed. As a final purification step, an anion exchange chromatography was performed. Chromatograms of the purification steps are illustrated in Figure 5.5.
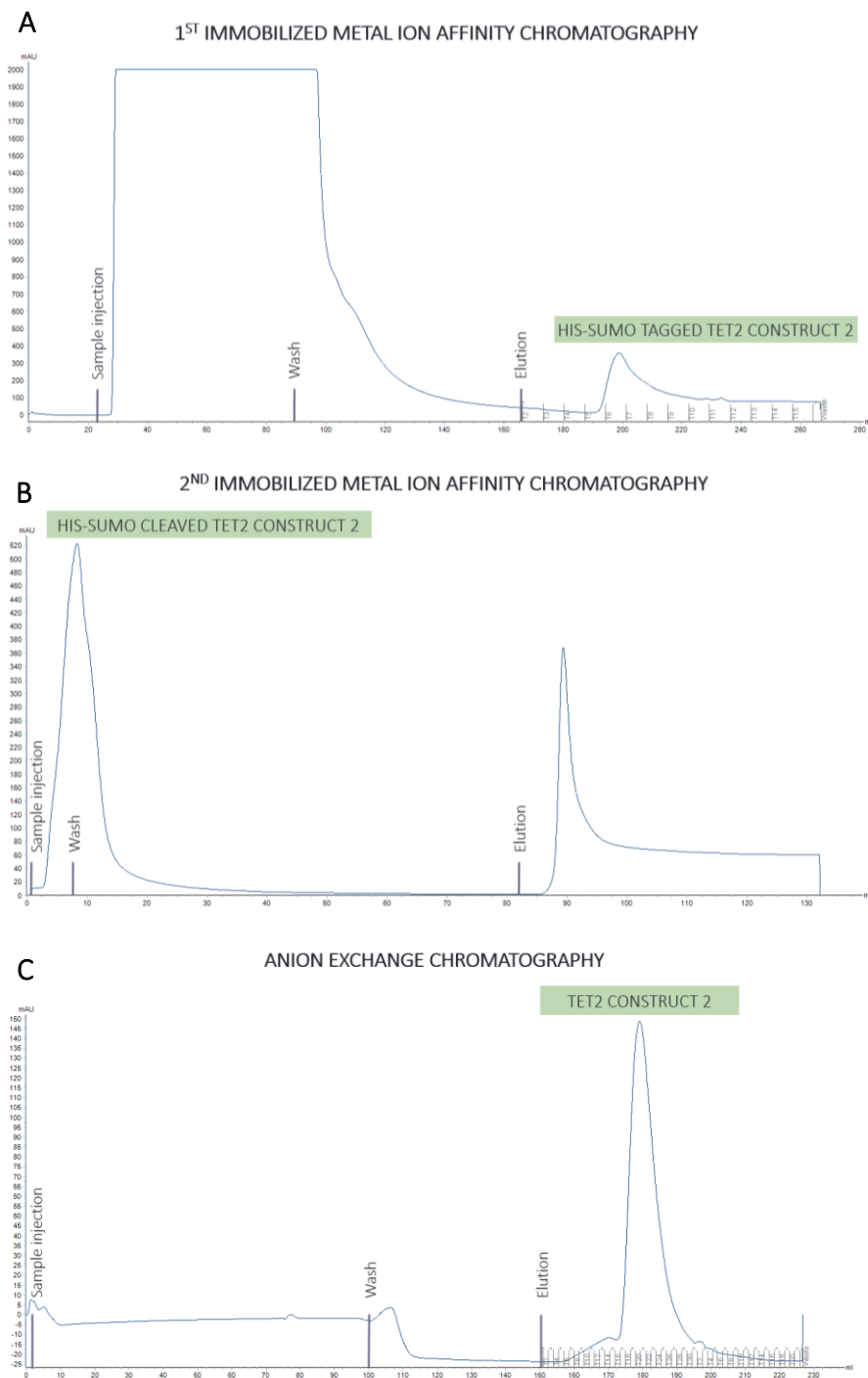
Figure 5.5. Chromatograms showing TET2 construct 2 purification procedure to perform crystallization experiments. A. First immobilized metal ion affinity chromatography. B. Second immobilized metal ion affinity chromatography. **C.** Anion exchange chromatography.

After the purification procedure a yield of 0.2 mg/L was obtained for TET2 construct 1, whereas a yield of 0.6 mg/L was achieved in the case of TET2 construct 2. The mass and purity of the protein were verified by SDS–PAGE and mass spectrometry (Figure 5.6).
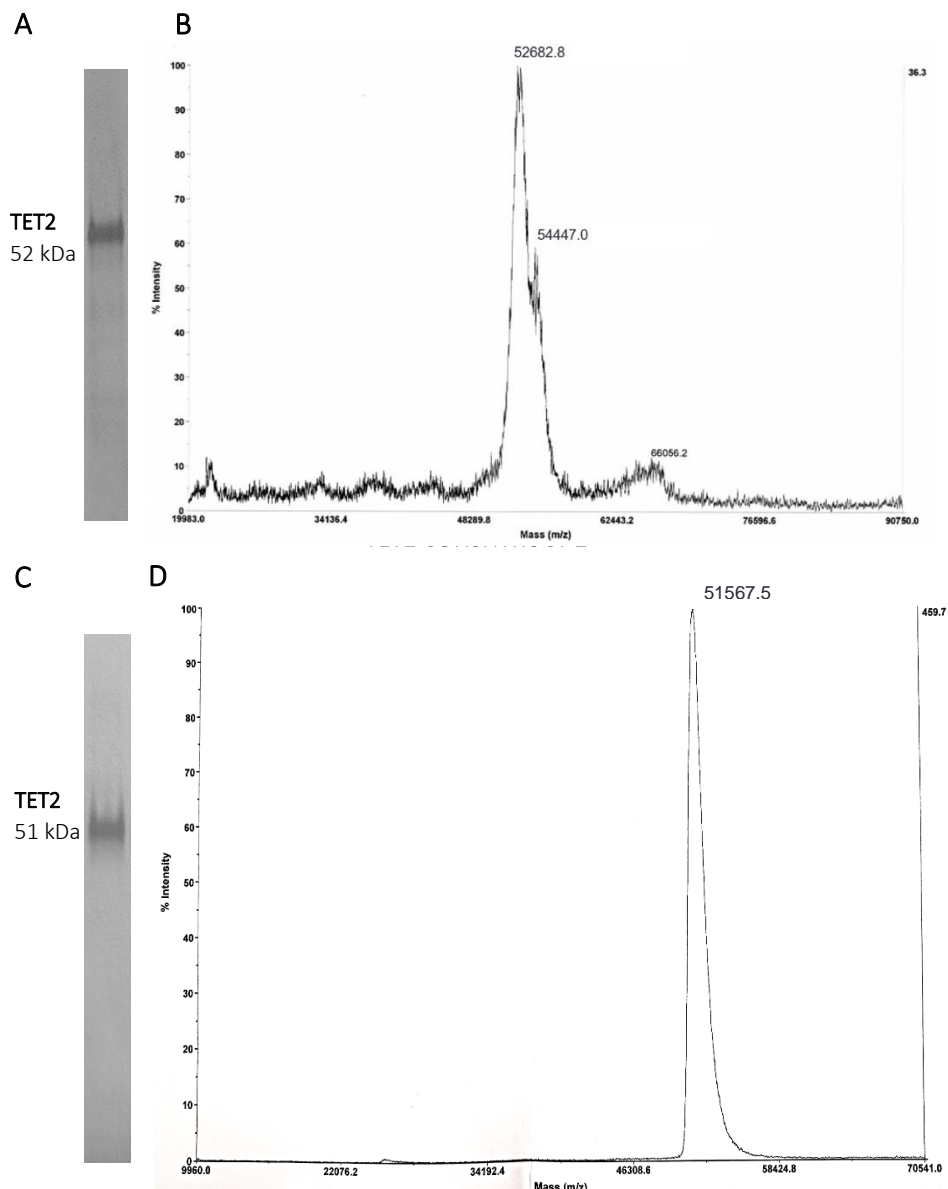
Figure 5.6. A. SDS polyacrylamide gel electrophoresis of TET2 construct 1 after purification**. B.** Matrix-Assisted Laser Desorption/Ionization (MALDI) of TET2 construct 1 (52 kDa). **C.** SDS polyacrylamide gel electrophoresis of TET2 construct 2 after purification. **D.** Matrix-assisted laser desorption/ionization (MALDI) of TET2 construct 2 (51 kDa).

5.4. Differential scanning fluorimetry of virtual screening hits and analogues

DSF was selected as a technique to assess binding to TET2 of the 92 virtual screening hits and the derivative analogues of the best hits according to enzymatic assays (performed by our collaborators in VHIO). Experiments were performed as described in Section 3.3.6.1. His-SUMO-TET2 produced from TET2 construct 1 (purification resulting from Section 5.3.1) was tested at a final concentration of 2.5 µM, which had a melting temperature of 40 ºC. Initially, 35 virtual screening hits were tested at 50 µM final concentration. In order to be considered significant, thermal shifts (ΔTm) of the protein-ligand complexes (ΔTm = Tm protein-ligand complex − Tm protein) had to be at least twice the standard deviation of the Tm of TET2. However, only in two of them (S19 and ONR-7B) an increase on protein stability was observed. For that reason, 57 virtual screening hits were then screened at 100 µM and only two additional compounds (3G, 2H) stabilized the protein. When comparing to enzymatic assays (Figure 5.2), only one of the positive compounds assessed stabilized TET2. In order to better understand that phenomenon, the three best compounds from enzymatic assays, 3D, 6H and ONR-7B, were tested in a dose-response manner at 10 µM, 50 µM, 100 µM and 200 µM (Figure 5.7). ONR-7B and 6H presented stabilization and destabilization, respectively, in in function of compound concentration (Figure 5.7 B and D).

Analogues of ONR-7B, 6H and 3D were purchased and tested at 100 µM. A summary of the results is illustrated in Figure 5.8. Analogues of 6H and 3D had a clear tendency to destabilize TET2. Unfortunately, no analogues presented a clear protein stabilization. In order to further study the binding of all compounds tested in DSF and since protein stabilization had been observed in few cases, SPR assays were performed.
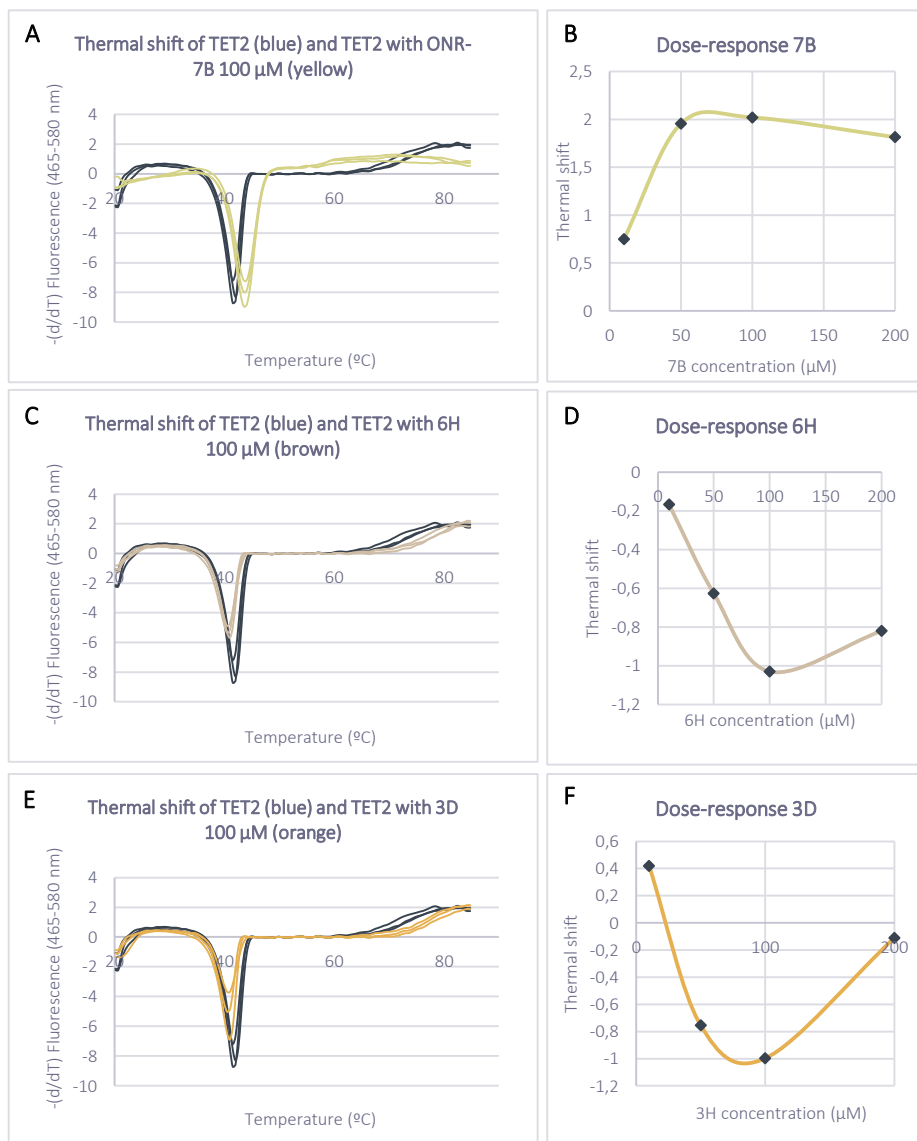
Figure 5.7. DSF graphics of initial hits. DSF graphics comparing melting temperatures of TET2 with TET2 in complex with 100 µM of ONR-7B (**A**), 100 µM of 6H (**C**) and 100 µM of 3D (**E**). Dose-response graphics that illustrate TET2 thermal shift derived from the interaction at increasing concentrations (10 µM, 50 µM, 100 µM and 200 µM) of ONR-7B (**B**), 6H (**D**) and 3D (**F**).
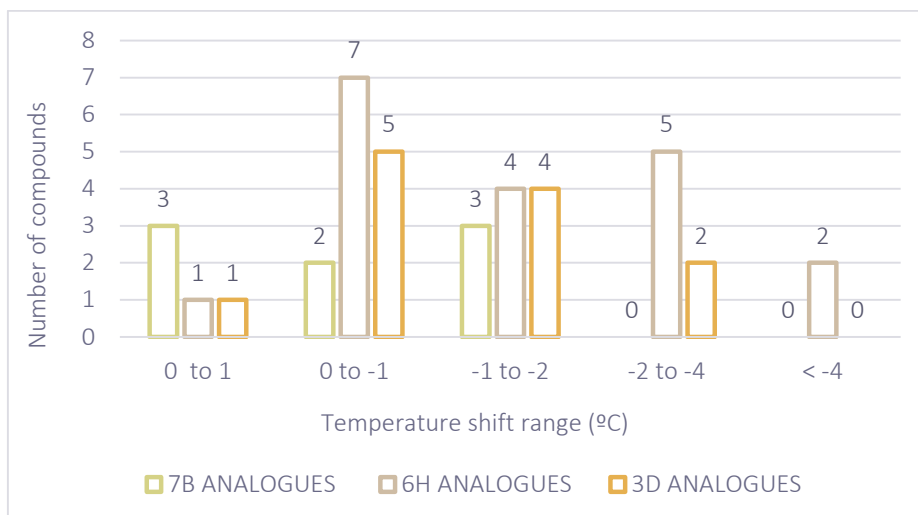
Figure 5.8. Summary of DSF results. Evaluation of TET2 interaction with 8 analogues of ONR-7B, 19 analogues of 6H and 12 analogues of 3D, all at 100 µM. Each column represents the number of compounds that produced a temperature shift in a particular range. Thermal shift at 100 µM of ONR-7B was 2 ºC, of 6H was -1 ºC and of 3H was -1 ºC.

5.5. Surface plasmon resonance of virtual screening hits and analogues

5.5.1. Surface plasmon resonance for hit identification and characterization

TET2 was immobilized in CM5 sensor chip with immobilization levels between 8000 and 9000 RU, following the protocol described in Section 3.3.6.2.1. Initially, the three best compounds from enzymatic assays (ONR-7B, 6H and 3D. Figure 5.2) were tested in a dose-response manner in a single replica in order to have a hint of the K_d. 3D saturated at 100 µM and K_d did not need to be calculated by fitting R_{max}. However, in the case of ONR-7B and 6H saturation was not achieved at soluble concentrations, forcing to calculate the affinity by fixing the R_{max}. This R_{max} consisted on the theoretical maximum response regarding the amount of ligand immobilized. In Figure 5.9 the initial dose-response curves and sensograms performed with 3D and 6H are illustrated. Besides, ONR-7B dose-response curve and sensograms in duplicates is also depicted.

A. 6H DOSE-RESPONSE CURVE

K_d: 61 µM
Chi$_2$: 9.2
R_{max} expected: 48 RU
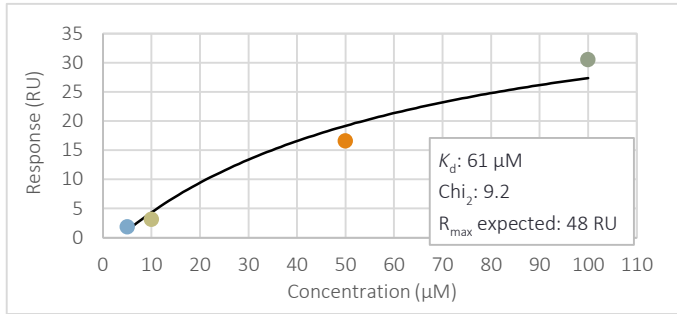
6H SENSOGRAM

B. 3D DOSE-RESPONSE CURVE

K_d: 29 µM
Chi$_2$: 2.1
R_{max} expected: 48 RU
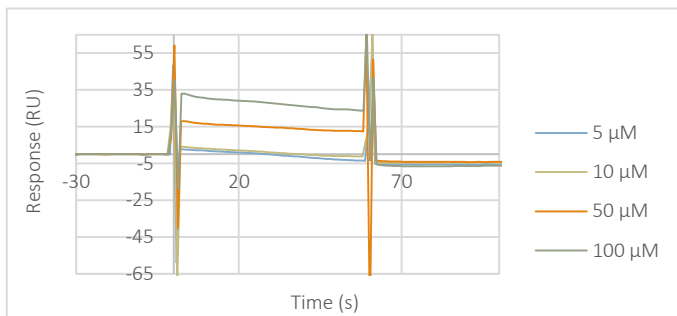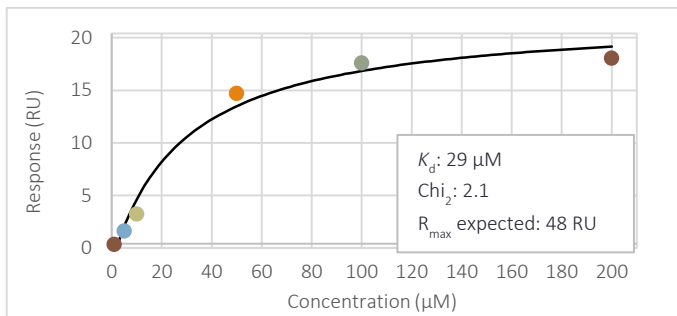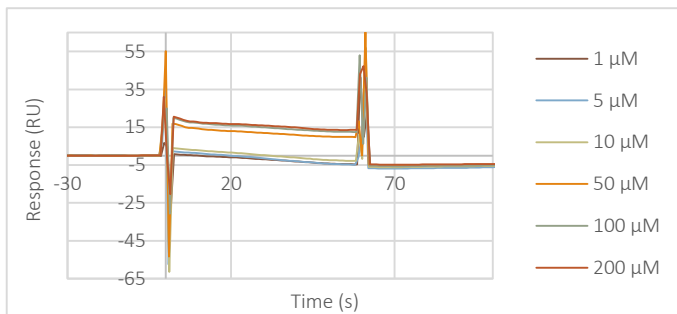
3D SENSOGRAM

C. 7B DOSE-RESPONSE CURVE
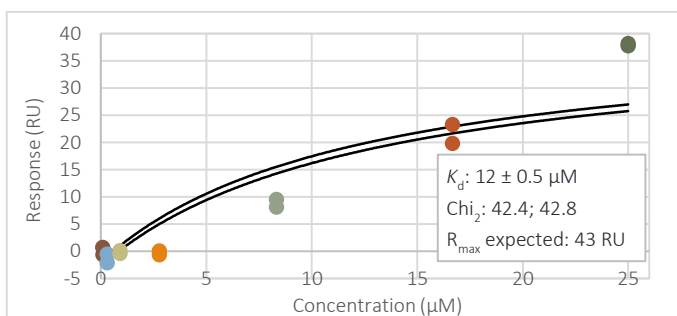
7B SENSOGRAM
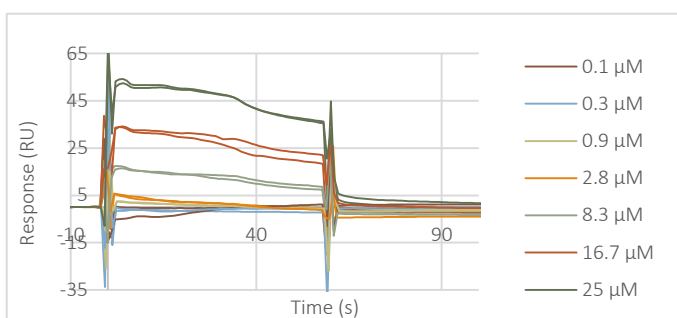
Figure 5.9. Dose-response curves and sensograms of 6H (A), 3D (B) and ONR-7B (C).

Seven analogues of 7B, previously tested by DSF, were screened in a single-dose assay at 10 µM. In contrast, twelve analogues of 6H and thirteen analogues of 3D were tested at a single-dose of 100 µM. Figure 5.10 A-C summarizes the obtained single-dose results regarding the percentage of theoretical R_{max}. Initially, no analogues presented single-dose response higher than ONR-7B. Indeed, in four of them negative results were obtained. These values were attributed to lack of solubility. Several analogues presented higher responses than 6H or 3D.

In parallel with SPR experiments, activity assays were also performed by our collaborators at VHIO. ONR-7B presented outstanding results in TET2 activation, making this compound and its analogues our priority for optimization and structure-activity relationship (SAR) studies. Dr Sergi Ruiz-Carmona, performed single-dose experiments at 50 µM for the eight ONR-7B analogues. Despite higher RU were obtained, some compounds presented solubility problems. For that reason, 25 µM was selected as an optimal dose, since it was a compromise between

solubility and relevant responses given by different compounds. 43 new ONR-7B analogues were purchased and synthetized to optimize the hit compound and study SAR. These compounds were screened at a single-dose, which results are illustrated in Figure 5.10 D. ONR-7B analogues that presented positive responses at a 10 µM and 25 µM experiments were tested in a dose-response manner to further characterize binding. In addition, compounds with structural interest were also screened. Overall, the affinity of 24 analogues was determined, being summarized in Table 5.1. Differences in response units between duplicates were observed. These were promoted due to an inappropriate blank subtraction, since the program only could subtract the average of blanks instead of each preceding blank for duplicate. Therefore, in order to proceed data treatment, each duplicate was analysed separately. K_d was determined by measuring the average of replicas and considering differences of separated K_d. R_{max} was fixed in all cases, as it was done with the parent compound ONR-7B. Chi^2 of both replicas were considered. Other considerations for the analysis are reported on Section 3.3.6.2.1. In SPR is established that Chi^2 should be less than 10% of R_{max} to be acceptable. In Table 5.1 it can be observed that some Chi^2 are outside the range. The purpose of these experiments was to discern between true and false binders and to rank them for activity assays performed at VHIO. For that reason, this data was already useful. Nevertheless, to have a more precise estimation of the affinity, more concentrations could be tested in the same range. None of the analogues improved remarkably the affinity of ONR-7B. Dose-response curves of representative analogues are depicted in Figure 5.11.

A

Compound	% Rmax
ONR-7B	66,9
A12	30,8
A29	26,3
A27	4,4
A33	-0,6
A18	-3,1
A21	-12,3
A31	-64,4

B

Compound	% Rmax
3D	36,0
A11	123,5
A19	109,0
A25	99,0
A24	63,3
A32	57,5
A2	51,9
A23	33,3
A22	32,5
A34	24,6
A26	24,0
A20	21,5
A35	20,0
A37	16,3

C

Compound	% Rmax
6H	62,9
A16	135,2
A8	99,2
A3	79,2
A14	77,3
A5	73,3
A6	71,3
A10	70,6
A1	69,2
A13	56,9
A2	51,9
A4	37,9
A7	14,0

D

Compound	% Rmax
ONR-7B	20,5
A12	36,7
A44	12,8
A33	12,3
FJPA256	10,9
FJPA260	10,6
OGD2	10,6
FJPA208	8,8
A49	8,3
FJPA207	8,1
FJPA204	8,0
FJPA258	7,6
FJPA203	7,6
FJPA201	6,9
FJPA213	5,8
FJPA205	5,5
FJPA238	3,6
FJPA212	1,4
FJPA250	0,4
OGD1	0,3
FJPA209	-0,2
FJPA234	-0,2
FJPA257	-0,3
A41	-0,6
A56	-0,7
A51	-0,7
FJPA237	-1,0
FJPA235	-1,3
A55	-1,6
FJPA236	-1,6
A46	-2,8
FJPA239	-3,6
A54	-5,0
A52	-5,1
A53	-5,1
A42	-6,2
A43	-6,3
FJPA211	-6,6
A45	-6,6
A47	-7,7
A40	-8,3
A50	-9,4
A48	-11,7
FJPA206	-12,4
FJPA202	-15,1
FJPA210	-26,9

Figure 5.10. Single-dose responses normalized by the percentage of the theoretical Rmax.
A. Percentage of Rmax of ONR-7B and analogues at 10 µM. **B.** Percentage of Rmax of 3D and analogues at 100 µM. **C.** Percentage of Rmax of 6H and analogues at 100 µM. **D.** Percentage of Rmax of ONR-7B and analogues at 25 µM.

Table 5.1. Summary of the SPR results obtained

Compound	K_d (µM)	Chi^2 replica 1	Chi^2 replica 2	R_{max}
7B	12 ± 0.5	42.4	42.8	43
A12	30 ± 19	72.9	7.8	43
A29	27 ± 1	5.7	16	43
A9	7 ± 0.9	146	333	43
A33	19 ± 0.3	73	57.3	43
A44	47 ± 4	7.3	9.8	43
A49	45 ± 2	14.4	11.9	43
FJPA201	83 ± 4	2.1	0.4	43
FJPA204	31 ± 3	0.2	0.6	43
FJPA205	69 ± 0.9	88.9	84.6	43
FJPA207	117 ± 6	8.2	9.4	43
FJPA208	93 ± 0.6	0.8	1.7	43
FJPA209	2300 ± 67	0.2	0.3	43
FJPA212	472 ± 95	0.2	0.4	43
FJPA213	2160 ± 56	0.8	2.4	43
FJPA234	221 ± 21	5	7.8	50
FJPA238	51 ± 0	16.2	14.7	50
FJPA239	4850 ± 5852	0.4	0.5	50
FJPA250	63 ± 5	103	100	50
FJPA256	16 ± 2	6.5	7.3	50
FJPA257	168 ± 39	6.1	15.6	50
FJPA258	17 ± 3	25.4	5.6	50
FJPA260	44 ± 22	0.4	0.8	50
OGD1	14 ± 0.4	0.6	0.4	50
OGD2	14 ± 10	50.2	10.2	50

K_d is the average of the replicas considering their deviation. Abbreviations: K_d, dissociation constant; Chi^2, chi square; R_{max}, maximum response.

A. A44 dose-response curve

K_d: 47 ± 4 μM
Chi$_2$: 7.3; 9.8
R$_{max}$ expected: 43 RU

B. A49 dose-response curve

K_d: 45 ± 2 μM
Chi$_2$: 14.4; 11.9
R$_{max}$ expected: 43 RU
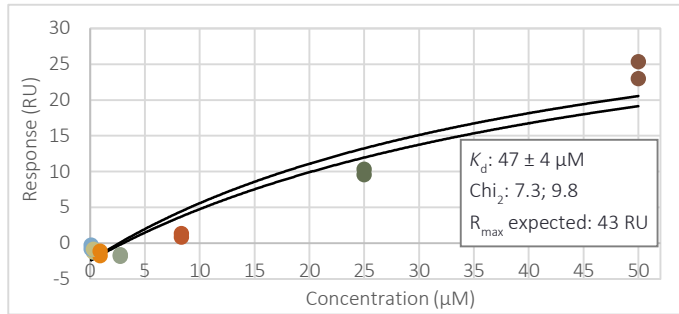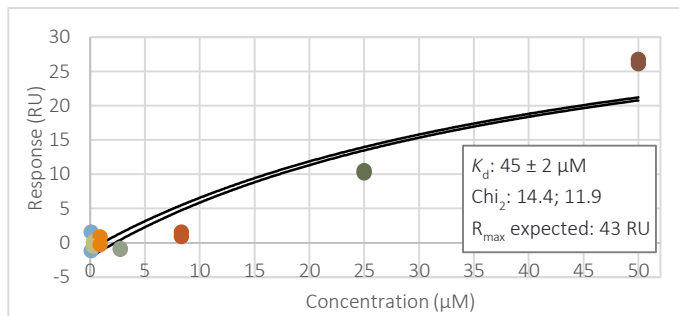
C. FJPA207 dose-response curve

K_d: 117 ± 6 μM
Chi$_2$: 8.2; 9.4
R$_{max}$ expected: 43 RU
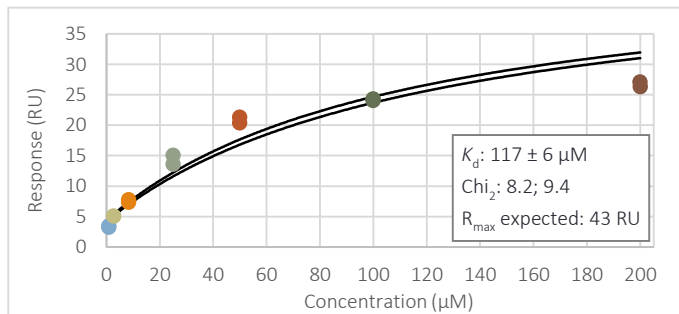
D. FJPA238 dose-response curve

K_d: 51 ± 0 μM
Chi$_2$: 16.2; 14.7
R$_{max}$ expected: 50 RU

E. FJPA250 dose-response curve

K_d: 63 ± 5 µM
Chi$_2$: 103; 100
R_{max} expected: 50 RU

F. FJPA256 dose-response curve

K_d: 16 ± 2 µM
Chi$_2$: 6.5; 7.3
R_{max} expected: 50 RU
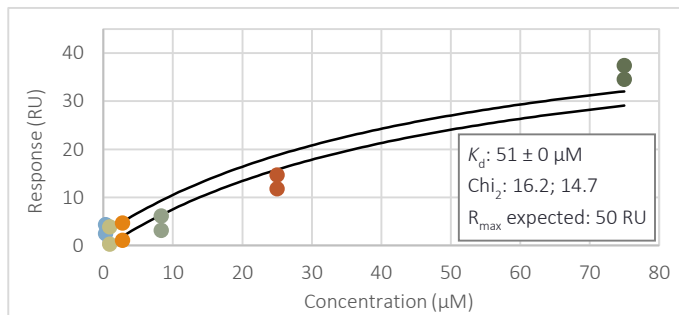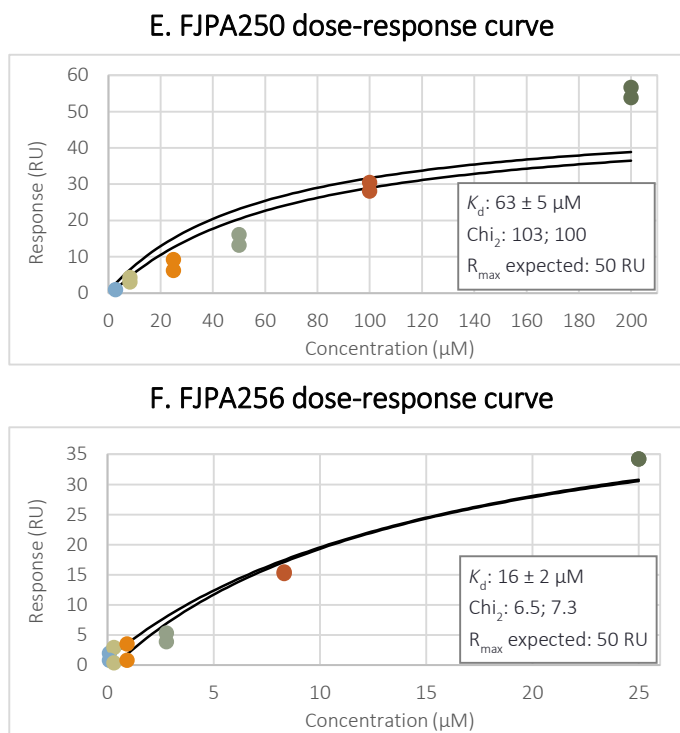
Figure 5.11. Dose-response curves of representative analogues: A44 (**A**), A49 (**B**), FJPA207 (**C**), FJPA238 (**D**), FJPA250 (**E**), and FJPA256 (**F**).

5.5.2. Surface plasmon resonance to study mechanism of action of binders

These experiments were done together with Dr Sergi Ruiz-Carmona. TET2 interaction with DNA has already been studied in SPR experiments[100], being the K_d described of 0.68 µM for the DNA containing 5-mC. In our case, a biotinylated DNA sequence was employed, which binding to TET2 had already been assessed by our collaborators. SPR experiment conditions are described in Section 3.3.6.2.2. TET2 affinity for DNA was initially assessed by testing different concentrations of protein. K_d obtained was of 1.3 µM (Figure 5.12). Mixtures of TET2 in complex with ONR-7B, 3D and 6H were screened to determine the effect of the compounds on DNA binding. Small differences were seen between the affinities of TET2 compared with TET2 in mixture with compounds (Figure 5.12). Consequently, we determined that those compounds did not affect DNA binding. Dr Sergi Ruiz-Carmona repeated those experiments by increasing the concentration range tested and confirming our conclusions.
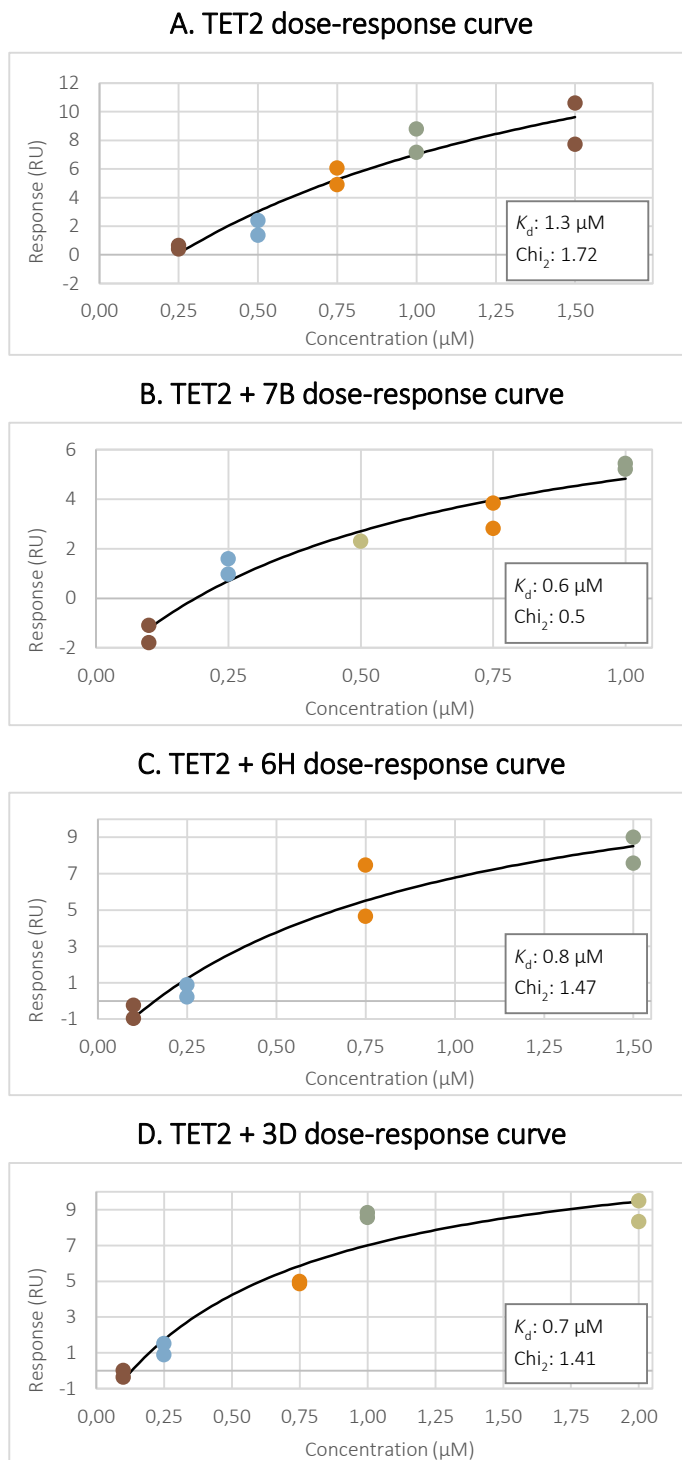
Figure 5.12. SPR assays immobilizing DNA. Dose-responses curves of TET2 binding to DNA (**A**) and TET2 in complex with ONR-7B (**B**), 6H (**C**) and 3D (**D**) binding to DNA.

5.6. X-ray crystallography of TET2 binders

X-ray crystallography trials of TET2 in complex with DNA and best binders were carried out through a placement in Dr Miquel Coll's laboratory (IRB).

5.6.1. TET2 crystallography of construct 1

Crystallization conditions of TET2 in complex with DNA have already been published[99,100]. Those conditions were reproduced using TET2 construct 1, being TET2 tested at 25 mg/ml in 1:1 ratio with DNA and with an inhibitor of α-ketoglutarate (in this case, 2-hydroxy-glutarate (2HG)) in excess (2 mM). Since no crystals were obtained, 96-well plate screenings with 300 nL drops (1:1 ratio) were performed using three different screens called PAC1, PAC9 and PAC10. These screens were performed by testing not only TET2, DNA, 2HG and ONR-7B complex, but also the following complexes: TET2 and ONR-7B; His-Sumo-TET2, DNA, 2HG and ONR-7B; His-Sumo-TET2 and ORB-7B. TET2, DNA and 2HG concentrations were the same as published and ONR-7B was tested at 1 mM and 2 mM (10% DMSO final concentration). All crystal growth was carried out at 4 ºC, except for the screenings with His-Sumo-TET2, in which 20 ºC was also tried since the protein should be more stable. From these trials, some resulting crystals were optimized in 24-well plates using hanging drop method. These conditions came from PAC1 well F11 and PAC9 well D9. The first consisted on 1.6 M ammonium sulphate, 0.1 M MES pH 6.5, 10% dioxane. PAC9-D9 contained 25 mM magnesium sulphate hydrate, 0.05 M TRIS pH 8.5 and 1.8 M ammonium sulphate. Two optimization plates were prepared, one for the complex of TET2, DNA, 2HG and ONR-7B, and the second for TET2, DNA, 2HG and FJPA260, another promising binder that is considered a negative allosteric modulator. Overall, the following crystals were brought to synchrotron: 3 crystals from optimization plate with ONR-7B, 4 from the optimization plate with FJPA260, 3 from PAC9 screening with TET2, DNA, 2HG and ONR-7B, and 6 of PAC9 screening with His-Sumo-TET2, DNA, 2HG and ONR-7B, 3 of which were from the plate stored at 4 ºC, whereas the other 3 were from the plate stored at 20 ºC. In all cases diffraction patterns were typical of salt or small molecule. We attributed these results to the fact that we were not using the same construct as crystallized.

5.6.2. TET2 crystallography of construct 2

Our collaborators in VHIO modified construct 1 to have exactly the same construct as published[99,100] (TET2 construct 2).

Described crystallization conditions were tried to be reproduced in 24-well plates with no relevant success. In parallel, 96-well plate screenings of 300 nL drops (1:1 ratio) were performed with PAC1, PAC9, PAC10, PAC21 and BCS screens with the following complexes: TET2, DNA, 2HG and ONR-7B; TET2, DNA, NOG (N-oxalylglycine, an inhibitor of α-ketoglutarate) and ONR-7B; TET2, DNA and 2HG; TET2 and ONR-7B. In these experiments, different TET2 concentrations were screened (maintaining 1:1 DNA ratio and 2mM of 2HG or NOG), as well as different ONR-7B concentrations. No crystals were observed in TET2 complexes without DNA. Crystals from PAC9 screening plate well D10 with TET2, DNA, 2HG and ONR-7B were brought to synchrotron (Table 5.2 A). Their reservoir was 5 mM magnesium sulphate hydrate, 0.05 M TRIS pH 8.5 and 35% 1,6-hexanediol. Besides, PAC10 screening also presented crystals in well G3 with TET2, DNA, 2HG and ONR-7B complex (Table 5.2 B). Its conditions consisted of 0.1 M magnesium chloride, 0.05 M TRIS pH 8.5, 20% PEG MME 2000. No conclusive results were obtained when bringing the crystal to synchrotron, probably because the crystal was too thin.

Optimization plates containing TET2, DNA, 2HG and ONR-7B complex were prepared from conditions PAC10-G3. Crystals resulting from this plate (Table 5.2 C) were brought to synchrotron and they resulted to be salt (Section 5.6.3). The same protein complex was tested for PAC10-C11 optimization conditions, which consisted of 0.1 M magnesium chloride, 0.05 M cacodylate pH 7 and 30% 2-Methyl-2,4-pentanediol. No crystals resulted from this optimization plate. Besides, optimization plates containing TET2, DNA and 2HG complex were prepared from PAC10-E12 conditions (0.2 M magnesium chloride, 0.05 M TRIS pH 8, 10% PEG200). These plates were further optimized varying crystallization conditions, protein concentration and complex-reservoir drop ratios. In parallel, since initially only tiny crystals were obtained, seeding plate was prepared by using as seeds these tiny crystals around conditions PAC10-E12 and PAC10-C11. Seeding screening was performed for TET2, DNA, 2HG complex and TET2, DNA, 2HG, ONR-7B complex. Crystals from PAC10-E12 optimization plates (Table 5.2 D) and from seeding screening were obtained and brought to ALBA synchrotron. Unfortunately, X-ray diffraction pattern of salt was observed (Section 5.6.3). Finally, PAC1-H3

consisted of 0.2 M magnesium chloride, 0.1 M TRIS pH 8.5, 3.4 M 1,6-hexanediol. An optimization plate around these conditions was prepared with TET2, DNA, NOG and ONR-7B complex. Resulting crystals (Table 5.2 E) were also brought to synchrotron. X-ray diffraction pattern showed that they were clearly not salt crystals (Section 5.6.3). Crystals from these conditions were rescreened confirming they were DNA crystals (Section 5.6.3).

In order to ensure the crystals contained 1:1 TET2 – DNA ratios and that the complex was properly formed, size-exclusion chromatography step was added before preparing crystal plates. To evaluate the results, absorptions at 260 and 280 nm were recorded, regarding DNA and protein absorption respectively. First, TET2 and DNA were loaded to the size-exclusion column separately in order to know their exact elution volume (Figure 5.13 A and B). Complex was then prepared likewise crystallization assays: 20 mg/mL of TET2 was mixed in 1:1 ratio with the DNA adding 2 mM of NOG. The sample was properly diluted to be injected to the column. Complex eluted at the same volume as the protein (9.7 mL), evidenced by the increase of absorption at 260 nm (Figure 5.13 C). The chromatogram also revealed an excess of DNA in the complex, which could be the reason why only DNA crystals were formed. Samples that contained the complex were pooled and concentrated until 26 mg/mL of TET2. 96-well screening plates (PAC1 and PAC2) were prepared with the previous mixture. For the moment, no crystals containing protein have been formed.
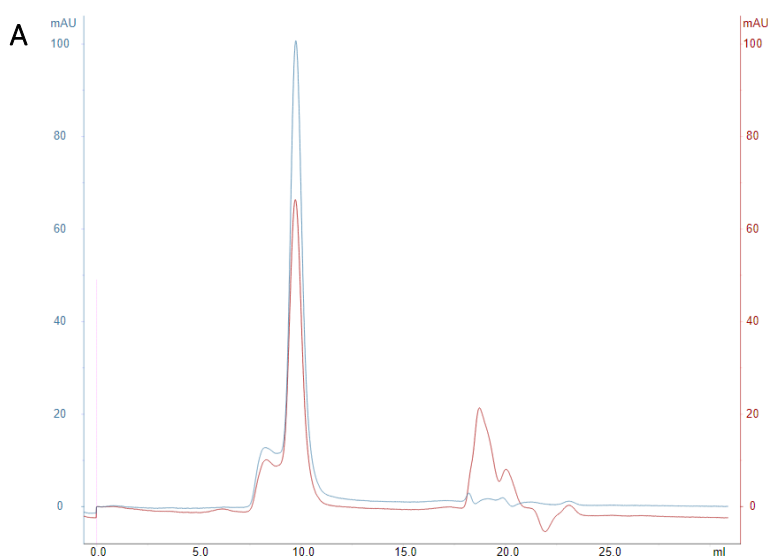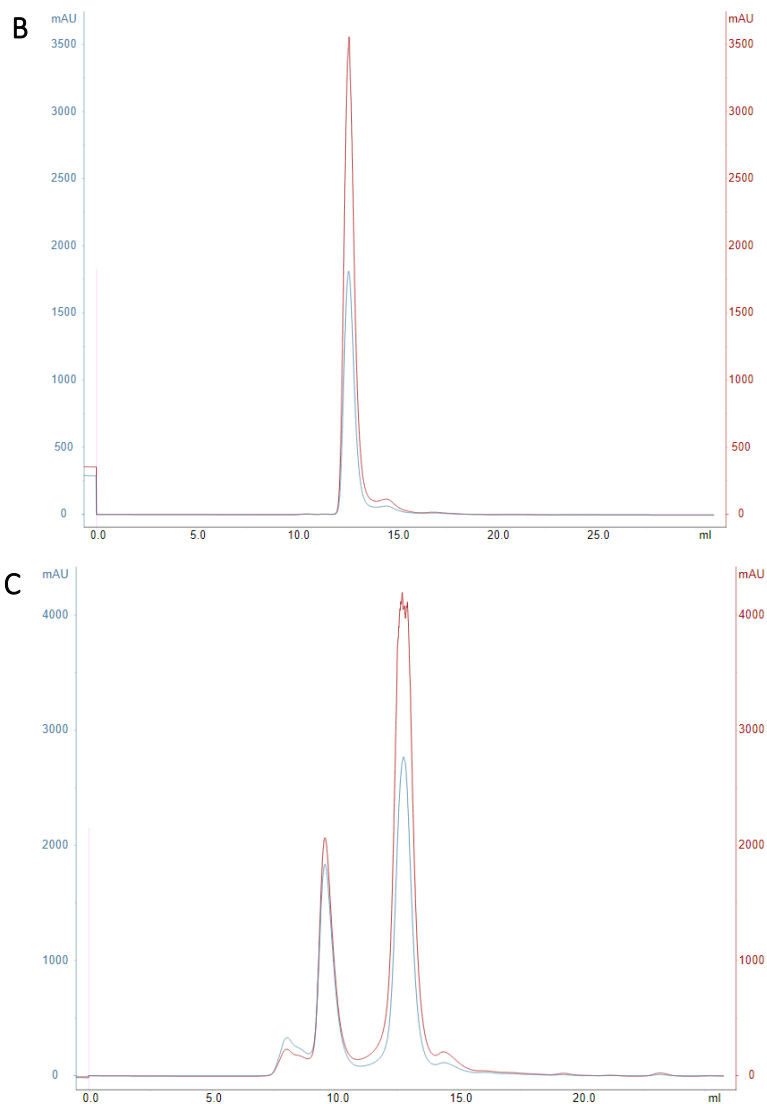
Figure 5.13. Size-exclusion chromatograms to verify TET2-DNA-NOG complex. Size-exclusion chromatograms of TET2 (**A**), DNA (**B**) and complex formed by TET2, DNA and NOG (**C**). Akta absorption units (mAU) in blue correspond to absorption at 280 nm, while in red correspond to absorption at 260 nm.

5.6.3. X-ray diffraction

Dr Albert Canals and Dr Montserrat Fabrega Ferrer, from the group of Dr Miquel Coll, were in charge of X-ray data collection of crystals from TET2 construct 1 and TET2 construct 2, respectively. X-ray diffraction data was collected at XALOC beamline at the ALBA synchrotron (Barcelona) or at

ID23-2 microfocus beamline at the European Synchrotron Radiation Facility (ERSF, Grenoble).

A total amount of 16 crystals of TET2 construct 1 were brought to XALOC beamline at ALBA synchrotron. Different protein complexes were assessed: 6 crystals that contained mixtures with TET2, DNA, 2HG and ONR-7B, 4 crystals that contained mixtures with TET2, DNA, 2HG and FJPA260, and 6 crystals that contained mixtures with His-Sumo-TET2, DNA, 2HG and ONR-7B. All of them presented diffraction patterns representative of salt or small molecules.

Initially, crystals containing TET2 construct 2 in complex with DNA, 2HG and ONR-7B were brought to ALBA synchrotron. They came from the screening plates PAC9 and PAC10, wells D10 and G3, respectively. While the one from PAC9-D10, clearly showed salt diffraction pattern, crystal from PAC10-G3 did not give conclusive results since it was too thin (Table 5.2 A and B). Optimization screening around this condition was performed and crystals were diffracted at ALBA synchrotron resulting to be salt (Figure 5.2 C). Crystals obtained from E12 optimization plate were also brought to ALBA synchrotron and presented salt diffraction pattern (Figure 5.2 D), as well as crystals from seeding plates. Finally, a crystal from PAC1-H3 was brought to ALBA synchrotron, giving patterns that clearly did not match with salt (Table 5.2 E). For that reason, another crystal from this optimization plate was brought to microfocus beamline (ID23-2) at ERSF, confirming it was a DNA crystal.

Table 5.2. Example of representative crystals brought to synchrotron with their diffraction pattern
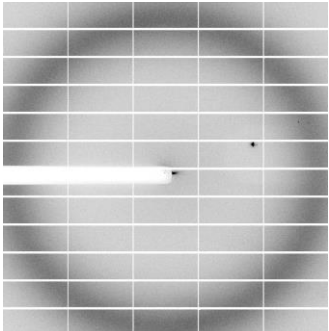
Crystal	Diffraction pattern
A	

Condition ID: PAC9-D10 screening
Exact conditions: 5 mM magnesium sulfate hydrate, 0.05 M TRIS pH 8.5 and 35% 1,6-hexanediol
Complex: TET2, DNA, 2HG, ONR-7B

Table 5.2 (cont.). Example of representative crystals brought to synchrotron with their diffraction pattern

B
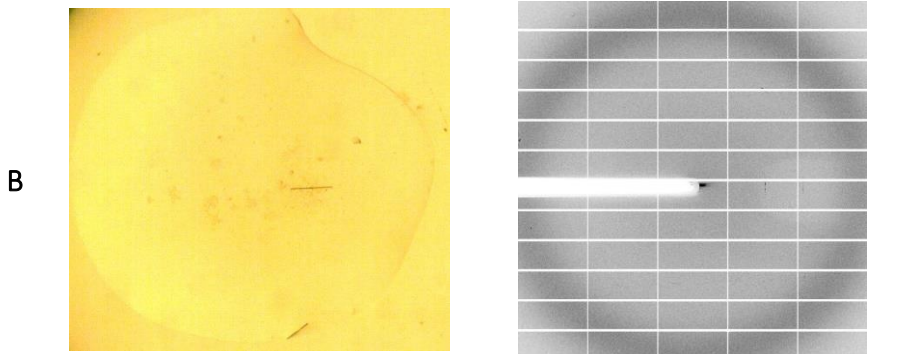
Condition ID: PAC10-G3 screening
Exact conditions: 0.1 M magnesium chloride, 0.05 M TRIS pH 8.5, 20% PEG MME 2000
Complex: TET2, DNA, 2HG, ONR-7B

C
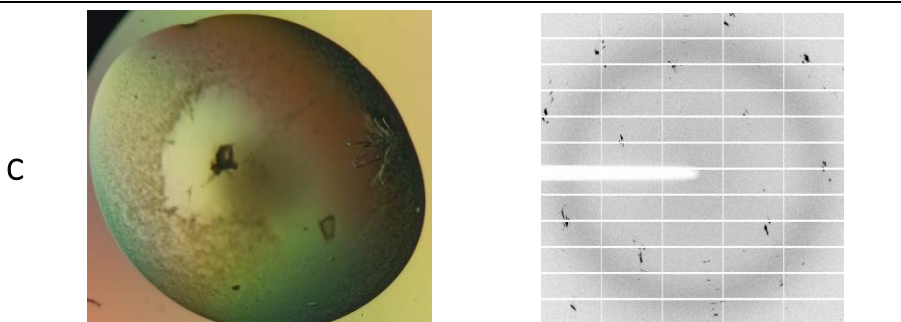
Condition ID: PAC10-G3 optimization
Exact conditions: 0.1 M magnesium chloride, 0.05 M TRIS pH 8.0, 10% PEG 2000 MME
Complex: TET2, DNA, 2HG, ONR-7B

D
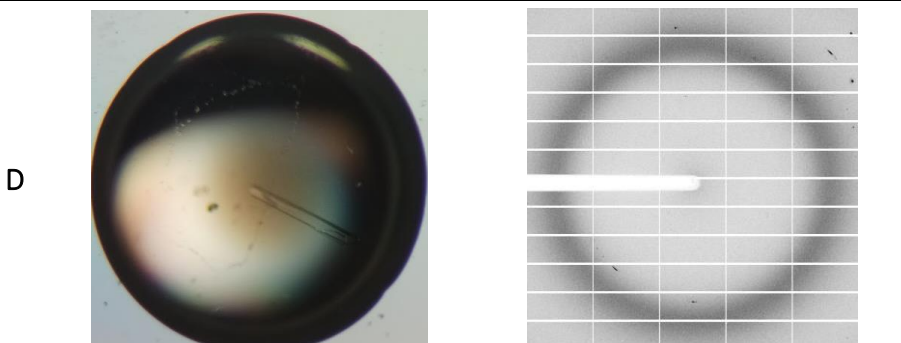
Condition ID: PAC10-E12 optimization
Exact conditions: 0.05 M magnesium chloride, 0.05 M TRIS pH 8.5, 5% PEG 200
Complex: TET2, DNA, 2HG

Table 5.2 (cont.). Example of representative crystals brought to synchrotron with their diffraction pattern

E
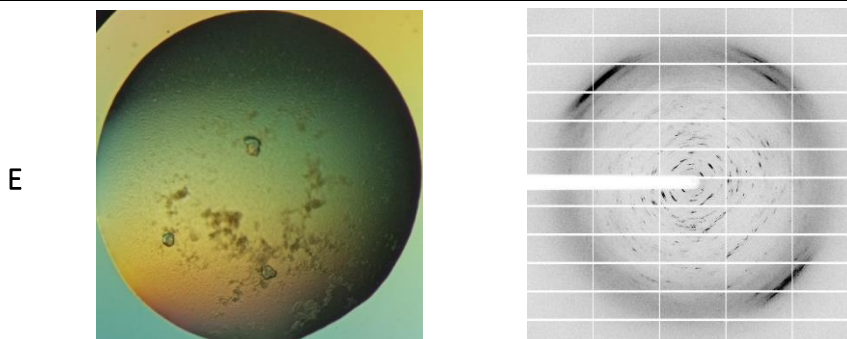
Condition ID: PAC1-H3 optimization
Exact conditions: 0.2M magnesium chloride, 0.05 M CHES pH 9, 3.8M 1,6-hexanediol
Complex: TET2, DNA, NOG, ONR-7B

5.7. Summary and future perspectives

The general goal of this multidisciplinary project is to perform a drug discovery approach to develop a first-in-class TET2 modulator. The contribution of the present thesis in this project was in the first instance to determine the optimal biophysical technique to study binding of TET2 modulators. DSF was considered as a primary screening, however few compounds could stabilize TET2. Indeed, a relevant number of compounds destabilized the protein. Despite it has been reported that destabilizing agents can be confirmed as true binders[70], we considered SPR as an alternative technique. SPR was performed, allowing us to assess not only binding but also the affinity of the compounds. SPR screening was performed of best virtual screening hits according to enzymatic assays and their analogues. Since ONR-7B presented outstanding results in TET2 activation, affinity of this compound and its analogues was characterized. Additional ONR-7B analogues were purchased and synthetized to improve their physicochemical properties, potency and to define a composition of matter patent. Even though an improvement of solubility was observed in ONR-7B analogues, none of them improved significantly its potency. Despite SPR data fitting was not optimal, compounds could be properly ranked in a priority list to be screened *in vitro* and *in vivo* by our collaborators at VHIO. Interestingly,

some ONR-7B analogues were found to be negative modulators of TET2 activity. SAR studies have been performed to better understand this phenomenon.

Structural information of ONR-7B binding would not only be crucial to understand the activity mechanism of TET2 modulators, but also to validate the structure-based approach employed in this drug discovery project. X-ray crystallography is the ideal technique for that. It is well known, as I experienced, that reproducing crystal conditions is not a straightforward process and performing screenings with other crystallization conditions should always be considered. On one hand, promising crystals were grown of TET2 in complex with DNA and 2HG from PAC10-E12 optimization plates. Unfortunately, they clearly were not protein crystals. On the other hand, crystals from PAC1-H3 optimization plate with TET2, DNA, NOG and ONR-7B clearly showed a non-salt diffraction pattern, although they resulted to be DNA crystals. In order to confirm the 1:1 ratio of the complex, size-exclusion chromatography was performed assessing complex formation. An excess of DNA was observed, which could explain the DNA crystals obtained. Two additional crystallization screenings were tried without the DNA excess with no success at the moment. Further crystallization experiments are going to be performed adding the size-exclusion chromatography step after complex formation. Despite the effort that can suppose obtaining an X-ray structure of a protein-ligand complex, crucial information can be provided. A structure of ONR-7B binding to TET2 would not only give a better understanding of the ligand mechanism, but would also provide valuable data for its optimization.

In parallel to this thesis, selected molecule ONR-7B was further studied by our collaborators in VHIO. ONR-7B not only was the most potent activator in enzymatic assays, but also confirmed its capacity to bind and activate TET2 on genomic DNA. ONR-7B blocked cancer cell proliferation and arrested leukemic cells in G1/G0 of the cell cycle. Besides, ONR-7B presented an ideal pharmacokinetic/pharmacodynamic PK/PD profile and was effective controlling the growth of AML tumours in NOD-SCID mice. Bone marrow transplant of HL60 leukemic cells in mice provoked their death in 28 to 34 days. One third of ONR-7B treated mice survived longer and even one animal was cured and lived more than three months

after cell injection and drug removal. Despite these encouraging results, further characterization of this small molecule would be of utmost importance in order to pursue with this drug discovery project.

CHAPTER 6: RESULTS

DEVELOPMENT OF NEW CHEMICAL ENTITIES THAT BIND TO THE BROMODOMAIN BRD4 BD1

RESULTS: DEVELOPMENT OF NEW CHEMICAL ENTITIES THAT BIND TO THE BROMODOMAIN BRD4 BD1

6.1. Background

6.1.1. Introduction to bromodomains

The term epigenetics, which literally means "above genetics", has gained popularity in the drug discovery field in the last few years. Epigenetics involve the dynamic and reversible regulation of gene expression, a process that can contribute to normal cellular phenotype or to human diseases[167,175]. Epigenetics mechanisms encompass covalent modifications of DNA and the proteins that package it, such as histones. Histone modifications include not only phosphorylation and ubiquitination, but also methylation and acetylation, being this last one the most abundant and studied. There are three classes of enzymes implicated in histone modifications: the writers, the erasers and the readers[175]. The bromodomain-containing family of proteins is a class of histone modification reader proteins that recognize acetylated lysine residues.
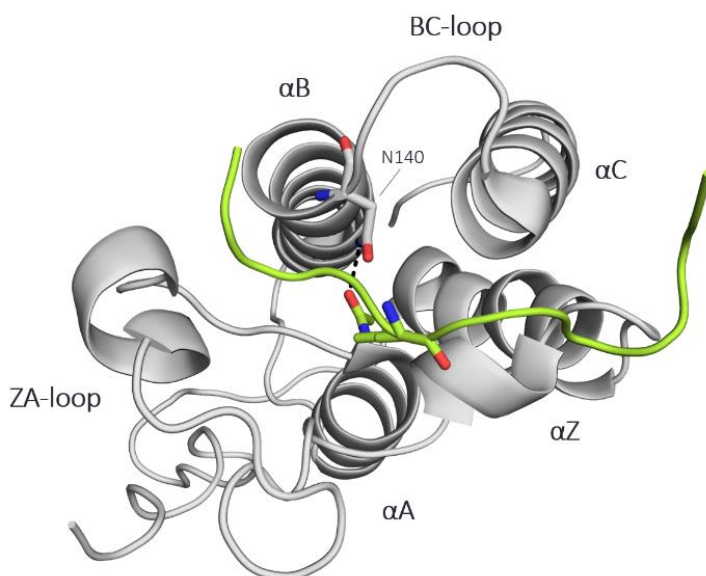
Figure 6.1. Structure of BRD4 BD1 (white) interacting with a diacetylated histone 4 peptide (green). PDB code: 3UVW[176]. Architecture of bromodomains is labelled and interaction between ASN (N) and acetylated peptide is shown in black slashes.

Bromodomains are composed of around 110 amino acids[177] and even they present large sequence variations, they all share conserved secondary structure. This common structure incorporates four alpha helices (aZ, aA, aB, aC), linked by diverse loop regions (ZA and BC loops) of variable length (Figure 6.1). Besides, in most cases, acetyl-lysine posttranslational modification is recognized within a hydrophobic pocket anchored by forming a conserved hydrogen bond with a highly conserved asparagine residue[104].

The human genome encodes 46 bromodomain-containing proteins. Each of them can incorporate between one and six bromodomains, resulting in a total amount of 61 diverse domains. Structure-based alignments of bromodomain containing proteins clustered the 46 proteins in eight diverse bromodomain groups (I-VIII)[104].

6.1.2. BET family of bromodomains

Class II of bromodomains is also referred as bromodomain and extra-terminal domain (BET) family. This name has been assigned since the protein members present two related tandem bromodomains, named BD1 and BD2, and a unique extra-terminal region in the C-terminal moiety[178]. In humans, this family consist of 4 members: BRD2, BRD3, BRD4 and BRDT, being BRD2-4 ubiquitously expressed and BRDT expressed in germ cells[179].

BET bromodomains recognize acetylated lysine residues in histones H3 and H4. Furthermore, they are also able to interact with different molecules with relevance in gene transcription. Potent and selective BET inhibitors have already been developed, exposing the large number of potential applications they have in drug discovery[179]. First generation of BET inhibitors were not selective among family members or BD domains. Nevertheless, compounds selective for BD1 or BD2 have recently been developed. Besides, selectivity has also been achieved with the development of PROTAC molecules (see Section 1.3), where the selectivity of PROTAC degradation has exceed the selectivity of the parent ligand. This full toolbox of compounds, which have shown robust preclinical activity and involvement in oncogenic transcriptional networks, have triggered to clinical activity of BET inhibitors, mainly in oncology applications[177].

6.1.2.1. BRD4 epigenetic enzyme and its interest in drug discovery

BRD4 is the most exhaustively studied member of BET family and even of the all readers of the histone code. BRD4 has been reported to be involved in cell cycle control and to influence cell proliferation, apoptosis and transcription. Consequently, it contributes to different human diseases including cancer, inflammation, human immunodeficiency virus infections, central nervous system disorders, and cardiovascular diseases. For that reason, BRD4 has been considered a promising target in drug discovery and the development of BRD4 inhibitors has been and is being pursued in many research groups[180]. Similarity between BET family members supposes a challenge to develop selective inhibitors for each protein and domain. In fact, practically only selective inhibitors for BD1 and BD2 domains have been disclosed[181]. Different approaches are being developed to increase protein selectivity, such as PROTACs (see Section 1.3). PROTAC technology has recently provided selectivity in BRD4 degradation over other BET family members[182]. Overall, despite more than 20 BRD4 inhibitors have already entered to clinical trials, more potent and selective inhibitors of BRD4 would clearly benefit the drug discovery field[181].

6.1.2.2. Structural network of water molecules in bromodomains

Substrate recognition of acetyl-lysine by bromodomains is mediated through a hydrogen bond with highly conserved asparagine, ASN140 in the case of BRD4 BD1, and an additional water-mediated hydrogen bond with a tyrosine, TYR97 for BRD4 BD1[181] (Figure 6.2). The latter is formed by the presence of a network of water molecules placed at the bottom of the cavity[178]. This water molecules have been extensively studied computationally and have been found to be preserved in holo and apo structures of bromodomains[183,184,185].

Despite preserved waters have already been displaced in BRD4 BD1[103,186], in none of them the displacement has been predicted rationally. Nonetheless, the most recent publication in 2021 has demonstrated that water displacement of preserved waters can provide selectivity among different BET members[186]. This fact manifests the relevance of water molecules in protein structures and molecular

recognition, making its prediction one of the current challenges in structure-based drug design.
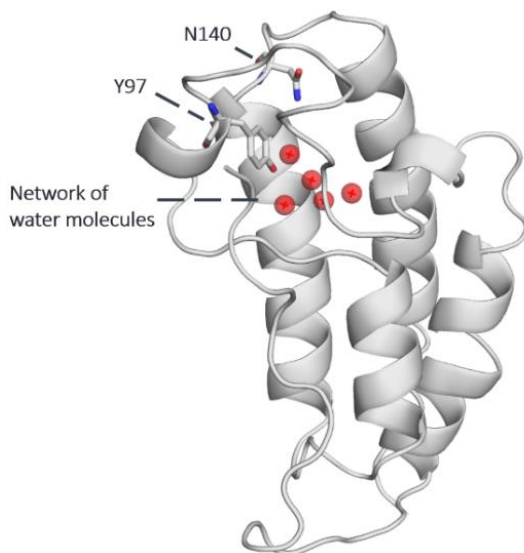
Figure 6.2. BRD4 BD1 structure with conserved water molecules.

6.2. Objectives of Chapter 6

In general terms, the objective of Chapter 6 is to develop and characterize new chemical entities that bind to the bromodomain BRD4 BD1. For that, two different strategies have been envisaged in the Lab.

- We aim to study the solvation preferences of BRD4 BD1 using computational tools. With the information obtained, an additional objective is to develop and characterize small molecules able to perform novel interactions by promoting the displacement of water molecules of BRD4 BD1 (Section 6.4 of this chapter).

- Two former PhD students in our Lab, Dr Serena Piticchio and Dr Moira Rachman, developed an automated platform based on fragment evolution to generate potent and diverse binders for BRD4 BD1 starting from low-potency fragment molecules (Section 6.5.1). Binding and structural characterization of evolved fragments is also an objective of the present thesis (Section 6.5 of this chapter).

6.3. General results

6.3.1. Protein production of BRD4 BD1: cloning, expression and purification

6.3.1.1. Cloning, expression and purification of BRD4 BD1 for biophysical assays

To perform biophysical assays BRD4 BD1 was expressed in ROS(DE3) cells after the induction of 0.4 mM IPTG, as mentioned in Section 3.4.4. Protein purification consisted on cell lysis and clarification followed by a His-affinity chromatography (metal ion affinity chromatography). Then, His-tag was removed by His-TEV protease cleavage overnight, and a second His-affinity chromatography was performed to remove the tag and TEV protease from the sample. Chromatograms of the purification steps are shown in Figure 6.3.
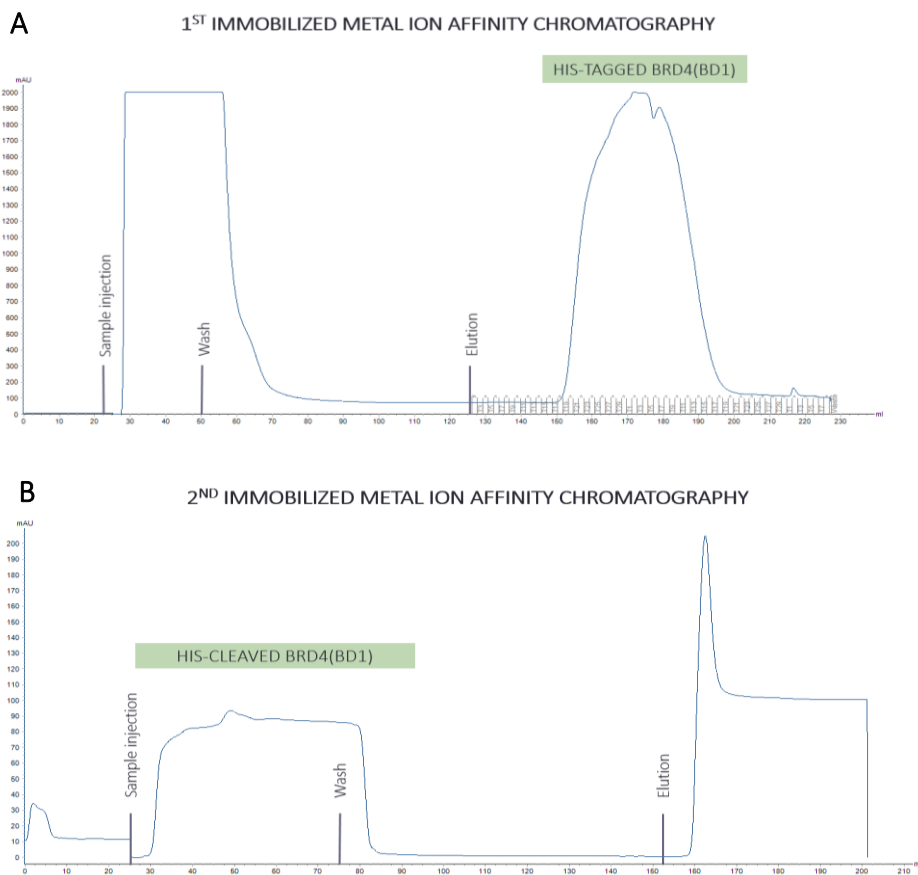
Figure 6.3. Chromatograms showing BRD4 BD1 purification procedure to perform biophysical experiments. A. First immobilized metal ion affinity chromatography. **B.** Second immobilized metal ion affinity chromatography.

After the purification procedure, a yield of 10 mg/L was obtained. The mass and purity of the protein were verified by SDS–PAGE and mass spectrometry (Figure 6.4).
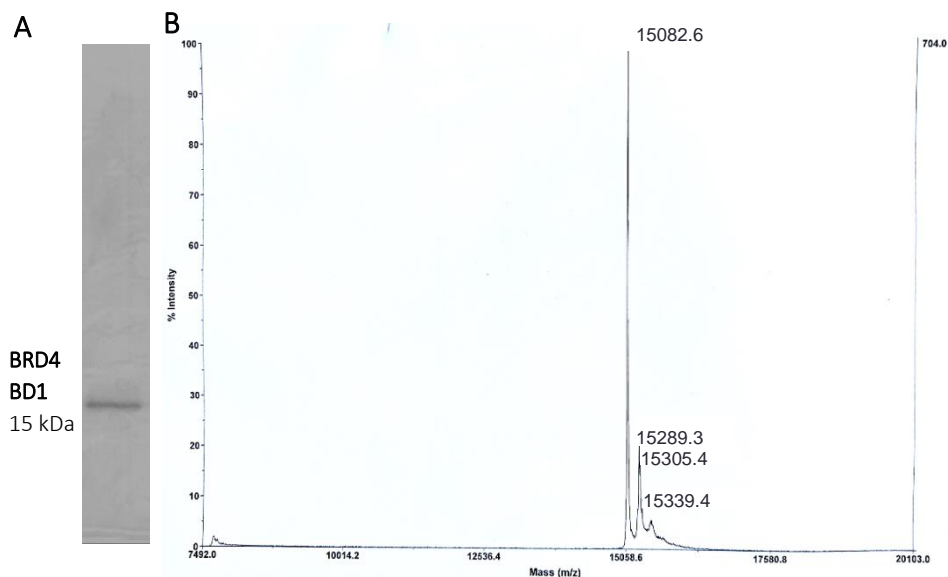
Figure 6.4. A. SDS polyacrylamide gel electrophoresis of BRD4 BD1 after purification. **B.** Matrix-assisted laser desorption/ionization (MALDI) of BRD4 BD1 (15 kDa).

6.3.1.2. Cloning, expression and purification of BRD4 BD1 for crystallization experiments

BRD4 BD1 production to perform crystallization experiments was carried out at the laboratories of Prof Frank von Delft and Prof Panagis Filippakopoulos (Structural Genomics Consortium (SGC), Oxford). BRD4 BD1 was expressed in ROS(DE3) cells after the induction of 0.4 mM IPTG, as mentioned in Section 3.4.4. For the purification process, cells were lysed using French Press. After clarification, the lysate was applied to a cobalt resin to separate His-BRD4 BD1. His-tag was removed by TEV protease cleavage overnight. The cleaved sample was then purified by size exclusion chromatography to have pure BRD4 BD1. After the purification procedure a yield of 100 mg/L was obtained. The mass and purity of the protein were verified by SDS–PAGE and mass spectrometry (Figure 6.5).
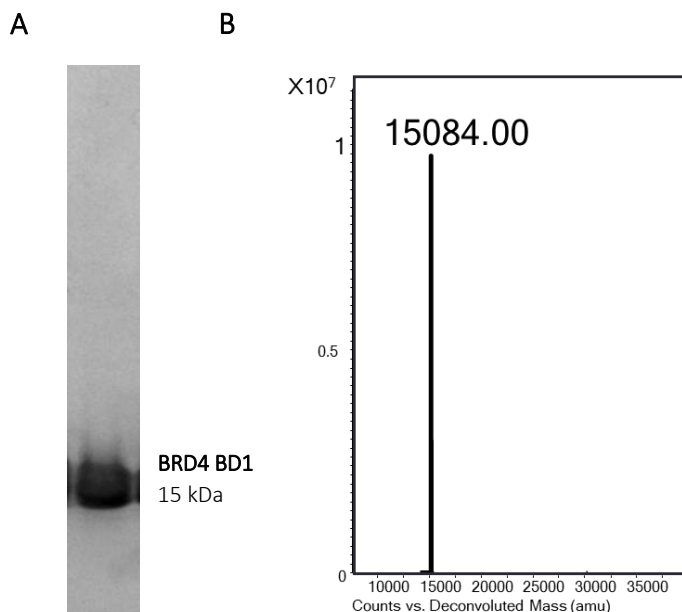
Figure 6.5. A. SDS polyacrylamide gel electrophoresis of BRD4 BD1 after purification. **B.** Electrospray ionization (ESI) of BRD4 BD1 (15 kDa).

6.4. Identification of novel small molecules disrupting the BRD4 BD1 water network

6.4.1. Study of solvation preferences of BRD4 BD1

Initially, waters of reference were determined as mentioned in Section 3.4.1.1. The coordinates of the resulting 15 waters of reference were compared with all PDB files that contained BRD4 BD1 in holo-conformation. From this data the fractional conservation (F, Equation 3.8) and the displacement fraction (D, Equation 3.9) were calculated (Section 3.4.1.1). Five conserved water molecules with F > 0.9 were found and selected as the most conserved ones (Table 6.1). Four of these water molecules have been found to be preserved among all bromodomains family and studied previously[183]. The structure of BRD4 BD1 with the five water molecules is depicted in Figure 6.2.

Table 6.1. Fractional conservation (F) and displacement fraction (D) of the different waters from the active site

HOH	F	D
1	0.959	0.041
2	0.934	0.066
3	0.942	0.058
4	0.942	0.058
5	0.058	0.942
6	0.694	0.306
7	0.488	0.512
8	0.198	0.802
9	0.041	0.959
10	0.182	0.818
11	0,041	0.959
12	0.074	0.926
13	0.066	0.933
14	0.975	0.024
15	0.752	0.248

The values with F higher than 0.9 are highlighted in blue.

6.4.2. Using MDMix results to predict water displacement

Displacement of the five most conserved waters was studied using MDMix[187] with the aim of finding small organic solvents able to displace them. For that, a former member of the group, Dr Daniel Álvarez, performed 20 ns molecular dynamic simulation with solvent-water (1:4) mixtures. Seven different solvents were used: acetonitrile, 2-chloroethanol, ethanol, isoxazole, acetamide, methyl sulphonamide and pyrazole. MDMix predicted free energy values (ΔG) of the different atom types from each solvent in the same coordinates of the conserved water molecules previously established (Table 6.2). Those results illustrated the preference of the different solvents to be in the same coordinates as waters while displacing them. Considering that when $\Delta G < 0$, the presence of the moiety of the solvent is favourable, all water molecules

could be favourably displaced by at least one solvent. Curiously, acetamide, a part from displacing conserved water 3, was the only solvent able to displace water 14.

Table 6.2. Free energy values of MDMix solvent moieties

Solvent moieties	Equivalent pharmacophore feature	HOH coordinates				
		1	2	3	4	14
ANT_C	Hydrophobic	0.26	-0.06	1.43	0.87	nan
ANT_N	HBA	2.07	0.43	0.41	0.67	nan
CLE_Cl	Hydrophobic	1.17	-0.88	-0.63	-1.06	nan
CLE_OH	HBD + HBA	0.55	0.07	0.31	-1.08	2.06
ETA_CT	Hydrophobic	0.05	-0.44	-0.15	-0.09	nan
ETA_OH	HBD + HBA	0.05	0.03	-0.12	-0.04	0,59
ISX_CA	Aromatic	-0.19	-0.18	0.02	0.64	nan
ISX_NO	HBA	1.43	-0.75	0.17	-0.61	nan
MAM_CT	Hydrophobic	1.80	-0.89	0.69	0.01	0.79
MAM_N	HBD	1.44	-0.83	-0.51	-0.15	nan
MAM_O	HBA	nan	-0.48	0.40	-0.11	-0.38
MSU_N	HBD	nan	-0.62	0.40	1.47	nan
MSU_O	HBA	nan	0.85	Nan	1.48	nan
PYZ_CA	Aromatic	0.10	0.21	0.20	0.55	nan
PYZ_N	HBA	1.15	-0.13	-0.41	-0.51	nan
PYZ_NH	HBD	0.33	0.28	-0.98	-1.15	nan

Free energy values (ΔG) in kcal/mol of the different moieties for each solvent tested with MDMix. HOH indicates the coordinates of the different water molecules where ΔG of the moieties were calculated. When $\Delta G < 0$ the presence of the moiety is favourable, whereas nan indicates that no solvent moiety could go in that coordinates. Abbreviations: ANT, acetonitrile; CLE, 2-chloroethanol; ETA, ethanol; ISX, isoxazole; MAM, acetamide; MSU, methyl sulphonamide; PYZ, pyrazole; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor.

In order to further study acetamide behaviour, CPPTRAJ from AmberTools12 was used to extract different frames of the simulation. The information provided allowed to determine how acetamide interacted when displacing water molecules and its exact position (Figure 6.6). In particular, when displacing water 3 and 14, acetamide would make and hydrogen bond with the backbone of methionine 105.

The interactions produced by acetamide were subsequently used to perform a docking-based virtual screening.
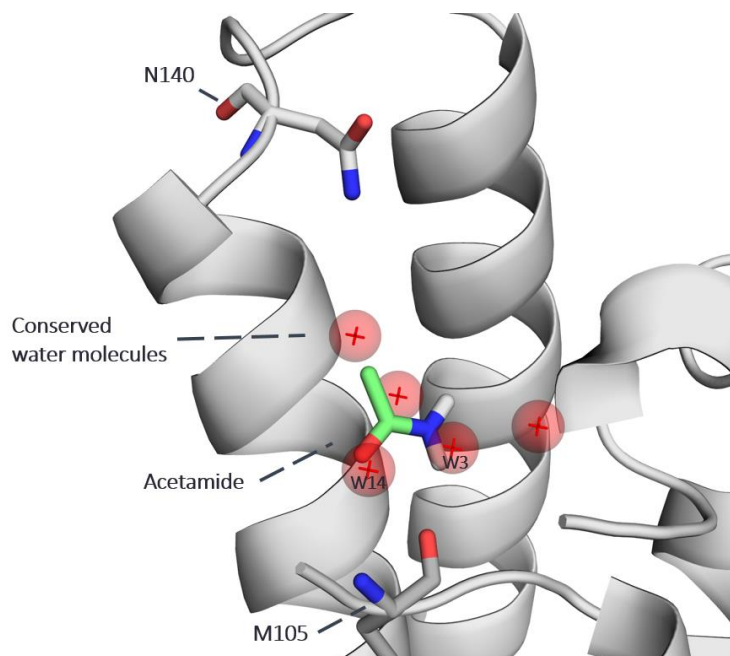
Figure 6.6. Acetamide molecule interacting with BRD4 BD1. Acetamide can interact with methionine 105 of BRD4 BD1 when displacing waters 14 and 3. Conserved water molecules are superposed and shown in red spheres.

6.4.3. Virtual screening to find binders with the ability to displace water molecules of BRD4 BD1

Virtual screening campaign started with the application of tethered docking in a library of approximately 6 million compounds. Tethered docking is used to enforce partial binding modes, in other words, one forced binding mode for all the ligands is established. In our particular case, this binding mode consisted on the acetamide disposition to displace waters 14 and 3. Consequently, the whole library was screened in order to keep only the molecules that contained an acetamide functional group in their structure. Approximately 200000 molecules were obtained, which were screened applying docking with a pharmacophoric restraint. The latter consisted on the hydrogen bond acceptor that BRD4 BD1 ligands perform with ASN140. Compounds with a "SCORE.INTER" (general docking score) and the "SCORE.RESTR.PHARMA" (pharmacophoric restraint score) of less than

-17.0 kJ/mol and 0.4 kJ/mol, respectively, were selected. As a result, 4026 compounds accomplished the scoring threshold, which were clustered by similarity (95% similarity) to finally obtain 2837 compounds. The following step implied performing DUck, which is a particular implementation of steered molecular dynamics that measures the work needed to break the main hydrogen bond between a protein and a binder[58]. The higher the work needed to break this hydrogen bond, the better the ligand is considered to be. Interaction with ASN140 was again chosen to apply DUck. 617 compounds passed the threshold of 6 kcal/mol, which were visually inspected to finally select 40 drug-like compounds. The final selection of compounds were bought to test them experimentally, validating our computational approach. The whole virtual screening protocol is illustrated in Figure 6.7.
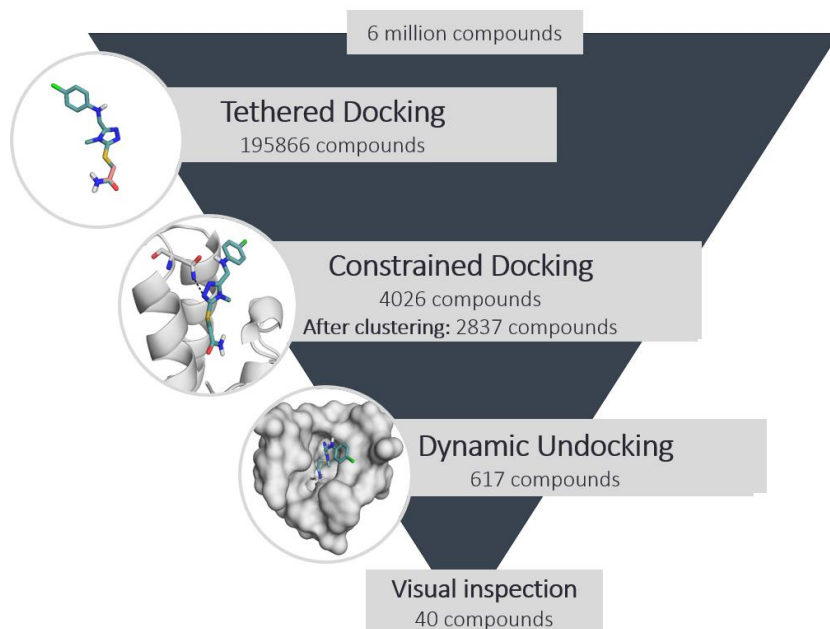
Figure 6.7. Virtual screening protocol to find compounds able to displace water molecule by mimicking acetamide interactions.

6.4.4. Differential scanning fluorimetry of virtual screening hits

DSF has been widely used in bromodomains field to assess compound binding[105,106,188], being the thermal profile of BRD4 BD1 previously described (44 ºC)[189]. Experiments were performed as described in Section 3.4.7.1. In each experiment the already described binder (+)-

JQ1[105] was tested to ensure the experiment reproducibility. Melting temperatures (Tm) of BRD4 BD1 and BRD4 BD1 in complex with compounds were determined in triplicate. In order to be considered significant, thermal shifts (ΔTm) of the protein-ligand complexes (ΔTm = Tm protein-ligand complex − Tm protein) had to be at least twice the standard deviation of the Tm of BRD4 BD1. On average, the threshold was of 0.5 ºC. The 40 compounds from virtual screening were tested at a final concentration of 200 µM. A summary of the results is illustrated in Figure 6.8 and all ΔTm are listed in Table 6.3.
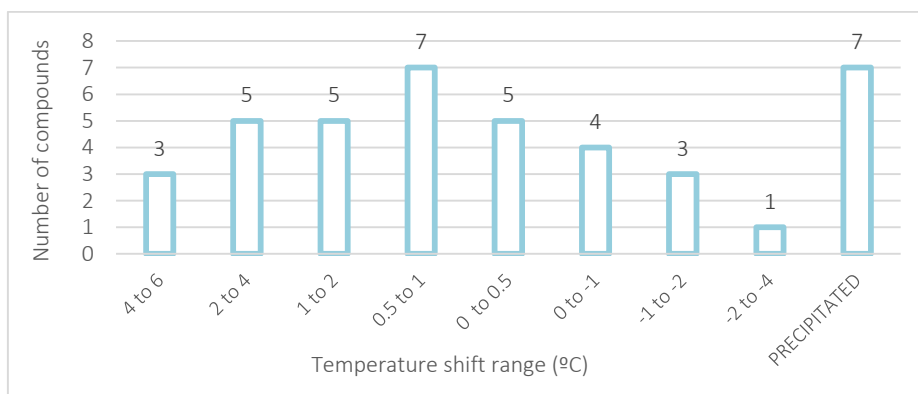
Figure 6.8. Summary of DSF results to evaluate BRD4 BD1 interaction with virtual screening hits at 200 µM. Each column represents the number of compounds that produced a temperature shift in a particular range.

Despite not finding any compound that produced a thermal shift similar than (+)-JQ1 (ΔTm 10 ºC, K_d 50 nM[105]), eight of them could stabilize the protein more than 2ºC. Derivative melting curves of the best binders and (+)-JQ1 are shown in Figure 6.9. In total, 20 out of 40 compounds stabilized BRD4 BD1, meaning that 50% of virtual screening hits passed the initial screening. Those binders and the compounds that precipitated, were rescreened at 50 µM. With that concentration, similar results were obtained and 17 compounds passed the threshold of 0.5 ºC. It has been described that thermal shifts correlate well with K_ds determined by ITC for the BET bromodomains[105]. Taking this into account, the K_d from our binders would be similar than midazolam, which crystal structure bound to BRD4 BD1 was elucidated even no K_d had been described[188]. Considering this information and due to the fact

that a crystal structure of our binders would be crucial to verify our hypothesis, X-ray crystallography experiments were performed.

Table 6.3. DSF results of virtual screening compounds at 200 μM

Compound ID	ΔTm [200 μM]	SD	Compound ID	ΔTm [200 μM]	SD
BRD4-MMC1	0.14	0.33	BRD4-MMC21	0.71	0.47
BRD4-MMC2	-0.01	0.30	BRD4-MMC22	-1.14	0.01
BRD4-MMC3	0.33	0.63	BRD4-MMC23	1.19	0.97
BRD4-MMC4	*Precipitated*		BRD4-MMC24	4.34	0.21
BRD4-MMC5	-0.07	0.78	BRD4-MMC25	*Precipitated*	
BRD4-MMC6	-0.48	-	BRD4-MMC26	*Precipitated*	
BRD4-MMC7	0.78	1.11	BRD4-MMC27	*Precipitated*	
BRD4-MMC8	1.38	0.47	BRD4-MMC28	-1.10	0.56
BRD4-MMC9	2.24	0.37	BRD4-MMC29	1.39	0.47
BRD4-MMC10	0.97	0.54	BRD4-MMC30	2.28	0.57
BRD4-MMC11	*Precipitated*		BRD4-MMC31	0.02	0.97
BRD4-MMC12	*Precipitated*		BRD4-MMC32	-0.57	0.42
BRD4-MMC13	6.16	0.26	BRD4-MMC33	*Precipitated*	
BRD4-MMC14	0.89	0.40	BRD4-MMC34	0.89	0.91
BRD4-MMC15	4.05	0.10	BRD4-MMC35	1.91	1.26
BRD4-MMC16	2.35	1.44	BRD4-MMC36	3.27	1.09
BRD4-MMC17	0.29	0.52	BRD4-MMC37	0.25	0.86
BRD4-MMC18	3.06	2.10	BRD4-MMC38	0.56	0.52
BRD4-MMC19	-2.19	0.94	BRD4-MMC39	1.41	0.48
BRD4-MMC20	-1.88	0.70	BRD4-MMC40	0.86	0.53

Abbreviations: ΔTm, thermal shift; SD, standard deviation.

A — Thermal shift of BRD4 BD1 (dark blue) and BRD4 BD1 with (+)-JQ1 10 µM (light blue) and with 200 µM (blue)

B — Thermal shift of BRD4 BD1 (dark blue) and BRD4 BD1 with BRD4-MMC13 200 µM (light blue)

C — Thermal shift of BRD4 BD1 (dark blue) and BRD4 BD1 with BRD4-MMC15 200 µM (light blue)

D — Thermal shift of BRD4 BD1 (dark blue) and BRD4 BD1 with BRD4-MMC18 200 µM (light blue)
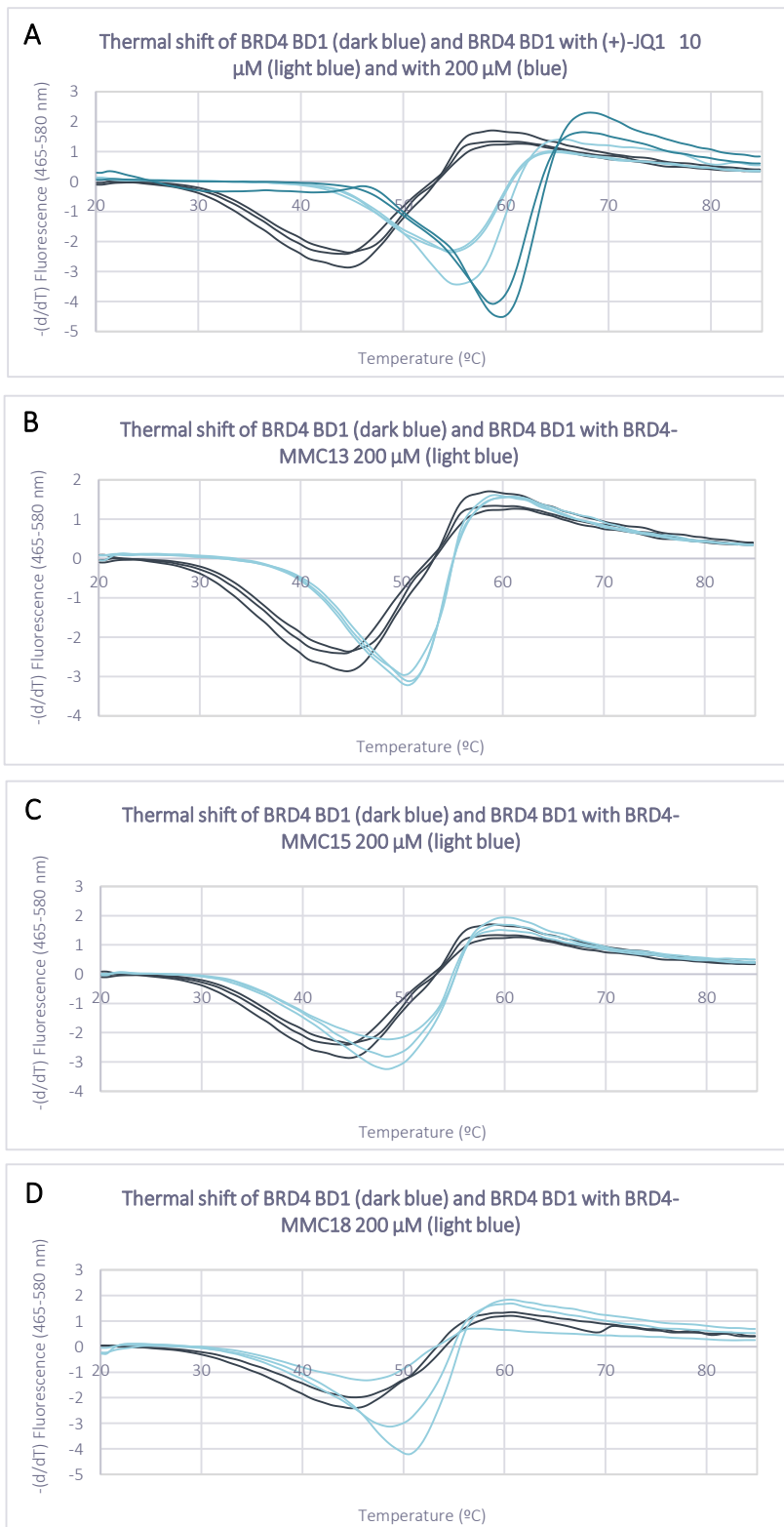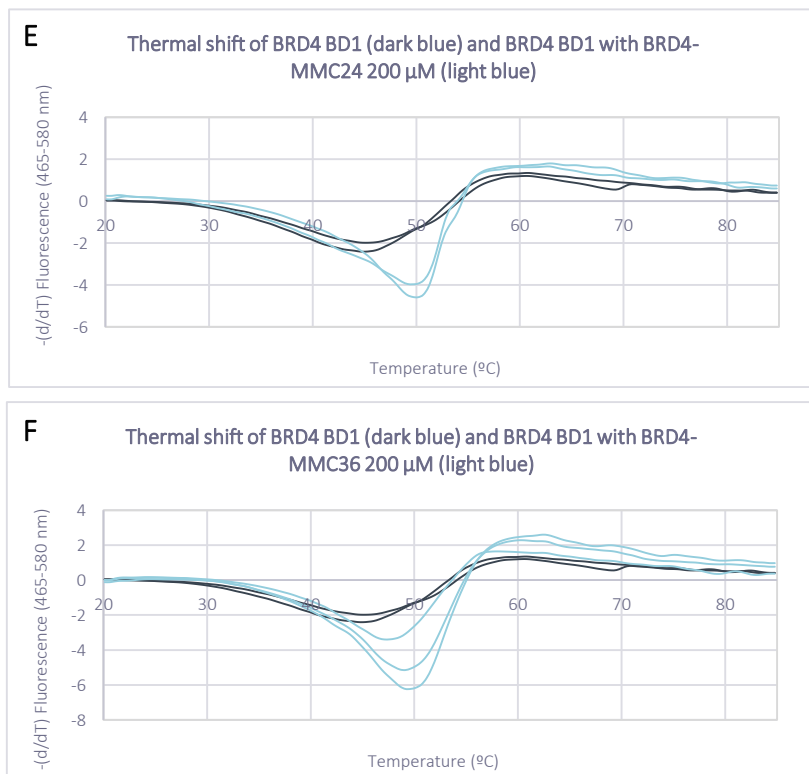
Figure 6.9. DSF graphics of virtual screening binders. DSF graphics comparing melting temperatures of BRD4 BD1 with BRD4 BD1 in complex with 10 µM and 200 µM of (+)-JQ1 (ΔTm 10 ºC and 15 ºC, respectively) (**A**), 200 µM of BRD4-MMC13 (ΔTm 6 ºC) (**B**), 200 µM of BRD4-MMC15 (ΔTm 4 ºC) (**C**), 200 µM of BRD4-MMC18 (ΔTm 3 ºC) (**D**), 200 µM of BRD4-MMC24 (ΔTm 4 ºC) (**E**), and 200 µM of BRD4-MMC36 (ΔTm 3 ºC) (**F**).

6.4.5. X-ray crystallography of BRD4 BD1 in complex with virtual screening hits

6.4.5.1. BRD4 BD1 crystallography in complex with virtual screening hits

X-ray crystallography trials of BRD4 BD1 in complex with compounds from virtual screening were performed through a 3 months placement in the laboratory of Prof Frank von Delft and Prof Panagis Filippakopoulos (SGC, Oxford).

Initially soaking experiments were attempted, by producing optimization plates that contained BRD4 BD1 crystals. Resulting crystals were used for soaking the 4 best positive compounds from DSF and the positive controls (+)-JQ1[105] and RVX-208[106]. Final concentrations of 23 mM for our compounds and 11.5 mM for the positive controls were tested.

Potential binders were not only diluted in DMSO, but also in ethylene glycol, which is a cryoprotectant that decreases compound solubility. In the case of positive controls, only stocks in DMSO were used. Soaked crystals were harvested and screened by X-ray crystallography after 24h and two weeks. Even though extending the time of harvesting usually helps to form the protein-ligand complex, neither the compounds from virtual screening, nor the positive controls were observed interacting with BRD4 BD1. Therefore, only apo structures were elucidated from electron diffraction patterns. All BRD4 BD1 protein in crystals have usually the same packing with a space groups of P 21 21 21. The fact that soaking did not work with (+)-JQ1 made us reconsider the soaking strategy and co-crystallization was performed instead.

Co-crystallization screenings of the 4 best positive compounds from DSF and the positive controls were performed, using final concentrations of 10 mM and 5 mM, respectively, all in 10% DMSO. For all of them LFS6 screening conditions were tested. Crystals grew within 24-48h in similar mixture conditions, which were considered to select one condition to design an optimization plate to carry out a co-crystallization screening. The condition selected consisted of: 20% PEG 3350, 10% ethylene glycol, 0.2 M sodium fluoride and 0.1 M bis-tris propane with a gradient of 12 different pH ranging from 6.5 to 8.7. Consequently, in an 8x12 well plate, 8 compounds could be tested in 12 different conditions. Co-crystallization screening was applied to the 40 compounds coming from the computation approach, using compound concentrations of 10 mM or 20 mM (10% or 20% DMSO). Unfortunately no crystal resulting from protein-compound complex was elucidated.

A second round of co-crystallization experiments was performed. This time, the 10 positive compounds from DSF were directly co-crystallized using different screening conditions already available in the protein crystallography facilities at SGC referred as: HIN3 and MIDAS. Besides, in order to increase the chances of obtaining crystals of compounds from the water displacement approach, BCS and JCSG7 screening plates were tested for co-crystallization for 6 and 3 compounds, respectively. In this second round of co-crystallization plates compounds were tested at concentrations ranging from 10 mm to 20 mM (10% DMSO). The rest of conditions from crystallization experiments can be found in Section 3.4.7.3.1. Crystals obtained from crystallization trials were flash cooled prior to X-ray diffraction studies.

6.4.5.2. X-ray diffraction of obtained crystals

Dr Tobias Krojer (SGC, Oxford) was responsible of the X-ray data collection and structure elucidation steps. X-ray diffraction data was collected on beamlines I03 and I04-1 at the Diamond Light Source (UK's national synchrotron). Data was integrated with XDS[190] and scaled with AIMLESS[191] as part of the XIA2 and autoPROC auto-processing pipelines[192,193]. Initial refinement and map calculation was carried out with DIMPLE[194] using a version of BRD4 BD1 (PDB code: 3MXF[195]) as starting model. Refinement and model building was performed with REFMAC[196] and COOT[197], respectively.

If a compound had been bound to BRD4 BD1 in the crystal, it would have been rapidly identified by inspection of differences between the electron density maps of the complex and the protein alone. Detailed examination of these differences and fitting of molecular models into the density enable the identification of the small molecule and its binding mode[67]. However, weakly bound ligands may cause weak signals that sometimes are considered as noise. Being in this situation, PanDDA analysis (Pan-Dataset Density Analysis[198]) was also used. This methodology was developed in the group of Prof Frank von Delft. PanDDA analysis takes multiple data sets from electron density maps of the apo protein (around 40) making an average of this state, which is subtracted to the electron density map of the protein bound to the ligand. Therefore, the electron density map of the ligand can clearly be determined. Despite all efforts, no structure of any compound bound to BRD4 BD1 protein was elucidated. Picture of a crystal in the loop of BRD4 BD1 prepared in mixture with one of the best compounds (BRD4-MMC18), and the electron density map of that crystal mounted is shown in Figure 6.10.

A

B
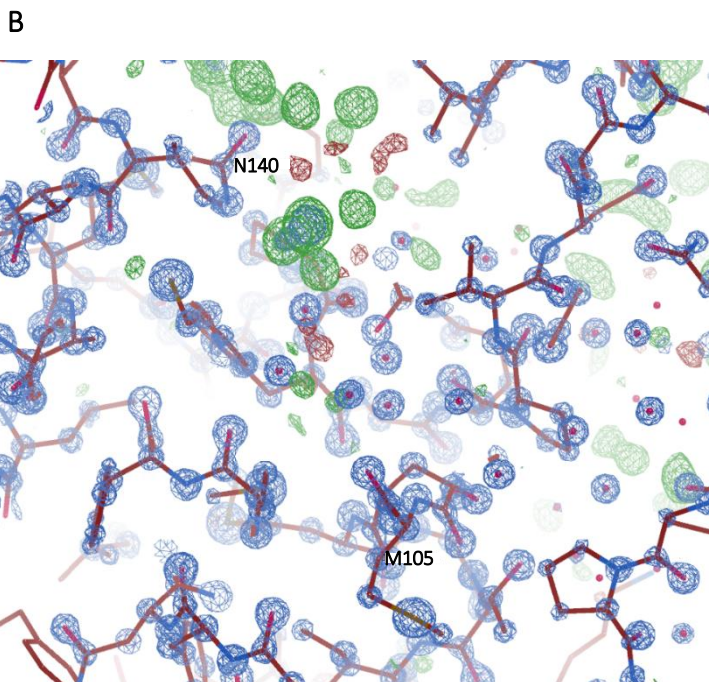
Figure 6.10. A. Loop with resulting crystal of mixing BRD4 BD1 within BRD4-MMC18 in LFS6 screen. **B.** Electron density map of crystal from A. It can be observed the electron density map of BRD4 BD1. In green, the unresolved electron density map is shown, which in this case, it does not fit with BRD4-MMC18. Residues that should interact with BRD4-MMC18, asparagine 140 (N140) and methionine 105 (M105), are labelled.

6.4.6. Docking of best compounds with network of water molecules.

In order to predict if compounds from virtual screening could still bind to BRD4 BD1 without displacing the network of water molecules, docking-based virtual screening of BRD4 BD1 with the 5 conserved waters was carried out for the 40 molecules purchased. Taking into account the same threshold as in constrained docking from Section 6.4.3 ("SCORE.INTER" and "SCORE.RESTR.PHARMA" less than -17 kJ/mol and 0.4 kJ/mol, respectively), 40% of purchased molecules were predicted to bind inefficiently to BRD4 BD1 without displacing water molecules. However, docking does not consider the cost of displacing the network of water molecules. Hence, structural information would still be of utmost importance in this project. Docking poses of some of the best molecules from DSF with and without the water network are depicted in Table 6.4.

Table 6.4. Docking poses of DSF best molecules with and without the water network
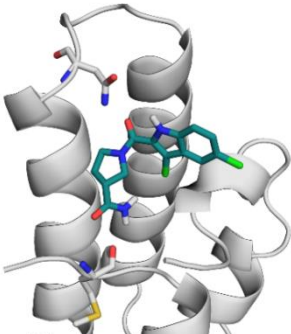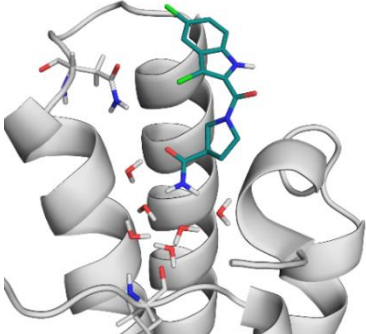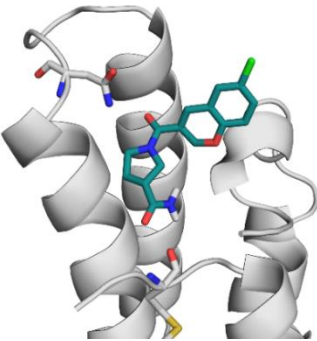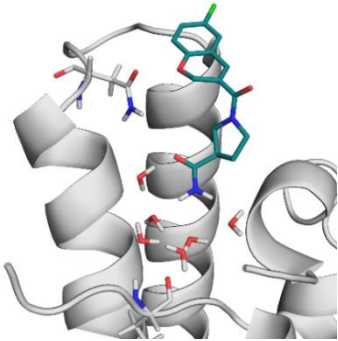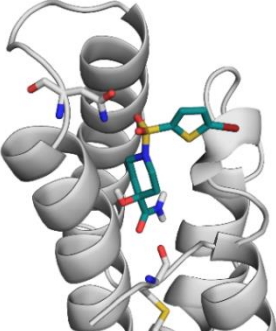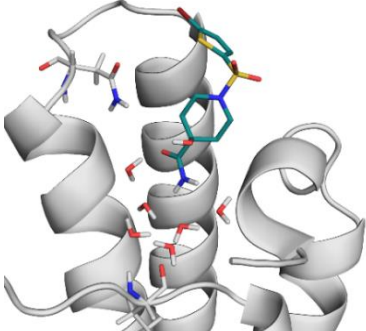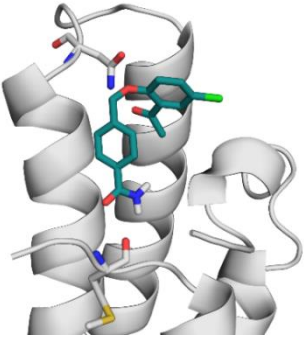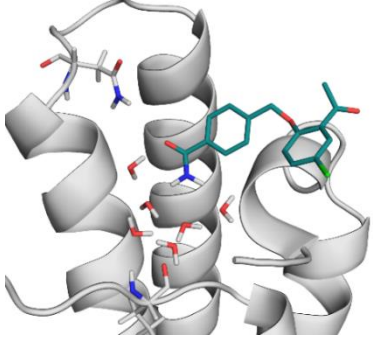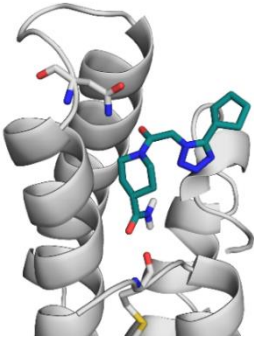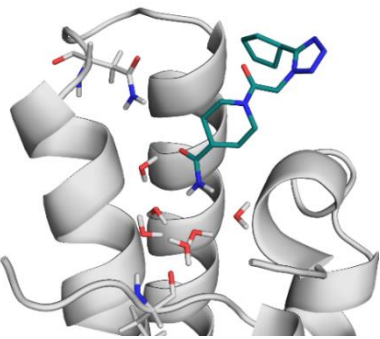
Water-displacement docking	Docking with water network

BRD4-MMC13 ΔTm (200 µM): 6.16 ºC

SCORE.INTER: - 20.345 kJ/mol
SCORE.RESTR.PHARMA: 0.003 kJ/mol

SCORE.INTER: - 20.299 kJ/mol
SCORE.RESTR.PHARMA: 0.152 kJ/mol

BRD4-MMC24 ΔTm (200 µM): 4.34 ºC

SCORE.INTER: - 18.931 kJ/mol
SCORE.RESTR.PHARMA: 0.186 kJ/mol

SCORE.INTER: - 16.116 kJ/mol
SCORE.RESTR.PHARMA: 0.065 kJ/mol

BRD4-MMC15 ΔTm (200 µM): 4.05 ºC

SCORE.INTER: - 18.245 kJ/mol
SCORE.RESTR.PHARMA: 0.000 kJ/mol

SCORE.INTER: - 16.645 kJ/mol
SCORE.RESTR.PHARMA: 0.791 kJ/mol

Table 6.4 (cont.). Docking poses of DSF best molecules with and without the water network

Water-displacement docking	Docking with water network
BRD4-MMC18 ΔTm (200 µM): 3.06 ºC	

SCORE.INTER: - 17.584 kJ/mol	SCORE.INTER: - 18.316 kJ/mol
SCORE.RESTR.PHARMA: 0.158 kJ/mol	SCORE.RESTR.PHARMA: 0.017 kJ/mol

BRD4-MMC9 ΔTm (200 µM): 2.24 ºC

SCORE.INTER: - 22.560 kJ/mol	SCORE.INTER: - 17.636 kJ/mol
SCORE.RESTR.PHARMA: 0.258 kJ/mol	SCORE.RESTR.PHARMA: 0.713 kJ/mol

Docking scores that would not pass the threshold are highlighted in red.

6.4.7 Summary and future perspectives (Identification of novel small molecules disrupting the BRD4 BD1 water network)

It has been evidenced that around two-thirds of interactions between a ligand and a protein rely on the involvement of at least one water molecule[199,200]. Indeed, the prediction of the displacement or the interaction with protein water molecules has been of significant relevance in the structure-based drug design field[201]. Due to the network of conserved water molecules in bromodomains, solvation of this protein family has been widely studied[183]. In the case of BET family,

202

several articles that involve water-mediated hydrogen bonds with this network have been reported[202,203], as well as water rearrangement[204]. During the present thesis, mixed solvent molecular dynamics simulations performed to predict water displacement of BRD4 BD1 have been published by other scientists. In that case, of the five solvents tested, only a solvent mixture that contained acetate and methyl ammonium was predicted to be able to displace conserved water molecules[205]. In contrast, our studies suggested that different types of solvents could displace at least one conserved water. An explanation for that could be that in our simulations organic solvents supposed 20% of the total amount of solvent, whereas they employed 5% organic/water mixtures. We selected 20% as a compromise between the amount of probe molecules needed to ensure efficient sampling, while remaining the environment predominantly aqueous and capturing the water solvation effect[187].

MDMix information allowed to perform rational design of compounds with the potential ability to displace the water network. 40 compounds were selected from the virtual screening to be studied biophysically. Hit rate from the primary screening was of 50% when testing the compounds at 200 µM, which can be considered a high success hit rate taking into account the expected for other structure-based approaches[164]. Information about the key interaction for BRD4 BD1 ligands and the deep knowledge of the system available in literature have promoted these positive results. Correlation between affinity of compounds and thermal shifts has been described in bromodomains, being our positive hits considered as weak binders[195]. Despite being more challenging, compounds presenting similar thermal shifts were crystallized in complex with BRD4 BD1. Using soaking to crystallize BRD4 BD1 with its ligands did not work, since the symmetry packing of the protein hinders the ligand access to the binding site. A prove of that is that co-crystallization has always been the method of choice to crystallize BRD4 BD1 ligands, even in structures with low affinity compounds[206]. In spite of all the struggle to obtain a crystal structure, none of the compounds from the virtual screening were co-crystallized with BRD4 BD1.

Docking-based experiments with the network of water molecules were performed as an alternative to predict if the compounds could still bind to BRD4 BD1 without displacing waters 3 and 14. Although 40% of the compounds presented to bind inefficiently, obtaining a structure would still be of utmost importance for the project. The fact that it was impossible to crystallize any compound from this approach, made us reconsider other techniques such as HSQC-NMR. This technique could be used to study the novel interactions that the compounds would perform with a particular amino acid backbone (MET105) once displacing the preserved water molecules. Even though different challenges would be faced when trying this experiments (such as protein labelling, proper visualization of chemical shifts, etc.), BRD4 BD1 NMR structure has already been assigned[207].

Interestingly, this year, a compound able to displace the network of water molecules of BRD4 BD1 with a trifluoromethyl group has been disclosed. Remarkably, this compound appeared to be selective for the BRD4 BD1 among the rest BET family members[186]. Being a hydrophobic group employed to displace this water molecules could mean that our compounds might not be able to place the acetamide moiety were expected. Nonetheless, further studies should be done, as aforementioned. Proving that the network of water molecules in bromodomains can be displaced in order to promote selectivity does not cease being encouraging information in the bromodomains field.

6.5. Biophysical characterization of BRD4 BD1 optimized fragments

6.5.1. Preliminary results

Two former PhD students in our Lab, Dr Serena Piticchio and Dr Moira Rachman, developed an automated platform based on iterative virtual screening using massive commercial libraries with the aim of evolving fragments to generate high-potency drug-like compounds. Concisely, an initial fragment with a known or predicted binding mode is provided as an input for the platform. With that information the protocol searches at different virtual libraries for compounds that are chemically related and slightly bigger in size. These are then tethered docked to the target

protein to identify those that are complementary (maintaining the key interaction(s) of the starting compound). Subsequently, DUck[58] is applied to filter out false positives. The resulting top candidates are selected and the process is repeated until the fragments are grown to drug-sized compounds.

This platform has been applied in two different fragment growing approaches that envisaged to find potent binders for BRD4 BD1 (Figure 6.11). One on hand, the approach led by Dr Serena Piticchio consisted on finding novel scaffolds to interact with BRD4 BD1 and, consequently, validate the method developed. In this particular case, the fragment of reference had a known binding mode (PDB code: 4LR6) and a described IC_{50} of 33 µM[208]. The computational platform was applied for 4 iterations to grow the fragment of reference. 30 compounds resulting from the different iterations of the applied platform were purchased or synthetized to be characterized biophysically (Sections 6.5.2, 6.5.3 and 6.5.4). This approach will be referred posteriori as "fragment evolution platform". On the other hand, the second approach consisted on predicting the binding mode of a low-affinity fragment, the structure of which could not be elucidated, by growing it to a more potent compound. For these approach, a fragment in literature reported to bind to BRD4 BD1 with no binding mode characterized was selected[209]. First, different potential poses were proposed by studying already published BRD4 BD1 structures that contained a similar ligand that the fragment of reference. These poses were then validated by DUck[58]. Two binding modes were selected and the automated platform was applied. 18 compounds were purchased to be biophysically characterized (Sections 6.5.2 and 6.5.4). This approach was led by Dr Moira Rachman and will be referred posteriori as "binding mode prediction platform".
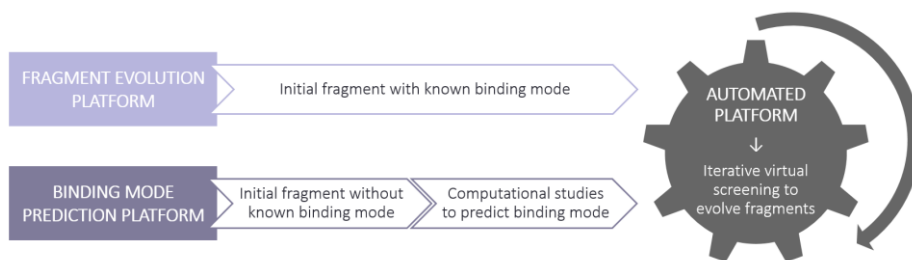
Figure 6.11. Fragment growing approaches applying the automated platform.

6.5.2. Differential scanning fluorimetry of optimized fragments

Binding of evolved fragments coming from both computational platforms was initially assessed by DSF. Experiments were performed as described in Section 3.4.7.1. As in DSF experiments of BRD4 BD1 mentioned in Section 6.4.4, the known binder (+)-JQ1[105] was tested to ensure the experiment reproducibility. Threshold for binding were also considered to be the thermal shift of the protein-ligand complexes that were twice the standard deviation of the replicas with the melting temperature of BRD4 BD1. On average, the threshold was of 0.5 ºC. The 30 compounds coming from the fragment evolution platform (including the reference compound, SPF-REF) were tested at final concentrations of 100 µM, 50 µM and 10 µM. The 18 compounds coming from the binding mode prediction platform (including the reference compound, BRD4-MREF) were tested at final concentrations of 100 µM and 10 µM.
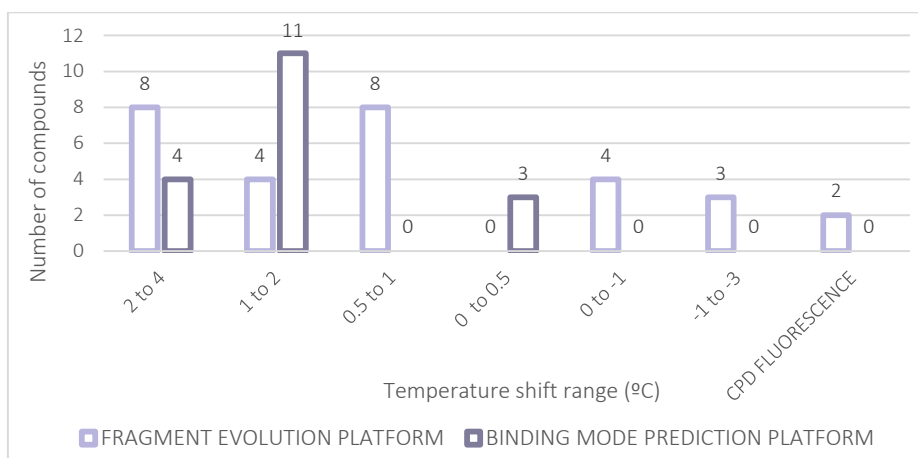
Figure 6.12. Summary of DSF results from the automated platform. BRD4 BD1 interaction was evaluated with 10 µM of evolved fragments from fragment evolution platform (light purple) and binding mode prediction platform (dark purple). Each column represents the number of compounds that produced a temperature shift in a particular range. Abbreviations: CPD, compound.

A summary all evolved fragments tested at 10 µM is depicted in Figure 6.12 and all ΔTm are listed in Table 6.5. Considering as binding thermal shifts higher than 0.5 ºC, initial hit rates of 73% and 83% were determined for fragment evolution and binding mode prediction platforms, respectively. Two compounds from the fragment evolution platform, presented intrinsic fluorescence. To confirm that, DSF control

experiments were performed by preparing the same mixture without the protein. Derivative melting curves of representative fragments from both approaches are shown in Figure 6.13.

Table 6.5. DSF results of fragment evolution platform and binding mode prediction platform evolved fragments at 10 µM

FRAGMENT EVOLUTION PLATFORM			BINDING MODE PREDICTION PLATFORM		
Compound ID	ΔTm [10 µM]	SD	Compound ID	ΔTm [10 µM]	SD
SPF-REF	4.55	0.62	BRD4-MREF	1.61	1.06
SPF1	0.89	0.22	BRD4-MR1	2.24	0.55
SPF2	-0.09	0.39	BRD4-MR2	1.33	0.65
SPF3	0.68	0.28	BRD4-MR3	1.49	0.49
SPF4	0.83	0.27	BRD4-MR4	1.47	0.12
SPF5	1.83	0.27	BRD4-MR5	1.62	0.32
SPF6	0.68	1.08	BRD4-MR6	3.22	0.32
SPF7	Intrinsic fluorescence		BRD4-MR7	2.69	0.14
SPF8	1.59	0.24	BRD4-MR8	1.28	0.25
SPF9	-1.04	0.23	BRD4-MR9	3.71	1.30
SPF10	0.88	0.19	BRD4-MR10	1.34	0.80
SPF11	-1.91	0.36	BRD4-MR11	1.86	0.78
SPF12	-0.09	0.37	BRD4-MRC1	1.06	0.54
SPF13	-0.95	0.62	BRD4-MRC2	1.75	0.22
SPF14	0.85	0.12	BRD4-MRC3	0.04	0.98
SPF15	Intrinsic fluorescence		BRD4-MRC4	0.47	0.49
SPF17	2.64	0.47	BRD4-MRC5	0.36	0.89
SPF18	3.41	0.38	BRD4-MRC6	1.64	0.57
SPF19	3.09	1.10			
SPF20	-0.31	0.41			
SPF21	1.70	1.47			
SPF22	2.36	0.52			
SPF23	2.88	0.16			
SSR1	1.36	0.71			
SSR2	0.75	0.25			
SSR3	-0.40	0.80			
SSR4	3.32	0.40			
SSR6	4.02	0.24			
SSR11	-2.35	0.48			
SSR12	0.53	0.91			

A Thermal shift of BRD4 BD1 (blue) and BRD4 BD1 with SPF1 10 μM (purple)

B Thermal shift of BRD4 BD1 (blue) and BRD4 BD1 with SPF5 10 μM (purple)

C Thermal shift of BRD4 BD1 (blue) and BRD4 BD1 with SSR4 10 μM (purple)

D Thermal shift of BRD4 BD1 (blue) and BRD4 BD1 with MR1 10 μM (purple)
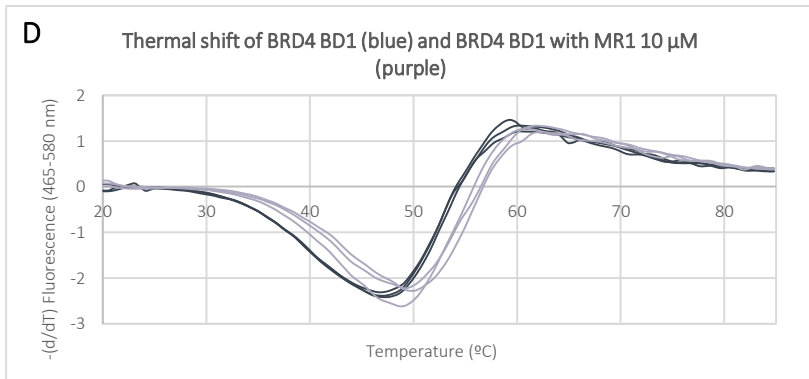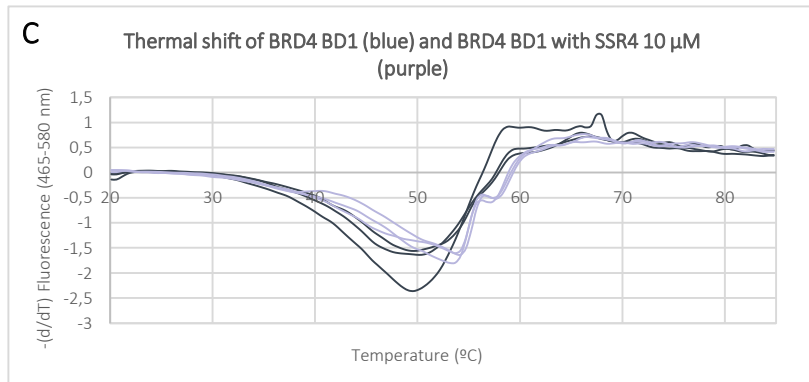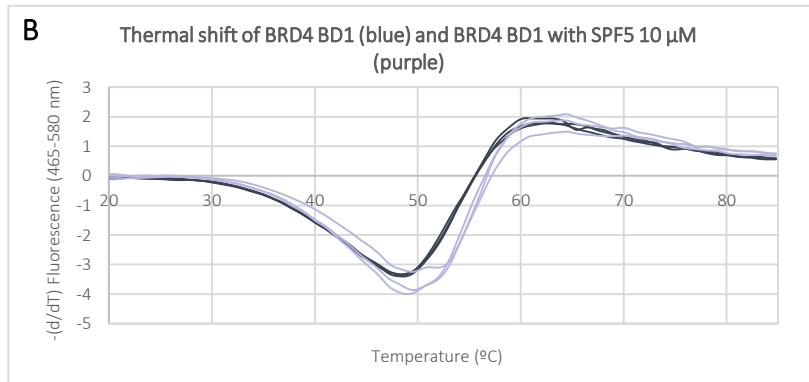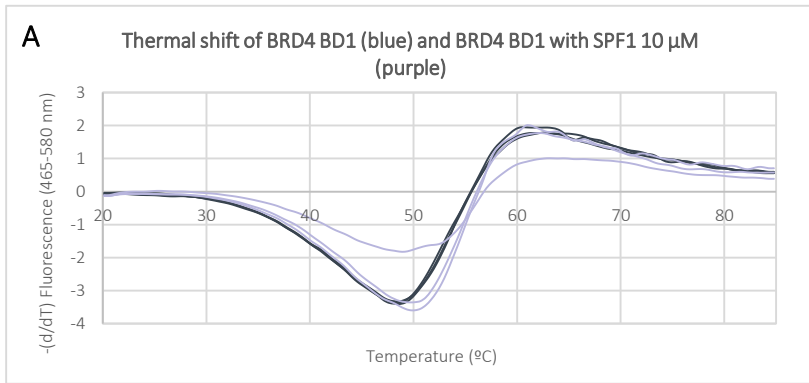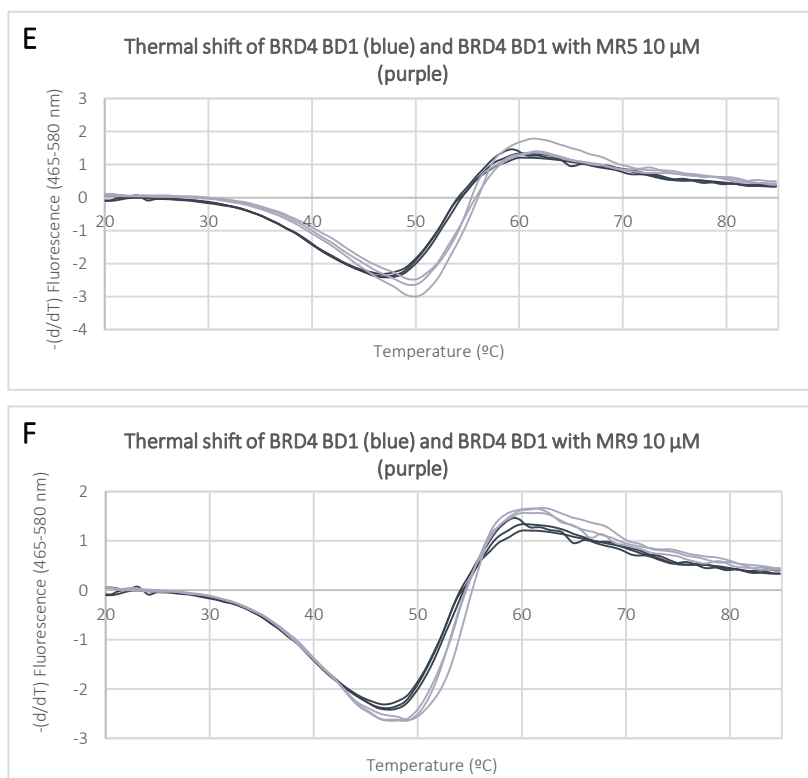
Figure 6.13. DSF graphics of evolved fragments. DSF graphics comparing melting temperatures of BRD4 BD1 with BRD4 BD1 in complex with 10 μM of SPF1 (ΔTm 0.9 °C) (A), SPF5 (ΔTm 1.8 °C) (B), SSR4 (ΔTm 3.3 °C) (C), MR1 (ΔTm 2.2 °C) (D), MR5 (ΔTm 1.6 °C) (E), and MR9 (ΔTm 3.7 °C) (F).

6.5.3. Isothermal titration calorimetry of optimized fragments

ITC experiments were performed with the aim of characterize binding of some of the best fragments and considering the ease of production of BRD4 BD1. Optimized fragments from the fragment evolution platform were studied by ITC. Initially, ITC titrations of the positive controls were performed. Binding of (+)-JQ1 was characterized in order to validate the experiment setup (Section 3.4.7.2). (+)-JQ1 ITC results are illustrated in Figure 6.14 A. K_d of the experiment was 112 nM, being 50 nM the one described in literature[105]. Nevertheless, these experiments were performed at different temperatures (25ºC in house experiments and 15ºC in literature) and K_d depends on temperature. Therefore, ΔG would be a more suitable parameter to compare both results. ΔG described was − 9.64 kcal/mol, whereas in our case ΔG was of − 9.48 kcal/mol. ITC was

also performed for the reference fragment (SPF-REF), which an IC_{50} of 33 µM was already described[208]. K_d obtained was of 12.4 µM and the ΔG of − 6.69 kcal/mol. Full thermodynamic profile is represented in Figure 6.14 and Table 6.6. ITC of five evolved fragments were performed: SPF5 (Iteration 2), SPF22 (Iteration 3), SPF14 (Iteration 4), SPF23 (Iteration 4) and SSR4 (Iteration 4). Several experiments were performed to determine the optimal concentrations to evaluate each fragment. From SPF14 and SPF23 no fitting data could be obtained, since no binding was observed. Thermodynamic profiles from ITC experiments are summarized in Table 6.6 and ITC thermograms are depicted in Figure 6.14. Despite stoichiometry did not need to be fixed in most cases, curves with proper sigmoidal shape could not be observed in ITC titrations. If we had known the fragments K_d, we could have predicted that by calculating the c value[78] (c value = n x [BRD4 BD1] x K_d $^{-1}$, where n is stoichiometry and [BRD4 BD1] the concentration of protein). Calculated c values are also listed in Table 6.6 from the obtained ITC K_ds. Considering K_ds of evolved fragments, titration shape could only be improved increasing concentrations of BRD4 BD1 and fragments, what would suppose reaching to the limits of solubility for fragments and the consumption of massive amounts of protein.

Table 6.6. Thermodynamic profiles obtained from ITC experiments at 25 ºC and thermal shifts from DSF experiments at 10 µM of compound.

Compound	ΔTm (ºC)	K_d (µM)	ΔH^{obs} (kcal/mol)	N	TΔS (kcal/mol)	ΔG (kcal/mol)	c
(+)-JQ1	10.5	0.11 ± 0.001	-7.34 ± 0.05	0.98 ± 0.004	2.14	-9.48	178
SPF–REF	4.5	12.41 ± 2.35	-2.75 ± 0.29	1.00 ± 0.067	3.94	-6.68	2.4
SPF5	1.8	17.15 ± 3.58	-1.70 ± 0.29	1.18 ± 0.135	4.80	-6.50	1.8
SPF22	2.4	3003 ± 504.8	-2154 ± 420	1.00 (Fixed)	-2149.66	-4.34	0.01
SSR4	3.3	6.06 ± 0.97	-1.91 ± 0.07	1.02 ± 0.022	4.98	-6.89	6.6

A. (+)-JQ1
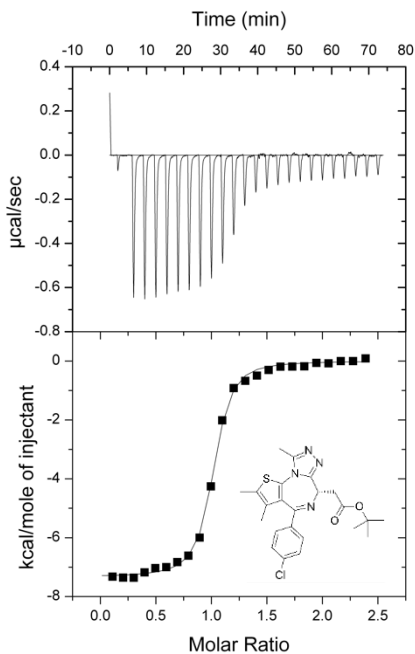
B. Reference Fragment
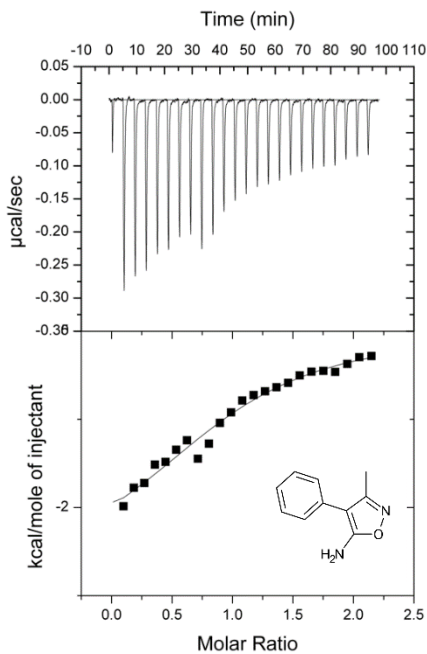
C. SPF5
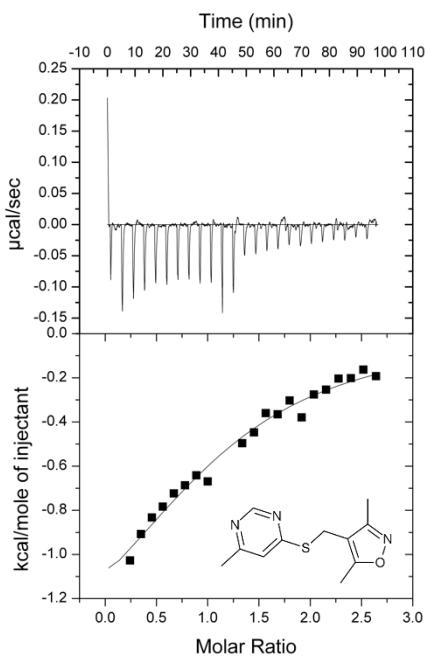
D. SPF22
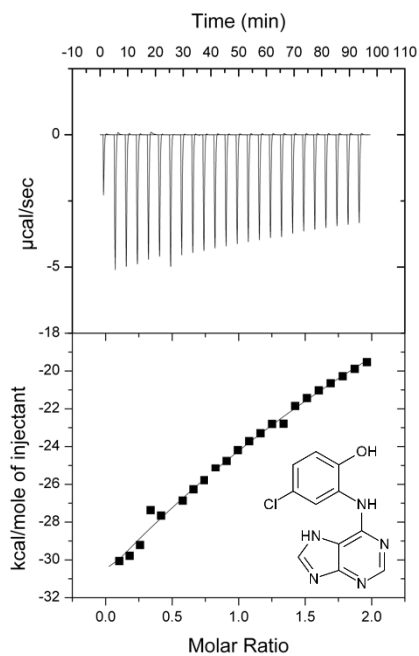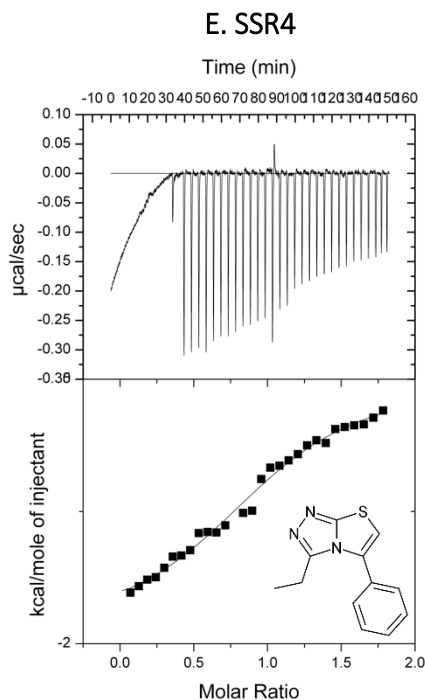
Figure 6.14. ITC thermograms of BRD4 BD1 interacting with (+)-JQ1 (A), SPF-REF (B), SPF5 (C), SPF22 (D), SSR4 (E).

6.5.4. X-ray crystallography of BRD4 BD1 in complex with optimized fragments

6.5.4.1. BRD4 BD1 crystallography in complex with optimized fragments

Crystallization experiments were performed in parallel with the ones described in 6.4.5.1 in the laboratory of Prof Frank von Delft and Prof Panagis Filippakopoulos (SGC Oxford).

Co-crystallization screening (Section 6.4.5.1 and 3.4.7.1) was tested for twelve evolved fragments coming from the fragment evolution platform. Protein-fragment mixtures were tested at 12 different reservoir conditions that consisted of 20% PEG 3350, 10% ethylene glycol, 0.2 M sodium fluoride and 0.1 M bis-tris propane with a gradient of 12 different pH ranging from 6.5 to 8.7. Fragments final concentrations were 10 mM 10% DMSO. SPF1 and SPF5 in complex with BRD4 BD1 were crystallized (Section 6.5.4.2). Reservoir conditions just differed in pH being 6.5 for SPF1 and 7.9 for SPF5. Second round of co-crystallization trials were

performed with a selection of new evolved fragments and two already tried with structural interest. Particularly, the already available crystallization screens BCS and MIDAS were tested to co-crystallize 10 and 12 fragments, respectively. Fragments final concentrations were 20 mM 10% DMSO. SSR2, SSR3 and SSR4 in complex with BRD4 BD1 were crystallized (Section 6.5.4.2). SSR2 reservoir conditions were from MIDAS and consisted of 0.2 M sodium chloride, 0.1 M MES pH 6, 30% v/v Jeffamine® ED-2003. SSR3 reservoir conditions were from MIDAS and consisted of 0.2 M potassium chloride, 0.1 M glycine pH 9.5, 20% v/v pentaerythritol. SSR4 reservoir conditions were from BCS and consisted of 0.1 M MES pH 6, 20% MPEGS, 0.15 M ammonium nitrate, 5% ethylene glycol.

Co-crystallization experiments were also performed with eleven compounds from the binding mode prediction platform. In this case, two crystallization screens, BCS and MIDAS, were assayed for all fragments at 20 mM, 10% DMSO. MR1, MR5 and MR7 in complex with BRD4 BD1 were crystallized (Section 6.5.4.2). MR1 and MR5 conditions were the same from MIDAS screen. They consisted of 0.2 M lithium citrate tribasic tetrahydrate, 35% v/v glycerol ethoxylate. MR7 reservoir conditions were from BCS and consisted of 28% MPEGS, 0.1 M HEPES pH 7.5, 0.05 M magnesium sulphate.

All crystals derived from 300 nL drops with different protein-fragment ratios. All crystallized fragments were in drops that contained protein-fragment 1:2 ratio, except from SSR2 and MR7 that had 2:1 and 1:1 ratio, respectively. The rest of conditions from crystallization experiments can be found in Section 3.4.7.1.

6.5.4.2. X-ray diffraction and structure elucidation

Dr Tobias Krojer (SGC, Oxford) was responsible for the X-ray data collection and structure elucidation steps. X-ray diffraction data was collected on beamlines I03 and I04-1 at the Diamond Light Source (UK's national synchrotron). Data was integrated with XDS[190] and scaled with AIMLESS[191] as part of the XIA2 and autoPROC auto-processing pipelines[192,193]. Initial refinement and map calculation was carried out with DIMPLE[194] using a ligand-free version of BRD4 BD1 (PDB code: 3MXF)[195] as starting model. Ligand restraints were generated with

ACEDRG[210] and GRADE[211]. Refinement and model building was performed with REFMAC[196] and COOT[197], respectively. Structure validation was performed with MolProbity[212]. BRD4 BD1 structure in complex with SPF1, SPF5, SSR2, SSR3, SSR4, MR1 and MR7 was elucidated and is illustrated in Figure 6.15. Notwithstanding, structure of BRD4 BD1 in complex with MR5 could not be properly determined. Despite finding density in the binding site that would indicate compound binding, MR5 did not fit properly, what would suggest that it was at low occupancy (Figure 6.16). Final models and structure factors of SPF1, SPF5 and SSR4 were deposited with PDB accession codes 6ZED, 6ZEL and 6ZF9, respectively. Data collection and refinement statistics are summarized in Table 6.7.
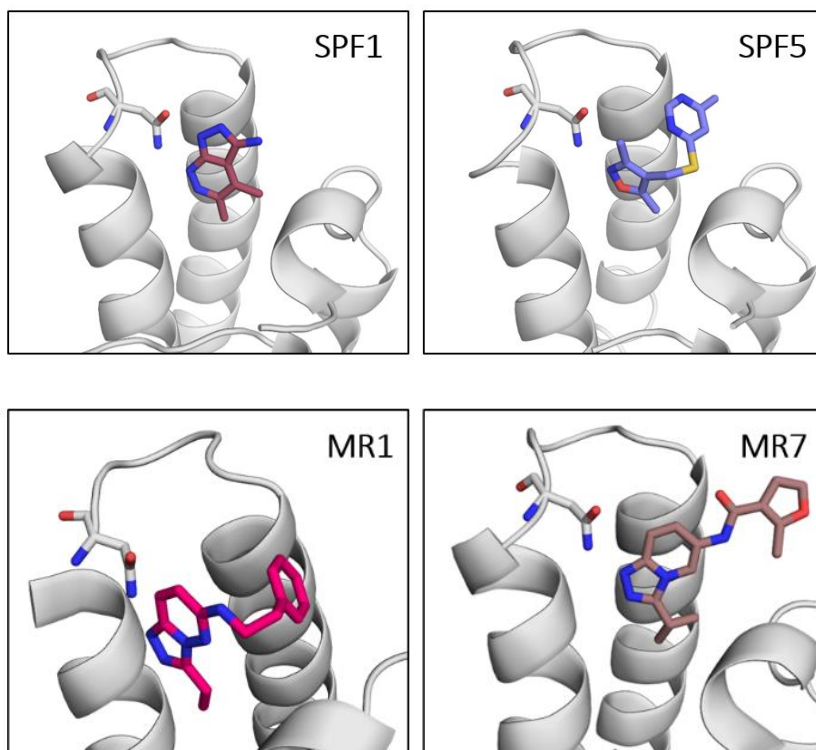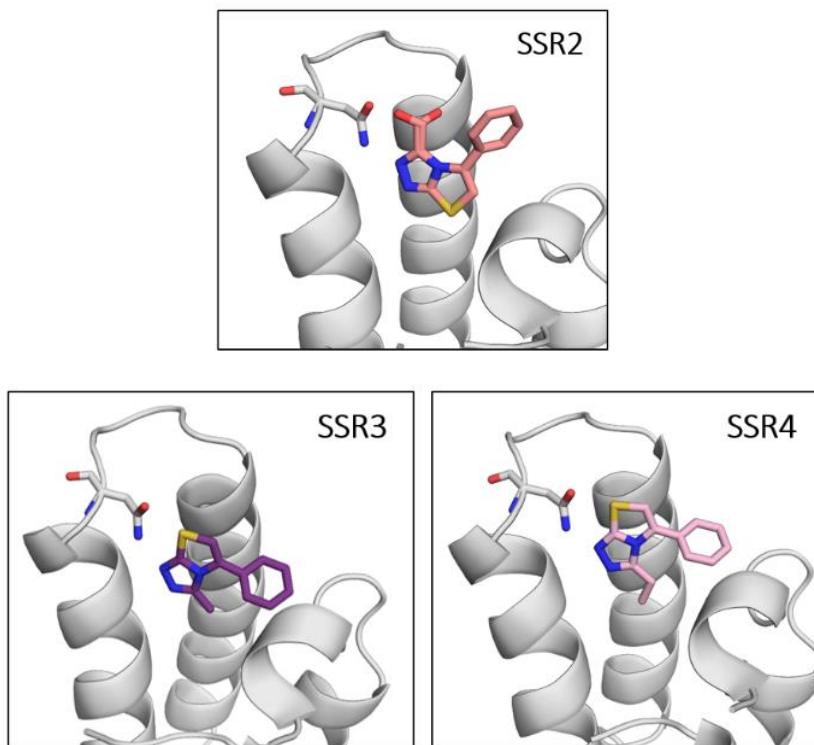
Figure 6.15. X-ray structure of BRD4 BD1 in complex with SPF1, SPF5, MR1, MR7, SSR2, SSR3 and SSR4.
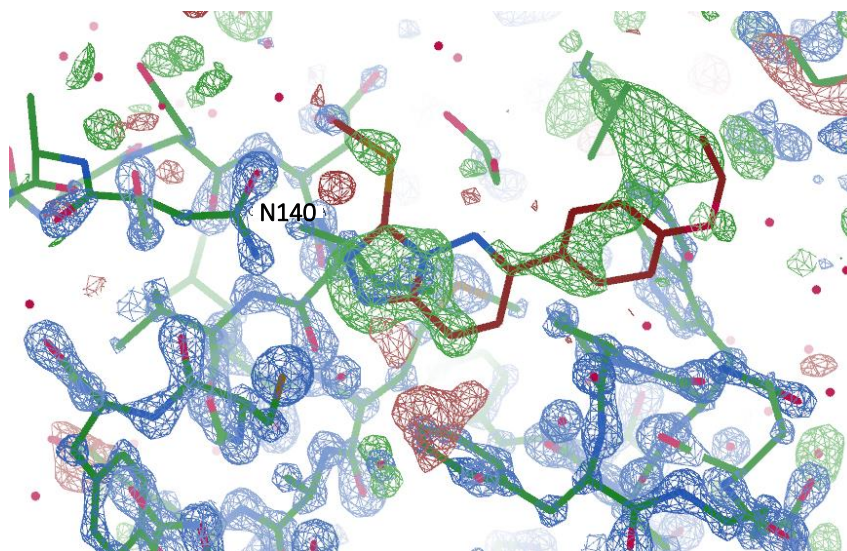
Figure 6.16. Electron density map of BRD4 BD1 co-crystallized with compound MR5. It can be observed the electron density map of BRD4 BD1. In green, the unresolved electron density map is shown, which in this case, it does not fit properly with MR5 (superposed in red).

Table 6.7. Data collection and refinement statistics of crystal structures of BRD4 BD1 interacting with SPF1 (6ZED), SPF5 (6ZEL), SSR4 (6ZF9).

PDB ID	6ZED	6ZEL	6ZF9
Data collection			
Beamline	DLS I04-1	DLS I03	DLS I04-1
Wavelength (Å)	0.9159	0.9763	0.9159
Resolution range (Å)	38.32 - 1.08 (1.12 - 1.08)	38.47 - 1.12 (1.16 - 1.12)	38.9 - 1.20 (1.24 - 1.20)
Space group	P 21 21 21	P 21 21 21	P 21 21 21
Cell dimensions:			
a, b, c (Å)	37.73 43.99 78.06	37.24 44.14 78.44	37.36 44.13 77.79
α, β, γ (deg)	90 90 90	90 90 90	90 90 90
Total reflections	315075 (19374)	336004 (22837)	199319 (5703)
Unique reflections	52334 (4466)	48007 (3979)	39156 (2810)
Multiplicity	6.0 (4.3)	7.0 (5.7)	5.1 (2.0)
Completeness (%)	92.43 (79.70)	94.98 (79.59)	95.47 (69.65)
Mean I/sigma(I)	24.68 (6.77)	26.10 (13.02)	22.68 (5.43)
Wilson B-factor	6,83	7.70	7,75
R-merge	0.0371 (0.1591)	0.04885 (0.1033)	0.03984 (0.1199)
R-meas	0.04052 (0.1813)	0.05282 (0.1133)	0.04378 (0.1571)
R-pim	0.016 (0.08521)	0.0197 (0.04521)	0.01779 (0.1)
CC1/2	0.999 (0.985)	0.998 (0.993)	0.999 (0.969)
CC*	1 (0.996)	1 (0.998)	1 (0.992)
Refinement			
Reflections used in refinement	52333 (4448)	48006 (3977)	39155 (2807)
Reflections used for R-free	2618 (216)	2405 (188)	1962 (146)
R-work	0.1232 (0.1105)	0.1249 (0.0958)	0.1235 (0.1601)
R-free	0.1433 (0.1421)	0.1466 (0.1299)	0.1485 (0.2001)
CC(work)	0.962 (0.846)	0.958 (0.861)	0.967 (0.908)

Table 6.7 (cont.). Data collection and refinement statistics of crystal structures of BRD4 BD1 interacting with SPF1 (6ZED), SPF5 (6ZEL), SSR4 (6ZF9).

PDB ID	6ZED	6ZEL	6ZF9
Number of non-hydrogen atoms	1311	1303	1313
macromolecules	1074	1086	1070
ligands	20	28	28
solvent	217	189	215
Protein residues	127	126	127
RMS(bonds)	0.012	0.244	0.044
RMS(angles)	1.65	4.17	3.09
Ramachandran favored (%)	99.2	99.19	99.2
Ramachandran allowed (%)	0.8	0.81	0.8
Ramachandran outliers (%)	0	0	0
Rotamer outliers (%)	0.83	0	0.84
Clashscore	1.84	1.8	2.76
Average B-factor	10.95	12.2	11.48
macromolecules	9.24	10.27	9.1
ligands	10.63	14.53	12.29
solvent	19.4	22.98	23.24

6.5.5. Summary and future perspectives (Biophysical characterization of BRD4 BD1 optimized fragments)

Fragment-based screening is a well-established methodology in drug discovery (Section 1.5.3.2). Several strategies are employed in order to optimize a fragment, being growing the most straightforward[213]. Computational tools can accelerate fragment to lead procedure in a rapid and cost-effective manner[214]. In that line, the automated platform based on iterative virtual screening developed by Dr Serena Piticchio and Dr Moira Rachmann was challenged to determine its performance in fragment to lead stage.

Hit rates of 73% and 83% were determined for fragment evolution and binding mode prediction platforms, respectively. These encouraging

results can be related to the fact that an initial known binder is selected and grown progressively while maintaining the main interaction with ASN140 (Section 6.1.2.2).

On one hand, in the fragment evolution platform no evolved fragments had thermal shifts higher than the reference fragment, SPF-REF. However, SSR4 demonstrated to be more potent than SPF-REF in ITC experiments. Even though correlation between affinity of compounds and thermal shifts has been described in bromodomains[195], in our experience, thermal shifts gave a clue to determine if a compound had weak or strong affinity but not in a quantitative manner. Different and novel chemical entities were evolved from the fragment evolution platform determining that it was not only able to increase the affinity of SPF-REF, but also navigate into chemical space for structural diversity. Atomic structures of 5 evolved fragments were resolved in the present thesis, three of which with very similar structure (SSR2, SSR3 and SSR4). Interestingly, while in SSR3 and SSR4 the alkyl group is pointing inside the binding site, in SSR2 the presence of the hydroxyl group rearranges the binding mode to have this moiety solvent exposed (Figure 6.15). That suggests the preference to place a hydrophobic group pointing through the cavity instead of a polar one to not disturb the network of conserved waters.

On the other hand, for the binding mode prediction platform 9 compounds presented thermal shifts higher than the reference fragment, BRD4-MREF. 3D structures of BRD4 BD1 interacting with two evolved fragments (MR1 and MR7) were obtained. However, their binding mode differed from the computationally predicted. Further optimization on initial steps of the platform were performed considering already crystallized binders. Overall, the automated platform provided new binders of BRD4 BD1 with impressive hit rates in both approaches. This encourages the application of this methodology in fragment-based campaigns for other targets.

CHAPTER 7: DISCUSSIONS

DISCUSSIONS

7.1. General discussion

The drug discovery process is an extremely arduous, long and expensive procedure and initial steps are crucial to ensure its success. The combination of computational and biophysical techniques is a strategy that can enhance our ability to modulate challenging and *undruggable* targets in early-stage drug discovery. In the present thesis, this approach was applied in three different projects with the aim to contribute to the development of novel small molecules able to interact with challenging proteins. Indeed, in Chapter 4 and Chapter 5 the targets of interest were *undrugged* proteins. In both cases, compounds able to interact with them were determined, resulting interesting points of departure to develop drugs able to modulate these targets. In contrast, the aim of Chapter 6 was focused on contributing to fundamental challenges faced in structure-based drug design. In the following paragraphs, a depth discussion of each project included in this thesis will be performed.

In Chapter 4 structure-based computational studies enabled the assessment of E3 ligases ligandability. E3 ligases specifically recognize other protein substrates for degradation. Even the popularity they have gained in recent years, especially for their involvement in PROTAC technology, only few of them have reported binders[122]. To contribute to the need of finding novel E3 ligases binders, their ligandability was studied implementing MDMix[53] methodology. MDMix consists on performing molecular dynamic simulations with organic co-solvents to identify hotspots over the protein surfaces. It relies on the demonstrated fact that proteins have the ability to unspecifically bind to small organic co-solvents in regions over the surface, and these regions correlate well with binding sites and important interactions for the drug binding affinity[52]. Besides, the use of these methodology was considered more robust than other tools for ligandability assessment, since it considers protein flexibility (while applying soft restraints) and predicts protein co-solvent preferences over time. E3 ligases ligandability was determined by the prediction of ligandable pockets with drug-like features. 22 out of the 23 E3 ligases tested presented at least one potential ligandable pocket, resulting in 2.3 ligandable pockets per E3 ligase. This evidences

the possibility to develop novel small molecules with the ability to interact with this target family.

We studied if MDMix was able to predict ligandable pockets in degron recognition sites. Actually, it predicted hotspots that could be employed as starting anchors to modulate some of these protein-protein interaction sites. Nonetheless, some of the MDMix pockets predicted to be in the degron site were not considered as ligandable. These results emphasized the importance of protein surface geometry in ligand design and the challenges faced when targeting protein-protein interactions. 66% of all predicted pockets were novel and allosteric, not being described before. Allosteric modulators would not compete with the natural substrate and this would be particularly interesting for the development of allosteric PROTAC molecules. Limitations of MDMix technology include the prediction of covalent ligands or cryptic pockets that underlie big conformational changes. In fact, MDMix was able to predict the opening of some pockets during the simulations. Retrospective validation of MDMix for the development of small molecules was done considering 7 described ligands of the selected E3 ligases. MDMix could successfully predict 6 out of 7 E3 ligases pockets with crystallized ligand. In fact, the non-predicted pocket was assessed by MDMix but not selected as ligandable for not accomplishing drug-like features.

Regarding FBW7 therapeutic interest and in order to prospectively validate our platform, a strategy to find FBW7 binders was planned. This could also be considered as a showcase of how MDMix results can be employed to develop novel E3 ligases modulators. First, the approach was based on the implementation of a virtual screening that involved docking experiments combined with dynamic undocking (DUck). While the first ensured that selected compounds accomplish the pharmacophore obtained from MDMix hotspots, the second was proved to be useful to detect true ligands and increase the success of virtual screening[92]. 7 million compound library was employed for this experiment, ending up with 41 compounds manually filtered to purchase and assess biophysically. This is a clear example of how computational tools help to speed up and reduce the cost of screening compounds,

leading to a final selection of a subset of molecules with more possibilities to interact with the target of interest.

Before testing the 41 compounds, protein had to be produced. Particularly, we produced FBW7 in complex with its adaptor SKP1. Several experiments were performed in order to improve the yield after protein purification with no success. For that reason, protein production was extensively performed during the whole thesis. DSF was selected as a primary screening for binding determination, since it is a versatile, fast and inexpensive methodology[65]. Despite all efforts in setting up the experiment, no compound presented any stabilization on the protein. Beside, difficulties were also faced when interpreting the results with the positive control or even the protein alone. The obtained results were attributed to the complexity of the protein we were working with, concluding that DSF was not the suitable technique to study small-molecule interactions with FBW7-SKP1 complex. SPR screening was performed instead, which compared to DSF requires less sample consumption, while allowing K_d determination. 9 compounds presented K_ds in the one to three-digit micromolar range, supposing a hit-rate of 22%. This is considered a desirable success rate taking into account the expected for other structure-based approaches[164] and that FBW7 had never been targeted. With the aim of finding more potent hits, *SAR by catalogue* was performed selecting 10 additional compounds that passed docking-based virtual screening filers. However, these new compounds had 70% or 80% of similarity to SPR virtual screening positive compounds. Therefore, some parent compounds shared few similarities with their analogues. All of them presented K_ds in the one to three-digit micromolar range, supposing a small increase in potency. In total 10 small molecules interacted with FBW7-SKP1 with a K_d lower than 100 µM.

Even these encouraging results, to validate our strategy it was crucial to characterize binding site of the positive compounds from the primary screening. To date 4 crystal structures containing FBW7-SKP1 have been released to PDB, being Dr Bing Hao (UconnHealth, USA) involved in all of them[165,98]. For that reason, we set up a collaboration with her to obtain crystal structures of the protein with our hits. However, no results have overcome until the moment. In parallel, SPR competitive assay was

performed. This experiments would show if compounds would compete for the same binding site. Actually, most binders seemed to compete between each other. To determine if they were able to interact in the predicted binding pocket, site-directed mutagenesis of different amino acids was performed. Five different single point mutations were performed in FBW7-SKP1 plasmid separately. Two consisted on mutating of the amino acid employed for DUck studies, while the other three pretended to block the cavity. SPR experiments were performed to compare the affinity data obtained when comparing WT and mutants. First, mutations regarding DUck amino acid were tested. No differences were observed between WT and mutants, probably because a close amino acid could replace the hydrogen bond formed. Differences in affinity were observed when comparing WT with the other mutations. Orthogonal techniques were performed to validate previous results. 4 of the best compounds were tested by MST experiments, comparing their interaction with FBW7-SKP1 and the mutated FBW7$_{A677I}$-SKP1. When comparing affinities between WT and the mutant A5_MMC17 and A6_MMC17 were observed to compete in both SPR and MST. ITC experiments were performed with A5_MMC17 with FBW7-SKP1 and FBW7$_{A677I}$-SKP1. Differences in affinity were confirmed, allowing us to conclude that this compound was actually binding to the predicted pocket. In contrast, discrepancies arose with A3_MMC21 and MMC21 when comparing SPR and MST results. ITC technique would be an ideal technique to further study this fact. Despite consuming a huge amount of protein, ITC does not require labelling or immobilization of the target. This technique would allow us to discard potential artefacts arising with the previous techniques and determine which of them is more appropriate to distinguish affinities between WT and mutants.

Mechanism of action of positive hits from SPR screening was initially assessed by fluorescence polarization assay. As a fluorescent probe, a labelled peptide of FBW7 natural substrate, DISC1, was selected. Therefore, this experiment allowed us to determine the effect of our small molecules in FBW7 substrate recognition. The predicted pocket was found in the interface between the WDR and the linker domains of FBW7, being the flexibility of this last described to play an important role in E3 ligases substrate ubiquitination[155]. Consequently, the fact that our

binders did not modulate substrate recognition was not surprising. The mechanism of action of A5_MMC17 and A6_MMC17 was further studied. Andrea Bertran, a PhD student in the group, performed cell assays following c-MYC, a substrate of FBW7, by western-blot. Both compounds presented the ability to decrease c-MYC levels, enhancing FBW7 activity in a proteasome-dependent manner. Notwithstanding, further assays need to be performed to confirm these results. Being A5_MMC17 and A6_MMC17 analogues, this family of compounds could be a starting point to develop drugs able to modulate FBW7 E3 ligase, a protein that has been considered *undruggable* until now, with no ligands described in the literature. A part from modulating FBW7 activity, described binders could also be employed for PROTAC construction. In fact, FBW7 was already successfully hijacked for protein degradation. This was achieved through an engineered fusion protein strategy[215].

TET2 epigenetic enzyme is the target of interest in Chapter 5. Modulation of TET2 has been described as a promising approach to treat cancer[170,173]. Our group has been collaborating with the group of Prof Héctor G. Palmer (VHIO, Barcelona) in order to perform a drug discovery project to develop a first-in-class TET2 modulator. To achieve that goal, a structure-based strategy similar than the one performed for the FBW7 E3 ligase was employed in the laboratory of Prof Xavier Barril. MDMix studies determined an allosteric pocket in TET2. Virtual screening was performed to select a set of small molecules with the potential to bind there, which were purchased. Enzymatic assays were performed at VHIO with the aforementioned compounds. Three different compounds were selected as hits: the positive allosteric modulator ONR-7B, and the negative allosteric modulators 3D and 6H. SAR by catalogue was employed to purchase analogues of them. In the present thesis, the first objective of this project was to determine the optimal biophysical technique to employ as primary screening. Initially, DSF was selected as primary screening. Nevertheless, similarly than in Chapter 4, few compounds could stabilize the protein and SPR was tried instead. SPR screening allowed us to better distinguish between true and false binders in single-dose experiments. Besides, K_d could be assessed for the three hits performing dose-response titrations. In parallel with these experiments, further studies related to the hits mechanism of action

were performed at VHIO and ONR-7B was selected as the principal hit. From that moment, analogues of ONR-7B were purchased or synthetized in order to improve hit potency, physicochemical properties and to define a composition of matter patent. SPR was performed at single-doses to select the binders that would subsequently be characterized in a dose-response manner. Even though some ONR-7B analogues presented improved physicochemical properties, none of them showed a relevant increase in affinity. The obtained SPR K_d allowed the ranking of ONR-7B analogues to perform further *in vitro* and *in vivo* experiments by our collaborators at VHIO. These activity assays showed that not all ONR-7B analogues were enhancing TET2 activity, since some of them were assessed as negative allosteric modulators.

Structural information of TET2 confirmed hits would be of utmost importance in order to validate our strategy. Furthermore, it would help to disentangle the activity mechanism related to the structure of ONR-7B analogues, and subsequently, allow the rational development of either positive or negative allosteric modulators. X-ray crystallography has been described as the most powerful, robust and routine method to determine atomic interactions between a target and its ligand[65]. Therefore and since TET2 had already been crystallized, we decided to perform this technique to characterize binding site and mode of ONR-7B. Three structures can be found in PDB that contain TET2 protein, all of them in complex with DNA and an α-ketoglutarate inhibitor[100,99]. The published conditions were tried to be reproduced with no success. Initially, we attributed that to the fact we were not using the same protein construct (TET2 construct 1). For that reason, our collaborators at VHIO modified the construct to have exactly the same as published (TET2 construct 2). Nevertheless, crystals could not be reproduced. Reproducing crystal conditions has been described to be a common pitfall. Consequently further crystallizations screens were also tried. This contained TET2 in different types of complex adding ONR-7B, without DNA, etc. Since crystals resulting from screening plates were tiny, no conclusive results were obtained in synchrotron and they were optimized. Optimized crystals were brought to synchrotron and unfortunately they resulted to be salt or DNA. This fact made us reconsider complex formation and size-exclusion chromatography of the

complex was performed. The results showed that complex was actually formed, however with an excess of DNA. Sample containing just the complex was pooled and used for preparing crystallization plates with no success at the moment. Nevertheless, this step will be included in the following crystallization experiments. All of these results evidence the difficulties faced in X-ray crystallization, which become even higher when pretending to crystallize a complex. The unique data this technique provides makes it worth the effort.

The project entailed in Chapter 5 is an example of an early stage drug discovery project lead by academia. In order to disentangle the mechanisms that underlie ONR-7B and its analogues a multidisciplinary team has been involved, which included medicinal chemists, who synthetize all the analogues, and biochemists, who assess the mechanism of action of the selected binders. Indeed, in parallel to this thesis, selected hit ONR-7B has been further studied by our collaborators at VHIO. Surprisingly, the primary hit ONR-7B already presented promising activity *in vivo* when tested in mice. Furthermore, it also presented an ideal pharmacokinetic and pharmacodynamic profile. These encouraging results motivate us to pursue in this drug discovery project.

A part from BRD4 BD1 therapeutic interest, its ease of production and its constant behaviour has made this system to be a good test protein for different computational approaches. Computer-aided drug design faces several challenges, being the prediction of solvation preferences and fragment evolution two of them. In Chapter 6, these challenges were tried to be assessed by implementing novel structure-based methodologies, employing BRD4 BD1 for prospective validation.

The prediction of the displacement or the interaction with protein water molecules has been of significant relevance in the structure-based drug design field[201]. BRD4 BD1 contains a network of conserved water molecules in the bottom of the substrate cavity. Actually, this network is conserved among all family members and it has been widely studied[183]. In the present thesis, we aimed to study solvation preferences of BRD4 BD1 water molecules by assessing water displacement using MDMix. Several articles describing water-mediated hydrogen bonds with the water network of BET family have been reported[202,203], as well as articles

that involve water rearrangement[204]. Furthermore, an approach that also employed mixed solvent molecular dynamic simulations to predict water displacement was published during the present thesis. The performed studies suggested that only a solvent mixture that contained acetate and methyl ammonium was able to displace conserved water molecules[205]. Nevertheless, our studies suggested that different types of solvents could displace at least one conserved water. These discrepancies could be caused by the percentage of solvent used in the simulations. They employed 5% organic/water mixture, while in our case it was of 20%. The percentage selected was an already described compromise between the amount of probe molecules needed to ensure efficient sampling, while remaining the environment predominantly aqueous and capturing the water solvation effect[187]. MDMix results and the information available in literature allowed us to perform a virtual screening for the rational design of compounds with the potential ability to displace the water network. After visual inspection, 40 compounds were selected to be studied biophysically. As in the other chapters, DSF was also selected as a primary technique to assess compound binding to the target. DSF has been widely used in bromodomains field and in this case it allowed us the proper determination of binding. Primary screening resulted in a hit rate of 50%, being considered a high success rate according to other structure-based approaches[164]. Information about the key interaction for BRD4 BD1 ligands and the deep knowledge of the system available in literature promoted these positive results. Thermal shifts provided by different ligands have been correlated with binding affinities in bromdomains[195]. According to that, our binders were predicted to be weak binders. X-ray crystallography was performed with compounds presenting similar thermal shifts. For that reason, we decided to perform this technique to finally validate our computational approach. Initially, soaking experiments were tried with no success. We related that to the symmetry packing of the protein, which hinders the ligand access to the binding site. In fact, co-crystallization has been the method of choice to crystallize BRD4 BD1 ligands. Co-crystallization experiments were exhaustively performed in order to determine the crystal structure of one of our binders bound to BRD4 BD1. Despite all efforts, only apo proteins could be crystallized. Computational studies involving docking were then performed to predict if compounds would

interact with BRD4 BD1 without displacing the network of water molecules. 40% of the selected compounds would bind inefficiently to BRD4 BD1. Nevertheless, structural information would also be necessary as a final assessment of our approach. Therefore, HSCQ-NMR is being considered as an alternative technique. With this experiment we would be able to determine novel interactions that our compounds would perform if displacing water molecules. Particularly, this would imply binding with the amino acid backbone of methionine 105. BRD4 BD1 structure has already been assigned by NMR[207], making the experiments more feasible, even we would still need to face the intrinsic challenges the technique implies.

Fragment-based screening is an approach widely used in early drug discovery, especially in recent years. Using smaller molecules than the drug-like ones allows the exploration of a larger amount of binding sites and proteins. The main challenge that this strategy faces is fragment optimization, mainly due to the requirement of structural information[61,62]. With the aim of contributing to this need, our group developed a computational tool that allowed the rational growing of a fragment. This automated platform was used during the present thesis in two different scenarios to address different challenges.

The first scenario involved the platform validation and has been referred in the thesis as fragment evolution platform. It consisted on growing an already crystallized fragment, referred as SPF-REF. 30 evolved fragments from the platform were either purchased or synthetized. As a primary screening DSF was also performed, providing a successful hit rate of 73%. Nevertheless, none of the evolved fragments provided thermal shifts higher than SPR-REF. ITC experiments were performed for some of the selected binders. SSR4 demonstrated to be more potent than SPF-REF, while SPF5 presented similar K_d than the latter. This showed that even DSF thermal shift can be correlated to affinity data, further characterization has to be performed in order to properly rank the molecules. The platform not only improved SPF-REF affinity, but also presented different and novel chemical entities, emphasizing its ability to navigate into chemical space for structural diversity. X-ray crystallography experiments were performed in order to characterize binding mode. The structures of BRD4 BD1 with 5 evolved fragments

were obtained. Three of them were analogues: SSR2, SSR3 and SSR4. Curiously, SSR3 and SSR4 have an alkyl group that is pointing inside the binding site, directly to the network of water molecules. In contrast, in SSR2 the presence of the hydroxyl group shifts the small molecule, rearranging the binding mode. That suggest the preference of the network of water molecules to place a hydrophobic group through the cavity.

The second strategy involving the automated platform has been referred as binding mode prediction platform. In this case, the structure of the reference fragment, MR-REF, was not determined. Consequently, the resulting approach involved a first step of binding mode prediction, followed by the application of the automated platform to develop small molecules with higher affinity and higher chances to be crystallized. This platform had an outstanding hit rate of 83%. 9 of the evolved fragments presented higher thermal shifts than MR-REF. X-ray crystallography experiments were also performed and the interaction of BRD4 BD1 with two evolved fragments could be characterized. Nevertheless, binding mode of the fragments was not the same as predicted by the platform. With the aim of improving the platform, further optimization on initial steps was performed considering obtained results.

Overall, the automated platform permitted the identification of novel binders of BRD4 BD1, providing impressive hit rates. These results encourage its application in further fragment-based drug discovery campaigns.

CHAPTER 8: CONCLUSIONS

CONCLUSIONS

8.1. Disentangling E3 ligases ligandability: application to FBW7 (Chapter 4)

- 23 E3 ligases were selected as representatives E3 ligases considering based on protein subfamilies and structural motifs.

- MDMix of the 23 E3 ligases was performed and its ligandability assessed.

- 22 E3 ligases presented at least one ligandable pocket, corresponding 2.3 pockets per E3 ligase. This evidenced the possibility to expand the toolbox of E3 ligase binders.

- MDMix was able to predict binding pocket of 6 out of the 7 ligands described of the selected E3 ligases.

- 66% of MDMix predicted pockets were novel and allosteric, not being described before.

- FBW7-Pocket G from MDMix was employed to perform a virtual screening. 41 small molecules were selected from virtual screening with the potential to bind to FBW7.

- For biophysical assays, FBW7-SKP1 was able to be expressed and purified.

- DSF experiment was set up to determine binding of 41 small molecules. However, it resulted not to be the suitable technique to study binding of FBW7-SKP1 complex.

- SPR demonstrated to be an appropriate technique for FBW7-SKP1 primary screening.

- 9 out of 41 compounds presented SPR K_ds in the one to three-digit micromolar range, supposing a hit-rate of 22%. SAR by catalogue was performed and 10 additional compounds were purchased. All presented SPR K_ds in the one to three-digit micromolar range. In total, we 10 compounds with a K_d lower than 100 µM were obtained.

- To validate our strategy, X-ray crystallography experiments were performed with no success for the moment.

- Alternatively, competitive SPR was performed to assess competition in the binding site. Almost all compounds appeared to compete between them.

- To determine the binding site, single-point mutations were performed in FBW7-SKP1, corresponding to: FBW7$_{N635A}$-SKP1, FBW7$_{N635I}$-SKP1, FBW7$_{A677I}$-SKP1, FBW7$_{A677F}$-SKP1 and FBW7$_{N679W}$-SKP1. Mutant proteins were able to be produced.

- SPR was performed to compare affinity between WT and mutants. No differences were observed between WT and FBW7$_{N635A}$-SKP1 or FBW7$_{N635I}$-SKP1. However, from 2 to 6 fold differences in affinity were observed for the other mutants.

- MST with FBW7-SKP1 and FBW7$_{A677I}$-SKP1 was performed for 4 of our best compounds. A5_MMC17, A6_MMC17 and A3_MMC21 presented differences between WT and mutant.

- Discrepancies were observed when comparing WT and mutant affinities for MMC21 and A3_MMC21 in SPR and MST assays.

- ITC was performed to characterize binding of A5_MMC17 with FBW7-SKP1 and FBW7$_{A677I}$-SKP1. Clear differences in affinity were observed.

- A5_MMC17 presented differences in affinity between FBW7-SKP1 and FBW7$_{A677I}$-SKP1 in three orthogonal techniques. This suggests that this compound binds to the predicted pocket.

- FP experiment was set up to assess how our compounds affect substrate recognition. No effect was observed to be performed by them.

- Cell-based assays following the FBW7 substrate c-MYC have been performed with A5_MMC17 and A6_MMC17. They both appeared to decrease c-MYC levels in a proteasome-dependent manner, supposing an enhancement of FBW7 activity. Further studies need to be performed to confirm that.

- Overall, the combination of computational approaches with biophysical techniques has led to the successful development of small molecules binding to FBW7, reported as undruggable.

- Being analogues, A5_MMC17 and A6_MMC17 family could be a point of departure to develop drugs able to modulate FBW7.

8.2. Pharmacological modulation of TET2 epigenetic enzyme (Chapter 5)

- We were able to express and purify TET2 construct 1 and construct 2.

- DSF was attempted as a primary screening of potential hits. However, it was not a suitable technique to assess TET2 binding.

- SPR resulted to be a more appropriate technique to determine TET2 binding.

- All 3D and 6H analogues tested at single-dose in SPR experiments could bind to TET2.

- 22 out of 52 ONR-7B analogues presented binding to TET2 according SPR single-dose experiments.

- 24 compounds were tested at dose-response with SPR and ranked for mechanism of action assessment. None of ONR-7B analogues was significantly more potent than ONR-7B.

- X-ray crystallography experiments were tried for TET2 construct 1 and TET2 construct 2. Unfortunately and despite all efforts, no crystals of TET2 were obtained.

- Size-exclusion chromatography of TET2 in complex with DNA, confirmed complex formation and showed the presence of an excess of DNA. This step will be included in further crystallization experiments.

- For the moment, binding mode of ONR-7B have not been able to be assessed by X-ray crystallography.

8.3. Development of new chemical entities that bind to the bromodomain BRD4 BD1 (Chapter 6)

- We were able to express and purify BRD4 BD1.

- BRD4 BD1 most conserved waters were calculated considering all the published structures in PDB.

- MDMix experiments of BRD4 BD1 with different co-solvents were performed. These results were used to assess the displacement of most conserved water molecules.

- Acetamide was selected as a co-solvent able to displace waters 14 and 3. A virtual screening of small molecules that could displace these waters by emulating acetamide was performed, considering the key interaction with ASN140. 40 drug-like molecules were manually selected to be tested biophysically.

- DSF was performed for the 40 selected small molecules, with a hit rate of 50% when testing the compounds at 200 µM.

- Despite all efforts, only apo BRD4 BD1 structures could be obtained and the binding mode of the positive binders could not be elucidated.

- Docking-based virtual screening including the network of water molecules was performed with positive binders. 40% of them were predicted to bind inefficiently in case they do not displace the water network.

- DSF experiments were performed for the 30 compounds from the fragment evolution platform project. A hit rate of 73% was obtained and novel chemical entities were developed.

- ITC experiments were performed with 5 fragments of the fragment evolution platform. SSR4 was observed to be more potent than SPF-REF.

- X-ray crystallography enabled binding mode characterization of 5 small molecules resulting from the fragment evolution platform.

– DSF experiments were performed for the 18 compounds of the binding mode prediction platform. An outstanding hit rate of 83% was obtained. 9 small molecules presented higher stabilization of BRD4 BD1 compared to MR-REF.

– Binding mode of 2 small molecules form the binding mode prediction platform could be successfully characterized by X-ray crystallography. However, they did not match with the predicted binding mode.

– The automated platform, which was employed in fragment evolution platform and binding mode prediction platform, demonstrated its capacity to successfully develop known fragments into more potent and diverse compounds.

CHAPTER 9: BIBLIOGRAPHY

BIBLIOGRAPHY

1. Rosa, A. C., Mio, M., Andreadou, I. & Sumbayev, V. V. Editorial: The Challenge of New Therapeutic Approaches for Unmet Therapeutic Needs. *Front. Pharmacol.* **11**, (2020).

2. Deore, A. B., Dhumane, J. R., Wagh, R. & Sonawane, R. The Stages of Drug Discovery and Development Process. *Asian J. Pharm. Res. Dev.* **7**, 62–67 (2019).

3. Wouters, O. J., McKee, M. & Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA - J. Am. Med. Assoc.* **323**, 844–853 (2020).

4. Hughes, J. P., Rees, S. S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **162**, 1239–1249 (2011).

5. Keseru, G. M. & Makara, G. M. Hit discovery and hit-to-lead approaches. *Drug Discov. Today* **11**, 741–748 (2006).

6. Mohs, R. C. & Greig, N. H. Drug discovery and development: Role of basic biological research. *Alzheimer's Dement. Transl. Res. Clin. Interv.* **3**, 651–657 (2017).

7. Friedman, L. M., Furberg, C. D., DeMets, D. L., Reboussin, D. M. & Granger, C. B. *Fundamentals of Clinical Trials*. (Springer International Publishing, 2015).

8. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–730 (2002).

9. Kana, O. & Brylinski, M. Elucidating the druggability of the human proteome with eFindSite. *J. Comput. Aided. Mol. Des.* **33**, 509–519 (2019).

10. Oprea, T. I. *et al.* Unexplored therapeutic opportunities in the human genome. *Nat. Rev. Drug Discov.* **17**, 317–332 (2018).

11. Vukovic, S. & Huggins, D. J. Quantitative metrics for drug–target ligandability. *Drug Discov. Today* **23**, 1258–1266 (2018).

12. Cancer research UK. Tackle cancers with substantial unmet need: our research strategy. Available at: https://www.cancerresearchuk.org/funding-for-researchers/our-research-strategy/tackle-cancers-with-substantial-unmet-need. (Accessed: 3rd July 2021)

13. Makurvet, F. D. Biologics vs. small molecules: Drug costs and patient access. *Med. Drug Discov.* **9**, 100075 (2021).

14. Mullard, A. 2020 FDA drug approvals. *Nat. Rev. Drug Discov.* **20**, 85–90 (2021).

15. Ni, D. *et al.* Combining Allosteric and Orthosteric Drugs to Overcome Drug Resistance. *Trends Pharmacol. Sci.* **41**, 336–348 (2020).

16. Monod, J., Changeux, J. P. & Jacob, F. Allosteric proteins and cellular control systems. *J. Mol. Biol.* **6**, 306–329 (1963).

17. Nussinov, R. & Tsai, C.-J. The Different Ways through Which Specificity Works in Orthosteric and Allosteric Drugs. *Curr. Drug Metab.* **18**, 1311–1316 (2012).

18. Lu, S. & Zhang, J. Allosteric Modulators. in *Comprehensive Medicinal Chemistry III* **2**, 276–296 (Elsevier, 2017).

19. Pommier, Y. & Marchand, C. Interfacial inhibitors of protein-nucleic acid interactions. *Curr. Med. Chem. - Anti-Cancer Agents* **5**, 421–429 (2005).

20. Scott, D. E., Bayly, A. R., Abell, C. & Skidmore, J. Small molecules, big targets: Drug discovery faces the protein-protein interaction challenge. *Nat. Rev. Drug Discov.* **15**, 533–550 (2016).

21. Jin, L., Wang, W. & Fang, G. Targeting protein-protein interaction by small molecules. *Annu. Rev. Pharmacol. Toxicol.* **54**, 435–456 (2014).

22. Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: Progressing toward the reality. *Chem. Biol.* **21**, 1102–1114 (2014).

23. Schreiber, S. L. The Rise of Molecular Glues. *Cell* **184**, 3–9 (2021).

24. den Besten, W. & Lipford, J. R. Prospecting for molecular glues. *Nat. Chem. Biol.* **16**, 1157–1158 (2020).

25. Gao, H., Sun, X. & Rao, Y. PROTAC Technology: Opportunities and Challenges. *ACS Med. Chem. Lett.* **11**, 237–240 (2020).

26. Galdeano, C. & Barril, X. CHAPTER 7. Targeted Protein Degradation Chemical Probes. in *The Discovery and Utility of Chemical Probes in Target Discovery* (eds. Brennan, P. & Rodriguez, S. V.) 150–181 (Royal Society of Chemistry, 2020).

27. Alabi, S. B. & Crews, C. M. Major advances in targeted protein degradation: PROTACs, LYTACs, and MADTACs. *J. Biol. Chem.* **296**, 100647 (2021).

28. Bondeson, D. P. *et al.* Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat. Chem. Biol.* **11**, 611–617 (2015).

29. Aymami, J., Barril, X., Rodríguez-Pascau, L. & Martinell, M. Pharmacological chaperones for enzyme enhancement therapy in genetic diseases. *Pharm. Pat. Anal.* **2**, 109–124 (2013).

30. Liguori, L. *et al.* Pharmacological Chaperones: A Therapeutic Approach for Diseases Caused by Destabilizing Missense Mutations. *Int. J. Mol. Sci.* **21**, 489 (2020).

31. Swinney, D. C. Phenotypic vs. Target-Based Drug Discovery for First-in-Class Medicines. *Clin. Pharmacol. Ther.* **93**, 299–301 (2013).

32. Moffat, J. G., Vincent, F., Lee, J. A., Eder, J. & Prunotto, M. Opportunities and challenges in phenotypic drug discovery: An industry perspective. *Nat. Rev. Drug Discov.* **16**, 531–543 (2017).

33. Kubota, K., Funabashi, M. & Ogura, Y. Target deconvolution from phenotype-based drug discovery by using chemical proteomics approaches. *Biochim. Biophys. Acta - Proteins Proteomics* **1867**, 22–27 (2019).

34. Ege, N., Bouguenina, H., Tatari, M. & Chopra, R. Phenotypic screening with target identification and validation in the discovery and development of E3 ligase modulators. *Cell Chem. Biol.* **28**, 283–299 (2021).

35. Eder, J., Sedrani, R. & Wiesmann, C. The discovery of first-in-class drugs: Origins and evolution. *Nat. Rev. Drug Discov.* **13**, 577–587 (2014).

36. Shortt, J., Ott, C. J., Johnstone, R. W. & Bradner, J. E. A chemical probe toolbox for dissecting the cancer epigenome. *Nat. Rev. Cancer* **17**, 160–183 (2017).

37. Arrowsmith, C. H. *et al.* The promise and peril of chemical probes. **11**, (2015).

38. Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe, E. W. Computational methods in drug discovery. *Pharmacol. Rev.* **66**, 334–395 (2014).

39. Willems, H., De Cesco, S. & Svensson, F. Computational Chemistry on a Budget: Supporting Drug Discovery with Limited Resources. *J. Med. Chem.* **63**, 10158–10169 (2020).

40. Batool, M., Ahmad, B. & Choi, S. A Structure-Based Drug Discovery Paradigm. *Int. J. Mol. Sci.* **20**, 2783 (2019).

41. Schmidtke, P. & Barril, X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J. Med. Chem.* **53**, 5858–5867 (2010).

42. Clark, J. J., Orban, Z. J. & Carlson, H. A. Predicting binding sites from unbound versus bound protein structures. *Sci. Rep.* **10**, 1–18 (2020).

43. Brylinski, M. & Skolnick, J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci.* **105**, 129–134 (2008).

44. Roche, D. B., Buenavista, M. T., Tetchner, S. J. & McGuffin, L. J. The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. *Nucleic Acids Res.* **39**, W171–W176 (2011).

45. Wass, M. N., Kelley, L. A. & Sternberg, M. J. E. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res.* **38**, W469–W473 (2010).

46. Schmidtke, P., Le Guilloux, V., Maupetit, J. & Tufféry, P. fpocket: Online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res.* **38**, 582–589 (2010).

47. Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **49**, 377–389 (2009).

48. Ravindranath, P. A. & Sanner, M. F. AutoSite: an automated approach for pseudo-ligands prediction—from ligand-binding sites identification to predicting key ligand atoms. *Bioinformatics* **32**, 3142–3149 (2016).

49. An, J., Totrov, M. & Abagyan, R. Pocketome via Comprehensive Identification and Classification of Ligand Binding Envelopes. *Mol. Cell. Proteomics* **4**, 752–761 (2005).

50. MacKerell, A. D., Jo, S., Lakkaraju, S. K., Lind, C. & Yu, W. Identification and characterization of fragment binding sites for allosteric ligand design using the site identification by ligand competitive saturation hotspots approach (SILCS-Hotspots). *Biochim. Biophys. Acta - Gen. Subj.* **1864**, (2020).

51. Yanagisawa, K., Moriwaki, Y., Terada, T. & Shimizu, K. EXPRORER: Rational Cosolvent Set Construction Method for Cosolvent Molecular Dynamics Using Large-Scale Computation. *J. Chem. Inf. Model.* **61**, 2744–2753 (2021).

52. Seco, J., Luque, F. J. & Barril, X. Binding Site Detection and Druggability Index from First Principles. *J. Med. Chem.* **52**, 2363–2371 (2009).

53. Alvarez-Garcia, D. & Barril, X. Relationship between Protein Flexibility and Binding: Lessons for Structure-Based Drug Design. *J. Chem. Theory Comput.* **10**, 2608–2614 (2014).

54. Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M. & Taranto, A. G. Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Front. Chem.* **8**, (2020).

55. Wang, Z. *et al.* Combined strategies in structure-based virtual screening. *Phys. Chem. Chem. Phys.* **22**, 3149–3159 (2020).

56. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).

57. Bianco, G., Goodsell, D. S. & Forli, S. Selective and Effective: Current Progress in Computational Structure-Based Drug Discovery of Targeted Covalent Inhibitors. *Trends Pharmacol. Sci.* **41**, 1038–1049 (2020).

58. Ruiz-Carmona, S. *et al.* Dynamic undocking and the quasi-bound state as tools for drug discovery. *Nat. Chem.* **9**, 2–7 (2016).

59. MacArron, R. *et al.* Impact of high-throughput screening in biomedical

research. *Nat. Rev. Drug Discov.* **10**, 188–195 (2011).

60. Scott, D. E., Coyne, A. G., Hudson, S. A. & Abell, C. Fragment-Based Approaches in Drug Discovery and Chemical Biology. *Biochemistry* **51**, 4990–5003 (2012).

61. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: The impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **15**, 605–619 (2016).

62. Parker, C. G. *et al.* Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* **168**, 527–541 (2017).

63. Atanasov, A. G. *et al.* Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.* **20**, 200–216 (2021).

64. Rodrigues, T., Reker, D., Schneider, P. & Schneider, G. Counting on natural products for drug design. *Nat. Chem.* **8**, 531–541 (2016).

65. Renaud, J. P. *et al.* Biophysics in drug discovery: Impact, challenges and opportunities. *Nat. Rev. Drug Discov.* **15**, 679–698 (2016).

66. Gao, K., Oerlemans, R. & Groves, M. R. Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophys. Rev.* **12**, 85–104 (2020).

67. Ciulli, A. Biophysical Screening for the Discovery of Small-Molecule Ligands. in *Methods Mol Biol* (eds. Williams, M. A. & Daviter, T.) **1008**, 357–388 (Humana Press, 2013).

68. Senisterra, G., Chau, I. & Vedadi, M. Thermal Denaturation Assays in Chemical Biology. *Assay Drug Dev. Technol.* **10**, 128–136 (2012).

69. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221 (2007).

70. Dai, R. *et al.* Fragment-Based Exploration of Binding Site Flexibility in Mycobacterium tuberculosis BioA. *J. Med. Chem.* **58**, 5208–5217 (2015).

71. Patching, S. G. Surface plasmon resonance spectroscopy for characterisation of membrane protein-ligand interactions and its potential for drug discovery. *Biochim. Biophys. Acta - Biomembr.* **1838**, 43–55 (2014).

72. Olaru, A., Bala, C., Jaffrezic-Renault, N. & Aboul-Enein, H. Y. Surface Plasmon Resonance (SPR) Biosensors in Pharmaceutical Analysis. *Crit. Rev. Anal. Chem.* **45**, 97–105 (2015).

73. GE Healthcare. Biacore systems in small molecule drug discovery. *Appl. Note* **2**, (2016).

74. Nanotemper Technologies Inc. User manual for the Monolith NT.115. 1–10 (2007).

75. Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S. Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev. Technol.* **9**, 342–353 (2011).

76. Jerabek-willemsen, M. *et al.* MicroScale Thermophoresis : Interaction analysis and beyond. *J. Mol. Struct.* **1077**, 101–113 (2014).

77. Asmari, M., Ratih, R., Alhazmi, H. A. & El Deeb, S. Thermophoresis for characterizing biomolecular interaction. *Methods* **146**, 107–119 (2018).

78. Turnbull, W. B. & Daranas, A. H. On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *J. Am. Chem. Soc.* **125**, 14859–14866 (2003).

79. Song, C., Zhang, S. & Huang, H. Choosing a suitable method for the

identification of replication origins in microbial genomes. *Front. Microbiol.* **6**, (2015).

80. Du, X. *et al.* Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **17**, 144 (2016).

81. Claveria-Gimeno, R., Vega, S., Abian, O. & Velazquez-Campoy, A. A look at ligand binding thermodynamics in drug discovery. *Expert Opin. Drug Discov.* **12**, 363–377 (2017).

82. Zhang, H., Wu, Q. & Berezin, M. Y. Fluorescence anisotropy (polarization): From drug screening to precision medicine. *Expert Opin. Drug Discov.* **10**, 1145–1161 (2015).

83. Rossi, A. M. & Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* **6**, 365–387 (2011).

84. Zheng, H. *et al.* X-ray crystallography over the past decade for novel drug discovery -where are we heading next? *Expert Opin. Drug Discov.* **10**, 975–989 (2015).

85. McPherson, A. Protein Crystallization. in *Nature* (eds. Wlodawer, A., Dauter, Z. & Jaskolski, M.) **1607**, 17–50 (Springer New York, 2017).

86. Maveyraud, L. & Mourey, L. Protein X-ray crystallography and drug discovery. *Molecules* **25**, (2020).

87. Müller, I. Guidelines for the successful generation of protein-ligand complex crystals. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 79–92 (2017).

88. Berman, H. M. *et al.* The protein data bank. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 899–907 (2002).

89. Chemical Computing Group ULC. Molecular Operating Environment (MOE), 2016. (2016).

90. Ruiz-Carmona, S. *et al.* rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.* **10**, 1–7 (2014).

91. Chemical Computing Group ULC. Molecular Operating Environment (MOE) 2014.09. (2015).

92. Ruiz-Carmona, S. *et al.* Dynamic undocking and the quasi-bound state as tools for drug discovery. *Nat Chem* **9**, 201–206 (2017).

93. Chen, X. & Reynolds, C. H. Performance of similarity measures in 2D fragment-based similarity searching: Comparison of structural descriptors and similarity coefficients. *J. Chem. Inf. Comput. Sci.* **42**, 1407–1414 (2002).

94. Durant, J. L., Leland, B. A., Henry, D. R. & Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **42**, 1273–1280 (2002).

95. Hao, B., Oehlmann, S., Sowa, M. E., Harper, J. W. & Pavletich, N. P. Structure of a Fbw7-Skp1-Cyclin E Complex: Multisite-Phosphorylated Substrate Recognition by SCF Ubiquitin Ligases. *Mol. Cell* **26**, 131–143 (2007).

96. Spurny, R. *et al.* Molecular blueprint of allosteric binding sites in a homologue of the agonist-binding domain of the α7 nicotinic acetylcholine receptor. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E2543–E2552 (2015).

97. Healthcare, G. Biacore ™ T200. *Manual* **200**, 1–8 (2015).

98. Yalla, K. *et al.* FBXW7 regulates DISC1 stability via the ubiquitin-proteosome system. *Mol. Psychiatry* **23**, 1278–1286 (2018).

99. Hu, L. *et al.* Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **155**, 1545–1555 (2013).

100. Hu, L. *et al.* Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527**, 118–122 (2015).

101. Pickering, J., Hammond, R. B., Ramachandran, V., Soufian, M. & Roberts, K. J. Synthonic Engineering Modelling Tools for Product and Process Design. in *NATO Science for Peace and Security Series A: Chemistry and Biology* **1**, 155–176 (2017).

102. Lucas, X. *et al.* 4-Acyl pyrroles: Mimicking acetylated lysines in histone code reading. *Angew. Chemie - Int. Ed.* **52**, 14055–14059 (2013).

103. Ember, S. W. J. *et al.* Acetyl-lysine Binding Site of Bromodomain-Containing Protein 4 (BRD4) Interacts with Diverse Kinase Inhibitors. *ACS Chem. Biol.* **9**, 1160–1171 (2014).

104. Filippakopoulos, P. & Knapp, S. The bromodomain interaction module. *FEBS Lett.* **586**, 2692–2704 (2012).

105. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).

106. Picaud, S. *et al.* RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proc. Natl. Acad. Sci.* **110**, 19754–19759 (2013).

107. Ng, J. T., Dekker, C., Reardon, P. & von Delft, F. Lessons from ten years of crystallization experiments at the SGC. *Acta Crystallogr. Sect. D Struct. Biol.* **72**, 224–235 (2016).

108. Wilkinson, K. D. Essay. *Cell* **119**, 741–745 (2004).

109. Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229 (2012).

110. Huang, X. & Dixit, V. M. Drugging the undruggables: exploring the ubiquitin system for drug development. *Cell Res.* **26**, 484–498 (2016).

111. Morreale, F. E. & Walden, H. Types of Ubiquitin Ligases. *Cell* **165**, 248–248 (2016).

112. Huang, X. & Dixit, V. M. Drugging the undruggables: Exploring the ubiquitin system for drug development. *Cell Research* **26**, 484–498 (2016).

113. Stewart, M. D., Ritterhoff, T., Klevit, R. E. & Brzovic, P. S. E2 enzymes: More than just middle men. *Cell Res.* **26**, 423–440 (2016).

114. Galdeano, C. Drugging the undruggable: targeting challenging E3 ligases for personalized medicine. *Futur. Med. Chem.* **9**, 347–350 (2017).

115. Zinzalla, G. & Thurston, D. E. Targeting protein-protein interactions for therapeutic intervention: A challenge for the future. *Future Med. Chem.* **1**, 65–93 (2009).

116. Galdeano, C. *et al.* Structure-guided design and optimization of small molecules targeting the protein-protein interaction between the von hippel-lindau (VHL) E3 ubiquitin ligase and the hypoxia inducible factor (HIF) alpha subunit with in vitro nanomolar affinities. *J. Med. Chem.* **57**, 8657–8663 (2014).

117. Davies, T. G. *et al.* 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. *J. Med. Chem.* **59**, 3991–4006 (2016).

118. Chan, C. H. *et al.* Pharmacological inactivation of Skp2 SCF ubiquitin ligase restricts cancer stem cell traits and cancer progression. *Cell* **154**, 556–568 (2013).

119. Guo, Z. *et al.* Small-Molecule Targeting of E3 Ligase Adaptor SPOP in Kidney Cancer. *Cancer Cell* **30**, 474–484 (2016).

120. Luo, M. *et al.* Chemoproteomics-enabled discovery of covalent RNF114-based degraders that mimic natural product function. *Cell Chem. Biol.* **28**, 559-566.e15 (2021).

121. Ward, C. C. *et al.* Covalent Ligand Screening Uncovers a RNF4 E3 Ligase Recruiter for Targeted Protein Degradation Applications. *ACS Chem. Biol.* **14**, 2430–2440 (2019).

122. Ishida, T. & Ciulli, A. E3 Ligase Ligands for PROTACs: How They Were Found and How to Discover New Ones. *SLAS Discov. Adv. Sci. Drug Discov.* **26**, 484–502 (2021).

123. Schapira, M., Calabrese, M. F., Bullock, A. N. & Crews, C. M. Targeted protein degradation: expanding the toolbox. *Nat. Rev. Drug Discov.* **18**, 949–963 (2019).

124. Liu, L. *et al.* UbiHub : a data hub for the explorers of ubiquitination pathways. **35**, 2882–2884 (2019).

125. Bateman, A. *et al.* UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

126. Sigrist, C. J. A. *et al.* New and continuing developments at PROSITE. *Nucleic Acids Res.* **41**, 344–347 (2013).

127. Groocock, L. M. *et al.* RNF 4 interacts with both SUMO and nucleosomes to promote the DNA damage response. **15**, 601–608 (2014).

128. Lucas, X., Van Molle, I. & Ciulli, A. Surface Probing by Fragment-Based Screening and Computational Methods Identifies Ligandable Pockets on the von Hippel-Lindau (VHL) E3 Ubiquitin Ligase. *J. Med. Chem.* **61**, 7387–7393 (2018).

129. Petzold, G., Fischer, E. S. & Thomä, N. H. Structural basis of lenalidomide-induced CK1α degradation by the CRL4 CRBN ubiquitin ligase. *Nature* **532**, 127–130 (2016).

130. Bussiere, D. E. *et al.* Structural basis of indisulam-mediated RBM39 recruitment to DCAF15 E3 ligase complex. *Nat. Chem. Biol.* **16**, 15–23 (2020).

131. Ma, B. *et al.* Design, synthesis and identification of novel, orally bioavailable non-covalent Nrf2 activators. *Bioorganic Med. Chem. Lett.* **30**, 126852 (2020).

132. Pallesen, J. S. *et al.* Deconstructing Noncovalent Kelch-like ECH-Associated Protein 1 (Keap1) Inhibitors into Fragments to Reconstruct New Potent Compounds. *J. Med. Chem.* **64**, 4623–4661 (2021).

133. Sackton, K. L. *et al.* Synergistic blockade of mitotic exit by two chemical inhibitors of the APC/C. *Nature* **514**, 646–649 (2014).

134. Asatsuma-Okumura, T., Ito, T. & Handa, H. Molecular mechanisms of cereblon-based drugs. *Pharmacol. Ther.* **202**, 132–139 (2019).

135. Liu, J. *et al.* CRL4ACRBN E3 ubiquitin ligase restricts BK channel activity and prevents epileptogenesis. *Nat. Commun.* **5**, 3924 (2014).

136. Schwefel, D. *et al.* Structural basis of lentiviral subversion of a cellular protein degradation pathway. *Nature* **505**, 234–238 (2014).

137. Mallery, D. L. *et al.* Antibodies mediate intracellular immunity through tripartite motif-containing 21 (TRIM21). **107**, 2–7 (2010).

138. James, L. C., Keeble, A. H., Khan, Z., Rhodes, D. A. & Trowsdale, J. Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proc. Natl. Acad. Sci.* **104**, 6200–6205 (2007).

139. Huynh, D. P., Scoles, D. R., Nguyen, D. & Pulst, S. M. The autosomal recessive juvenile Parkinson disease gene product, parkin, interacts with and ubiquitinates synaptotagmin XI. *Hum. Mol. Genet.* **12**, 2587–2597 (2003).

140. Kumar, A. *et al.* Parkin-phosphoubiquitin complex reveals cryptic ubiquitin-binding site required for RBR ligase activity. *Nat. Struct. Mol. Biol.* **24**, 475–483 (2017).

141. Wauer, T., Simicek, M., Schubert, A. & Komander, D. Mechanism of phospho-ubiquitin-induced PARKIN activation. *Nature* **524**, 370–374 (2015).

142. Dong, C. *et al.* Molecular basis of GID4-mediated recognition of degrons for the Pro/N-end rule pathway. *Nat. Chem. Biol.* **14**, 466–473 (2018).

143. Dong, C. *et al.* Recognition of nonproline N-terminal residues by the Pro/N-degron pathway. *Proc. Natl. Acad. Sci.* **117**, 14158–14167 (2020).

144. Rusnac, D.-V. *et al.* Recognition of the Diglycine C-End Degron by CRL2KLHDC2 Ubiquitin Ligase. *Mol. Cell* **72**, 813–822 (2018).

145. Hao, B. *et al.* Structural basis of the Cks1-dependent recognition of p27 Kip1 by the SCF Skp2 ubiquitin ligase. *Mol. Cell* **20**, 9–19 (2005).

146. Uljon, S. *et al.* Structural Basis for Substrate Selectivity of the E3 Ligase COP1. *Structure* **24**, 687–696 (2016).

147. Filippakopoulos, P. *et al.* Structural Basis for Par-4 Recognition by the SPRY Domain- and SOCS Box-Containing Proteins SPSB1 ,. *J. Mol. Biol.* **401**, 389–402 (2010).

148. Kung, W., Ramachandran, S., Makukhin, N., Bruno, E. & Ciulli, A. Structural insights into substrate recognition by the SOCS2 E3 ubiquitin ligase. *Nat. Commun.* **10**, 2534 (2019).

149. Yoshimura, A. *et al.* A novel cytokine-inducible gene CIS encodes an SH2-containing protein that binds to tyrosinephosphorylated interleukin 3 and erythropoietin receptors. *EMBO J.* **14**, 2816–2826 (1995).

150. Radaszkiewicz, T. & Bryja, V. Protease associated domain of RNF43 is not necessary for the suppression of Wnt/β-catenin signaling in human cells. *Cell Commun. Signal.* **18**, 91 (2020).

151. Balasubramaniam, D. *et al.* How the Ankyrin and SOCS Box Protein, ASB9, Binds to Creatine Kinase. *Biochemistry* **54**, 1673–1680 (2015).

152. Hatakeyama, S. TRIM Family Proteins : Roles in Autophagy, Immunity, and Carcinogenesis. *Trends Biochem. Sci.* **42**, 297–311 (2017).

153. Zeng, L. *et al.* Structural insights into human KAP1 PHD finger–bromodomain and its role in gene silencing. *Nat. Struct. Mol. Biol.* **15**, 626–633 (2008).

154. Liu, J. & Nussinov, R. The Mechanism of Ubiquitination in the Cullin-RING E3 Ligase Machinery: Conformational Control of Substrate Orientation. *PLoS Comput. Biol.* **5**, e1000527 (2009).

155. Liu, J. & Nussinov, R. Molecular Dynamics Reveal the Essential Role of Linker Motions in the Function of Cullin-RING E3 Ligases. *J. Mol. Biol.* **396**, 1508–1523 (2010).

156. Liu, J. & Nussinov, R. Flexible Cullins in Cullin-RING E3 Ligases Allosterically. **286**, 40934–40942 (2011).

157. Maser, R. S. *et al.* Chromosomally unstable mouse tumours have genomic alterations similar to diverse human cancers. *Nature* **447**, 966–971 (2007).

158. Yumimoto, K. & Nakayama, K. I. Recent insight into the role of FBXW7 as a tumor suppressor. *Semin. Cancer Biol.* **67**, 1–15 (2020).

159. Akhoondi, S. *et al.* FBXW7/hCDC4 is a general tumor suppressor in human cancer. *Cancer Res.* **67**, 9006–9012 (2007).

160. Wang, Z. *et al.* The two faces of FBW7 in cancer drug resistance. *BioEssays* **33**, 851–859 (2011).

161. Shimizu, K., Nihira, N. T., Inuzuka, H. & Wei, W. Physiological functions of FBW7 in cancer and metabolism. *Cell. Signal.* **46**, 15–22 (2018).

162. Agilent Technologies. QuickChange Primer Design. Available at: https://www.agilent.com/store/primerDesignProgram.jsp.

163. Hodge, K., Have, S. Ten, Hutton, L. & Lamond, A. I. Cleaning up the masses: Exclusion lists to reduce contamination with HPLC-MS/MS. *J. Proteomics* **88**, 92–103 (2013).

164. Barril, X. Computer-aided drug design: time to play with novel chemical matter. *Expert Opin. Drug Discov.* **12**, 977–980 (2017).

165. Hao, B., Oehlmann, S., Sowa, M. E., Harper, J. W. & Pavletich, N. P. Structure of a Fbw7-Skp1-Cyclin E Complex: Multisite-Phosphorylated Substrate Recognition by SCF Ubiquitin Ligases. *Mol. Cell* **26**, 131–143 (2017).

166. Jensen, T. J. *et al.* Multiple proteolytic systems, including the proteasome, contribute to CFTR processing. *Cell* **83**, 129–135 (1995).

167. Zhang, L., Lu, Q. & Chang, C. *Epigenetics in Health and Disease*. *Advances in Experimental Medicine and Biology* **1253**, (Springer Singapore, 2020).

168. Nakajima, H. & Kunimoto, H. TET2 as an epigenetic master regulator for normal and malignant hematopoiesis. *Cancer Sci.* **105**, 1093–1099 (2014).

169. Solary, E., Bernard, O. A., Tefferi, A., Fuks, F. & Vainchenker, W. The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia* **28**, 485–496 (2014).

170. Cimmino, L. *et al.* Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. *Cell* **170**, 1079–1095 (2017).

171. Bray, J. K., Dawlaty, M. M., Verma, A. & Maitra, A. Roles and Regulations of TET Enzymes in Solid Tumors. *Trends in Cancer* **7**, 635–646 (2021).

172. Jiang, S. Tet2 at the interface between cancer and immunity. *Commun. Biol.* **3**, 1–6 (2020).

173. Puig, I. *et al.* TET2 controls chemoresistant slow-cycling cancer cell survival and tumor recurrence. *J. Clin. Invest.* **128**, 3887–3905 (2018).

174. Chua, G. N. L. *et al.* Cytosine-Based TET Enzyme Inhibitors. *ACS Med. Chem. Lett.* **10**, 180–185 (2019).

175. Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K. & Schapira, M. Epigenetic protein families: A new frontier for drug discovery. *Nat. Rev. Drug Discov.* **11**, 384–400 (2012).

176. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214–231 (2012).

177. Cochran, A. G., Conery, A. R. & Sims, R. J. Bromodomains: a new target class for drug development. *Nat. Rev. Drug Discov.* **18**, 609–628 (2019).

178. Gallenkamp, D., Gelato, K. a., Haendler, B. & Weinmann, H. Bromodomains and their pharmacological inhibitors. *ChemMedChem* **9**, 438–464 (2014).

179. Chaidos, a., Caputo, V. & Karadimitris, A. Inhibition of bromodomain and extra-terminal proteins (BET) as a potential therapeutic approach in haematological malignancies: emerging preclinical and clinical evidence. *Ther. Adv. Hematol.* **6**, 128–141 (2015).

180. Liang, D., Yu, Y. & Ma, Z. Novel strategies targeting bromodomain-containing protein 4 (BRD4) for cancer drug discovery. *Eur. J. Med. Chem.* **200**, 112426 (2020).

181. Lu, T., Lu, W. & Luo, C. A patent review of BRD4 inhibitors (2013-2019). *Expert Opin. Ther. Pat.* **30**, 57–81 (2020).

182. Zengerle, M., Chan, K. H. & Ciulli, A. Selective Small Molecule Induced Degradation of the BET Bromodomain Protein BRD4. *ACS Chem. Biol.* **10**, 1770–1777 (2015).

183. Aldeghi, M. *et al.* Large-scale analysis of water stability in bromodomain binding pockets with grand canonical Monte Carlo. *Commun. Chem.* **1**, 19 (2018).

184. Zhong, H. *et al.* Importance of a crystalline water network in docking-based virtual screening: A case study of BRD4. *Phys. Chem. Chem. Phys.* **21**, 25276–25289 (2019).

185. Nittinger, E. *et al.* Water molecules in protein–ligand interfaces. Evaluation of software tools and SAR comparison. *J. Comput. Aided. Mol. Des.* **33**, 307–330 (2019).

186. Cui, H. *et al.* Selective N-Terminal BET Bromodomain Inhibitors by Targeting Non-Conserved Residues and Structured Water Displacement. *Angew. Chemie - Int. Ed.* **60**, 1220–1226 (2021).

187. Álvarez-García, D. & Barril, X. Molecular Simulations with Solvent Competition Quantify Water Displaceability and Provide Accurate Interaction Maps of Protein Binding Sites. *J. Med. Chem.* **57**, 8530–8539 (2014).

188. Filippakopoulos, P. *et al.* Benzodiazepines and benzotriazepines as protein interaction inhibitors targeting bromodomains of the BET family. *Bioorganic Med. Chem.* **20**, 1878–1886 (2012).

189. Matthias, G. *et al.* A bump-and-hole approach to engineer controlled selectivity of BET bromodomain chemical probes. *Science (80-.).* **346**, 638–641 (2014).

190. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).

191. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).

192. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).

193. Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 293–302 (2011).

194. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).

195. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).

196. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).

197. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

198. Pearce, N. M. *et al.* A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.* **8**, 24–29 (2017).

199. Lu, Y., Wang, R., Yang, C. Y. & Wang, S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.* **47**, 668–675 (2007).

200. Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov. Today* **11**, 580–594 (2006).

201. Spyrakis, F. *et al.* The Roles of Water in the Protein Matrix: A Largely Untapped Resource for Drug Discovery. *J. Med. Chem.* **60**, 6781–6827 (2017).

202. Shadrick, W. R. *et al.* Exploiting a water network to achieve enthalpy-driven, bromodomain-selective BET inhibitors. *Bioorganic Med. Chem.* **26**, 25–36 (2018).

203. Hewings, D. S. *et al.* Progress in the Development and Application of Small Molecule Inhibitors of Bromodomain–Acetyl-lysine Interactions. *J. Med. Chem.* **55**, 9393–9413 (2012).

204. Crawford, T. D. *et al.* Diving into the Water: Inducible Binding Conformations for BRD4, TAF1(2), BRD9, and CECR2 Bromodomains. *J. Med. Chem.* **59**, 5391–5402 (2016).

205. Graham, S. E., Smith, R. D. & Carlson, H. A. Predicting Displaceable Water Sites Using Mixed-Solvent Molecular Dynamics. *J. Chem. Inf. Model.* **58**, 305–314 (2018).

206. Millan, D. S. *et al.* Design and Optimization of Benzopiperazines as Potent Inhibitors of BET Bromodomains. *ACS Med. Chem. Lett.* **8**, 847–852 (2017).

207. Abner, E. *et al.* A New Quinoline BRD4 Inhibitor Targets a Distinct Latent HIV-. *J. Virol.* **92**, 1–22 (2018).

208. Gehling, V. S. *et al.* Discovery, design, and optimization of isoxazole azepine BET inhibitors. *ACS Med. Chem. Lett.* **4**, 835–840 (2013).

209. Yu, J. L. *et al.* NMR-based platform for fragment-based lead discovery used in screening BRD4-targeted compounds. *Acta Pharmacol. Sin.* **37**, 984–993 (2016).

210. Long, F. *et al.* AceDRG: A stereochemical description generator for ligands. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 112–122 (2017).

211. Smart, O. S. *et al.* grade. (2011).

212. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).

213. Kirsch, P., Hartman, A. M., Hirsch, A. K. H. & Empting, M. Concepts and Core Principles of Fragment-Based Drug Design. *Molecules* **24**, 4309 (2019).

214. de Souza Neto, L. R. *et al.* In silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery. *Front. Chem.* **8**, 1–18 (2020).

215. Lim, S. *et al.* bioPROTACs as versatile modulators of intracellular therapeutic targets including proliferating cell nuclear antigen (PCNA). *Proc. Natl. Acad. Sci.* **117**, 5791–5800 (2020).