In silico tools to aid medicinal chemistry: optimising bromodomain inhibitors

Joseph P. Bluck

Exeter College

University of Oxford

A thesis submitted in partial fulfilment for the degree of

Doctor of Philosophy at the University of Oxford

Trinity 2019

For Colin Brown.

This thesis would not exist without his time, advice and encouragement.

Acknowledgements

It is hard to put into words how grateful I am to my supervisors for everything that they have taught me. I want to thank Prof. Phil Biggin for always having an open door and allowing me to explore areas of computational research. I always feel I learned something from our meetings and always felt enthusiastic after them. I also want to thank Prof. Stuart Conway for making me feel very welcome within his group, and helping me explore medicinal chemistry from a different angle. Through Stuart I was given a board view of drug discovery and has definitely pushed me to continue in the field. Finally, my thanks go to Prof. Garrett Morris who supported me in both his roles of supervisor and course co-director. He always offered a different point of view and also had more useful tricks in python. I would also like to extend my gratitude to my industrial sponsors at GSK. Both Dr. Phil Humphreys and Francesco Rianjongdee worked as my co-supervisors and always offered valuable feedback, especially from an industrial perspective. In addition, I would also like to thank Prof. Charlotte Deane, for welcoming me to the Oxford Protein Informatics Group (OPIG) and helping me see a breadth of science within the group.

I have had the pleasure of working in three fantastic research groups during my DPhil, with too many people to thank in this all too brief section. However, there are a few I owe extra special thanks to. Firstly, to Dr Irfan Alibay for going above and beyond in helping me get to through the DPhil. Naturally, to Coco Laurin for her constant enthusiasm and support. Special thanks go to Dr Matteo Aldeghi, for introducing me to *in silico* work and his support in pointing me in the right direction when I got lost. I also owe a lot of thanks to the fantastic TRIM team: Dr. Larissa See, Amy Scorah and Jess Reynolds. I couldn't have asked for a better group of people to tackle a project with. To Andrew Boczek, for being a fantastic Part II student and an annoyingly good squash partner. Also, to Susan Leung for answering all my RDKit questions and being a fantastic housemate, along with Alex Skates.

There are also many people to thank for making such a fantastic atmosphere to work in. Starting with Dr Catherine Haslam, for helping me with synthesis and for all the great fun we had on the Kef antibiotic project, despite it not going in this thesis. The same goes for other brilliant members of the Conway group: Dr Ewen Calder, Dr Deborah Sneddon, Dr Amelie Joffrin, Dr James Clayton, Dr Matthias Schiedel, Dr Anthony Chan, Dr Sarah Collins, Dr Catherine Haslam,

Mustafa Morologu, Holly Bunce, Adam Thomas, Darius McArdle and Katrina Andrews. I also have my fellow residents of the SBCB to thank for their support: Will Glass, Dr Afroditi Maria Zaki, Dr Maria Musgaard, Dr Naushad Velgy, Max Epstein, Marc Dämgen, Rocco Meli, Jodie Collingridge and Mina Moniri. It is also hard to forget all the members of OPIG, for taking me out of my comfort zone and providing excellent cakes. And a massive thank you to everyone I've had to miss off, otherwise we would never get to the research.

After eight years at Oxford, I would also like to pass on my thanks to Exeter College. It has served many roles and shaped my experience of Oxford. I want to especially thank the MCR for letting me represent you and for providing me with life long friends. Also to all the college staff, who have been extremely helpful and made the College such a welcoming place. I am also extremely grateful for the financial support provided by College, and by other awarding bodies. This DPhil would not have been possible without the generous support of GSK, the EPSRC and the MRC.

I want to end by thanking those closest to me. Firstly, I am forever grateful to both my Mum and Dad for their unending support over the years. To my brothers, for supporting me in their own brotherly way and for becoming decent drinking partners, along with my Grandparents for checking in on me weekly. Finally, I want to thank my wonderful partner, Henry. He has been incredibly supportive and understanding from the moment I met him, and helped me through through writing this thesis. I am an extremely lucky man to have him in my life.

Abstract

Development of ligands for bromodomains has expanded our understanding of gene regulation and epigenetic links to disease. These acetylated lysine 'reader' domains are found within larger macromolecules that are involved in gene transcription and degradation. This thesis looks to employ a range of computational tools to compliment the development of inhibitors to a selection bromodomain targets. The first target was TRIM33β, an E3 ubiquitin ligase that also contains a methylated lysine reader, known as a plant homeodomain. After identifying putative binding locations for ligands identified in a previously performed high-throughput screen through docking and molecular dynamics studies, one ligand was taken forward for further optimisation. Subsequent free energy calculations on both the ligand and water molecules were used to justify a proposed binding model, in the absence of a co-crystal structure. This model identified a stable water molecule in the ZA channel and was used to design the highest affinity ligands known for the TRIM33β bromodomain. The model was also scrutinised to understand why vector elaborations above the binding site did not yield a higher affinity probe molecule.

The second set of targets were bromodomains within parasites responsible for a range of tropical diseases. In the absence of co-crystal structures, homology models and molecular dynamics studies were used to design binding models for a set of existing bromodomain ligands. These models were used to aid development of tool compounds, based on the BRD4(1)/PLK1 inhibitor BI-2536, and propose future modifications to increase affinity for parasite bromodomains. Two of the homology models were later confirmed by a subsequent co-crystal structure. Finally, several computational methods are employed to understand the trends in binding affinity for a range of ligands against the CREBBP bromodomain and BRD4(1). These studies highlighted the importance of an intramolecular hydrogen bond in analogues of OXFBD02, the role of a ZA channel water molecule on imposing selectivity for CREBBP ligands, and supported the role of a proposed cation-π interaction.

Contents

List of Figures

List of Tables

List of Abbreviations

ABFE . . . . . Absolute binding free energy

ALK . . . . . Anaplastic lymphoma kinase

AlphaScreen™ Amplified luminescent proximity homogeneous assay screen

AML . . . . . Acute myeloid leukaemia

BAZ . . . . . Bromodomain adjacent to zinc finger domain

BCP . . . . . Bromodomain containing protein

BD1 . . . . . First bromodomain of two

BD2 . . . . . Second bromodomain of two

BET . . . . . Bromodomain and extra terminal domain

BRD . . . . . Bromodomain

CADD . . . . . Computer aided drug design

CCDC . . . . . Cambridge crystallographic data centre

CPU . . . . . Computer processing unit

CREBBP . . . cAMP-response element-binding protein

DDQ 2,3-Dichloro-5,6-dicyano-1,4-benzoquinone

eq . . . . . Equivalent

eqn . . . . . Equation

FEP . . . . . Free energy perturbation

FRET . . . . . Fluorescence resonance energy transfer

GAFF . . . . . General Amber force field

GCI . . . . . Grand canonical integration

GCMC . . . . Grand canonical Monte Carlo

GPU . . . . . Graphics processing unit

h . . . . . Hour(s)

HAT . . . . . Histone acetyltransferase

HDAC . . . . . Histone deacetylase

HRMS . . . . High resolution mass spectrometry

HTS . . . . . High-throughput screen(ing)

IFD . . . . . Induced fit docking

ITC . . . . . Isothermal titration calorimetry

KAc . . . . . Acetylated lysine residue

KDE . . . . . Kernal density estimate (plot)

KMe0 . . . . . Non-methylated lysine residue

KMe2 . . . . . Di-methylated lysine residue

KMe3 . . . . . Tri-methylated lysine residue

Ld . . . . . *Leishmania donovani*

LJ . . . . . Lennard Jones

LJ-PME . . . Lennard Jones Particle-Mesh

LRMS . . . . . Low resolution mass spectrometry

mBAR . . . . Multistate Bennett's acceptance ratio

MC . . . . . Monte Carlo

MD . . . . . Molecular dynamics

min . . . . . Minute(s)

MM . . . . . Molecular mechanics

MMPBSA . . Molecular mechanics Poisson-Boltzmann surface area

MOE . . . . . Molecular operating environment

m.p. . . . . . Melting point

NMC . . . . . Nut-midline carcinoma

NMR . . . . . Nuclear magnetic resonance

NOE . . . . . Nuclear Overhauser effect

PAINS . . . . Pan-assay interference compounds

PDB . . . . . Protein data bank

PHD . . . . . Plant homeodomain

PLK1 . . . . . Polo-like kinase 1

PME . . . . . Particle-mesh Ewald

PROTAC . . . Proteolysis-targeting chimera

PTM . . . . . Post-translational modification

QSAR . . . . . Quantitive structure-activity relationship

RBFE . . . . . Relative binding free energy

R_f . . . . . Retention factor

RING . . . . . Really interesting new gene (domain)

RMSD . . . . . Root-mean-square deviation

RMSF . . . . . Root-mean-square fluctuation

rt . . . . . Room temperature

SAR . . . . . Structure-activity relationship

SBDD . . . . . Structure-based drug design

SFI . . . . . Structure factor index

Sm . . . . . *Schistosoma mansoni*

SMART . . . . SMILES arbitrary target specification

SMILE . . . . Simplified molecular-input line-entry system

Tb . . . . . *Trypanosoma brucei*

Tc . . . . . *Trypanosoma cruzi*

THQ . . . . . Tetrahydroquinoline

TLC . . . . . Thin-layer chromotography

TRIM . . . . . Tripartite motif containing

vdW . . . . . van der Waals

VHL . . . . . von Hippel-Lindau

VS . . . . . Virtual screen(ing)

WaterLOGSY Water-ligand observed *via* gradient spectroscopy

WT . . . . . Wild type

Are you really going to spend all night finding quotes for your thesis?

— Henry Sawczyc

1 Introduction

Contents

This chapter will introduce bromodomains as a therapeutic target, both in human cancers and as a putative tropical disease treatment, providing the relevant background needed to contextualise the research in this thesis. Firstly, the chapter will introduce several key structural aspects of the bromodomain before introducing the targets of interest in this thesis. The discussion will then move to onto highlighting how computer aided drug design (CADD) methods can be used to

enhance medicinal chemistry projects. Finally, the chapter will give an overview of the assays used by colleagues to obtain the experimental data used within this thesis.

1.1 Epigenetic 'readers'

The genetic material within each eukaryotic cell is extensive, requiring a complex packaging solution in the form of chromatin. Chromatin relies on building blocks, known as nucleosomes, to wind the DNA around.[1] These nucleosomes are an octamer of four histone proteins: H2A, H2B, H3 and H4. Each of these histone proteins forms part of the nucleosome core, shown in Figure 1.1, with an extended histone tail that protrudes from the core.[2] Both the nucleosomes and the DNA itself can be subjected to covalent post translational modifications (PTMs), which can alter the charge and size of components within chromatin. These changes can led to subsequent alterations in DNA packaging of the nucleosomes, and hence chromatin.[3] These PTMs, especially those exposed on the histone tails, can also act as recognition and binding sites for proteins such as transcription factors. These PTMs and their impact on genetic regulation are being increasingly linked to disease states, including cancer,[4] and so understanding the roles of each modification is vital. This is especially important when considering the multitude of PTM combinations possible, which has become known as the histone code.[5]

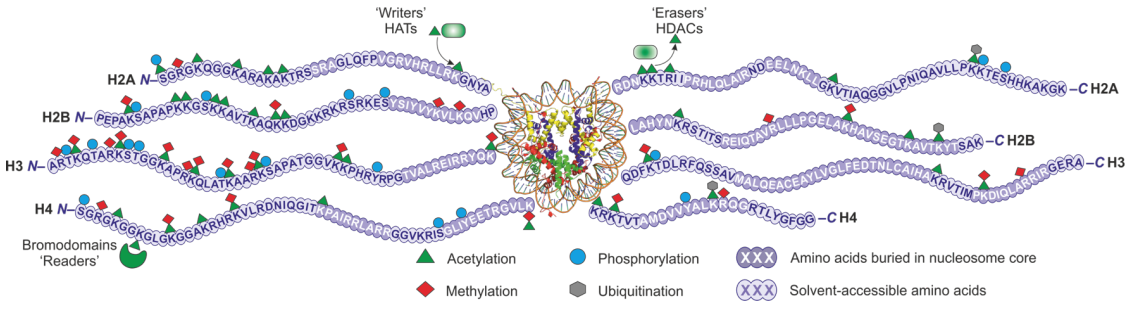


Figure 1.1: A cartoon representation of the nucleosome octamer. Approximately 150 base pairs of DNA are wrapped around the nucleosome core, which contains an octamer of 4 different histone proteins. Each of these proteins contain a tail that protrudes from the core which is subject to a range of PTMs, which have been highlighted. Figure reproduced with permission from Jennings *et al.*, Future Medicinal Chemistry.[6] Copyright 2014 Future Science.

Understanding this histone code is the focus of the field of epigenetics, which looks at the heritable changes in phenotype that are not due to changes of the underlying DNA sequence.[4] These changes can occur on the histone tail residues, shown in Figure 1.1. The tails can undergo various covalent modifications including methylation of lysine, acetylation of lysine and arginine, phosphorylation and ubiquination.[7] These marks are placed with 'writer' domains such as histone acetyltransferases (HATs), detected by 'reader' domains like the bromodomains (BRDs) and removed with 'eraser' domains including histone deacetylases (HDACs). These domains are all part of larger protein bodies that often are associated with a particular function.[4]

The complex regulation of these marks is important for cellular health and therefore is an area of therapeutic interest.[3] For example, acetylation of lysine is linked to formation of less densely packed chromatin (euchromatin) and consequently expression of genes including tumor suppressors. Reader domains facilitate the recruitment of macromolecules to chromatin that are involved in gene regulation, through selective recognition of PTMs, or 'marks', on the histone proteins. These include tudor domains and plant homeodomains (PHDs) which bind to methylated lysine residues, and bromodomains (BRDs) that bind to acetylated lysine residues. This role has meant many reader domains have been linked to various disease states. These include the bromodomain and extra terminal domain (BET) family of bromodomain containing proteins (BCPs), comprising bromodomain-containing proteins 2-4 (BRD2-4) and the testis-specific BRDT. These have been linked to various cancers, including Nut-midline carcinoma (NMC) and breast cancer.[8] However, the disease link for many reader domains and their associated proteins are yet to be understood fully, partially due to a lack of selective chemical probes.

1.1.1 Multiple-domain-containing proteins

Multiple-domain-containing proteins are often joined *via* flexible linkers, allowing for flexibility in the macromolecule conformation and hence orientation of the different domains within the protein.[9] This allows for proteins to contain more

than one epigenetic reader, writer or eraser domain and ultimately carry out complex functions.[3] For example, members of the BET family each contain two adjacent canonical bromodomains. This complicates the understanding the role of each protein within the histone code, as each domain within these proteins plays a separate role in recognition.

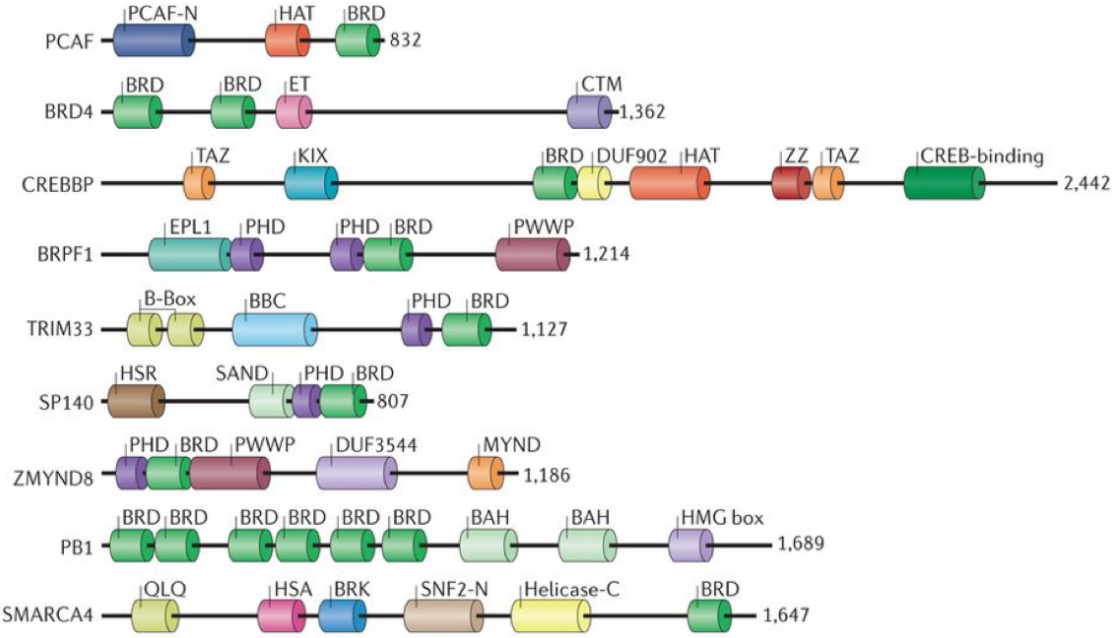


Figure 1.2: Example of multiple domain proteins involved within epigenetics. Reprinted with permission from Fujisawa *et al.*, Nature Reviews Molecular Cell Biology.[3] Copyright 2017 Nature publishing group.

1.1.2 Plant homeodomains

Plant Homeodomains (PHDs) have been identified as readers of methylated lysine and are found within a variety of proteins that regulate gene expression.[10] Since the PHD was first identified in 1995 as a chromatin-related mediator of transcriptional regulation by Aasland *et al.*,[11] over a hundred of putative PHD sequences have been identified and loosely grouped into functional families.[12, 13] PHDs have been found to predominantly read the methylation state of the K4 residue of the H3 histone peptide (H3K4 - $H3K4Me_3$, $H3K4Me_2$ or $H3K4Me_0$), and to a lesser extent the H3K9 and H3R2 positions.[14, 15]

PHDs are zinc finger containing proteins that consist of 50-80 amino acids built into a two-strand anti-parallel β-sheet motif, and are stabilised by two buried Zn^{2+} ions. The typical zinc finger motif of a collection of Cys_2His_2 residues co-ordinating Zn^{2+}, is contained within multiple chromatin mediating proteins.[16] The PHD has a slightly different motif to chelate Zn^{2+}, instead using Cys_4–His–Cys_3 combination of residues to make a cross-braced fold, shown in Figure 1.3.[14] This fold is typically retained despite a larger range of sequence variation between PHDs.[13] There have also been some exceptions where the Cys and His chelating residues have been interchanged in some PHDs with little impact on the tertiary structure.[17] The PHD structure is reliant on Zn^{2+} co-ordination, where Pascual *et al.* showed that protein production in a low zinc environment or purification in the presence of EDTA led to no protein production or the unfolded protein, respectively.[18] This observation was also seen within the MD studies within this thesis (Chapter 3).

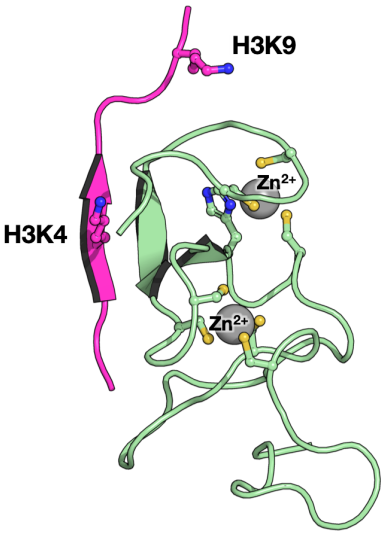


Figure 1.3: Structure of the plant homeodomain (PHD). The zinc ion co-ordinating residues are shown as sticks and the H3 peptide is shown in magenta. This figure shows the PHD of TRIM24 (PDB ID: 3O37).[19]

Binding of the PHD to the H3K4 mark is *via* β-sheet formation on the PHD surface of the N-terminus of the H3-peptide, shown in Figure 1.3. Typical recognition

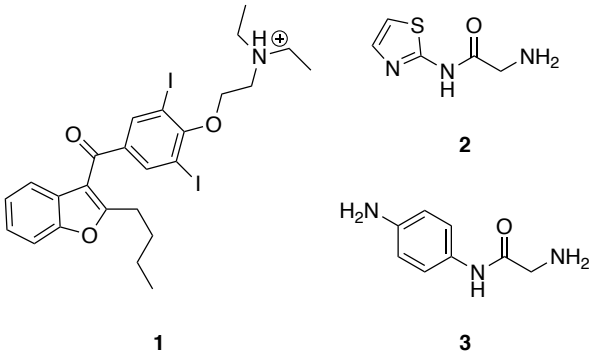


Figure 1.4: Ligands that bind to PHD domains.[20, 21]

of the H3K4Me$_3$ mark is *via* an aromatic cage of three residues adjacent to the H3 binding region, although the composition of this cage varies amongst PHDs. In the case of H3K4Me$_0$ binding, recognition of the mark is through hydrogen bonds between the ζ-nitrogen of the lysine and backbone carbonyl groups of nearby residues. In a subset of PHDs, recognition of the K9Me$_3$ mark is through a cation-π interaction with an aromatic residue perpendicular to the H3K4 binding region.[15]

Given the surface protein-protein interaction-like nature of lysine recognition by the PHD, there are few ligands that inhibit PHDs. Wagner *et al.* identified a set of inhibitors for the PHD of JARID1A, a protein which recognises the H3KMe$_3$ mark, which is important for the progression of acute myeloid leukaemia (AML). The ligands identified are analogues of the anti-arrhythmia drug Amiodarone (**1**), and only show partial selectivity amongst differing PHDs.[20] A similar issue is seen in a fragment-based study performed by Amato *et al.* to identify ligands for the PHDs found within BAZ2A and BAZ2B.[21] These are members of the bromodomain adjacent to zinc finger domain (BAZ) family of proteins (BAZ1A, BAZ1B, BAZ2A and BAZ2B) and have been established as a potential therapeutic targets.[22–24] For PHD inhibition, a scaffold that could displace the H3 peptide was developed, shown by the ligands **2** and **3**.[21] However, both these ligands target the H3K4Me$_3$ mark, and there are currently no published ligands to inhibit the H3K9Me$_3$ interaction.

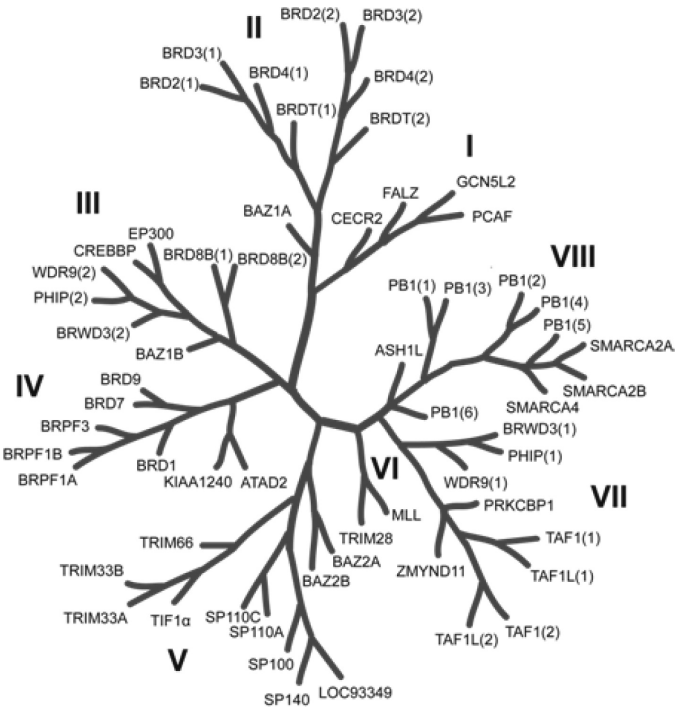


Figure 1.5: Human bromodomain phylogenetic tree showing bromodomain sub-families. The phylogenetic free shows the 61 bromodomains found within 46 proteins. The figure has been subdivided into families, labelled using roman numerals. Reprinted with permission from Hewings *et al.*, J. Med. Chem.[27] Copyright 2012 American Chemical Society.

1.1.3 Bromodomains

Bromodomains (BRDs) are another class of epigenetic readers domains, which bind to acetylated lysine (KAc) residues on histones.[25] There have been 61 BRDs identified within 46 bromodomain-containing proteins (Figure 1.5). They have been clustered into eight families based of their sequence homologies.[3] These are evolutionarily-conserved proteins that were first identified in the *brahma* gene within *Drosophila melanogaster* by Haynes *et al.*[26] As with PHDs, these proteins are often associated with regulation of gene expression. Bromodomain-containing proteins can act as transcription factors, scaffolds for the assembly of larger protein complex, or perform catalytic functions that drive chromatin modification. These proteins have also been show to vary in expression levels in different tissues, indicating a context-dependent nature to their cellular roles.[3]

The BRD structure is relatively simple, comprising of around 110 amino acid

residues. Bromodomains consist of a left handed bundle of four α-helices (α_Z, α_A, α_B and α_C) which are joined by the ZA, AB and BC loops, respectively. The KAc-binding site is found between the BC and ZA loops, where the composition of these loops determines the selectivity of the histone mark read.[7] In the majority of bromodomains, the KAc residue binds to a highly conserved Asn residue within the binding site. Below this residue, at the bottom of the binding site sit five conserved water molecules, which will be labelled W0-W4 and shown in Figure 1.6. These water molecules are typically non-displacable, however, a computational study on 35 by Aldeghi *et al.* showed that some water networks can be interrupted to allow binding.[28] This work has been validated in later studies, such as an inhibitor for the tri-partite motif containing (TRIM) protein TRIM24 developed by Liu *et al.* which contains a polar group that disrupts the water network.[29]

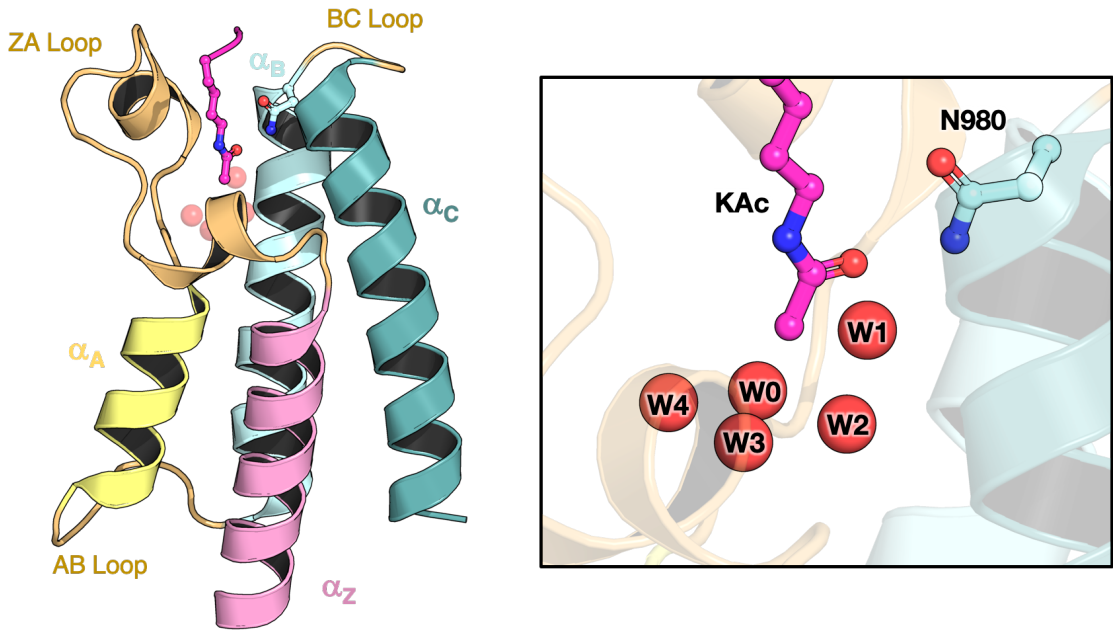


Figure 1.6: Structure of the canonical bromodomain (BRD). The acetylated lysine (KAc) residue (magenta) binds to a highly conserved Asn residue in the BRD binding site. Within this binding site are a network of conserved water molecules. Note that the colour scheme for secondary structure elements is conserved throughout the thesis. The figure shows the BRD of TRIM24 shown (PDB ID: 3O36).[19]

1.1.4 Inhibition of bromodomains

Unlike in the PHDs, the deeper binding sites found within BRDs have led to many successful probes and drug candidates, which will be described in this sub-section. Vidler *et al.* provide a scale of the druggability of 24 of the 61 known BRDs using available crystallographic structures. This predicts that the BET family of BRDs are easily druggable, with the CREBBP protein slightly harder to target and proteins such as TRIM24 comparatively difficult to target.[30] This observation aids in rationalising why several bromodomains are yet to have high affinity, selective ligands and the challenge faced in future development.

The BET family of BRD containing proteins are linked to several cancerous states through oncogene regulation, genetic alteration and transcriptional reprogramming. This family has been the target of multiple probes and there are currently 18 BET ligands within clinical trials.[31]. Trials of BET inhibitors in patients with NMC showed cases of tumour regression or disease stabilisation or partial clearance within a subset of AML patients.[8] The BET proteins contain two BRDs at their N-terminus, known as BD1 and BD2. However, the similarity between the different BET family members makes designing selective ligands a challenge. Two ligands for the BET BRDs, (+)-JQ1 (**4**) and I-BET762 (**5**), were identified using phenotypic screens and published in 2010.[32, 33] Both these ligands contained a diazepine scaffold, which acts as a KAc mimic, binding to the Asn residue in the binding site, and is also capable of interacting with water molecule W1.[5] Both **4** and **5** bind both BD1 and BD2 of the tandem BRD4(1/2) construct, with an IC_{50} of 36 nM (determined by FRET) and 77 nM (determined by AlphaScreen™), respectively.[32, 33]

Selectivity between BD1 and BD2 in the BET family was first achieved using a variant of I-BET762, **6**. This was combined with the L94A mutation of BRD4(1) to impose selectivity over the tandem BRDs, by ensuring the ligand could only fit in the mutated pocket. This 'bump and hole' approach allowed for the discovery that inhibition BD1 in BRD4 is sufficient to prevent chromatin binding.[34] Selectivity of the wildtype BET proteins was achieved with GSK340 (**7**), which is BD2 selective

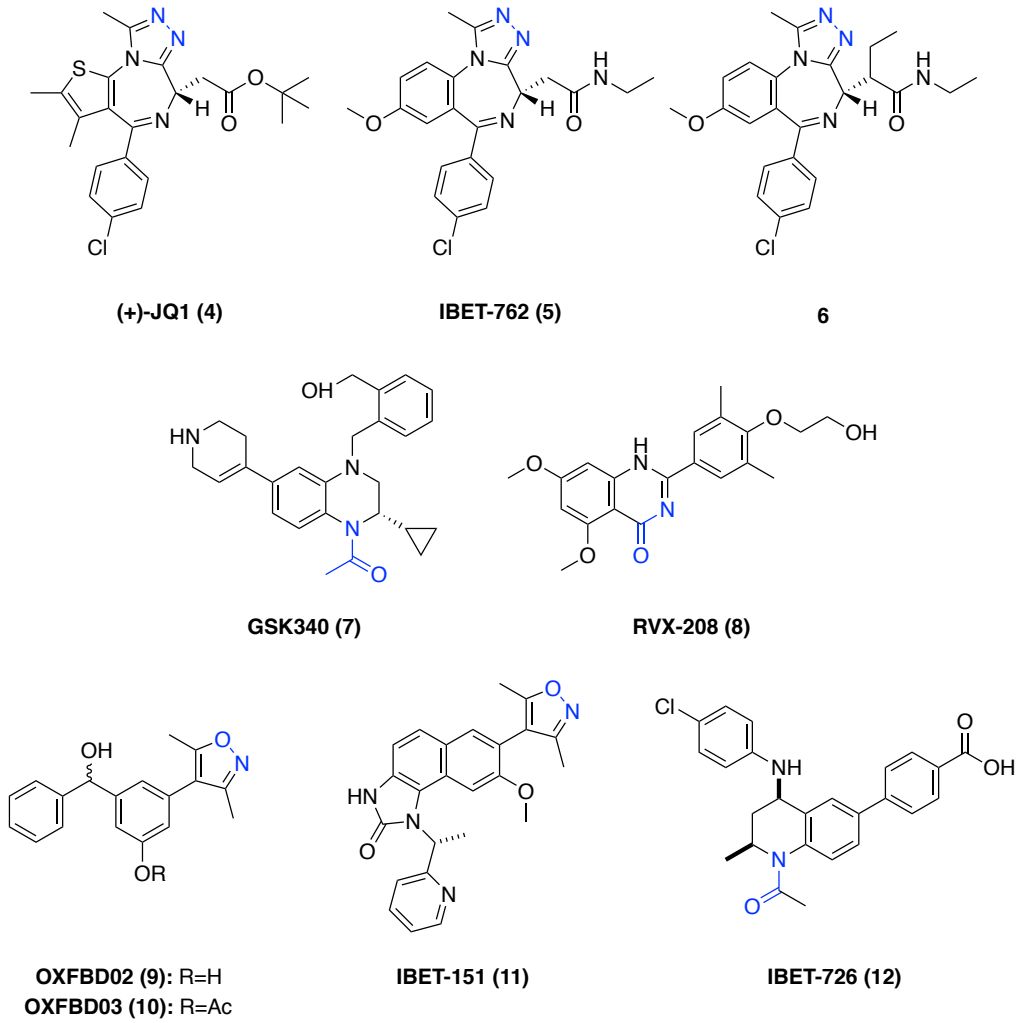


Figure 1.7: BET bromodomain ligands. KAc mimicking moieties are highlighted in blue.

within the BET family while also retaining >300 fold selectivity over non-BET BRDs.[35] In addition Apabetalone, also known as RVX-208 (**8**), is another BET ligand that holds a 40-fold selectivity for BD1 over BD2 in BRD4. In this ligand an amide group acts as the KAc mimic.[36, 37] This compound underwent Phase II clinical trials for atherosclerosis and is currently undergoing Phase III trials.[38, 39]

The 3,5-dimethylisoxazole group has been used as a KAc mimic in both BET and non-BET ligands. This group features in both OXFBD02 (**9**) and OXFBD03 (**10**). These ligands were developed by Hewing *et al.* and optimised for BET binding through addition of a lipophilic phenyl group.[40] The geometry of this group allows

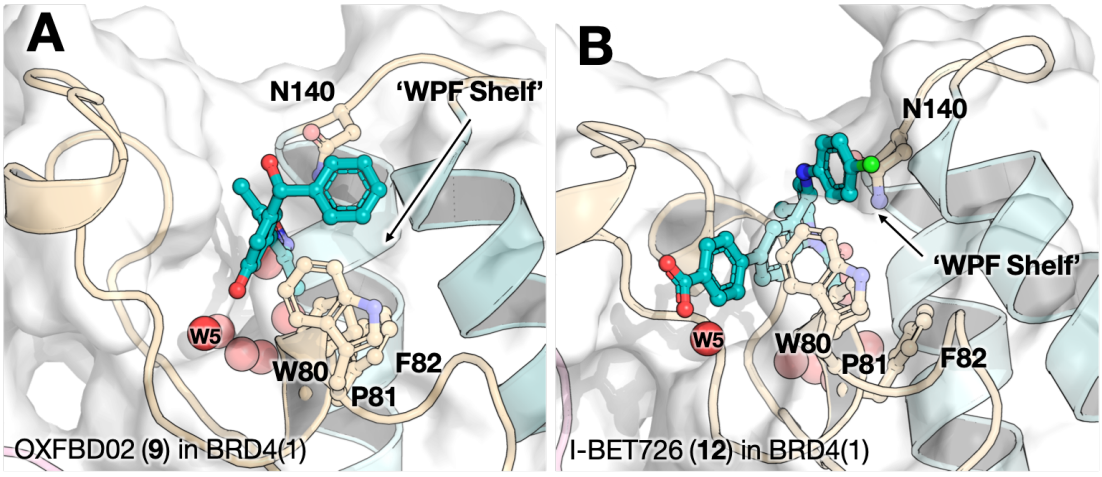


Figure 1.8: Co-crystal structures of (*R*)-OXFBD02 (**9**) and I-BET726 (**12**) binding to BRD4(1), taken from PDB accessions 4JOR and 4BJX, respectively.[40, 45] Binding of **9** is also dependent on a water molecule within the ZA channel, termed W5, which co-ordinates with the hydroxyl group off the core of **9**

for additional hydrophobic interactions with a hydrophobic shelf adjacent to the ZA channel, shown in Figure 1.8. This shelf, known as the WPF shelf (consisting of W81, P82 and F83 in BRD4(1)), is conserved among the BET proteins and offers a means of imposing BET selectivity. Independently, work at GSK led to the development of I-BET151 (**11**) which also contained an isoxazole group as the KAc mimic. This ligand, with a binding affinity of 36 nM (K_d by fluorescence anisotropy) is still BET selective, with the pyridyl group forming interactions with the WPF shelf.[41] This ligand has been shown to impede human ILC2 activation and dampens inflammation from type-2 immune response.[42] I-BET726 (**12**), which was patented in 2010 and published in 2016, also binds the tandem BRD4(1/2) construct with a K_d of 4 nM (by ITC).[43, 44] This structure uses an amide group as the KAc mimic, with a chlorobenzene group that is able to form interactions with the WPF shelf, shown in Figure 1.8.

While many ligands have been developed for the BET BRDs, fewer have been developed for the non-BET proteins. CREBBP is a non-BET human BRD-containing protein that , along with its paralogue p300, are acetyl-lysine transferases that have been shown to have over 400 interacting protein partners.[46, 47] These

partners include the tumour supression protein p53, where recognition of the KAc PTM on p53 is required for co-activator recruitment after DNA damage.[48] This complex interactome has made CREBBPs domains a target for inhibitor development, with several ligands of its BRD published in the literature. Zhou *et al.* identified several weakly binding fragments of the CREBBP BRD, MS2126 (**13**) and MS7972 (**14**), which were shown to block the interactions of CREBBP with p53.[49] Rooney *et al.* published a ligand that contained a modified version of this KAc-mimic in **15**, which also appended a tetrahydroquinoline (THQ) group to the core. The THQ group was found to form a cation-π interaction with R1173, which is adjacent to the binding site.[50] Genentec have also patented a series of similar compounds that are designed for use in cancer immunotherapy, due to the compounds role in T-cell regulation.[31, 51, 52]

As with the BET family, some notable probes have not been purely selective for just one BRD. IACS-9571 (**16**) is a dual TRIM24-BRPF1 binder. Palmer *et al.* used a structure-guided approach to development of a TRIM24 ligand, which has a benzimidalzone core that containing a KAc mimic. This ligand bound to both TRIM24, discussed in the next section, and BRPF1 which is a scaffolding protein used in assembly of acetyltransferase complex with affinities of 31 nM and 14 nM (by ITC), respectively.[53] This probe has since been used to develop a TRIM24 proteolysis-targeting chimera (PROTAC) molecule. This chimera of a TRIM24 ligand and a von Hippel-Lindau (VHL) protein ligand allows for whole, in-cell degradation of TRIM24 and highlighted a dependency of acute leukaemia on TRIM24.[54] Similar dual-specificity is seen in the both GSK2801 (**17**) and BAZ2-ICR (**18**) which are both inhibitors of BAZ2A/B.[55, 56] Interestingly, the initial compound used in the BAZ2-ICR optimisation did not require a KAc mimic to bind the BAZ2 proteins, shown within the co-crystal structure (PDB ID: 4XUA). Pan-bromodomain ligands have also been developed, notably seen with the SGC's bromosporine ligand (**19**) which targets multiple bromodomains.[57] This ligand was also used in a large pan-bromodomain computational study by Aldeghi *et al.* and

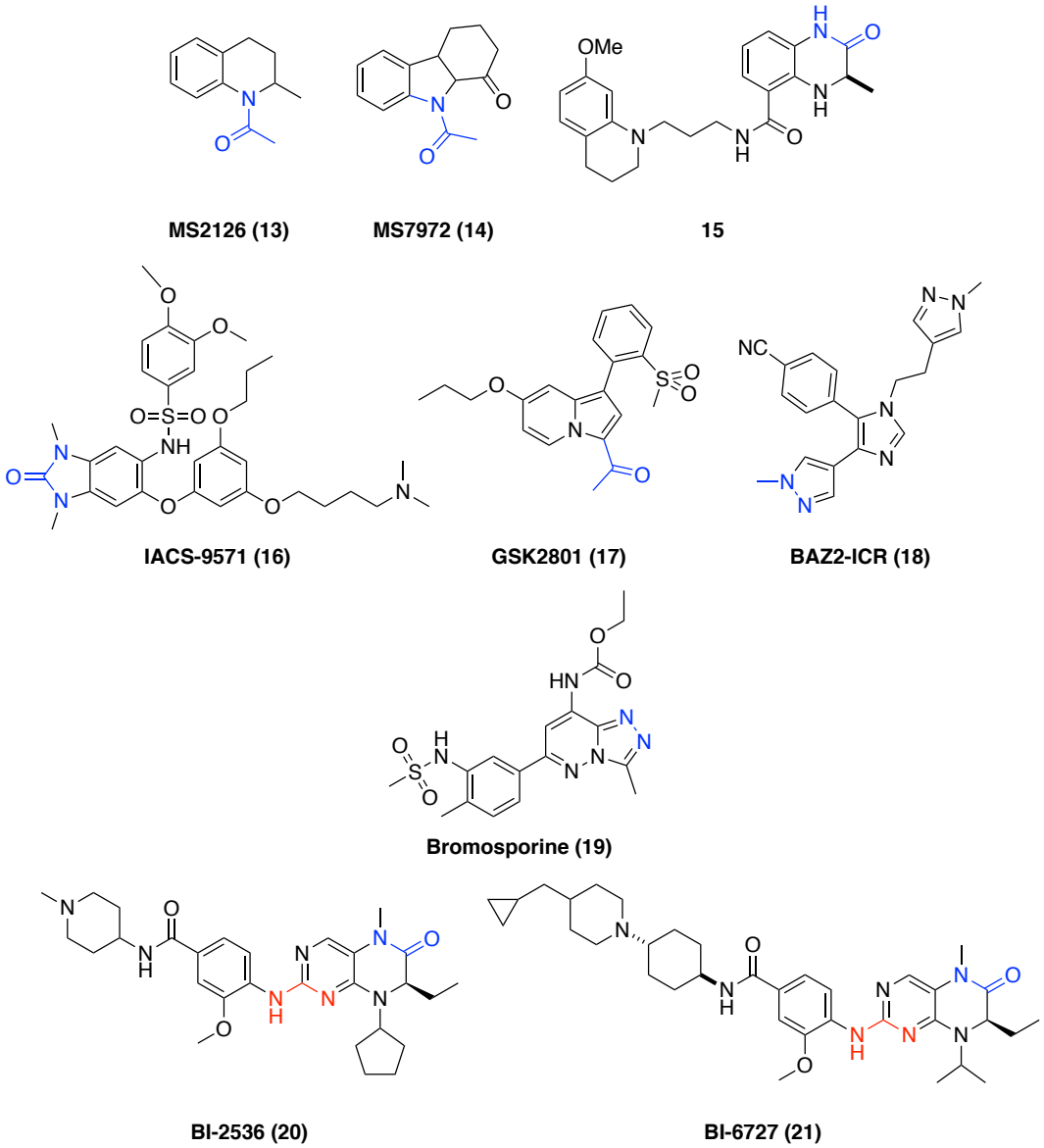


Figure 1.9: Ligands that bind to non-BET proteins. KAc mimic moieties have been highlighted in blue and the kinase binding hinge regions have been highlighted in red.

showed free energy predictions could correctly determine bromosporine selectivity for differing bromodomains and rank the potenital binding poses.[58]

Some BRD ligands also bind to non-BRD targets, such as BI-2536 (**20**) which is a dual BRD4 (37 nM) and polo-like kinase 1 (PLK1) kinase inhibitor (<2.6 nM).[59, 60] This compound contains a KAc mimic that facilitates binding to BRD4. In addition **20** contains a classical hinge motif, often observed in kinase ligands, which binds a Cys residue in the PLK1 binding site.[61] This versatile core has

been the subject of several studies to alter the kinase binding and selectivity. Watts *et al.* tuned the ligand to promote anaplastic lymphoma kinase (ALK) and BRD4 binding over PLK1, through modifications to the anisole group and hydrophobic bulk previously occupied by the cyclopentane moiety.[60] Structure-activity relationship (SAR) studies by Chen *et al.* and an additional patent showed that kinase selectivity can be prevented through removal of nitrogens within the hinge binding motif.[62, 63] An analogue of BI-2536, BI-6727 (**21**), shows an improvement in its pharmacokinetic profile, including a longer half life in mice allowing for greater therapeutic potential.[64]

1.2 TRIM24 and TRIM33: E3 ligases containing two reader domains

The tri-partite motif containing (TRIM) family of proteins have been identified as E3-ubiquitin ligases, due to the majority containing a really interesting new gene (RING) domain at the N-terminus.[65] E3 ligases are involved in many cellular processes including signalling, apoptosis, autophagy, and carcinogenesis. The E3-ligase is involved in substrate recognition, as the RING domain of TRIM mediates the transfer of ubiquitin, a 76 amino acid group, from E2 ubiquitin-conjugating enzymes to the ζ-group of a lysine residue of the substrate.[66, 67] These proteins also contain zinc-finger B-box domains and associated coiled-coil region. More than 80 TRIM proteins have been identified within humans and can be classified from C-I to C-XI, depending on the domains at the C-terminus on the proteins.[68, 69] The work in this thesis focuses on TRIM proteins in the C-VI family. This family, also known as the TIF family, consists of TRIM24, TRIM28 and TRIM33; which all contain tandem PHD and BRD reader domains at their C-terminus, shown in Figure 1.10A.

TRIM24, also known as TIF1α, is a down-regulator of p53, through ubiquitination, and hence a regulator of autophagy.[70] This protein is itself down-regulated in castration-resistant prostate cancer (CRPC), making it an appealing drug target.[71] The tandem PHD/BRD cassette, shown in Figure 1.10B, recognises multiple marks on the H3 histone peptide. The PHD recognises H3K4Me$_0$ and H3K9Me$_3$ *via* a

cation-π interaction with W828, as shown in Figure 1.10B, while the BRD is able to recognise H3K23Ac.[19] Crystallographic structures exist of the *apo* protein, the protein binding the H3 peptide along with several structures accompanying the structure based design of IACS-9571.[19, 53] These structures and associated PDB accession have been summarised in Table A.2. Additional weak TRIM24 binders were identified in 2017 by Liu *et al.* using an NMR-based screening assay. Despite the low affinity of these compounds, they use a range of polar warheads to promote deeper binding in the BRD site and a possible means to impose TRIM24 selectivity in future compounds.[29] As previously discussed, a dual BRD inhibitor has been published of both TRIM24 and BRPF1, along with a PROTAC analogue.[53, 54] A highly analogous ligand was also developed by Bennett /et al..[**Bennett2016a**] However, no selective inhibitors of the TRIM24 BRD exist, nor any inhibitor the TRIM24 PHD.

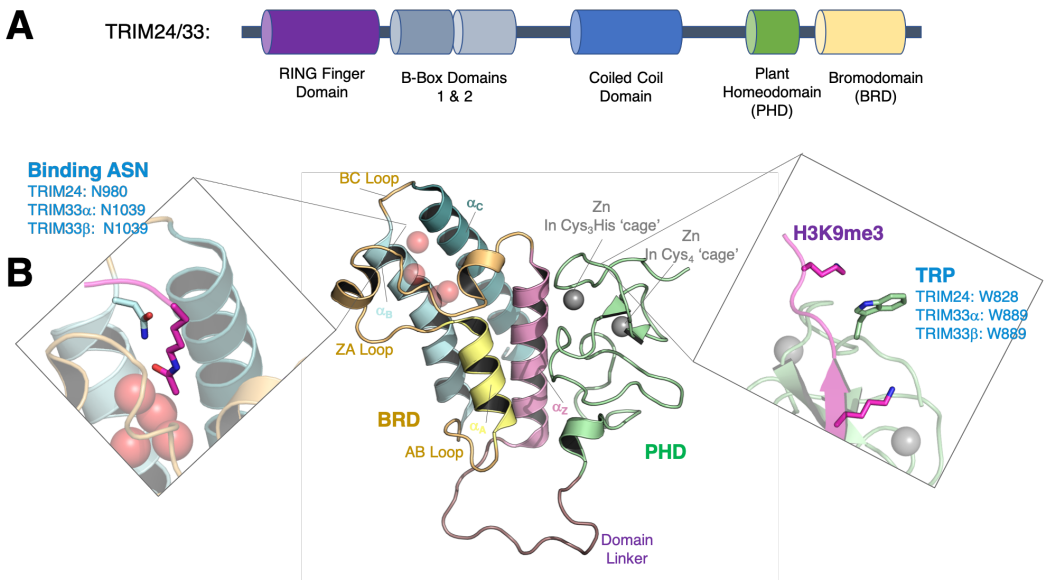


Figure 1.10: TRIM24/33 structural features. **A**: Domains found within the homologous TRIM24/33 proteins. **B**: TRIM24 (PDB: 3O33) with the interactions with H3 peptide highlighted in the boxes (PDB: 3O35 & 3O37).[19]

TRIM33, which is also known as TIF1γ, is a homologue of TRIM24. This TRIM protein has been linked to carcinogenesis through SMAD4 down-regulation, as well as to the DNA damage response pathway.[72–74] This protein is less selective than TRIM24, able to bind both H3K18Ac and H3K23Ac *via* the BRD and forming the

same interactions with the PHD.[75] The TRIM33 BRD is found in two isoforms: TRIM33α which contains a 17 amino acid extension of the α_B and α_C helices, which is not found in the TRIM33β isoform. Four structures of the TRIM33α PHD/BRD cassette exist in the *apo* and peptide bound states, listed in Table A.2.[76] However, the residues corresponding to the amino acid insertion are partially unresolved. In addition, the extension of TRIM33α means the Asn residue, that would normally bind to KAc, is no longer pointing into the binding pocket in the crystallographic structures with the histone peptide bound.[76] It has been hypothesised that the TRIM33α isoform, containing the non-canonical BRD, corresponds to the 'off' state or inactive form of the BRD. Meanwhile, there is one available structure of the TRIM33β PHD/BRD cassette, which contains the *apo* BRD and the H3 peptide bound to the PHD (PDB ID: 5MR8). There are no published inhibitors of either the TRIM33 BRDs or the PHD. In addition, it is not known the importance of each reader domain on TRIM33 function within the cell.

1.3 Bromodomains as a target for tropical disease

Many tropical diseases are caused by invasion of a host by a parasite and can be fatal, especially with a lack of effective treatments. Some of these diseases are caused by trypanosomes, a genus of protozoan parasites that are responsible for thousands of deaths annually. *Trypanosoma brucei* is the cause of African sleeping sickness transferred by the tsetse fly,[77] with 1446 cases recorded in 2017 in 36 sub-saharan countries.[78] The disease can take two forms depending on the parasite species: *T. brucei rhodesiense* and *T. brucei gambiense*, both of which are fatal if not treated. Currently, treatments are limited to a combination of pentamidine and suramin, both of which have toxic side effects.[77]

Trypanosoma cruzi is the cause of Chagas disease, which was first identified in 1909. Since then, very little has been achieved in identifying an effective treatment and the disease was responsible for 7,900 deaths in 2017.[79] Unfortunately, the disease is asymptomatic until the later stages, which often leads to devastating

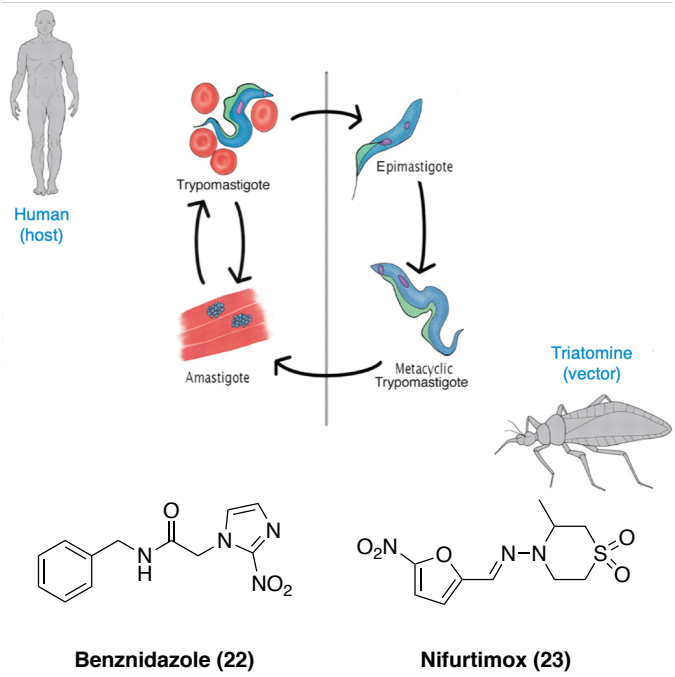


Figure 1.11: Lifecycle and current treatments of *T. cruzi*. The current treatments are benznidazole (**22**) and nifurtimox (**23**). Adapted with permission from Atwood *et al.*, Science.[81] Copyright 2005 The American Association for the Advancement of Science.

and irreversible damage to the heart and digestive systems. The lifecycle of *T. cruzi* can be seen in Figure 1.11, where the transfer between hosts begins by blood ingestion from the origin host by an insect vector. Within the insect vector, the parasite undergoes a transformation into the epimastigote stage and can replicate in the midgut *via* binary fission. After moving to the hindgut, the epimastigotes differentiate into metacyclic trypomastigotes that are later released onto the skin of hosts. Through mucous surfaces of wounds, the typomastigotes enter the host and invade nucleated cells *via* a lysozome-based mechanism. Yet another transformation occurs within the cells into the amastigote, where subsequent in-cell replication occurs. The amastigotes exit the cell after a final transformation into trypomastigotes which spread the disease to vital organs and picked up by additional vectors *via* blood bites.[80, 81] The current treatments for Chagas disease are benznidazole (**22**) and nifurtimox (**23**), both which are ineffective against the later, chronic disease stage and are subject to parasite resistance.[82]

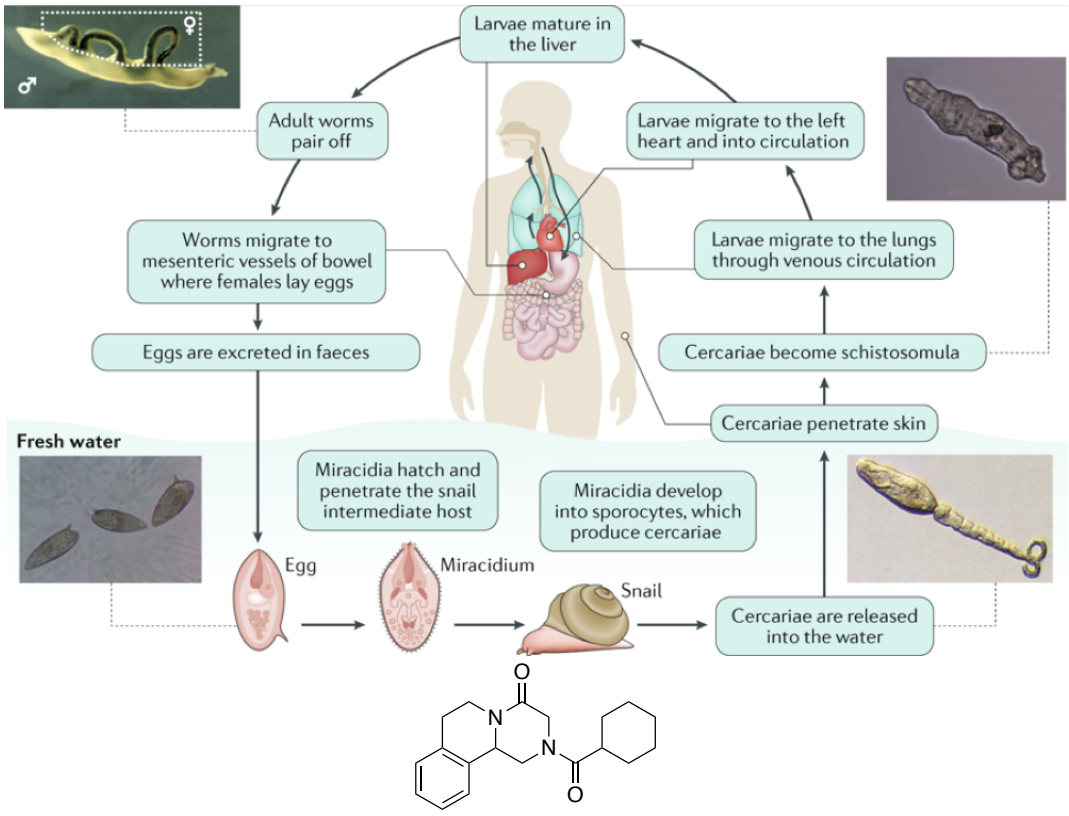


Figure 1.12: Lifecycle and praziquatel (**24**), the only treatment available for *S. mansoni*. Adapted with permission from McManus *et al.*, Nature Reviews Disease Primers.[84] Copyright 2018 Nature Publishing Group.

A similar lifecycle is seen with *Schistosoma mansoni*, one of the parasites responsible for schistosomiasis. This time the vector, or intermediary host, is the fresh-water snail.[83] The *S. mansoni* lifecycle, shown in Figure 1.12, is more complex than in *T. cruzi*, where reproduction and production of eggs requires pairing of opposite-sex adult worms in the liver. This pairing leads to production of eggs, which leave the host *via* excretion.[84] Praziquantel (Biltricide - **24**) is an anti-parasitic agent and the only known treatment of schistosomiasis, however wide spread usage of this treatment has led to cases of resistance.[85]

1.3.1 Epigenetic regulation of parasitic lifecycles

As eukaryotic organisms, simpler yet similar epigenetic machinery has been identified within parasite species. Within *T. brucei*, histone-modifying enzymes including

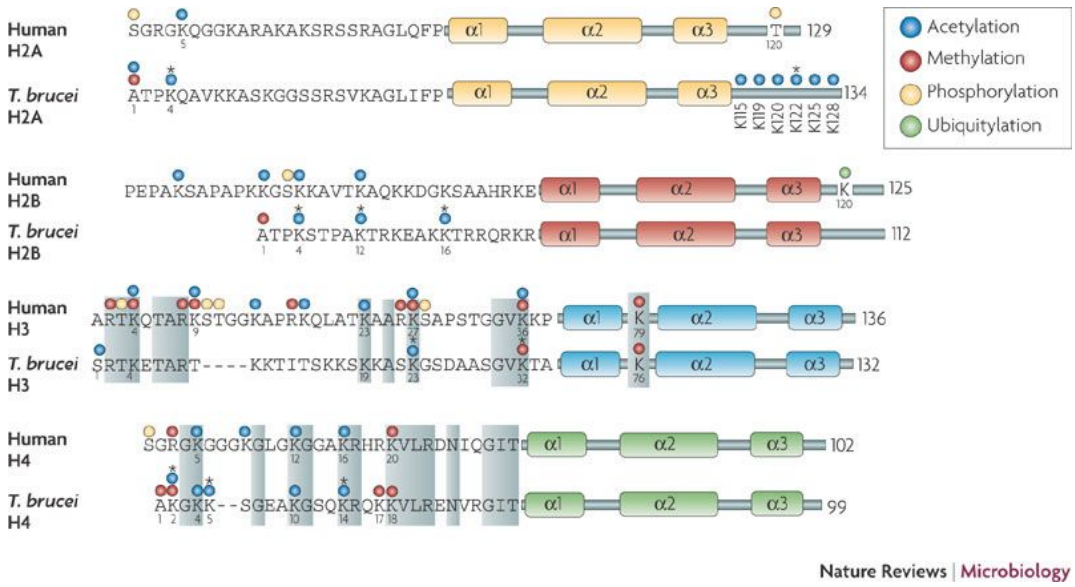


Figure 1.13: Histone tails and known modifications within *T. brucei* compared to humans. Reprinted with permission from Figueiredo *et al.*, Nature Reviews Microbiology.[86] Copyright 2009 Nature Publishing Group.

HATs and HDCs as well as histone-binding protein such as BRDs, PHDs and Tudor domains have been identified. *T. brucei* also contains homologous histone cores and tails, shown in Figure 1.13, that are subject to multiple different PTMs.[86] Since the 1990s, it has been known that the histone tails in *T. brucei* are subjected to PTMS at different stages of their lifecycle, showing that epigenetics could play a role in trapanazome transformations.[87, 88] This observation could offer a means to control tropical disease progression, through understanding and inhibition of the epigenetic processes that could be involved in parasite transformations.

Recently, x-ray co-crystal structures of bromodomains within *T. brucei* have shown binding to the human bromodomain ligands I-BET151 (**11**) and BI-2536 (**20**).[89] This suggests that human bromodomain ligands can be used as a basis for designing parasite bromodomain ligands. Through sequencing of the parasite genomes, bromodomains have also been identified within *T. cruzi* and *S. mansoni*.[90, 91], and a more extensive introduction to these domains is given at the start of Chapter 5. It is also clear that to elucidate the roles that each of these bromodomains hold, and their relevance in tropical disease treatment, selective probes need to

be developed and ultimately remove any human binding.

1.4 Computer aided drug-design

Chemical space is vast, and identifying ligands that have both a high affinity and selectivity for a given target is a challenge.[92] High-throughput screens (HTS), such as phenotypic screening, are valuable tools that can screen large compound libraries and identify compounds that bind to the target.[93] However, these assays are limited to only available compounds and a useable protein construct. Luckily, over the past few decades several predictive tools have been developed to aid this search.[94] Virtual screening (VS) has emerged as one alternative technique, which can test vast virtual libraries of compounds against a target structure quickly. These libraries can be available in-house, purchasable, or compounds that are tractable.[95, 96] This allows for a large region of chemical space to be tested and can be experimentally validated afterwards.[97] Docking is the most popular method for VS, which identifies potential binding modes of the ligands and uses a scoring function, which can quickly rank large numbers of compounds.

Once an initial hit has been identified, there are several approaches that can be used for ligand optimisation. Structure-based drug design (SBDD) aims to engineer ligand optimisations using existing 3D information about the target. This information is obtained from X-ray crystallography, CryoEM or NMR studies.[98] However, if no information is available it can be obtained from a homology model, which predicts a target structure based on known structures containing similar sequences.[99] It is worth noting that the *apo* and co-crystal binding sites seen in a protein will differ, due to the induced binding caused by the ligand. Hence, it is often desirable to dock to a co-crystal structure of the protein.[100] With the target structure obtained, a range of *in silico* techniques can be used to guide ligand optimisation. These include further docking studies to predict binding modes and roughly rank compound analogues, MD to understand the dynamics of binding and free energy techniques to predict potential energetic improvements of compounds.

Despite the work in this thesis using a structure-based drug design (SBDD) approach, ligand-based drug design can also be employed to guide ligand optimisation. This approach is popular in medicinal chemistry, which looks to understand the structure-activity relationships of different structures and their physiochemical properties. This can be enhanced using *in silco* techniques, which can screen large libraries of compounds to identify multiple ligands that bind the target. This allows for the building of quantitive structure-activity relationship (QSAR) models based on the data.[101] This field has recently made large advances in the era of machine learning. With large experimental datasets available, the processing and extrapolation of the data to predict new, biologically active compounds has now been made possible. These also offer a means of predicting other significant factors that can affect latter-stage drug discovery projects and drive high attrition rates.[102]

Molecular dynamics simulations, which are based on Newton's laws of motion, allow for both protein and ligand flexibility.[103] This is important as binding is a dynamic event, and so offers insights into the protein-ligand interactions that cannot be obtained from a co-crystal or docked structure. Crystallographic structures provide a static, low energy conformation of the target and docking is often performed with a rigid binding site, which does not account for the induced fit that can occur upon ligand binding. Through simulations, the protein and ligand can accommodate each-other and sample possible binding modes. This induced fit is particularly important when an *apo* structure or homology model of the target is used for docking studies. Where only *apo* structures are available, MD simulations can be implemented to extract multiple binding site conformations that can be used for docking. This approach, known as the relaxed complex scheme (RCS) has been used in the identification of ligands for or RNA-editing ligase 1 in *T. brucei* by Amaro *et al.*[104] The sampling of multiple binding site conformations also allows for the identification of potential cryptic binding sites, which are often not observed in the crystallographic state and require a ligand to allow for site formation. Simulations of the target, for example using enhanced sampling techniques or co-solvents, have allowed for cryptic binding sites to be identified in target structures.[105]

1.4.1 Binding free energy calculations

The free energy of ligand binding, $\Delta G_{binding}$, is a metric that is vital in drug discovery. This quantity can be determined experimentally, through techniques including ITC and WaterLOGSY. However, this requires synthesis of the ligands, soluble recombinant protein, and subsequent assay optimisation for the protein system. Computationally 'cheap' *in silico* techniques have been developed to give an estimate of the free energy in the form of scoring functions, and these have been incorporated into docking software and employed in high throughput screening.[106] However, scoring functions suffer from large errors, in the range of 2-3 kcal mol^{-1}.[107] An improvement to the scoring functions has been found in end-point methods such as MM-PBSA (molecular mechanics Poisson-Boltzmann surface area) calculations. These techniques use molecular dynamics simulations to calculate an improved estimate of $\Delta G_{binding}$.[108] However, these calculations are still estimates and lack the ranking accuracy and mathematical rigour of alchemical techniques.[109]

Two alchemical methods have become popular within the field of ligand development: firstly, relative binding free energy calculations (RBFE) that predict the $\Delta\Delta G_{binding}$ of two similar ligands to a given target.[110, 111] Secondly, absolute binding free energy (ABFE) calculations that give an absolute value of $\Delta G_{binding}$. Both methods use a non-physical thermodynamic cycle and molecular dynamics simulations to perform the rigorous calculation of $\Delta G_{binding}$ and $\Delta\Delta G_{binding}$.[112]

ABFE calculations have proven to be beneficial and worth the high computational cost associated with the calculations in two main areas. Firstly, the calculation of $\Delta G_{binding}$ with a reduced error over the aforementioned techniques.[109] Secondly, in giving an accurate ranking of ligand binding poses. Given the size of errors involved in scoring functions, determining the correct pose in docking studies is often challenging, with multiple possibilities. Through ABFE calculations, it has been shown that docked poses can be accurately ranked and even reproduce crystallographic poses.[58, 113] This is a powerful tool, offering confidence in the predicted binding mode of a ligand in place of a crystal structure, which can then be used for a structure based drug design approach.

Recently, free energy calculations have also been extended to look at water molecules. Using Monte Carlo simulations, which differ from classical molecular dynamics simulations, the free energy of binding of water molecules can be calculated.[28, 114] This offers an improvement on the WaterDock software, which used a scoring function based approach to predict the locations of water molecules.[115]

1.5 *In vitro* ligand binding assays

Bromodomains are excellent proteins to test against *in vitro*, given that many bromodomains have been expressed as a stable construct outside of their endogenous protein.[116] The experimental assays used by colleagues to obtain the experimental data for this thesis have been outlined below, with a brief discussion on their limitations.

1.5.1 AlphaScreen™

The amplified luminescent proximity homogenous assay screen (AlphaScreen™) has been relied upon heavily to evaluate the IC_{50} values of ligands within this thesis. This is a peptide-ligand competition assay, where a protein-binding peptide is linked to a streptavidin-coated donor bead containing a phthalocyanine photosensitiser. This sensitiser can excite the ambient oxygen into its singlet state ($^1O_2$) when irradiated at 680 nm. Meanwhile the protein construct is linked to a Ni^{2+}-chelating acceptor bead *via* a polyhistidine tag on the protein termini, shown in Figure 1.14. This bead contains a series of compounds that undergo a photochemical pathway to release fluorescence in the range 520-620 nm. If the peptide is associated to the protein, the donor bead and acceptor bead are in close proximity, and fluorescence is observed. As the ligand competes with the peptide, there is an associated reduction in fluorescence. This change in fluorescence can be incorporated into a titration experiment and an IC_{50} value extracted.[116, 117]

The main limitation of this assay is a high false-positive rate, which can be assessed using a further assay known as the TruHits test.[117] The assay also shows

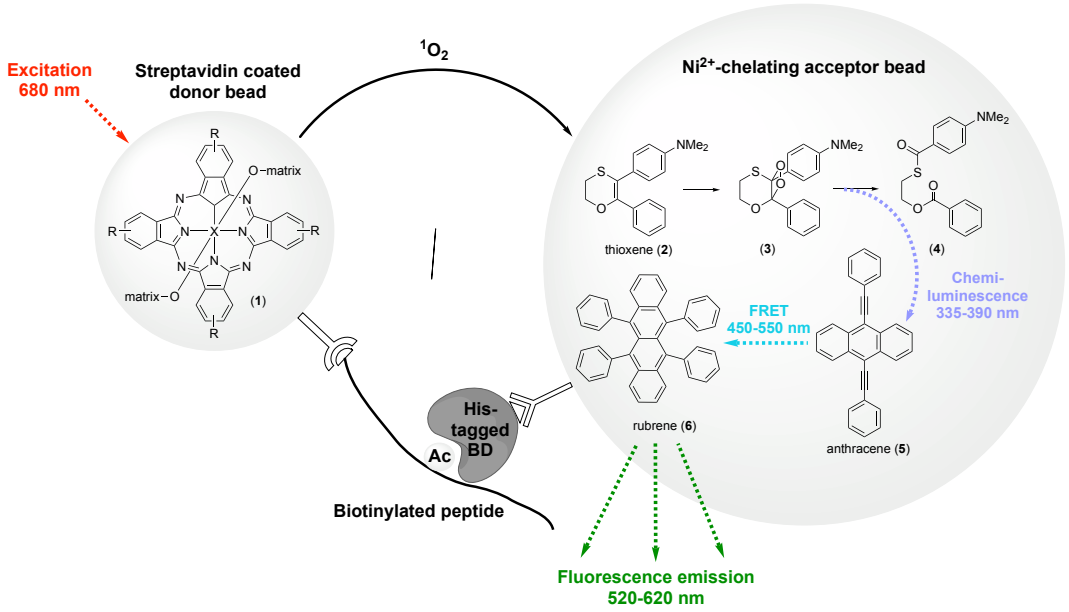


Figure 1.14: AlphaScreen™ schematic. Image produced by Dr M. Schiedel.

poor precision, often giving fluctuations in IC_{50} values between repeats, though the assay still performs well at ranking hits against a target.[116, 117]

The IC_{50} values determined can be difficult to compare between protein targets, given the peptides have different affinities between proteins. This target dependence can be negated through conversion to K_i values. This normalises the IC_{50} against a measurement of peptide affinity, as per equation 1.1, where the IC_{50} value from a displacement assay is converted using the K_d of the displaced ligand at a given concentration L.[118]

$$K_i = \frac{IC_{50}}{(L/K_d + 1)} \tag{1.1}$$

1.5.2 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a low-throughput assay that uses calorimetry to elucidate the thermodynamic properties of binding. It allows the calculation of the binding constant (K_d), as well as the binding enthalpy (ΔH) and stoichiometry (n) of the reaction. Through basic thermodynamic equations it is then possible to calculate the binding entropy (ΔS) and ultimately the binding free energy (ΔG).[119]

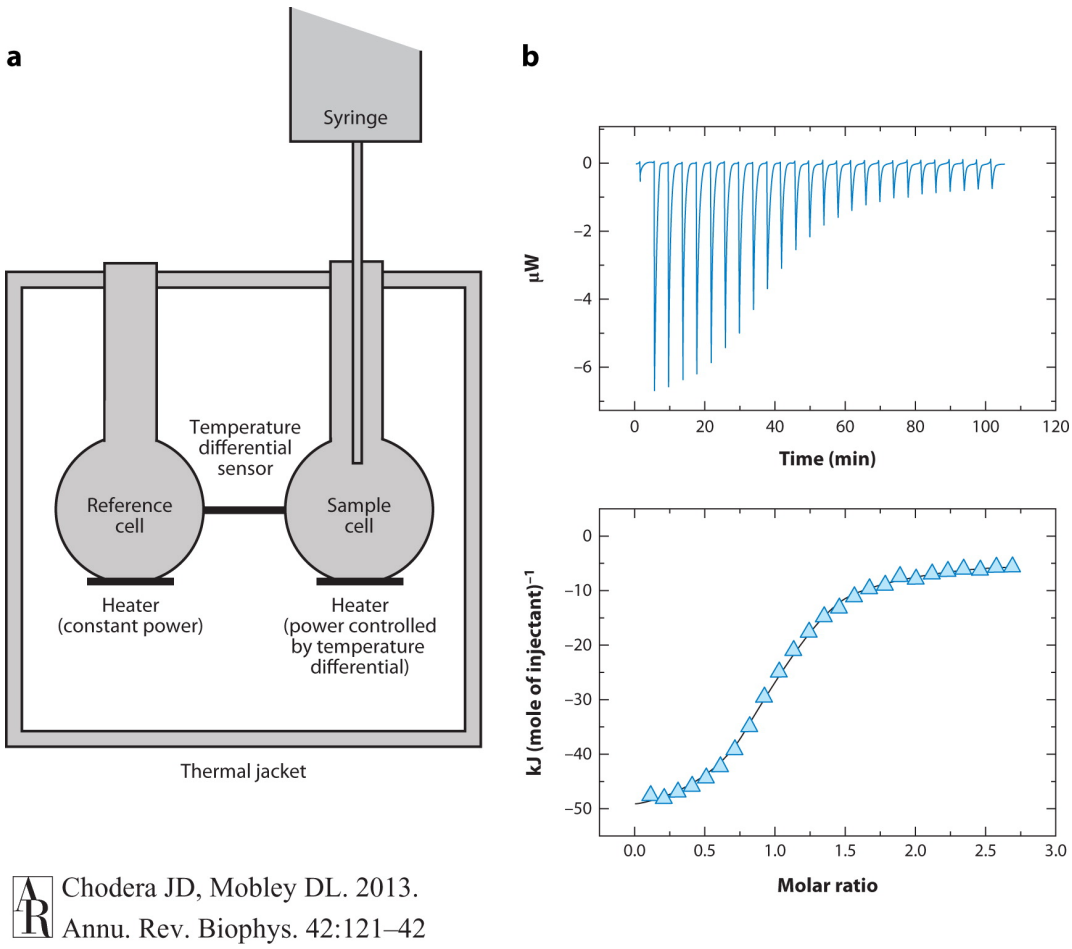


Figure 1.15: Schematic of an ITC experiment. **(A)** shows the set up of a power-compensating ITC experiment. **(B)** gives the data collected from an ITC experiments. The top shows the raw output, with power given as a function of time which is then integrated to give the plot below. Reprinted with permission from Chodera *et al.*, Annual Review of Biophysics.[120] Copyright 2013 Annual Reviews.

Modern calorimetry experiments rely on power compensation to measure the heat changes. As shown in Figure 1.15, the system contains two cells, which are maintained at a constant temperature through a thermal jacket. Both cells have electric heaters attached, and the temperatures between the cells are measured using a temperature differential sensor. One cell, termed the 'reference cell', undergoes constant heating from the heat source. Whereas the sample cell is subject to a varying amount of heat depending on the temperature differential sensor. The aim of this sensor is to return the sample cell's temperature to the predetermined value. As one half of a protein-ligand system is incrementally injected to the other half in

the sample cell, there is an associated temperature change. The cell temperature is then brought back to that of the reference cell, *via* the heat source. The power needed to do this is the raw experimental measurement, as shown in Figure 1.15. Integrating the signal over time can extract the heat associated with each injection and allow calculation of the thermodynamic properties.[120, 121]

This assay has a large detection window, helped by its large reading sensitivity. The lack of assay co-factors adds increased versatility to the range of assay conditions that can be used and increases the confidence in results. Unfortunately, there are large sources of error caused by buffer mismatch, inaccurate measurement of initial concentrations and poor ligand solubility to get the high initial concentrations required for experiments.[120, 121]

1.5.3 WaterLOGSY

Ligand and protein binding is an equilibrium process, with ligands inter-changing in the binding site with those in bulk solvent. The greater the ligand affinity for the protein, the slower this exchange will be. This feature can be exploited through WaterLOGSY, an NMR based assay that can detect ligand binding through nuclear Overhauser effect (NOE) experiments.[122]

In the typical WaterLOGSY setup, shown in Figure 1.16, the water is magnetised. The water then passes on its NOE signal to other neighbouring bodies through magnetisation transfer, where one transfer pathway is from bulk water to the protein. As the NOE signal passes through the protein, it is able to pass it to any bound ligands. As these ligands leave the binding site at equilibrium, they take the magnetisation with them with the same NOE signal as the protein because it tumbled while bound to the protein. This sign will be the opposite of a ligand that remains in solution which will be tumbling quickly and shown by the inversion of peaks in Figure 1.16.[122, 123] Through ligand titration, and measurement of the intensity of the NOE signal it is possible to extract a K_d value, as described by Dalvit *et al.*[122]

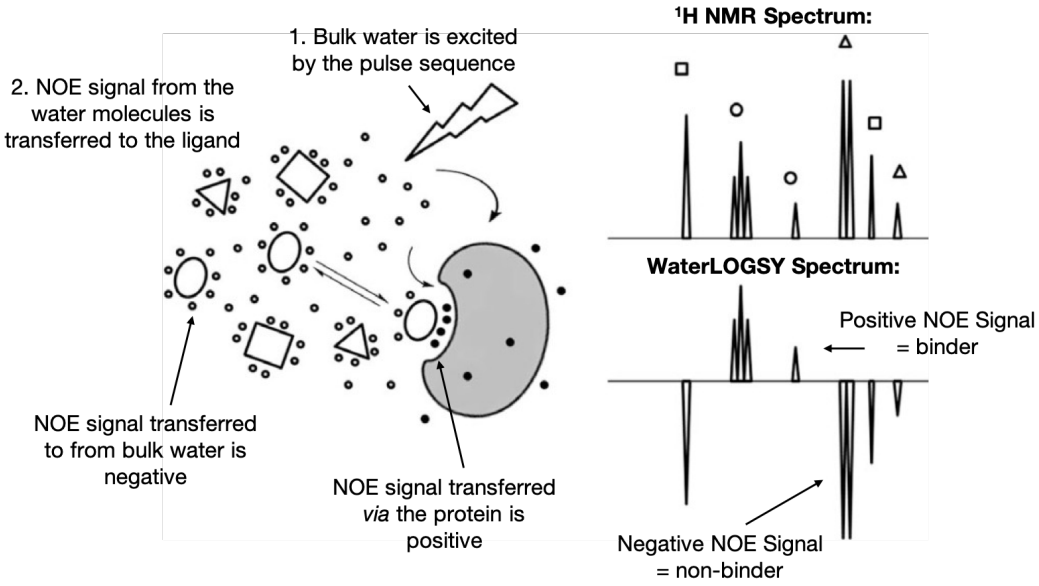


Figure 1.16: Schematic of an WaterLOGSYexperiment. Adapted with permission from del Carmen Fernandez-Alonso *et al.*, Current Protein and Peptide Science.[123] Copyright 2013 Eureka Science.

1.6 Aims and motivation

This thesis looks to use a range of *in silico* tools to aid rational ligand design within multiple active medicinal chemistry projects. In particular, how docking, molecular dynamics and ABFE calculations can be used in a prospective manner to improve ligand affinity. The thesis will start on the TRIM proteins, where there are no published inhibitors of TRIM33β. Through a HTS performed by Dr A. Sekirnik, several ligands that bind both the TRIM24 and TRIM33 PHD/BRD cassette were identified.[124] Chapter 3 of this thesis looks at identifying the binding sites of these ligands and the key interactions that they form. Chapter 4 uses these binding site predictions to aid in a SBDD approach to optimising one of the ligand hits. These techniques will then be applied to a slightly different target: parasite bromodomains, where a lack of structural information has prevented SBDD so far. Finally, simulations will be used to understand the SAR observed in binders of both CREBBP and OXFBD02, where subtle differences between ligands have large impacts on the binding affinity.

...if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms.

— Richard Feynman - *The Feynman Lecture Series*

2

Theory and Methods

Contents

The work within this thesis focuses on the use of *in silico* methods, namely

docking, molecular dynamics simulations and free energy calculations, to understand ligand binding events in active medicinal chemistry projects. This chapter looks to introduce the methods used within this thesis and cover their implementation in Chapters 3-6. Although a brief summary, excellent reviews by Leach and Allen & Tildesley go on to further explain these techniques and are the basis of this chapter.[125, 126]

2.1 Molecular mechanics

Many biological processes are the result of macromolecular interactions between proteins, ions and ligands. But on closer inspection, these processes are the result of interactions between atomic nuclei and electrons. Modelling these processes from first principles requires a rigorous *ab initio* approach at a quantum mechanical (QM) level and, when investigating long timescale motions, come with an associated computational cost beyond the scope of most projects. The cost can be reduced through application of the Born-Oppenheimer approximation, which assumes that electrons re-adjust their position near instantaneously relative to nuclei. The Born-Oppenheimer approximation allows electrons to be ignored and thus the energy of the system is then a property of nuclei positions. This simplifies the calculations and reduces the costs, however the cost only becomes feasible when the systems are modelled using a molecular mechanics (MM) approach.

A MM approach allows particle motion to be described by simple functions, often based on Newtonian physics.[103] These models, or "force fields", describe molecular motion in terms of bond stretching, bending and dihedrals. Moreover, these force fields have a high degree of transferability, which allows force fields for large molecules to be parameterised from small molecules.[125, 126]

2.1.1 Force fields

Force fields have been developed and refined since the early 20^{th} Century,[127] in efforts to allow calculations to reproduce experimental observations accurately and efficiently. A typical biomolecular force field used today describes the potential

energy, U, of a system as the sum of its bonded and nonbonded interactions. This is described in equation 2.1, where $\mathbf{r}^N$ are the atomic positions. The bonded and nonbonded terms are described using classical mechanics equations, where the equation constants, or *parameters*, are derived from x-ray spectroscopic measurements or *ab initio* calculations such as QM or density functional theory (DFT) calculations.[125, 126]

$$U(\mathbf{r}^N) = \sum U_{bonded} + \sum U_{nonbonded} \tag{2.1}$$

Bonded interactions

$$\sum U_{bonded} = \sum U_{bonds} + \sum U_{angles} + \sum U_{dihedrals} \tag{2.2}$$

The force field terms for the bonded interactions look to understand the energetic contributions caused by bond stretching, along with angle and dihedral deviations. Equation 2.2 demonstrates the components of the potential energy, where an energetic penalty is imposed should the movement deviate from its 'reference' position:

$$U_{bonds} = \sum_{bonds} \frac{k_r}{2}(r_{ij} - r_0)^2 \tag{2.3}$$

$$U_{angles} = \sum_{angles} \frac{k_\theta}{2}(\theta_{ijk} - \theta_0)^2 \tag{2.4}$$

$$U_{dihedrals} = \sum_{dihedrals} \sum_{n} k_{\theta,n}\Big[1 + cos(n\theta_{ikjl} - \delta_n)\Big] \tag{2.5}$$

Bond stretching is often modelled using a simple harmonic potential, where r_0 is the reference bond length, r_{ij} is the distance between atoms i and j, and k_r determines the penalty imposed for deviation from this length over each pair of atoms ij. This model is cheaper than the Morse potential, which is a better fit for true bond stretching energetics, but is still accurate assuming that bond length is never far from equilibrium.[125] Angles are modelled in a similar fashion, where θ_0 represents the reference bond angle with a similar constant k_θ between the three atoms ijk.[126]

Torsions were a late introduction in force field development,[127] however are vital in incorporating barriers to bond rotation into the model which arise from staggered and eclipsed conformations. The torsional potentials are defined by four interconnected atoms, $ijkl$, with a periodic function. Equation 2.5 models the dihedrals using a cosine function with a multiplicity n that governs the number of maxima and minima. $k_{\theta,n}$ is a factor in the *barrier height* for rotation, and the *phase factor*, δ controls the location of the energy minima. This is the functional form used within the Amber force fields, where n is capped at 3.[128] It is also important to note that all these terms can be improved with the introduction of higher order terms, but at an increased calculation cost that often outweighs the benefit of improved accuracy. Lastly, *improper torsions* can be used to impose additional restrictions such as planarity or stereochemistry on a molecule and can take a periodic or harmonic form.[125]

Nonbonded interactions

$$U_{nonbonded} = U_{vdW} + U_{Coulombic} \tag{2.6}$$

The nonbonded interactions reflect the interactions within and inbetween molecules. These are modelled as the summation of electrostatic interactions, modelled using Coulomb's law, and repulsion/dispersion interactions (or van der Waal (vdW) forces) which are modelled using the Lennard Jones (LJ) potential. These exclude the terms covered by bonded and angle terms, though will contribute to the torsional energies of molecules.[126]

$$U_{vdW} = \sum_{nonbonded\ pairs} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right] \tag{2.7}$$

$$U_{Coulombic} = \sum_{nonbonded\ pairs} \left[\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right] \tag{2.8}$$

Coulomb's law gives the potential energy arising from the interactions between charged particles, shown in 2.8. Here q_i and q_j are the partial charges on particles i and j which are at a separation of r_{ij} and ϵ_0 is the permittivity of free space. The vdW interactions are modelled using the LJ potential, shown in 2.7. This has

a shorter distance dependence than the Coulombic interactions. The two terms describe the balance between repulsion at short distances, with a rapid r^{-12} cutoff, and the attractive forces at larger distances, seen with the r^{-6} term. A and B are parameters optimised against experimental or *ab initio* data.[126]

2.1.2 Water models

Water molecules are vital in biological systems, for both the mechanism of action and the medium that proteins are often soluble in. This environment is mimicked within simulations, either through including explicit water molecules in the simulation or mimicking their impact using an implicit water model. If simulations are run with an explicit water model, a large proportion of the atoms in the simulation will be water. Hence it is important that a computationally efficient water model is used.[126]

The work within this thesis uses the TIP3P model exclusively, which has been shown to reproduce important thermodynamic properties of water at 298K, especially within free energy calculations.[129] The efficiency of this model comes from its rigid structure, discussed more in § 2.2.2, with only three fixed electrostatic interaction sites on the oxygen and hydrogen atoms. More intricate models, such as the TIP4P and OPC model look to improve the accuracy, however the additional model complexity increases the simulation cost of using these models.[130, 131]

2.1.3 Energy minimisation

When a system is constructed for simulation it often requires the bringing together of multiple components, such as addition of solvent molecules, docking of ligands and explicit protons. This can lead to unfavourable geometries or geometric clashes of atoms and thus render the system unstable for simulation. Hence, before simulations it important to perform a geometry optimisation to remove these clashes. One of the simplest forms is to search for a stationary point, or local minimum, on the potential energy surface of the whole system. Mathematically this takes the form of:

$$\frac{\partial U}{\partial \mathbf{r}} = 0 \quad \text{and} \quad \frac{\partial^2 U}{\partial \mathbf{r}^2} > 0 \tag{2.9}$$

These derivatives cannot be solved analytically due to the many body nature of the equations and hence is solved numerically. One of the simplest methods for solving the equation is the *steepest descent* algorithm which guides the system to an energetic minimum iteratively by moving down the potential energy gradient. The process is often stopped after a set number of steps or when the maximum force experienced by an atom is below a certain threshold.[125]

2.2 Molecular dynamics

MD essentially employs Newton's second law (eqn 2.10) to propagate the movement of atoms over time, given an initial set of co-ordinates and atomic velocities. This is possible from taking the Lagrangian formulation of Classical Mechanics, which provides a formal structure that states the forces on all particles are vector quantities that can be derived from the potential energy function, $\mathbf{F}_i\left(\mathbf{r}_1,\ldots,\mathbf{r}_N\right) = -\nabla_i U\left(\mathbf{r}_1,\ldots,\mathbf{r}_N\right)$. This formulation, represented using dot notation in eqn 2.11, allows for the particle positions to be derived through integration of the potential energy function, shown in eqn 2.13. However this integral is extremely hard to solve analytically, and hence solved numerically in discrete time steps.[132]

$$\mathbf{F}_i = m\mathbf{a}_i \tag{2.10}$$

$$-\nabla_{\mathbf{r}_i} U\left(\mathbf{r}_i\right) = m\ddot{\mathbf{r}}_i(t) \tag{2.11}$$

$$\iint -\nabla_{\mathbf{r}_i} U\left(\mathbf{r}_i\right) dtdt = m\mathbf{r}_i(t) \tag{2.12}$$

$$\mathbf{r}_i(t) = \frac{1}{m}\iint -\nabla_{\mathbf{r}_i} U\left(\mathbf{r}_i\right) dtdt \tag{2.13}$$

As this formulation is continuous, an initial set of atomic positions and velocities of each atom are required for integrating eqn 2.13.[132] The atomic positions are generated from crystal structures, or if these are not known then a combination of homology models and docking solutions. Solvent and ions are added to neutralise the overall system charge are also generated and avoid artefacts arising from a

residual charge. Meanwhile, initial velocities are drawn from the Maxwell-Boltzmann distribution at the given system temperature through eqn 2.14:

$$p\left(\mathbf{v}_i\right) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \cdot \exp\left(-\frac{1}{2}\frac{m_i \mathbf{v}_i}{k_B T}\right) \tag{2.14}$$

where $p\left(\mathbf{v}_i\right)$ is the probability distribution of the velocities $(\mathbf{v}_i)$, m_i is the mass of the particle, T is the temperature and k_B is the Boltzmann constant.[125]

2.2.1 Integration algorithms

As discussed, equation 2.13 cannot be solved analytically, as it is a many body problem. Instead, several algorithms have been proposed to solve the second order differential equation using *finite difference* methods. This leads to the equation being solved over several discretised steps and thus propagation of atomic positions and velocities over several timesteps (Δt). A key assumption to this approach is that all forces on atoms are constant within a given Δt. Using the chosen force field, the forces on each atom are calculated at time t. This force is then used to calculate the positions $(\mathbf{r})$, velocities $(\mathbf{v})$ and accelerations $(\mathbf{a})$ of all particles at time $t + \Delta t$ using the *finite difference* algorithm of choice. Whatever the choice of algorithm, it is important that it is: i) of low computational cost; ii) conserves energy and momentum; iii) permits a long time-step, Δt, to allow for longer simulations.

All integration algorithms are built on the assumption that positions, velocity and acceleration can be approximated by a Taylor expansion:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 + \frac{1}{6}\mathbf{b}(t)\Delta t^3 + \frac{1}{24}\mathbf{c}(t)\Delta t^4 + \ldots \tag{2.15}$$

$$\mathbf{v}(t+\Delta t) = \mathbf{v}(t) + \mathbf{a}(t)\Delta t + \frac{1}{2}\mathbf{b}(t)\Delta t^2 + \frac{1}{6}\mathbf{c}(t)\Delta t^3 + \ldots \tag{2.16}$$

$$\mathbf{a}(t+\Delta t) = \mathbf{a}(t) + \mathbf{b}(t)\Delta t + \frac{1}{2}\mathbf{c}(t)\Delta t^2 + \ldots \tag{2.17}$$

$$\begin{aligned} \mathbf{b}(t+\Delta t) &= \mathbf{b}(t) + \mathbf{c}(t)\Delta t + \ldots \\ &\vdots \end{aligned} \tag{2.18}$$

The *Verlet* algorithm is a popular finite differences integration method,[133] though combining different order terms can led to imprecision.[126] An improved

method is the *Velocity Verlet* algorithm,[134] which uses the positions and accelerations at time t and the positions from time $t - \Delta t$, without any need for explicit velocities, to calculate new positions at time $t + \Delta t$:

$$\mathbf{v}\left(t + \frac{1}{2}\delta t\right) = \mathbf{v}(t) + \frac{1}{2}\delta t\mathbf{a}(t) \tag{2.19}$$

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t\mathbf{v}\left(t + \frac{1}{2}\delta t\right) \tag{2.20}$$

$$\mathbf{v}(t + \delta t) = \mathbf{v}\left(t + \frac{1}{2}\delta t\right) + \frac{1}{2}\delta t\mathbf{a}(t + \delta t) \tag{2.21}$$

and after ignoring higher order terms the equations can be added to give:

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \mathbf{a}(t)\Delta t^2 \tag{2.22}$$

where particle velocities can then be calculated.[126]

A cheap and common method for integration is the *leap-frog* algorithm.[135, 136] Here, the particle velocities are calculated half a time step ahead ($\mathbf{v}\left(t + \frac{\Delta t}{2}\right)$) using the acceleration to allow the particle positions at $t + \Delta t$ to be calculated. The velocities are then approximated using interpolation. Essentially, the velocity leaps before position by $t + \frac{\Delta t}{2}$, then positions leap over velocities to get to $t + \Delta t$:

$$\mathrm{v}\left(t + \frac{\Delta t}{2}\right) = \mathrm{v}\left(t - \frac{\Delta t}{2}\right) + \mathrm{a}(t)\Delta t \tag{2.23}$$

$$\mathrm{r}(t + \Delta t) = \mathrm{r}(t) + \mathrm{v}\left(t + \frac{\Delta t}{2}\right)\Delta t \tag{2.24}$$

$$\mathrm{v}(t) = \frac{1}{2}\left[\mathrm{v}\left(t + \frac{\Delta t}{2}\right) + \mathrm{v}\left(t - \frac{\Delta t}{2}\right)\right] \tag{2.25}$$

This method explicitly includes velocities and hence useful for kinetic energy calculations. Its should also be noted that other algorithms, such as *Beeman's* algorithm, exist however they have not been used within this work.[126, 134, 137] The molecular dynamics package used within this thesis, GROMACS, uses a modified version of the leap-frog algorithm that allows for Langevin dynamics to be performed.[138]

2.2.2 Consequences of solving by integration

As discussed in the previous sub-section, a key assumption to integration approaches is that all forces on atoms are constant within a given Δt. For this to hold, the size of Δt is limited to the highest frequency motion that is being simulated. A larger Δt would lead to increased equilibrium bond distances between the pair of atoms concerned, leading to increased forces acting upon the bond and ultimately affect the stability of the simulation.[126]

The bonds between a hydrogen and heavy atom are often of a high frequency and hence a bottleneck on the size of Δt. As these bonds are of little interest in macromolecular simulations, these can be constrained to their equilibrium bond value with several algorithms developed to correct for deviations from equilibrium bond length over a timestep. GROMACS uses the SETTLE algorithm,[139] though other methods such as P-LINCS exist and also widely used.[140] These algorithms allow a timestep of 2 fs to be used. Recently, this limiting timestep has been improved using hydrogen bond mass repartitioning (HMR) which now allows for an increase of timestep to around 4 fs to be used.[141, 142]

2.2.3 Statistical mechanics: micromolecular to macromolecular

Simulations of macromolecules are run at the microscopic level, however it is often macroscopic observables that are of interest, such as pressure and changes in free energy. Statistical mechanics describes the relationship between the distribution and motion of N particles in a system to its macroscopic properties.

An ensemble is a collection of points in phase space satisfying the conditions of a particular thermodynamic state, where for an N particle system, phase space consists of $6N$ dimensions. Through the stepwise solving of eqn 2.13, a series of points in phase space are generated that belong to the same ensemble. Another way to look at this is the stepwise MD solutions represent a collection of systems which have different microscopic states but have an identical macroscopic or thermodynamic state (e.g. temperature (T), pressure (P) and volume (V))

Several ensembles are useful within MD simulations, especially ones that better mirror physiological conditions such as the canonical ensemble (NVT) or the isobaric-isothermal ensemble (NPT).[132]

Statistical mechanics derives observables as ensemble averages, $\langle O \rangle_{\text{ensemble}}$. In reality it is hard to calculate an ensemble average from an MD simulation as the simulation would need to pass through all possible states. Luckily, the *ergodic* principle states that the time average equates to the ensemble average ($\langle O \rangle_{\text{ensemble}} = \langle O \rangle_{\text{time}}$):

$$\langle O \rangle = \frac{1}{M} \sum_{i=1}^{M} O\left(\mathbf{p}^N, \mathbf{r}^N\right) \tag{2.26}$$

where $\mathbf{p}^N$ are the particle momenta, M is the number of observed timesteps and $\mathbf{r}^N$ are the particle positions.[125]

2.2.4 Temperature and pressure coupling

As discussed in § 2.2.3, the NVT and NPT ensembles are useful for modelling physiologically relevant systems. Ensuring constant N or V is trivial, however control of P and T require more complex algorithms that are discussed in the following sub-sections.

Thermostats

As the temperature of a system is linked to the kinetic energy of the particles:

$$K = \frac{1}{2} \sum_{i=1}^{N} m_i \mathbf{v}_i^2 = \frac{3}{2} N k_B T \tag{2.27}$$

$$T = \frac{2}{3} \frac{K}{N k_B} = \frac{1}{3 N k_B} \sum_{i=1}^{N} m_i \mathbf{v}_i^2 \tag{2.28}$$

it is possible to alter the temperature of the system through changing the particle velocities in the system. It would be easy to assume that the velocities could be simply scaled by a factor of $\lambda = \sqrt{T_{ref}/T}$, where T_{ref} is the constant temperature of the ensemble and T is the instantaneous temperature calculated from the kinetic

energies scaled.[143] However this deterministic method, like many temperature control algorithms, alters the dynamics of the system and does not accurately produce the thermodynamic properties of the canonical ensemble.[126] Instead, more complex solutions have been found where the system is coupled to an external heat bath with the aim of maintaining an average temperature of T_{ref}, which is applied to each particle *via* altering the particle's velocity or modifying Newton's equations of motion, building on the method just seen.[144]

Often the best algorithm to use is system dependent, with both stochastic and deterministic solutions proposed. The *Berendsen* thermostat sets a time scale for updating velocities,[145] rather at each step:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}\left(T_{ref} - T(t)\right) \tag{2.29}$$

where τ is the coupling parameter that scales the coupling between the system and external bath, which is at T_{ref}. Hence the change in temperature and the scaling factor each step are:

$$\Delta T = \frac{\Delta t}{\tau}\left(T_{ref} - T(t)\right) \tag{2.30}$$

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_{ref}}{T(t)} - 1\right)} \tag{2.31}$$

This method still fails to reproduce the correct kinetic energy distribution and lead to violations in the principle of energy equipartition,[146] though has still been shown to be a useful coupling method.[144] More rigorous methods have been developed and include the *Nose-Hoover* and *Andersen* stochastic thermostat.[134, 147]

The Langevin stochastic approach looks at altering the equations of motion, starting with the Langevin equation of motion:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = -m_i \xi_i \frac{d\mathbf{r}_i}{dt} + \mathbf{F}_i\left(\mathbf{r}_i\right) + \hat{r}_i \tag{2.32}$$

where ξ_i is a frictional constant and $\hat{r}_i$ is a stochastic noise term proportional to ξ_i and T_{ref}. This then allows for the velocities to be rescaled at each timestep:

$$\Delta \mathbf{v}_i = \mathbf{a}_i \Delta t - \gamma_i \mathbf{v}_i \Delta t + \xi_i \sqrt{2 k_B T_{ref} \gamma_i \Delta t / m_i} \tag{2.33}$$

where γ_i is a frictional coefficient to dampen motion of the particles. This returns the correct kinetic energy distributions for decoupled states,[146] which is particularly important for free energy calculations where ligands can be totally removed from the systems in a stepwise manner.[138] All simulations performed within this thesis used the Langevin thermostat for temperature coupling.

Barostats

The instantaneous pressure of a system can be calculated from the following:

$$P = \frac{1}{V} \left[N k_B T - \frac{1}{3} \sum_{i=1}^{N} \sum_{j>i}^{N} \mathbf{r}_{ij} \cdot \mathbf{F}_{ij} \right] \tag{2.34}$$

where the interatomic forces are already calculated in the simulation, so the calculation comes at a low computational cost. Given the dependence on V, it is possible to scale P through scaling the volume of the simulation box, and hence scaling the atomic co-ordinates.[132]

The *Berendsen* barostat uses a similar setup to that seen in the thermostats, employing an external pressure 'bath' coupled to the system to scale the atomic co-ordinates and maintain an average pressure:

$$\frac{dP(t)}{dt} = \frac{1}{\tau} \left(P_{ref} - P(t) \right) \tag{2.35}$$

$$\lambda = 1 - \kappa \frac{\Delta t}{\tau} \left(P - P_{ref} \right) \tag{2.36}$$

$$\mathbf{r}_i' = \lambda^{1/3} \cdot \mathbf{r}_i \tag{2.37}$$

where κ is the isothermal compress-ability.[148] However, although this keeps the correct average pressure, it fails to give the exact NPT ensemble. Although this algorithm can be useful for equilibration, the *Parrinello-Rahman* barostat is more useful for production MD simulations.[149] This is applied throughout this thesis.

2.2.5 Periodic boundary conditions

The computational cost of any simulation scales with the number of particles, N, by a factor of $NlogN$ when using Ewald methods. To reduce the cost, only small systems are studied, often comprising of the macromolecule of interest that is surrounded by solvent and lipids, if required. This system is unable to hold together in space, and so the temptation is to place a 'box' around the box comprising of a potential to hold the simulation together. However, this setup leads to surface effects, where a water molecule at the box edge will experience very different forces compared to a solvent molecule within the solution bulk.

The solution is to impose periodic boundary conditions, which essentially allow all particles to be treated as if they were in bulk solution. The simulation box can be replicated through all space in a lattice, each image containing the system of interest. If a molecule strays out of its simulation box and into its mirrored neighbour, it is also re-entering the original simulation box on the other side. This setup allows for conservation of N, and can be visualised in Figure 2.1. Cubic boxes are the easiest to work with and visualise, but a dodecahedral box is more efficient and reduces the number of solvent particles within the system.

There are some considerations that are needed to ensure that the microscopic properties are consistent, especially when looking at intermolecular interactions. In particular, to remove artefacts associated with the Lennard Jones potential a box length of $L \approx 6\sigma$ is required which prevents a particle interacting with its own image in the lattice, where σ is the separation at which the inter-particle potential is zero.[126]

2.2.6 Treatment of long range interactions

Calculations of the nonbonded interactions are the most time consuming aspect of MD calculations, which scales in a N^2 fashion with the number of atoms, N. By implementing periodic boundary conditions, it initially implies an infinite number of pair-wise interactions to calculate. This scheme would be impossible to perform, and hence distance cut-offs are applied to limit the number of pair-wise

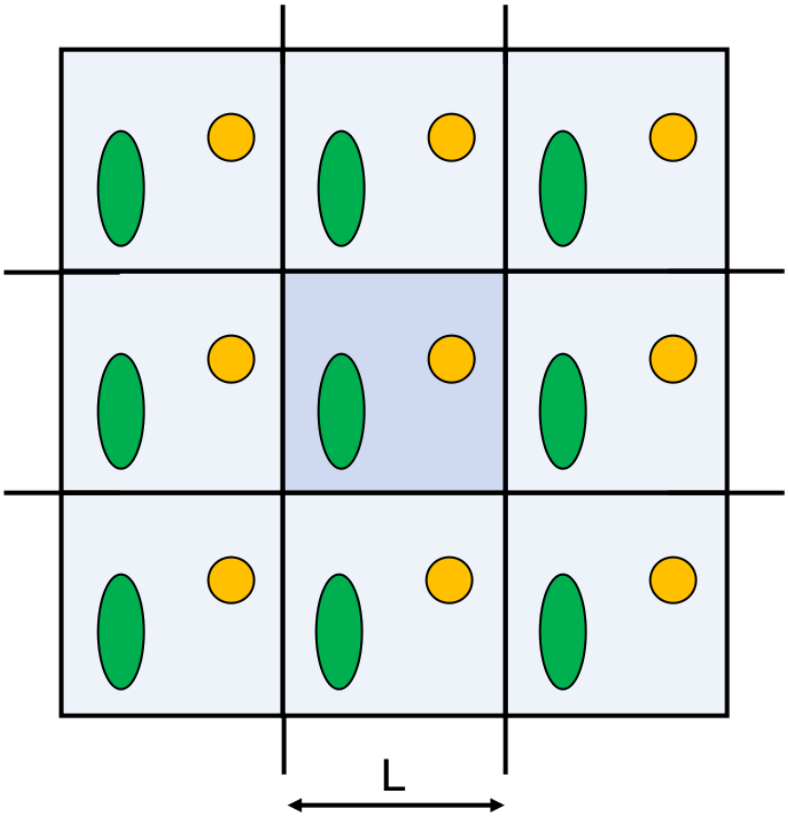


Figure 2.1: Cartoon example of 2D periodic boundary conditions

interactions calculated in a periodic system. However, this requires additional considerations for the impacts on the relevance of these cut-offs and the impact on thermodynamic properties.[126]

Coulombic cutoffs

Electrostatic interactions decay slowly with distance, proportional to r^{-1}, hence the errors associated with applying a cutoff are larger.[150] There are several schemes for evaluating the long range interactions in periodic systems, with Ewald summation methods the most common in MD simulations, due to their accuracy and convergence.[151] With this treatment, the interaction potential is split into into two terms:

$$\varphi(\mathbf{r}) \overset{\text{def}}{=} \varphi_{sr}(\mathbf{r}) + \varphi_{\ell r}(\mathbf{r}) \tag{2.38}$$

where φ_{sr} are the short range interactions which are easily calculated in real space, while $\varphi_{\ell r}$ are the long range interactions which converge rapidly in Fourier space.[152] Hence the cost of calculating $\varphi(\mathbf{r})$ is rapidly reduced through computation in two parts. The calculation costs have been further improved using the PME method, allowing the computational scaling to be reduced from N^2 to $NlnN$.[153] The PME method has been used for all simulation work within this thesis.

Lennard-Jones cutoffs

The interaction energy of two particles with the LJ potential decays rapidly, proportional to r^{-6}. The simplest solution is to set a cutoff distance for the distance of atoms accounted for in the MD calculations. All interactions beyond this distance are then assumed to have an interaction energy of zero. However, this cut-off can introduce artefacts, especially during free energy calculations. GROMACS has an inbuilt analytical corrections to remove artefacts, however these assume an isotropic distribution of solvent beyond the cutoff which is not the case when the tessellated images contain large macromolecules.[154] Shirts *et al.* propose the "EXP-LR" and "WHAM" corrections, to account for the differences in free energy imposed by a vdW cut-off. The *EXP-LR* method uses a two state exponential average calculation to find the $\Delta G^{\mathrm{LR}}(\lambda)$. The exponential average looks at the energies calculated between the cutoff used $u^{\mathrm{SR}}(\lambda)$ and the highest possible cutoff $u^{\mathrm{LR}}(\lambda)$:

$$\Delta G^{\mathrm{LR}}(\lambda) = -k_B T \ln \left\langle \exp\left[-u^{\mathrm{LR}}(\lambda) - u^{\mathrm{SR}}(\lambda)\right]\right\rangle_{\mathrm{SR}} \tag{2.39}$$

$$\Delta G^{\mathrm{Correction}} = \Delta G^{\mathrm{LR}}(\lambda = 1) - \Delta G^{\mathrm{LR}}(\lambda = 0) \tag{2.40}$$

where λ is the phase factor (discussed in § 2.4.2), T is temperature and k_B is the Boltzmann constant.[155]

Recently, the Lennard Jones Particle-Mesh (LJ-PME) method has been developed and available in GROMACS (however, it is not currently supported with GPU acceleration). This calculates the interactions beyond the cutoff with only a small error, effectively removing the need for a cutoff.[156] This method was used

to treat all long range interactions in all free energy calculations performed with GROMACS 2016. Where GROMACS 2019 is used for free energy calculations, the EXP-LR method was used to correct for differences in free energy, to allow for GPU acceleration of the calculations.

2.2.7 Simulation setup

Ligand parameters

Ligands, by default, were parameterised using the general Amber force field (GAFF) (v. 1.8), with AM1-BCC charges,[157, 158] using *antechamber* and *parmchk2* within the AmberTools16 package.[159, 160] The force field files were converted to co-ordinate and topology filetypes for GROMACS using *tleap*, from AmberTools16, and finally *acpype* (v. 2016-07-16).[161]

Where stated, ligands were alternatively parameterised to have RESP charges. The ligand structure was subject to geometry optimisation, at the HF/6-21G(d) level of theory using Gaussian16 (A. 03). A subsequent molecular electrostatic potential calculation was performed at the same level of theory. The partial charges were extracted using *antechamber* and combined with the GAFF parameters. The final ligand coordinates and topology files were converted for GROMACS using *acpype*.

For simulations of the H3 peptide, post-translationally modified amino acid parameters were acquired from FF-PTM. The PTMs parameters are derived from a solvent model as an extension of the ff03 force field.[162] A comparison of charges by Dr. O Andersen showed that there were no significant differences between these parameters and AMBER99SB-ildn, and so the parameters were merged within *tleap*. The peptide parameters were then constructed using *tleap*, found within the AmberTools16 package.[159]

System setup

All protein structures were parameterised using the AMBERff99sb-ildn force field,[163] with protonation states assigned within the GROMACS software.[164]

Missing residues were modelled using the WHATIF web interface.[165] If crystallographic waters were resolved, all were removed, with the exception of waters W0-5 within the BRD binding site. Simulations containing the PHD were adjusted to use the deprotonated Cys parameters within AMBERff99sb-ildn for the Zn co-ordinating residues, where the deprotonated form is expected to be co-ordinating zinc.[166, 167] Ligand protonation states were assigned as described in Section 2.6.1.

Water molecules were modelled using the TIP3P model,[168] within a solvated dodecahedral box that had a minimum boundary of 1.2 nm from any protein atoms. Sodium and chloride ions were added to neutralise any residual charge, and bring the salt concentration to 150 mM. The systems were first subjected to energy minimisation using the steepest decent algorithm of 10,000 steps, with a maximum force cut off of 100 $kJ\ mol^{-1}\ nm^{-1}$. The systems then underwent 200 ps equilibration in the isothermal-isobaric ensemble. The temperature was coupled using a Langevin thermostat, with a target temperature of 300 K, and the pressure was coupled using the Berendsen weak coupling algorithm to a target pressure of 1 atm.[136, 138, 148] Simulations were then carried out for 50 ns in triplicate. Post-production analysis was performed using PyMOL (v.1.8) and VMD (v.1.9.2) for visualisation and the MDAnalysis python library.[169, 170] Where simulations were clustered, the k-means clustering algorithm (`analysis.encore.cluster`) within MDAnlaysis was used, selecting 7 clusters based on the co-ordinates of the ZA loop.[170] The centroid of each cluster was then identified and extracted for further analysis.

This work spanned several iterations of the GROMACS MD package,[164] where adaptations to the setup were made to allow for increased GPU acceleration of the simulations.[171] Simulations in Chapters 3 and 4 were performed in GROMACS 2016.3, and used a LJ-PME order of 6 and the LJ-PME cutoff correction. With the release of GROMACS 2018, several alterations were made to the base code to offload parts of the calculation to the GPU for non-FEP calculations. This improvement came at the expense of using a long-range cutoff and reducing the PME order to 4. This change significantly improved the speed of the MD simulations, and used for the work performed in the Chapters 5 and 6.[171]

2.3 Monte Carlo simulations

Monte Carlo simulations are another type of MM based simulation method that generates a series of microstates. However, instead of a dynamic evolution of the system overtime the system states are generated through repeated random sampling. This section gives a quick overview of Metropolis Monte Carlo, which can be used to sample physical systems.[126, 172, 173]

Each random substitution to generate a new system state is subject to an acceptance criteria test, which ensures that the resulting distribution of micro-states in accordance with their Boltzmann factor. The selection criteria ensure that higher energy states are less likely to be selected, while more likely to select lower energy states. Each MC is a random movement of translating and rotating particles where the force field can be used to calculate the energy of the new system state. If the energy is lower, and ΔE is negative, it is accepted. If not, then it is accepted under the following criteria:

$$rand(0,1) \leq e^{-\Delta E/k_B T} \tag{2.41}$$

After a large number of random moves it is possible to extract observables from all the micro-states using a statistical mechanics approach:

$$\langle O \rangle = \frac{1}{M} \sum_{i=1}^{M} O\left(\mathbf{r}^N\right) \tag{2.42}$$

where O is the observable of interest, $\mathbf{r}^N$ are all the coordinate particles and M is the number of accepted state configurations.[125] The work within this thesis uses the MC implementation within ProtoMS package.[174]

2.3.1 Grand canonical Monte Carlo simulations

The Grand Canonical ensemble (μVT), where the chemical potential μ, volume and temperature are kept constant, is another statistical ensemble that can be used.[132] This ensemble allows the number of particles in the system to change, and hence lends itself to MC based techniques where each state is independent. The

added/removed particles are stored in an external reservoir in their ideal gas state. Hence the number of particles in the simulation are controlled by the chemical potential of the system, where the equilibrium is reached when the reservoir and system are at equipotential.[114] Insertion and deletion is also subject to selection criteria in MC simulations, which are dependent on the Adam's value:

$$B = \frac{\mu}{k_B T} + \ln\left(\frac{V}{\Lambda^3}\right) \tag{2.43}$$

where Λ is the de Broglie wavelength of the molecule and V is the volume in which insertion/deletion occurs. The Adam's value is therefore proportional to the chemical potential, μ, and therefore influences the acceptance criteria of the simulation and the number of particles in the simulation volume at equilibrium.[175]

This ensemble is vital within the free energy calculations on water, which are performed in this thesis.[28] Here the particle being inserted/deleted is water (modelled as TIP3P) and the volume is a region around the protein binding site.

2.4 Free energy calculations

The free energy of binding (ΔG_{bind}) is a vital thermodynamic quantity within drug discovery, quantifying the strength of a ligand binding its target. Under the isothermal and isobaric ensemble, this can be related directly to the dissociation constant (K_d) through:

$$\Delta G_{bind} = -k_B T \ln \frac{K_d}{c^0} \tag{2.44}$$

where c^0 is the is the standard concentration, 1 molar by convention and normally omitted, and K_d is the dissociation constant of the ligand to the protein.[176]

This quantity can be experimentally determined, as discussed in Section 1.5, however this requires the ligand to be synthesised before ΔG_{bind} is known. To mitigate the need for synthesis, a range of computation methods to predict ΔG_{bind} have been developed to either predict or calculate ΔG_{bind}, each with a trade off between computational cost and accuracy. Section 2.6 will introduce docking,

which uses scoring functions, which are a computationally inexpensive method for estimating the ligand binding affinity. However, these methods are often inaccurate due to multiple simplifications and failings of the model.[177] In this section, three computationally more expensive techniques to calculate ΔG_{bind} will be introduced. Firstly, end point methods, which improve the accuracy but are not mathematically rigorous, then moving to the rigorous absolute binding free energy calculations (ABFE) calculations and ending with Grand Canonical Integration (GCI) as a method for estimating binding energies. Each subsection will also cover the methods used to implement the techniques within this thesis.

2.4.1 End point methods: MM-PBSA

Endpoint methods such as *molecular mechanics Poisson-Boltzmann surface area* (MMPBSA) post-process MD simulations to extract multiple system configurations and estimate the binding energy. These are computationally cheap as they employ an implicit solvent model and are limited to a relatively small number of configurations. A similar method, *molecular mechanics Generalised Born surface area* (MMGBSA), is also used but not discussed in this thesis.[178] The underlying principle is that ΔG_{bind} can be estimated with the following approximation:

$$\Delta G_{bind} = \left\langle G^{PL} \right\rangle_{PL} - \left\langle G^{P} \right\rangle_{P} - \left\langle G^{L} \right\rangle_{L} \tag{2.45}$$

$$\approx \left\langle G^{PL} - G^{P} - G^{L} \right\rangle_{PL} \tag{2.46}$$

where it can be separated into the ensemble average of the Gibbs energy of the protein ligand (PL), the protein (P) and ligand (L). A common shortcut used is to only simulate the PL system and then separate the P and L conformations to perform the isolated component calculations. This is less rigorous and can fail to account for energy changes caused by large binding site changes upon binding, though often is as accurate as three-component simulations.[179]

Aldeghi *et al.* showed that the MM-PBSA ranks bromodomain ligand affinities with a Pearson rank correlation of $\overline{r}_p = 0.39$, which is less when compared to

ABFE (discussed in the next section) with an $\overline{r}_p = 0.64$.[109] The ability to rank is increased through utility of the entropy correction proposed by Duan *et al.* which increased the correlation to $\overline{r}_p = 0.55$,[180] or through including the nearest 20 water molecules as part of the protein component $\overline{r}_p = 0.53$.[109] This work has been built on by Wright *et al.*, who showed an increased ability to correctly rank BRD4 ligands, with MM-PBSA, when using a significantly large number of replicas, termed the *ESMACS* approach.[181]

Methods used for MM-PBSA calculations

All MMPBSA calculations were carried out using the bash scripts provided with GMXPBSA 2.1.2.[182] Equidistant snapshots every 80 ps were extracted from 50 ns simulations, where the first 10 ns was discarded as an equilibration period, yielding 500 frames for analysis. The vdW contribution to the energy was calculated using GROMACS 5.1,[164] while the Coulombic contribution was calculated using APBS (v. 1.5).[183] It was decided to use the values calculated using a water solvation shell of 30 water molecules, as recommended by Aldeghi *et al.*[109]

2.4.2 Absolute binding free energy calculations

Absolute Binding Free Energy (ABFE) calculations employ MD simulations and alchemical transformations to rigorously calculate the binding energy of a ligand to a protein, *via* a non-physical pathway. Through using atomistic MD, all solvent and entropic effects, which are neglected by scoring functions and MM-PBSA, are fully accounted for.

Another alchemical calculation known Relative Binding Free energies (RBFEs), which look at the $\Delta\Delta G$ associated with binding ligand A to ligand B, have proven more popular within drug discovery. These require less intermediate states on the alchemical pathway and benefit from multiple error cancellations. The FEP+ package, from Schrodinger, has successfully commercialised this method which is rapidly being adopted in industry.[111] However, there are scenarios where ABFE can be the better alternative. Firstly, it can provide the ΔG_{bind} value of a given

docked pose directly allowing accurate ranking of poses. Secondly, RBFE methods are limited by the number, type and charges of heavy atom changes between ligand A and B, while ABFE has no limit to the differences in atoms between the two ligands being compared.[58]

Alchemical cycles for ABFE

ΔG_{bind} is a state function, meaning the binding pathway of the ligand has no effect on the value of ΔG_{bind} as long as the start and end states are the same. There are no limitations of the pathway taken, allowing non-physical (or alchemical) pathways to be used. Figure 2.2 shows an alchemical cycle used within ABFE calculations. ΔG_{bind} is calculated from performing multiple alchemical transformations to intermediate states around this non-physical cycle. The energy change to each intermediate state is from independent MD simulations modelling a discrete point inbetween these states.

Walking through the cycle in Figure 2.2, initially the ligand is free in solution. The ligand then undergo several alchemical changes to have its Coulombic charges then vdW interactions decoupled. This ligand is essentially 'naked' to the protein. Restraints are applied to this ligand to ensure its orientation is held relative to the binding site, regardless of the lack of interactions. This non-interacting, restrained ligand in solution is essentially the same as the bare ligand in the binding site and so no work is needed to move it into the binding site. The protein bound ligand can now have its vdW and Couloumbic interactions slowly reintroduced before the restraints are finally released. This path has taken the ligand from solution to protein bound, and so the ΔG_{bind} is the sum of individual energy changes around this pathway.[176, 177]

As just discussed, it is important to restrain the ligand relative to the binding site to ensure it doesn't wander away from the binding site. There are multiple methods to hold the ligand, including spherical and flat-bottom restraints. However, a set of restraints developed by Boresch *et al.* are particularly useful in ABFE calculations. These restraints involve fixing one bond, two angles and three dihedrals.

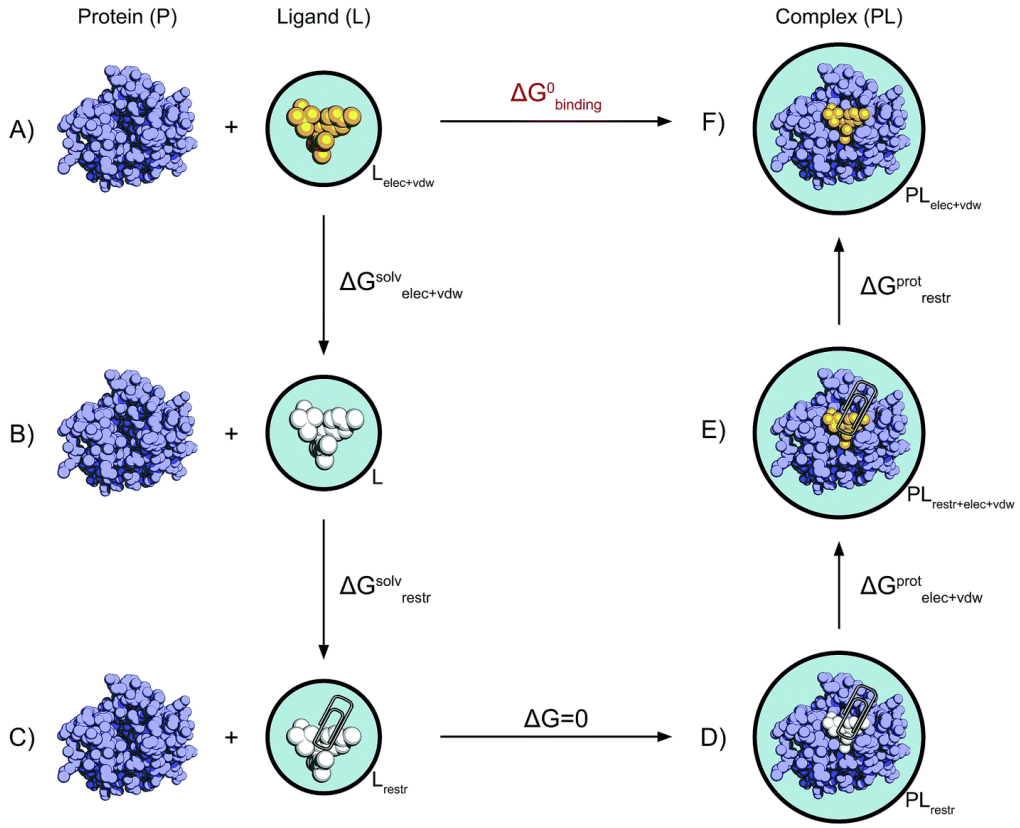


Figure 2.2: Alchemical cycle for ABFE calculations. Reproduced with permission from Aldeghi *et al.*, Chemical Science.[177] - Published by The Royal Society of Chemistry.

The energy loss of these degrees of freedom in the ligand from this combination can be calculated analytically from statistical mechanics, allowing the number of MD simulations used for ligand decoupling to be reduced.[184] The energy penalty calculation for loss of freedom for the ligand is shown in eqn 2.47:

$$\Delta G^{solv}_{restr} = -k_B T ln \left[\frac{8\pi V^0}{r_0^2 sin\theta_{A,0} sin\theta_{B,0}} \cdot \frac{(K_r K_{\theta_A} K_{\theta_B} K_{\phi_A} K_{\phi_B} K_{\phi_C})^{\frac{1}{2}}}{(2\pi k_B T)^3} \right] \tag{2.47}$$

where K is the force constant applied in the restraints, R is the equilibrium bond length restrained, θ_A and θ_B are the equilibrium values of the two angle restraints, and ϕ_A, ϕ_B, and ϕ_C are the equilibrium values of the three dihedral restraints.[184]

Estimating ΔG_{bind} from intermediate states

Calculation of ΔG_{bind} using alchemical methods requires sufficient sampling of phase space, over the λ space of the transformation. Practically, this can be be

done by sampling λ-space over several discrete windows that ensure sufficient phase space overlap with each other. Hence, the free energy changes are performed using individual MD simulation windows, each at a different stage along the decoupling/recoupling λ pathway. The position on this pathway is defined by the variable λ, which can hold any value between 0 and 1.

Thermodynamic integration (TI) is a widely used gradient method for extracting the free energy values, calculating the ensemble average of $\frac{\Delta G}{\Delta \lambda}$ for a series of discrete λ values. The overall free energy change over all intermediate steps is then extracted through numerical integration, normally using the trapezoid rule:

$$\Delta G_{0,1} = \int_{\lambda=0}^{i=1} \frac{\partial U(\lambda, x)}{\partial \lambda} \tag{2.48}$$

This method has been used by Bhati *et al.* in their TIES approach to use an increased number of replicas to better estimate $\frac{\Delta G}{\Delta \lambda}$ for each intermediate step.[185]

The *Zwanzig relation* is a perturbation approach that uses an exponentially weighted ensemble average of the differences in potential energy between two states $(0, 1)$ with sufficient phase space overlap to estimate the change in free energy, shown in eqn 2.49. Although being mathematically exact, this method shows slow convergence and the direction of calculation can lead to hysteresis. An improvement on these exponential averaging techniques is *Bennett's acceptance ratio*, which is a maximum likelihood estimate of the free energy from two samples, in two states. This method requires less samples to converge. The results are even more improved using the multistate Bennett's Acceptance Ratio (MBAR), proposed by Shirts *et al.*,[186] which uses a series of weighting functions to reduce the uncertainty of the values obtained. Thus MBAR has been used for all free energy calculations in this thesis, as implemented in the `alchemical-analysis.py` script.[187]

$$\Delta G_{0,1} = -k_{\mathrm{B}} T \ln \left\langle e^{-\beta(U_1(x) - U_0(x))} \right\rangle_0 \tag{2.49}$$

Replica exchange has been shown to aid convergence of calculations and improve the accuracy of the values identified. This technique enhances the sampling through multiple replicas of the same system run at different thermodynamics states. Using

a Monte Carlo selection criteria for state exchange, this method allows large energy barriers to be overcome in the system while maintaining the correct distribution of states. This technique also allows λ to be varied, through *Hamiltonian Replica Exchange*, increasing the mixing between states and thus improving the phase overlap. These methods come at an increased computational cost and require all intermediate states to be run in parallel to allow communication, given the resources available this technique was not used.

Removing artefacts: Charge correction

For systems that contained a charged ligand, removal of the ligand introduces a change of net charge over the system and thus leads to artefacts. Although still a major issue in alchemistry fields, there are several techniques that have been employed to correct for this observation.[188, 189] The semi-analytical correction scheme set out by Rocklin *et al.* uses Poisson-Boltzmann calculations on the complex and free ligand systems in implicit solvent to estimate the overall free energy difference imposed due to the introduction of charge.[189] This was implemented using APBS (v.1.5).[183]

Methods for performing ABFE

All simulations of differing λ values were parameterised and solvated as described in § 2.2.7, where simulations were performed for 15 ns in triplicate without replica exchange, after 200 ps of NVT equilibration and 200 ns of NPT equilibration. The non-physical cycle used for the the calculation of absolute binding free energies consists of three stages of calculations: firstly the ligand vdW interactions were decoupled and Coulombic interactions annihilated over 32 MD simulations, with the free energy of this transformation calculated using the mBAR method, as implemented in the *alchemical_analysis.py* script.[187] In GROMACS 2019, this stage was split into 11 windows for annihilation of Coulombic interactions and 21 windows for vdW decoupling, each with a separate mBAR calculation. The protocol for calculations in both 2016.3 and 2019.0 were slightly different to that used by Aldeghi *et al.*, removing the now defunct requirement for λ-space transitions to

dummy atoms and instead fully decoupling the ligand from its environment. This allowed for total ligand annihilation to the gas phase, rather than conversion to an 'off' state. The LJ-PME cutoff correction was also applied to remove artefacts derived from using a long-range cutoff within the simulations.[156]

Secondly, the loss of degrees of freedom caused by addition of Boresch restraints are accounted for through an analytical correction, discussed in Section 2.4.2. Finally, the free energy of reintroducing the nonbonded interactions and removal of the Boresch restraints is calculated by a second mBAR calculation over 42 windows for GROMACS 2016. The phasing in/out of nonbonded interactions was performed using a linear transformation of $\Delta\lambda = 0.05$ for the van der Waals and $\Delta\lambda = 0.1$ for the Coulombic transformations. The addition/removal of ligand restraints used 12 non-uniformly distributed λvalues (0.0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.3, 0.5, 0.75, 1.0). In GROMACS 2019 the same setup is used, however separated into a set of 21, 11 and 12 λ windows for the vdW recoupling, reintroduction of the Coulombic interactions and removal of the restraints, respectively. These spacings of λ values was proposed by Aldeghi *et al.*[58]

Proteins and ligands were prepared and parameterised as described in Section 2.2.7, where ligand **41** (discussed in § 4.2.1) was parameterised with RESP charges and **45** was eventually parameterised using AM1-BCC charges (discussed in § 4.2.2).[157, 158] Initial ligand poses were generated through docking, and then subjected to 5 ns of production MD to obtain an equilibrated system. Where averaged Boresch restraints were used, a 15 ns equilibrium MD simulation was performed and the bonds/angles/dihedrals averaged over the last 5 ns of the simulation.

Large speed increases were available for free energy calculations with the release of GROMACS 2019, where all simulations not involving Coulombic decoupling/recoupling could be GPU accelerated. This provided an opportunity to increase the speed of ABFE calculations and the ability to perform them on GPU heavy, rather than CPU heavy, high-performance computing (HPC) resources. The protocol described below was applied to the calculations described in §

4.4.2 and § 4.5.1. A brief overview of re-examining the bromosporine systems is discussed in § B.1.

It should be noted that when using GROMACS 2019.0, errors arose when the systems were run on HPC resources with large numbers of CPUs. The large number of CPUs lead to multiple mpi threads to speed up the calculations, however a bug present in GROMACS 2019.0 to GROMACS 2019.2 failed to impose any restraints when more than one MPI thread was used (Redmine Bug Number: 2953 - `https://redmine.gromacs.org/issues/2953`). Through enforcing one MPI thread, a slight decrease in speed was observed. However, this allowed correct addition of the restraints.

2.4.3 Grand canonical integration to estimate water binding energies

It has been shown that the binding energies of waters, which are inserted/deleted in a region in Grand Canonical Monte Carlo (GCMC) simulations, can be extracted using Grand Canonical Integration (GCI). Through varying the Adam's value over multiple systems, hence the chemical potential of the systems, it is possible to plot the Adam's value against the number of water molecules in the box. Integration under the curve, hence the GCI name, it is possible to estimate the energy of bringing N water molecules from the ideal gas reservoir into the GCMC region, $\Delta G_{trans}(N)$. Through placing this value in a two step thermodynamic cycle then allows the value of the binding energy to be calculated;

$$\Delta G_{bind}(N) = \Delta G_{trans}(N) - \Delta G_{hyd}(N) \tag{2.50}$$

where $\Delta G_{hyd}(N)$ can be calculated by double decoupling for a given water model and available in the literature.[114] This technique was adapted by Aldeghi *et al.* to rigorously extract the binding energies of both the entire network and individual water molecules within the binding site and then estimate the binding affinities of individual water molecules.[28] This method has been used in this thesis to probe the binding affinities of water molecules in the ZA channel of bromodomain protein.

Method for implementation

The protocol for water binding energy calculations was adapted from the work by Aldeghi *et al.*.[28] Simulation setup started with all structures being aligned to a structure of BRD4(1) (PDB: 2OSS), using PyMOL (v. 1.8), to ensure the definition of the search area remained the same as previously published. Structures were checked for missing residues using the WHAT-IF web-interface,[165] and all non-protein atoms were removed from the structure. The protein structure was then protonated using the *reduce* tool within AmberTools16.[160] The pKa values for Asp, Glu, Lys and Cys residues were predicted using *PROPKA3* (v. 3.1),[190] where these residues were modelled in the structure as deprotonated if the pKa was found to be >1.0 units beyond the pH of solution.[191] The system then underwent a short 1000 step energy minimisation using the *sander* package within AmberTools16.[160]

Two GC regions were defined, one encompassing the binding site waters and another covering the ZA channel of the protein. Proteins were parameterised with the AMBER ff14sd force field.[192] Only the TIP3P water model was used,[168] unlike the weighted average of the TIP3P, TIP4P and SPC water models previously performed by Aldeghi *et al.*. The bulk hydration enthalpies for the TIP3P model was set at -6.18 kcal/mol.[28] Residues more than 20 Å away from the GC region were removed for the simulation, and a solvation shell of radius 30 Å was imposed around the centre of the GC region. An equilibration set of 50M solvent only moves were performed, where GC region waters were removed before the production simulation.

A total of 32 GCMC replica exchange simulations were performed using ProtoMS 3.3 in triplicate,[174] where the replicas covered a range of Adam's values from -1 to -32, in integer increments. All simulations were performed at 298 K with a long-range interaction cut-off of 10 Å. Before production MC, 15M moves of the TIP3P waters proceeded a further 5M moves of both the solvent and protein. Production simulations were performed using 50M moves, with data exchange allowed every 0.2M steps, as described by Aldeghi *et al.*[28]

Energies were calculated using Grand Canonical Integration (GCI), using scripts within ProtoMS 3.3.[174] The equilibration period was calculated through

an adapted version of the *calc_series.py* script, and the average equilibrium period over all 32 windows was used for the set of simulations. The lowest integer Adam's value was found using the *calc_gci.py* script, and this value was to identify the most stable waters through clustering with *calc_clusters.py*. To find the energies of each crystallographic water molecule, the clusters best representing the crystallographic waters were selected. A non-rigorous localised GCI calculation, looking at the energy difference between bulk and a 2 Å sphere around each cluster was performed using *calc_local_gci.py*, written by Dr G. Ross and available at `github.com/gregoryross/watertoolbox-protoms.git`. This yielded an approximation of the binding energy of each water molecule.

2.5 Molecular feature prediction

Cavity detection and pharmacophoric feature search

The Cambridge Crystallographic Data Centre's (CCDC) SuperStar (2019 v.1) was used to predict cavities and probe pharmacophores.[193] *Apo* structures of the PHD/BRD from TRIM24, TRIM33α & TRIM33β were taken from the PDB files 4YAC, the completed structure of 3U5M and 5MR8, respectively.[53, 76]. Cavity detection was performed over the entire protein at physiological pH, using the 'shallow/normal' cavity type, a grid spacing of 0.7 Å and a minimum cavity volume of 10 $Å^3$. Pharmacophores were detected through screening a range of atom type propensities against entire protein structures.

pKa prediction

pKa predictions of amine groups was performed using ACD labs (ACD, version 2018.1, Advanced Chemistry Development, Toronto, Canada, www.acdlabs.com, 2019) and ChemAxon's *cxcalc* tool. For all simulations, the major microspecies was modelled unless stated, which was identified using the *majormicrospecies* tool within the Marvin Suite.

Geometry optimisation

The optimised geometry calculation of the urea containing compounds, **27** and **28** (discussed in § 3.3.1) was performed in Gaussian 2003 (Rev. E.01), at the HF/6-31G(d) level of theory. Input Gaussian files were prepared and converted using *antechamber*, found within AmberTools16.[194]

2.5.1 Tractable library building

For design of BI-2536 analogues in § 5.3.1, all 3922 available acyl chlorides and carboxylic acids were downloaded from the Sigma Selected Structures website (April 2019). These were joined to **100** using the following reaction SMART: `[C:1](=[O:2])-[C1D1,OD1].[N!H0:3] >> [C:1](=[O:2])[N:3]`. The resulting 4497 compounds were then subject to a PAINS filter, using a substructure match in RDKit (v. 2017.03.3 - `www.rdkit.org`) against the PAINS smarts available at `github.com/rdkit/rdkit/blob/master/Data/Pains/wehi_pains.csv` based on the publication by Saubern *et al.*[195] The remaining 4415 compounds were then subjected to a structure factor index (SFI) < 11 filter, implemented with ChemAxon's cxcalc tool to remove 76 compounds. The remaining compounds were prepared for docking with GOLD, with a H-bond constraint to ensure the KAc mimic binds the conserved asparagine.

For building of benzimidalazone analogues, the entire MOE linker library was screened against the selected C-H bond.[196] If a substitution was spatially allowed, it was then scored within MOE. The resulting compounds were then subject to the aforementioned PAINS filter to remove likely false positive compounds. The resulting compounds were subject to a diversity filter, to reduce the number of MD simulations to perform. This filter was implemented using the *MaxMinPicker* within RDKit, which selects compounds with the greatest distance between their Morgan fingerprints.

2.6 Docking

Docking is a powerful *in silico* tool for predicting binding orientations of one molecule to another using a search algorithm, which is often combined with a scoring function to measure the strength of this association. This has been employed with large macromolecular targets, probing protein protein interactions and DNA binding.[197] However, docking has shown its strength in protein-ligand docking, where entire libraries of compounds can be screen against a protein target in relatively short time periods in Virtual Screening.

Within docking, there are several restraints used to reduce the time taken to dock. The search region, or box, is often centred around the binding site with a finite boundary to reduce the search space. This can reduce the docking time significantly. However, if the binding site of the protein is not known, then the search region is often increased. This method, known as 'blind docking' is useful at identifying binding sites in proteins. Binding sites are dynamic, but sampling all protein conformations is time consuming.

$$\Delta G = \Delta G_{vdW} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{conform} + \Delta G_{tor} + \Delta G_{sol} \tag{2.51}$$

The first major component of docking is the searching for low energy poses. The aim here is to quickly probe the energetic landscape of ligand and protein binding to find a minimum. Different softwares approach this in different ways, with AutoDock4 choosing to use a Lamarckian Genetic algorithm. The second component is the scoring functions, which provides an estimation of the strength of the given low energy pose. Force field approaches have been used in scoring functions, however these are often expensive. Empirical based scoring functions are a fraction of the computational cost and still perform well at ranking poses and giving a scale of the possible interaction strength. Equation 2.51 shows the empirical structure based function that is used within AutoDock4. This has the standard bonded and nonbonded interactions that you would expect in a force field,

see section 2.1.1, but with additional terms to account for the ligand desolvation (ΔG_{solv}) and loss of degrees of freedom upon binding (ΔG_{tor})[106]

2.6.1 Application of docking within this thesis

The work within here looks at using the AutoDock4 package and MOE to perform molecular docking. Protein models were protonated and water networks of the binding site and ZA channel waters optimised within the Molecular Operating Environment (MOE) package (v. 2018.2) with the exception of the docking performed in Chapter 3. For the docking within this Chapter only waters W0-W5 were retained in the protein structure, all other crystallographic waters, ions and peptides were removed. Protein residue protonation states were assigned within GROMACS 2016.3, with water orientations optimised using a steepest decent energy minimisation protocol with a cut-off of 100 $kJ\ mol^{-1}$. This protocol used the TIP3P water model,[168] and the AMBERff99sb-ildn force field for the protein.[161]

All protein structures were aligned to Chain B of the *apo* TRIM24 structure (PDB ID: 3O33), except structures used for blind docking which were rotated to fit a Cartesian axis using the *save_transformed.py* PyMOL plugin. Binding site definitions for site specific docking and blind docking can be found within the Appendix, Table A.3 & Table A.4.

AutoDock4

Site specific docking was performed using AutoDock4,[107] using a population size of 100, 2,500,000 GA energy evaluations and a grid spacing of 0.35 Å. Blind docking, over the entire protein structure, was performed using a grid spacing of 1.00 Å with all aforementioned parameters. Blind docking was not performed in AutoDock Vina,[198] as this software failed to reproduce ligand binding poses of TRIM24 ligands in the TRIM24 BRD (data not shown). Ligand protonation states at pH 7.4 were assigned using the *cxcalc* tool within the Marvin Suite 16.16.6.0, 2016, ChemAxon (`https://www.chemaxon.com`). Protein and ligand structures were converted to the required PDBQT format using the *prepare_ligand4.py* and

prepare_receptor4.py scripts, found within the MGLTools package (v. 1.5.6).[107] Ligand poses were clustered using a 2 Å RMSD cut-off, which was able to differentiate between different binding poses, with the lowest energy pose of each cluster considered for further studies.

MOE

Docking within MOE (v. 2018.2) used the default settings,[196] with a pharmacophore query to require an acceptor atom to H-bond with N1039 in TRIM33β.

2.7 Homology modelling

In the absence of a crystal structure, homology modelling can be used to build a 3D model based on the structures of similar proteins. This comparative modelling technique relies on the identification of known structures (templates) with a high degree of sequence homology to the sequence of the unknown structure (target). Through alignment of the sequences, a model of the target can be built from one or more templates.[199, 200]

Methods used for homology modelling

MODELLER (v.9.9) was used for all homology model generations, with the generic protocol being the generation of 100 models created, where the model with the lowest DOPE score was taken forward to validation.[201, 202] MD simulations of the *apo* protein were used to check the stability of the models, through assessing the root mean squared deviation (RMSD) of the protein backbone atoms over time.[203] Homology modelling was also employed within Chapter 3 to model the missing BC and inter-domain linker loop residues in TRIM33α, using the partial structure from PDB ID: 3U5M. In this case a consensus protein structure prediction was performed using the Prabi server(`npsa-prabi.ibcp.fr`) to dictate structural restraints. The model was generated in two parts: i) The predicted α-helices were added as an extension to α_B and α_C, using PDB ID: 3U5M as the template ii) The BC and linker loops were then added, with the α-helices restrained.

Why has pharmaceutical research and development lagged so far behind other industries in the development and application of simulation and modelling for research and development?

— Walter S. Woltosz, *If we designed airplanes like we design drugs...*[94]

3

Determining the binding modes of TRIM33 ligands

Contents

The work presented in this chapter aims to understand the binding modes of ligands identified in a previous high-throughput screen for TRIM33. This work relies on molecular docking and subsequent MD studies to assess the docked poses.

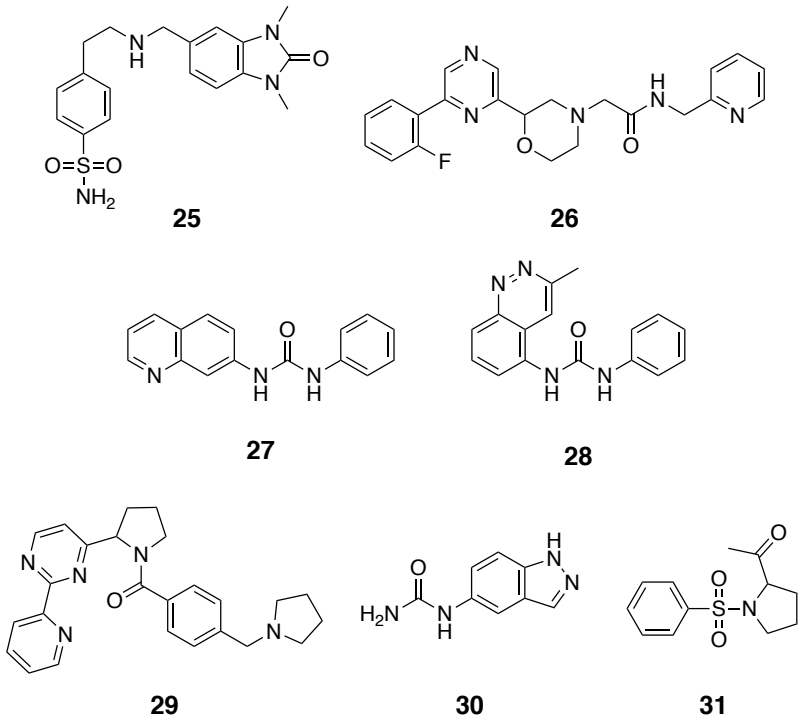


Figure 3.1: Final ligands from high-throughput screen. Over 30,000 compounds from the Maybridge and PPI-NET library were screened against TRIM24 and TRIM33 PHD/BRD cassette, with 6 compounds remaining after refinement.[204] All ligands in the figure show binding to the TRIM proteins through the AlphaScreen™ assay, with determined IC_{50} values shown in Table 3.1.

3.1 Introduction

In a series of AlphaScreen™ and WaterLOGSY assays, described in Section 1.5, seven compounds from the PPI-Net and Maybridge libraries were identified as binders of the TRIM24/TRIM33 PHD and BRD reader domains by Dr A. Sekirnik and Dr L. See.[124, 204] Four of these compounds - **25**, **26**, **27** and **28** - were found to bind in both orthogonal assays. Compound **29** was found to not bind by WaterLOGSY, despite showing binding within the AlphaScreen™ assay.[204] Two further compounds - **30** and **31** - were either not soluble in the WaterLOGSY assays conditions or gave ambiguous results, despite showing binding within the AlphaScreen™ assay.

The AlphaScreen™ assay data are summarised in Table 3.1. The AlphaScreen™ data shows IC_{50} values calculated from displacement of three modified H3 peptides. These

Table 3.1: AlphaScreen™ data for TRIM24/33 HTS hits. IC_{50} values were determined for two of the HTS hit compounds. It should be noted that differing affinities of the H3 peptides to the proteins prevents comparison of the values between H3 peptides or proteins. Data was collected by Dr L. See,[204] H3 protein-peptide affinities are shown in Appendix Table A.1. All hits were verified as true positives using the TruHits™ validation kit.[117]

Ligand	Peptide	IC_{50}(AlphaScreen™)/μM		
		TRIM24	TRIM33α	TRIM33β
25	H3(BRD)	> 250	2.79 ± 0.83	18.4 ± 2.06
	H3(PHD)	> 250	> 250	> 250
	H3(Dual)	> 250	14.1 ± 1.62	> 250
26	H3(BRD)	41.7 ± 7.82	51.8 ± 3.87	59.4 ± 5.73
	H3(PHD)	41.8 ± 5.11	43.2 ± 5.67	78.6 ± 9.27
	H3(Dual)	62.2 ± 4.15	56.5 ± 10.3	111 ± 13.5

peptides contain modifications recognised by the reader domains: either K18Ac (known as H3(BRD)), $K9Me_3$ (known as H3(PHD)) or both simultaneously (known as H3(Dual)).[116] Through the differences in peptide displacement, the location of ligand binding can be estimated. However, each modified peptide has differing affinities to the same protein, and the same peptide will have differing affinities to the different proteins. Hence, the IC_{50} values can not be compared between peptides or proteins. It should be noted that conversion to of IC_{50} to K_i values could make these values comparable. Unfortunately, the IC_{50} curves fail to reach 100% peptide displacement, preventing this numeric conversion.[204]

Wild type (WT) and mutant WaterLOGSY data are shown in Table 3.2. Studies on the N to F and W to A mutants sought to determine if the ligands bound in the vicinity of the BRD binding site, or in the vicinity of the PHD's Trp residue. These mutant studies removed either an Asn residue, which is vital to KAc binding in the BRD site, or the Trp residue that binds the $K9Me_3$ mark on the H3 peptide.

The data indicate that **25** binds the TRIM24/33 BRD sites, but is not selective between TRIM24 and TRIM33. Compound **26** is TRIM33 selective, but binding dependence on both N1039 and W890, shown by WaterLOGSY, makes the binding site location hard to identify. Compounds **27** and **28** show no selectivity between TRIM proteins, and the WaterLOGSY data indicate no dependence on either N1039 or W890 for binding making the binding site ambiguous. However, a

Table 3.2: WaterLOGSY data for TRIM24/33 HTS hits. The ✓symbol denotes when binding was detected, a ✗ notes when binding was not detected and ∽ denotes an ambiguous result. Data were collected by Dr L. See.[204]

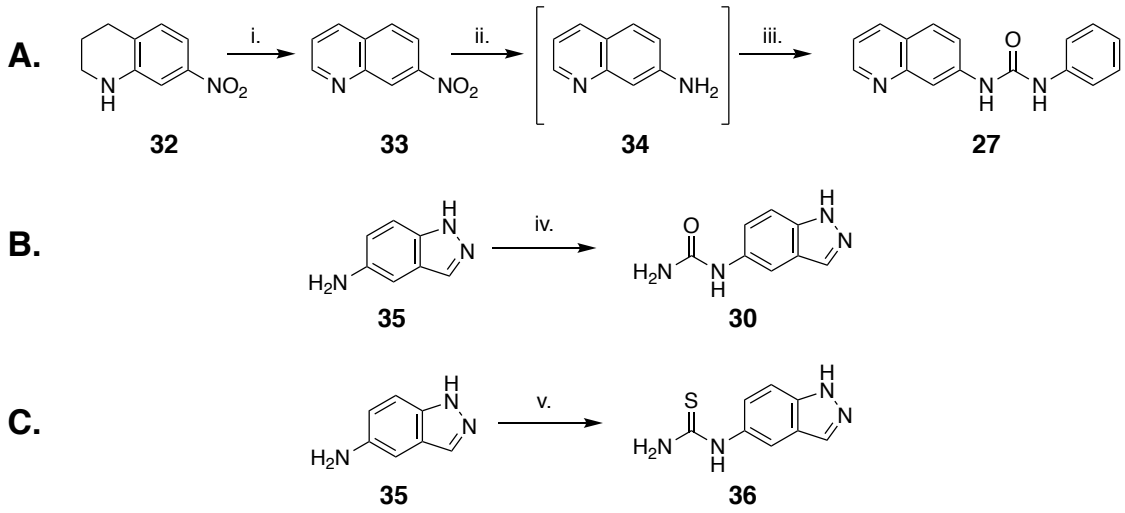
Ligand	TRIM24			TRIM33α			TRIM33β		
	WT	N980F	W828A	WT	N1039F	W889A	WT	N1039F	W889A
25	✓	✗	✓	✓	✓	✓	✓	∽	✓
26	✗	✗	✗	✓	✓	✓	✓	✗	✗
27	✓	✓	✓	✓	✓	✓	✓	✓	✓
28	✓	✓	✓	✓	✓	✓	✓	✓	✓
29	✗	-	-	✗	-	-	✗	-	-
31	∽	-	-	∽	-	-	∽	-	-

positive AlphaScreen™ result for ligands **27** and **28** indicate that the binding sites should be in the vicinity of a one of the H3 binding regions, leading to peptide displacement and assay detection.[204]

There is no crystallographic data of these ligands binding to either TRIM24 or TRIM33, and the assay data only confirm the binding locations of one compound: **25**. The work presented in this chapter initially looks at identifying the binding locations through co-crystallisation. It then pursues an *in silico* approach to identifying potential binding locations of the HTS ligands through blind docking and subsequent MD studies. All synthetic experimental details, can be found at the end of the chapter, with spectroscopic supporting data available in the Appendix.

3.2 Co-crystallisation efforts to identify ligand binding modes

To elucidate the binding locations of the HTS ligands within the PHD/BRD cassette, it was decided to synthesise the compounds and attempt co-crystallisation against TRIM33β, through a collaboration with Prof. S. Knapp. TRIM33β was selected for two reasons, firstly that the TRIM24 BRD already had published inhibitors,[53]. Secondly, the crystal structures of H3 bound TRIM33α (Table A.2) shows that the KAc binding Asn residue resides on the α-helix in a position that is not within the binding site and so likely to be the inactive form of the binding site.[76]



Scheme 3.1: Synthesis of 1-phenyl-3-(quinolin-7-yl)urea (**27**), 1-(1*H*-indazol-5-yl)urea (**30**) & 1-(1*H*-indazol-5-yl)thiourea (**36**). *Reagents and conditions*: i) DDQ, CH_2Cl_2, 60 °C, 2 h, 54% ii) Fe, AcOH, 80°C, 2 h iii) phenyl isocyanate, acetone, rt, 16 h, 37% iv) Sodium isocynate, AcOH, H_2O, rt, 20 hr, 54%, v) 1,1-Thiocarbonyldiimidizole, CH_2Cl_2, rt, 16 h, 25%.

Synthesis of compound **27** was performed from **32**, as shown in Scheme 3.1A, *via* the formation of a quinoline, **33**, that was subsequently reduced to give an unstable amine, **34**. This compound was immediately reacted with phenyl isocyanate. Other compounds shown to bind in the HTS (**25**, **26** & **31**), were synthesised by Dr L. See and A. Scorah. Despite these efforts, no co-crystal structures were obtained through the course of this project.

Further crystallisation attempts expanded to include compound **30**, which was previously excluded from further studies due to a lack of solubility in assay conditions.[204] However, given similarity to compound **27**, a co-crystal structure could aid identifying the binding location of the urea-containing compounds. A one step synthetic route was devised and performed through reaction of the amide (**35**) with sodium isocyanate (Scheme 3.1B). Unfortunately, this compound did not lead to any successful co-crystal structures.

Due to the insolubility in the AlphaScreen™ assays compound **30** was also not confirmed as a true positive binder of TRIM33. To verify this ligand a soluble thiourea analogue of compound **30** was proposed, giving compound **36**, and synthesised as shown in Scheme 3.1C. However, **36** did not show a positive response

in the AlphaScreen™ assay, indicating it has a low affinity for TRIM33β. This ruled **36**, and its predecessor **30**, out of all subsequent studies.

3.3 *In silico* approaches to locate ligand binding sites

With no co-crystallisation studies able to identify the binding locations of the HTS hits, and only compound **25** having assay data indicating its binding site, an orthogonal, *in silico* approach was adopted. This approach commenced with experiments to determine the regions of the protein that could be potential binding sites. These sites were then the focus of cavity detection and H3 peptide MD to elucidate possible binding sites of **26**, **27** and **28**.

Before the crystal structures could be used in studies (TRIM24: 4YAX; TRIM33α: 3U5M; TRIM33β: 5MR8), structures with missing residues needed to be completed. All crystallographic structures of the TRIM33α PHD/BRD are missing residues on the α_B and α_C loops, which are in proximity to the binding site, likely due to this region pointing towards empty space within the crystal packing. The structure was completed, as described in Section 2.7, and shown in Figure 3.2A. The modelled BRD maintains a stable RMSD during equilibrium MD (data not shown). Both the homology model and crystal structure show that the 17 amino-acid insertion positions the KAc binding residue, N1039, away from the binding site. This supports the hypothesis that TRIM33α could be the non-chromatin-binding isoform of the BRD. However, crystal structures (PDB ID: 3U5O) still show H3 binding to the BRD *via* an alternative interaction between E981 and H3R17.[76] ITC experiments by Dr L. See, Table A.1, show that there is a ten-fold decrease in affinity for the H3(BRD) peptide for TRIM33α when compared to TRIM33β.

The AlphaScreen™ assay relies on peptide displacement by a ligand. Hence the result implies that the ligand binds in the same regions of the protein as the H3 peptide. This principle narrowed the search for binding regions of ligands detected by AlphaScreen™. This is not true for ligands that give a positive result by WaterLOGSY, where ligand could bind anywhere on the construct. However,

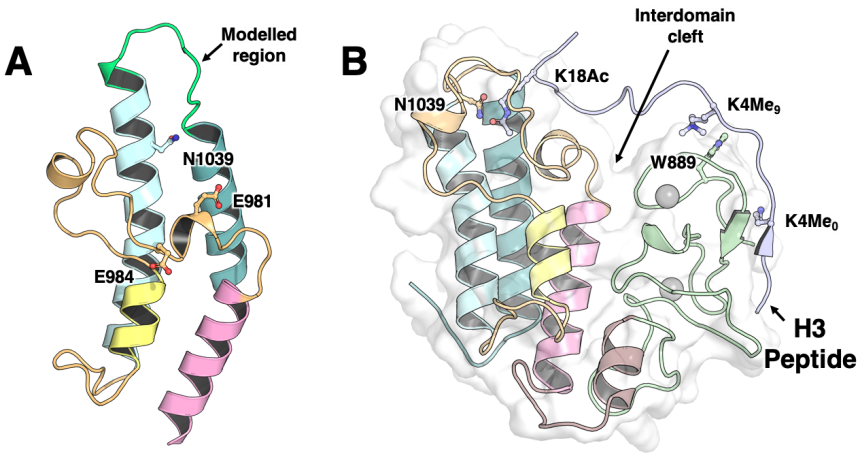


Figure 3.2: TRIM33 structures. All crystallographic structures of the TRIM33αBRD are missing the BC loop, which was modelled (**A**) using the homology modelling protocol described in §2.7. All TRIM24/33 tandem PHD/BRD structures have the same quaternary structure, demonstrated by (**B**) TRIM33β (PDB: 5MR8) with the H3 protein from a TRIM33α structure (PDB: 3U5O) imposed.

this assay can be combined with mutant studies to provide information on which residues are required for ligand binding.

There are no complete structures of the H3 peptide binding TRIM33, with the TRIM33α structure (PDB ID: 3U5O) missing the BC loop of binding site, and TRIM33β missing all H3 residues beyond K9. The two structures can be superimposed, Figure 3.2B, to model the structure of TRIM33β with the residues 1-21 of H3 histone tail. In the TRIM33α structure, the H3 peptide binds in the BRD binding site *via* K18Ac. It extends over the inter-domain cleft, forming a cation-π interaction between W889 through H3K9Me_3, then forming an extended β-sheet-type interaction with the PHD at the K4Me_0 position. The superimposed TRIM33β H3 peptide model was subjected to more rigorous MD studies to probe whether these interactions are conserved.

Simulations of TRIM33β binding H3K18AcK9Me_3K4Me_0 showed that the peptide remains bound in all simulations. Multiple anchoring points on the TRIM cassette were consistent with the crystallised H3 peptide structures. No rearrangement was observed of the peptide and the anchoring points on the PHD/BRD cassette. The KAc mimic was observed to interact with H3K18Ac

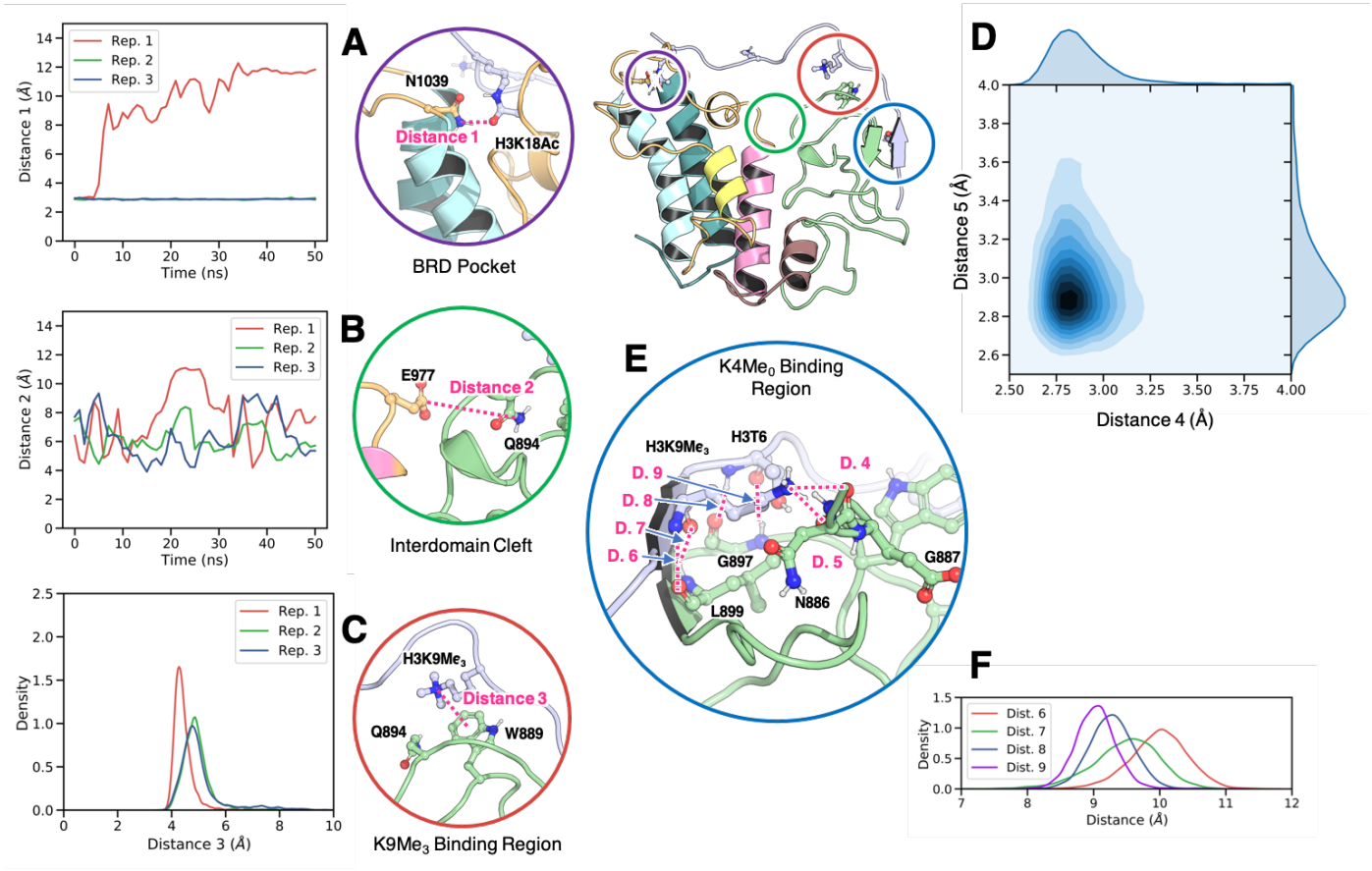


Figure 3.3: MD simulations of the H3 peptide bound to TRIM33β PHD/BRD cassette. (**A**) The distance between the KAc and N1039. (**B**) The distance between E977 and Q894, at the top of the interdomain cleft. (**C**) The H3K9Me$_3$ distance to W889. (**D**) Comparison of distances between ζ-nitrogen of H4Me$_0$ and backbone oxygen of N886 and G887. (**E**) The hydrogen bonds formed between the H3 peptide and the PHD (**F**) Distances of the H-bonds formed between H3K4 to L889 and H3T6 to G897.

(Figure 3.3A) confirming that ligands which bind to the BRD could displace the H3 peptide. H3R17 is found to sample multiple interactions around the BRD pocket (data not shown). The H3K14 residue points into solution and rarely towards the inter-domain cleft, which is unsurprising as it is not resolved in any of the crystal structures.[76] The $K9Me_3$ group maintains its charged interaction with W889, which is seen in the crystal structures (Figure 3.3C). There are a further six polar interactions in the $K4Me_0$ binding site, the first two were formed between the chain of $H3K4Me_0$ and the backbone atoms of G887 and N886 (Figure 3.3D). The remaining four interactions were in the β-sheet-like interaction between H3T6-G897 and H3K4-L889 backbone atoms (Figure 3.3E). The inter-domain cleft is gated near K3H14 through a hydrogen bond between E977 and Q894, which is open in the crystal structure (PDB: 5RM8), giving a range of cleft conformations (Figure 3.3B). Furthermore, comparing these simulations to $H3K18AcK9Me_0K4Me_0$ shows a lack of dependence of $K9Me_3$ for peptide binding (data not shown), as shown in previous experimental work.[124]

It was also important to identify any additional ligand binding sites in the protein construct beyond the H3 contact points. If additional sites are present, these could explain the assay data for urea containing compounds. This data showed binding through WaterLOGSY, but with a lack of dependence on W889 or N1039. The available crystal structures of the TRIM24, TRIM33α and TRIM33β PHD-BRD cassettes (Appendix Table A.2) have most of residues present in the assay constructs resolved. Alignment of the crystal structure sequences with the experimental recombinant protein constructs show relatively short peptide chains missing from either terminus. Hence it was appropriate to use the crystal structures for *in silico* characterisation.

Applying a cavity detection tool to the TRIM24 and TRIM33 cassettes, multiple shallow binding cavities were detected (Figure 3.4). All the structures show a deep cavity in the BRD, which contains the KAc-binding Asn residue. A long, thin cavity is seen in the inter-domain cleft between the PHD and BRD, extending along the PHD and BRD boundary beyond the H3-binding region. There is a shallow

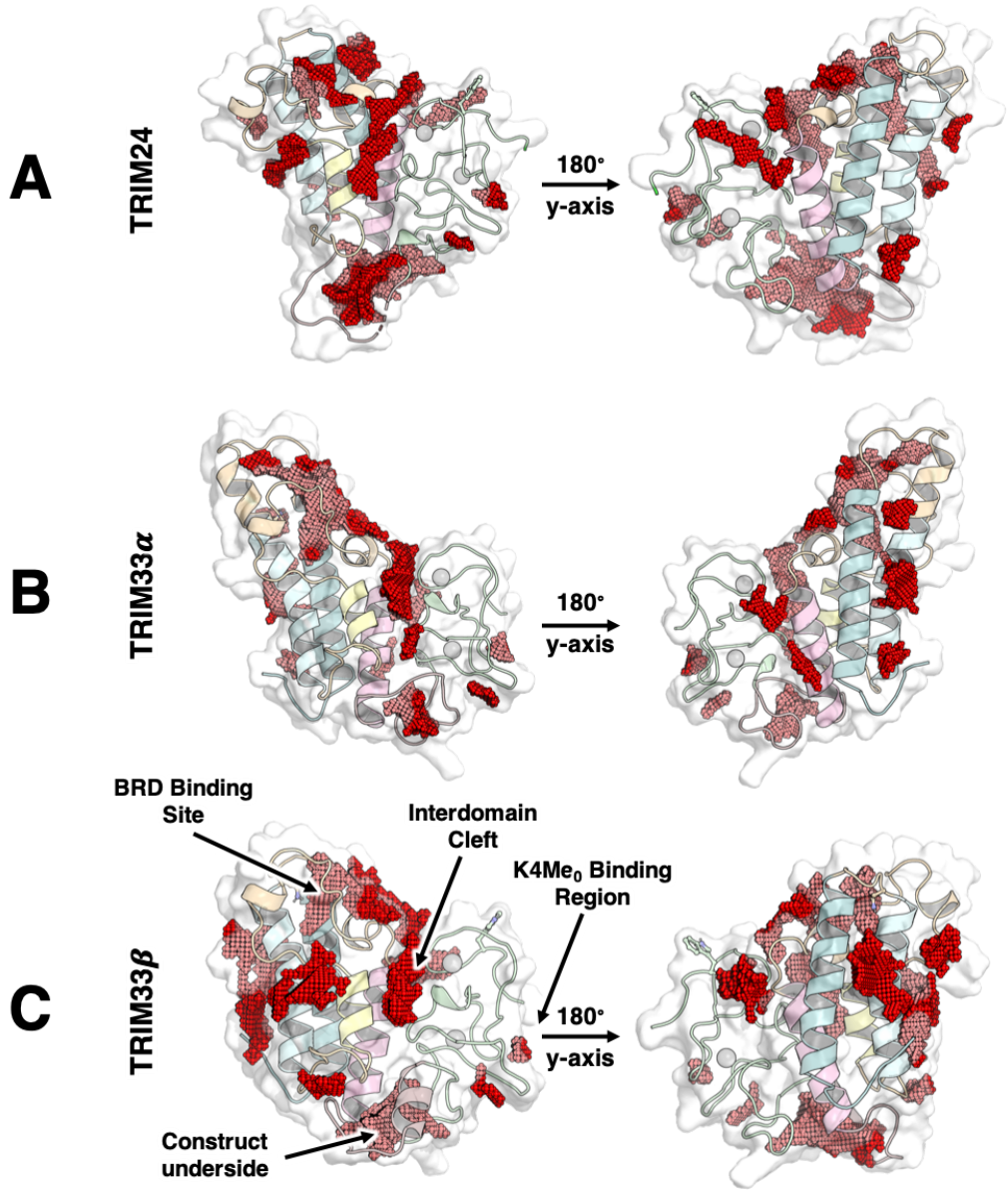


Figure 3.4: Cavity detection search on TRIM24/33 PHD/BRD cassettes. A cavity search was performed using the CCDC Superstar cavity detection search tool,[205] highlighting putative binding regions on (**A**) TRIM24, (**B**) TRIM33α and (**C**) TRIM33β. For all three proteins cavities were observed in the BRD-KAc binding region, the inter-domain cleft, a shallow site surrounding the K3Me$_0$-binding region, a large cavity in the construct underside, and regions between the BRD α-helices.

cavity in the K4Me$_0$-binding region and no cavity adjacent to W890, which binds K9Me$_3$. Each construct shows small cavities in the regions between α-helices on the BRD. Finally, all structures show a larger cavity on the opposite sides to the

PHD and BRD binding sites, which we refer from here as the 'underside' cavity. This cavity may not be biologically relevant in the context of the entire TRIM protein quaternary structure, however, it remains a possible binding site on the constructs used in the experimental assays.

3.3.1 Identifying ligand binding sites through blind docking studies

As discussed in Section 3.2, given the novelty of the target and lack of published ligands when compared to TRIM24, [53] we decided to focus all subsequent studies on the TRIM33β PHD/BRD cassette.

Without a co-crystal structure, and a range of possible binding sites for each ligand, with the exception of **25**, a series of blind docking studies were performed to identify potential binding sites for each ligand. The search region encompassed the entire protein above the underside cavity, which was independently docked in studies described later. Compounds **27**, **28** & **26** were each docked 100 times against the TRIM24/33 cassettes, with the ligand centre of masses shown as a green sphere in Figure 3.6. Ligand **30** was also reintroduced for this study, where docked poses could help identify urea-binding regions. The ligand model torsions were set to allow all non-conjugated bonds to rotate, along with the urea-aryl bonds in compounds **27** & **28**. This decision was supported through optimised geometry calculations, Figure 3.5, which showed the urea-containing compounds adopted a folded conformation to reduce steric repulsion of hydrogen atoms, and allowing partial stacking of the conjugated rings. Though it should be noted that these geometries were calculated in the gas phase.

Blind docking of compound **30** shows that the mono-capped urea could act as a KAc mimic and bind within the BRD pocket. This group could also bind at the K4-binding site on the PHD, mimicking the amide backbone that binds the PHD to H3K4Me$_0$. There is also possibility of it binding in the inter-domain cleft. When the urea is capped at both ends, as in **27** and **28**, resulting in folded conformations, there is an increase in the number of possible binding sites. Both **27**

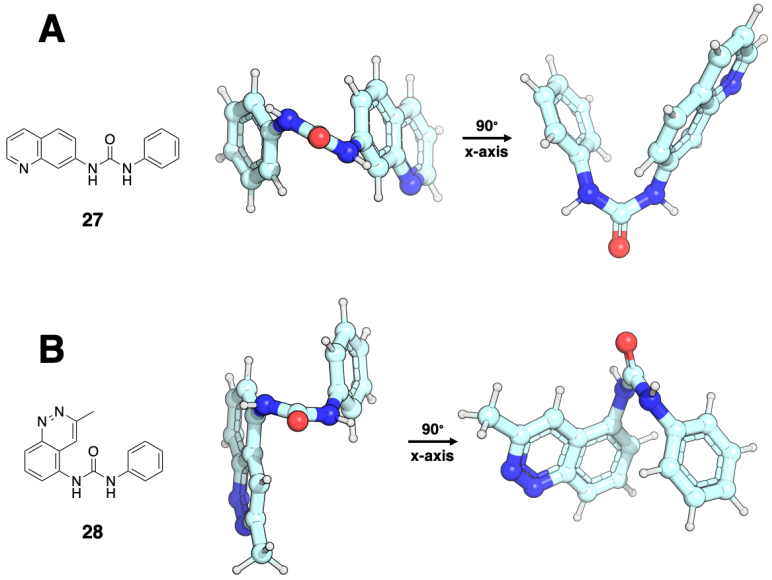


Figure 3.5: Optimised geometry performed for compounds **27** & **28**. DFT calculations were used to calculate the optimised geometry for compounds **27** (**A**) and **28** (**B**), which showed a folded conformation over a linear one.

and **28** present a range of possible binding sites. The spheres in Figure 3.6, which represent the ligand centre of masses, do not overlap perfectly, indicating multiple binding modes within these sites. There were also multiple sites identified between the α-helices in the bromodomain, which could be an artefact of the poor scoring of ring systems in docking. Comparing these observations with the assay data, the lack of N1039 dependence means the docks to the BRD could be ignored, as could the sites on the α-helices. Hence, the blind docking showed that further studies into the inter-domain cleft and the K4Me$_0$ binding site are needed. Binding in the inter-domain cleft or at the K4 site fits with the AlphaScreen™ data, as these sites could lead to H3 displacement. This could occur either through bulk at the domain interface or competing against the β-sheet formation at the H3K4Me$_0$ site. Blind docking of compound **26** indicated that its binding site is in the proximity of the BRD and top of the inter-domain cleft. This is consistent with the WaterLOGSY data which showed dependence on N1039 and W889, which is close to this region. This region will be the focus of future studies.

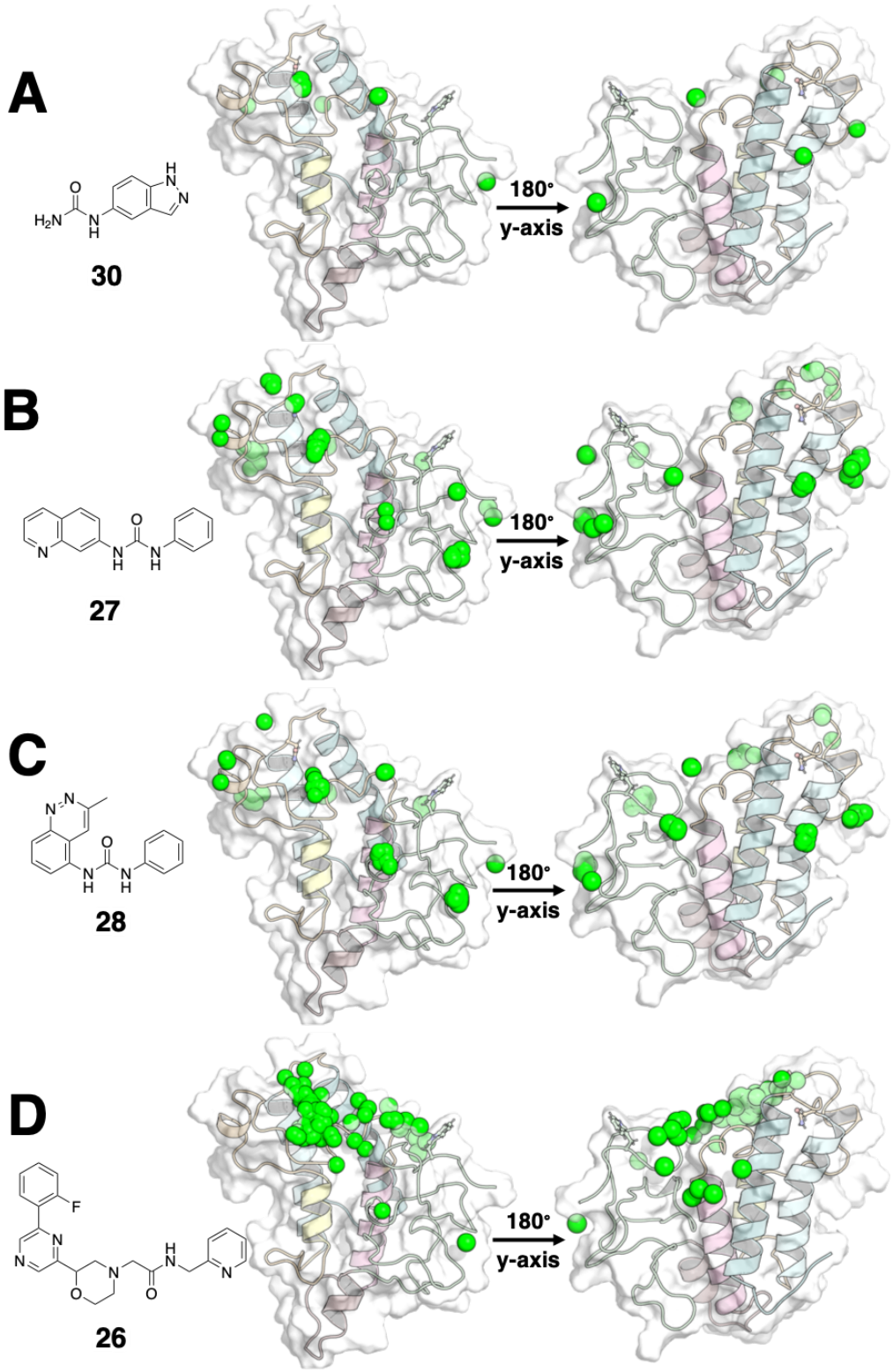


Figure 3.6: Blind docking against TRIM33β. Ligands **30** (**A**), **27** (**B**), **28** (**C**) and **26** (**D**) underwent blind docking studies against TRIM33β(shown) and TRIM24 (Appendix). For clarity, the green spheres represent the ligand centre of mass,

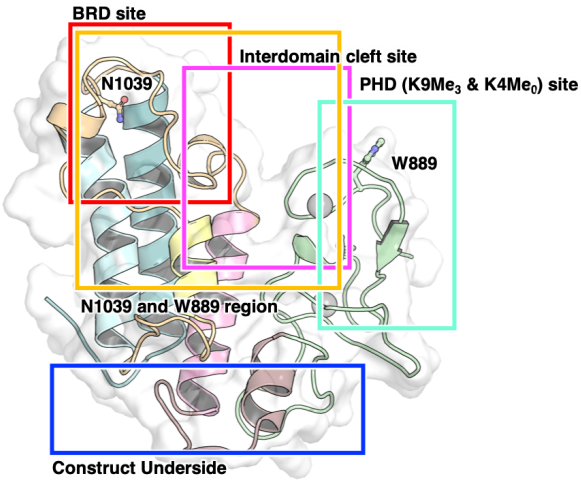


Figure 3.7: Defined docking regions used for site directed docking of TRIM33β (PDB: 5RM8). A range of binding regions were selected for site directed docking studies i) The BRD site (red) for compound **25**; ii) The BRD site and adjacent ZA channel (orange) for compound **26** (orange); iii) the interdomain cleft (pink) for compounds **27** & **28**; iv) the PHD (cyan) encompassing the K9Me$_3$ binging Trp residue and the K4Me$_0$ binding interface for compounds **27** & **28**; v) The 'underside' cavity.

3.3.2 Site specific docking and MD studies to identify ligand binding poses

To validate the sites identified through blind docking, a two step approach was adopted. Firstly, site-specific docking was performed at the regions identified by blind docking. This docking was completed using a smaller grid spacing of 0.375 Å to increase the resolution. Poses were inspected, manually filtered, and finally the pose stability was assessed through molecular dynamics simulations. Assessing pose stability through MD has been shown by Liu *et al.* to be a useful measure of whether ligand binding is feasible. They show that 95% of correct ligand poses are retained an RMSD of < 2.0 Å in equilibrium MD simulations, while 25-44% of decoy poses could be ruled out through MD using the same criteria.[206]

Compound **25** was docked to the KAc-binding site. Compounds **27** & **28** were docked in three grids centred around: i) the inter-domain cleft, ii) a region encompassing the H3H4Me$_0$ and H3K9Me$_3$ binding sites iii) an additional docking region encompassing the underside cavity. Finally, compound **26** was docked to a larger region encompassing both the BRD and inter-domain cleft cavities. This

decision was based on the large region of binding in the blind docking studies. These regions are depicted in Figure 3.7 and each docking study is expanded on in the subsequent sections.

Studies on the benzimidazolone hit: 25

Compound **25** contains an obvious KAc mimic within its benzimidazolone core. This core has previously been used in BRD ligands, including the series developed against TRIM24 which have been accompanied by co-crystal structures.[53] **25** bound to TRIM24 and both isoforms of TRIM33 with low micromolar affinity Table 3.1, in AlphaScreen™ assays. Docking of **25** to the TRIM33β BRD yields two poses that have the KAc mimic interacting with N1039. Both these poses placed the amine containing vector (at the 5-position) projecting from the core towards (Figure 3.8B) or away from the binding site (Figure 3.8C). Confidence in these poses is increased when compared to crystal structures of similar ligands bound to TRIM24, Figure 3.8A. In this structure the KAc mimic sits in the BRD pocket in an identical fashion, with the methyl groups keeping the core vertical within the binding site. There is no clear preference for one pose over the other, visually or by docking score.

The docking shows that E981 can form a polar interaction with the benzylic amine in either pose, an interaction unique to the TRIM33 BRDs, and offers an opportunity to obtain ligand selectivity. Whether the 5-position points towards or away from the pocket, an additional polar interaction *via* the chain's terminal amine can be formed. When the 5-position points towards the pocket, the docking suggests an interaction with the solvent-exposed E984 is possible. Meanwhile, if the 5-position points away from the binding site then an interaction with the backbone oxygen of I980 could be possible.

MD studies showed that in both docked poses, the interactions of the benzylic amine with E981, which was predicted to be charged, is retained for the majority of the simulations (D1 in Figure 3.9). Both poses also retain an interaction with N1039, *via* the benzimidazolone carbonyl oxygen. However, the average RMSD of the ligands are above 2.0 Å, due to the sulfonamide group and linking aryl groups

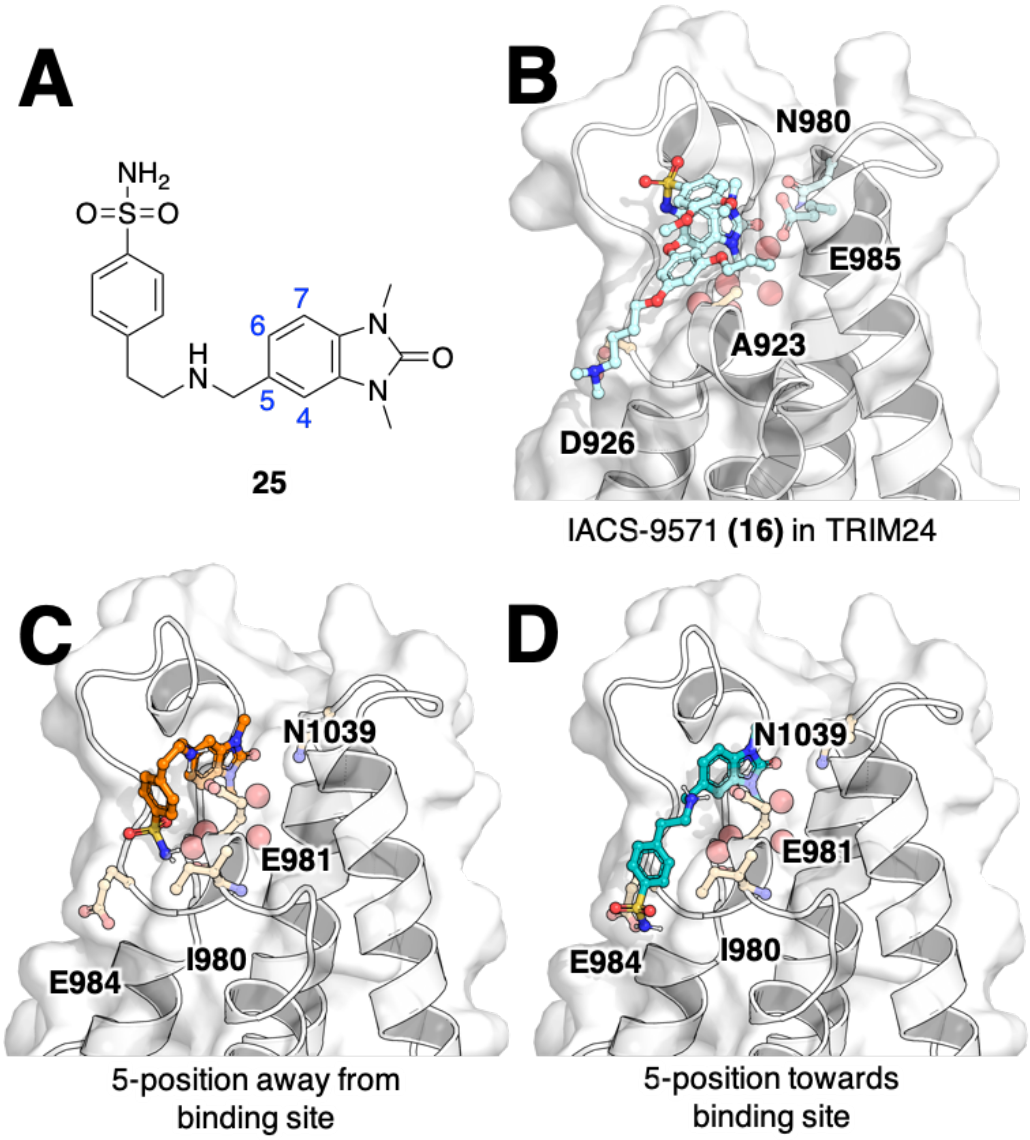


Figure 3.8: Compound **25** docked to TRIM33β and crystal structure of IACS-9571(**16**) bound to TRIM24. (**A**) Compound **25** with vector labelled. (**B**) IACS-9571 (**16**), which contains the same KAc mimic, crystallised in the homologous TRIM24 BRD (PDB: 4YC9). Docking **25** to TRIM33β gave two possible binding modes: (**C**) the vector at the 5-position pointing away from the binding site or (**D**) where it points towards the binding site.

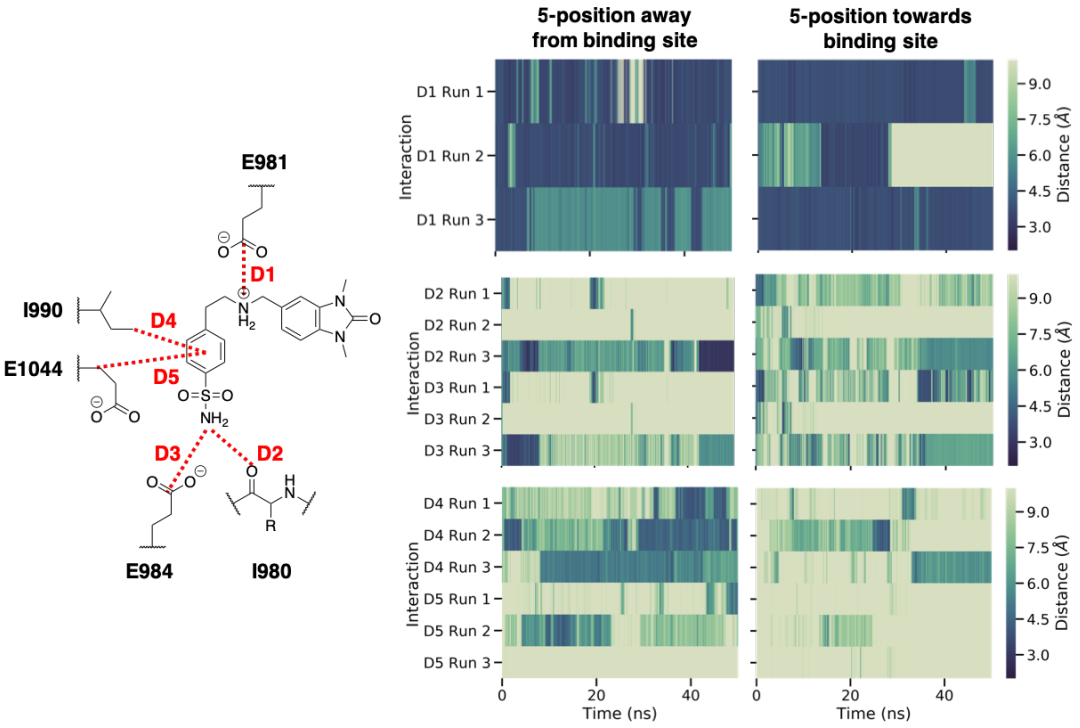


Figure 3.9: MD studies on compound **25** docked to the TRIM33β BRD. Compound **25** was docked to the TRIM33β BRD in two poses, which were subject to MD simulations, in triplicate. Five distances, D1 to D5, between the ligand and protein were analysed for both poses, and the distances have been represented as barcode plots.

forming more transient interactions. This group often remains solvent exposed while sampling several residues. Distances D2 and D3 in Figure 3.9 show that interactions of the sulfonamide amine with E984 and the backbone of I980, both beyond the ZA channel, are possible in both docked modes. The aryl linker is capable of forming hydrophobic interactions in the vicinity of E1044, near the top of the αC loop, or near I990, which is found above the binding pocket on the ZA loop. This is found to be more common in the simulations where the sulfonamide containing vector is pointing away from the binding site.

Studies on the urea-containing hits, 27 & 28

The urea containing HTS hits, compounds **27** & **28**, were docked to the three putative binding regions of the TRIM33β PHD/BRD cassette which are indicated in Figure 3.7: i) the interdomain cleft; ii) the PHD site (encompassing $K9Me_3$

and K4me$_0$ binding regions) and iii) the construct underside. Docking of both compounds identified one feasible pose for each compound in the interdomain cleft, 4 feasible poses for each compound in the construct underside and additional feasible poses at the K4Me$_0$ site. All poses are visualised within the appendix (Figure A.1 and Figure A.3).

MD studies of the **27** docked poses showed only two stable poses, Figure 3.10A, with the RMSD from the original pose compared in Figure 3.10B. The pose identified in the underside of the construct, where the hydrophobic rings are pointing into the site with the urea solvent exposed, is extremely stable during MD simulations. A docking pose on the K4Me$_0$ binding site was shown to stabilise, but after rotation of the ligand in the binding site. Two poses showed a stable RMSD, shown in Figure 3.10C. Both show that the urea can form polar interactions with either the C901 backbone or the E903 side chain. The rings are involved in a π-π interaction with F921, and the quinoline nitrogen interacts with the backbone of L899. The RMSD data of all docked poses are available in the appendix (Figure A.2).

MD simulations of **28** showed only two poses (Figure 3.11A) retained the ligand for the majority of the simulation, Figure 3.11B. Both poses failed to stabilise with a RMSD under 2 Å. The 'underside' pose was not accommodated hydrophobic pocket, unlike **27**. This is likely due to a larger 3D structure within **28**, which can be compared in Figure 3.5. Meanwhile, the poses identified at the K4Me0 binding site are stable after reorganisation, as seen with **27**, and formed π-πinteractions with F921. The RMSD data of all docked poses are available in the appendix, Figure A.4.

Studies of the amide-containing HTS hit: 26

Compound **26** was docked in a large region involving both the BRD, inter-domain cleft and W889 residue, as shown in Figure 3.7. This identified 5 putative poses of the ligand that all involved the inter-domain cleft, however, none involved interactions with N1039 or W889. These poses are shown in the appendix (Figure A.5). Through the same MD approach, Figure A.6, it was observed that none of the docked poses

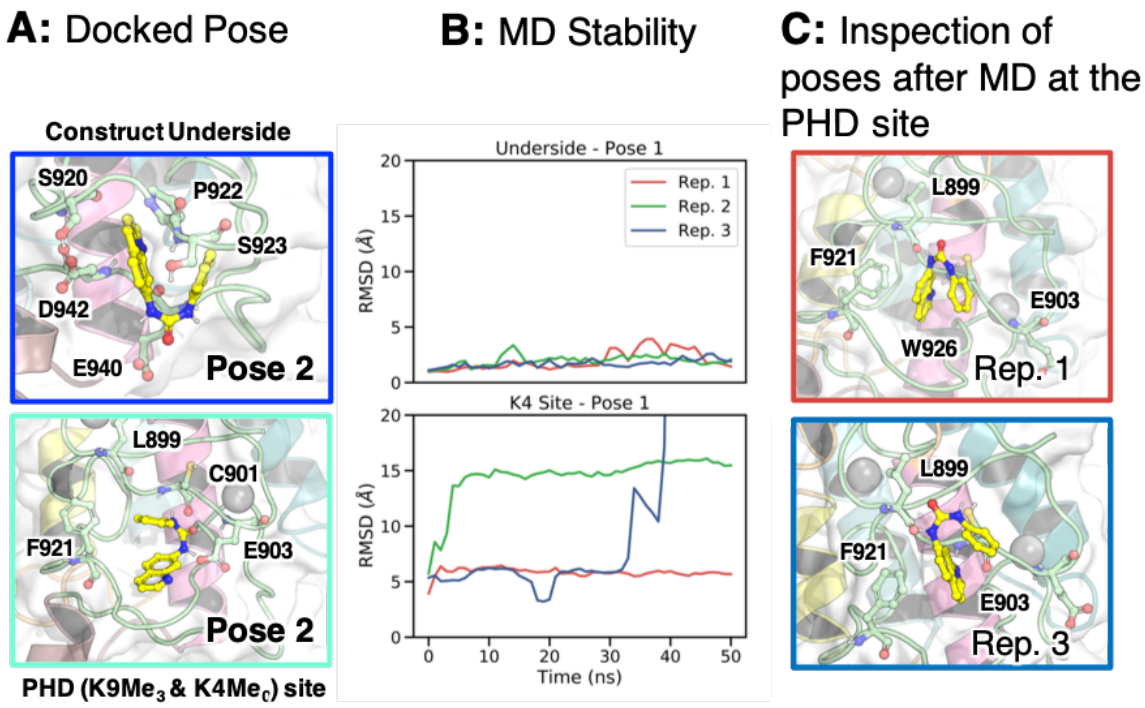


Figure 3.10: Selected docking and MD studies on compound **27** against the TRIM33β PHD/BRD cassette. (**A**) Two docked poses showed low RSMD values during MD simulations: Pose 2 in docking studies towards the underside of the construct (blue) and pose 2 identified in docking studies against the $K4Me_0$ binding site (cyan). (**B**) The RMSD values of the ligands over time. (**C**) Docked poses at the $K4Me_0$ site showed rotation during the simulation to give alternative stable poses.

were stable during simulation. The majority of simulations only stabilised with an average RMSD of over 10 Å, after some form of partial unbinding event.

3.4 Discussion

Crystallography is vital to determining the binding modes of ligands within protein structures, though unfortunately it was not possible to obtain any co-crystal structures of the HTS ligands against TRIM33β. Instead a series of modelling steps were needed to predict the ligand binding modes, where work focused on the typical, and previously uninhibited TRIM33β isoform of the PHD/BRD cassette.

The most plausible results are from the studies of **25**, where the binding site had already been experimentally determined, and binding mode indicated in a

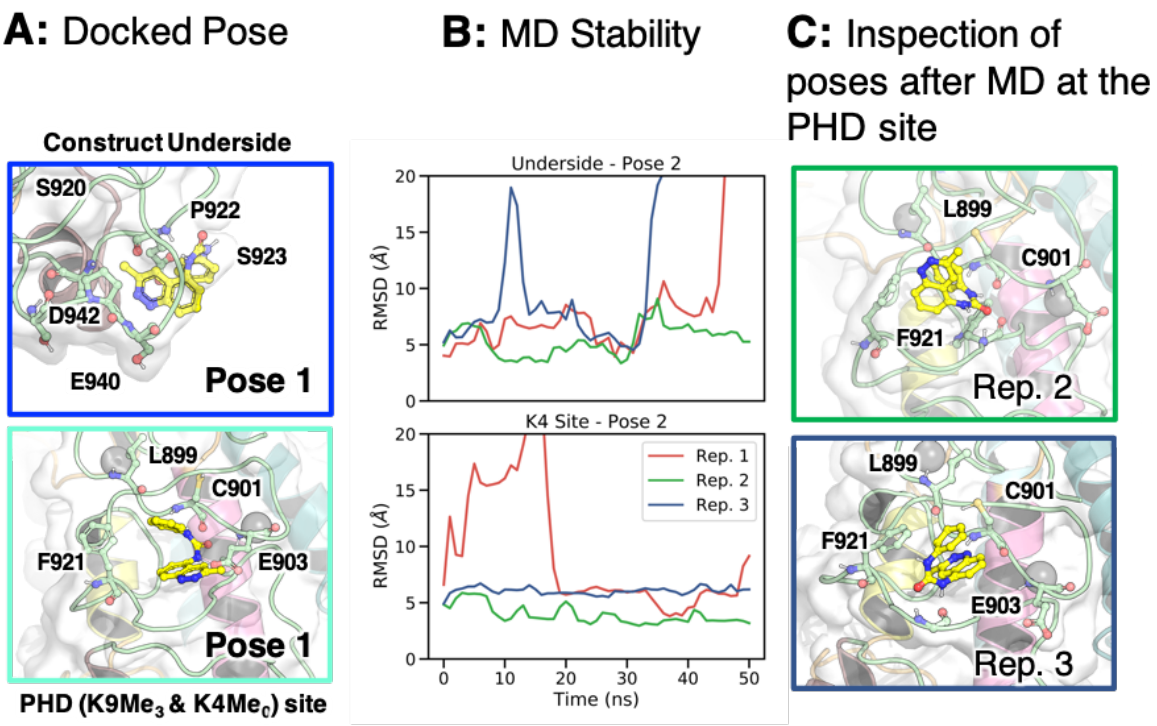


Figure 3.11: Selected docking and MD studies on compound **28** against the TRIM33β PHD/BRD cassette. (**A**) Two docked poses did not escape the their respective binding sites during simulation. Pose 1 in docking studies towards the underside of the construct (blue) and pose 1 identified in docking studies against the $K4Me_0$ binding site (cyan). (**B**) The ligand RMSD values during the simulations. (**C**) Snapshots of the MD simulations at 40 ns show rotation of the ligand from the initial pose.

similar protein.[53, 204] The benzimidazolone core docks in an analogous manner to IACS-9571 (**19**) in the TRIM24 BRD,[53] with the KAc mimic moiety interacting with the KAc recognising Asn residue. Residue E981, which is Ala in TRIM24, is key to both the ligand binding. This residue also offers a chance to obtain selectivity with the benzylic amine. There are few polar residues in the vicinity of the TRIM33β BRD site, which makes obtaining additional polar interactions difficult, though the model shows there is potential for the sulfonamide to interact with E984. One should also keep in mind that this could be an artefact of poor sulfonamide group parameters in the MD simulations, which have been shown previously to effect simulated observables in MD.[58]

The binding modes of the urea-containing compounds have remained more elusive, though MD simulations have shown that compound **27** could bind to the

underside of the construct with little fluctuation in pose. This finding is less exciting, given that the underside could not be biologically relevant in the context of the entire protein. MD studies of the $K4Me_0$ binding site showed that there is possibility for the urea-containing compounds to form surface interactions *via* H-bond interactions with backbone atoms, and π-π interactions with F921. Future mutant studies of F921 residue could determine whether the ligand is, at least partially, residing in this site. It could be possible that this low affinity ligand binds in multiple positions. This could be explored using the Binding Modes of Ligands Using Enhanced Sampling (BLUES) approach developed by Gill *et al.* and use an ensemble of poses to explain the binding events observed.[207] If so, this would be an exciting prospect as there have only been two previously developed PHD ligand series.[20, 208] There is also the danger of off-target effects, where a series of compounds that are analogous of **27** have been found to be weak CYP1B1 inhibitors.[209]

The studies failed to identify a binding mode of **26**, which was shown to be TRIM33 selective. This could be a matter of docking studies failing to locate an appropriate binding mode, or the MD approach failing to account for the ligand binding. Future work should look at adopting more sampling of the inter-domain cleft,[210] where the assay data and blind docking studies indicate is the region of ligand binding.

The molecular docking methods adopted in this chapter are limited by a lack of receptor flexibility, which fails to take into account the ligand and protein reorganisation that occurs upon binding known as the induced fit.[211] To enhance the results, molecular dynamics has been used to allow reorganisation of the protein and used the stability of the binding pose as a measure of docking success, as previously shown by Liu *et al.*[206] This approach differs from an alternative MD approach known as ensemble docking, where different binding conformations are generated through MD before docking to these various conformations.[211]. An alternative approach to improve docking studies is to allow residue movement during docking, known as Induced Fit Docking (IFD), though this suffers from poor reliability and larger computational costs. Future work could follow that of Clark

et al. who have taken this even further and combined IFD with metadynamics simulations to measure H-bond persistence and RMSD over time to give a more accurate IFD score.[210]

3.5 Conclusions

The work presented in this chapter looked at unsuccessful co-crystallisation, and subsequent *in silico* studies to identify potential binding modes for ligands found in a previous HTS. The works has highlighted the BRD binding site, the inter-domain cleft, the K4Me$_0$-binding region and the construct underside as potential binding sites. Through focusing on the previously uninhibited, active form of the TRIM33 cassette, TRIM33β, a series of docking and MD studies looked to identify potential ligand binding modes. From previous experimental data, it was known that compound **25** binds the BRD binding site of TRIM33β, and it has now been shown that the ligand could bind in two differing modes at the site. Binding of this compound is dependent on E981, and capable of forming transient polar interactions with E984 or hydrophobic interactions in the vicinity of the BC loop. The urea-containing compounds are capable of binding at either the K4Me$_0$ site or the construct underside, though further experimental studies are needed to validate this. The binding mode of **26** remains unresolved, with all docked poses failing to form stable interactions in MD. The work covered in the next two chapters looks at the optimisation of compound **25** as a TRIM33β BRD inhibitor, with a focus of retaining the E981 interaction.

3.6 Synthetic experimental details

3.6.1 General synthetic experimental

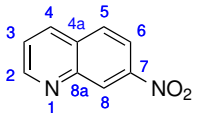
Proton and carbon nuclear magnetic resonance spectra were recorded on Bruker AVIII HD 400 (400 MHz), Bruker AVII 500 (500 MHz) or Bruker AVIII 500 (500 MHz) spectrometers using deuterated solvents as reference for the internal deuterium lock. ^{1}H and ^{13}C spectra were fully assigned using 2D NMR experiments

including COSY, HSQC, and HMBC. The chemical shift data for each signal are given as δH in units of parts per million (ppm) relative to residual chloroform, DMSO-d_6 or D_2O. The number of protons (n) for a given resonance signal is indicated by nH. Identical proton coupling constants (J) are averaged in each spectrum and reported in Hertz (Hz) to the nearest 0.1 Hz. Infrared spectra were obtained using a diamond ATR module. The spectra were recorded on a Bruker Tensor 27 spectrometer. Absorption maxima are reported in wavenumbers (cm^{-1}). Melting points were recorded using a Leica Galen III hot stage microscope and are uncorrected. The solvent(s) from which the sample was crystallised is given in parentheses. Low resolution mass spectra (LRMS) were recorded on a LCT Premier XE using electrospray ionisation (ES) from solutions of methanol. High resolution mass spectra (HRMS) were recorded on a Bruker microTOF. m/z values are reported in Daltons. Specific optical rotations were measured on a Schmidt Haensch UniPol 2000 polarimeter in cells with a path length of 1.0 dm at 589 nm and at 25 °C with concentrations (c) given in g/100 mL. Analytical thin layer chromatography (TLC) was carried out on normal phase Merk Kieselgek 60F254 precoated aluminium backed plates and visualised with a combination of the following: 254 nm UV lamp or thermal development after dipping in an aqueous solution of $KMnO_4$ (5% $KMnO_4$ in 1 M NaOH). Analytical HPLC was carried out on a PerkinElmer Flexar system with a Binary LC Pump and UV/VIS LC Detector. For determination of compound purity, a Dionex Acclaim 120 column (C18, 5 μm, 120 Å, 4.6 × 150 mm) was employed, using a 10-minute gradient of 95:5 H_2O/MeCN to 95:5 MeCN/H_2O (with or without TFA (0.1%)), a flow rate of 1 mL/min and detection at 254 nm. CH_2Cl_2 and DMF were dried through a column of active basic alumina according to Grubbs' procedure. All other solvents were used as supplied (analytical or HPLC grade) without further purification. Where appropriate, all non-aqueous reactions were performed in a flame-dried flask under an inert atmosphere of Ar. *In vacuo* refers to the removal of solvent on a Buchi under reduced pressure in a water bath at 40 °C. Brine refers to a saturated aqueous solution of sodium chloride. Petroleum

ether refers to the fraction boiling between 40–60 °C.

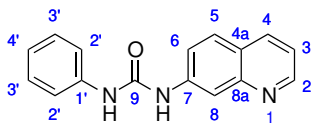
3.6.2 Synthetic procedures

7-Nitroquinoline (33)[212]



2,3-Dichloro-5,6-dicyano-1,4-benzoquinone (244 mg, 1.22 mmol, 2.0 eq.) was added to a solution of 7-nitro-1,2,3,4-tetrahydroquinoline (100 mg, 0.561 mmol, 1.0 eq.) in CH_2Cl_2 (5 mL) at rt and then stirred for 2 hours at 60 °C. The reaction mixture was cooled and the filtrate collected under vacuum before being added to 15 mL of warm EtOAc and black charcoal for 5 mins. This mixture was then filtered through Celite and concentrated *in vacuo*. Purification by silica chromatography (CH_2Cl_2:EtOAc, 100:0 to 0:100) afforded 7-nitroquinoline (**33**) as a light yellow solid (53 mg, 54 %): R_f 0.58 (EtOAc); m.p. 114-116 °C (EtOAc); ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.63 (s, 1H, H-8), 6.55 – 6.47 (m, 2H), 6.38 – 6.28 (m, 2H), 6.01 (d, J = 7.3 Hz, 1H, H-6); LRMS *m/z* (ESI$^+$) 175 ([M+H]$^+$, 100 %); HPLC Retention time 11.0 min, 97%. The spectroscopic data are in good agreement with the literature values. [212, 213]

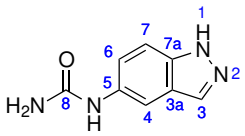
1-Phenyl-3-(quinolin-7-yl)urea (27)



Fe^0 was activated in aqueous HCl (3 M) for 3 hours before being filtered under vacuum and washed with 10 mL CH_2Cl_2. The mixture was then added to a mixture of 7-nitroquinoline (50 mg, 0.29 mmol, 1.0 eq.) in 5 mL glacial AcOH and stirred for 3 hours at rt. The mixture was then filtered through Celite under vacuum, and the

Celite washed with H_2O (20 mL), followed by 20 mL EtOAc. The aqueous fraction was washed with 2 × 20 mL EtOAc and the organic fractions combined, dried with anhydrous $MgSO_4$, filtered, and concentrated under vacuum to yield quinolin-7-amine which was used immediately. Phenyl isocyanate (28.2 µL, 0.258 mmol, 1.0 eq.) was added dropwise to a solution of quinolin-7-amine (37.2 mg, 0.258 mmol, 1.0 eq.) in acetone (2.5 mL) and stirred for 18 hours at rt. The precipitate was isolated by filtration under vacuum and washed with acetone (5 mL) and then CH_2Cl_2 to yield 1-phenyl-3-(quinolin-7-yl)urea (**27**) as a colourless solid (25.7 mg, 37 %): R_f 0.73 (EtOAc); m.p. 194-196 °C (CH_2Cl_2); $\tilde{\nu}_{max}$ (thin film)/cm^{-1} 3343 (N-H, w), 2980 (m), 2885 (w), 1693, 1673, 1626, 1597 (C=O, s), 1557, 1534, 1497, 1461, 1439, 1393, 1359, 1314, 1296, 1250, 1227, 1177, 1154, 1118; ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.24 (s, 1H, N-H), 7.97 (s, 1H, N-H), 7.96 (dd, J 4.2, 1.7 Hz, 1H, H-2), 7.41-7.39 (m, 1H, H-4) 7.38 (d, J 2.1 Hz, 1H, H-8), 7.03 (d, J 8.8 Hz, 1H, H-5), 6.76 (dd, J 8.8, 2.1 Hz, 1H, H-6), 6.69 – 6.62 (m, 1H, H-2'), 6.51 (dd, J 8.1, 4.2 Hz, 1H, H-3), 6.50 – 6.42 (m, 2H, H-3'), 6.15 (tt, J 7.2, 1.1 Hz, 1H, H-4') ^{13}C NMR (126 MHz, DMSO-d_6) δ 152.3 (C-9), 150.6 (C-2), 148.6 (C-8a), 140.5 (C-7), 139.3 (C-1'), 135.3 (C-4), 128.7 (C-3'), 128.3 (C-5), 123.4 (C-4a), 121.9 (C-4a'), 119.8 (C-6), 119.3 (C-3), 118.2 (C-2'), 114.0 (C-8); HRMS m/z (ESI$^+$) [Found: 262.09847, $C_{16}H_{12}O_1N_3$ requires $[M+H]^+$ 262.09859]; LRMS m/z (ES$^+$) 264 ($[M+H]^+$, 100 %), 286 ($[M+Na]^+$, 21 %); HPLC Retention time 4.3 min, >99%.

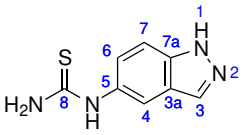
1-(1*H*-Indazol-5-yl)urea (30)



Sodium thiocyanate (97 mg, 1.5 mmol, 2.0 eq.) was added to a room temperature solution of 5-aminoazide (100 mg, 0.75 mmol, 1.0 eq.) in 10 mL of a 1:1 acetone:H_2O mixture. After stirring for 5 minutes, AcOH (33 µL, 5.25 mmol, 7 eq.) was added and the reaction stirred at rt for 20 hours. The solvents were removed under

vacuum and lyophilised. Purification of the resulting solid by silica chromatography (CH_2Cl_2:EtOAc, 12.5:77.5 to 0:100) afforded 1-(1*H*-indazol-5-yl)urea **30** as a violet solid (31.2 mg, 23%) after trituration with toluene: R_f 0.12 (EtOAc); m.p. 196-198 °C (EtOAc); $\tilde{\nu}_{max}$ (thin film)/cm^{-1} 3381, 3268, 3186, 2956 (w), 2925(w) , 2854(w), 1738, 1654, 1591, 1548, 1509, 1473, 1390, 1367, 1350, 1304, 1285, 1261, 1232, 1165, 1120, 1081, 1044, 1017, 948; ^{1}H NMR (500 MHz, DMSO-d_6) δ 12.83 (s, 1H, N-H), 8.44 (s, 1H, N-H), 7.93 (s, 1H, H-3), 7.87 – 7.83 (m, 1H, H-4), 7.39 (dd, *J* 8.8, 1.0 Hz, 1H, H-7), 7.23 (dd, *J* 8.8, 2.0 Hz, 1H, H-6), 5.75 (s, 2H, N-H); ^{13}C NMR (126 MHz, DMSO-d_6) δ 156.9 (C-8), 136.7 (C-5), 134.0 (C-7a), 133.4 (C-3), 123.5 (C-3a), 120.4 (C-6), 110.4 (C-7), 108.1 (C-4); LRMS *m/z* (ESI$^-$) 211 M+Cl; HRMS *m/z* (ESI$^-$) [Found: 175.06238, $C_8H_7O_1N_4$ requires [M-H]$^-$ 175.06253]; HPLC Retention time 4.2 min, 97%.

1-(1*H*-Indazol-5-yl)thiourea (36)[214]



1,1'-Thiocarbonyldiimidazole (214 mg, 1.2 mmol, 1.2 eq.) was added to a solution of 5-aminoindizole (133 mg, 1.0 mmol, 1.0 eq.) in CH_2Cl_2 (5 mL) and stirred at rt. After 3 hours, aqueous ammonia solution (20 mL of 25%) was added and the reaction mixture was left to stir for 16 hours. The solid was filtered under vacuum and washed with H_2O (5 mL) and MeOH (5 mL). This afforded 1-(1*H*-indazol-5-yl)thiourea (**36**) as a violet solid (48 mg, 25%): R_f 0.53 (EtOAc); m.p. 198-200 °C (CH_2Cl_2) [lit[214] 192-193 °C]; ^{1}H NMR (500 MHz, DMSO-d_6) δ 13.04 (s, 1H, N-H), 9.61 (s, 1H, N-H), 8.04 (s, 1H, H-3), 7.68 (s, 1H, H-4), 7.50 (d, *J* 8.8 Hz, 1H, H-7), 7.23 (d, *J* 8.8 Hz, 1H, H-6); LRMS *m/z* (ESI$^-$) 191 M$^-$; HPLC Retention time 4.6 min, 96%. The spectroscopic data are in good agreement with the literature values. [214]

The great tragedy of Science – the slaying of a beautiful hypothesis by an ugly fact.

— T. H, Huxley, *English Biologist*

4

Optimisation of a benzimidazolone ligand for TRIM33β BRD inhibition

Contents

This chapter looked at improving the affinity of the benzimidazolone ligand studied in the previous chapter. Investigations started on the amine-containing vector and then focussed additional vectors off the benzimidazolone core. This required a series of computational methods to build a binding model that fits

experimental data. The final model was then used to identify the highest affinity binders known for the TRIM33β BRD.

4.1 Introduction

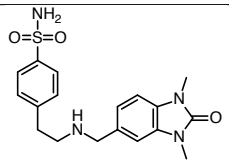
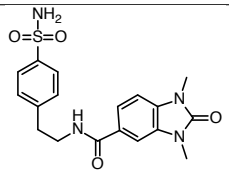
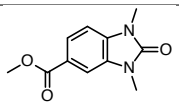
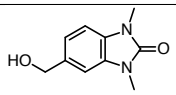
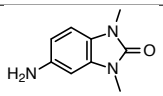
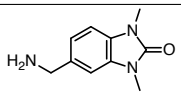
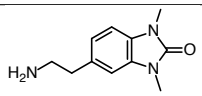
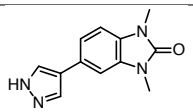
The studies performed in Chapter 3 identified two possible binding poses the benzimidazolone containing ligand, **25**, could adopt in the TRIM33β BRD site. Both poses held the KAc mimic in the vicinity of N1039, and also allowed the benzylic amine to interact with E981 (Figure 3.9). This was one of the few polar residues in the vicinity of the binding site and also a means of obtaining selectivity over TRIM24. Meanwhile, beyond the benzylic amine it was shown by MD that the sulfonamide group formed no long-lived interactions, but was capable of interacting with E984 when benzylic amide was projecting from the benzimidazolone core towards the binding site.

Further synthesis and testing of analogues of **25** by Dr L. See and A. Scorah, Table 4.1, identified that the benzylic amine adjacent to the core is key for both binding and gaining selectivity over the homologous TRIM24 BRD site. Conversion of the amine to an amide, seen in **37**, prevents ligand binding. This could be due to the N-H bond conjugated into an unfavourable orientation or reduced H-bond donating potential. Conversion to the ester, compound **38**, also prevents binding. Interestingly, the free alcohol, compound **39**, did not bind which indicated that the interaction with E981 is charge based. The optimal distance between the amine and benzimidazolone core was subsequently probed using compounds **40**, **41** and **42**. It was found that the benzylic amine lead to the best IC_{50}. This affinity was matched by compound **43**, however this had a lower ligand efficiency than **41**, where ligand efficiency is calculated through:

$$LE = \frac{1.37}{nHA} \times pIC_{50} \tag{4.1}$$

where nHA is the number of non-hydrogen atoms.[215] As this was the start of a ligand optimisation project, the ligand efficient compound **41** was taken forward for ligand optimisation.

Table 4.1: Analogues of **25**. These analogues were used to probe the importance of the β-amine group. IC_{50} values are reported from AlphaScreen™ assays with the H3(BRD) peptide (introduced in § 3.1), pKa values were predicted using the ACD labs webtool `http://ilab.psds.ac.uk/`.

Ligand	Structure	pKa	LE	IC_{50}(AlphaScreen™)/μM	
				TRIM24	TRIM33β
25		8.6	0.26	40.26±1.88	11.88±1.09
37		-	-	N.B	> 250
38		-	-	N.B	> 250
39		16	-	N.B	> 250
40		4.6	0.40	-	168.3±1.96
41		9.3	0.44	N.B	30.58±1.09
42		9.4	-	-	> 250
43		3.4	0.37	-	26.52±1.10

4.2 Addition of a second hydrogen bond donor to improve affinity

There are no published co-crystal structures of **41** bound to TRIM33, with the closest related structure placing the ligand in a tudor domain.[216] Hence, it was important to establish whether the ligand held a similar pose to those predicted for **25**. These predictions showed the vector at the 5-position pointed either towards (down) or away (up) from the binding site. Docking studies showed a similar outcome, with the

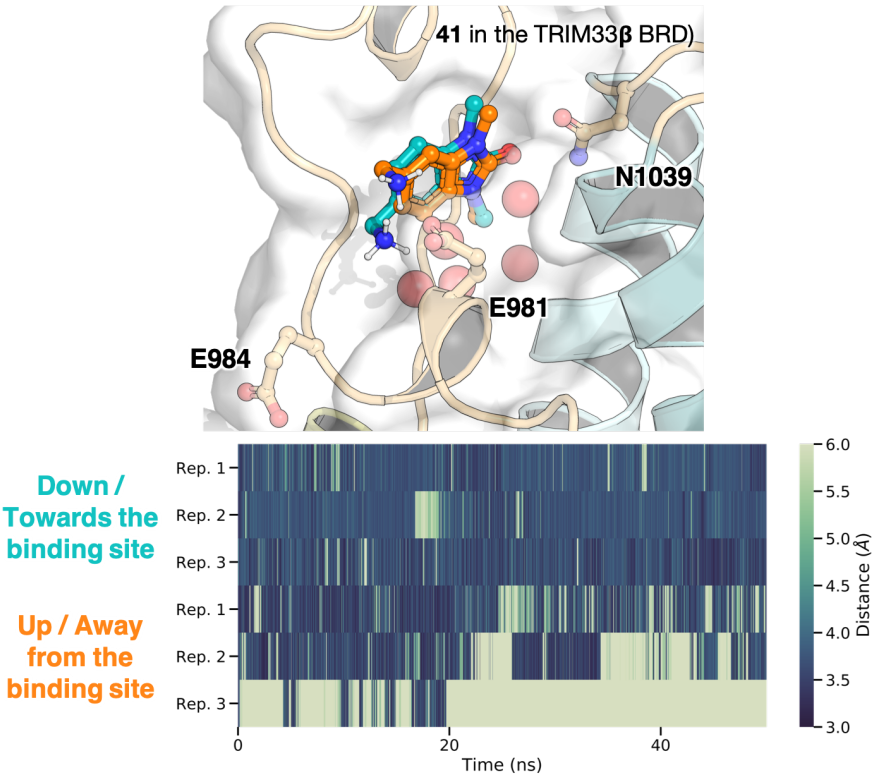


Figure 4.1: MD studies of **41**. Both docked poses of **41** were subject to triplicate MD for 50 ns. The distances of the charged, benzylic amine and E981 were measured as shown as barcode plots.

benzlyic amine able to adopt both binding poses shown in Figure 4.1. However, the ranking of docked poses suffers from the large errors associated with scoring functions, which can be in the range of 2-3 kcal/mol.[107] Hence, MD studies were performed on both docked poses of **41** using the approach used in Chapter 3 to elucidate the preferred pose. The distances of the charged amine to the E981 are shown as barcode plots in Figure 4.1. There was no noticeable difference in the interactions between the two poses, excluding the latter part of the 'away' pose of replica 3. Without a co-crystal structure it was difficult to design potential optimisations and so the use of more rigorous free energy calculation methods were investigated.

4.2.1 Free energy calculations to probe the optimal benzylic amine interaction with E981

Absolute binding free energy calculations are rigorous MD based methods to calculate the $\Delta G_{binding}$ of ligands.[217] These had been previously shown by Aldeghi *et al.* to be able to rank docked poses accurately, greater than MM-PBSA calculations but at a larger computational cost.[109] The calculations here also required optimisation of the Boresch restraints, that hold the ligand within the binding site during the calculations. After restraints to N1039 proved too unstable for simulations, a set of stable restraints were identified that held the benzimidalazone to the ZA wall (*via* the backbone atoms of V896). With the calculations stable and optimised, they were able to calculate the $\Delta G_{binding}$ of both ligand poses.

The calculated free energies are shown in Table 4.2, where a breakdown of the energetic contributions can be found in Appendix Table B.1. The results for the **41** poses implied that both poses are degenerate, concurring with the MD result that both poses retain a polar interaction with E981. Hence, ligand optimisation could proceed factoring the ligand pointing either towards or away from the binding site.

Table 4.2: Summary ABFE calculations performed on **41**, **45** and **58** with GROMACS 2016.3. ΔG values are reported as the mean of three independent repeats in kcal/mol, and the errors reported as 95% confidence intervals. The intervals were calculated by multiplying the standard error by the relevant t-statistic. The * indicates that the simulation was performed with only one replica, and the error is the standard deviation of the mBAR boot-strap error estimate. **45** is discussed in § 4.2.2 and **58** is discussed in § 4.4.1.

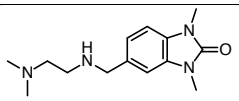
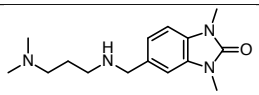
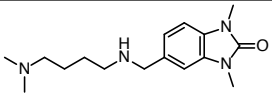
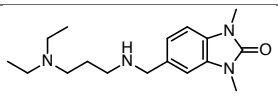
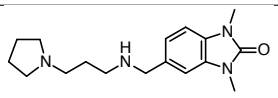
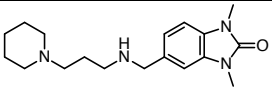
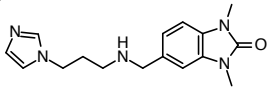
Ligand	Pose	ΔG_{calc} (kcal/mol)
41	up	-5.67 ± 0.90
41	down	-5.94 ± 1.28
45	up	-5.90 ± 3.00
45	down	-8.85 ± 1.23
58	-	-4.36 ± 0.14 *

4.2.2 MD to probe polar groups interactions with E984

It was hypothesised that additional polar interactions could increase both the affinity and selectivity of the ligand. Surrounding the binding pocket, there were

only a few polar residues or exposed backbone atoms. The most attractive polar residue beyond the binding site was E984, which sat at the other end of the ZA channel from the binding site. From the MD simulations of **25** in Chapter 3, it was shown that the sulfonamide group was able to form transient interactions with E984 (Figure 3.9). Therefore, a series of ligands were designed that reintroduced a second amine group to the benzylic amine in **41**. It was predicted that this extension of the amine containing vector would lead to favourable interactions with E984, which would also lead to the displacement of water molecules in the ZA channel and obtain additional affinity. The TRIM24 BRD inhibitor, IACS-9571 (**16**) adapted a similar approach and contains an extended amine chain which is shown to interact with the analogous D928 within the crystal structure (PDB ID: 4YAC).[53]

Table 4.3: Compounds to probe a second amine group. IC_{50} values are reported from AlphaScreen™ assays with the H3(BRD) peptide. K_d values are calculated using ITC (performed by L. See). The pKa values of the amine furthest from the core (pKa(1)) and the benzylic amine (pKa(2)) were predicted using ACD labs.

Ligand	Structure	IC_{50} (µM)	K_D (µM)	pKa(1)	pKa(2)	L.E
44		15.24±1.07		9.0	5.8	0.35
45		5.72 ± 0.65	18.6 ± 4.60	9.7	7.6	0.36
46		2.73 ± 1.14	18.3 ± 2.71	10.1	8.6	0.36
47		2.02 ± 1.06	11.3 ± 2.55	10.2	7.8	0.36
48		9.29 ± 1.47		10.1	7.8	0.31
49		9.24 ± 1.06		9.8	7.7	0.30
50		Unknown		6.6	9.7	-

A series of reductive animations were performed on the carbonyl precursor of

the core, allowing installation of vectors containing two amine groups (synthesis performed by Dr L See and A. Scorah). The compounds synthesised along with their assay data are listed in Table 4.3. The table also shows pKa predictions of both amine groups, from ACD labs. Ligands **44**, **45** and **46** installed small tertiary amines on varying linker lengths. The pKa predictions indicated the protonation state of the benzylic amine was dependent on the linker used. These also showed a longer linker was favoured. The size of the additional amine was varied in compounds **47**, **48** and **49**. These showed that the size of the amine had little impact on the ligand affinity. The imidazole group was installed in **50**, though the protonation state of the amine was dubious and studies were performed on both the protonated (p) and non-protonated forms of the imidazole. Overall, the pKa calculations predicted the major species of most compounds would be dual-charged, both on the benzylic and terminal amine groups. However, subsequent docking and MD simulations showed that the dual-charged species was not stable in the binding site (data not shown). Orthogonal docking studies using MOE were unable to produce poses that placed the KAc mimic near N1039, an issue which was only resolved when the benzylic amine was modelled as uncharged. Hence, modelling of these compounds placed a single charge on the terminal amine and left the benzylic amine, with a predicted pKa near 7, uncharged.

All compounds showed that the addition of a second amine improved the IC_{50} against TRIM33β, with all di-amine compounds, except **50**, showing increased binding affinity relative to **41**. Some of these compounds were able to undergo subsequent ITC studies by Dr L. See to determine K_d values.[204] The increase in affinity was only 10-fold, when comparing **46** to **41** with the assay data forming a plateau at low micromolar IC_{50} values. This showed a decrease in ligand efficiency for a small addition to the original compound, which ranged from 0.30 to 0.36 for the compounds in Table 4.3.

To confirm that the additional amine led to interactions with E984, docking and MD studies were performed on the compounds in Table 4.3. Where docking allowed, poses with the amine tail pointing towards (down) and away from (up) the

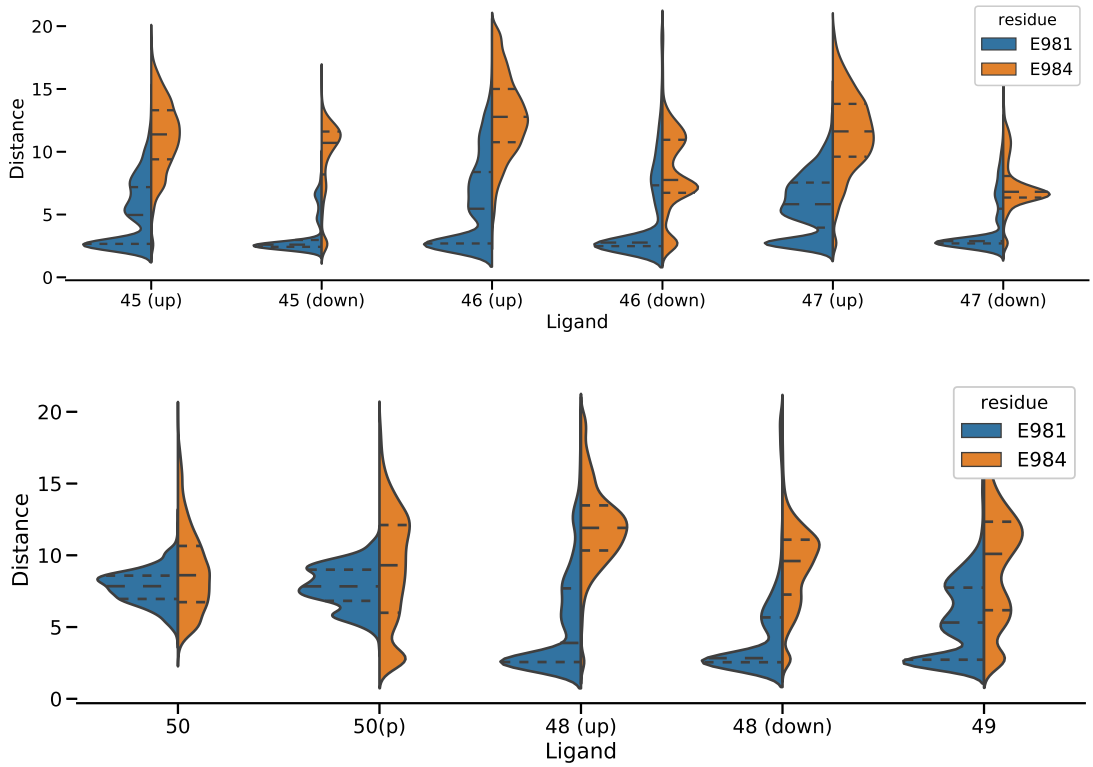


Figure 4.2: Distances of the terminal amine group to E981 and E984. Distances were extracted from triplicate MD simulations, and given in Å. Dashed lines mark the median and quartile values.

binding site were both studied. The plots in Figure 4.2 were surprising, initially showing that the charged, terminal amine had a preference for interacting with E981 over E984 (holding a distance of <4 Å). The only exception was **50**, when modelled with the imidazole group protonated (**50**-(p)). The plots also showed that when the ligand is pointing down/towards the binding site the terminal amine group is more associated with E984. This indicated that there could now have been a preference for the vector to point towards the binding site.

To determine whether the dual-amine containing vector had an energetic preference for pointing towards or away from the binding site, ABFE calculations were used again. The ABFE studies, which are computationally very expensive, focused on **45**. This compound was found to be a good representation of the compound developed and one of the most ligand efficient. The model used for

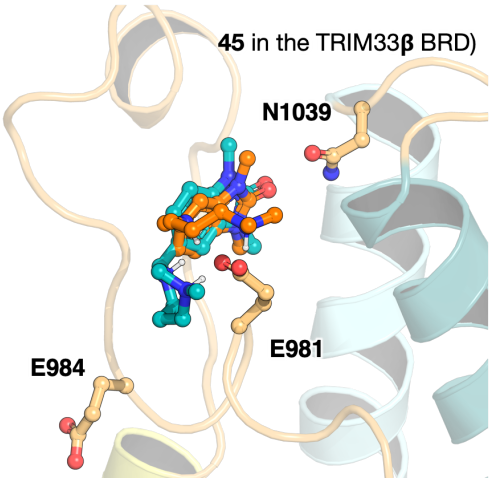


Figure 4.3: Equilibrated poses of **45** in TRIM33β. Docking of the ligand to TRIM33β showed that the di-amine containing vector can point either towards (down - cyan) or away from (up - orange) the binding site.

both poses held the terminal amine in the vicinity of E981, as this was the major observation from Figure 4.2 and shown in Figure 4.3.

The ABFE calculations required a restraint to keep the core near the ZA wall, but used an angle restraint to ensure the amine tail was kept away from the core. The restraint distances and angles were taken from average of a short MD simulation, which also equilibrated the system. Initially, the ABFE model used RESP charges for the ligand, based on the single conformational fit protocol set used by Aldeghi *et al.*[58] However, observations of the equilibrating simulations showed that the terminal amines rarely formed H-bonding interactions with either E981 or E984. This observation was also seen when simulations were extended to 50 ns, in triplicate, to mirror the data in Figure 4.2. To verify the charges assigned by the RESP model, the models were visualised within VMD and compared to AM1-BCC charges. The comparison shows that the method used for RESP parametrisation fails to assign charges that were intuitive to the models, shown in the charges assigned to nitrogen atoms in Figure 4.4, and so AM1-BCC charges were used for all future calculations. This resolved the issue as AM1-BCC charge assignments are not conformationally dependent. To use RESP charges in future studies, a multi-conformational fit would be needed.

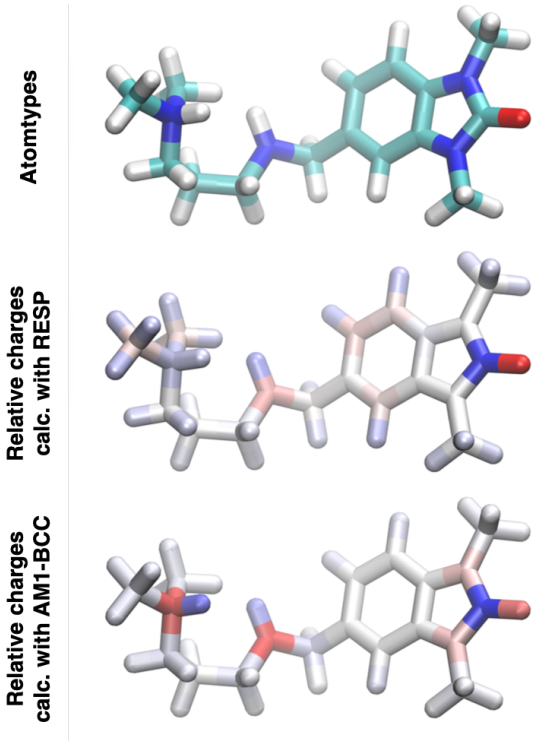


Figure 4.4: Charge models for ligand **45**. Relative charges assigned to atoms using both the RESP and AM1-BCC charge models. Figure produced in VMD v.1.9.1.

The $\Delta G_{binding}$ of both binding modes was calculated through further ABFE calculations, Table 4.2. The predicted $\Delta G_{binding}$ for pointing towards the binding site was higher than that determined by ITC; where $\Delta G_{calc} = -8.85 \pm 1.23$ kcal/mol and $\Delta G_{ITC} = -6.45 \pm 0.31$ kcal/mol. However, this increased magnitude has been shown in other ABFE studies.[58] It should be noted that the 95% confidence intervals for the 'up' and 'down' poses, Table 4.2, allow for the possibility that both poses are degenerate. Given that the overlap in errors is small, and the large error of the 'up' pose resulted largely from only one repeat of the ligand simulations it was decided the interpret the result as a preference for the 'down' pose. However, this result still failed to explain why the ligand does not form interactions with E984.

4.3 GCMC studies on the ZA channel water molecules

The protein model for TRIM33β used for docking had, so far, excluded the crystallographic waters within the ZA channel, assuming them to be non-vital

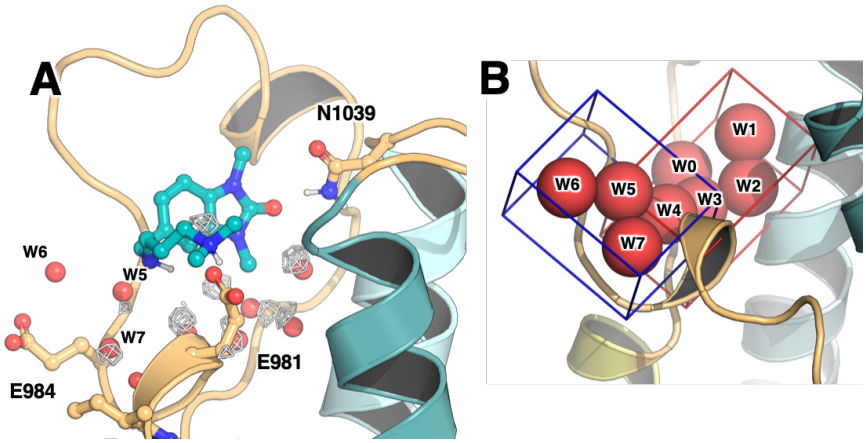


Figure 4.5: Water densities for the **45**:TRIM33 complexed during MD simulations and the GCMC search regions. (**A**) The regions of high water molecule occupancy during a simulation of **45** are shown as a grey mesh, with spheres showing the location of crystallographic water molecules. (**B**) GCMC calculations to identify the water molecule binding free energies were performed within two regions: the binding site (red) and the ZA channel (blue).

to binding. This small network of waters, W5-7, sit outside the binding pocket in a channel within the ZA loop (Figure 4.5). Given the results in the previous section, where there were few interactions of the ζ-amine to residues within the ZA channel or beyond the channel (E984), the water molecules were subject to a more rigorous analysis. When looking at regions of high water occupancy during simulations of **45** bound to TRIM33β, all three ZA water molecules are seen to be present along with the binding site waters, shown by the white mesh in Figure 4.5A. These water molecules were not positioned before the MD studies, so spontaneously occupied these regions during the simulation. Hence, this indicated that the ZA channel water molecules are stable and are not being displaced by the ligand for the majority of the simulation.

To probe whether the lack of displacement was due to the waters being tightly bound, in a similar way exhibited by the binding site waters, water binding affinity calculations were employed. This used another protocol adapted from Aldeghi *et al.*, employing a series of GCMC simulations for free energy calculations.[28] From these simulations a grand canonical integration (GCI) calculation can be used to extract the binding energies of the water molecules. This calculation protocol was

Table 4.4: GCMC Calculations on binding site and ZA channel waters. GCMC Calculations were performed on both the binding site and ZA regions, shown in Figure 4.5B calculating the free binding energies of the water molecules within those regions.

Water	Affinity (kcal mol^{-1})
W1	-6.35 ± 0.09
W2	-2.52 ± 0.14
W3	-0.96 ± 0.10
W4	-2.03 ± 0.11
W0	-1.89 ± 0.13
W5	$+1.71 \pm 0.09$
W6	$+0.82 \pm 0.14$
W7	-3.33 ± 0.10

used twice, on both the binding site and the ZA channel regions, specified by red and blue boxes in Table 4.4, respectively. The results, shown in Table 4.4, showed that W0-W5 in the binding site had a negative binding energy and thus contain an energetic penalty for displacement. This is expected, as similar values were reported for most BRD binding sites and these waters are rarely displaced.[28] However, the results for the ZA channel waters are surprising. These indicated that W5 and W6 could be displaced with an energetic gain, but W7 is tightly bound. This water molecule is predicted to H-bond to the carbonyl backbone of I980, backbone nitrogen of E984, W5 and W6.

This result has an impact on the dual-E981 binding mode of the charged terminal amine seen in **45**, and analogues. Figure 4.6 shows an overlay of two MD snapshots, where the terminal amine is interacting with E981 and then with E984. Overlaying the crystallographic ZA water molecules with these snapshots, it becomes clearer why the amine tail compounds form a 'pincer' interaction to E981. For the terminal amine to interact with either E984 or the backbone carbonyl of I980, would require displacement of W7 to allow for the rotation of E984 and subsequent H-bonding. The snapshot shows E984 placed in the position of W7, meaning the conformation change has expelled W7 and so incurring an energetic penalty. This explains why only a small proportion of simulations formed H-bonds with E984 and why the ZA channel water molecules are observed during MD simulations. Future docking

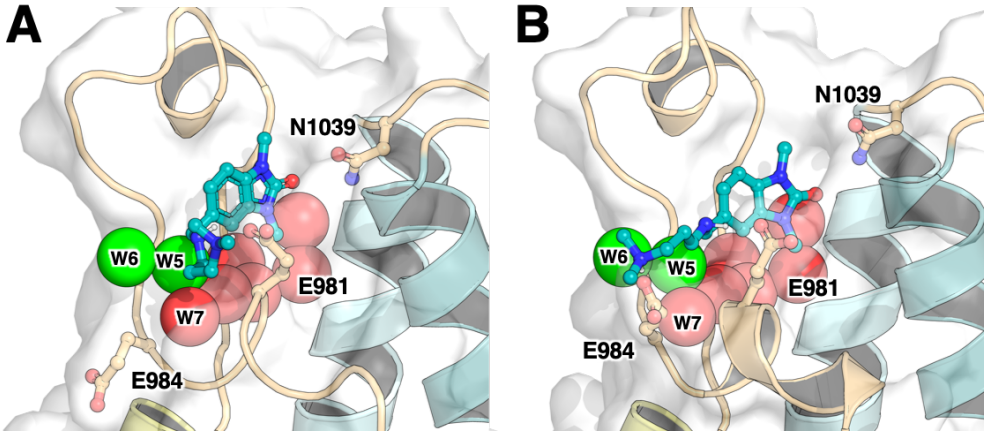


Figure 4.6: Overlay of crystallographic waters with MD snapshots of **45**. MD snapshots shown show interactions of the terminal amine group with both E981 and E984 in MD. these structures were aligned to the crystal structure of TRIM33β and allowing the crystallographic water positions to be shown.

models should account for W7 being present, and high energy barrier to an E984 conformational change. This could also explain why the protonated model of **50** showed interactions with E984, as this has a linker between the amines long enough to not require E984 to rotate towards the ZA channel.

4.4 Ligands to exploit the displaceable ZA channel waters

The identification of W7 as a non-displaceable water molecule, while W5 and W6 remain displaceable, offers an opportunity to increase both affinity and selectivity of ligands for the TRIM33β BRD. It was hypothesised that displacing water molecules W5 and W6 with a bulky and lipophilic group, which are adjacent to P985 in the ZA channel, could increase ligand affinity from loss of water and increased lipophilic contact. To probe this hypothesis, two approaches were taken. Firstly, to incorporate these modifications into the **45** ligand. Secondly, given the difficulties seen with **45**, instead using the original ligand efficient fragment **41** as a starting point for vector optimisation.

4.4.1 Incorporating modifications into 45

To gain additional affinity from the 5-position vector it was important to maintain and enhance the pincer interaction with E981. Bulk pointing towards the ZA channel would prevent the terminal amine group from interacting with E984 and ensure this. A new docking model of the protein was constructed that included only W7 in the ZA channel, which allowed the size of hydrophobic substituent to be assessed. There was a non-exhaustive search of methyl and phenyl substitutions, investigating both enantiomers, using molecular docking and molecular dynamics. This search identified a suitable phenyl substitution adjacent to the benzylic amine, **51**. This ligand was found to maintain hydrophobic bulk in the ZA channel, adjacent to P985, especially in the *(S)*-enantiomer during MD simulations. As shown in Figure 4.7, two out of three simulations showed a low RMSD and showed the interactions of the terminal amine and are well retained with E981. Importantly, the new added phenyl ring resides against P985 and displaced the water molecules W5 and W6.

Synthesis and testing of compound **51** identified a 2-fold increase in binding affinity over the non-substituted compound **45** (synthesis and testing by A. Scorah). Further enantioselective synthesis by A. Scorah validated that the *(S)*-enantiomer had a slightly higher affinity, shown in Table 4.5, and identifying *(S)*-**51** as the second highest affinity binder of the TRIM33β BRD. Placement of the Phenyl group at the 5'-position, **52**, decreased the compound affinity, which is expected as it is further from P985 and closer to water molecule W7. At this stage of the project, further optimisation of the di-amide containing vector. This focused on increasing the pKa of the terminal amine and led to the synthesis and testing (by A. Scorah) of **53**, which is narrowly the highest affinity ligand known for the TRIM33β BRD.

4.4.2 Modifying 41 to target ZA waters

The work in § 4.2.2 showed that addition of a second amine was likely to lower the pKa of the benzylic amine removing a charged interaction with E981. It was hypothesised that taking the original small fragment **41** and adding bulk to target W6 and W7 would lead to a more ligand efficient binder while also maintaining

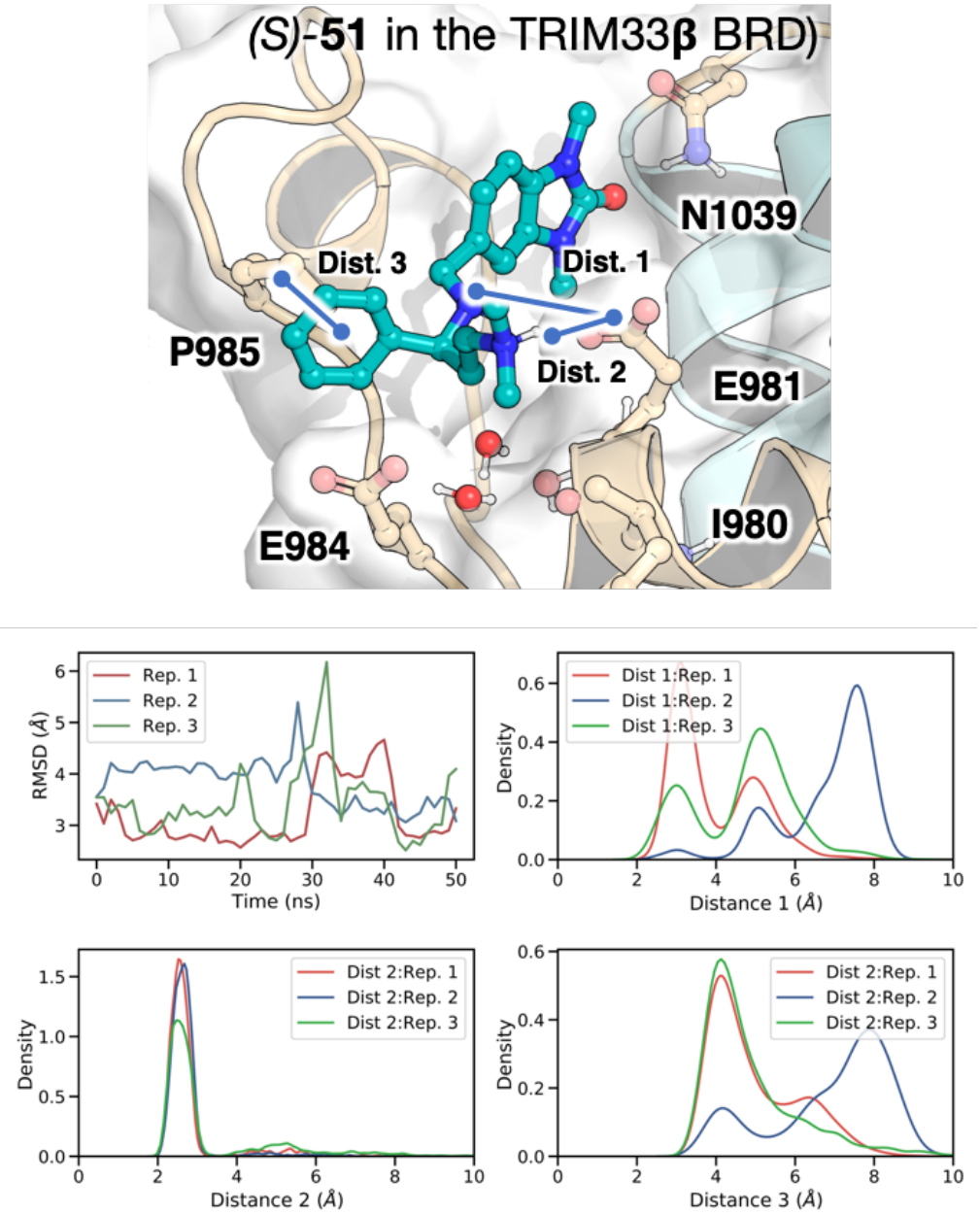
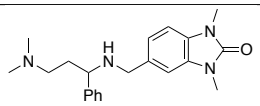
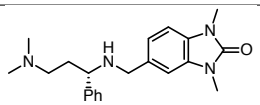
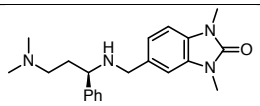
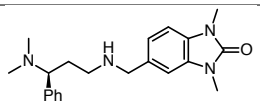
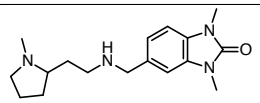


Figure 4.7: MD studies of *(S)*-**51**. (Top) Snapshot of a 50 ns MD simulation, showing the phenyl group resides adjacent to P985 which maintaining the amine interactions with E981. Water molecule W7 is retained in the simulation. The plots below show the RMSD and distances of ligands during three simulations.

Table 4.5: Compounds to probe whether the ZA water molecules could be displaced with phenyl groups. IC_{50} values are reported from AlphaScreen™ assays with the B2 peptide. K_d values are calculated using ITC (performed by L. See)

Compound	Structure	IC_{50} (µM)	K_D (µM)
51		4.83 ± 0.51	N.D
(S)-**51**		4.52 ± 0.60	7.33 ± 1.18
(E)-**51**		7.75 ± 0.60	N.D
52		16.90 ± 2.25	N.D
53		1.61 ± 0.14	7.28 ± 2.55

the charge on the benzylic nitrogen. In addition, it was also predicted that if the bulk contained heteroatoms, there was a possibility of incorporating interactions with W7 to increase affinity.

An equilibrated MD structure of **41** was taken where there was a clear interaction between E981 and the charged benzylic amine. The C-H bond pointing towards W5 was selected for substitution against the MOE growth library.[196] The *growth* tool in MOE was used to perform structure-based substitutions at this hydrogen position and identified 2738 compounds as viable substitutions. The majority of compounds consisted of 5-membered rings containing various combinations of heteroatom substitutions. The compounds designed were subject to a PAINS filter, a MOE score cutoff of -5.8, a minimum of 3 hydrogen bond donors and acceptors and finally a maximum number of 3 torsional bonds. These filters were used to encourage a selection of heterocycles that interact with W7 in the ZA channel and do not contain long flexible groups which reduce the accuracy of docking studies. These filters retained a large number of compounds that could not all be tested

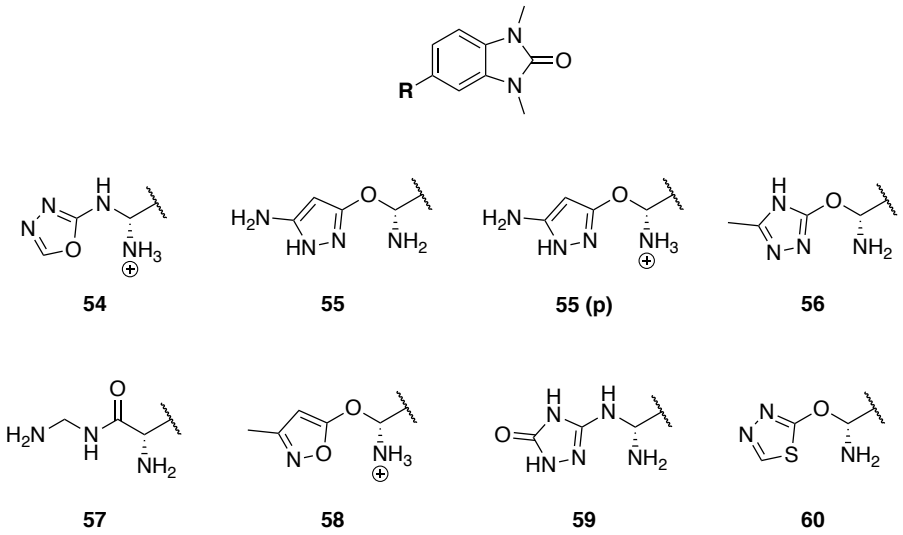


Figure 4.8: A diverse set of compounds designed to probe water molecule displacement. Compounds were built within MOE (v. 2018.2) and subject to a set of filters to select the ligands shown.

using molecular dynamics, so the *MaxMinPicker* tool with RDKit was used to pick 8 diverse compounds for further testing. These ligands are shown in Figure 4.8.

MD studies on these 8 compounds, **54** to **60**, all showed at least one simulation where the ligand pose remained stable, shown in Figure 4.9. **55** and **55 (p)**, which held the same heterocycle but differing charges on the nitrogen, both had two MD runs that held low RMSD poses. Further study of **55**, shown in Figure 4.10, shows that hydrogen bonds of the heterocycles to the backbone of P985 holds the 5 membered ring above W7 and remains in the ZA channel. pKa predictions of the benzylic amine model the major microspecies as deprotonated, which was used for ABFE calculations to assess the feasibility of this compound. The calculation, performed with three repeats and the GPU accelerated protocol, gave a ΔG of -3.76 ± 1.64 kcal/mol. This shows that the compound, at least in its uncharged form, is not an improvement on **41**. Interestingly, an orthogonal, single ABFE calculation using the GROMACS 2016.3 setup showed that **58** had no increase in affinity relative to **41**, reported in Table 4.2. Only a single simulation was used due to resources available at the time and provided an indication if further simulations would be useful. Overall, these calculations showed that these modifications to **41**, despite tolerated in the binding site, did not offer a large increase in binding

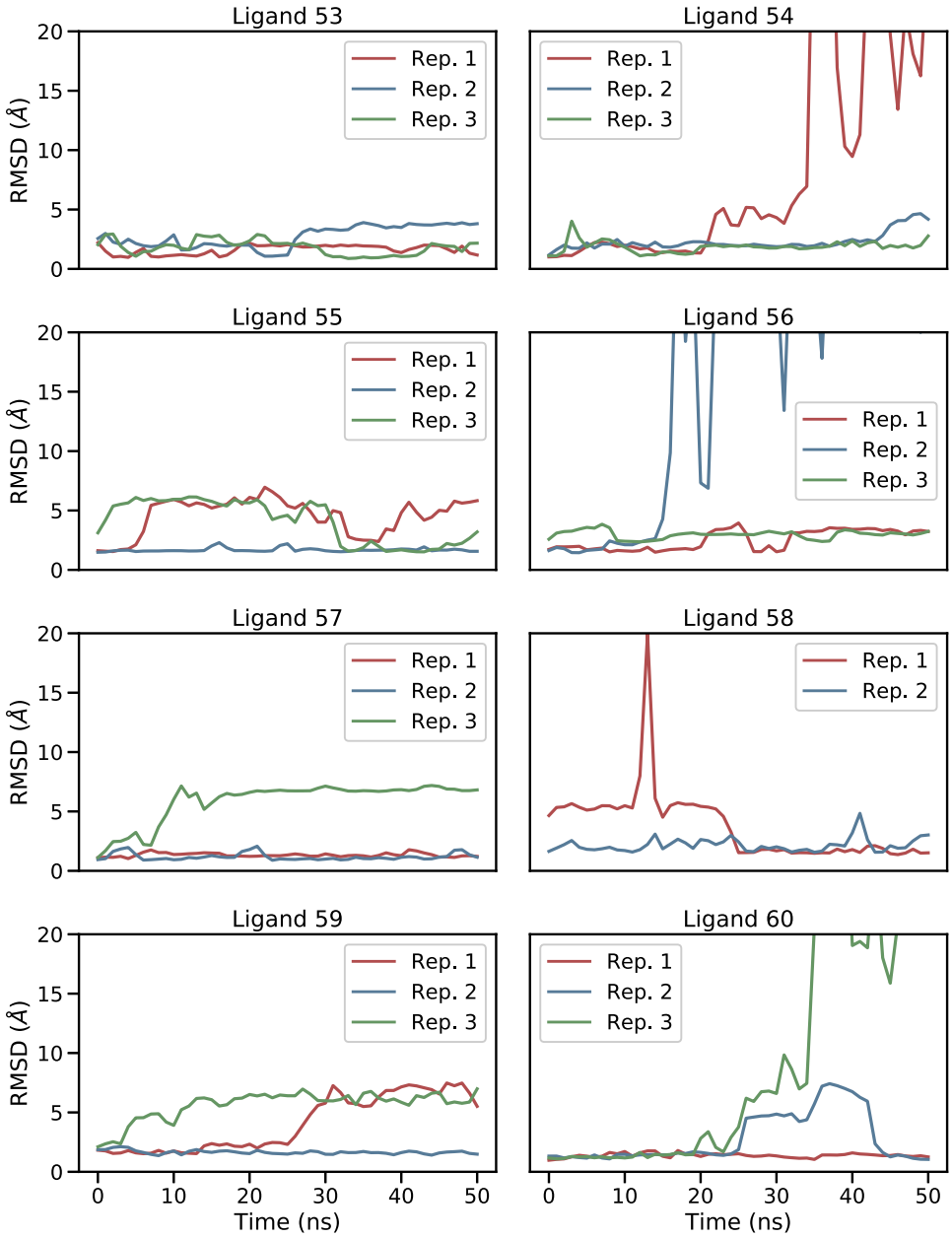


Figure 4.9: MD studies on diverse compounds selected from the MOE growth search. Situations were performed in triplicate with RMSD of the ligand shown with respect to time.

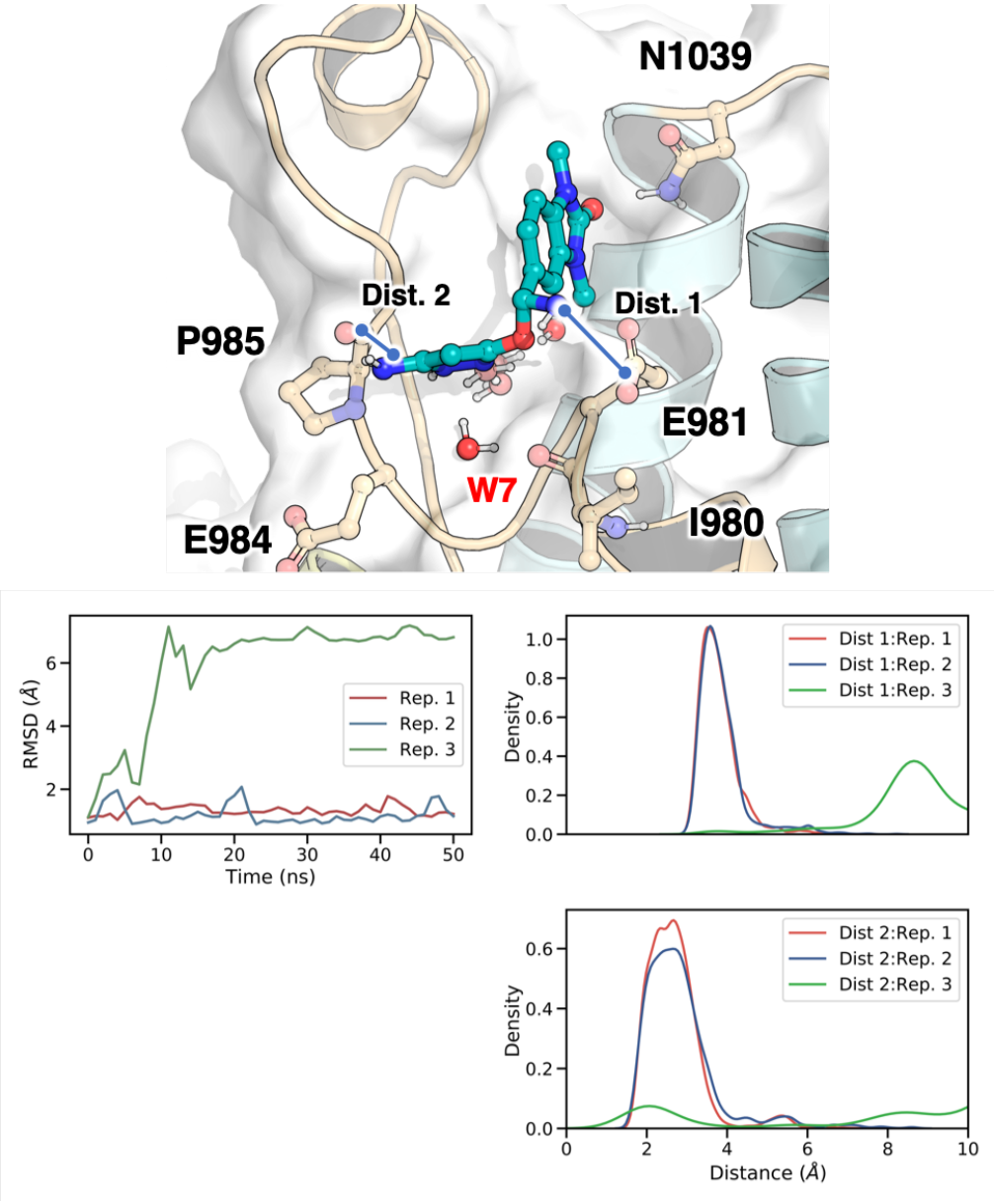


Figure 4.10: MD studies of **57**. The docked pose, shown at the top, underwent MD simulations of which two runs were stable. The stances of the two amine groups were measured (distances 1 and 2) and plotted below.

affinity even when charged. In addition, these ligand designs contained unstable aminal and hemiaminal linkers at the point of addition. If this region of chemical space is to be further explored, work on these compounds should look to replace these with bioisosteres to improve their stability in solution.

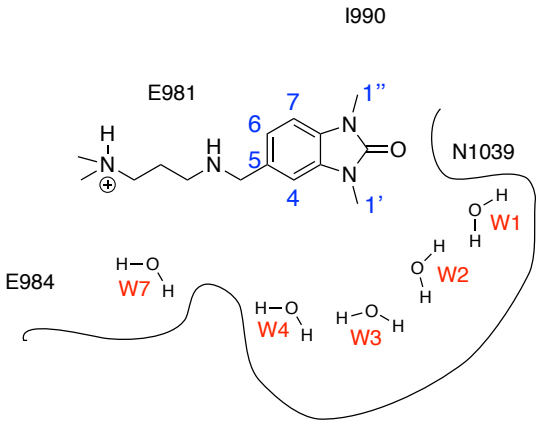


Figure 4.11: Cartoon representation summarising work in the first half of Chapter 4. Vector positions are labelled in blue, and waters are identified using red lettering.

4.5 Optimisation of additional vectors

Exploration of the 5-position vector had shown that expansion beyond a benzylic amine with an addition H-bond donor favoured the vector pointing towards the binding site. However, other positions on the benzimidalzone core were yet to be explored. Given the benzimidalazone core is common in several compounds, which do functionalise these positions,[218] it was hypothesised that additional affinity could be gained from building at several of these positions. A cartoon summary of the positions investigated are shown in Figure 4.11.

As previously discussed, there are few polar residues around the binding site. Hence, it was hypothesised that affinity could be obtained from improved hydrophobic interactions. An aliphatic carbon propensity search, which indicates regions of increased likelihood of finding a particular atom type, was performed around the region of the binding site. The results, shown in Figure 4.12, indicate an extended region, above the ligand and in the same plane as the ligand that could hold an aliphatic carbon atom, especially in the region around I990. Hence, this showed that there is are regions beyond positions 6,7 and 1” that could be used to add additional lipophilic groups. The work in the following subsections looked at how *in silico* methods aided exploring the SAR at the 5- and 1”-positions.

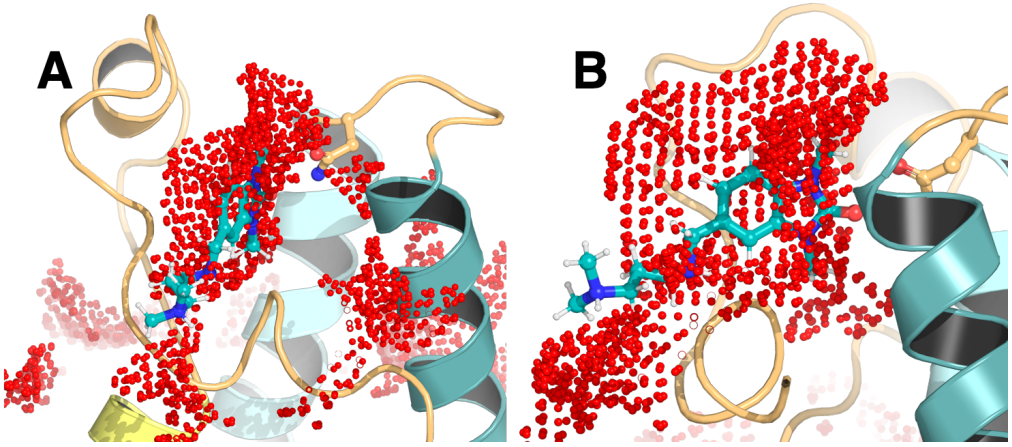


Figure 4.12: Aliphatic carbon propensity map for TRIM33β. Red dots highlight regions of higher than twice the average probability of finding an aliphatic carbon at that position. **(A)** and **(B)** show the binding site, containing **45** at different angles.

4.5.1 Building at the 1"-position

The largest region proposed by the propensity map was above the 1"-position. This corresponded to the region where the acetylated side chain of lysine leaves the H3 peptide into the binding site. This region also contained no crystallographic water molecules. The compounds in Figure 4.13 were designed to cover a range of hydrophobic bulk and polar groups to obtain additional polar interactions. Initially, **61** and **62** were subject to docking and MD simulations and both compounds were found to be accommodated in the *apo* binding site. Simulations of **61** showed that the hydroxyl group was capable of forming interactions with the carbonyl of N1039. Synthesis and testing of this compound, and the methylated version as a negative control (**63**), by Dr. L. See showed slightly weaker binding of **61** than the methyl equivalent. Meanwhile, binding was not observed in the negative control, compound **63**, showing the additional H-bond was important to binding. Data for experimental testing is shown in Table 4.6. MD simulations of **62** showed the ligand was not stable in the binding site. A 1,5 clash between a phenyl hydrogen and the hydrogen at the 7-position on the core prevented the phenyl ring sitting in the plane of the core and thus unable able to interact optimally with the hydrophobic ZA wall formed by I990. Hence smaller rings that do not suffer the 1,5-hydrogen clash were proposed.

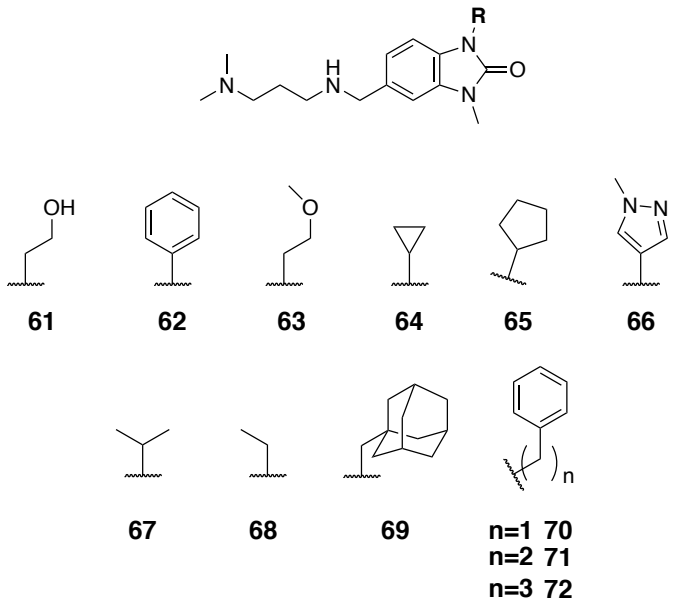


Figure 4.13: 1,3-analogues to probe interactions with I990. These compounds assume the amine tail should be pointing towards the pocket, on the same side of the core as the hydrophobic group at the 1-position.

Table 4.6: Affinities of compounds designed to probe the hydrophobic region above the binding pocket. IC_{50} values are reported from AlphaScreen™ assays with the B2 peptide.

Compound	IC_{50} (µM)
45	5.72 ± 0.65
61	20.61 ± 1.94
62	N.D
63	N.D
64	95.99 ± 24.94
65	60.19 ± 7.42
66	N.D
67	19.56 ± 1.64
68	14.74 ± 1.47
69	N.D
70	85.97 ± 1.15
71	59.72 ± 1.12
72	99.95 ± 1.49

Smaller 5- and 3-membered rings rings were docked and studied in MD within **64**, **65** and **66**. Simulations of **66** showed the ring could remain in the same plane at the core, as shown in Figure 4.14A. Meanwhile the 3- and 5- membered rings remained partially perpendicular to the core. Interestingly, **65** placed its ring inbetween the side chains of I990, in a style similar to that seen in a leucine zipper motif. Despite all poses being stable in MD, synthesis and testing by Dr L. See found that all compounds bound with a lower affinity than **45**, as shown in Table 4.6. This decrease was even seen with small substitutions of *i*-Pr (**67**) and an ethyl group (**68**) One hypothesis was that the hydrophobic bulk was not large enough to gain adequate affinity. To test this, the large adamantine group was modelled with **69**. This was accommodated by both docking and MD studies, however did not show any binding to the TRIM33β BRD.

Despite the range of bulk and polar groups, the compound affinity affinity was always reduced by replacing the methyl group at the 1"-position. ABFE calculations were performed with GROMACS 2019 to determine whether the simulations were capable of predicting this reduction in affinity. The calculations were performed using the GPU accelerated setup and using **65** as the representative ligand. The calculations correlated with experiment, showing a reduction in free energy relative to **45**, with a smaller value of -6.57 ± 0.47 kcal/mol (a graphical breakdown is available in the appendix). This result indicates that the simulations could justify the reduction in affinity. Visual comparison of simulations identified a change on the orientation of F1038 within the simulations where bulk was placed at the 1"-position. This residue forms a side-on π-π interaction with Y993 in the binding site (Figure 4.14A). From inspection, it appeared that F1038 was being pushed away from Y993 to accommodate the larger bulk at the 1"-position. This is best represented by the dihedral angles of F1038 shown in Figure 4.14B-D, for compounds **64**, **66** and **69**, respectively. These show that **45** displaces the dihedral from the crystal structure, though this is partially due to the system not being in the crystallographic state in simulation conditions. However, the bulky compounds

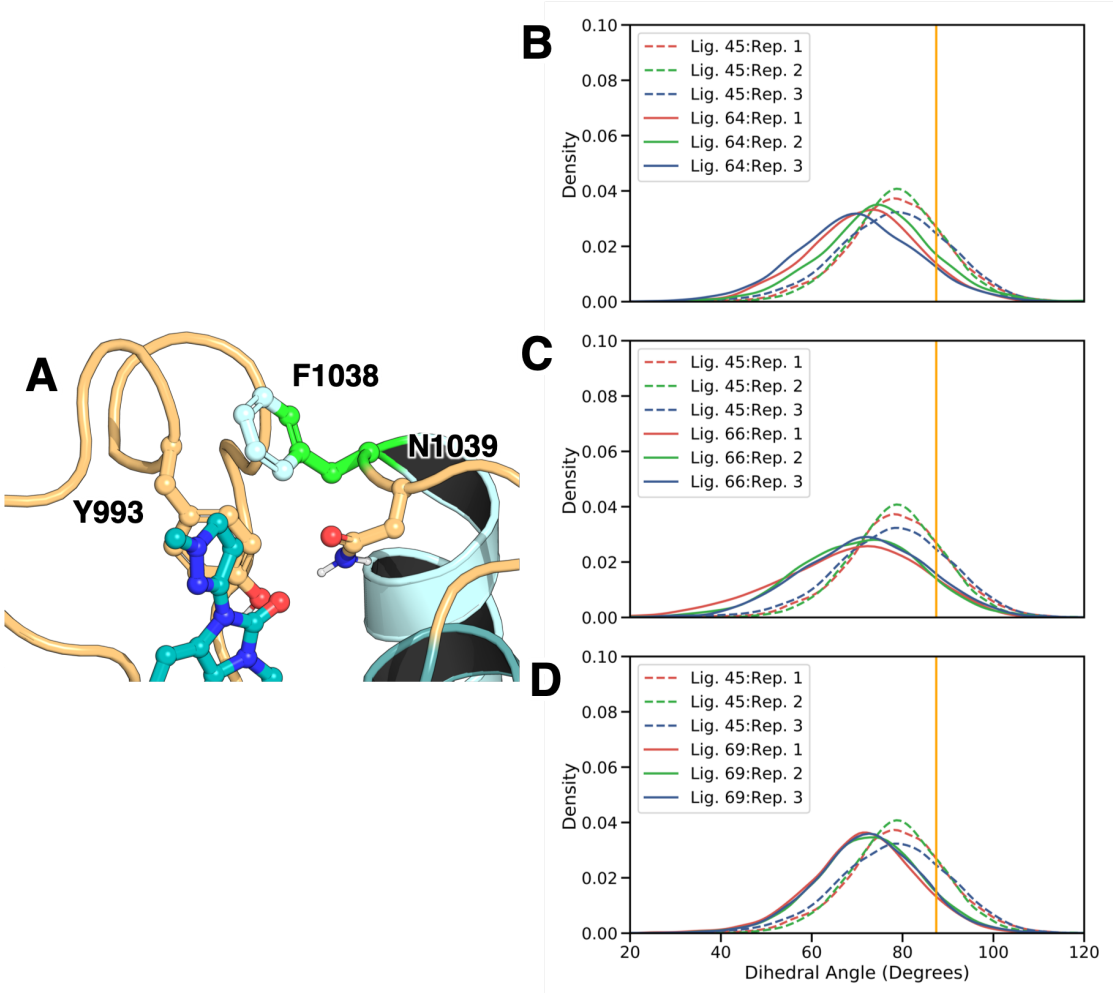


Figure 4.14: Dihedral angles in compounds designed to interact with I990. **(A)** Example of a compound designed to form hydrophobic contacts near I990 using a larger group, shown with **66**. **(B-D)** Plots showing the side chain dihedral angles within F1038 when various ligands are bound: **(B) 64**; **(C) 66**; **(D) 69**. The atoms used for the calculation are shown as green in **(A)**. Dashed lines are the same dihedral angle extracted from simulations of **45** and the orange line shows the dihedral angle in the *apo* crystal structure.

clearly lead to a larger rotation which would reduce the interaction between F1038 and Y993 and could partially explain the decrease in affinity of the compounds.

Given the position of F1038 above the binding site, it was hypothesised that a π-π interaction could be gained from an extended aromatic group at the 1”-position. To allow this several models were created that placed a phenyl group on a variable length alkyl linker (**70** (n=1), **71** (n=2) and **72** (n=3)). Unfortunately, these compounds again showed a reduction in binding. However, this could indicate that the linkers were not long enough to form a π-stacking interaction with F1038.

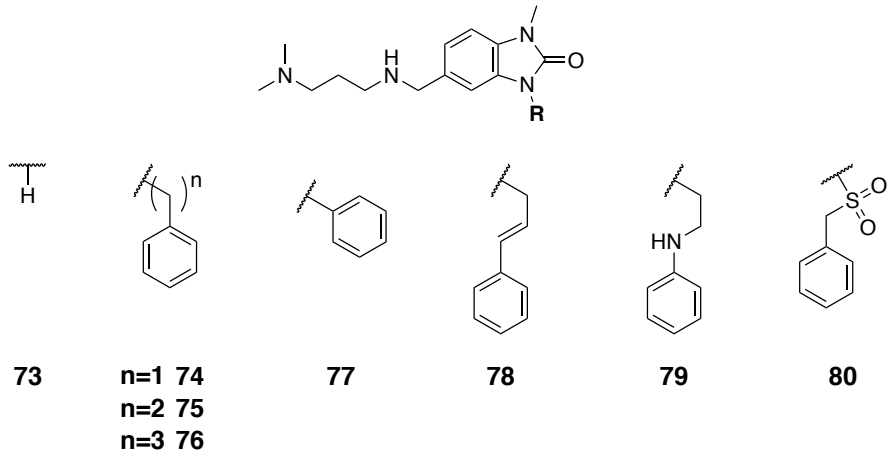


Figure 4.15: 1,5-analogues to probe interactions with I990. These compounds assume the amine tail should be pointing away from the pocket, on the same side of the core as the hydrophobic group at the 1-position.

During the ligand design process, mis-communication lead to a series of compounds synthesised and tested that placed the hydrophobic bulk on the 1'-position, rather than the 1"-position, **73** to **80** shown in Figure 4.15. These compounds showed no binding to the TRIM33β BRD and served as an important positive control that the amine tail binds when the vector points towards the binding site and the hydrophobic bulk can only be accommodated when pointing away from the binding site.

4.5.2 Building at the 6-position

There are several published ligands of the TRIM24 BRD, have a second vector leaving the benzimidazalone core at the 6-position. For example IACS-9571 (**16**), is shown in Figure 3.8B has an aromatic system that is placed in the hydrophobic region above the core. This uses a sulfonamide group to force a turn in the chain and place the group above the benzimidalzone core. This orientation places the ring above the other hydrophobic vector of IACS-9571 (**16**), forming a local 3D conformation.[53] It was hypothesised that given the homology between TRIM24 and TRIM33, a similar sulfonamide group could be used at the 6-position to increase affinity to TRIM33 through lipophilic interactions. However, simulations of **81**, which is a chimera of **45** and the additional vector from IACS-9571 (**16**) showed that the aromatic system at the 6-position does not form any fixed interactions. Instead

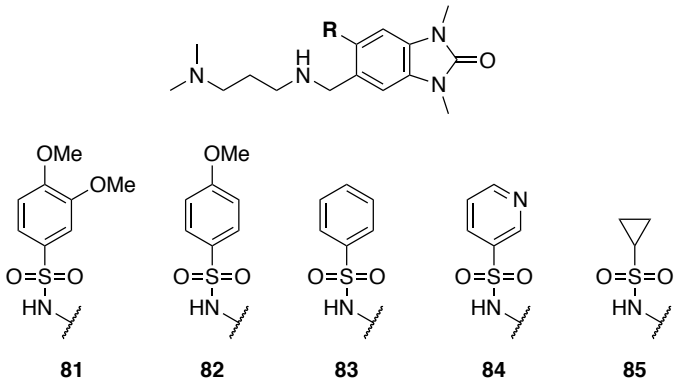


Figure 4.16: 5,6-analogues to probe interactions with I990.

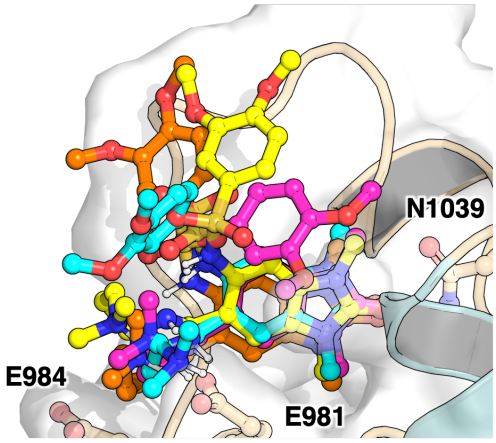


Figure 4.17: Snapshots of simulations of the TRIM33β and **81** complex. The amine containing vector retains its interactions with E981, but the aryl group does not form an stable interactions above the binding pocket.

81 samples the hydrophobic region above the binding site. This can be shown in the various MD snapshots shown in Figure 4.17, where the amine containing vector remains near E981 but the anisole ring moves location. Despite this observation, **81** and various analogues were synthesised and tested by J. Reynolds (**81** to **85**). Only one ligand, **82**, displayed binding to TRIM33β with an IC_{50} of 77.73 ± 13.9 µM. This decrease in affinity relative to **45** ruled these bulky substitutions at the 6-position out of future studies.

4.5.3 Searching commercial space for ligand analogues

Commercial chemical space is ever increasing in size, with the ZINC library containing over 230,000,000 compounds.[219] Rudling *et al.* have shown that

identifying, purchasing and testing commercial space analogues of HTS hits can increase the success of ligand optimisation projects.[220] A similar approach was used here, searching just the ChEMBL database, which contains bioactive compounds with drug like properties, for compounds with the benzimidazalone core.[221, 222] In May 2018, this produced 2,438 compounds with the same sub-structure. Malhotra *et al.* identified that when a fragment is elaborated 87% of compounds contained a substructure where the binding of a sub-structure was conserved.[223] This result can be used to assume the KAc mimic binding pose will be retained in the the binding site. Hence, a pharmacotherapy requirement was imposed during docking in MOE which required the KAc mimic of the benzimidazolone core to remain adjacent to N1039.

The top 100 docked docks were inspected within MOE, paying particular attention to the polar contacts formed between the ligand and protein. The seven ligands shown in Figure 4.18, **86** to **91**, were chosen based on likely interactions between the ligand and protein, with a subset of these ligands mimicking the H3 binding pose. These docks were subject to the same MD stability screening process discussed in the previous sections. Unfortunately, none of the ligands identified retained stable interactions with the protein, where RMSD plots can be found within the appendix (Figure B.4).

4.6 Discussion

The ligands designed and tested within this project were based off a series *in silico* models, which led to the discovery of the highest affinity ligands known for the TRIM33β BRD. This shows that medicinal chemistry projects can be benefit by computational design, especially in the absence of a co-crystal structure. This chapter builds off early experimental work that showed the benzylic amine on the benzimidalazone core is important for binding to the TRIM33β BRD. The ligand efficient version of the benzimidalazone, **41**, has also been seen as a binder of a tudor domain.[216] Hence, it was important that fragment optimisation retained the benzylic amine but looked to improve the selectivity of the compound as well as increasing the affinity. Testing these compounds against the tudor

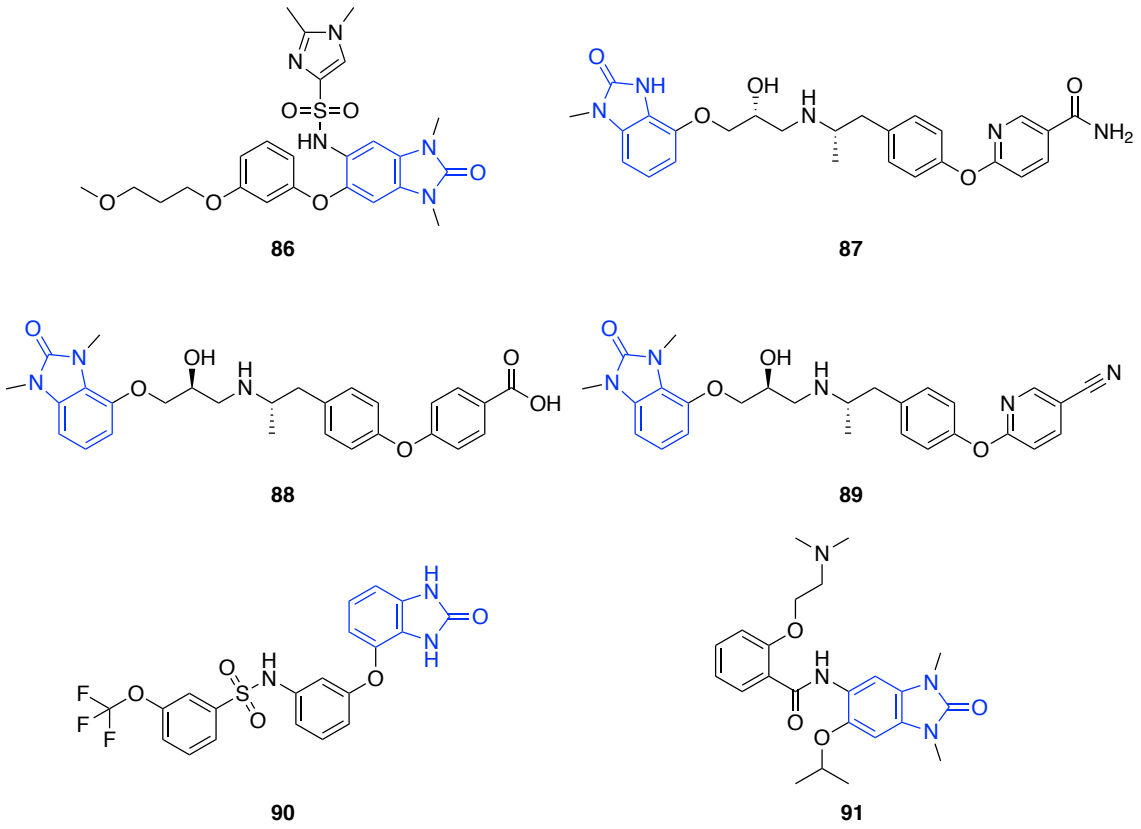


Figure 4.18: Analogues of **41** found within the ChEMBL library. The common core used for the search is highlighted in blue.

domain was not performed in this project, but will be required to verify that selectivity had been achieved.

Addition of a second amine to the benzylic amine, to form a range of dual-amine containing vectors, increased affinity of the compounds, but raised issues regarding building an explicit *in silico* model. The empirical-based pKa predictions by ACD Labs suggest that both amines on this vector would be protonated and charged. This prediction tool was shown by Liao *et al.* to be robust at calculating pKa, but this method performed poorly at tertiary amine centres.[224] This could explain why the the protonation model predicted was unstable in both docking and MD studies. This instability is not surprising as this would have placed two like charges close in space. The models used to elucidate the SAR trends instead used a single charge, placed on the terminal amine and this allowed for several important discoveries. These findings included that the binding mode placed the dual-amine vector towards

the binding site and that E981 was the residue the charged amine interacted with, not E984. Ideally, future work would look to replace this di-amine containing vector to reduce the number of torsions and ambiguous charge state. Meanwhile, this dual-amine containing ligand has since been used as the linking point in PROTAC ligands developed by A. Scorah and J. Reynolds. This route could allow for the TRIM33β BRD to be validated as a therapeutic target for future drug development.

The MD simulations showed that the E984 residue was capable of ligand interactions, despite the desolvation penalty and loss of the tightly bound W7 water molecule. This rarer binding event was further explored in a study by Dr L. See to develop potential covalent binders for the TRIM33β BRD, which would still contain the E981 residue required for ligand binding. A mutant protein, containing E984C, was designed after MD validation showing that this substitution did not impact protein stability. Ligands designed by Dr L. See replaced the terminal amine group with groups that could covalently bind to the E984C. Preliminary results against the mutant protein showed alkylations of the mutant protein not seen on the WT protein, however further studies are required to confirm the site of alkylation.[204] Covalent binding of the ligands could increase the chances of obtaining a crystal structure of the benzimidalazone series, which were not possible with the ligands discussed in Chapter 3.

Future optimisations should aim to explore the possibility of placing a hydrophobic group at the 7-positions on the benzimidalazone core, which was not addressed within this thesis. Placing bulk at the 1"-position came at the expense of the π-π interaction between F1038 above the binding site and Y993 in the binding site. However, the propensity maps show that hydrophobic bulk can also be placed at the 7-position and interact with the top of the ZA wall, in the vicinity of I990. Work can also focus on the 6-position, which was only briefly explored in this chapter with large complex groups.

This chapter also used ABFE calculations to rank predicted binding poses and determine the impact of select chemical elaborations. Through adaptation of existing protocols, these calculations can be run on much smaller, and affordable,

HPC resources. For example, one repeat took only 477 hours on a HPC node with 6 cores (at 2.6 GHz) with an NVIDIA 1080Ti GPU. These speed increases enabled the calculations to be run on local resources and could be used to test future designs for TRIM33β ligands. However, further efforts are needed to improve the error produced in these calculations, which were all above 1 kcal/mol. These are likely due to under-sampling, that can be overcome with extended simulations. Alternatively, recent work by Wan *et al.* have shown in their *TIES* approach that increasing the number of replicas can improve error estimates[225]. This increased replica approach could improve the confidence intervals and is more feasible given the improvements in GPU acceleration.

4.7 Conclusion

This chapter focused on understanding the binding of a benzimidalone core with an additional benzylic amine. Through ABFE calculations, it was determined that the benzylic amine had no energetic preference for pointing towards or away from the core. A second amine was built onto the benzylic amine, with an aim to incorporate an additional polar interaction between this group and E984. These compounds showed a small increase in affinity experimentally, though MD simulations showed the additional amine group preferred to form an interaction with E981 instead. Further ABFE calculations showed that the dual-amine containing vector now had a preference for pointing towards the binding site. When combined with the observation in GCMC GCI calculations that one of the ZA channel water molecules is not displaceable, a binding model that explains why the terminal amine group disfavours interacting E984 is possible. This model is now being used for PROTAC ligand design.

The observation that other ZA channel waters are displaceable to expose hydrophobic surfaces allowed for compounds that displaced water molecules to be designed and tested, with **51** becoming one of the highest affinity compounds for the TRIM33β BRD. Other ligands that were removed from the dual amine

containing series were designed, but failed to show a promising increase in affinity with ABFE calculations.

Finally, the hydrophobic region above the binding site was explored both *in silico* and experimentally. Despite docking and MD showing stable binding poses, these imposed strain on F1038 and hence reduced affinity both in ABFE calculations and in assay conditions. Future work on this series should look at building at the 7-position to increase contacts of the ligand to the ZA wall.

Humanity shares a common ancestry with all living things on Earth. We often share especially close intimacies with the microbial world. In fact, only a small percentage of the cells in the human body are human at all. Yet, the common biology and biochemistry that unites us also makes us susceptible to contracting and transmitting infectious disease.

— Brenda Wilmoth Lerner *Infectious Diseases: In Context*

5

Probing BRD inhibition in Parasites

Contents

This chapter describes a set of homology models generated for parasite bromodomains found in proteins from *S. mansoni* and *T. cruzi*. These models were then used in a series of studies to rationalise trends in observed SAR, and predict compounds for future study. Some of this work has been subsequently validated through a recently solved X-ray structure.

5.1 Introduction

Epigenetic machinery that is homologous to that found within humans, including bromodomains, has been identified within parasite genomes. It has been hypothesised that inhibition of this machinery could influence the lifecycles of parasitic

lifeforms, and therefore offer a potential avenue for treatment of tropical diseases.[86] Before this machinery can become a viable therapeutic target, chemical probes are required to determine the role of the components. The work in this chapter looks at inhibition of bromodomains found within parasite species: *Trypanasoma cruzi* and *Schistosoma mansoni*.

Trypanasoma cruzi and *Schistosoma mansoni* are two disease-causing parasites, associated with Chagas disease and schistosomiasis, respectively.[80, 226] The genomes of both species have been sequenced and subjected to domain characterisation.[83, 227] From this characterisation, putative bromodomains have been identified within both of these species. There have been 6 BRDs labelled in *T. cruci*: *Tc*BDF1-4,[90, 228] and a tandem bromodomain cassette in *Tc*BDF5(1/2).[229] Meanwhile in *S. mansoni* three homologous bromodomain containing proteins (BCPs) have been identified: *Sm*CBP1 and *Sm*CBP2, which are analogous to human CREBBP/p300, and a tandem BCP *Sm*BRD(1/2) (identified within the Conway group) which is analogous to *Hm*BRD3.[91] Full protein sequences can be found in the Appendix.

Through the efforts of Dr M. Schiedel, Dr A. Chan and C. Laurin, the following soluble parasite bromodomain constructs have been produced: *Tc*BDF2, *Tc*BDF3, *Tc*BDF5(1),*Sm*BRD3(1/2) and *Sm*BRD3(2). Sequences can be found in C.1. These constructs were screened against a series of human BRD ligands, given the literature precedent for binding.[89] The screening results are shown in shown in Figure 5.1. The data show a range of binding affinities, from non-binding to sub-micromolar affinity, as shown in Table 5.1. Additional ligands of *Tc*BDF3 were reported by Ramallo *et al.*[230] However, docking studies against the homology models described in this chapter failed to identify any poses stable in MD (work by A. Boczek) and testing by ITC failed to detect binding (work by C. Laurin).

These human bromodomain ligands are heavily optimised for human BRD binding, and for kinase binding in BI-2536 (**20**) and BI-6727 (**21**), hence large remodelling of these scaffolds will be needed if these are ever to be used for treatment of tropical diseases. However, these compounds offer an opportunity to validate

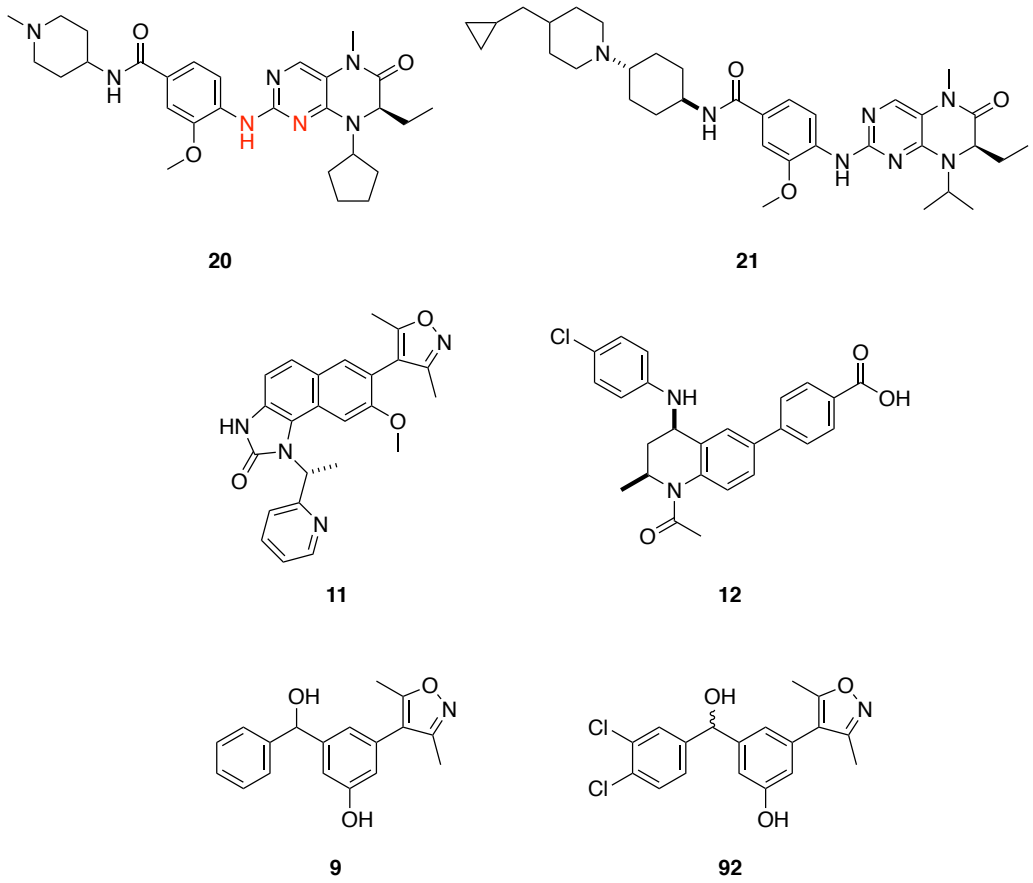


Figure 5.1: Bromodomain ligands that were identified as binders of one or more parasite BRD. The kinase hinge binding motif of **BI-2536** is is shown in red.

Table 5.1: Human bromodomain ligands that have been shown to bind parasite bromodomains. K_d values are given in µM and were calculated by ITC, except for data marked with a *, where K_d values were determined by 1H NMR ligand observed protein titration, values are given as a range. A dash line denotes no binding detected and blank entries denote the experiment was not performed. Data were collected by Dr M. Schiedel and C. Laurin.

Ligand	*Sm*BRD3(1/2)	*Sm*BRD3(2)	*Tc*BDF3	*Tc*BDF5(1)
BI-2536 (**20**) [59]	1.99 ± 0.28	3.81 ± 0.58	14.6 ± 3.0	-
BI-6727 (**21**) [64]			57.1 ± 4.9	-
I-BET151 (**11**) [41]	2.83 ± 0.28	6.16 ± 0.79	$263 - 313*$	-
I-BET726 (**12**) [44]	1.52 ± 0.52	1.85 ± 0.52	-	-
OXFBD02 (**9**) [40]	0.68 ± 0.17	-	$445 - 508*$	$73 - 1000*$
N28992-57 (**92**) [231]	-	17.8 ± 9.72	-	-

Table 5.2: Summary of parasitic bromodomain homology modelling. All proteins models were constructed as described in § 2.7). Only the gene range corresponding to the BRD was modelled, the entire gene can be found using the GenBank ID.

Protein	GenBank ID	Template Protein	Template Structure	Identity
*Tc*BDF2	XP_803713.1	*Tb*BRD2	5CZG (Chain A)	65%
*Tc*BDF3	XP_812334.1	*Tb*(427)	5C8G (Chain B)	65%
*Tc*BDF5(1)	XP_818752.1	*Ld*BPK(1)	5TCM (Chain B)	54%
*Sm*BRD3(2)	CCD76183.1	*Hm*BRD4(2)	4Z93 [233]	51%

parasite BRDs as a viable therapeutic target in disease treatment. The work presented in this chapter looked at developing models of the ligands binding the BRDs which has led to several key compounds that will aid the development of chemical probes to validate targeting parasite BRDs as a therapeutic target.

5.2 Homology models the parasite bromodomains

Prior to this work, there were no published crystal structures of any BRDs from either *S. mansoni* or *T. cruzi*. Hence, it was important to produce an accurate homology model from which to rationalise the observed SAR and guide future ligand design. The sequences of the individual bromodomains in *Sm*BRD3 and *Tc*BDF5 were identified using BLAST and used for templates searches.[232] Suitable templates for all the experimentally expressed constructs, with the exception of *Sm*BRD3(1), were identified using the online sequence search tool available on `www.rcsb.org`, where the highest ranking results are shown in Table 5.2. Templates were identified for all proteins, all of which had over 51% sequence identity. The *T. cruzi* proteins had templates within other parasite bromodomains that had been crystallised, while the sequence of *Sm*BRD3(2) best matched with *Hm*BRD4(2).[233] Interestingly, the templates for *Tc*BDF3 (PDB ID: 5C8G) and *Tc*BDF5(1) (PDB ID: 5TCM) were both co-crystallised with BI-2536.

Homology models were constructed as described in § 2.7, and are shown in Figure 5.2. All models appeared to be canonical bromodomains, where the KAc recognising Asn residue pointed into the binding site. The model of *Sm*BRD3(2)

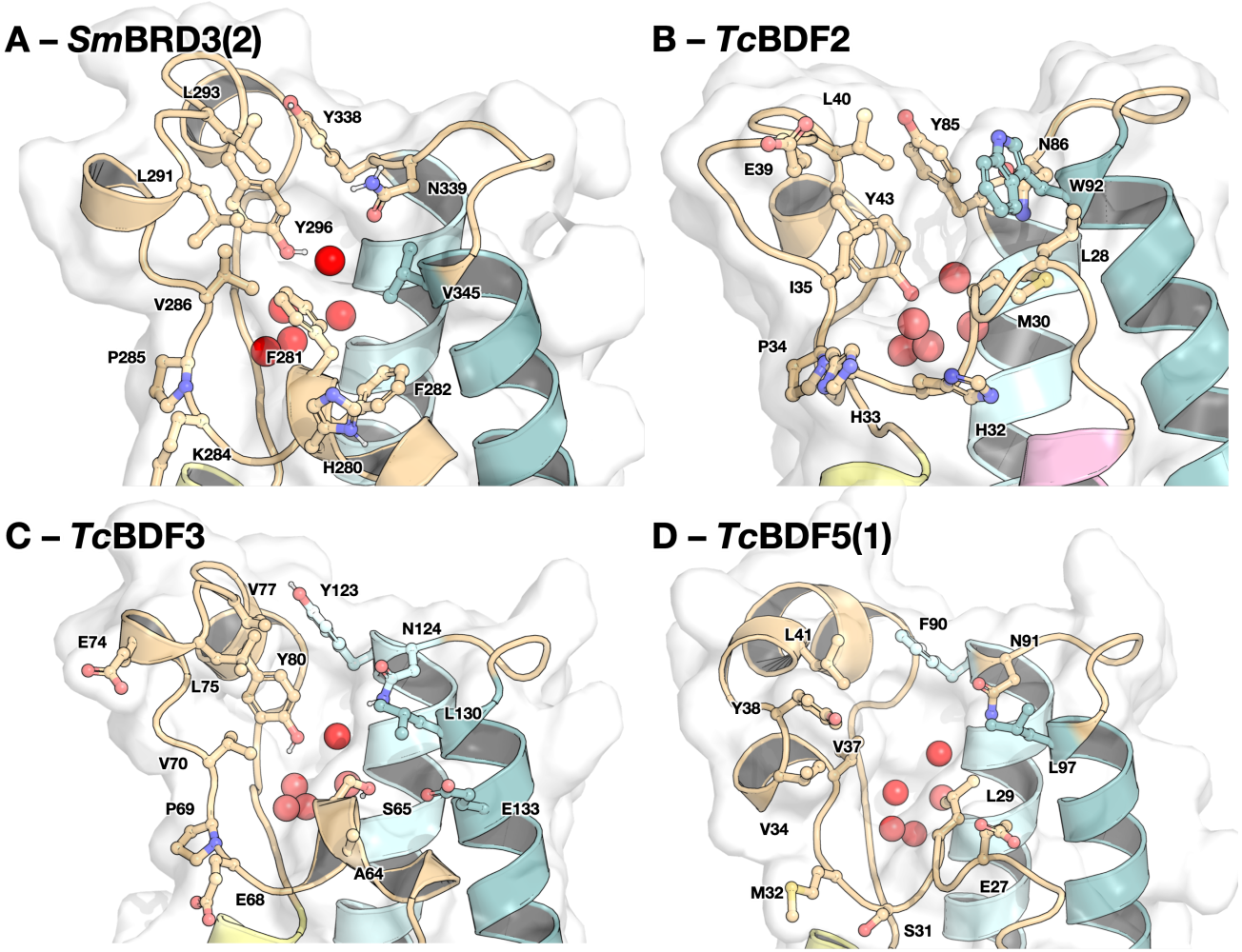


Figure 5.2: Binding sites of *S. mansoni* and *T. cruzi* homology models. (**A**) *Sm*BRD3(2). (**B**) *Tc*BDF2. (**C**) *Tc*BDF3. (**D**) *Tc*BDF5(1).

showed a hydrophobic pocket adjacent to the binding site consisting of H280, F281 and F282, similar to the WPF shelf seen in *Hm*BRD4(1).[40]. However, this region differed from the WPF shelf, shown previously in Figure 1.8, as the bulky F281 residue is one position closer to the α_C helix than W81 is within *Hm*BRD4(1), making the shelf slightly smaller. The models of the *T. cruzi* bromodomains did not have WPF shelf equivalents. Instead, *Tc*BDF2 and *Tc*BDF5(1) have a more horizontal hydrophobic surface formed from a Met and Leu residue, respectively. *Tc*BDF5(1) has a bulky Tyr (Y38) residue that partially-occludes the top of the ZA channel. In place of the WPF region, *Tc*BDF3 model had a very shallow surface relative the the WPF shelf, where the Trp (W81) and Pro (P82) residues are instead Ala (A64) and Ser (S65) residues. This opens up one side of the ZA channel to be more solvent exposed.

The stability of the homology models during equilibrium MD was assessed to ensure they remained folded during simulation. The data show no major unfolding of the protein during simulations, shown in Figure 5.3. The majority of movement within the protein within the ZA loop residues, shown by the RMSF plots in Figure 5.3 and the RMSD values of the non-ZA loop residues having a relatively lower RMSD. With stable homology models constructed, the molecular dynamics simulations were further studied to verify if the parasite bromodomains contained the typical binding site water network. A density analysis was performed over the *apo* simulations to identify regions of high water occupancy, shown in Figure 5.4. Overlaying these densities with the water molecules from the template proteins, it was observed that the water positions within the templates could be used to model the water positions for docking studies. Regions of high water density can also be seen in the ZA channels of the *T. cruzi* models, seen to the left of the images in Figure 5.4. These positions were noted for future studies. The homology models were prepared for docking studies using the MOE package, which allowed for water molecule orientation optimisation, with protonation states predicted at pH 7.4.

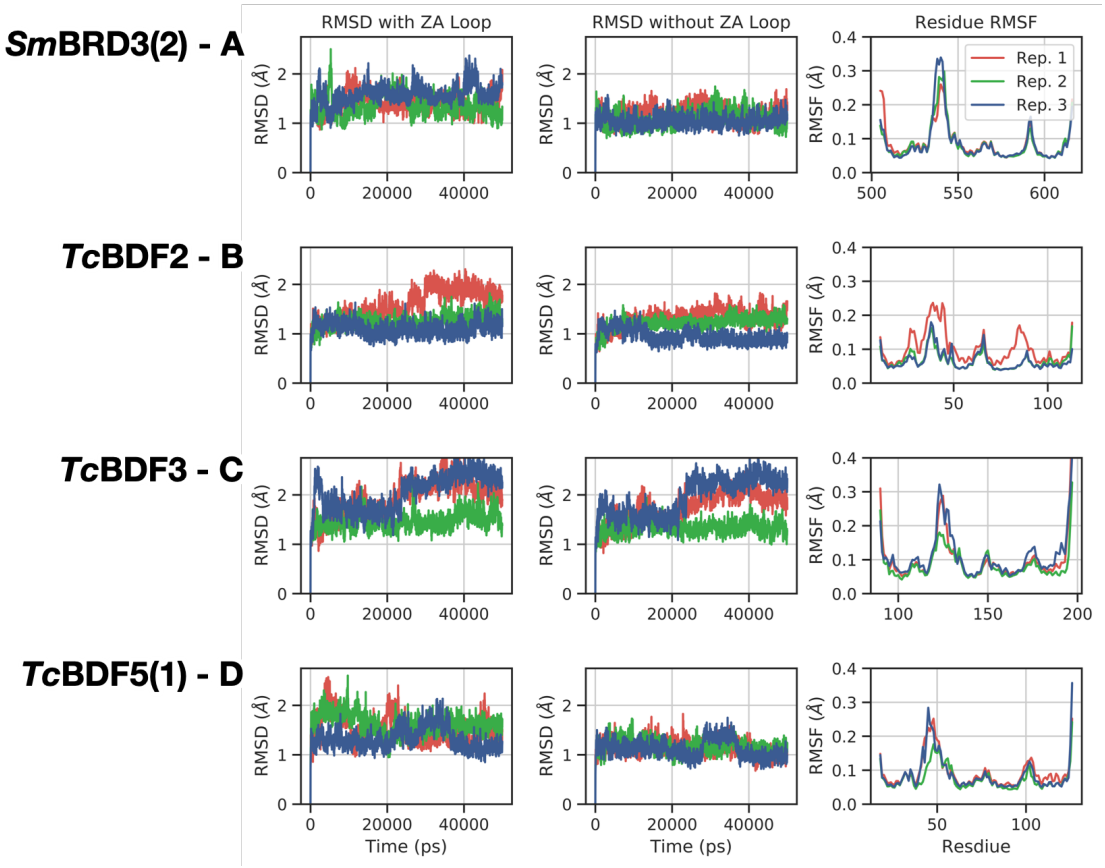


Figure 5.3: MD assessment of homology model stability. The stability of (**A**) *Sm*BRD3(2), (**B**) *Tc*BDF2, (**C**) *Tc*BDF3, (**D**) *Tc*BRF5(1) homology models were assessed using RMSD of the whole protein, the protein without the ZA loop and root mean squared fluctuation (RMSF). Simulations were performed in triplicate, with the different replicas shown in red, green and blue.

5.3 Studies on the model of *Tc*BDF3

The homology model of *Tc*BDF3 identified a canonical BRD site, with the Asn residue (N139) capable of interacting with the KAc mimic conserved in the structure. Docking studies were performed on binders of the *Tc*BDF3 construct, focusing on **20** which was identified as the highest affinity binder known, and **21** which was a slightly weaker analogue. Docking studies identified several poses of **20**. However, only one pose mirrored the crystallographic structures in *Hm*BRD4(1) (PDB ID: 4O74) and *Tb*BRD3 (PDB ID: 5C8G) and placed the KAc mimic next to N139, shown in Figure 5.5A. This pose placed the amide and attached piperidine ring away from the binding site as solvent exposed.

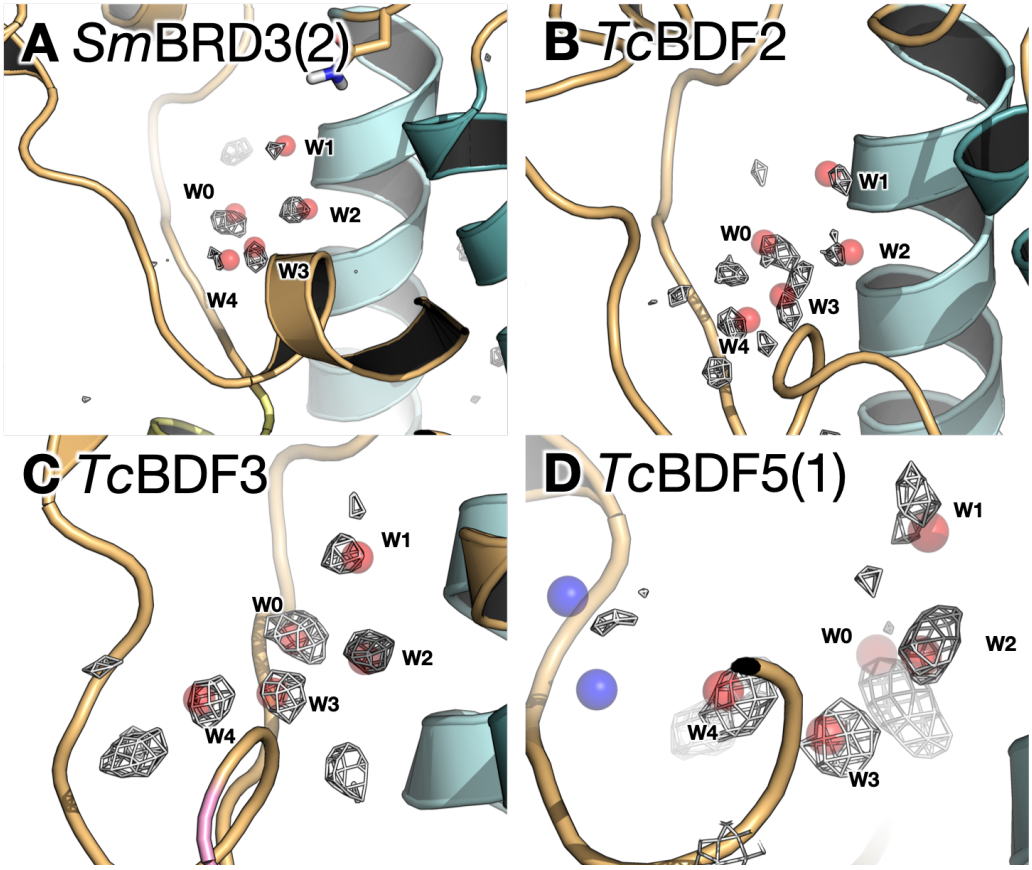


Figure 5.4: Regions of high water occupancy in *apo* simulations of homology models. Regions of high water density in (**A**) *Sm*BRD3(2), (**B**) *Tc*BRF2, (**C**) *Tc*BDF3, (**D**) *Tc*BDF5(1) are shown using the wire mesh. Water molecules from the homology model templates have been imposed for reference. Red water molecules indicate binding site waters, blue water molecules indicate waters in the ZA channel.

As discussed, the homology models showed that one side of the ZA channel is partially solvent exposed in *Tc*BDF3. Despite the increase in solvent exposure, MD simulations showed that the anisole ring remained largely on the hydrophobic wall within the ZA channel, represented by the distance to L75 in Figure 5.5B. In contrast, in *Hm*BRD4(1) (PDB ID: 4O74) there is a bulky Trp (W81) residue which forms a side on π-π interaction with the central anisole ring and shields it from solvent, shown in Figure 5.5C. The plot in Figure 5.5D shows that the central anisole ring is effectively pincered within the ZA channel between the ZA wall, represented by the distance to L92, and the WPF shelf. Note that the distances from the second MD simulation were omitted from this plot, as it showed the ligand

exiting the binding site. The lack of the ZA channel walls pincering **20** could partially justify why there is a large difference in binding affinity for **20** between *Hm*BRD4(1) (K_d = 37 nM) and *Tc*BDF3 (K_d = 15 µM).[60] The piperidine group was modelled as charged, acting as a solubilising group. This group remained largely in solvent, but was able to form transient interactions with E74.

Docking and MD simulations on **21** in *Tc*BDF3, identified a similar pose of the compound core, shown in Figure 5.5E, though a different regioisomer of the anisole core. The docked model placed the group beyond the amide linker within solvent, and this was maintained during MD simulations. The anisole group was found to be less attached from the ZA wall, shown in Figure 5.5F.

The observations that the solvent-exposed group in **20** formed no vital interactions with *Tc*BDF3, and likely acted as a solvating group, led to the hypothesis that the compound could be truncated to just feature the BI-2536 core, resulting in compound **93**. This compound was found to be a weaker binder of *Tc*BDF3, with a K_d of 335-365 µM (determined from NMR titration, C. Laurin). Further synthetic work by C. Laurin added the Alexa Fluor™ 488 group (ThermoFisher) at the amide linker to produce compound **94**. This compound was used in florescence polarisation (FP) assays as a high-throughput means of identifying *Tc*BDF3 binders. The K_d values of the probe was determined using FP by C. Laurin - TcBDF3: 3.36 ± 1.17 µM; *Hm*BRD4:9.9 ± 2.5 nM; HmBRD4(2): 14.9 ± 1.7 nM; PLK1: 5.5 ± 1.5 nM. However, the affinity of the compound could not be determined by ITC despite two attempts by C. Laurin (Appendix Figure C.4).

5.3.1 Guiding optimisation of BI-2536 (20)

To further improve the ligand affinity to *Tc*BDF3, additional hydrophobic or polar interactions needed to be built into **20**. The work in the previous section showed that the core is held within the binding site and that the anisole group sat adjacent to L75, however, no more contacts are made with the protein after this group. This is due to a turn in the ZA hydrophobic wall beyond L75, from the binding site. Changing the group beyond the core could allow the ligand to bind around

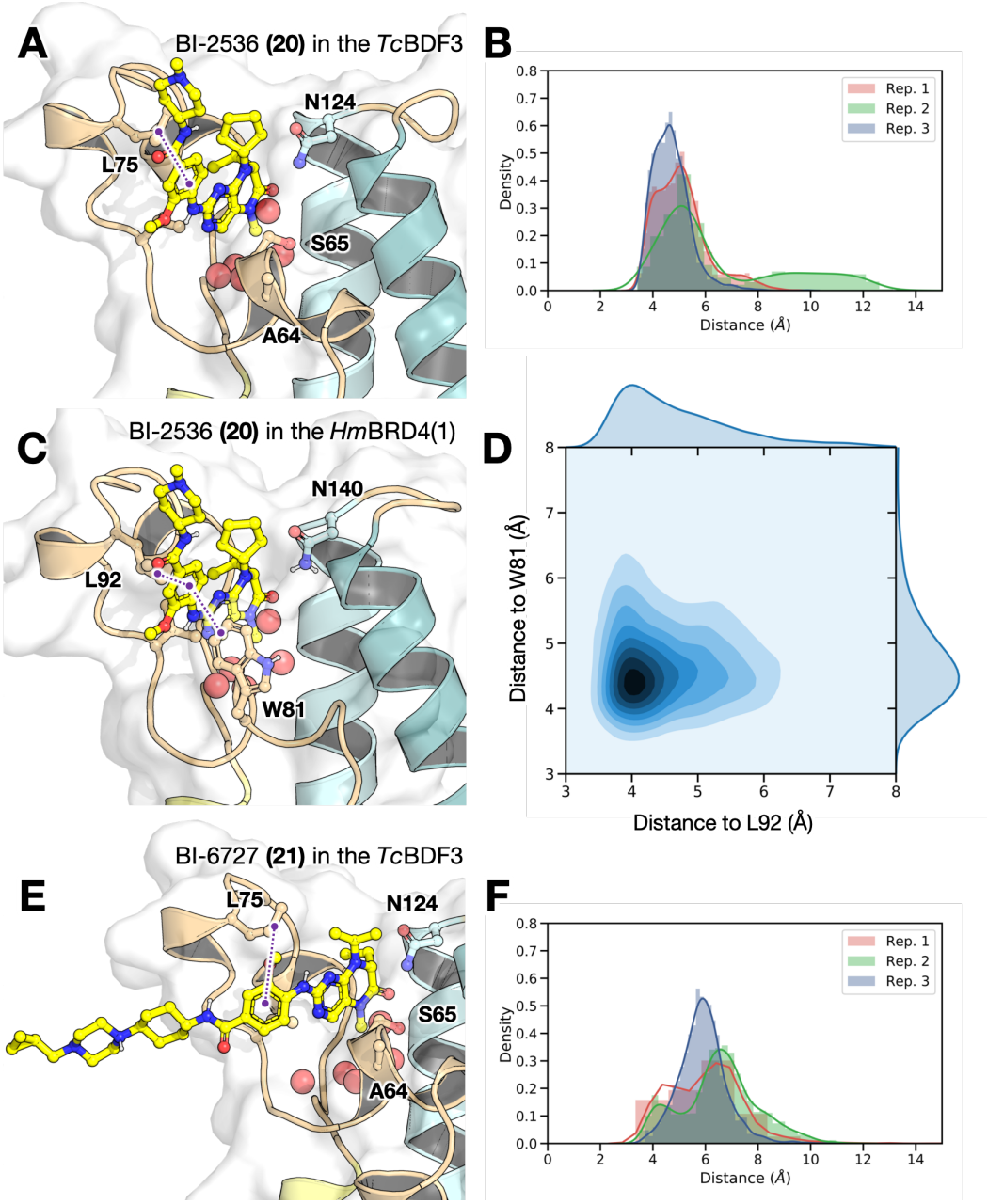
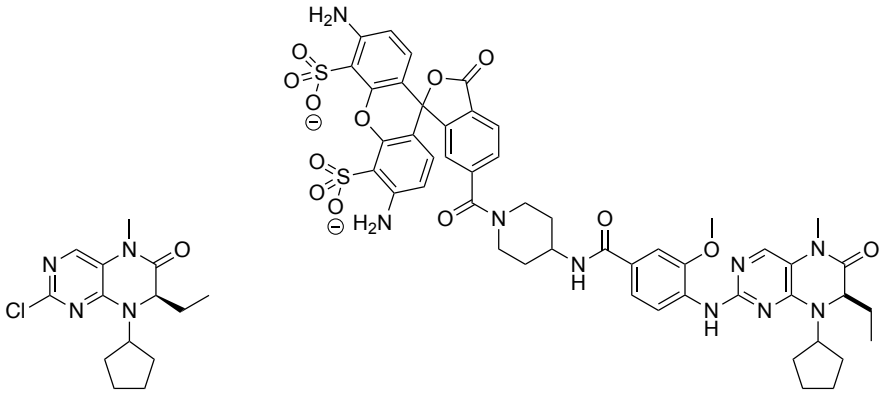


Figure 5.5: Studies of **20**/**21** in *Hm*BRD4(1) and *Tc*BDF3. (**A**) Docked structure of **20** in *Tc*BDF3. (**B**) KDE plot of the distance of the central anisole ring to L75. (**C**) Crystal structure of **20** in *Hm*BRD4(1) (PDB ID: 4O74). (**D**) Jointplot showing the distances of the anisole ring to L92 and W81 in *Hm*BRD4(1), which are on opposite sides of the ZA channel.(**E**) Docked structure of **21** in *Tc*BDF3. (**F**) KDE plot of the distance of the central anisole ring to L75.



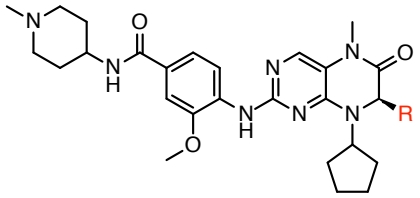
93 **94**

Figure 5.6: Derivatives of BI-2536 (**20**). Compound **93** was a designed as a minimal binder of **20** and showed strong binding. Compound **94** was a flourophore designed by C. Laurin and used in a FP displacement assay. [234]

this corner and increase the surface area of hydrophobic contacts. Around this corner there are also two polar residues, E74 and N71, which are solvent exposed but could be explored to form polar interactions.

With the absence of a hydrophobic shelf on the adjacent side of the ZA channel building hydrophobic groups perpendicular to the molecule, such as those in OXFBD02 (**9**) and I-BET151 (**11**), was unlikely to gain affinity. Especially as simulations by A. Boczek showed the pyridyl group of **11** failed to reside on this region (unpublished). The ZA channel was an additional region that could have been targeted for maximise affinity, however, there was no co-crystal data available to understand which ZA waters are displaceable.

It became clear that designing analogues of **20** was a multivariate problem, where the ligand needed to be adapted to improve affinity but also altered to remove selectivity for kinases and human bromodomains. Firstly, it was hypothesised that *Hm*BRD4(1) selectivity could be removed by alteration of the KAc mimic *(R)*-ethyl group. This group mimics the methyl group on the KAc residue and studies by Vollmuth *et al.* have shown that the binding affinity for *Hm*BRD4(1) to KAc is reduced whin the methyl group was replaced by larger groups such as *n*-propyl or *n*-butyl groups.[235] Chen *et al.* had only probed the substitution of an ethyl group to either a proton or a benzyl group in *Hm*BRD4 and removing the



Compound	R Group	$\Delta\Delta G_{MMPBSA}$(*Hm*BRD4)	$\Delta\Delta G_{MMPBSA}$(*Tc*BDF3)
20	*(R)*-Et	0	0
95	*(R)*-*n*Pr	$+5.58 \pm 2.93$	-1.13 ± 3.32
96	*(R)*-*i*Pr	-2.00 ± 1.01	-1.17 ± 2.73
97	*(R)*-CyPr	$+1.12 \pm 1.52$	$+2.07 \pm 2.47$
98	*(R)*-CyPe	$+2.49 \pm 3.83$	-
99	*(R)*-*n*Bu	-4.68 ± 1.64	-1.67 ± 2.86

Figure 5.7 & Table 5.3: Analogues of **20**. The *(R)*-Et groups were replaced to increase bulk and scored in both *Hm*BRD4 and *Tc*BDF3 using MM-PBSA. All scores are presented as relative to BI-2536 binding the respective protein. Units are reported in kcal/mol. CyPr = cyclopropane and CyPe = cyclopentane. ΔG_{MMPBSA} values were calculated in triplicate, where simulations containing unbinding were removed. $\Delta\Delta G_{MMPBSA}$ errors are reported as the standard error.

stereochemistry, showing that a benzyl group prevents binding.[63] The binding site in *Tc*BDF3 was visibly larger than in *Hm*BRD4(1) and hence it was hypothesised that selectivity for only *Tc*BDF3 could be retained by replacing the *(R)*-ethyl group in **20** with larger groups. MD simulations of differing R groups, **95** to **99**, were prepared by by A. Boczek and allowed for subsequent scoring using MM-PBSA. These calculations showed large errors, reducing confidence in its ranking ability. In addition, it predicted that larger groups were tolerated, to varying degrees, in both *Hm*BRD4(1) and *Tc*BDF3, shown in table Figure 5.7. However, the scoring by molecular mechanics Poisson-Boltzmann surface areas (MM-PBSA) free energy estimations indicated that *n*Pr and *n*Bu were substitutions that could improve selectivity. Applying these modifications to the histone tail lysine has already been shown by Vollmuth *et al.* to reduce *Hm*BRD4(1) affinity while the scores show that these modifications could maintain or improve *Tc*BDF3 affinity.[235]

Secondly, to address the low affinity of **20**, it was important to design ligands that were of higher affinity and also easy to synthesise for testing. Compound **100** was a truncated analogue of **20** that was synthesised by C. Laurin. It was hypothesised that it could act as an amide coupling partner to test a variety of

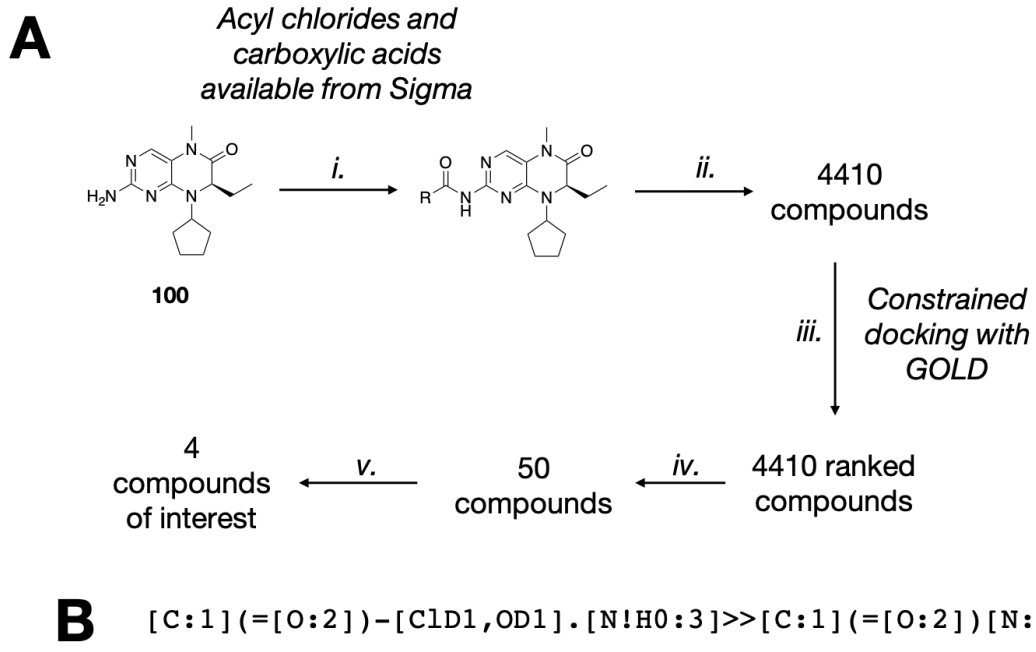


Figure 5.8: Workflow used to generate one-step amide analogues. (**A**) The workflow for generating amides: i) The structure of **100** was combined with all acyl chlorides and carboxylic acids available from Sigma Aldrich (January 2019) using the reaction SMART shown in panel **B**; ii) The resulting compounds were subject to a structure-factor index filter of < 11 and checked against known PAINS sub-structures.[195] iii) Remaining compounds were docked against the *Tc*BDF3 homology model using H-bond contained docking between the amine group of N124 and the carbonyl group of **100**; (iv) The top 20% of compounds were subject to manual inspection and all reagents with possible side reactions removed. (v.) The filtered compounds were subject to MD simulations to assess stability and identify new interactions.

different one-step amide couplings and allow chemical space around the anisole ring to be rapidly investigated.

The workflow used to design the series of all possible one-step amide coupling products is shown in Figure 5.8. The amides were produced using an amide coupling reaction SMART, adapted from Oakley *et al.*,[236] and using all acyl chlorides and carboxylic acids available from Sigma in January 2019. These were subject to a PAINS filter to reduce the number of false positive compounds, where substructures were taken from the work of Saubern *et al.*,[195] and a structure factor index filter (SFI=clogD + (No. aromatic rings)) of < 11 to remove insoluble compounds. Initial docking attempts using AutoDock4 regularly failed to place the **20** core in the

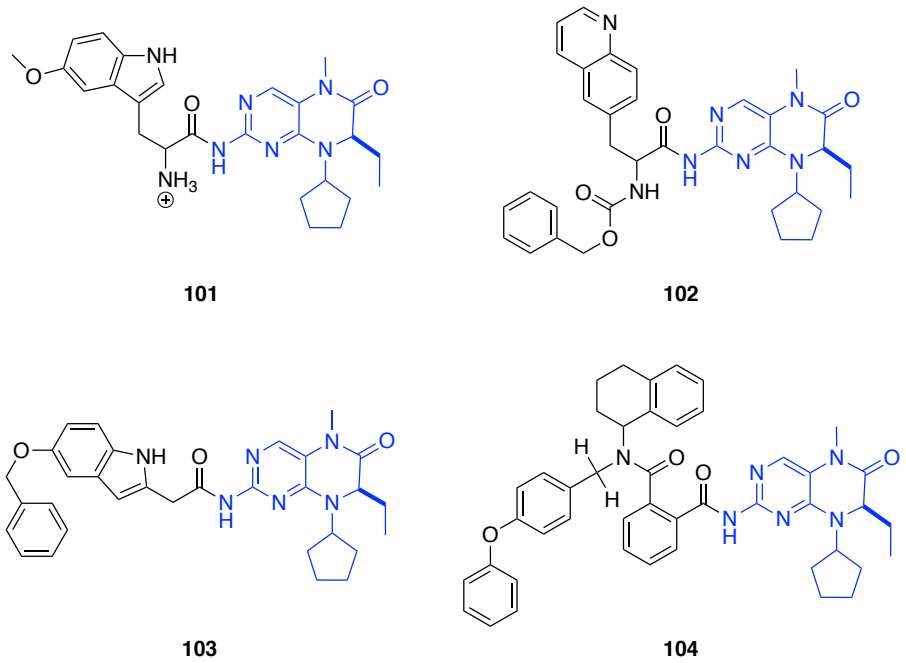


Figure 5.9: Synthetically tractable amides that were stable during MD simulations. MD studies are shown in Figure 5.10 and Figure 5.11. The **100** core has been highlighted in blue.

binding site, and hence restrained docking alternatives were investigated. GOLD, from the CCDC, was used instead as this allowed for H-bond constraints to be imposed within docking while maintaining a small calculation cost.[237] The atom number of the core carbonyl oxygen atom was needed to facilitate the constrained GOLD docking, and was identified using a substructure search with RDKit.

Initially, compounds were scored and ranked using the GOLD ChemPLP score. The top 20% were then subject to manual visualisation to remove compounds that: contained unfeasible geometries,[238] would have undergone side reactions upon synthesis or would not undergo amide coupling. Of the compounds generated, 20 were taken forward for further study in MD simulations. Four compounds maintained carbonyl binding to N124 in at least one of the simulations and were then visually inspected to understand the interactions gained (Figure 5.9).

101 projected two groups out from the amide bond: a charged amine towards the ZA channel and an indole group on an sp^3 linker, shown in Figure 5.10A. MD showed there were two fates for this group: either it was able to turn the L75 corner

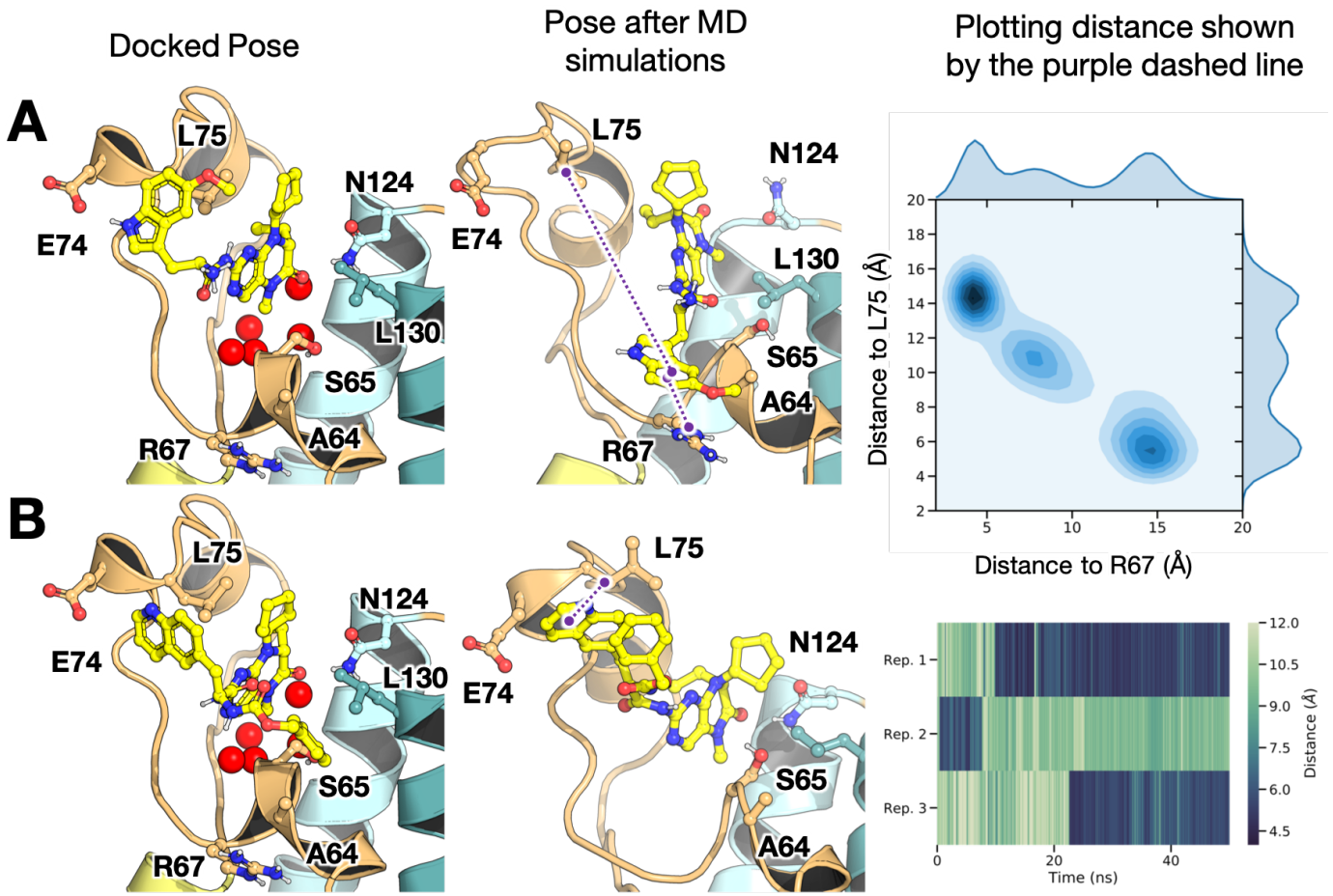


Figure 5.10: Docking and MD studies of tractable amides **101** and **102**. The amides shown in Figure 5.9 were docked to the *Tc*BDF3 homology model, where **101** and **102** are shown in the left column. These underwent triplicate MD simulations, which revealed different binding modes, shown by the central column. Distances (marked by purple dashes) were measured during the simulations and shown by the plots on the right. **(A) 101**, **(B) 102**.

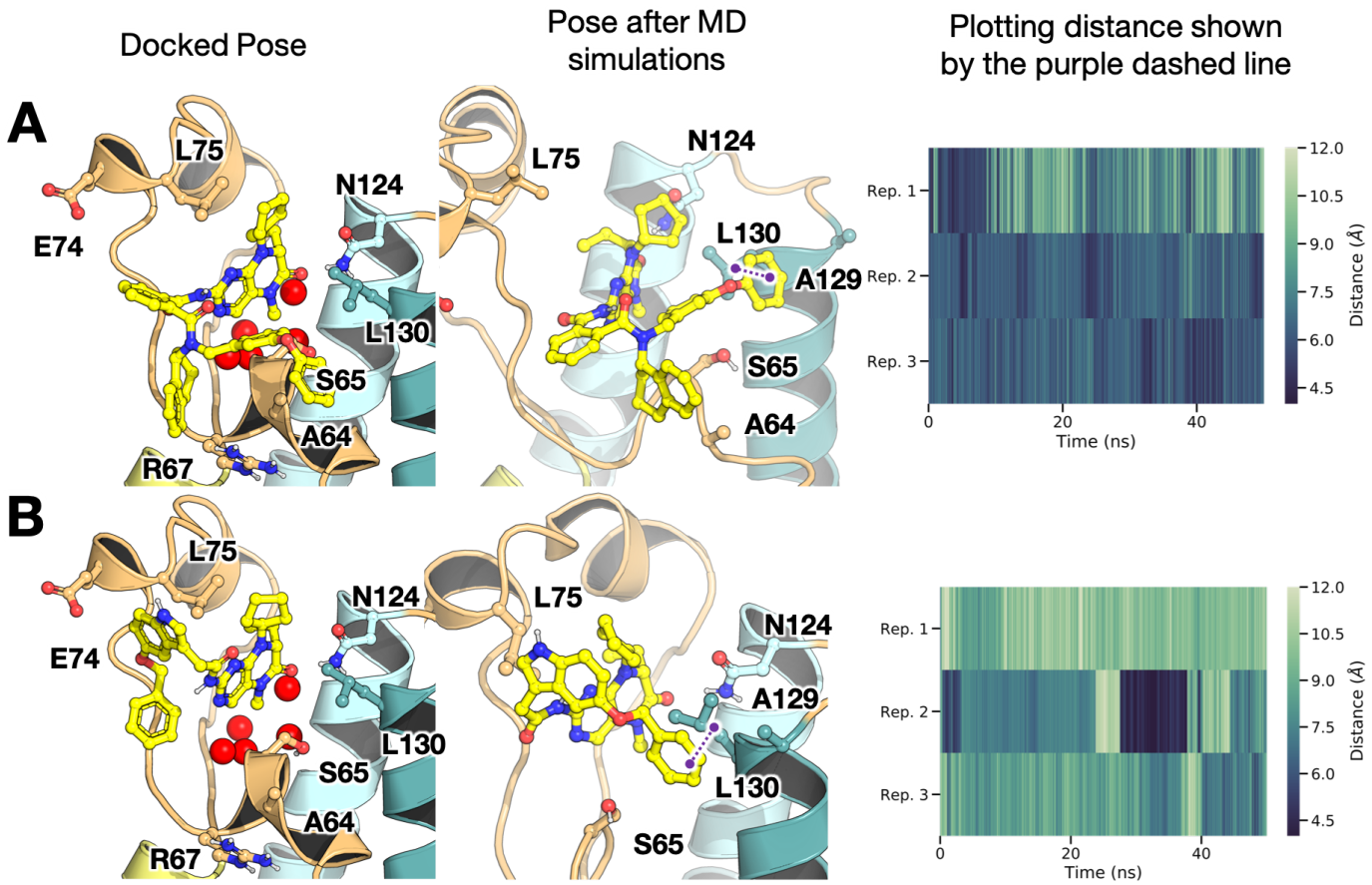


Figure 5.11: Docking and MD studies of tractable amides **103** and **104**. As seen in Figure 5.10, ligands **(A) 104** and **(B) 103** underwent docking and subsequent MD studies.

and place an indole group in the vicinity of L75 or form a cation-π interaction with R67, shown by the two islands in the 2D histogram in Figure 5.10A. Meanwhile **102** projected two aromatic groups away from the amide bond: a quinoline towards L75 and a Cbz group towards the ZA channel, Figure 5.10B. Simulations show the Cbz group formed no permanent interactions and pointed towards solution, while the quinoline group resided near L75 (shown by the barcode plot in Figure 5.10B). **103** contained a similar indole system to **101**, however, was extended to include another phenyl group *via* an ether linker. Despite the phenyl group starting docked in the ZA channel, in all simulations this compound folded over to some extent. This conformational change allowed the phenyl group to reach across S65 and form hydrophobic interactions with a hydrophobic pocket formed by L130 and A129, as shown in Figure 5.11B. Surprisingly, this pose do not allow the ethyl linker to form polar interactions with S65 (data not shown). A similar binding pose was seen with **104**, where the phenyl group resided in the same hydrophobic pocket during simulations, also shown as a barcode plot (Figure 5.11B). This ligand also placed the central, linking benzyl group in the hydrophobic region surrounding L75. However, this interaction prevented the tetralin group from forming lipophilic interactions and instead pointed unfavourably towards solvent, seen in Figure 5.11A. In summary, analysis of these compounds has shown that a sp^3 centre adjacent to the amide bond could allow the compound to wrap around L75 and form increased hydrophobic contacts. In addition, phenyl groups on extended linkers could be used to probe a previously unexplored, hydrophobic region of the binding site formed by L130 and A129. Finally, that a cation-π interaction is possible beyond the ZA channel with aromatic systems that can adopted the correct geometry.

The workflow used has two outstanding issues; firstly, the ChemPLP score was visibly biasing towards high molecular weight (MW) compounds. Secondly, performing the simulations with GROMACS 2019.0 prevented improved scoring of all ligands taken forwards for MD studies with MM-PBSA, due to the removal of implicit solvent from the codebase. Future work should look to re-weigh the GOLD scores according to molecular weight, so investigate smaller perturbations

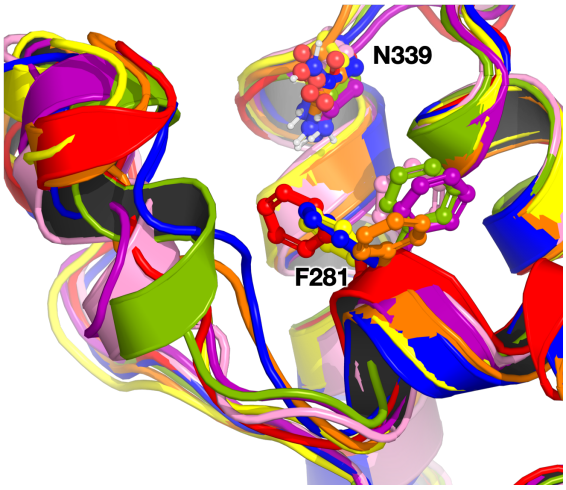


Figure 5.12: *K*-means clustering of the *Sm*BRD3(2) homology model after MD

to the initial fragment. In addition, if MM-PBSA studies are used for improved scoring, an older version of GROMACS should be used.

5.4 Investigating *Sm*BRD3(2) binders

Molecular docking studies of BI-2536 (**20**), I-BET726 (**12**) and I-BET151 (**11**) were performed in *SM*BRD3(2), to understand the putative binding modes of these ligands and guide future ligand design. Initial docking studies failed to place the ligand within the binding site, instead placing ligands above the binding site (data not shown). Closer inspection of the homology model found that F281 was pointing directly into the binding site, and thus occluding it. To identify different binding site conformations, *apo* simulations of *Sm*BRD3(2) were subject to *k*-means clustering into 7 representative clusters, shown in Figure 5.12 and described in Chapter 2. This identified three representative clusters where the F281 residue is pointing perpendicular to the binding site and thus no longer occluded. The largest representative cluster was selected for future docking studies (purple in Figure 5.12) and was able to successfully dock ligands within the binding site, and with the KAc mimic interacting with N339.

Docking of **12** produced only one binding pose of the compound that placed the KAc mimic in the vicinity of N339, shown in Figure 5.13A. This pose failed to fully

resemble the pose seen in *Hm*BRD4(1), where the *para*-chlorobenzene group would reside on the WPF shelf. This *Hm*BRD4(1) ligand pose in *Hm*BRD4(1) has been overlaid with the homology model in Figure 5.13B and shows the benzylic clash with F281. This is due to the cluster used for docking studies having the F281 residue in close proximity to helix α_C, relative to W81 of the *Hm*BRD4(1) WPF shelf, hence the hydrophobic nook between F281 and α_C is quite tight. The docked pose instead had a different ring conformation, the *para*-chlorobenzene group pointing away from the binding site and into solution and the benzoic acid group perpendicular to the core.

The clash of F281 with the hydrophobic chlorobenzene group, and different ring conformation, suggested that the homology model was not in the optimal conformation to bind **12**. Therefore, the docked pose was subject to MD simulations to allow the binding site and ligand to interact further. Only one simulation retained the ligand in a stable binding pose, where the ring kinked to the form in *Hm*BRD4(1) and the *para*-chlorobenzene group was able to move and reside on the nook on the HFF shelf, shown in Figure 5.13C. This adjustment in the protein and ligand showed that the binding mode in *Hm*BRD4(1) was indeed possible and additional affinity can be gained from the HFF shelf. It also shows that the carboxylic acid group is capable of interacting with K284 and H280. Later co-crystallisation studies by Dr M. Schiedel resulted in crystallisation of **12** in *Sm*BRD3(2). The structure (unpublished) is shown in Figure 5.13D. There is a small difference in organisation of the LLY region of the ZA loop, where the crystal structure shows L291 forming additional hydrophobic contacts with the ligand, rather than pointing into solution. This difference stems from the cluster chosen for docking studies, where the *apo* protein undergoes larger ZA loop movements in the absence of ligand. However, this structure has an RMSD difference of 0.93 Å from the snapshot shown in Figure 5.13C, validating the model generated through docking and MD to allow for an induced fit.

Compound **11** also relies on gaining affinity from the WPF shelf in *Hm*BRD4(1) (PDB ID: 3ZYU), where the pyridyl group residues on the shelf.[41] An alternative binding pose has been seen in *T. brucei* (PDB ID: 4PKL), where the benzimidizalone core pointing into the binding site as the KAc mimic, which has been hypothesised as

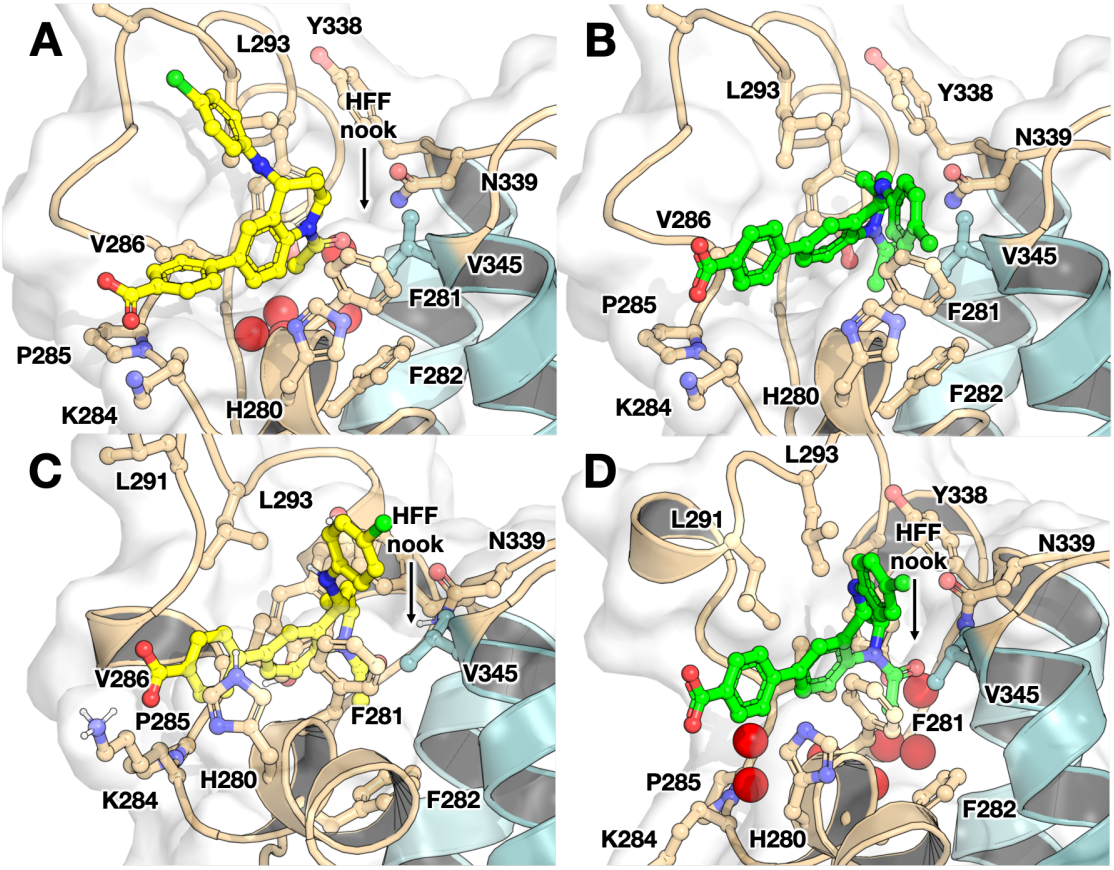


Figure 5.13: Docking, MD, and crystallographic binding modes of I-BET726 (**12**). **(A)** Only docked pose of **12** that placed the KAc mimic in the vicinity of N339. **(B)** Overlay of the crystallographic pose of **12** in *Hm*BRD4(1) (PDB ID: 4BJX) on the *Sm*BRD3(2) homology model. **(C)** A snapshot 40 ns into an MD simulation of the docked pose shown in panel A. **(D)** The binding site of the co-crystal structure of **12** in *Sm*BRD3(2), obtained by Dr M. Schiedel (Chain A, 1.2 Å, unpublished).

being due to a π-stacking interaction of this group with a Trp adjacent to the binding site.[89] Molecular docking studies identified both these poses in *Sm*BRD3(2): pose 1 shown in Figure 5.14A and pose 4 shown in Figure 5.14B. However, the pose seen in *Hm*BRD4(1), pose 1, was not fully recreated as the pyridyl group was unable to occupy the HFF nook and instead remained solvent exposed, Figure 5.14A. MD studies showed both poses were stable in the binding site, shown by the RMSD plots in Figure 5.14C for pose 1 and Figure 5.14D for pose 4. The major movement within the ligand is the pyridyl group. Figure 5.14E shows that for pose 1 the pyridyl group remains in the vicinity of V289 and does not move into the HFF

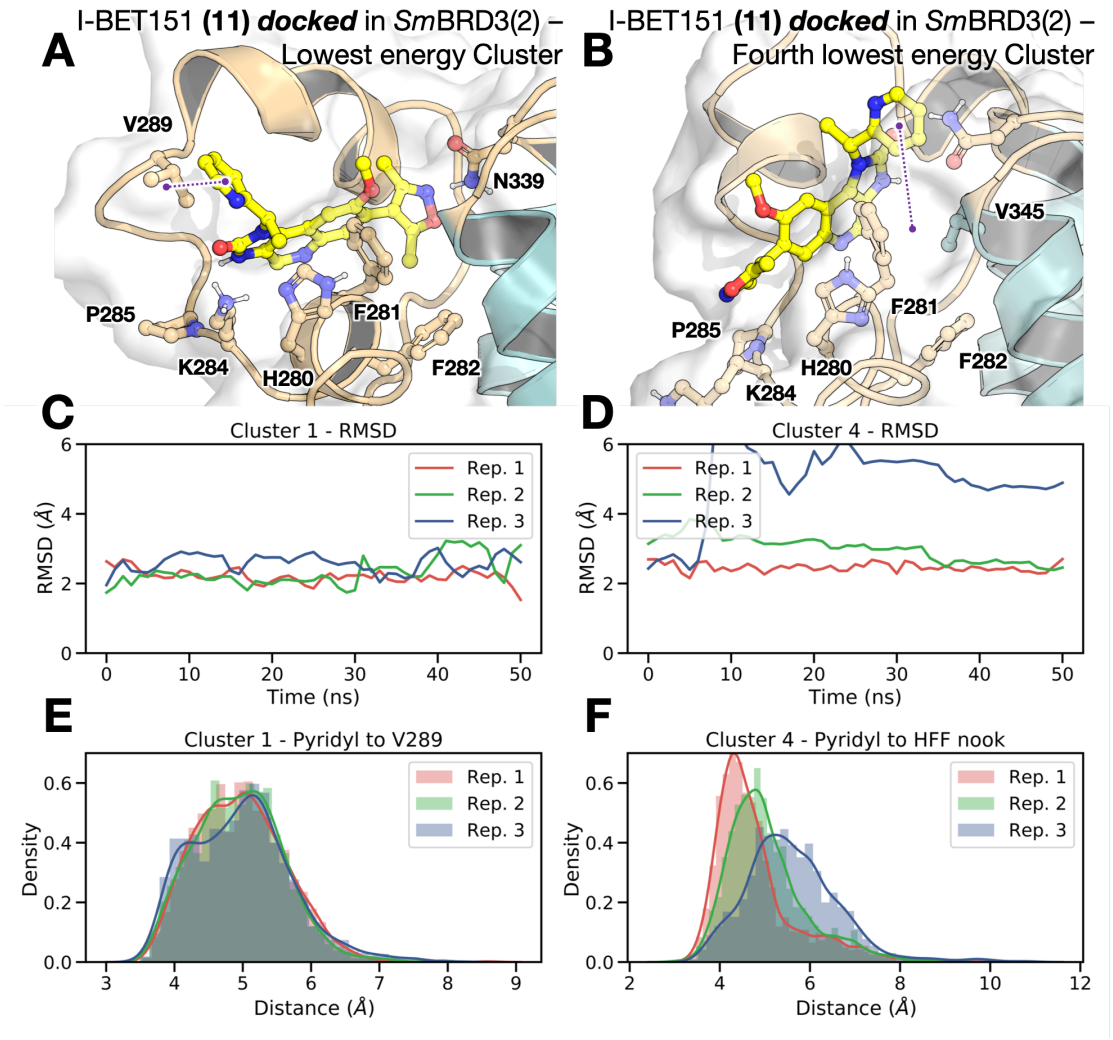


Figure 5.14: Clustering of docking studies of I-BET151 (**11**) against *Sm*BRD3(2) gave two clusters that resembled known **11** binding modes: Cluster 1 **(A)** and Cluster 4 **(B)**. MD studies showed both poses were stable in the binding site, **(C)** and **(D)**, with the exception of the pyridyl group which sampled local space. Plots **(E)** and **(E)** show the distance of this ring to V289 and the midpoint of the HFF nook during MD simulations of the respective docked poses. Simulations were performed in triplicate, shown by the red, blue and green lines.

nook. While in Figure 5.14F, which measures the distance of the pyridyl group to the bottom of the HFF nook, it is shown that in pose 4 the pyridyl group partially resides in the HFF nook. With both poses stable and with sub-optimal binding of the pyridyl moiety, more rigorous studies would be needed to understand which pose is the most energetically favourable. Once the lower energy pose is obtained, exploring the SAR around this pyridyl group would be the highest priority.

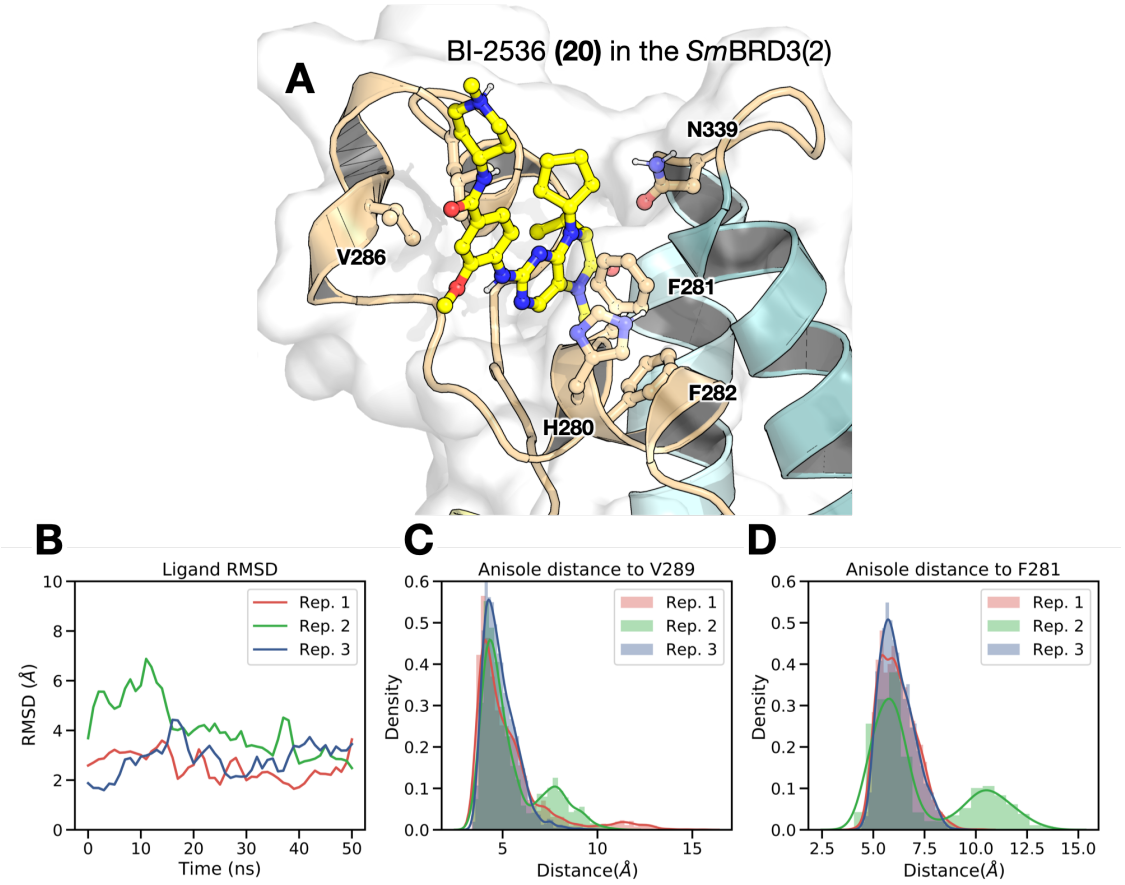


Figure 5.15: Docking and MD studies of BI-2536 (**20**) binding in *Sm*BRD3(2). **(A)** Docked pose of **20** binding in *Sm*BRD3(2) that placed the core in the binding site. **(B)** The ligand RMSD during the simulations. **(C)** The distance fo the anisole group to V289, on the hydrophobic wall of the ZA channel. **(D)** The distance of the anisole ring to F281, which is on the opposite side of the ZA channel to V289.

The docked pose generated of **20** was observed to be similar to that seen in *Tc*BDF3, shown in Figure 5.15A. The ten fold increase in binding affinity for **20** to *Sm*BRD3(2) over *Tc*BDF3 could be explained by the reintroduction of a bulky aromatic group adjacent to the the anisole group, such as that seen in *Hm*BRD4(1). Disregarding the second simulation replica, which shows partial ligand displacement in Figure 5.15B, it can be seen the anisole group remains pincered between V289, Figure 5.15, and F281, Figure 5.15D. This binding pose would also predict that the compound would work with **94**, the fluorescent probe, which was later confirmed by Dr M. Schiedel. This exhibited the same behaviour in ITC, where a K_d could not be determined.

In the final study, OXFBD02 (**9**) was expected to be a binder of *Sm*BRD3(2) given the HFF nook provides a similar hydrophobic region to the WPF shelf in *Hm*BRD4(1), albeit slightly closer to the α_C helix. However, binding of **9** was not observed. Docking studies were unable to reproduce any sensible binding pose that involved N339 interacting with a KAc mimic. Given the importance of W5 on binding **9** in *Hm*BRD4(1), the water density analysis in the ZA channel was re-examined and a ZA channel water molecule added to the protein model used in docking. Docking was still unable to reproduce likely poses, as F281 prevented the phenyl group from residing in the HFF nook. An overlay with the I-BET726 (**12**) crystal structure, shown in Figure 5.16, shows that the phenyl group is unable to adopt the required geometry in the binding site to interact with N339 and placing the phenyl group in the HFF nook, and so unable to dock in both sets of studies.

Overall, these studies have shown that *Sm*BRD3(2) was able to bind ligands designed for *Hm*BRD4(1), due to the presence of the HFF nook. This is an equivalent of the WPF region seen in *Hm*BRD4(1). However, this nook is placed closer to the α_C-helix which prevents some *Hm*BRD4(1) ligands from binding. This difference, however, offers a unique opportunity to remove human selectivity from the ligands in future.

5.5 Enantioselective binding to *Tc*BDF5(1)

Only one compound has been identified as a binder of *Tc*BDF5(1): OXFBD02 (**9**). Chiral separation of this compound and subsequent testing by FP showed only a slight difference in binding affinity between enantiomers.[234] The homology model of *Tc*BDF5(1) showed a shallower hydrophobic shelf, formed by L29 lying flat across the shelf. Hence, it was expected that **9** would bind in the binding mode seen in *Hm*BRD4(1) (PDB IDs: 4JOS/4JOR).[40] Initial docking studies using just the binding site waters failed to reproduce the docking poses, which was expected as W5 is key to binding the hydroxyl group. Again, inspection of the water densities in Figure 5.4D showed a water molecule at the start of the ZA channel as it exits the binding site, shown in Figure 5.4D. The water molecule from the *Ld*BPK template

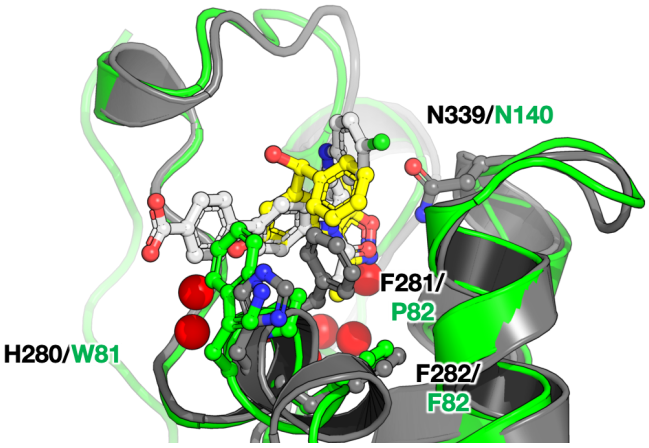


Figure 5.16: Comparison of OXFBD02 (**9**) (yellow) and I-BET726 (**12**) (white) binding in *Sm*BRD3(2). The binding mode of **9** (taken from a *Hm*BRD4(1) structure, PDB ID: 4JOS - green) aligned to and overlaid by the crystal structure of **12** binding in *Sm*BRD3(2) (grey - unpublished).[40]

closest to this density was introduced for docking and further docking studies gave the two poses shown in Figure 5.17A and B. Docking of both enantiomers showed the phenyl group could be placed on the shelf made by L29, with the hydroxy group on the linker interacting with either the backbone of G28 or the hydroxy group on Y38.

Through MD studies, the interactions of the hydroxy group to G28 and Y38 were investigated, as well as the proximity of the phenyl group to the L29 shelf. All simulations were stable and retained the ligand in the binding site for both enantiomers. Figure 5.17C compares the interactions of (*S*)-**9** with both these residues. This shows that in the majority of the simulation, the group is in close proximity to Y38 which places the phenyl group over L29 and on the hydrophobic shelf. Meanwhile, Figure 5.17D does the same comparison for (*R*)-**9** and shows the opposite, that the linker hydroxyl group interacts with G28 to place the phenyl group over L29. These observations explain why binding of both enantiomers is detected at similar affinities.

As discussed in § 5.2, the template of the *Tc*BDF5(1) homology model contained BI-2536 (**20**) in the binding site. Yet, this compound was shown not to bind *Tc*BDF5(1) by C. Laurin. Comparison of the binding sites shows that region above

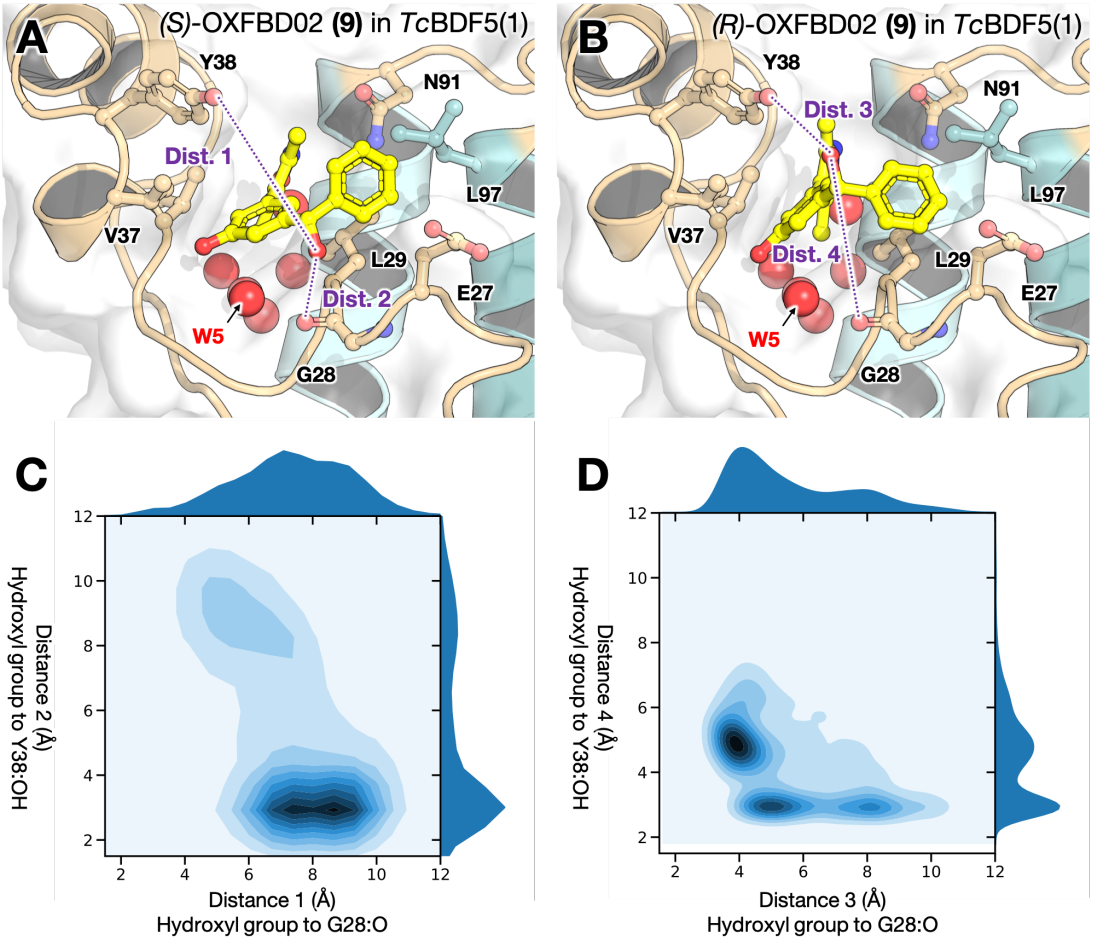


Figure 5.17: Both enantiomers of OXFBD02 (**9**) docked to *Tc*BDF5(1). (**A**) (*S*)-**9** and (**B**) (*R*)-**9** were docked and only the poses shown placed the isoxazole KAc mimic in the proximity to N91. Both docked poses places the phenyl group on the hydrophobic region above L29. A comparison of the distances between the backbone carbonyl of G28 and the hydroxyl group of Y38 are shown in Distances marked are shown in (**C**) for (*S*)-**9** and (**D**) for (*R*)-**9**.

the *Tc*BDF5(1) binding site is partially occluded by L97, shown in Figure 5.18. This bulky residue could clash with the cyclopentane moiety and likely prevent binding. While in *Ld*PBK(1) this residue is a Tyr, which is pinned away from the binding site through a polar interaction with H23. Further testing of **21**, or **20** analogues without the cyclopentane, against this compound would test this hypothesis. This also explains why **94** failed to bind in the FP assays, tested by C. Laurin.

After these studies were completed, an *apo* structure of *Tc*BDF5(1) was deposited in the PDB (PDB ID: 6NEY - 2.17 Å resolution). This was accompanied by two

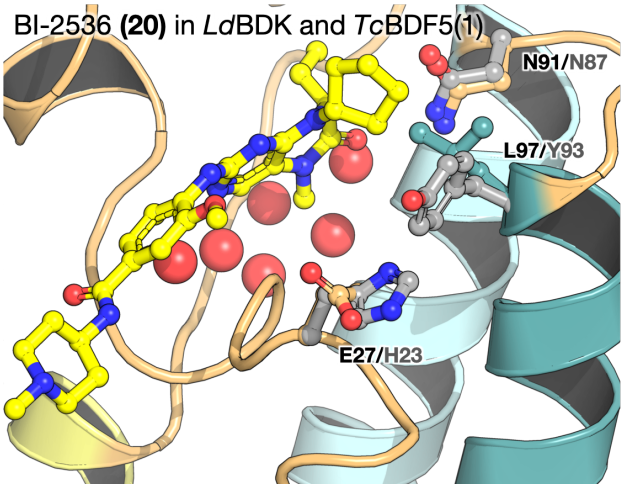


Figure 5.18: Overlay of BI-2536 (**20**) binding *Ld*BPK(1) and *Tc*BDF5(1) homology model. The coloured model of *Tc*BDF5(1) is aligned and overlaid with the crystal structure of *Ld*BPK1 (grey). The aromatic Y93 is pinned back away from the binding site through interaction with H23. In *Tc*BDF5(1) this residue is a bulkier Leu (L97) and partially blocks the top of the binding site.

structure of *Tc*BDF2 in its *apo* state and co-crystallised with bromosporine (**19**) (PDB ID: 6NP7 and 6NIM, respectively). The structure of *Tc*BDF5(1) was a close match with the homology model used in the previous studies, where the only visible difference was an extension to the α_C helix in the crystal structure. The RMSD of the homology model was 0.60 Å, and after 25 ns of MD the RMSD increased to only 0.85 Å. This small deviation from the crystal structure shows the homology model was a suitable substitute for the crystal structure. Hence, the observations from the MD studies are still valid.

5.6 Discussion

BI-2536, a dual *Hm*BRD4(1) and PLK1 kinase binder, was shown to bind all parasite bromodomain constructs, but *Tc*BDF5(1). This has also been shown to bind *L. donovani* and *T. brucei* bromodomains (from crystal structures awaiting accompanying publication), hence provides a starting point for developing potential tropical disease treatments. Despite the differing binding sites in *Hm*BRD4(1), *Sm*BRD3(2) and *Tc*BDF3, this ligand is accommodated in the binding site, gains

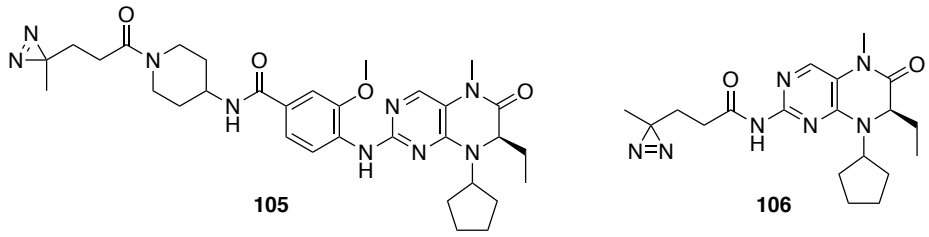


Figure 5.19: Diazerine derivatives of **BI-2536**. Compounds **105** and **106** were derived from BI-2536 and used in a LCMS based assay to determine additional *Tc*BDF3 binders.

affinity from placing the anisole ring on the hydrophobic ZA wall and has the solubilising group facing into solution. Confirmation that the solubilising group points to solution has led to the fluorescent probe **94**, by C. Laurin, which can aid in ligand development of **20** binders.

*Tc*BDF3 has a lower affinity for BI-2536 (**20**), likely due to the central anisole ring being solvent exposed on one side. This suggests the anisole group is less vital to obtaining affinity through hydrophobic interactions. This observation has been validated by testing of **93**, which removes all but the KAc mimic and retains binding. This observation has also contributed to the design of two diazerine linkers, **105** and **106**, which have been synthesised by C. Laurin. These were used as controls in a LCMS based library search of diazerine containing compounds to identify binders of *Tc*BDF3, where binders were able to cross link to the protein and led to a mass shift.[234]

Before analogues of **20** can be used as a probe in development of tropical disease treatments, it is important to remove its kinase binding ability. Especially if there are equivalent kinase proteins within *T. cruzi* and other parasite species. Kinase binding is dependent on the hinge binding motif, where two nitrogen atoms in the centre of **20**, shown in red in Figure 5.1, bind the backbone of a cysteine residue, C133 in PLK1.[60] It has been shown in literature,[62, 63] and subsequent validation by M. Keller (unpublished work), that removal of the hinge binding nitrogen atoms can remove kinase binding ability. MD simulations by A. Boczek showed that removal of one or both of these hinge binding atoms prevented the ligand binding deep in the PLK1 binding site, and MM-PBSA calculations of these simulations

showed positive $\Delta\Delta G_{MM-PBSA}$ values, shown in Appendix Figure C.6. Hence, future optimisation of **20** should also focus on removal of ideally both hinge binding nitrogens, which are shown in these models to be non-vital for *Tc*BDF3 binding.

Further to these modification, synthetic validation of replacing the KAc ethyl with *n*-propyl or *n*-butyl could provide a way to remove human selectivity in **20** analogues. The *in silico* validation, through triplicate MM-PBSA calculations, suffered large errors which were often larger than the magnitude of the value and so still ambiguous at to whether the modification would increase selectivity. However, if further computational validation is needed, there are two approaches that could be used. Firstly, using the MM-PBSA calculation setup seen in the *ESMACS* method, proposed by Wright *et al.*[181] This approach uses a larger set of replica simulations, though using simulation data from 4 ns compared to the 35 ns used within this work, to bester estimate the error and increase sampling. Alternatively, these systems could be subject to ABFE calculations which have been shown by Aldeghi *et al.* to improve the ranking accuracy of compounds by predicted ΔG.[109]

The summary of all these proposed modifications to **20** are shown in Figure 5.20. Future work should look at identifying suitable R' groups to increase hydrophobic contacts with the ZA wall, incorporate in a cation-π interaction with R67 or possibly build across to L130 and A129. There are multiple methods that could be used to design suitable ligands to incorporate these. One approach would be to identify small fragments that bind well within these proposed regions and then identify suitable linkers to join them to the **20** core. Many tools exist to perform this, for example Yamaotsu *et al.* have developed a package called *Fsubsite* which uses a knowledge based approach to identify putative binding locations of common ligand substructures.[239] Co-solvent simulation solutions have also been published which identify protein regions favoured by a particular solubilised fragment during simulation.[240]

Docking studies of I-BET726 (**12**) did not determine the binding mode in the homology model of *Sm*BRD3(2), without MD being used to relax the binding site and allow the *para*-chlorobenzene group to move into the HFF nook. The resulting

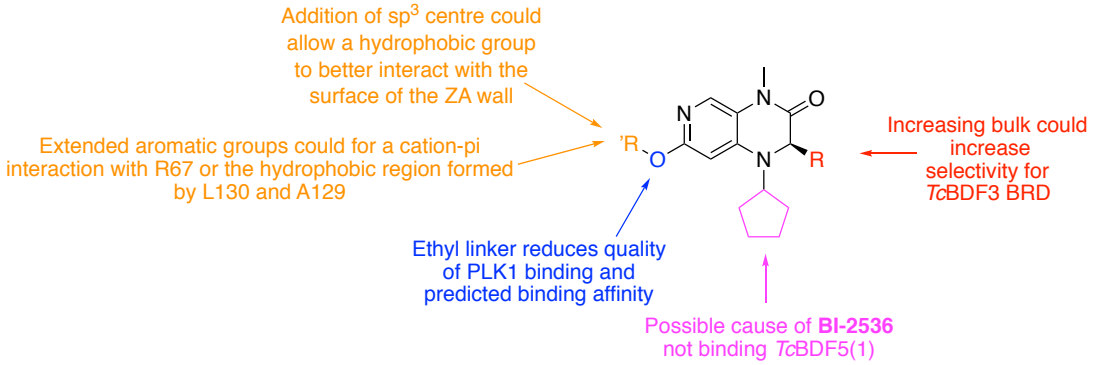


Figure 5.20: Summary of suggested modifications for BI-2536 (**20**)

model has subsequently been confirmed by a co-crystal structure and shows close agreement. Both the model and crystal structure show that there is scope to build off the benzylic acid ring to increase its hydrophobic surface area, towards L293, and hence hydrophobic interactions with the ZA wall.

The lack of binding of **20** by *Tc*BDF5(1), which could be caused by L97 preventing the cyclopentane moiety from binding, is useful in developing a *Tc*BDF3 selective probe. On top of building in modifications to remove *Hm*BRD4(1) and kinase binding, the bulky cyclopentane group should be retained to ensure the ligand cannot bind *Tc*BDF5(1). Meanwhile, large modifications of **9** are needed to remove its excellent human BET protein binding and increase its affinity for *Tc*BDF5(1). Modifications could focus on Y38, which could be used to gain either a cation-π or π-π interaction.

5.7 Conclusions

In summary, homology models of *Tc*BDF3, *Tc*BDF5(1) and *Sm*BRD3(2) were used in this chapter to understand the binding of human bromodomain ligands to parasite bromodomains. BI-2536 (**20**) was able to bind both *Tc*BDF3 and *Sm*BRD3(2), where the KAc mimic interacts with the binding site Asn residue and the central anisole ring sits against the hydrophobic wall of the ZA channel. The affinity of **20** is increased in *Sm*BRD3(2) due to F281 forming a π-π interaction with the other side of the anisole ring, while in *Tc*BDF3 this side is left solvent exposed. The

observation that the solubilising group on **20** formed no key interactions has led to the development of diazerine linkers and a fluorescent probe to aid in development of ligands for **20** binders. *In silico* techniques were employed to generate analogues of the the **20** core, which led to the observation that sp^3 centres off the core could increase the hydrophobic surface area of ligands by turning the ligand around L75, a cation-π interaction is possible with R67 and there is a hydrophobic pocket near the α_C helix which could all be exploited in future ligand design.

The binding mode of I-BET726 (**12**) to *Sm*BRD3(2) was correctly identified from combining docking and MD studies to allow the a relaxed version of the homology model to adjust to the ligand, a model that was subsequently confirmed with a co-crystal structure. This showed that the *para*-chlorobenzene group was able to reside in the nook created between F281, F280 and the α_C helix. This nook is too close to the α_C helix for the phenyl group of **9** or the pyridyl group of I-BET151 (**11**) to bind in this region, offering a chance to obtain selectivity over the human BET proteins. The **11** binding mode could not be determined, however, removal of the pyridyl group could aid in developing a better scaffold for *Sm*BRD3(2).

Despite not binding *Tc*BDF3, **9** was able to bind *Tc*BDF5(1). Binding of **9** was dependent on W5 in the ZA channel and the hydrophobic shelf adjacent to the wall in the ZA channel formed by L29. The hydroxyl linker in **9** was able to interact with both G28 and Y38 in the ZA channel, an additional H-bond interaction that also facilitated the phenyl group residing on the hydrophobic shelf formed by L29.

Overall, these models have offered insights into the binding modes of human bromodomain ligands. The insights can now be used to improve selectivity over the human bromodomains, particularly using the HFF nook in *Sm*BRD3(2). These can also be used to increase affinity for the parasite targets, such as incorporating additional interactions to *Tc*BDF3 in **20**.

Tout est poison, rien n'est poison, tout est une question de dose - Everything is poisonous, nothing is poisonous, it is all a matter of dose.

— Claude Bernard *Pathologie expérimenta*

6

Understanding the SAR of CREBBP and BRD4(1) binders

Contents

The aim of the work presented in this chapter was to use MD models to elucidate differences in binding affinity of OXFBD02 (**9**) analogues for BRD4(1) and to rationalise selectivity between BRD4(1) and CREBBP of a CREBBP BRD ligand series. Part of this work has already been published and can be used to expand on the details covered in this chapter.[231]

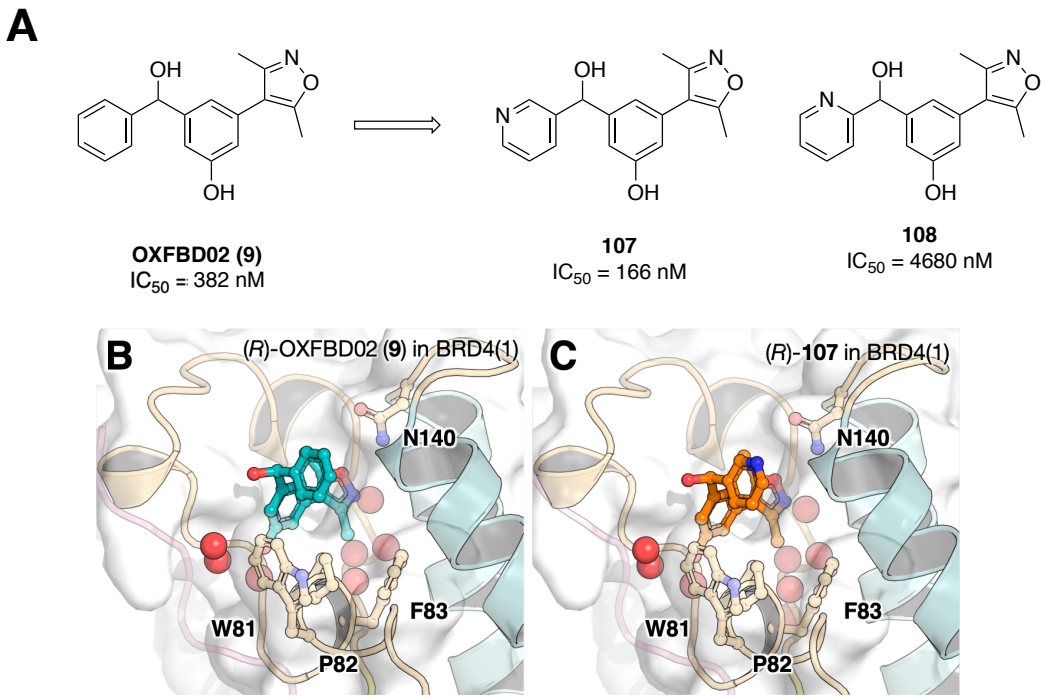


Figure 6.1: BRD4(1) Ligands and crystallographic binding modes. **(A)** OXFBD02 (**9**) and its pyridyl derivatives: the 2-pyridyl (**108**) and the 3-pyridyl (OXFBD04, **107**). IC_{50} values reported were determined by AlphaScreen™.[231] **(B)** Structure of *(R)*-OXFBD02 bound to BRD4(1) (PDB ID: 4J0R).[40] **(C)** Structure of *(R)*-**107** bound to BRD4(1) (PDB ID: 6FSY).[231]

6.1 Introduction

There are 61 known bromodomains (BRDs) which exist in 46 bromodomain containing proteins (BCPs) within the human body.[241] Over the last few years, several probes have been successfully developed for a number of these bromodomains.[5, 27, 242] The majority of these ligands target the bromodomain and extra terminal domain (BET) family of BCPs, which contain two adjacent BRDs (BD1 and BD2), and have proven to be therapeutic targets.[242] However, BRD ligands have suffered from a lack of selectivity amongst the 61 known BRDs, with BET ligands targeting all members of the family, and showing differing selectivity between BD1 and BD2.[243]

6.1.1 OXFBD02 derivatives

OXFBD02 (**9**) is a BET BRD ligand developed by Hewings *et al.*, which binds to BRD4(1) with nanomolar affinity (IC_{50}=382 nM).[231] This compound contains a

terminal phenyl ring that resides in the WPF shelf of BRD4(1),[40] a hydrophobic cleft formed by W81, P82 and F83 (Figure 6.1B and C). This interaction improves both the affinity and selectivity of the ligand for the BET family of BRDs. **9** also contains an isoxazole moiety that acts as a KAc mimic.[244] Further SAR studies to explore the WPF binding moiety found that changing the substitution from a phenyl to a pyridyl group increased the metabolic stability of the compounds. It was also found that this substitution could enhance the affinity, as seen with ligand **107**. However, moving the nitrogen one position closer to the linker led to a dramatic reduction in affinity, as seen with ligand **108**.[231] It was predicted that an intramolecular H-bond within **108** could be responsible, but the reasons for the impact of these interactions on binding affinity were not clear. A crystal structure of *(R)*-**107** shows an almost identical binding pose to (**9**) with the pyridyl nitrogen solvent exposed (Figure 6.1C). The initial work within this chapter looks to explain these differences in binding affinity through docking and MD studies.

6.1.2 CREBBP ligands show differing selectivity over BRD4(1)

CREBBP is another BCP, that does not belong to the BET family. It functions as a lysine acetyltransferase (KAT) and is vital for normal human development.[50] As such, it has been linked to a range of cancers.[48] A series of ligands have been published that bind to the BRD of CREBBP. The ligands contain three distinct features: i) a KAc mimic that binds the KAc reading N1168; ii) a tetrahydroquinoline (THQ) group that forms a cation-π interaction with R1173 and sits on the LPF shelf (formed by residues L1109, P1110, F1111), an equivalent of the WPF shelf; iii) a linker that joins these two groups.[50, 52] These have been highlighted in Figure 6.2.

Work by Dr M. Brand identified that 7-methoxy-THQ was the optimal π-donor group, based on DFT studies of the the electrostatic surface potential.[245] These groups can be seen in the ligands shown in Figure 6.2. The KAc mimic was attached to the linker using an amide group, seen in **109**. It was hypothesised that the internal H-bond present in **109** (Figure 6.2) was broken upon binding to the proteins, given the change in dihedral angles of the bound and unbound crystal structures.

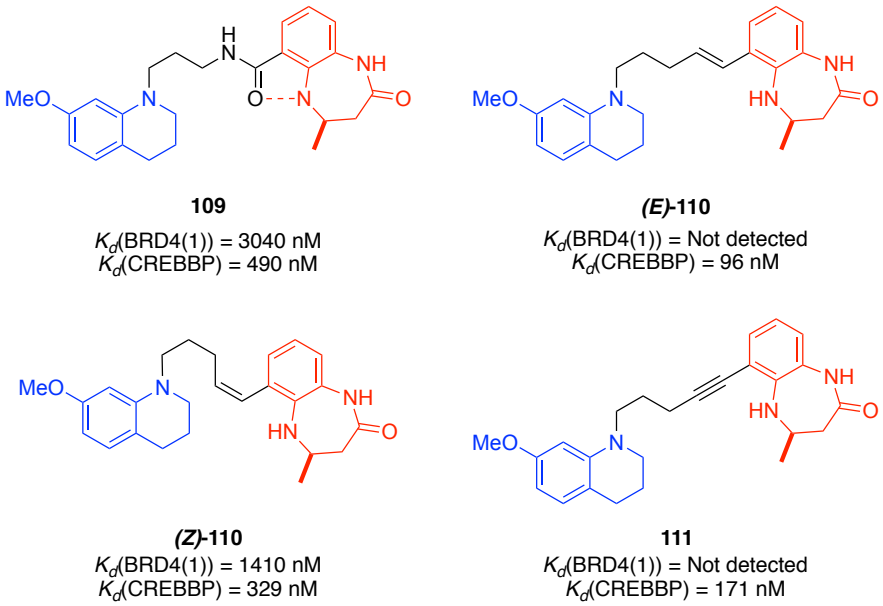


Figure 6.2: Ligands optimised for the CREBBP bromodomain. All ligands contain three components: i) The optimised 7-methoxy-tetrahydroquinoline (THQ) group that interacts with R1173 via a cation-π interaction; ii) a 4,5-dihydrobenzodiazepione group that acts as the KAc mimic; iii) a linker between the two domains containing an amide or amide bioisostere. K_d values were determined using Dr M. Schiedel and M. Morologu *via* ITC. The KAc containing ring has been highlighted in red, the linker in black and the cation-π forming THQ group is in blue.

Therefore, additional work was performed by Dr. J. Clayton to replace this group with a bioisostere that did not have this hypothesised enthalpic binding penalty. This led to two alkene isomers, *(E)*-**110** and *(Z)*-**110** and an alkyne, **111** (Figure 6.2).

The ITC data for both the CREBBP BRD and BRD4(1) showed varying degrees of ligand selectivity for CREBBP over BRD4(1), all dependent on the linker used. At the start of this project, only two crystal structures were available through collaboration with Dr. P. Filippakopoulos (The Structural Genomics Consortium, Oxford). The first was an unusual structure of two *(E)*-**110** ligands binding two CREBBP proteins, with each KAc bound ligand interacting with the R1173 of the cognate protein (Figure 6.3A). An additional structure showed **109** bound to CREBBP, forming the expected interactions of the THQ group with R1173 and the KAc mimic bound in the BRD site and forming a H-bond with N1168 (Figure 6.3B).

The later work presented in this chapter aimed to confirm the binding modes

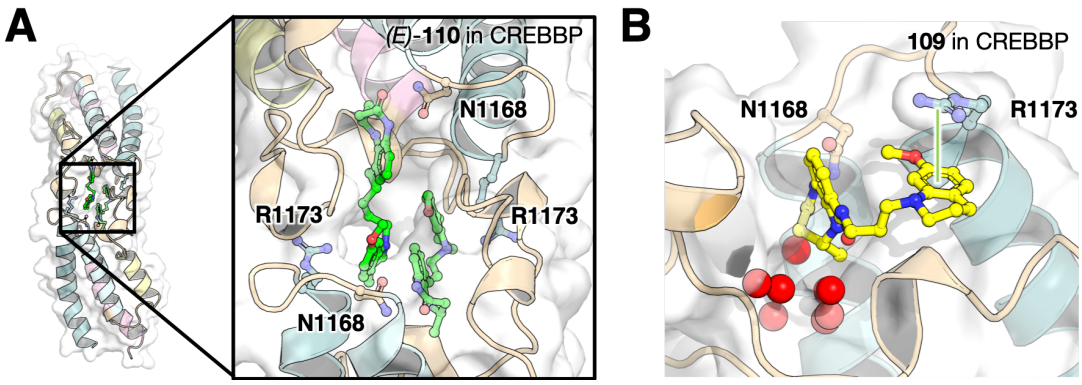


Figure 6.3: Crystal structures of CREBBP binding *(E)*-**110** and **109**. (**A**) Dimeric crystal structure of *(E)*-**110** binding CREBBP, where ligands interact with both proteins. (**B**) Crystal structure of **109** binding the CREBBP BRD, where the cation-π interaction has been highlighted in green. Structures were produced through collaboration with P. Filippakopoulos and ligands were synthesised by Dr J. Clayton and Dr M. Brand.

of *(E)*-**110**, *(Z)*-**110** and **111** in CREBBP, as the crystallographic data were either ambiguous or missing. Modelling was then employed to understand the possible ligand binding modes of **109**, *(E)*-**110**, *(Z)*-**110** and **111** to BRD4(1) and rationalise the differences in affinity.

6.2 The role of intramolecular hydrogen bonds on BRD4(1) binding affinity of OXFBD02 derivatives

Substitution of the phenyl group of OXFBD02 (**9**) for a 2-pyridyl group, **108**, allows for an internal H-bond to be formed, which is not possible within the 3-pyridyl analogue, **107**, shown in Figure 6.4. ^{1}H NMR studies have shown that this H-bond is present in the solution state structure of **108**.[231] Given the drop in affinity between the 2- and 3-pyridyl compounds, it was hypothesised that if the intramolecular bond was retained upon binding it would place the pyridyl group in an unfavourable orientation for interaction with the WPF shelf. To probe this hypothesis, simulations were performed on both enantiomers of **107** and **108**, where initial ligand coordinates for the *(R)*- and *(S)*- enantiomers were taken from PDBs 4J0R and 4J0S, respectively.[40]

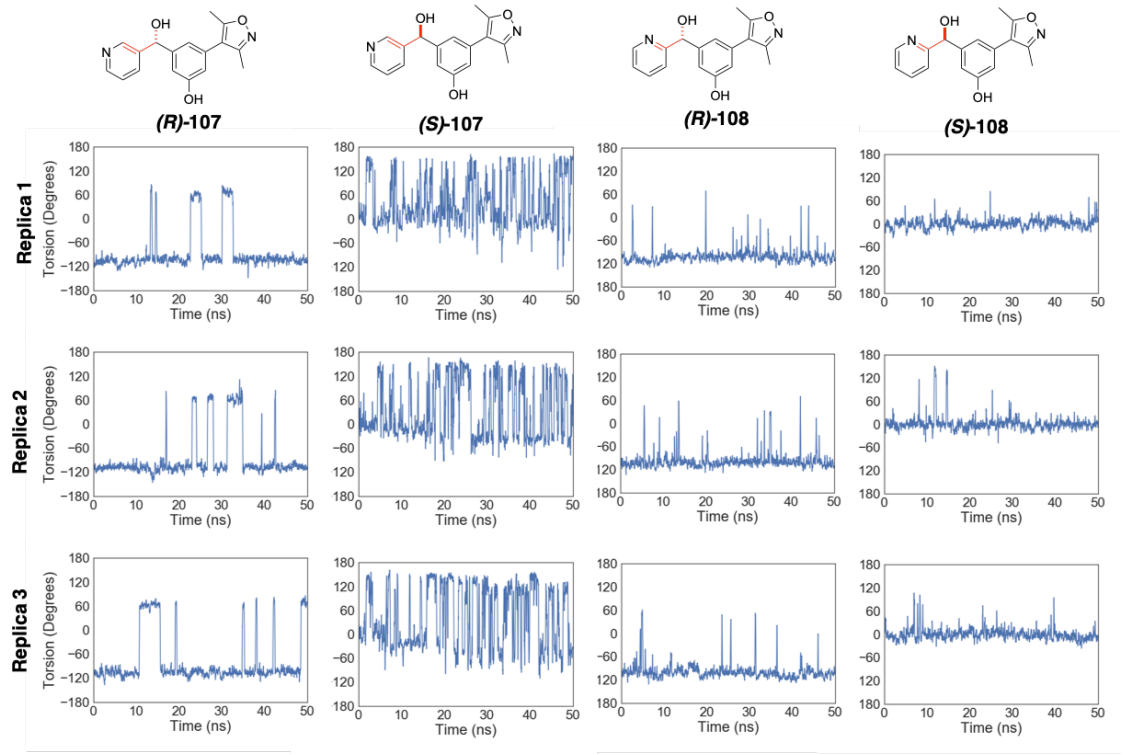


Figure 6.4: Torsions of protein bound **107** and **108**. (**A**) *(S)*- **108** which held the intramolecular H-bond. (**B**) *(S)*-**107** which did not have an internal H-bond. (**C**) *(R)*-**108** which has broken the intramolecular H-bond. (**C**) *(R)*-**107**. The torsion was determined between the four atoms indicated in red during simulations in triplicate.

The torsions of the hydroxyl group to the 2-position are shown in Figure 6.4, where the intramolecular H-bond would cause a torsional angle of 0 degrees. A torsional angle of 0 degrees is needed for the *(S)*-enantiomer, while a torsional angle of approximately 110 degrees is needed for the *(R)*-enantiomer to bind the WPF shelf optimally. The plots also show that the phenyl groups are capable of the occasional ring flip. The average dihedral angle of 0 degrees for *(S)*-**108** shows that the intramolecular H-bond is retained upon binding, and so could act to pre-organise the ligand prior to binding. This orientates the pyridyl nitrogen atom towards the solvent and away from the hydrophobic WPF shelf, which is likely to increase the ligand's affinity and is the orientation seen in the crystal structure (Figure 6.1C). A similar effect is also seen with the I-BET151 (**11**) ligand (PDB ID: 3ZYU). However, in *(R)*-**108** it is shown that the intramolecular H-bond would be broken to interact optimally with the WPF shelf and allowing the nitrogen atom to

point away from W81. This implies that there has either been an enthalpic penalty to binding, caused by the need to overcome this H-bond, or that the intramolecular bond is not broken and hence there is a sub-optimal interaction with the WPF shelf. The comparably higher affinity of **107** can be rationalised as the 3-pyridyl compound does not suffer from these penalties.

6.3 Validation of a cation-π interaction guiding CREBBP binding

Prior to this work, there were only two unpublished structures of the ligand series bound to CREBBP (Figure 6.2) which are similar to structures seen of previous CREBBP ligands.[50] The structure with **109** bound (unpublished - Figure 6.3) shows the key interaction of the THQ group with R1173. However, the structure of *(E)*-**110** (unpublished) formed a dimeric complex, rather than the monomeric complex expected. Is was hypothesised that this was an artefact and size exclusion chromatography indicated that protein was purified as a monomeric structure (work by Dr M. Schiedel). To identify whether *(E)*-**110** was capable of forming the monomeric binding pose seen in the **109** structure, a single monomeric unit's co-ordinates were isolated from the crystal structure, with *(E)*-**110** interacting *via* the KAc-binding pocket and N1168.

Figure 6.5A shows the ligand pose of *(E)*-**110** in CREBBP when a monomeric unit was isolated from the dimeric co-crystal structure along with a pose from equilibrium MD simulations, showing that the expected binding pose was recovered. This is also shown in Figure 6.5B, where the distance of the THQ aromatic ring to P1110, part of the LPF shelf underneath the R1173 cationic residue was measured. The simulations show that two of the three simulations settle with the THQ group residing over P1110, at distances similar to that seen in the crystal structure of **109** bound to CREBBP. Note that selecting atoms that best represent the cation-π interaction was difficult and gave a poor representation of the binding events.

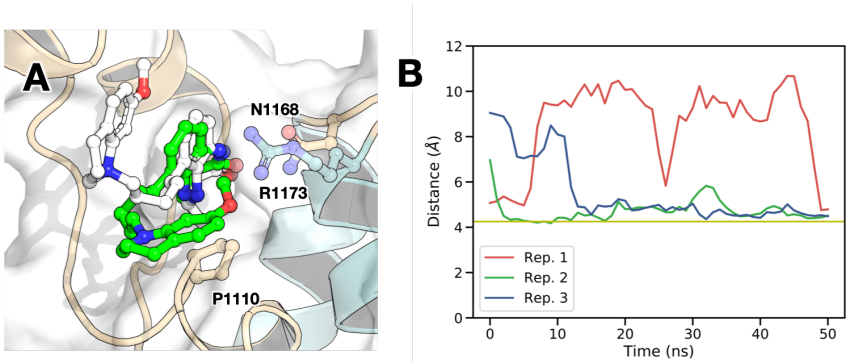


Figure 6.5: Recovery of the electrostatic interaction with R1173 through MD simulations. (**A**) The initial (white) and equilibrium (green) MD pose of *E*-**110**. (**B**) The distance between P1110, the residue under R1173, and THQ group of *(E)*-**110** drops close to 4.2 Å (yellow line) during equilibrium MD simulations. Simulations were carried out in triplicate.

6.3.1 MD showed the cation-π interaction is stable for all CREBBP ligands

Docking studies identified potential binding modes for *(Z)*-**110** and **111** bound to CREBBP. These poses, along with the equilibrium pose of *(E)*-**110** all mirrored the structure of **109** bound to CREBBP, shown in Figure 6.6A-D. It should also be noted that two conformers of the 7-membered ring within the KAc mimic were identified in the initial conformer search, however, all co-crystal structures show only the higher energy ring conformer bound within CREBBP. These differ from the small molecule crystal structures identified by Dr J. Clayton, shown within the Appendix (Figure D.1). The studies in this chapter therefore enforced the co-crystallised ring conformer for docking and MD studies.

The docked poses underwent triplicate MD to ensure that the binding poses were stable and the cation-π interaction was retained. During ligand parameter assignment with the GAFF forcefield the amide nitrogen in the 7-membered ring was manually modelled as sp^3, despite being adjacent to an aromatic system.[159] This observation was based on the small molecule crystal structures, which show a trigonal bipyramidal nitrogen centre. These structures are shown in the appendix, Figure D.1.

The distance of the THQ group in each ligand to P1110 in MD simulations is shown in Figure 6.6E-H. These graphs show that the THQ groups spend the

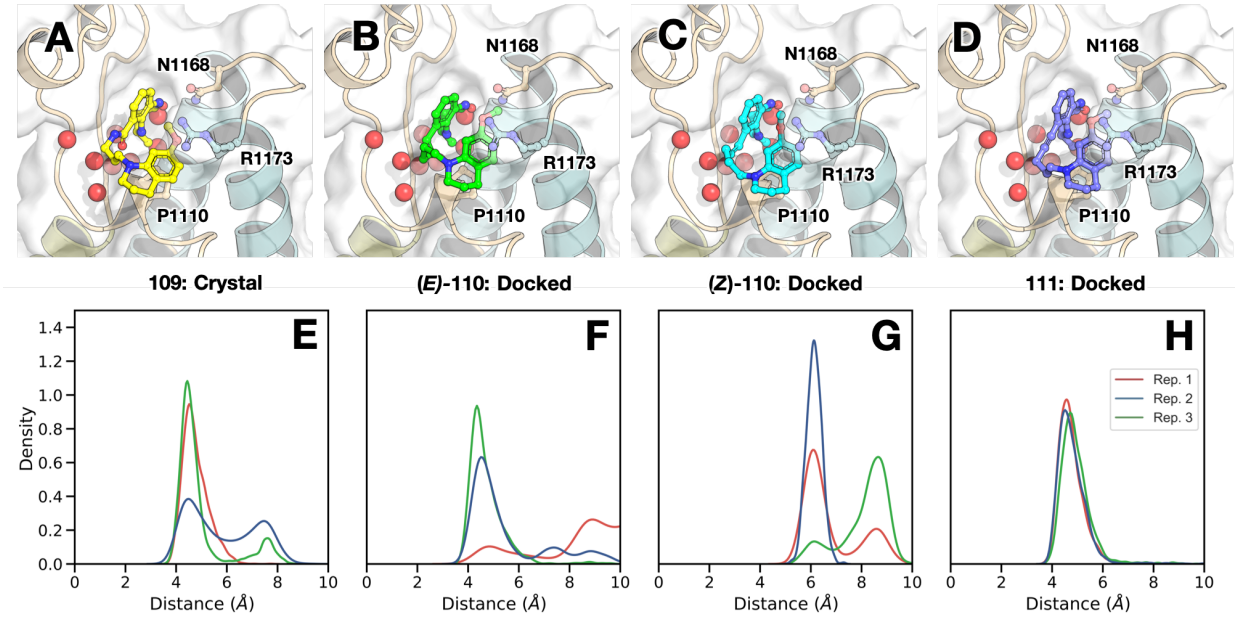


Figure 6.6: CREBBP ligands interact with P1110. (**A**) The crystal structure of **109** bound to the CREBBP BRD. (**B-D**) Docked structures of *(E)*-**110**, *(Z)*-**110** and **111** bound to the CREBBP BRD. Structures were subjected to 50 ns of MD simulation, in triplicate. (**E-H**) Distances to the LPF shelf are shown in the kernel density estimate (KDE) plots, where the distance is measured from the aromatic ring in the THQ group to P1110.

majority of each simulation residing in contact with P1110. Where the distance increases by around 3 Å this corresponds to the THQ aromatic system moving to an alternative pose, where this group is located near L1109. The first replicate of the simulation of *(E)*-**110** binding CREBBP, Figure 6.6F, showed ligand dissociation and hence this replicate can be ignored for analysis. **111** shows no movement to L1109, while simulations of both weaker binders, **109** and (*Z*)-**110**, show this movement away from P1110, and hence R1173, in two of the three replicas.

6.4 Alternative binding modes to BRD4(1) lead to steric clashes

ITC data for ligands **109**, (*Z*)-**110**, (*E*)-**110** and **111** revealed surprising binding data for BRD4(1). Both (*E*)-**110** and **111** showing either no binding to BRD4(1). Meanwhile, (*Z*)-**110** and **109** showed weaker binding with K_d values of 1.4 µM and

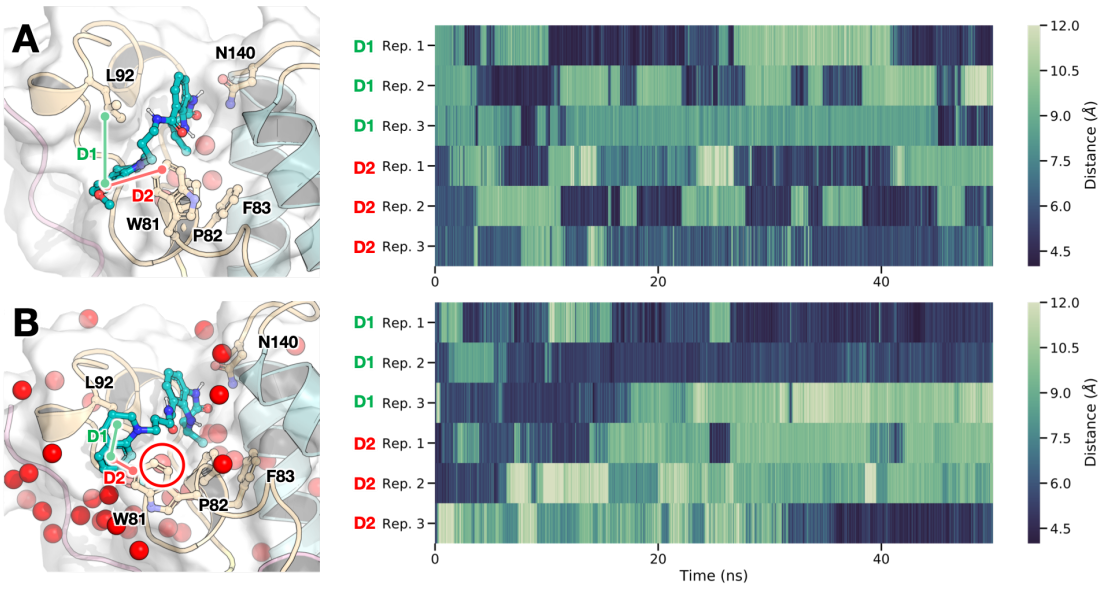


Figure 6.7: Interactions of **109** docked and crystallised in BRD4(1). Both the docked and crystallised structures of **109** in BRD4(1) were subjected to MD simulations. The distances (D1,D2) of the aromatic system in the THQ group were measured against the edge of W81 and the centre of L92. These distances are shown as barcode plots for each of the three simulation repeats. (**A**) Docked structure. (**B**) Crystallised structure, including all crystallographic water molecules and W5 highlighted *via* a red circle.

3.0 µM, respectively. This shows the ligands hold differing degrees of selectivity to CREBBP over BRD4(1), as shown in Figure 6.2. To understand the differences, molecular docking was employed to predict the binding modes of the ligands. Docking identified the THQ group residing in the ZA channel and forming a hydrophobic interaction with L92, instead of on the WPF shelf in the equivalent position to its binding mode in the CREBBP BRD.

Docking of **109** to BRD4(1) (Figure 6.7A), identified that the THQ could reside in a shallow pocket adjacent to L92. MD simulations showed that the THQ group was capable of forming a hydrophobic interaction with L92 or a side on π-π interaction with W81, demonstrated by the mutually exclusive dark patches in the barcode plots of D1 and D2 in Figure 6.7A, respectively. After the initial studies were performed, a co-crystal structure of **109** bound to BRD4(1) was solved. This showed the same interaction of the THQ group with L92. MD simulations of the crystallographic structure showed that the THQ group resided closer to L92. This

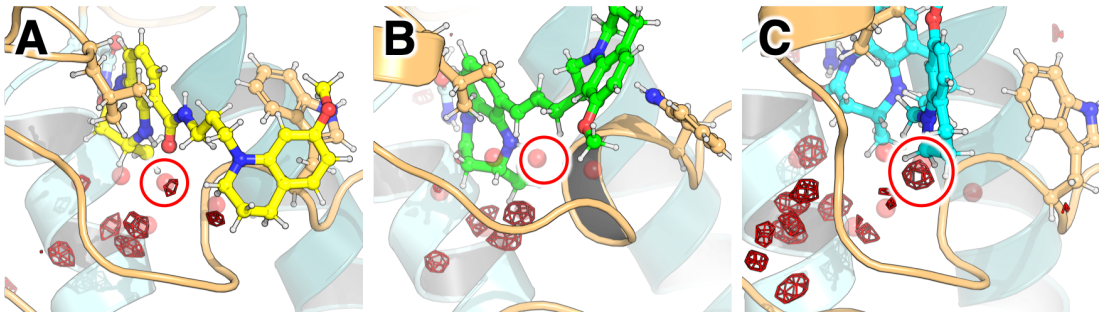


Figure 6.8: Water densities for **109** and **110**. Snapshots (at 25 ns) from MD simulations of (**A**) **109** crystallised, (**B**) (*E*)-**110** docked and (**C**) (*Z*)-**110** docked, with areas of high water density shown as a red mesh. Crystallographic waters of BRD4(1) have been included as transparent spheres, with water molecule W5 highlighted with a red ring.

is likely due to the initial rotamer of W81 in the simulations, which is more solvent exposed and away from the THQ group in the crystal structure.

Docking of *(E)*-**110** showed an identical ligand placement to **109**. However, MD simulations identified that the linker did not hold a stable orientation relative to the KAc mimic, shown through the comparison of dihedral angles in Figure 6.9. Further analysis of the MD simulations focused on the ZA channel waters. Figure 6.8A shows areas of high water density during a simulation of **109**, with the ZA channel water positions remaining heavily occupied during the simulations and capable of forming a H-bond with the linker amide carbonyl. When compared to the water occupancy of (*E*)-**110**, Figure 6.8B, the ZA channel waters densities are absent and show that waters have been displaced by the alkene proton.

Docking studies of **110** identified two poses, Figure 6.10A and B, where the THQ ring could reside against either L92 or W81. Both poses were subjected to MD studies, which showed that both placements of the THQ group allowed it to form interactions with: i) L92; ii) W81; iii) P82, after moving over W81. These can be observed as the mutually exclusive dark patches in Figure 6.10. Both of these poses hold the ZA alkene bond perpendicular to the KAc mimic and hence do not lead to displacement of the ZA channel water molecules (as shown in Figure 6.8C).

Despite mirroring the interactions with L92 during docking studies, **111** was unable to sustain interactions with either L92 or W81 during simulation. The

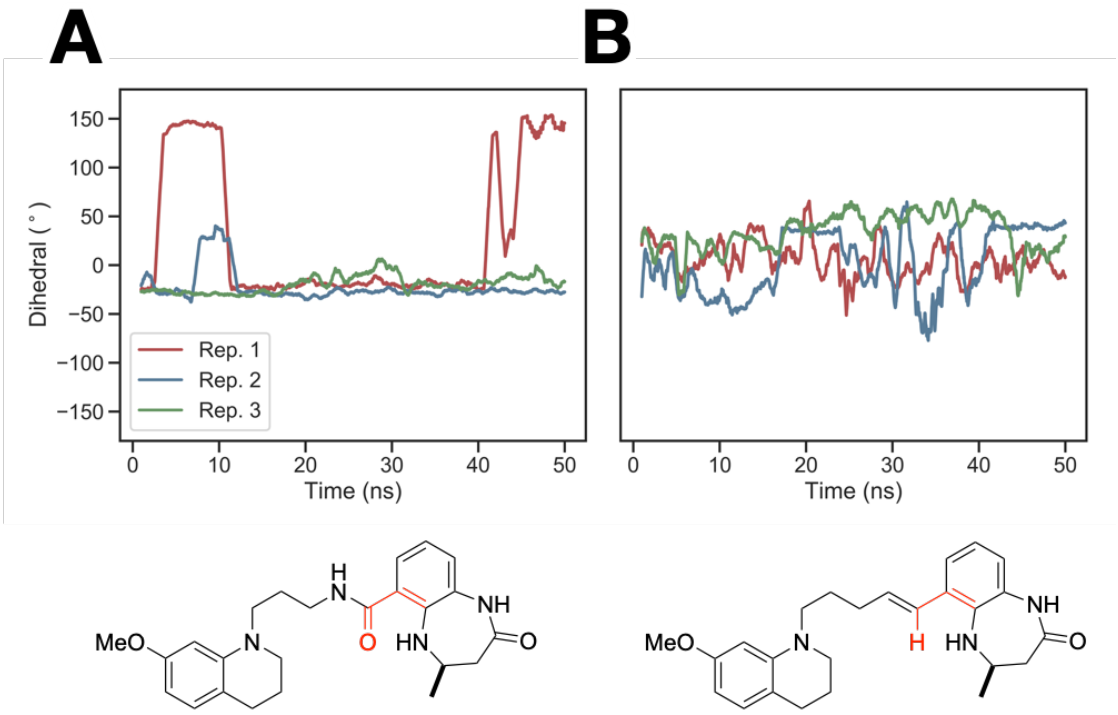


Figure 6.9: Dihedral angles within MBX01 and JCX03E during binding of BRD4(1). The dihedral angles within (A) MBX01 and (B) JCX03E are shown in red on the structures below.

instability of the ligand can be seen through its high ligand RMSD during simulations, as shown in appendix Figure D.3.

6.5 Discussion

The work within this chapter has shown the importance of modelling within medicinal chemistry projects. Especially where co-crystal structures cannot be obtained, or contain artefacts. This chapter uses molecular docking, complimented by molecular dynamics simulations, to increase our understanding of ligands that bind to CREBBP and BRD4(1).

Firstly, these methods were employed to understand why there was a decrease in affinity for **108**. This was not obvious from the crystal structures of OXFBD02 (**9**) and *(R)*-**107** and required a modelling approach. Using MD to study the systems allowed for the optimal orientation of the pyridyl systems to be identified. Notably,

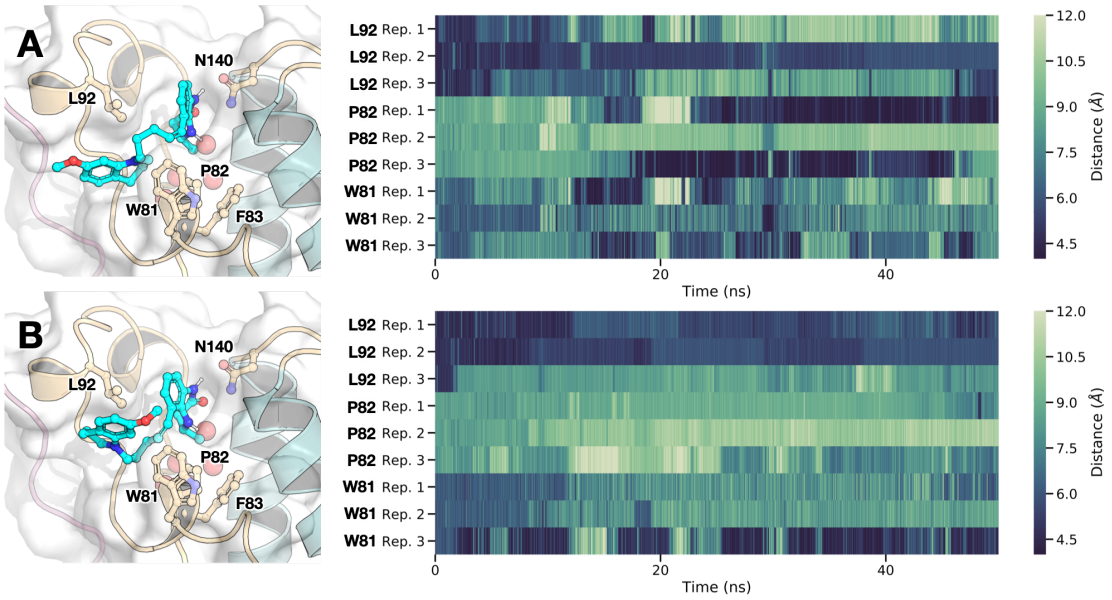


Figure 6.10: Interactions of docked *(E)*-**110** in BRD4(1) in MD simulations. Two poses of *(Z)*-**110** were predicted for BRD4(1) from molecular docking studies. (**A**) shows the THQ group interacting with L92. (**B**) Shows a side-on π-π interaction between the THQ group and W81. Both poses were subject to 50 ns simulations in triplicate. The barcode plots show the movement of the THQ ring to interact with L92, W81 and P82 during the simulations.

analysis showed that the H-bond is retained in the (*S*)-enantiomer of **108** and could bind BRD4(1) without breaking an internal H-bond. The enthalpic penalty of breaking the intramolecular bond to bind one enantiomer explains the low affinity of the racemic mixture. However, this also implied that there should be a large difference in affinity between enantiomers. Separation of enantiomers and testing could confirm this hypothesis, where it would be predicted that (*S*)-**108** would bind with a high affinity, given the pre-organisation of the intramolecular H-bond.

Similar *in silico* studies were important in elucidating the differences in binding affinity of **109**, (*Z*)-**110**, (*E*)-**110** and **111** to BRD4(1). (*E*)-**110** is a non-binder and therefore it was not possible to use co-crystallisation to understand the structural features that prevented BRD4(1) binding. The work in Chapter 4 showed the importance of the ZA channel waters upon binding, and led to a similar approach within this work. This highlighted the importance of BRD4 binding of **109** and (*E*)-**110** on W5 within the ZA channel. This water molecule has been shown to

be key for binding of **9**, where the phenyl alcohol group binds *via* W5,[40] shown in Figure 6.1B and C. Further validation could be achieved through performing GCMC calculations on the ZA channel waters, as seen in § 4.3.

Finally, molecular dynamics was important in understanding the binding of (*E*)-**110** to CREBBP. Given the dimeric co-crystal structure, it was not possible to determine if the cation-π interaction with R1173 was retained. Luckily, MD studies showed that this interaction was restored, placing the THQ group on the LPF shelf and forming a cation-π interaction with R1173.

6.6 Conclusions

The work presented in this chapter has highlighted the importance of key H-bonds in binding BRD4(1). When investigating pyridyl derivatives of **9** through molecular dynamics simulations, it was shown that the binding energy of (*R*)-**108** would be enthalpically impacted by the breaking of the intramolecular H-bond to bind the WPF shelf. This intramolecular H-bond does not impact the (*S*)-enantiomer. Enantioselective synthesis of **108** and ITC testing would confirm this model. The absence of this intramolecular bond in the 3-pyridyl analogue retained the metabolic benefits of having a pyridyl system whilst removing the intramolecular interaction.

Studies of a series of CREBBP ligands showed that H-bonds are also important when binding BRD4(1), where the inability of the (*E*)-alkene to H-bond to the ZA channel waters relative to the amide linker prevents binding. As a non-binder, this observation would not have been possible without *in silico* models. Meanwhile, the (*Z*)-alkene, **110**, is still capable of binding to BRD4(1). This is possible by having the alkene orthogonal to the KAc mimic and forming hydrophobic interactions between the THQ group and the WPF shelf or L92. Hence, selectivity for CREBBP over BRD4(1) is achieved through displacement or retention of the ZA channel water molecules. Studies of these ligands in the CREBBP BRD showed that all bio-isosteres of the amide linker are capable of placing the electrophilic THQ group on the LPF shelf and in close proximity to R1173, likely leading to a cation-π interaction.

To succeed, jump as quickly at
opportunities as you do at conclusions.

— Benjamin Franklin

7 Summary

The work within this thesis looked at how *in silico* tools could be used to aid ligand design prospectively in various medicinal chemistry projects. Within these projects a lack of co-crystal structures, or sometimes even *apo* structures, made understanding the SAR within each project difficult. However, though a combination of computational methods and iterative drug ligand design, the SAR could still be explored. The models created and refined were key to identifying the highest affinity ligands known for the TRIM33β bromodomain, tool compounds within parasite bromodomains and understanding how waters within BRD4(1) can impose non-BET selectivity on CREBBP ligands.

Chapter 3 introduced four ligands that had previously been identified as binders of the TRIM24/33 tandem PHD/BRD cassette. After re-synthesis and co-crystallisation attempts on these ligands failed, a computational modelling approach was adopted. Several potential binding sites were identified on the cassette: the BRD binding site, the region inbetween the domains, the K4 binding site of the PHD and the 'underside' of the experimental construct. Docking studies agreed with orthogonal mutant assay data that the benzimidazolone containing ligand bound in the BRD site. Two binding modes of this ligand were identified, both forming a hydrogen bond between the benzylic amine and E981. The urea containing ligands were identified as being capable of binding to the construct either on the

'underside' of at the K4 binding region of the PHD, though further mutant studies are needed to confirm these computational predictions. The studies were unable to identify the binding regions of the fourth ligand.

The benzimidazolone ligand was further explored within Chapter 4. Within the TRIM33β binding site, several key residues were identified as important for ligand development. The most important is E981, which is adjacent to the ZA channel. This residue offers both a means of selectivity and affinity. A second polar residue, E984, was a tempting target to obtain a second polar interaction. However, extension of the benzylic amine with an additional amine did not show a strong interaction with this residue. Free energy calculations on the ligand poses showed that this extended vector favoured pointing towards the binding site. Further free energy calculations on the ZA channel waters molecules identified one stable molecule that would be displaced by E984 interacting with the introduced amine. Instead the new amine group showed interactions to E981 with the benzylic amine. This new vector, now containing two amine groups, gave an ambiguous protonation state. However, MD studies showed that if both amines are charged then the ligand binding can be unstable. Future studies should look to optimise the interactions of this vector with E981.

With a lack of polar interactions, additional affinity and selective had to come from shape complimentary hydrophobic interactions in the TRIM33β binding site. The hydrophobic region above the binding site, in the vicinity of I990, was shown to be viable for forming hydrophobic interactions. Unfortunately, F1038 was shown to form cap above the binding site, through a π-π interaction with Y993. This residue means that building large bulk above the benzimidazone core, in the vicinity of I990 comes at the expense of the binding site stability, which was validated experimentally. This offers two positions on the benzimidazolone core to functionalise. Future studies should focus on the 7-position, directly above the benzimidalazone core. This position would still allow hydrophobic interactions but not cause strain to F1038. Designing a range of hydrophobic groups at this position and subsequent ABFE calculations would determine if this was a viable route.

Developing ligands for parasite bromodomains was explored in Chapter 5, as a putative target in tropical disease treatment. These studies focused on bromodomains with *Trapanasoma cruzi* and *Schistosoma mansoni*, which at the start of the studies had no *apo* or co-crystal structures. After generating stable homology models and validating the water network is retained, docking studies of human bromodomain ligands known to bind were performed. Studies of *Tc*BDF3 focused on the ligand BI-2536, which MD studies showed had a low affinity due to the ligand being more solvent exposed relative to BRD4(1). The binding mode observed from these studies were used to design tool several tool compounds that can be used in future ligand testing. Through screening of amide containing analogues, it was shown that additional affinity and selectivity could be obtained from improved shape complementarity to the ZA wall, and through hydrophobic or cation-π interactions beyond the binding site. Future work should look at synthesis and testing of these designs, to increase confidence in the model and remove BRD4(1) binding. Homology models of *Sm*BRD3(2) needed further refinement to achieve the induced protein-ligand fit. However, these models showed that a hydrophobic shelf was present and able to bind a sub-set of ligands designed for interaction with BRD4(1)'s WPF shelf. Obtaining selectivity over BRD4(1) could be achieved by exploiting the differences in hydrophobic shelf. Hence, the geometry of some ligands was not compatible with the shelf positioned slightly closer to the α_C helix. This model was later validated by a co-crystal structure, as was the model of *Tc*BDF5(1).

The final chapter looked at understanding the SAR of two sets of ligands. The first were a set of analogues of OXFBD02, which replaced the phenyl ring with pyridyl groups. Through MD studies, it was shown that the (*S*)-enantiomers of these ligands kept the ring in the same plane as the hydroxyl group on the linker, but the (*R*)-enantiomers were not. This would cause the disruption of an internal hydrogen-bond in the 2-pyridyl analogue and hence caused a surprising reduction in affinity. The second set of ligands were designed to bind to the BRD of CREBBP. MD studies were able to show that a cation-π interaction between the ligands and R1173 was retained in analogues, despite a very curious co-crystal structure.

Docking studies to BRD4(1), which were partially confirmed by a later co-crystal structure, showed the ligands adopted different binding modes in this BRD. This new binding mode placed the linking group in the ZA channel. Further MD studies showed that the (*S*)-alkene analogue of the original amide group disrupted the water network within the ZA channel while the amide was able to interact with it *via* the carbonyl. This difference allows the alkene to be selective over BRD4(1).

Overall, application of computational methods have allowed SAR models to be generated and guided important synthetic decisions. The models will continue to be important in continuing to optimise ligands in TRIM33, parasite bromodomains, CREBBP and BRD4(1), especially in the absence of co-crystal structures. Interestingly, within all the chapters, it has been shown that the ZA channel water molecules are important factors in bromodomain ligand design. These waters, which have been shown to be stable or displaceable depending on the protein, are often omitted from discussions on bromodomain structure. However, it is clear that further work is needed to map out the different waters and their potential role in ligand binding.

Appendices

A

Appendix for Chapter 3

Table A.1: AlphaScreen™ H3 peptide affinities for TRIM proteins. Peptide affinities were determined using ITC, performed by Dr L. See.[204] These differing affinities of the peptides means IC_{50} values from AlphaScreen™ assays are not comparable.

		H3(PHD)	H3(BRD)	H3(Dual)
TRIM24	N	0.892	0.640	0.647
	K_D	3.90 µM	695 nM	309 nM
	ΔH (kcal/mol)	−5.51	−10.8	−14.7
	ΔG (kcal/mol)	−7.38	−8.40	−8.88
	TΔS (kcal/mol)	−1.87	2.44	5.77
TRIM33α	N	0.496	0.158	0.598
	K_D	9.46 µM	54.7 µM	5.58 µM
	ΔH (kcal/mol)	−2.94	−80	−8.80
	ΔG (kcal/mol)	−6.85	−5.82	−7.17
	TΔS (kcal/mol)	−3.91	74.2	1.63
TRIM33β	N	0.547	0.411	0.449
	K_D	9.25 µM	4.92 µM	1.11 µM
	ΔH (kcal/mol)	−3.65	−10.5	−10.5
	ΔG (kcal/mol)	−6.87	−7.24	−8.12
	TΔS (kcal/mol)	−3.22	3.22	2.41

Table A.2: The table below summarises all the structural information on the relevent TRIM proteins, correct March 2018. The structure of TRIM33β was donated by Stefan Knapp and is not yet avaliable on the PDB. All structures are determined through X-Ray crystallography, with all TRIM33α structures containing an unresolved BC loop.

Protein	PDB ID	Description	Resolution (Å)
TRIM24	3O33	*apo*	2.0
	3O34	H3K23Ac	1.9
	3O35	H3K27ac	1.76
	3O36	H3K16Ac	1.7
	3O37	H3K4	2.0
	4YAB	*Holo* structure	1.9
	4YAD	*Holo* structure	1.73
	4YAT	*Holo* structure	2.18
	4YAX	*Holo* structure	2.25
	4YBM	*Holo* structure	1.46
	4YBS	*Holo* structure	1.83
	4YBT	*Holo* structure	1.82
	4YC9	*Holo* structure	1.82
	4ZQL	*Holo* structure	1.79
	5H1T	*Holo* structure	1.95
	5H1U	*Holo* structure	1.90
	5H1V	*Holo* structure	2.00
TRIM33α	3U5M	*apo*	3.08
	3U5N	H3K9me3K14ac	1.95
	3U5O	H3K9me3K14acK18ac	2.70
	3U5P	H3K9me3K14acK18acK23ac	2.80
TRIM33β	5MR8	H3K9me3	1.74

A.1 TRIM24/33 sequences aligned

Sequences of TRIM24/33 recombinant proteins were aligned using ClustalOmega.[246]

```
TRIM24    MHHHHHHSSGVDLGTENLYFQSPNEDWCAVCQNGGELLCCEKCPK   846
TRIM33a   MHHHHHHSSGVDLGTENLYFQS..MDWCAVCQNGGDLLCCEKCPK   907
TRIM33b   MHHHHHHSSGVDLGTENLYFQS..MDWCAVCQNGGDLLCCEKCPK   907

TRIM24    VFHLSCHVPTLTNFPSGEWICTFCRDLSKPEVEYDCDAPSHNSEK   891
TRIM33a   VFHLTCHVPTLLSFPSGDWICTFCRDIGKPEVEYDCDNLQHSKKG   952
TRIM33b   VFHLTCHVPTLLSFPSGDWICTFCRDIGKPEVEYDCDNLQHSKKG   952

TRIM24    KKTEGLVKLTPIDKRKCERLLLFLYCHEMSLAFQDPVPLTVPDYY   936
TRIM33a   KTAQG...LSPVDQRKCERLLLYLYCHELSIEFQEPVPASIPNYY   994
TRIM33b   KTAQG...LSPVDQRKCERLLLYLYCHELSIEFQEPVPASIPNYY   994

TRIM24    KIIKNPMDLSTIKKRLQEDY.SMYSKPEDFVADFRLIFQNCAEFN   980
TRIM33a   KIIKKPMDLSTVKKKLQKKHSQHYQIPDDFVADVRLIFKNCERFN  1039
TRIM33b   KIIKKPMDLSTVKKKLQKKHSQHYQIPDDFVADVRLIFKNCERFN  1039

TRIM24    E.................PDSEVANAGIKLENYFEELLKNLYPE.  1007
TRIM33a   EMMKVVQVYADTQEINLKADSEVAQAGKAVALYFEDKLTEIYSDR  1084
TRIM33b   E.................ADSEVAQAGKAVALYFEDKLTEIYSDR  1067

TRIM24    ..............................................  1007
TRIM33a   TFAPLPEFEQEEDDGEVTEDSDEDFIQPRRKRLKSDERPVHIK  1127
TRIM33b   TFAPLPEFEQEEDDGEVTEDSDEDFIQPRRKRLKSDERPVHIK  1110
```

X acidic (−)
X aliphatic
X aliphatic (small)
X amide
X aromatic
X basic (+)
X hydroxyl
X imino
X sulfur

A.2 Docking parameters

Table A.3: Grid parameters used for AutoDock4. Grid spacing was 0.375 Å.

Protein	Region	x, y, z spacings	x, y, z co-ordinates
TRIM24	BRD	50, 50, 40	39.032, 36.136, -13.331
TRIM24	Cleft	50, 56, 50	48.267, 20.946, -16.445
TRIM24	K4	50, 56, 50	43.234, 1.887, -16.084
TRIM33α	BRD	50, 50, 38	39.098, 38.535, -12.352
TRIM33α	Cleft	56, 50, 48	45.855, 22.909, -13.765
TRIM33α	K4	56, 50, 50	44.266, 3.238, -16.097
TRIM33β	BRD	50, 50, 40	39.032, 36.136, -13.331
TRIM33β	Cleft	50, 56, 50	48.267, 20.946, -16.445
TRIM33β	K4	50, 56, 50	43.234, 1.887, -16.084

Table A.4: Grid parameters used for blind docking with AutoDock4. Grid spacing was 1 Åfor both programmes.

Protein	x, y, z spacings	x, y, z co-ordinates
TRIM24	58 36 50	-3.832 -22.955 18.065
TRIM33α	58 36 50	-3.832 -22.955 18.065
TRIM33β	58 46 50	-3.832 -18.895 18.065

A.3 Full docking results

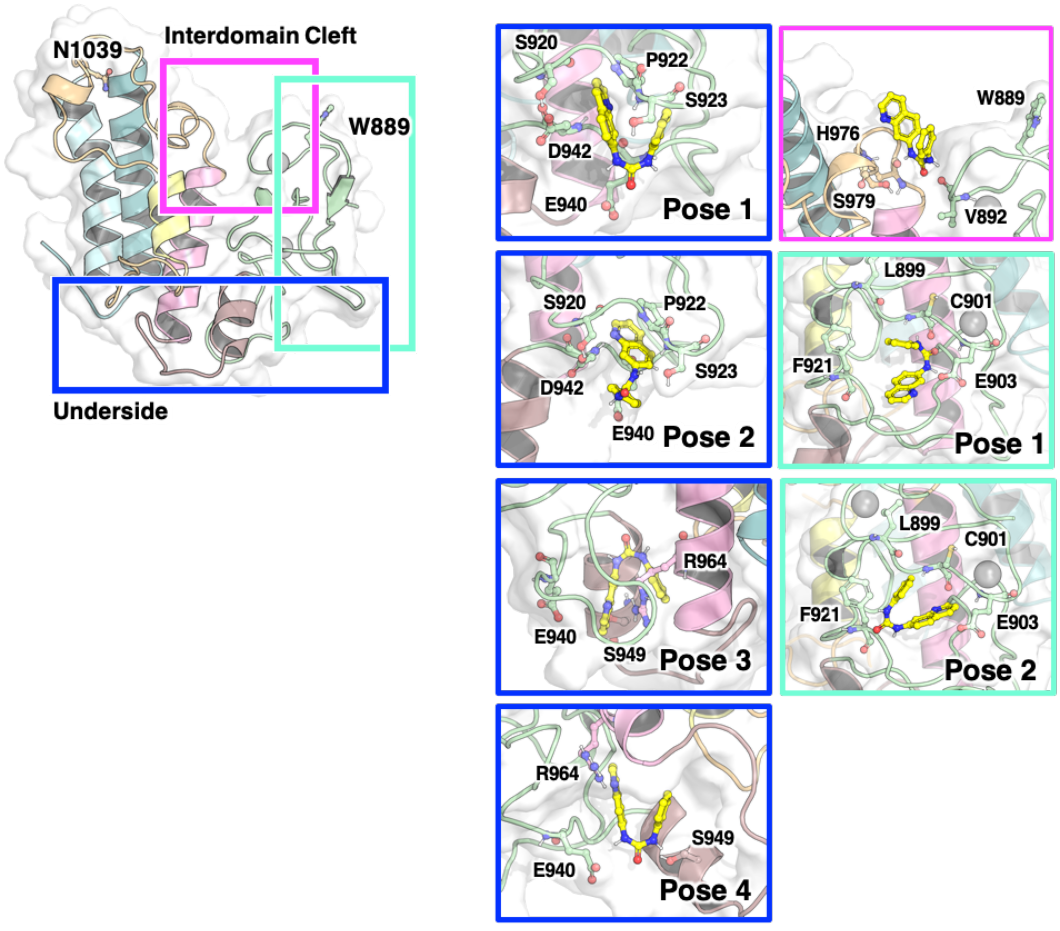


Figure A.1: Results of docking compound **27** against TRIM33β

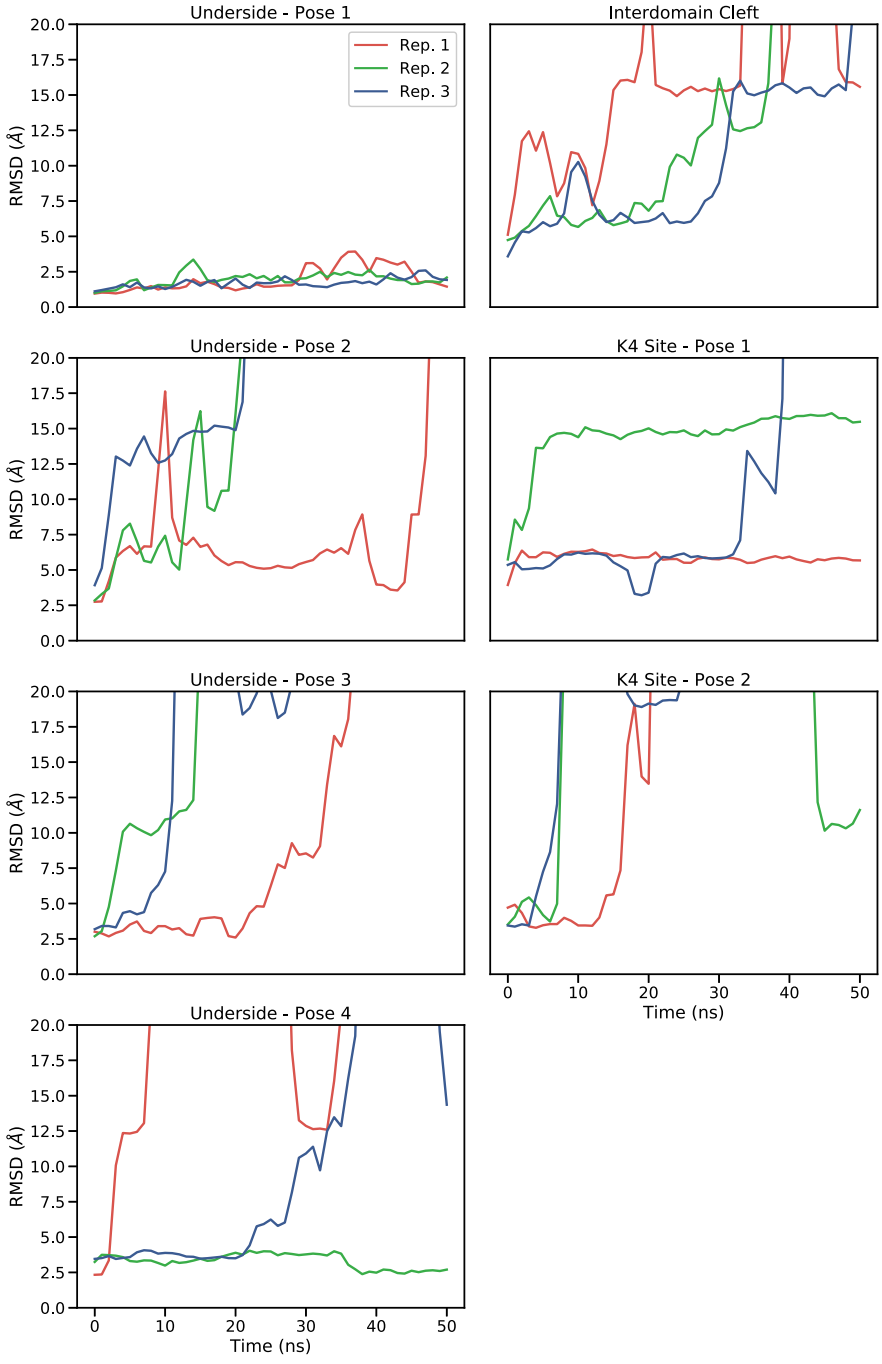


Figure A.2: MD studies on docked poses of **27** bound to TRIM33.

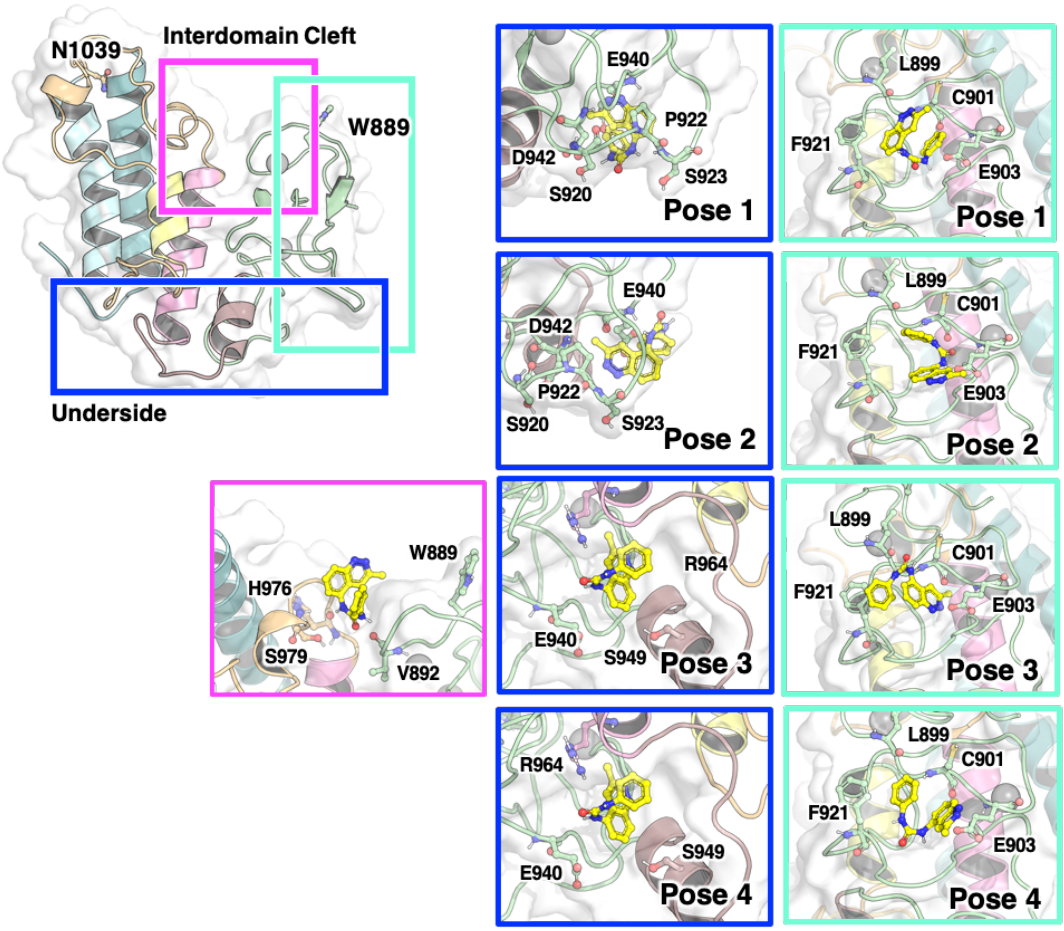


Figure A.3: Results of docking compound **28** against TRIM33β

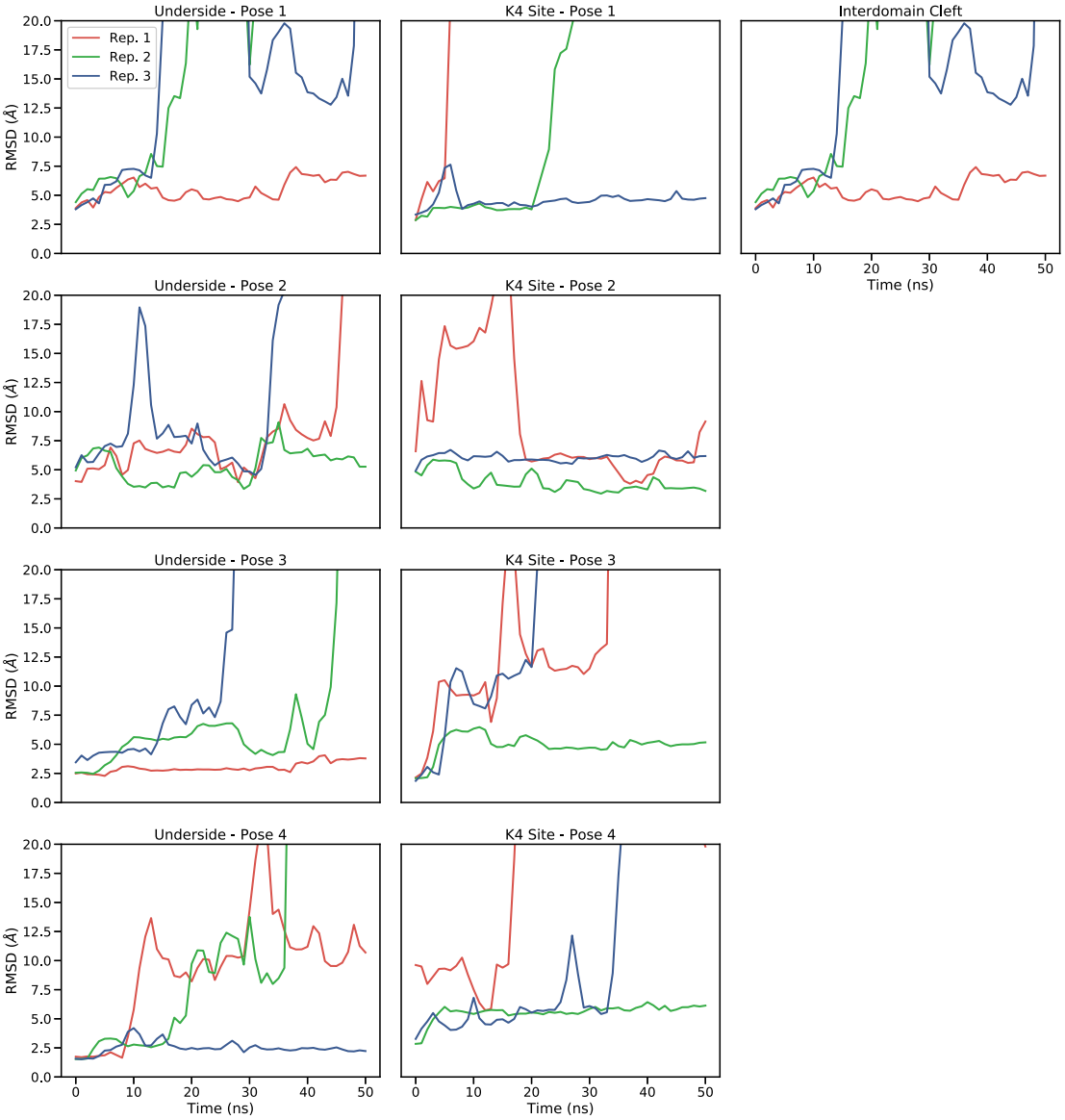


Figure A.4: MD studies on docked poses of **28** bound to TRIM33

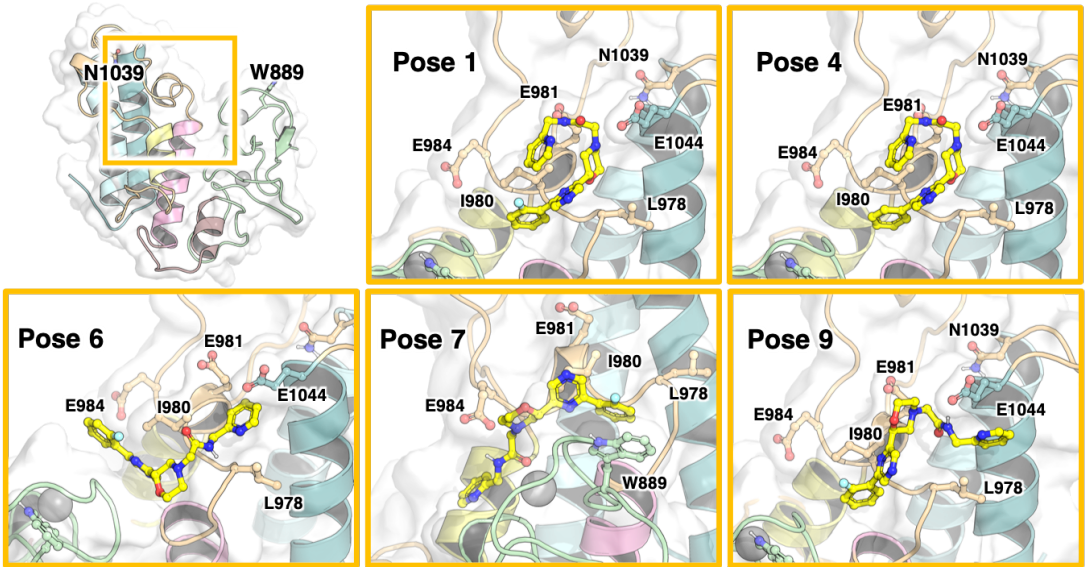


Figure A.5: Results of docking compound **28** against TRIM33β

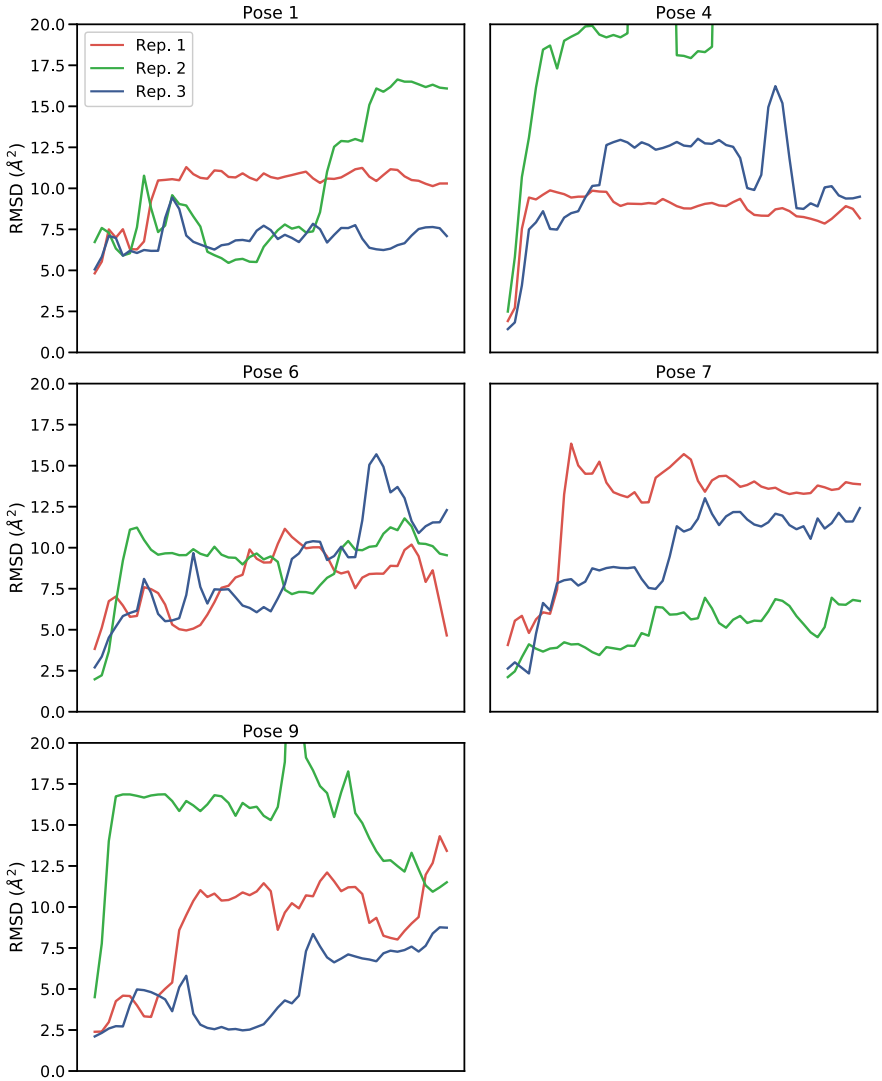


Figure A.6: MD studies on docked poses of **26** bound to TRIM33

B

Appendix for Chapter 4

Contents

B.1 GPU accelerated ABFE investigations

This section briefly overviews several observations from investigaing the use of GPU accelerated ABFE calculations on bromosporine (**19**) in BRD4(1). This experiment used the same co-ordinates and ligand topology used by Aldeghi *et al.* in 2016 [58], but the made alterations to the MD parameters described in Chapter 2. These alterations also included removal of the dummy atom co-ordinates by used by Aldeghi *et al.* These dummy atoms have no charge or mass, and represented the final 'annihilated' ligand. The requirement to convert to dummy atoms was made redundant in later versions of GROMACS, which allowed a direct removal of the ligand to the gas phase. Comparison of the energy terms between the GPU accelerated protocol an the protocol by Aldeghi *et al.* highlighted differences in the energies calculated from total ligand annihilation/recoupling compared to

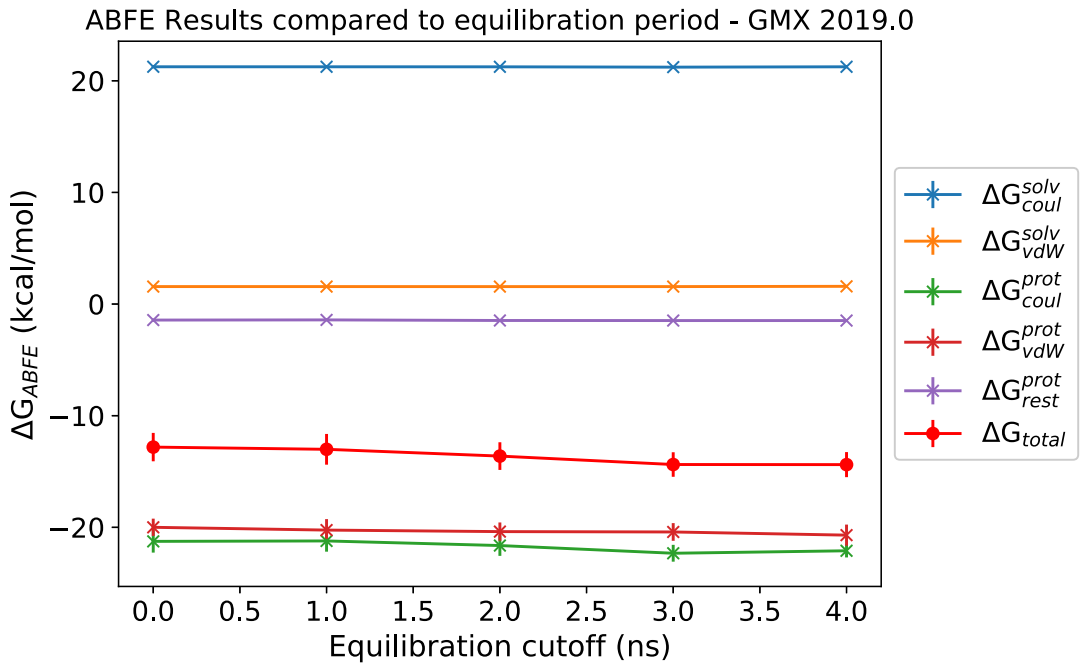


Figure B.1: Energetic breakdown of ABFE calculations of bromosporine binding BRD4(1). ABFE calculations were analysed discarding a an initial equilibration period of 1-4 ns.

conversion to dummy atoms.[176] Although it was not possible to fully investigate this observation for this thesis, a quick summary of the observations is shown below.

The calculated values given with this protocol differ from those shown by Aldeghi *et al.*, who predicted an affinity of -11.3 ± 0.3 kcal/mol.[58] The initial concern was that the magnitude of the value was greater than those predicted, hence the components of the thermodynamic cycle was investigated. Aldeghi *et al.* reported a ligand 'desolvation' energy (ΔG^{solv}) value of 167.545 ± 0.227 kcal/mol and a complex re-coupling value (ΔG^{prot}) of -185.162 ± 0.114 kcal/mol.[58] These differ from the values predicted using the revised method presented, where of the terms for comparison gives $\Delta G^{solv} = 22.81 \pm 0.035$ kcal/mol and $\Delta G^{prot} = -43.49 \pm 1.238$ kcal/mol.

To test if the breakdown of the calculations into individual components (from two to five) was responsible, the ligand energy contribution (ΔG^{solv}) were recalculated in one single set of MD simulations. This new simulation set underwent subsequent mBAR calculations. The values obtained agreed within error, showing

the breakdown of calculations was not responsible. Hence, the difference is likely due to the change from conversion to dummy atoms to complete annihilation in the new procedure. Future work should look at running the GROMACS 2019 protocol with dummy atoms for comparison, as well as running a calculation for taking non-interacting dummy atoms to totally annihilated atoms.

In addition, the error calculated was larger than that reported by Aldeghi *et al.* and stems from the Coulombic set of simulations of the complex, where the errors arising from differences between replicas is large. A rigorous analysis of the cause of the errors is beyond the scope of this thesis, but adopting an increase in the number of replicas could add confidence to the results obtained. The decomposition of the calculations also allows for the Coulombic simulations to be run for longer, and could allow for improved convergence.

B.2 ABFE data breakdown

Table B.1: Breakdown of non-GPU accelerated ABFE calculations performed. All components are shown in units of kcal/mol, with all uncertainties being one standard-deviation.

Ligand	$\Delta G^{solv}_{elec+VdW}$	ΔG^{solv}_{restr}	$\Delta G^{prot}_{elec+VdW+restr}$	$\Delta\Delta G_{corr.}$	ΔG_{Total}
41 (up)	54.91 ± 0.03	7.05	-66.58 ± 0.06	-0.49	
	54.94 ± 0.04	7.05	-67.43 ± 0.07	-0.51	-4.65 ± 0.47
	54.85 ± 0.03	7.05	-67.32 ± 0.06	-0.50	
41(down)	54.91 ± 0.03	7.06	-67.98 ± 0.06	-0.53	
	54.94 ± 0.04	7.06	-66.67 ± 0.05	-0.50	-4.93 ± 0.67
	54.85 ± 0.03	7.06	-67.51 ± 0.07	-0.49	
45(up)	44.68 ± 0.11	6.87	-56.16 ± 0.10	-0.52	
	45.58 ± 0.12	6.87	-56.48 ± 0.07	-0.47	-4.93 ± 1.68
	45.22 ± 0.03	6.87	-59.35 ± 0.06	-0.43	
45(down)	44.68 ± 0.11	6.97	-60.42 ± 0.09	-0.44	
	45.58 ± 0.12	6.97	-60.60 ± 0.11	-0.47	7.90 ± 0.62
	45.22 ± 0.03	6.97	-60.97 ± 0.09	-0.47	

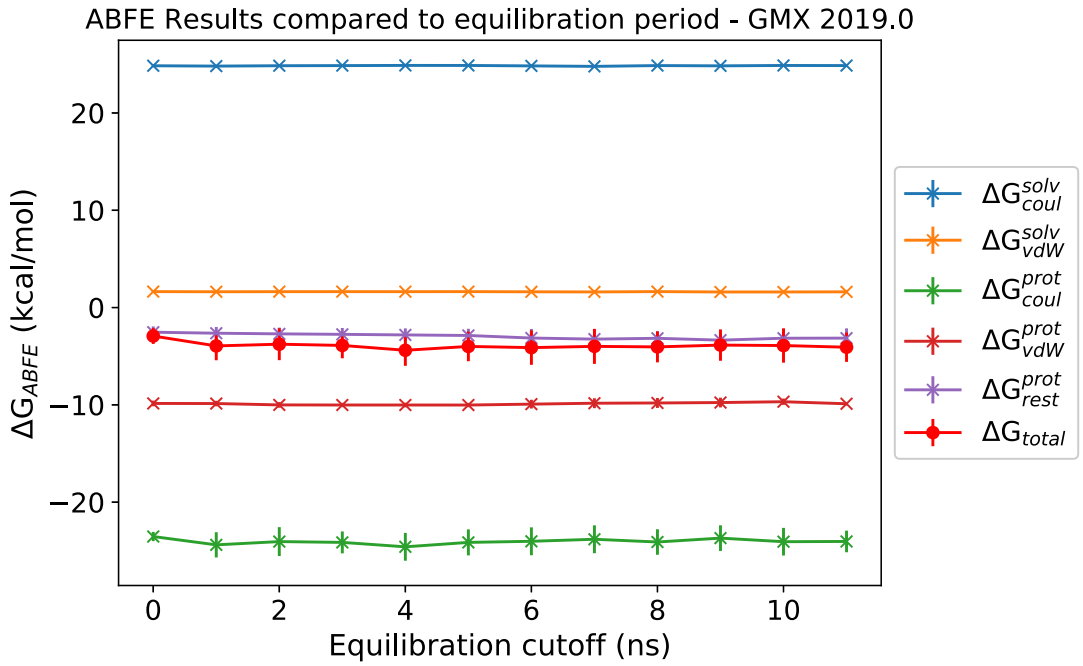


Figure B.2: Energetic breakdown of ABFE calculations of **57** binding TRIM33β.

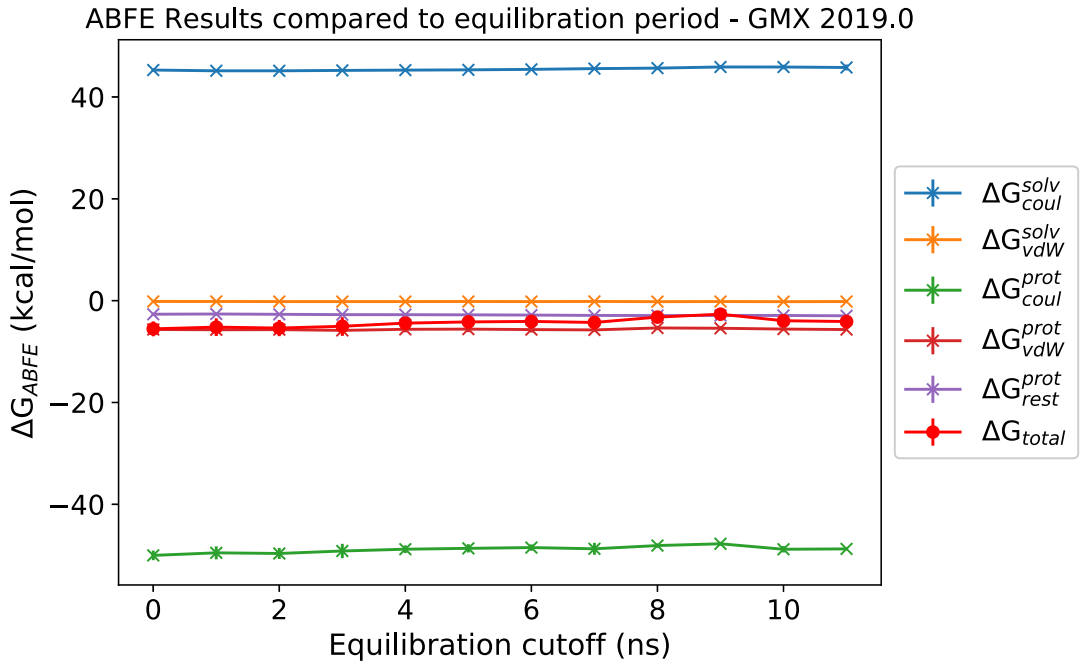


Figure B.3: Energetic breakdown of ABFE calculations of **65** binding TRIM33β.

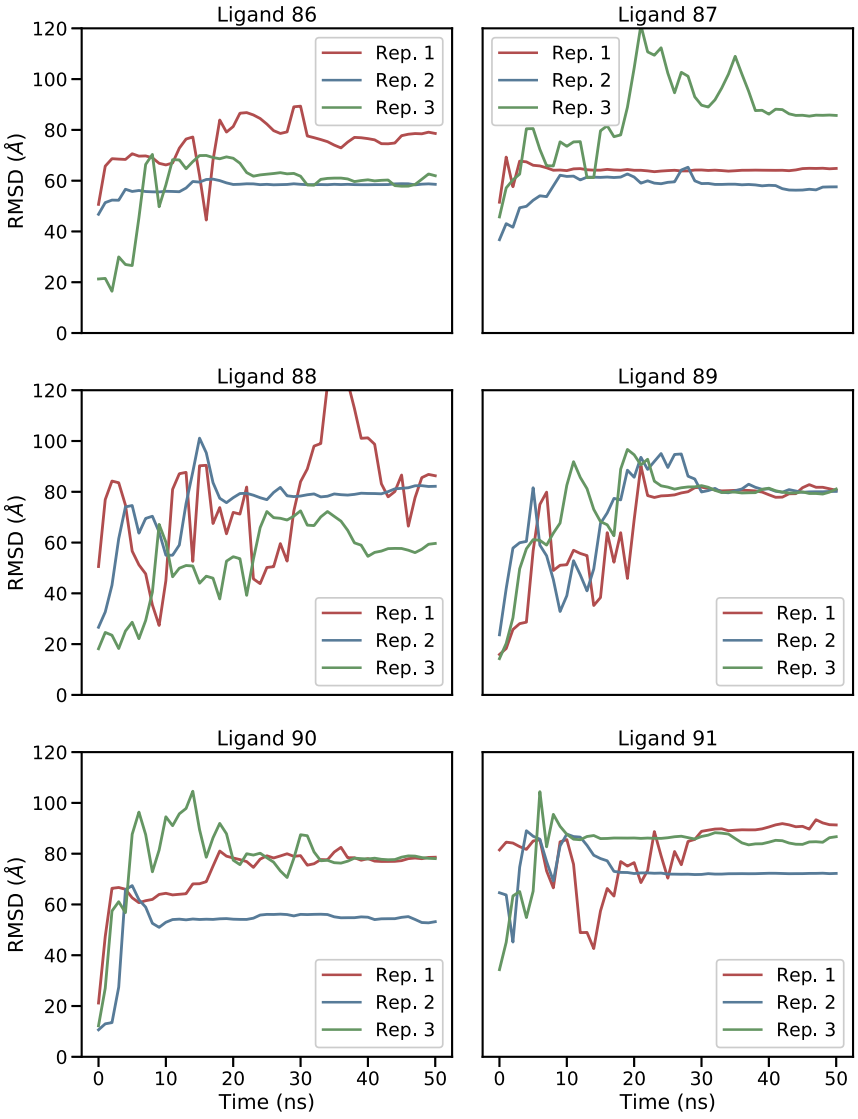


Figure B.4: RMSD plots of ChEMBL analogues of **41**, from triplicate MD simulations.

C

Appendix for Chapter 5

Contents

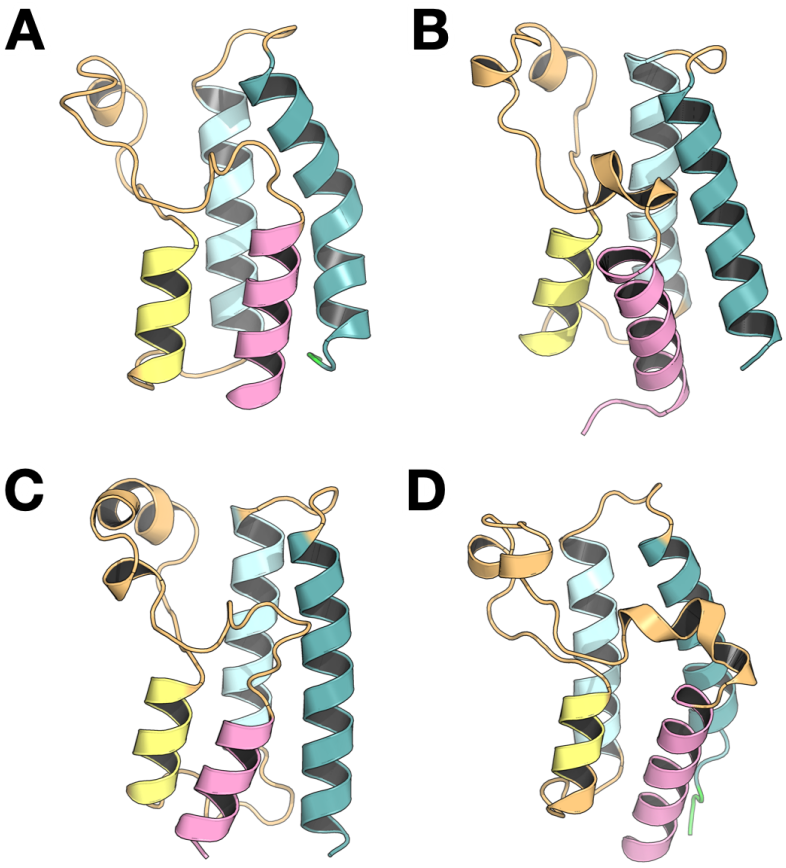


Figure C.1: Homology models of parasite bromdomains: (**A**) *Sm*BRD3(2), (**B**) *Tc*BDF2, (**C**) *Tc*BDF3 and (**D**) *Tc*BDF5(1)

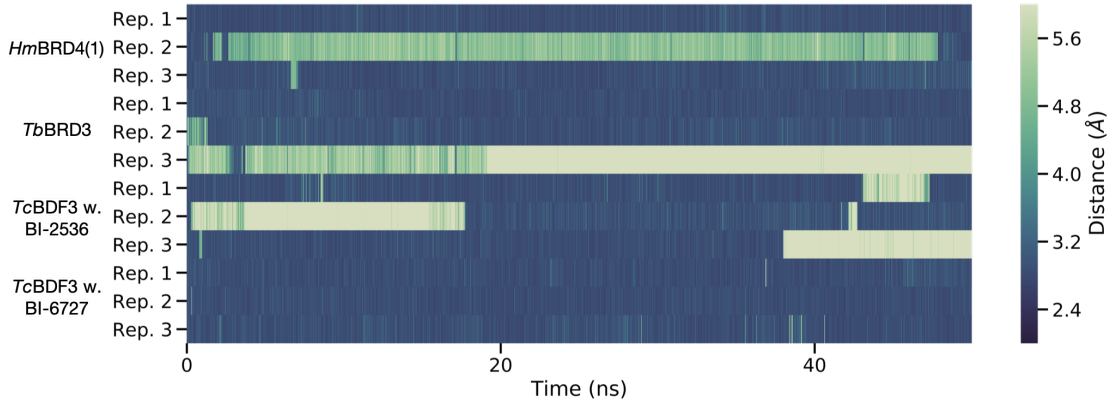


Figure C.2: Distances of the BI-2536/BI-6727 (**20/21**) core carbonyl to the binding site Asn residue.

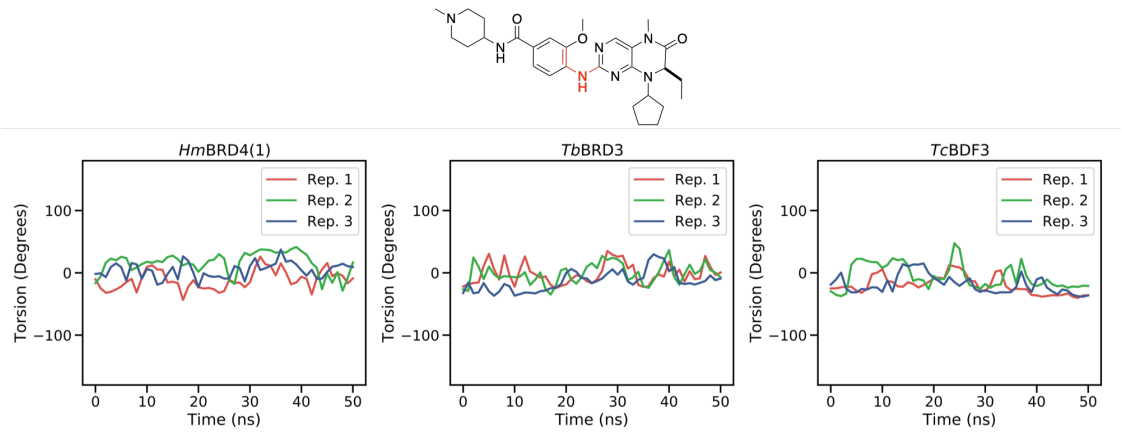


Figure C.3: Anisole torsions of **20** in differing proteins. The torsions, shown in red, have been measured during MD simulations bound to *Hm*BRD4(1), *Tb*BRD3 and *Tc*BDF3.

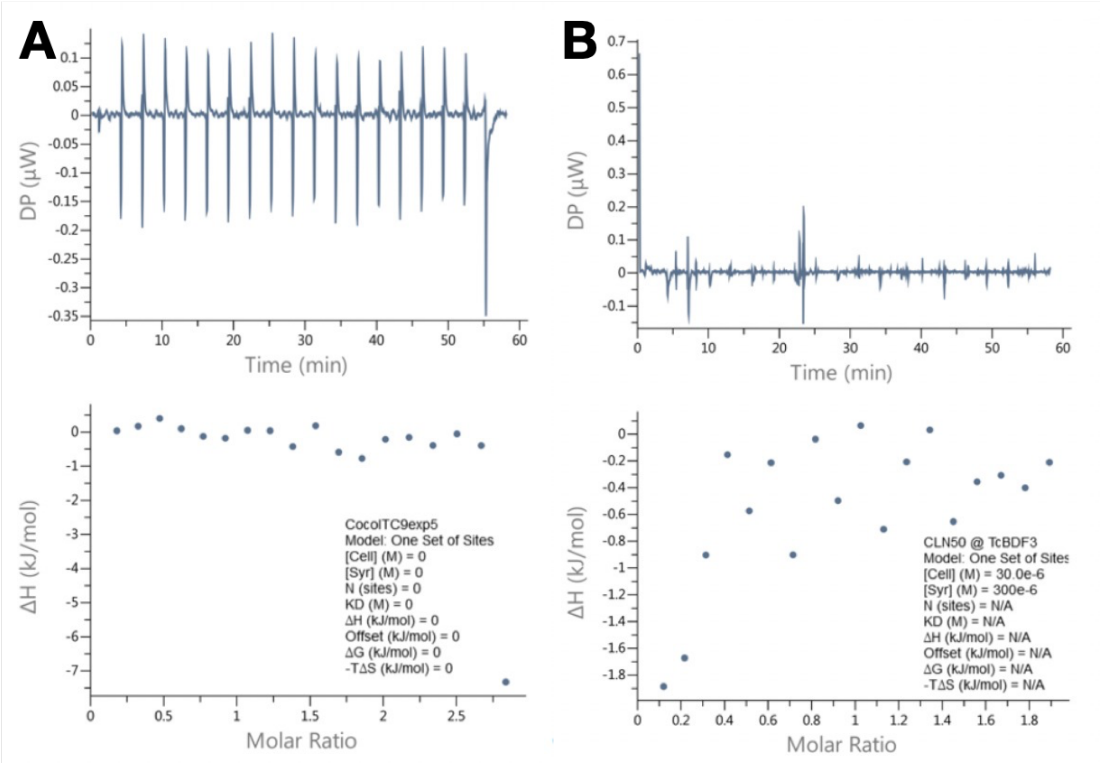


Figure C.4: ITC results of **94** binding *Tc*BDF3. ITC was performed by C. Laurin.

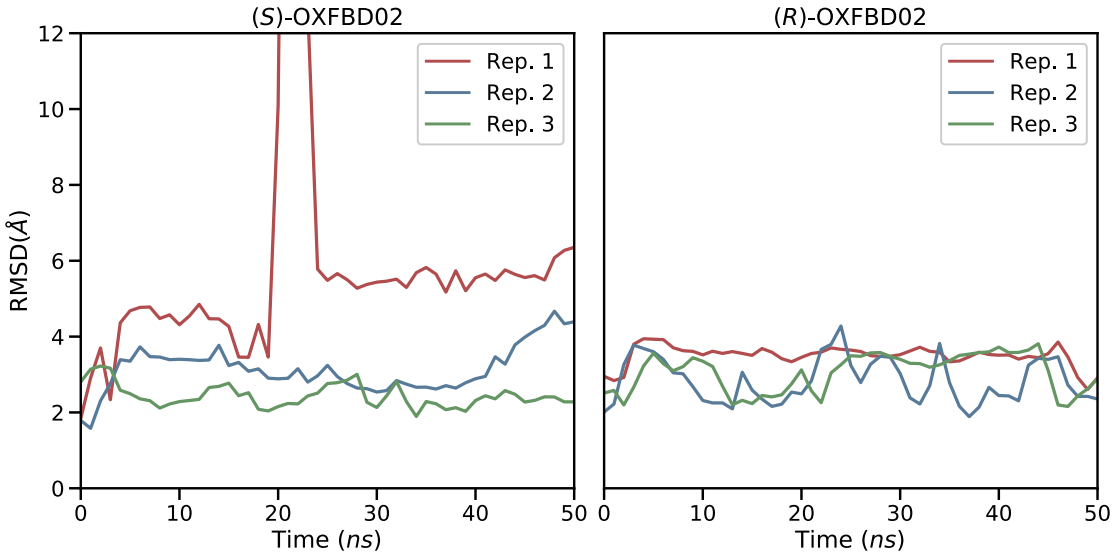


Figure C.5: RMSD of ligand atoms in OXFBD02 (**9**) to *Tc*BDF5(1)

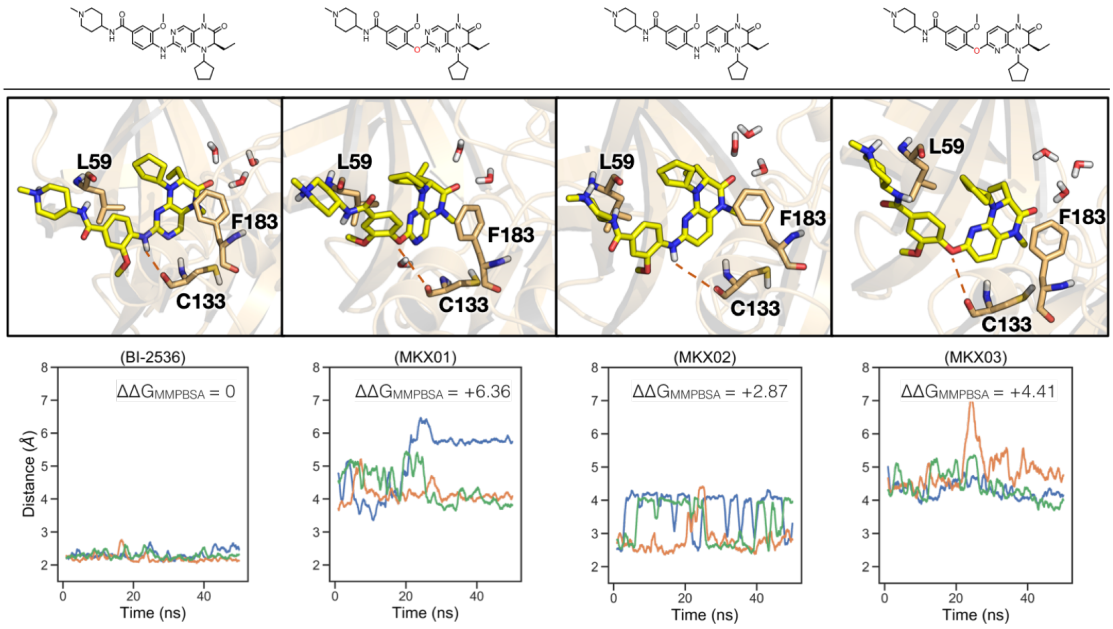


Figure C.6: MD and MM-PBSA experiments performed on PLK1 binding **20** and its analogues. **20** and analogues removing vital hinge binding motifs were subject to MD simulations in triplicate by A. Boczek. The plots show the distance of the ligands from C133, a residue at the base of the PLK1 binding site. MM-PBSA calculations of these simulations showed positive ΔΔG predictions, shown within the respective plots.

C.1 Alignments for homology modelling

*Tc*BDF2

```
TcBDF2        MHHHHHHSSGVDLGTENLYFQSMGKRGREGTLDKPRCLAF    50
TbBDF2:5CZG   ....................GMSKNERDTSFNKNGCLVF    18

TcBDF2        VHQLWDKDKLKMFHHPISAAELPDYHKVINYPVDLSTIRQ    90
TbBDF2:5CZG   VSRLWDLDKLGMFHHPVSAEELPDYHTVIKRPVDLSSIRD    58

TcBDF2        GIESGKYDSDADVQNAVAQMIANALEYNAKGTEWHQQALS   130
TbBDF2:5CZG   GIEKGTYATDVDVQNDVARMITNALEYNAKGSTWYQEAMS    98

TcBDF2        FRNIYLDVARQCGLS.   145
TbBDF2:5CZG   FRKTYLDLARQSGLVV   114
```

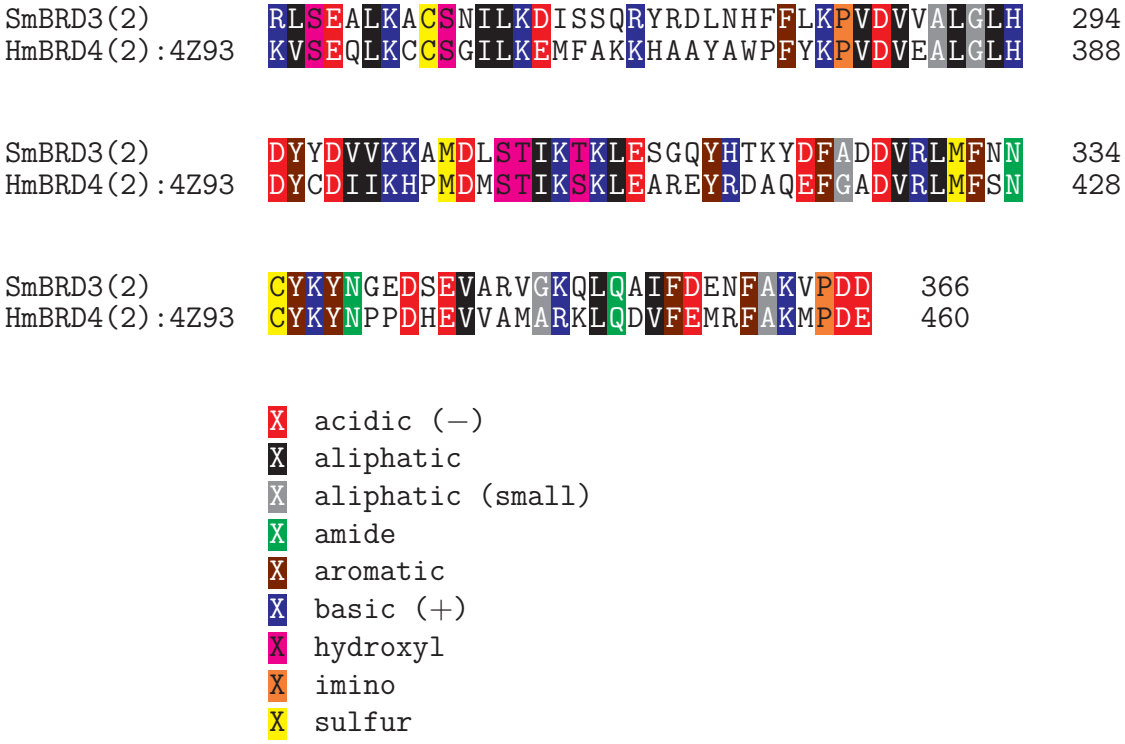
*Tc*BDF3

```
TcBDF3        MHHHHHHSSGVDLGTENLYFQSMGSTGRKAVDEINHWIAY    30
Tb(427):5C8G  ....................................      -1

TcBDF3        IDCALSHPHPLPKGKHVFRSDLSTVPEVRDIYDCLYKLYA    70
Tb(427):5C8G  ......HPLPLPKNKHVFHSDQRLAPEIRDLYDCLYKLYA    33

TcBDF3        EESASASFREPVNALELGVFNYYEVVTEPMSLRTVLDRIA   110
Tb(427):5C8G  EESASEYFREPVDALRVGAWNYYSVITEPMSLRTVLDYIV    73

TcBDF3        EGGHYSQASQVLADVEKIWSNCEKYNGADSALVKEAKKCQ   150
Tb(427):5C8G  QGGRYSHVEQIMNDVELIWKNCERYNGAESHLAADARRCR   113

TcBDF3        GILTRLRERL.........   160
Tb(427):5C8G  AILEKHRERLADEQTAPAAEVD   135
```

*Tc*BDF5(1)

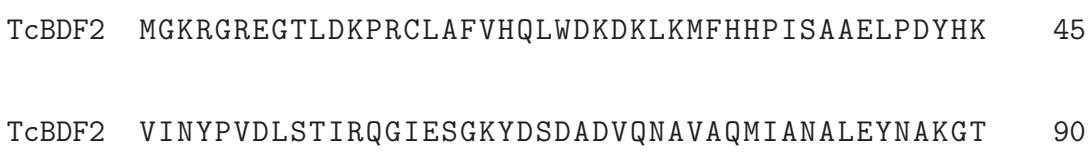
```
TcBDF5(1)     MHHHHHHSSGVDLGTENLYFQSMEERRQKFYPEEELVALV    18
LpBPK1:5TCM   ........................FQGYNEADVAALV    14
```

*Sm*BRD3(2)



C.2 Protein sequences

*Tc*BDF2 - GeneBank ID: XP_803713.1



```
TcBDF2  EWHQQALSFRNIYLDVARQCGLSVDDDAAYIPSAAFKDDESTLRK   135

TcBDF2  AEKKIAENLDVVLKDLEEEKAVPLEELKAKYKRAKKKKGGKNADG   180

TcBDF2  SDSDDSDDSDDSDESEESEESDEIADDSDDDSDDDSDDSDDSDEE   225

TcBDF2  KFYYDDESDDE   236
```

***Tc*BDF3 - GeneBank ID: XP_812334.1**

```
TcBDF3  MGSTGRKAVDEINHWIAYIDCALSHPHPLPKGKHVFRSDLSTVPE    45

TcBDF3  VRDIYDCLYKLYAEESASASFREPVNALELGVFNYYEVVTEPMSL    90

TcBDF3  RTVLDRIAEGGHYSQASQVLADVEKIWSNCEKYNGADSALVKEAK   135

TcBDF3  KCQGILTRLRERLAEEQPAPNAELDKIISAFESADESVLGELEAY   180

TcBDF3  FRREDPSLIISNGDVDLTALRVKHLKAMKAILERAMNGGGGRG   223
```

***Tc*BDF5(1) - GeneBank ID: XP_818752.1**

```
TcBDF5(1/2)  MEERRQKFYPEEELVALVRSLDRPQDEGLFSMDVLVVYPH    40

TcBDF5(1/2)  LEQEYTRVCPKRCDLATAAEKAANEAYSYDVNLTALREDI    80

TcBDF5(1/2)  KLMVNNCYRFNGTKGPLANIAERFEAFAKEQIDAYVTKKA   120

TcBDF5(1/2)  GGRRLSSLRLSAVTATAAETSRGKRQRVESKDTSGVHGSQ   160

TcBDF5(1/2)  KPQPHRGGDKSRSGEMDGLSATSAEMVRLVDSLNRREDQG   200

TcBDF5(1/2)  AFAVDVAEKYPELKKSYEEVCPVKMNLTILRERAASGYYL   240
```

```
TcBDF5(1/2)  GIPQPTTTSKNPEPQTQSKRRQLLASDASPSSVFGSTTAE  280

TcBDF5(1/2)  SLVLLRNDVELMVSNCVRFNSNVQSWVELARSFQSFAHKK  320

TcBDF5(1/2)  IDDFVIRHVPALKGTKSGVDAYVTGEKQQSQQQLLSQASK  360

TcBDF5(1/2)  SQQLHFTSTQDDEQMVTSTHSFASPSFSLLSTATATTYAM  400

TcBDF5(1/2)  KGRAAEAASPPLRAPAVTVANYITPTIVPTALQPVFATPD  440

TcBDF5(1/2)  ALRRRLVSDHLHRETLRARLIRSSTRSHNDKGDNKNTINN  480

TcBDF5(1/2)  GTNNNKEEEEGKEEEEEEDGADKKEEKNSVQHTYNSAFSC  520

TcBDF5(1/2)  RAVLCAFAASVREFYRLKRECQNFVNPFEYPRQEEQLYMD  560

TcBDF5(1/2)  CIVLIEQQFERLFLHTLLYDKEKADFYDWNAEKAVRKIVG  600

TcBDF5(1/2)  SSEVTSLKEKDHEGFLCGFGSSWLDDAHLCYLVRFLQHLP  640

TcBDF5(1/2)  QLMGLACATTVHLTGTELKDARPALQITEVAKGVIEKLAR  680

TcBDF5(1/2)  ITEELLTFIAQYEQEAMNMESHAEIGKEVS   710
```

_Sm_BRD3(2) - GeneBank ID: CCD76183.1

```
SmBRD3(1/2)  MILANAMSSEANLDSSNSIKSEPHAAKSHSSKITTNQLEY   40

SmBRD3(1/2)  IKKEVVGRLFKEKIVWPFTKPVDHQRLNLPDYPKIIKHPM   80

SmBRD3(1/2)  DLGTIKQRLNLKFYHSSSECLDDLFTMFRNCYIFNKPGDD  120

SmBRD3(1/2)  VVAMAMKLEQIARERLKFMPTPETEICPQKTPKSIRPIGA  160
```

```
SmBRD3(1/2)  PLQVHPPIEPIHTAASTNHTEGLNGSAVSVDQTTLPFRPS   200

SmBRD3(1/2)  VTSTSTKKASKKKSDSTIDELPSTPQSYDDLSRDRRQIKK   240

SmBRD3(1/2)  PKREYEERNVGKRLRLSEALKACSNILKDISSQRYRDLNH   280

SmBRD3(1/2)  FFLKPVDVVALGLHDYYDVVKKAMDLSTIKTKLESGQYHT   320

SmBRD3(1/2)  KYDFADDVRLMFNNCYKYNGEDSEVARVGKQLQAIFDENF   360

SmBRD3(1/2)  AKVPDDESDPAASPDGRPVDQNMYQLIQNAIKEHQKLTNQ   400

SmBRD3(1/2)  FQRFSEDLQKSTANLNSILSSLSMAVRKAPIGHNTPHINS   440

SmBRD3(1/2)  LPPTQTGLPTVPRPTMNDIEDVNITKRGRQSQSKTKYRQS   480

SmBRD3(1/2)  GLSAAAPVLNAPCAPVSSSTVNMSSTHSQPIPVPGYATDE   520

SmBRD3(1/2)  EMSENNVRPMTYDEKRQLSLDINKLPGEKLGRVVQIIQQR   560

SmBRD3(1/2)  EPSHRDCNPDEIEIDFETLQHTTLRELEKYVKSVLQKTKS   600

SmBRD3(1/2)  GSRKYVKKGLSGVPPGKTREECMKEKTEELENRLREIGGL   640

SmBRD3(1/2)  PQGAPGYHNAHNPKKSHGTNRLSASSSSSSDSESTSDSSE   680

SmBRD3(1/2)  SDSSDSKSYSES   692
```

D

Appendix for Chapter 6

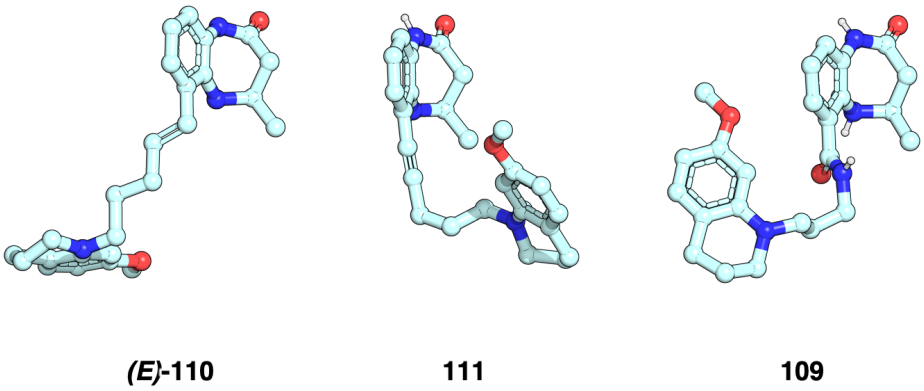


Figure D.1: Small molecule crystal structures of *(E)*-**110**, **111** & **109**. Determined by Dr M. Brand and M. Morologu.

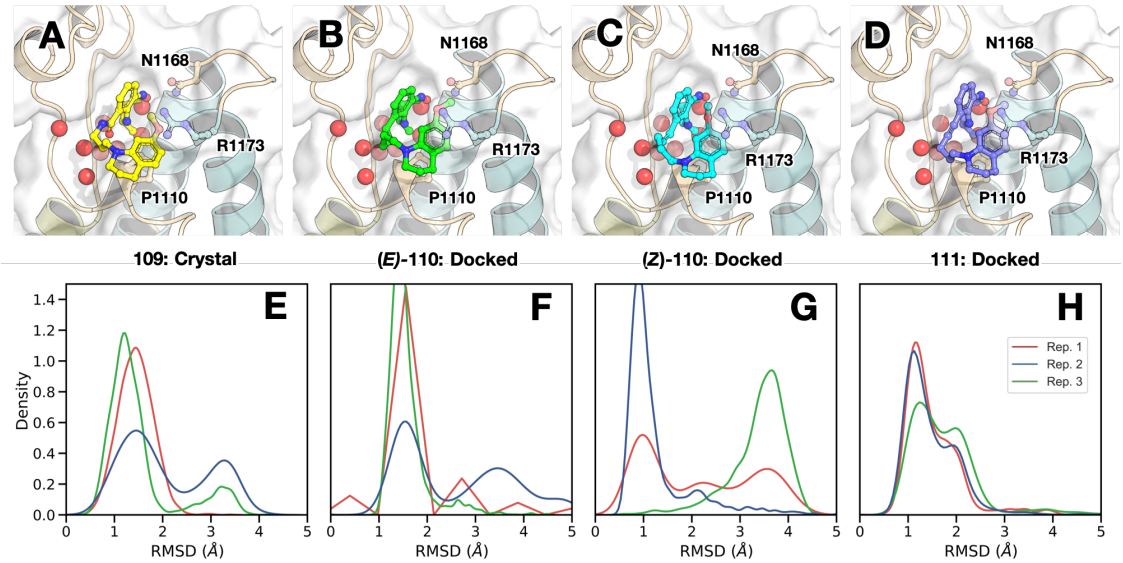


Figure D.2: RMSD of CREBBP ligands. (**A**) The crystal structure of **109** bound to the CREBBP BRD. (**B-D**) Docked structures of *(E)*-**110**, *(Z)*-**110** and **111** bound to the CREBBP BRD. Structures were subject to 50 ns of MD simulation, in triplicate. (**E-H**) RMSD values of the ligands are shown as KDE plots.

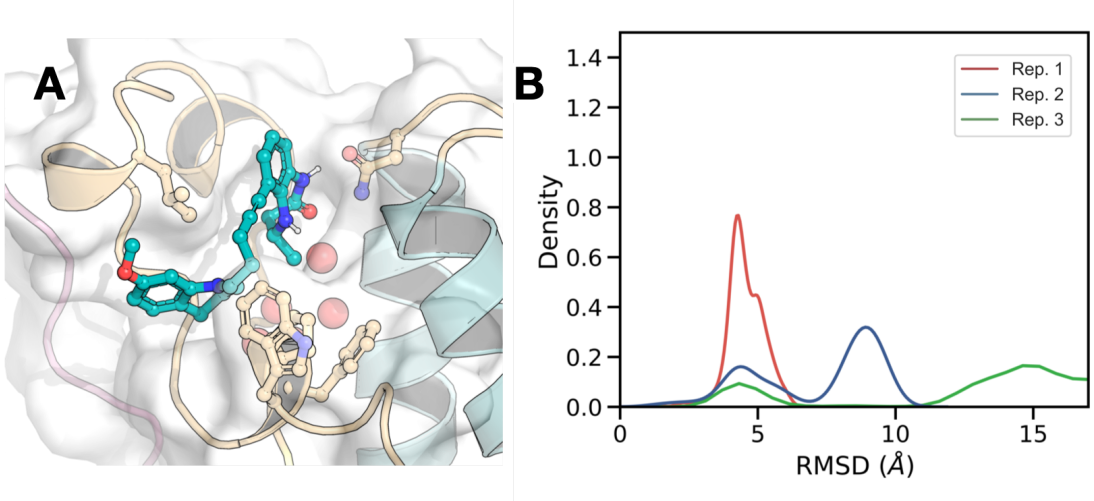


Figure D.3: RMSD of **111** docked in BRD4(1). (**A**) The docked pose of **111** in BRD4(1). (**B**) The RMSD values of the ligand throughout the simulations are shown as a KDE plot.

The first kind of intellectual and artistic personality belongs to the hedgehogs, the second to the foxes . . .

— Sir Isaiah Berlin

References

1. Musselman, C. A., Lalonde, M.-E., Côté, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nature Structural & Molecular Biology* **19,** 1218–1227 (2012).
2. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389,** 251–260 (1997).
3. Fujisawa, T. & Filippakopoulos, P. Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nature Reviews Molecular Cell Biology* **18,** 246–262 (2017).
4. Xiao Pfister, S. & Ashworth, A. Marked for death: targeting epigenetic changes in cancer. *Nature Reviews, Drug Discovery* **16,** 241–263 (2017).
5. Brand, M. *et al.* Small molecule inhibitors of bromodomain - Acetyl-lysine interactions. *ACS Chemical Biology* **10,** 22–39 (2015).
6. Jennings, L. E., Measures, A. R., Wilson, B. G. & Conway, S. J. Phenotypic screening and fragment-based approaches to the discovery of small-molecule bromodomain ligands. *Future medicinal chemistry* **6,** 179–204 (2014).
7. Smith, S. G. & Zhou, M. M. The bromodomain: A new target in emerging epigenetic medicine. *ACS Chemical Biology* **11,** 598–608 (2016).
8. Mohammad, H. P., Barbash, O. & Creasy, C. L. Targeting epigenetic modifications in cancer therapy: erasing the roadmap to cancer. *Nature Medicine* **25** (2019).
9. Pack, L. R., Yamamoto, K. R. & Fujimori, D. G. Opposing chromatin signals direct and regulate the activity of lysine demethylase 4C (KDM4C). *Journal of Biological Chemistry* **291,** 6060–6070 (2016).
10. Mellor, J. It takes a PHD to read the histone code. *Cell* **126,** 22–24 (2006).
11. Aasland, R., Gibson, T. J. & Stewart, F. The PHD Finger: implications for chromatin mediated transcriptional control. *Trends in Biochemical Sciences* **20,** 56–59 (1995).
12. Krishna, S. S., Majumdar, I. & Grishin, N. V. Structural classification of zinc fingers. *Nucleic Acids Research* **31,** 532–550 (2003).
13. Slama, P. & Geman, D. Identification of family-determining residues in PHD fingers. *Nucleic Acids Research* **39,** 1666–1679 (2011).
14. Sanchez, R. & Zhou, M. M. The PHD finger: A versatile epigenome reader. *Trends in Biochemical Sciences* **36,** 364–372 (2011).
15. Spiliotopoulos, D., Spitaleri, A. & Musco, G. Exploring PHD fingers and H3K4Me0 interactions with molecular dynamics simulations and binding free energy calculations: AIRE-PHD1, a comparative study. *PLoS ONE* **7,** 1–13 (2012).

16. Laity, J. H., Lee, B. M. & Wright, P. E. Zinc finger proteins: New insights into structural and functional diversity. *Current Opinion in Structural Biology* **11,** 39–46 (2001).

17. Gibbons, R. J. *et al.* Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain. *Nature Genetics* **17,** 146–148 (1997).

18. Pascual, J., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *Journal of Molecular Biology* **304,** 723–729 (2000).

19. Tsai, W. W. *et al.* TRIM24 links a non-canonical histone signature to breast cancer. *Nature* **468,** 927–932 (2010).

20. Wagner, E. K., Nath, N., Flemming, R., Feltenberger, J. B. & Denu, J. M. Identification and characterization of small molecule inhibitors of a plant homeodomain finger. *Biochemistry* **51,** 8293–8306 (2012).

21. Amato, A., Lucas, X., Bortoluzzi, A., Wright, D. & Ciulli, A. Targeting ligandable pockets on PHD zinc finger domains by a fragment-based approach. *ACS Chemical Biology* **13,** 915–921 (2018).

22. Strohner, R. *et al.* NoRC - A novel member of mammalian ISWI-containing chromatin remodeling machines. *EMBO Journal* **20,** 4892–4900 (2001).

23. Gu, L. *et al.* BAZ2A (TIP5) is involved in epigenetic alterations in Prostate cancer and its overexpression predicts disease recurrence. *Nature Genetics* **47,** 22–33 (2015).

24. Arking, D. E. *et al.* Identification of a sudden cardiac death susceptibility locus at 2q24.2 through genome-wide association in european ancestry individuals. *PLoS Genetics* **7,** 1–9 (2011).

25. Ferri, E., Petosa, C. & McKenna, C. E. Bromodomains: Structure, function and pharmacology of inhibition. *Biochemical Pharmacology* **106,** 1–18 (2016).

26. Haynes, S. R. *et al.* The bromodomain: A conserved sequence found in human, Drosophila and yeast proteins. *Nucleic Acids Research* **20,** 2603 (1992).

27. Hewings, D. S. *et al.* Progress in the development and application of small molecule inhibitors of bromodomain-acetyl-lysine interactions. *Journal of Medicinal Chemistry* **55,** 9393–9413 (2012).

28. Aldeghi, M. *et al.* Large-scale analysis of water stability in bromodomain binding pockets with grand canonical Monte Carlo. *Communications Chemistry* **1,** 1–12 (2018).

29. Liu, J. *et al.* The polar warhead of a TRIM24 bromodomain inhibitor rearranges a water-mediated interaction network. *The FEBS Journal* **284,** 1082–1095 (2017).

30. Vidler, L. R., Brown, N., Knapp, S. & Hoelder, S. Druggability analysis and structural classification of bromodomain acetyl-lysine binding sites. *Journal of Medicinal Chemistry* **55,** 7346–7359 (2012).

31. Theodoulou, N. H., Tomkinson, N. C., Prinjha, R. K. & Humphreys, P. G. Clinical progress and pharmacology of small molecule bromodomain inhibitors. *Current Opinion in Chemical Biology* **33,** 58–66 (2016).

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

33. Nicodeme, E. *et al.* Suppression of inflammation by a synthetic histone mimic. *Nature* **468,** 1119–1123 (2010).

34. Baud, M. G. J. *et al.* A bump-and-hole approach to engineer controlled selectivity of BET bromodomain chemical probes. *Science* **346,** 638–641 (2014).

35. Law, R. P. *et al.* Discovery of tetrahydroquinoxalines as bromodomain and extra-terminal domain (BET) inhibitors with selectivity for the second bromodomain. *Journal of Medicinal Chemistry* **61,** 4317–4334 (2018).

36. Picaud, S. *et al.* RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 19754–19759 (2013).

37. McLure, K. G. *et al.* RVX-208, an inducer of ApoA-I in humans, is a BET bromodomain antagonist. *PLoS ONE* **8,** 1–12 (2013).

38. Nicholls, S. J. *et al.* Efficacy and safety of a novel oral inducer of apolipoprotein A-I synthesis in statin-treated patients with stable coronary artery disease: A randomized controlled trial. *Journal of the American College of Cardiology* **57,** 1111–1119 (2011).

39. Tsujikawa, L. M. *et al.* Apabetalone (RVX-208) reduces vascular inflammation in vitro and in CVD patients by a BET-dependent epigenetic mechanism. *Clinical Epigenetics* **11,** 1–21 (2019).

40. Hewings, D. S. *et al.* Optimization of 3,5-dimethylisoxazole derivatives as potent bromodomain ligands. *Journal of Medicinal Chemistry* **56,** 3217–3227 (2013).

41. Dawson, M. A. *et al.* Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* **478,** 529–533 (2011).

42. Kerscher, B. *et al.* BET bromodomain inhibitor IBET151 impedes human ILC2 activation and prevents experimental allergic lung inflammation. *Frontiers in Immunology* **10,** 1–15 (2019).

43. Demont, H. E. & Gosmini, R. L. M. *Thetrahydroquinolines derivatives as bromodomain inhibitors (WO2011054848)* 2011.

44. Gosmini, R. *et al.* The discovery of I-BET726 (GSK1324726A), a potent tetrahydroquinoline ApoA1 up-regulator and selective BET bromodomain inhibitor. *Journal of Medicinal Chemistry* **57,** 8111–8131 (2014).

45. Wyce, A. *et al.* BET inhibition silences expression of MYCN and BCL2 and induces cytotoxicity in neuroblastoma tumor models. *PLoS ONE* **8,** 1–16 (2013).

46. Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. & Nakatani, Y. The Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases Vasily. *Cell* **87,** 953–959 (1996).

47. Bannister, A. J. & Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature* **384,** 641–643 (1996).

48. Bedford, D. C., Kasper, L. H., Fukuyama, T. & Brindle, P. K. Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases. *Epigenetics* **5,** 9–15 (2010).

49. Sachchidanand *et al.* Target structure-based discovery of small molecules that block human p53 and CREB binding protein association. *Chemistry and Biology* **13,** 81–90 (2006).

50. Rooney, T. P. C. *et al.* A series of potent crebbp bromodomain ligands reveals an induced-fit pocket stabilized by a cation-π interaction. *Angewandte Chemie - International Edition* **53,** 6126–6130 (2014).

51. Albrecht, K. B. *et al. Use of CBP/EP300 bromodomain inhibitors for cancer immunotherapy (WP2015054642A2)* 2015.

52. Taylor, A. M. *et al.* Fragment-based discovery of a delective and cell-active benzodiazepinone CBP/EP300 bromodomain inhibitor (CPI-637). *ACS Medicinal Chemistry Letters* **7,** 531–536 (2016).

53. Palmer, W. S. *et al.* Structure-Guided Design of IACS-9571, a Selective High-Affinity Dual TRIM24-BRPF1 Bromodomain Inhibitor. *Journal of Medicinal Chemistry* **59,** 1440–1454 (2016).

54. Gechijian, L. N. *et al.* Functional TRIM24 degrader *via* conjugation of ineffectual bromodomain and VHL ligands. *Nature Chemical Biology* **14,** 1–8 (2018).

55. Chen, P. *et al.* Discovery and Characterization of GSK2801, a Selective Chemical Probe for the Bromodomains BAZ2A and BAZ2B. *Journal of Medicinal Chemistry* **59,** 1410–1424 (2016).

56. Drouin, L. *et al.* Structure enabled design of BaZ2-ICR, a chemical probe targeting the bromodomains of BaZ2a and BaZ2B. *Journal of Medicinal Chemistry* **58,** 2553–2559 (2015).

57. Picaud, S. *et al.* Promiscuous targeting of bromodomains by bromosporine identifies BET proteins as master regulators of primary transcription response in leukemia. *Science Advances* **2,** 15–20 (2016).

58. Aldeghi, M., Heifetz, A., Bodkin, M. J., Knapp, S. & Biggin, P. C. Predictions of ligand selectivity from absolute binding free energy calculations. *Journal of the American Chemical Society* **139,** 946–957 (2016).

59. Steegmaier, M. *et al.* BI 2536, a potent and delective inhibitor of polo-like kinase 1, inhibits tumor growth *in vivo*. *Current Biology* **17,** 316–322 (2007).

60. Watts, E. *et al.* Designing dual inhibitors of Anaplastic Lymphoma Kinase (ALK) and Bromodomain-4 (BRD4) by tuning kinase selectivity. *Journal of Medicinal Chemistry* **62,** 2618–2637 (2019).

61. Kothe, M. *et al.* Selectivity-determining residues in Plk1. *Chemical Biology and Drug Design* **70,** 540–546 (2007).

62. Bradner, J. E., Gray, N., Qi, J., McKeown, M. R. & Buckley, D. *Dihydropteridinone Inhibitors of BRD4 (WO 2015/117055 A1)* 2015.

63. Chen, L. *et al.* BRD4 structure-activity relationships of fual PLK1 kinase/BRD4 bromodomain inhibitor BI-2536. *ACS Medicinal Chemistry Letters* **6,** 764–769 (2015).

64. Rudolph, D. *et al.* BI 6727, a polo-like kinase inhibitor with improved pharmacokinetic profile and broad antitumor activity. *Clinical Cancer Research* **15,** 3094–3102 (2009).

65. Reymond, A. *et al.* The tripartite motif family identifies cell compartments. *EMBO* **20,** 2140–2151 (2001).

66. Hatakeyama, S. TRIM family proteins: Roles in autophagy, immunity, and carcinogenesis. *Trends in Biochemical Sciences* **42,** 297–311 (2017).

67. Napolitano, L. M. & Meroni, G. TRIM family: Pleiotropy and diversification through homomultimer and heteromultimer formation. *International Union of Biochemistry and Molecular Biology: Life* **64,** 64–71 (2012).

68. Short, K. M. & Cox, T. C. Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding. *Journal of Biological Chemistry* **281,** 8970–8980 (2006).

69. Ozato, K., Shin, D. M., Chang, T. H. & Morse, H. C. TRIM family proteins and their emerging roles in innate immunity. *Nature Reviews Immunology* **8,** 849–860 (2008).

70. Allton, K. *et al.* Trim24 targets endogenous p53 for degradation. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 11612–11616 (2009).

71. Groner, A. C. *et al.* TRIM24 Is an oncogenic transcriptional activator in prostate cancer. *Cancer Cell* **29,** 846–858 (2016).

72. Bai, X. *et al.* TIF1γ controls erythroid cell fate by regulating transcription elongation. *Cell* **142,** 133–143 (2010).

73. Kulkarni, A. *et al.* Tripartite motif-containing 33 (TRIM33) protein functions in the poly(ADP-ribose) polymerase (PARP)-dependent DNA damage response through interaction with amplified in liver cancer 1 (ALC1) protein. *Journal of Biological Chemistry* **288,** 32357–32369 (2013).

74. Yu, C., Ding, Z., Liang, H., Zhang, B. & Chen, X. The roles of TIF1γ in cancer. *Frontiers in Oncology* **9,** 1–12 (2019).

75. Gallenkamp, D., Gelato, K. A., Haendler, B. & Weinmann, H. Bromodomains and their pharmacological inhibitors. *ChemMedChem* **9,** 438–464 (2014).

76. Xi, Q. *et al.* A poised chromatin platform for TGF-beta access to master regulators. *Cell* **147,** 1511–1524 (2011).

77. Kennedy, P. G. Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *The Lancet Neurology* **12,** 186–194 (2013).

78. World Health Organisation. *Trypanosomiasis - Human African sleeping sickness*

79. Organisation, W. H. *Chagas disease (American trypanosomiasis)*

80. Tzizik, D. M. & Borchardt, R. A. Chagas disease. *Journal of the American Academy of Physician Assistants* **31,** 30–33 (2018).

81. Atwood, J. A. *et al.* The *Trypanosoma cruzi* proteome. *Science* **309,** 473–476 (2005).

82. Ribeiro, A. L., Nunes, M. P., Teixeira, M. M. & Rocha, M. O. Diagnosis and management of Chagas disease and cardiomyopathy. *Nature Reviews Cardiology* **9,** 576–589 (2012).

83. Protasio, A. V. *et al.* A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases* **6,** 1–13 (2012).

84. McManus, D. P., Dunne, D. W., Sacko, M., Zhou, N. & Vennervald, B. J. Schistosomiasis. *Nature Reviews Disease Primers* **4,** 1–19 (2018).

85. De Rycker, M., Baragaña, B., Duce, S. L. & Gilbert, I. H. Challenges and recent progress in drug discovery for tropical diseases. *Nature* **559,** 498–506 (2018).

86. Figueiredo, L. M., Cross, G. A. M. & Janzen, C. J. Epigenetic regulation in African trypanosomes: A new kid on the block. *Nature Reviews Microbiology* **7,** 504–513 (2009).

87. Schlimme, W., Burri, N., Bender, K., Hart, B. & Hecker, H. *Trypanosoma brucei brucei*: differences in the nuclear chromatin of bloodstream forms and procyclic culture forms. *Parasitology* **107,** 237–247 (1993).

88. Burri, M., Schlimme, W., Betschart, B. & Hecker, H. Characterization of the histones of *Trypanosoma brucei brucei* bloodstream forms. *Acta Tropica* **58,** 291–305 (1994).

89. Schulz, D. *et al.* Bromodomain proteins contribute to maintenance of bloodstream form stage identity in the African trypanosome. *PLoS Biology* **13,** 1–38 (2015).

90. Villanova, G. V. *et al. Trypanosoma cruzi* bromodomain factor 2 (BDF2) binds to acetylated histones and is accumulated after UV irradiation. *International Journal for Parasitology* **39,** 665–673 (2009).

91. Bertin, B. *et al. Schistosoma mansoni* CBP/p300 has a conserved domain structure and interacts functionally with the nuclear receptor SmFtz-F1. *Molecular and Biochemical Parasitology* **146,** 180–191 (2006).

92. Mullard, A. The Drug-Maker's Guide To the Galaxy. *Nature* **549,** 445–447 (2017).

93. Priest, B. t. & Erdemli, G. Phenotypic screening in the 21st century. *Clinical Pharmacology and Therapeutics* **5,** 1–2 (2014).

94. Woltosz, W. S. If we designed airplanes like we design drugs. *Journal of Computer-Aided Molecular Design* **26,** 159–163 (2012).

95. Ebejer, J. P., Finn, P. W., Wong, W. K., Deane, C. M. & Morris, G. M. Ligity: A non-superpositional, knowledge-based approach to virtual screening. *Journal of Chemical Information and Modeling* **59,** 2600–2616 (2019).

96. Chevillard, F. & Kolb, P. SCUBIDOO: A large yet screenable and rasily searchable database of computationally created chemical compounds optimized toward high likelihood of synthetic tractability. *Journal of Chemical Information and Modeling* **55,** 1824–1835 (2015).

97. Batiste, L. *et al.* Chemical space expansion of bromodomain ligands guided by in silico virtual couplings (AutoCouple). *ACS Central Science* **4,** 180–188 (2018).

98. Anderson, A. C. The process of structure-based drug design. *Chemistry & Biology* **10,** 787–797 (2003).

99. Cappel, D. *et al.* Relative binding free rnergy calculations applied to protein homology models. *Journal of Chemical Information and Modeling* **56,** 2388–2400 (2016).

100. Ban, F. *et al.* Best practices of computer-aided drug discovery: Lessons learned from the development of a preclinical candidate for prostate cancer with a new mechanism of action. *Journal of Chemical Information and Modeling* **57,** 1018–1028 (2017).

101. Mandal, S., Moudgil, M. & Mandal, S. K. Rational drug design. *European Journal of Pharmacology* **625,** 90–100 (2009).

102. Lo, Y. C., Rensi, S. E., Torng, W. & Altman, R. B. Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today* **23,** 1538–1546 (2018).

103. Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC biology* **9,** 71 (2011).

104. Amaro, R. E. *et al.* Discovery of drug-like inhibitors of an essential RNA-editing ligase in *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 17278–17283 (2008).

105. Oleinikovas, V., Saladino, G., Cossins, B. P. & Gervasio, F. L. Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *Journal of the American Chemical Society* **138,** 14257–14263 (2016).

106. Morris, G. M. *et al.* Automated docking using a Lamarckian genetic algorithm and an rmpirical binding free rnergy function. *Journal of Computational Chemistry* **19,** 1639–1662 (1998).

107. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **31,** 2967–2970 (2009).

108. Rastelli, G. *et al.* Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *Journal of Computational Chemistry* **31,** 797–810 (2009).

109. Aldeghi, M., Bodkin, M. J., Knapp, S. & Biggin, P. C. A statistical analysis on the performance of MMPBSA versus absolute binding free energy calculations: bromodomains as a case study. *Journal of Chemical Information and Modeling* **57,** 2203–21 (2017).

110. Cournia, Z., Allen, B. & Sherman, W. Relative binding free energy valculations in drug discovery: Recent advances and practical considerations. *Journal of Chemical Information and Modeling* **57,** 2911–2937 (2017).

111. Abel, R., Wang, L., Harder, E. D., Berne, B. J. & Friesner, R. A. Advancing drug discovery through enhanced free energy calculations. *Accounts of Chemical Research* **50,** 1625–1632 (2017).

112. Williams-Noonan, B. J., Yuriev, E. & Chalmers, D. K. Free energy methods in drug design: Prospects of "alchemical perturbation" in medicinal chemistry. *Journal of Medicinal Chemistry* **61,** 638–649 (2018).

113. Evoli, S., Mobley, D. L., Guzzi, R. & Rizzuti, B. Multiple binding modes of ibuprofen in human serum albumin identified by absolute binding free energy calculations. *Physical Chemistry Chemical Physics* **18,** 32358–32368 (2016).

114. Ross, G. A., Bodnarchuk, M. S. & Essex, J. W. Water sites, networks, and free energies with grand canonical Monte Carlo. *Journal of the American Chemical Society* **137,** 14930–14943 (2015).

115. Sridhar, A., Ross, G. A. & Biggin, P. C. Waterdock 2.0: Water placement prediction for Holo-structures with a pymol plugin. *PLoS ONE* **12,** 1–17 (2017).

116. Schiedel, M. *et al.* Chemical epigenetics: the impact of chemical- and chemical biology techniques on bromodomain target validation. *Angewandte Chemie International Edition* (2019).

117. Yasgar, A., Jadhav, A., Simeonov, A. & Coussens, N. P. *AlphaScreen-Based Assays: Ultra-High-Throughput Screening for Small-Molecule Inhibitors of Challenging Enzymes and Protein-Protein Interactions. In: Janzen W. (eds) High Throughput Screening.* 77–98 (Humana Press, New York, NY, 2016).

118. Cer, R. Z., Mudunuri, U., Stephens, R. & Lebeda, F. J. IC50-to-Ki: a web-based tool for converting IC50 to Ki values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Research* **37,** 441–445 (2009).

119. Freyer, M. W. & Lewis, E. A. Isothermal titration calorimetry: Experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods in Cell Biology* **84,** 79–113 (2008).

120. Chodera, J. D. & Mobley, D. L. Entropy-Enthalpy compensation: Role and ramifications in biomolecular ligand recognition and design. *Annual Review of Biophysics* **42,** 121–142 (2013).

121. Ladbury, J. E., Klebe, G. & Freire, E. Adding calorimetric data to decision making in lead discovery: A hot tip. *Nature Reviews Drug Discovery* **9,** 23–27 (2010).

122. Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M. & Stockman, B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *Journal of Biomolecular NMR* **21,** 349–359 (2001).

123. del Carmen Fernandez-Alonso, M. *et al.* Protein-carbohydrate interactions studied by NMR: From molecular recognition to drug design. *Current Protein and Peptide Science* **13,** 816–830 (2013).

124. Measures, A. *Probing tandem epigenetic reader domains* PhD thesis (University of Oxford, 2016).

125. Leach, A. R. *Molecular modelling: Principles and applications* 2nd (Harlow: Prentice Hall, 2001).

126. Allen, M. P. & Tildesley, D. J. *Computer simulation of liquids* Second (Oxford, 2017).

127. Dauber, P. & Hagler, A. *Biomolecular force fields: where have we been, where are we now, where do we need to go and how do we get there?* (Springer International Publishing, 2018).

128. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function and Bioinformatics* **65,** 712–725 (2006).

129. Shirts, M. R. & Pande, V. S. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *Journal of Chemical Physics* **122,** 1–12 (2005).

130. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water liquid water. *Journal of Chemical Physics* **79,** 926–935 (1983).

131. Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building water models : A different approach. *Journal of Physical Chemistry Letters* **5,** 3863–3871 (2014).

132. Tuckerman, M. E. *Statistical Mechanics: Theory and Molecular Simulation* (Oxford University Press, 2010).

133. Verlet, L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review* **159,** 98–103 (1967).

134. Swope, W. C. *et al.* A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *Journal of Chemical Physics* **637** (1998).

135. Hockney, R. W., Goel, S. P. & Eastwood, J. W. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics* **14,** 148–158 (1974).

136. Van Gunsteren, W. F. & Berendsen, H. J. C. A leap-frog algorithm for stochastic dynamics. *Molecular Simulation* **1,** 173–185 (1988).

137. Beeman, D. Some multistep methods for use in molecular dynamics calculations. *Journal of Computational Physics* **20,** 130–139 (1976).

138. Goga, N., Rzepiela, A. J., De Vries, A. H., Marrink, S. J. & Berendsen, H. J. Efficient algorithms for langevin and DPD dynamics. *Journal of Chemical Theory and Computation* **8,** 3637–3649 (2012).

139. Miyamoto, S. & Kollman, P. A. SETTLE : An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13,** 952–962 (1992).

140. Hess, B. P-LINCS : A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation* **4,** 116–122 (2008).

141. Feenstra, K. A., Hess, B. & Berendsen, H. J. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry* **20,** 786–798 (1999).

142. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *Journal of Chemical Theory and Computation* **11,** 1864–1874 (2015).

143. Woodcock, L. V. Isothermal molecular dynamics calculations for liquid salts. *Chemical Physics Letters* **10,** 257–261 (1971).

144. Basconi, J. E. & Shirts, M. R. Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations. *Journal of Chemical Theory and Computation* **9,** 2887–2899 (2013).

145. Berendsen, H. J. C. & van Gunsteren, W. F. in *Molecular-Dynamics Simulations of Statistical-Mechanical Systems* 43–65 (1986).

146. Harvey, S. C., Tan, R. K. & Cheatham III, T. E. The flying ice cube : Velocity rescaling in molecular dynamics leads to violation of energy equipartition. *Journal of Computational Chemistry* **19,** 726–740 (1997).

147. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **52,** 255–268 (1984).

148. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *Physical review letters* **81,** 3684–3690 (1984).

149. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals : A new molecular dynamics method. *Journal of Applied Physics* **52,** 7182–7190 (1981).

150. Bergdorf, M. *et al.* Influence of cut-off truncation and artificial periodicity of electrostatic interactions in molecular simulations of solvated ions : A continuum electrostatics study. *Journal of Chemical Physics* **119,** 9129–9144 (2003).

151. Frenkel, D. & Smit, B. *Understanding Molecular Simulation* 2nd (Academic Press, 2001).

152. Ewald, P. P. Be Berechmmg optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **369,** 253–287 (1918).

153. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald : An N log (N) method for Ewald sums in large systems in large systems. *Journal of Chemical Physics* **98,** 10089–10092 (1993).

154. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **26,** 1701–1718 (2005).

155. Shirts, M. R., Mobley, D. L., Chodera, J. D. & Pande, V. S. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *Journal of Physical Chemistry B* **111,** 13052–13063 (2007).

156. Wennberg, C. L., Murtola, T., Abraham, M. J., Hess, B. & Lindahl, E. Direct-space corrections enable fast and accurate Lorentz-Berthelot combination rule Lennard-Jones lattice summation. *Journal of Chemical Theory and Computation* **11,** 5737–5746 (2015).

157. Jakalian, A., Bush, B. L., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *Journal of Computational Chemistry* **21,** 132–146 (2000).

158. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry* **23,** 1623–1641 (2002).

159. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **25,** 1157–1174 (2004).

160. D.A. Case *et al. AMBER 2016*

161. Sousa da Silva, A. W. *et al.* ACPYPE - AnteChamber PYthon Parser interfacE. *BioMed Central Research Notes* **5,** 367 (2012).

162. Khoury, G. A., Thompson, J. P., Smadbeck, J., Kieslich, C. A. & Floudas, C. A. Forcefield_PTM: Ab Initio charge and AMBER forcefield parameters for frequently occurring post-translational modification. *Journal of Chemical Theory and Computation* **9,** 5653–5674 (2013).

163. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics* **78,** 1950–1958 (2010).

164. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2,** 19–25 (2015).

165. Vriend, G. WHAT-IF: a molecular modelling and drug design program. *Journal of Molecular Graphics* **8,** 52–56 (1990).

166. Brandt, E. G., Hellgren, M., Brinck, T., Bergman, T. & Edholm, O. Molecular dynamics study of zinc binding to cysteines in a peptide mimic of the alcohol dehydrogenase structural zinc site. *Physical Chemistry Chemical Physics* **11,** 975–983 (2009).

167. Zheng, P. & Li, H. Direct measurements of the mechanical stability of zinc-thiolate bonds in rubredoxin by single-molecule atomic force microscopy. *Biophysical Journal* **101,** 1467–1473 (2011).

168. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **79,** 926–935 (1983).

169. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *Journal of Molecular Graphics and Modelling* **14,** 33–38 (1996).

170. Gowers, R. J. *et al.* MDAnalysis: A python package for the rapid analysis of molecular dynamics simulations. *Proceedings of the 15th Python in Science Conference* **SCIPY,** 98–105 (2016).

171. Kutzner, C. *et al.* More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *Journal of Computational Chemistry* **40,** 2418–2431 (2019).

172. Metropolis, N. The Monte Carlo method. *Journal of the American Statistical Association* **44,** 335–341 (1953).

173. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21,** 1087–1092 (1953).

174. *ProtoMS (v.3.3)* 2017.

175. Adams, D. J. Grand canonical ensemble Monte Carlo for a Lennard-Jones fluid. *Molecular Physics* **29,** 307–311 (1975).

176. Aldeghi, M., Bluck, J. P. & Biggin, P. C. in *Computational Drug Discovery and Design, Methods in Molecular Biology* (eds Gore, M. & Jagtap, U. B.) 199–232 (Humana Press, New York, NY, 2018).

177. Aldeghi, M., Heifetz, A., Bodkin, M. J., Knapp, S. & Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chem. Sci.* **7,** 207–218 (2016).

178. Wang, E. *et al.* End-point binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and applications in drug design. *Chemical Reviews* **119,** 9478–9508 (2019).

179. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* **10,** 449–461 (2015).

180. Duan, L., Liu, X. & Zhang, J. Z. Interaction entropy: A new paradigm for highly efficient and reliable computation of protein-ligand binding free energy. *Journal of the American Chemical Society* **138,** 5722–5728 (2016).

181. Wright, D. W. *et al.* Application of ESMACS binding free energy protocols to diverse datasets: Bromodomain-containing protein 4. *Scientific Reports* **4,** 1–15 (2019).

182. Paissoni, C., Spiliotopoulos, D., Musco, G. & Spitaleri, A. GMXPBSA 2.1: A GROMACS tool to perform MM/PBSA and computational alanine scanning. *Computer Physics Communications* **186,** 105–107 (2015).

183. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *PNAS* **98,** 10037–10041 (2001).

184. Boresch, S., Tettinger, F., Leitgeb, M. & Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *Journal of Physical Chemistry B* **107,** 9535–9551 (2003).

185. Bhati, A. P., Wan, S., Wright, D. W. & Coveney, P. V. Rapid, accurate, precise and reliable relative free energy prediction using ensemble based thermodynamic integration. *Journal of Chemical Theory and Computation* (2016).

186. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *Journal of Chemical Physics* **129,** 124105–124110 (2008).

187. Klimovich, P. V., Shirts, M. R. & Mobley, D. L. Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design* **29,** 397–411 (2015).

188. Chen, W. *et al.* Accurate calculation of relative binding free energies between ligands with different net charges. *Journal of Chemical Theory and Computation* **14,** 6346–6358 (2018).

189. Rocklin, G. J., Mobley, D. L., Dill, K. A. & Hünenberger, P. H. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. *Journal of Chemical Physics* **139** (2013).

190. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *Journal of Chemical Theory and Computation* **7,** 525–537 (2011).

191. Søndergaard, C. R., Olsson, M. H., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p K a values. *Journal of Chemical Theory and Computation* **7,** 2284–2295 (2011).

192. Maier, J. A. *et al.* ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation* **11,** 3696–3713 (2015).

193. Verdonk, M. L., Cole, J. C., Watson, P., Gillet, V. & Willett, P. SuperStar: Improved knowledge-based interaction fields for protein binding sites. *Journal of Molecular Biology* **307,** 841–859 (2001).

194. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **25,** 247–260 (2006).

195. Saubern, S., Guha, R. & Baell, J. B. KNIME workflow to assess PAINS filters in SMARTS format. Comparison of RDKit and indigo cheminformatics libraries. *Molecular Informatics* **30,** 847–850 (2011).

196. *Molecular Operating Environment (MOE), 2018.2* Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.

197. Van Dijk, M., Visscher, K. M., Kastritis, P. L. & Bonvin, A. M. J. J. Solvated protein-DNA docking using HADDOCK. *Journal of Biomolecular NMR* **56,** 51–63 (2013).

198. Trott, O. & Olson, A. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry* **31,** 455–461 (2010).

199. Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics* **47,** –32 (2006).

200. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10,** 845–858 (2016).

201. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spacial restraints. *Journal of Molecular Biology* **234,** 779–815 (1993).

202. Sali, A. & Shen, M.-Y. Statistical potential for assessment and prediction of protein structures. *Protein Science* **15,** 2507–2524 (2006).

203. Braun, E. *et al.* Best practices for foundations in molecular simulations : (v1.0). *LiveCoMS* **1,** 1–27 (2018).

204. See, L. *Development of Ligands for the TRIM33 Bromodomain* PhD thesis (University of Oxford, 2018).

205. Verdonk, M. L., Cole, J. C. & Taylor, R. SuperStar: A knowledge-based approach for identifying interaction sites in proteins. *Journal of Molecular Biology* **289,** 1093–1108 (1999).

206. Liu, K. & Kokubo, H. Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations: A cross-docking study. *Journal of Chemical Information and Modeling* **57,** 2514–2522 (2017).

207. Gill, S. *et al.* Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes Using Nonequilibrium Candidate Monte Carlo. *The Journal of Physical Chemistry B* **122,** 5579–5598 (2018).

208. Miller, T. C. R. *et al.* Competitive binding of a benzimidazole to the histone-binding pocket of the pygo PHD finger. *ACS Chemical Biology* **9,** 2864–2874 (2014).

209. Mohd Siddique, M. U. *et al.* Biphenyl urea derivatives as selective CYP1B1 inhibitors. *Org. Biomol. Chem.* **14,** 8931–8936 (2016).

210. Clark, A. J. *et al.* Prediction of protein ligand binding poses *via* a combination of induced fit docking and metadynamics simulations. *Journal of Chemical Theory and Computation* **12,** 2990–2998 (2016).

211. De Vivo, M., Masetti, M., Bottegoni, G. & Cavalli, A. Role of molecular dynamics and related methods in drug discovery. *Journal of medicinal chemistry* **59,** 4035–4061 (2016).

212. Norman, M. H. *et al.* Novel Vanilloid Receptor-1 antagonists : 1. conformationally restricted analogues of *trans-cinnamides*. *Journal of medicinal chemistry* **50,** 3497–3514 (2007).

213. Bang, S. B. & Kim, J. Efficient dehydrogenation of 1,2,3,4-tetrahydroquinolines mediated by dialkyl azodicarboxylates. *Synthetic Communications* **48,** 1291–1298 (2018).

214. Kang, I. J. *et al.* Synthesis, activity, and pharmacokinetic properties of a series of conformationally-restricted thiourea analogs as novel hepatitis C virus inhibitors. *Bioorganic and Medicinal Chemistry* **18,** 6414–6421 (2010).

215. Hopkins, A. L., Keserü, G. M., Leeson, P. D., Rees, D. C. & Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. *Nature reviews. Drug discovery* **13,** 105–21 (2014).

216. Liu, J. *et al.* Structural plasticity of the TDRD3 Tudor domain probed by a fragment screening hit. *FEBS Journal* **285,** 2091–2103 (2018).

217. Chodera, J. D. *et al.* Alchemical free energy methods for drug discovery: progress and challenges. *Current Opinion in Structural Biology* **21,** 150–160 (2011).

218. Borysko, P. *et al.* Straightforward hit identification approach in fragment-based discovery of bromodomain-containing protein 4 (BRD4) inhibitors. *Bioorganic and Medicinal Chemistry* **26,** 3399–3405 (2018).

219. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: A free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling* **52,** 1757–1768 (2012).

220. Rudling, A. *et al.* Fragment-based discovery and optimization of enzyme inhibitors by docking of commercial chemical space. *Journal of Medicinal Chemistry* **60,** 8160–8169 (2017).

221. Gaulton, A. *et al.* The ChEMBL database in 2017. *Nucleic Acids Research* **45,** D945–D954 (2017).

222. Davies, M. *et al.* ChEMBL web services: Streamlining access to drug discovery data and utilities. *Nucleic Acids Research* **43,** 612–620 (2015).

223. Malhotra, S. & Karanicolas, J. When Does Chemical Elaboration Induce a Ligand To Change Its Binding Mode? *Journal of Medicinal Chemistry* **60,** 128–145 (2017).

224. Liao, C. & Nicklaus, M. C. Comparison of nine programs predicting pKa values of pharmaceutical substances. *Journal of Chemical Information and Modeling* **49,** 2801–2812 (2009).

225. Wan, S., Bhati, A. P., Zasada, S. J., Wall, I. & Green, D. Rapid and reliable binding affinity prediction for analysis of bromodomain inhibitors : a computational study. *Journal of Chemical Theory and Computation,* 1–30 (2016).

226. Wang, W., Wang, L. & Liang, Y. S. Susceptibility or resistance of praziquantel in human schistosomiasis: A review. *Parasitology Research* **111,** 1871–1877 (2012).

227. El-Sayed, N. M. *et al.* The genome sequence of *Trypanosoma cruzi*, etiologic agent of chagas disease. *Science* **309,** 409–416 (2005).

228. Alonso, V. L., Ritagliati, C., Cribb, P., Cricco, J. A. & Serra, E. C. Overexpression of bromodomain factor 3 in *Trypanosoma cruzi* (*Tc*BDF3) affects differentiation of the parasite and protects it against bromodomain inhibitors. *FEBS Journal* **283,** 2051–2066 (2016).

229. Jeffers, V., Yang, C., Huang, S. & Sullivan, W. J. Bromodomains in protozoan parasites: Evolution, function, and opportunities for drug development. *Microbiology and Molecular Biology Reviews* **81,** 1–17 (2017).

230. Ramallo, I. A., Alonso, V. L., Rua, F., Serra, E. & Furlan, R. L. A bioactive *Trypanosoma cruzi* bromodomain inhibitor from chemically engineered extracts. *ACS Combinatorial Science* **20,** 220–228 (2018).

231. Jennings, L. E. *et al.* BET bromodomain ligands: Probing the WPF shelf to improve BRD4 bromodomain affinity and metabolic stability. *Bioorganic and Medicinal Chemistry* **26,** 2937–2957 (2018).

232. Boratyn, G. M. *et al.* Domain enhanced lookup time accelerated BLAST. *Biology Direct* **7,** 12 (2012).

233. Ran, X. *et al.* Structure-based design of γ-carboline analogues as potent and specific BET bromodomain inhibitors. *Journal of Medicinal Chemistry* **58,** 4927–4939 (2015).

234. Laurin, C. M. C. *Chemical tools to probe the role of bromodomains in the parasite Trypanosoma cruzi* PhD thesis (University of Oxford, 2019).

235. Vollmuth, F. & Geyer, M. Interaction of propionylated and butyrylated histone H3 lysine marks with Brd4 bromodomains. *Angewandte Chemie - International Edition* **49,** 6768–6772 (2010).

236. Cox, O. B. *et al.* A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chemical Science* **7,** 2322–2330 (2016).

237. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein–ligand docking using GOLD. *PROTEINS: Structure, Function, and Genetics* **52,** 609–623 (2003).

238. Korb, O., Stützle, T. & Exner, T. E. Empirical scoring functions for advanced Protein-Ligand docking with PLANTS. *Journal of Chemical Information and Modeling* **49,** 84–96 (2009).

239. Yamaotsu, N. & Hirono, S. In silico fragment-mapping method: a new tool for fragment-based/structure-based drug discovery. *Journal of Computer-Aided Molecular Design* **32,** 1229–1245 (2018).

240. Arcon, J. P. *et al.* Molecular dynamics in mixed solvents reveals protein-ligand interactions, improves docking and allows accurate binding free energy predictions. *Journal of Chemical Information and Modeling* **57,** 846–863 (2017).

241. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149,** 214–231 (2012).

242. Wu, Q. *et al.* A chemical toolbox for the study of bromodomains and epigenetic signaling. *Nature Communications* **10,** 1915 (2019).

243. Cochran, A. G., Conery, A. R. & Sims, R. J. Bromodomains: a new target class for drug development. *Nature Reviews Drug Discovery* **18,** 609–628 (2019).

244. Hewings, D. S. *et al.* 3,5-Dimethylisoxazoles act as acetyl-lysine-mimetic bromodomain ligands. *Journal of Medicinal Chemistry* **54,** 6761–6770 (2011).

245. Brand, M. *Developing small molecule ligands for the study of bromodomain-histone interactions* PhD thesis (University of Oxford, 2016).

246. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7,** 1–6 (2011).

Copyright Clearance Center

Confirmation Number: 11856696
Order Date: 10/03/2019

Customer Information

Customer: Joseph Bluck
Account Number: 3001529571
Organization: University of Oxford
Email: joseph.bluck@dtc.ox.ac.uk
Phone: +44 1865613200
Payment Method: Invoice

This is not an invoice

Order Details

Future medicinal chemistry

Billing Status: **N/A**

Order detail ID: 72029114
ISSN: 1756-8919
Publication Type: Journal
Volume:
Issue:
Start page:
Publisher: Future Science

Permission Status: **Granted**
Permission type: Republish or display content
Type of use: Thesis/Dissertation
Order License Id: 4681500854878

Requestor type	Academic institution
Format	Print
Portion	image/photo
Number of images/photos requested	1
The requesting person/organization	Joseph Bluck
Title or numeric reference of the portion(s)	Chapter 1, Figure 1.1
Title of the article or chapter the portion is from	Introduction
Editor of portion(s)	Joseph Bluck
Author of portion(s)	Joseph Bluck
Volume of serial or monograph	N/A
Page range of portion	1-26
Publication date of portion	October 2019
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	U.K. and Commonwealth (excluding Canada)
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity	

of new product	Up to 499
Title	DPhil Student
Institution name	University of Oxford
Expected presentation date	Nov 2019

Note: This item was invoiced separately through our **RightsLink service.** More info **$ 0.00**

Total order items: 1

Order Total: $0.00

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Oct 03, 2019

This Agreement between University of Oxford -- Joseph Bluck ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4681530438169
License date	Oct 03, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Molecular Cell Biology
Licensed Content Title	Functions of bromodomain-containing proteins and their roles in homeostasis and cancer
Licensed Content Author	Takao Fujisawa, Panagis Filippakopoulos
Licensed Content Date	Jan 5, 2017
Licensed Content Volume	18
Licensed Content Issue	4
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	DPhil Student
Institution name	University of Oxford
Expected presentation date	Nov 2019
Portions	Figure 1b
Requestor Location	University of Oxford 72B Walton Street Oxford, other OX2 6AG United Kingdom Attn: University of Oxford
Total	0.00 GBP
Terms and Conditions	

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you

Title: Progress in the Development and Application of Small Molecule Inhibitors of Bromodomain–Acetyl-lysine Interactions

Author: David S. Hewings, Timothy P. C. Rooney, Laura E. Jennings, et al

Publication: Journal of Medicinal Chemistry

Publisher: American Chemical Society

Date: Nov 1, 2012

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE LICENSE TERMS AND CONDITIONS

Oct 03, 2019

This Agreement between University of Oxford -- Joseph Bluck ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

License Number	4681531145746
License date	Oct 03, 2019
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	The *Trypanosoma cruzi* Proteome
Licensed Content Author	J. A. Atwood,D. B. Weatherly,T. A. Minning,B. Bundy,C. Cavola,F. R. Opperdoes,R. Orlando,R. L. Tarleton
Licensed Content Date	Jul 15, 2005
Licensed Content Volume	309
Licensed Content Issue	5733
Volume number	309
Issue number	5733
Type of Use	Thesis / Dissertation
Requestor type	Scientist/individual at a research institution
Format	Print
Portion	Text Excerpt
Number of pages requested	1
Order reference number	
Title of your thesis / dissertation	DPhil Student
Expected completion date	Nov 2019
Estimated size(pages)	1
Requestor Location	University of Oxford 72B Walton Street Oxford, other OX2 6AG United Kingdom Attn: University of Oxford
Total	0.00 GBP

Terms and Conditions

American Association for the Advancement of Science TERMS AND CONDITIONS Regarding your request, we are pleased to grant you non-exclusive, non-transferable permission, to republish the AAAS material identified above in your work identified above, subject to the terms and conditions herein. We must be contacted for permission for any uses

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Oct 03, 2019

This Agreement between University of Oxford -- Joseph Bluck ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4681540060540
License date	Oct 03, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Disease Primers
Licensed Content Title	Schistosomiasis
Licensed Content Author	Donald P. McManus et al
Licensed Content Date	Aug 9, 2018
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	DPhil Student
Institution name	University of Oxford
Expected presentation date	Nov 2019
Portions	Figure 1
Requestor Location	University of Oxford 72B Walton Street Oxford, other OX2 6AG United Kingdom Attn: University of Oxford
Total	0.00 GBP
Terms and Conditions	

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Oct 03, 2019

This Agreement between University of Oxford -- Joseph Bluck ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4681540335063
License date	Oct 03, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Microbiology
Licensed Content Title	Epigenetic regulation in African trypanosomes: a new kid on the block
Licensed Content Author	Luisa M. Figueiredo et al
Licensed Content Date	Jul 1, 2009
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	DPhil Student
Institution name	University of Oxford
Expected presentation date	Nov 2019
Portions	Figure 3
Requestor Location	University of Oxford 72B Walton Street Oxford, other OX2 6AG United Kingdom Attn: University of Oxford
Total	0.00 GBP
Terms and Conditions	

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

Copyright Clearance Center

Back to view orders

Copy order

Print this page
Print terms & conditions
Print citation information
(What's this?)

Confirmation Number: 11856732
Order Date: 10/03/2019

Customer Information

Customer: Joseph Bluck
Account Number: 3001529571
Organization: University of Oxford
Email: joseph.bluck@dtc.ox.ac.uk
Phone: +44 1865613200

Search order details by: Choose One Go

This is not an invoice

Order Details

Annual review of biophysics

Billing Status: **N/A**

Order detail ID: 72029299
ISSN: 1936-1238
Publication Type: e-Journal
Volume:
Issue:
Start page:
Publisher: Annual Reviews

Permission Status: Granted
Permission type: Republish or display content
Type of use: Thesis/Dissertation
Order License Id: 4681540906450

Hide details

Requestor type	Academic institution
Format	Print
Portion	image/photo
Number of images/photos requested	1
The requesting person/organization	Joseph Bluck
Title or numeric reference of the portion(s)	Figure 4
Title of the article or chapter the portion is from	N/A
Editor of portion(s)	Joseph Bluck
Author of portion(s)	Joseph Bluck
Volume of serial or monograph	N/A
Page range of portion	1
Publication date of portion	October 2019
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing	

privileges	yes
For distribution to	U.K. and Commonwealth (excluding Canada)
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	Up to 499
Title	DPhil Student
Institution name	University of Oxford
Expected presentation date	Nov 2019

Note: This item was invoiced separately through our **RightsLink service.** More info **$ 0.00**

Total order items: 1

Order Total: $0.00