RATIONALISING THE INHIBITION OF *M. TUBERCULOSIS* MSHB BY A SERIES OF INHIBITORS CONSTRUCTED FROM PLUMBAGIN CONJUGATED *VIA* A VARIABLE ALKYL LINKER TO A PHENYL THIOGLYCOSIDE

IAN L. ROGERS

RATIONALISING THE INHIBITION OF *M. TUBERCULOSIS* MSHB BY A SERIES OF INHIBITORS CONSTRUCTED FROM PLUMBAGIN CONJUGATED *VIA* A VARIABLE ALKYL LINKER TO A PHENYL THIOGLYCOSIDE

Dissertation presented to the

UNIVERSITY of CAPE TOWN

In fulfilment of the requirements for the degree of

MASTER OF SCIENCE

By

IAN L. ROGERS

Supervisor: Professor Kevin J. Naidoo

Co-supervisor: Assoc. Professor David W. Gammon

Scientific Computing Research Unit

University of Cape Town

2011

DECLARATION

I declare that RATIONALISING THE INHIBITION OF *M. TUBERCULOSIS* MSHB BY A SERIES OF INHIBITORS CONSTRUCTED FROM PLUMBAGIN CONJUGATED *VIA* A VARIABLE ALKYL LINKER TO A PHENYL THIOGLYCOSIDE is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

_____

Ian Rogers

ACKNOWLEDGEMENTS

Firstly I would like to express my sincere gratitude to Professor Kevin Naidoo, who provided the computational resources and valuable guidance for this dissertation. I am indebted to my co-supervisor Assoc. Prof. David Gammon for his understanding and his help in exploring (the many!) opportunities in chemistry. I acknowledge the funding provided to me by the NRF and UCT. Finally I would like to thank the SCRU group for all their help and companionship, especially Richard Matthews who is always willing to help, and Chris Barnett for his system administrations.

ABSTRACT

Infection by *M. tuberculosis* results in an estimated 1.7 million TB related deaths worldwide. Mycothiol is produced in *M. tuberculosis* as the dominant low molecular weight thiol and is thought to protect the bacteria against oxidative stress. Since mycothiol is unique to *Actinomycetes* and is also proposed to play an important role in the dormant state of *Mycobacteria*, the pseudo-dissacharide is seen as a potential target for novel anti-tuberculars. Targeting the mycothiol redox cycle has led to MshB inhibition by a series of substrate analogues. These inhibitors comprise plumbagin conjugated to a phenyl thioglycoside *via* an alkyl chain of variable length. Kinetics studied showed that the competitive inhibition increased when the alkyl linker was lengthened. The binding of the inhibitors was investigated using computational techniques in order to rationalise the observed trend in inhibition.

The binding of the MshB natural substrate, 1-D-*myo*-inosityl-2-acetamido-2-deoxy-α-D-glucopyranoside (GlcNAc-Ins), is unconfirmed in the literature. Therefore, the MshB active site was characterised and the enzyme was docked with GlcNAc-Ins. Atomistic molecular dynamics (MD) simulations were then used to identify key interactions. The GlcNAc binding mode was used as a template for docking the substrate-based inhibitors. Key interactions between MshB and the inhibitors were elucidated using MD, and the validity of the model was tested using free energy methods. The ranking of the inhibitor series based on their relative free energy of binding was compared with experimental inhibition constants. The free energy calculations showed that the binding of the linker and plumbagin does not occupy a hydrophobic cavity adjacent to the putative active site, but rather a kink in the alkyl chain directs the plumbagin towards the active site entrance.

.

ABBREVIATIONS

2D	Two dimensional
3D	Three Dimensional
Å	Angstroms
BOG	β-octyl glucoside
CHARMM	Chemistry at Harvard Macromolecular Mechanics
COM	Centre of Mass
EPS	Electrostatic Potential Surface
ELE	Electrostatic
FEP	Free Energy Perturbations
GABC	General-acid-base catalysis
GlcNAc	*N*-Acetylglucosamine
GlcNAc-Ins	1-D-*myo*-inosityl-2-acetamido-2-deoxy-α-D-glucopyranoside
GlcNAc-SPh	Phenyl-2-deoxy-2-acetamido-1-thio-b-D-glucopyranoside
IE	Interaction Energy
MM	Molecular Mechanics
MD	Molecular Dynamics
MSH	1-*O*-[2-[(2R)-2-acetamido-3-mercaptopropanamido]-2-deoxy-α-D-glucopyranosyl]-D-myo-inositol (Mycothiol)
Mca	Mycothiol-*S*-conjugate amidase (Mca)
MshA	glycosyltransferase
MshB	GlcNAc-Ins deacetylase
MshC	Mycothiol ligase
MshD	Mycothiol synthase
Mtr	Mycothiol disulfide reductase
ns	Nanosecond
NVT	Canonical Ensemble
ps	Picosecond
QM	Quantum Mechanics

RMSD	Root Mean Square Deviation
SBD	Stochastic Boundary Dynamics
SG	Slow Growth
TI	Thermodynamic Integration
vdW	van der Waals

LIST OF SOFTWARE PROGRAMS USED IN THIS DISSERTATION

APBS, version 1.3

CASTp, webserver

CHARMM, version ch33b2

GAUSSIAN, version G09

Glide, version 5.6

Jmol, version 12.0.17

Maestro, version 9.1

PyMol, version 1.3r3pre

SiteMap, version 2.4

VMD, version 1.8.7

CONTENTS

CHAPTER 1:

CHAPTER 2:

COMPUTER SIMULATION METHODS

CHAPTER 4:

MSHB ACTIVE SITE AND NATURAL SUBSTRATE BINDING

CHAPTER 5:

RATIONALISING THE INHIBITION OF MSHB BY A SERIES OF PLUMBAGIN CONJUGATES

CHAPTER 6

APPENDIX

CHAPTER 1

1.1. MYCOTHIOL AS A TARGET FOR NOVEL ANTI-TUBERCULARS

1.1.1. TUBERCULOSIS

Tuberculosis (TB) is a highly contagious disease caused by *Mycobacterium tuberculosis* [1]. The symptoms of active TB are coughing blood with sputum, chest pains, weakness, weight loss, fever and night sweats. It is estimated that one-third of the world's population is infected with TB bacilli [1], which are transmitted *via* droplets from the throat and lungs of hosts with the active respiratory disease. Once inhaled into the lung the bacteria are phagocytosed by alveolar macrophages and induce an inflammatory response (Figure 1-1). Mononuclear cells from neighbouring blood vessels are recruited and form the building blocks for a granulum (or tubercule). In the containment phase non-replicating *M. tuberculosis* cells lie dormant within the granuloma and the host displays no symptoms. When the host's immune system is compromised the granuloma ruptures and spills thousands of viable, infectious bacilli into the airways. Such a change in the immune status is one which reduces the number, or impairs the function of $CD4^+$ T cells [2]. Thus active TB has a close association with age, malnutrition and acquired immunodeficiency deficiency virus (AIDS) [1]. TB has been classified by the World Health Organisation (WHO) as a pandemic and in 2009 the disease resulted in an estimated 1.7 million deaths worldwide [3]. TB is most prevalent in Asia and southern Africa. In 2009 South Africa reported the third highest number of incidence cases behind India and China [3].

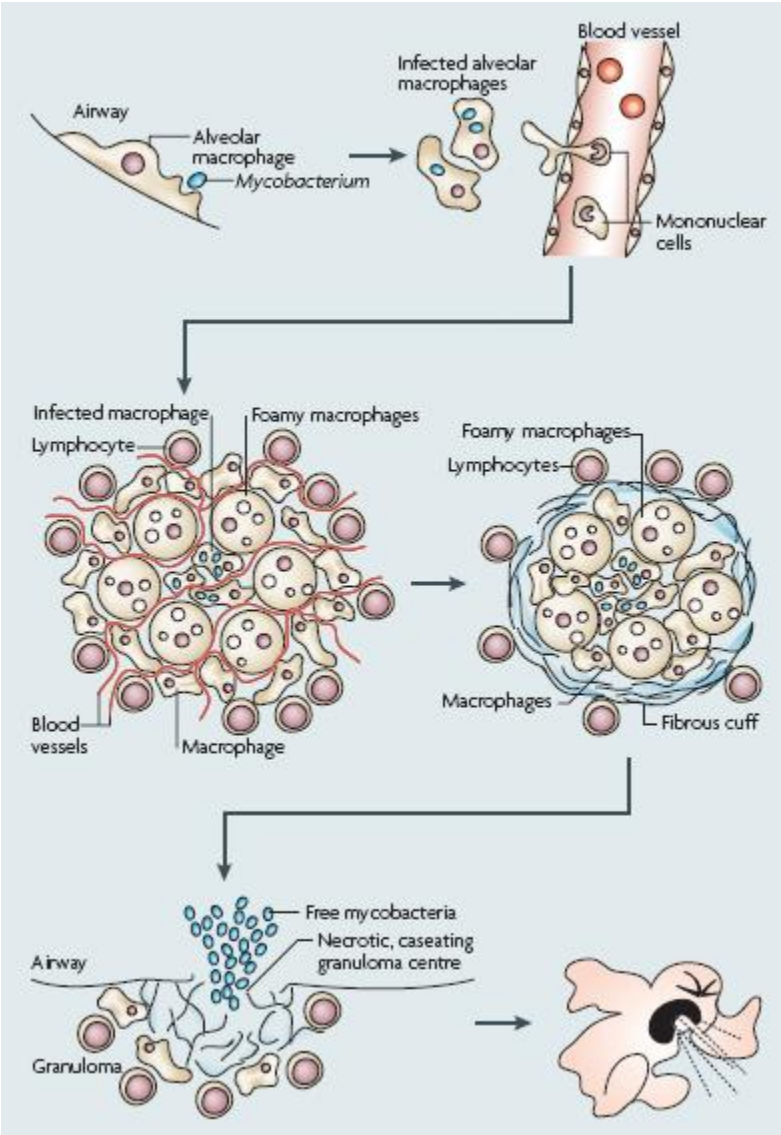


FIGURE 1-1 Pathology of Tuberculosis. Figure taken from Russell [2].

Traditionally TB has been treated by a six month course of isoniazid and rifampicin which are known as first-line drugs. Partial or inconsistent treatment of TB has led to antibiotic resistance in *M. tuberculosis* and the emergence of multidrug-resistant TB (MDR-TB[1]) and extensively drug-resistant TB (XDR-TB[2]). Treatment of these cases require extensive chemotherapy using drugs that are costlier and more toxic than the first line drugs [1, 4]. MDR-TB and XDR-TB are a serious emerging threat to global public health, especially, in countries with a high prevalence of

1 MDR-TB is defined as TB resistant to the two main first-line drugs (isoniazid and rifampicin)

2 XDR-TB is defined as TB resistant to multiple drugs as well as to any one of the flouroquinolone drugs and to at least one of the three injectable second-line drugs (amikacin, capreomycin or kanamycin)

the human immunodeficiency virus (HIV). In 2009 South Africa had the fourth largest number of estimated cases of XDR-TB (13 000) behind China, India and the Russian Federation [3].

It is feared that XDR-TB could replace the current TB epidemic of mostly drug-susceptible TB with a form of TB with severely restricted treatment options [4]. In 2006, WHO launched the new STOP TB Strategy which aims to reverse the incidence of TB by 2015 and eliminate TB as a public health problem by 2050. A key point of this strategy is the research and development of new diagnostics, drugs and vaccines against drug resistant strains [1].

Thus, development of a new treatment for TB has become an international priority. Furthermore, effectiveness against latent *M. tuberculosis* is considered a prime criterion for new tuberculosis drugs [5]. Many drugs that are lethal to growing *M. tuberculosis* cells fail to kill the bacterium in the latent or dormant state (non-replicating and persistant). When treatment is discontinued reactivation can result in a further episode of active tuberculosis. In this context, there has been much focus on mycothiol as a target for novel anti-tuberculars.

1.1.2. MSH: STRUCTURE, DISTRIBUTION AND FUNCTION

STRUCTURE AND BIOSYNTHESIS

Mycothiol (MSH, Figure 1-2), or 1-*O*-[2-[(2R)-2-acetamido-3-mercaptopropanamido]-2-deoxy-α-D-glucopyranosyl]-D-myo-inositol, comprises a cysteine residue in which the amino group is acetylated and the carboxyl group is linked by a unique amide bond to D-glucosamine, which is in turn α(1-1) linked to D-myo-inositol. MSH has been found to occur only in *Actinobacteria*, the class which includes *M. tuberculosis*, where it is mostly produced as the dominant low-molecular-weight (LMW) thiol [5]. MSH cycles between its reduced and oxidised disulfide forms ($MSSM + 2H^+ + 2e^- \rightleftharpoons 2MSH$), but the reducing action of the constitutively active MSH-specific reductase, Mtr, maintains a high intracellular MSH/MSSM ratio.

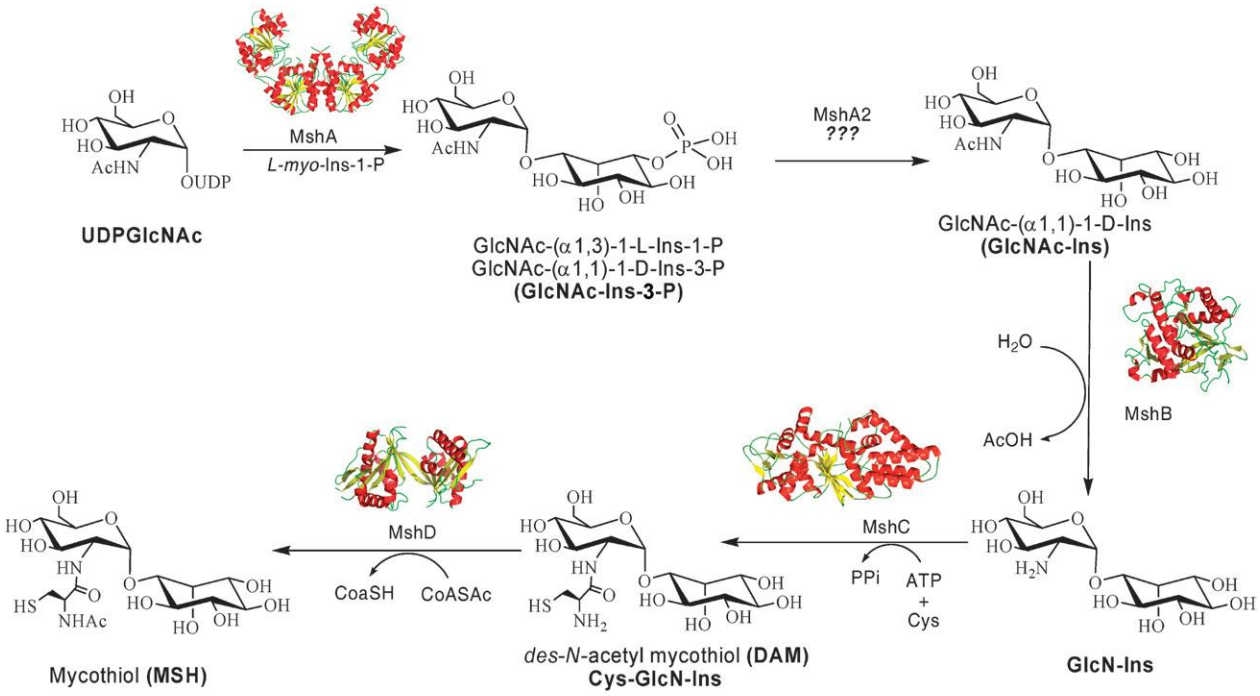


FIGURE 1-2 Biosynthesis of MSH. Initially 1L-*myo*-inositol-1-phosphate is produced from glucose-6-phosphate by inositol phosphate synthase (Ino1). MSH biosynthesis is then accomplished in five steps. The enzymes involved in the five-step pathway include a glycosyltransferase (MshA), a phosphatase (MshA2), a deacetylase (MshB), an ATP-dependent ligase (MshC), and an acetyltransferase (MshD). Figure taken from Hamilton *et al.* [6].

The biosynthetic pathway for MSH has been elucidated and is presented in Figure 1-2. Measurement of thiol levels in *M. smegmatis* mutants with targeted disruptions in each of *mshA*, *mshB*, *mshC* and *mshD* have shown that MshA and MshC are essential to MSH biosynthesis in Mycobacteria [5]. On the other hand the *M. tuberculosis mshB* mutant produced 20% of normal MSH levels during exponential growth, but showed increased levels over the wild-type with prolonged culture [7]. This indicates that the cell contains an additional activity capable of deacetylating GlcN-Ins. The *M. tuberculosis mshD*::Tn5 mutant produced ~1% of normal MSH levels by transacetylation of Cys-GlcN-Ins by acetyl-CoA, but formyl-GlcN-Ins was detected at high yields and may partially substitute for MSH in MshD-deficient mutants [8].

FUNCTION OF MSH IN *M. TUBERCULOSIS*

Many of the functions of MSH in *Actinobacteria* have been extrapolated by analogy to glutathione (GSH), which is the major LMW thiol in most other organisms [6, 9]. Comparison

with GSH suggests that MSH serves as the major cellular redox buffer and helps to maintain an effective reducing environment within the cell. This is essential for metabolism in aerobic cells and the redox status must be maintained in the face of the oxidising threat from metabolic intermediates. *M. tuberculosis* must also protect itself against the highly oxidising extracellular environment of host macrophages, which has been shown to make the boundary to the central region of granulomas hypoxic [2]. The retaining mechanisms of MSH are not fully known, but it is understood that high concentrations of the thiol form (maintained through the action of Mtr) makes a large contribution to the reducing environment [5, 6]. Further MSH-dependent mechanisms that protect the *M. tuberculosis* cells from oxidation stress include: detoxifying reactive oxygen and nitrogen species; detoxifying electrophillic xenobiotics; and serving as a reservoir of stabilised cysteine that is less susceptible to metal catalysed auto-oxidation and resultant hydrogen peroxide formation [5, 6]. The activities of the well characterised MSH-dependent enzymes are summarised in Figure 1-3. It has been shown that MSH S-conjugate Amidase, Mca, may play an important role in drug resistance of *M. tuberculosis* [6, 9]. The enzyme does so by cleaving conjugates of MSH and thiol-reactive drugs (MSR in Figure 1-3), and excreting the mercapturic acid produced.

1.1.3. MSH AS A DRUG TARGET

A key trait of a drug target is its essentiality to the pathogen. Hence, the presence of MSH as the dominant thiol in *M. tuberculosis* has sparked much interest in the pseudo-disaccharide as a target for novel anti-tuberculars. Although MSH plays an important function as redox buffer, it is still unclear as to whether it is in fact essential [5]. Early attempts that were unsuccessful in isolating *M. tuberculosis* MSH-free mutants by a targeted disruption in *mshA* [10] and *mshC* [11] led to the conclusion that the thiol is needed for growth. Recently, however MshD mutants have been reported [12, 13] and cast doubt on the previous assertions [5]. Nevertheless, targeting the redox cycling of MSH remains an important drug-discovery strategy.

Recently a study has shown that most bactericidal antibiotics ultimately act by generating hydroxyl radicals [14]. Since MSH plays a fundamental role in protection against oxidising

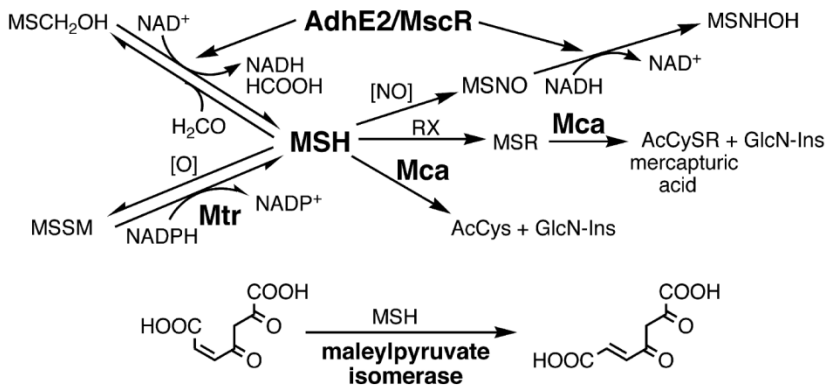


FIGURE 1-3 Summary of well characterised reactions where MSH is either a substrate of co-factor. Illustration taken from Fahey *et al.* [5].

stress, the MSH redox pathway appears to be a good target for combinatorial chemotherapy with existing antibiotics. Indeed, Mycobacterial MSH mutants have been shown to be more sensitive to a number of antibiotics including erythromycin, azithromycin, penicillin G and vancomycin [15]. Several other factors highlight MshB as an important drug target:

- Chemotherapy: Since there is no mammalian counterpart to the MSH pathway it should be possible to achieve selective inhibition of MSH biosynthesis and metabolism.

- Targeting antibiotic resistance: *S*-conjugation and subsequent action of Mca has been identified as a possible source of antibiotic resistance [5, 6]. MSH inhibition could inhibit this activity allowing efficacy of existing first line antibiotics.

- Targeting of dormant state: The significant increase in MSH levels observed in *M. tuberculosis* cultures in the transition from log to stationary phase suggests that MSH may be required for *M. tuberculosis* in attaining a state of dormancy. Furthermore, there is evidence that the enzymes of MSH biosynthesis and metabolism are active in the dormant state [5, 6]. Drugs targeting MSH may be effective against dormant tuberculosis, which has been identified as an important goal for novel anti-tuberculars.

In targeting the MSH pathway in order to develop novel anti-tubercular drugs, lead compounds should inhibit MSH biosynthesis and/or the MSH-dependent enzymes that are critical to the

survival of replicating and dormant cells of *M. tuberculosis*. The former strategy would simultaneously inactivate all MSH-dependent metabolic processes. Therefore, the biosynthetic enzymes that are essential for MSH production, namely MshA and MshC, are prime targets. However, the search for drugs directed against them has been limited by the inability to readily generate active, purified protein, and by lack of available crystal structures [5, 6]. MshB and MshD are less promising [5, 6], since mutants with targeted disruptions in *mshA* and *mshD* do not represent MSH-free states. Nevertheless, as explained above, reducing MSH levels in *M. tuberculosis* cells may still yield the pathogen susceptible to antimicrobial factors of the host immune response and antibiotics. This hypothesis is supported by research which showed that a MshD mutant of *M. tuberculosis* failed to grow in cultures of primary murine macrophages that modelled the host immune response. Thus, MshA and MshB remain potential drug targets, especially if relevant drugs were administered in combination with known bactericidal antibiotics. Research into targeting the MSH redox pathway has focused on the design of substrate-based inhibitors of deacetylase MshB and related Mca.

1.2. MSHB

1.2.1. CRYSTAL STUCTURE

GlcNAc-Ins *N*-deacetylase or MshB is the third enzyme in the MSH biosynthetic pathway and catalyses hydrolysis of the *N*-acetyl amide bond of GlcNAc-Ins to give the free amino sugar GlcN-Ins. MshB has also been shown to have overlapping amidase activity with Mca, although the rate of amide cleavage is significantly lower in MshB for most substrates [5].

Two crystal structures of *M. tuberculosis* MshB have been reported [16, 17]. The first structure, solved by James *et al.* (1Q74) [16], contains four protein molecules in the unit cell and the topology of Molecule A is given in Figure 1-4. In the tertiary structure a nine-stranded mixed β-sheet is arranged with an outer layer of α-helices into α/β domain. The first five strands of the

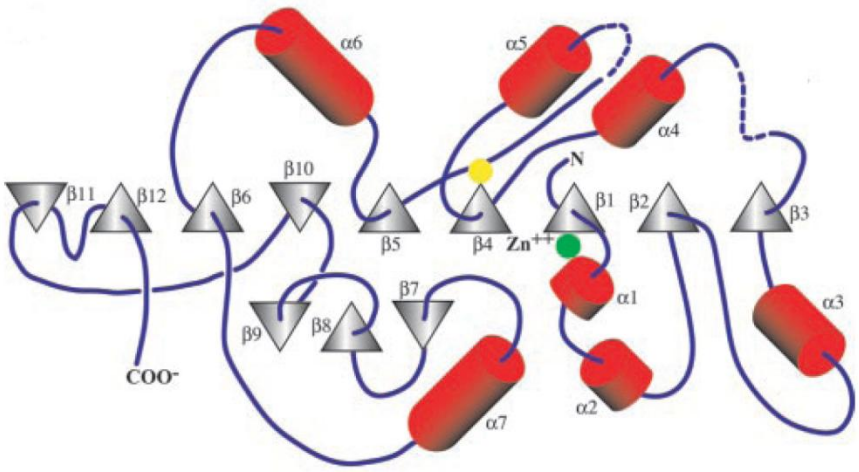


FIGURE 1-4 Secondary structure of MshB 1Q74. The *green circle* denotes the location of two of the zinc binding ligands (His13 and Asp16) and of the general base for the proposed catalytic mechanism (Asp15). Illustration taken from James *et al.* [16].

core β-sheet, together with helices α_1 through to α_5, adopt a supersecondary structure similar to the Rosman fold. A three-stranded anti-parallel β-sheet (β_7, β_8, β_9) forms a subdomain with α_7. His13 from the loop linking β_1 and α_1, Asp16 from the C-terminus of β1 and His147 from the N-terminus of helix α_5 comprise a zinc-binding site. The holoenzyme showed two water molecules that coordinate to the Zn^{2+} ion (Figure 1-5B). Molecule A is the most complete structure from 1Q74 but is missing residues 1-2, 100-103, 164-169, 300-303. Baker *et al.* have also published a crystal structure (1Q7T) [17] in which the first of two molecules in the asymmetric unit is missing only the loop formed by residues 165-170 (Figure 1-5). β-octylglucoside, used in the crystallisation process, was co-crystallised near the zinc binding site and its binding is illustrated in Figure 1-6A. The β-octyl glucoside is an analogue of the glucose unit in the natural substrate, and was used to model GlcNAc binding in accordance with the proposed catalytic mechanism discussed below (Section 1.2.2, Figure 1-6C). In this model a zinc binding site consistent with 1Q74 was generated by flipping the His13 side chain so that its $N^{\delta 1}$ approached Zn^{2+}. The authors were then able to build a 2-C-acetylamido group onto the BOG template such that the carbonyl was coordinated to the zinc. The resulting orientation of GlcNAc is illustrated in (Figure 1-6B).

The sugar binding is mediated by hydrogen bonding. The oxygens of O3 and O4 are hydrogen bonded to Arg68, while the hydrogens at O4 and O6 are hydrogen bonded to the carboxylate of Asp95. A hydrogen bond is also modelled between His144 and the oxygen at O6.

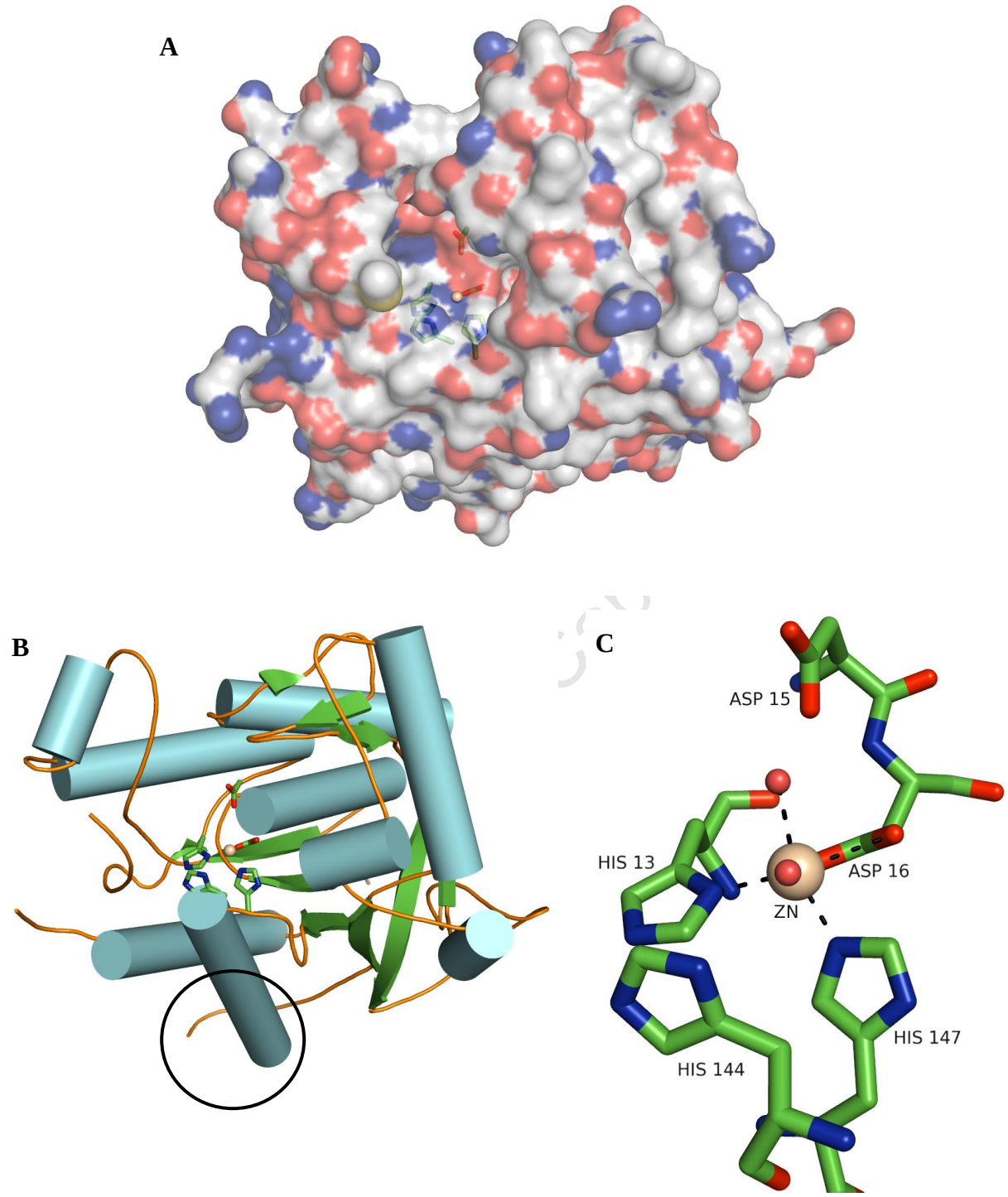


FIGURE 1-5 **(A) Surface representation of the crystal structure of molecule A in 1Q74. The catalytic zinc is shown as a wheat coloured sphere. (B) Molecule A illustrated in cartoon representation with the same angle of view as Figure 1-5A. The catalytic zinc is shown as a wheat coloured sphere. The circle indicates the position of the loop comprising residues 164-170 which is missing in this structure and in 1Q7T. (C) The zinc-binding pocket is magnified and shows the two zinc coordinated waters. The same angle of view as in Figure 1-5A and Figure 1-5B is used.**

1.2.2. CATALYTIC MECHANISM AND NATURAL SUBSTRATE BINDING

MshB is a zinc-dependent deacetylase. Typically the catalytic mechanism of this class of enzymes follows nucleophillic attack of a zinc-coordinated water with zinc stabilisation of the intermediate and general-acid-base catalysis (GABC) [18].

A catalyst increases the rate of a reaction by decreasing the free energy of the transition state. In general acid catalysis this is achieved by partial proton transfer from an acid, while in general base catalysis the free energy of the transition state is lowered by partial proton abstraction by a base [19]. When both processes are present the catalysis is referred to as GABC. Enzymes are well suited to GABC since several catalytic residues (either acidic or basic) can be arranged around the substrate. In a large number of cases, the mechanism of zinc-dependent deacetylases is initiated with a basic residue partially abstracting a hydrogen from a zinc-coordinated water [18]. The water then undergoes a nucleophillic attack on the acetyl carbonyl group of the substrate. Finally, an acidic residue transfers a proton to the resulting tetrahedral intermediate to promote conversion to the deacetylated product and acetic acid. There are two types of GABC mechanisms. In the first mechanism a single bifunctional residue serves as the basic residue and once protonated can function as the acidic residue. In the second mechanism a pair of amino acids provide the acid base catalysis.

The proposed GABC mechanism for MshB includes a GABC pair and is shown in Figure 1-6C. The mechanism is based on the similarity of the active site residues in MshB with those found in well characterised metalloproteases. It is postulated that the substrate binds to the active site, so that the carbonyl oxygen of the *N*-acetyl coordinates to the Zn^{2+} cation, and displaces a water molecule in the process. The polarised C=O then undergoes a nucleophilic attack to form a

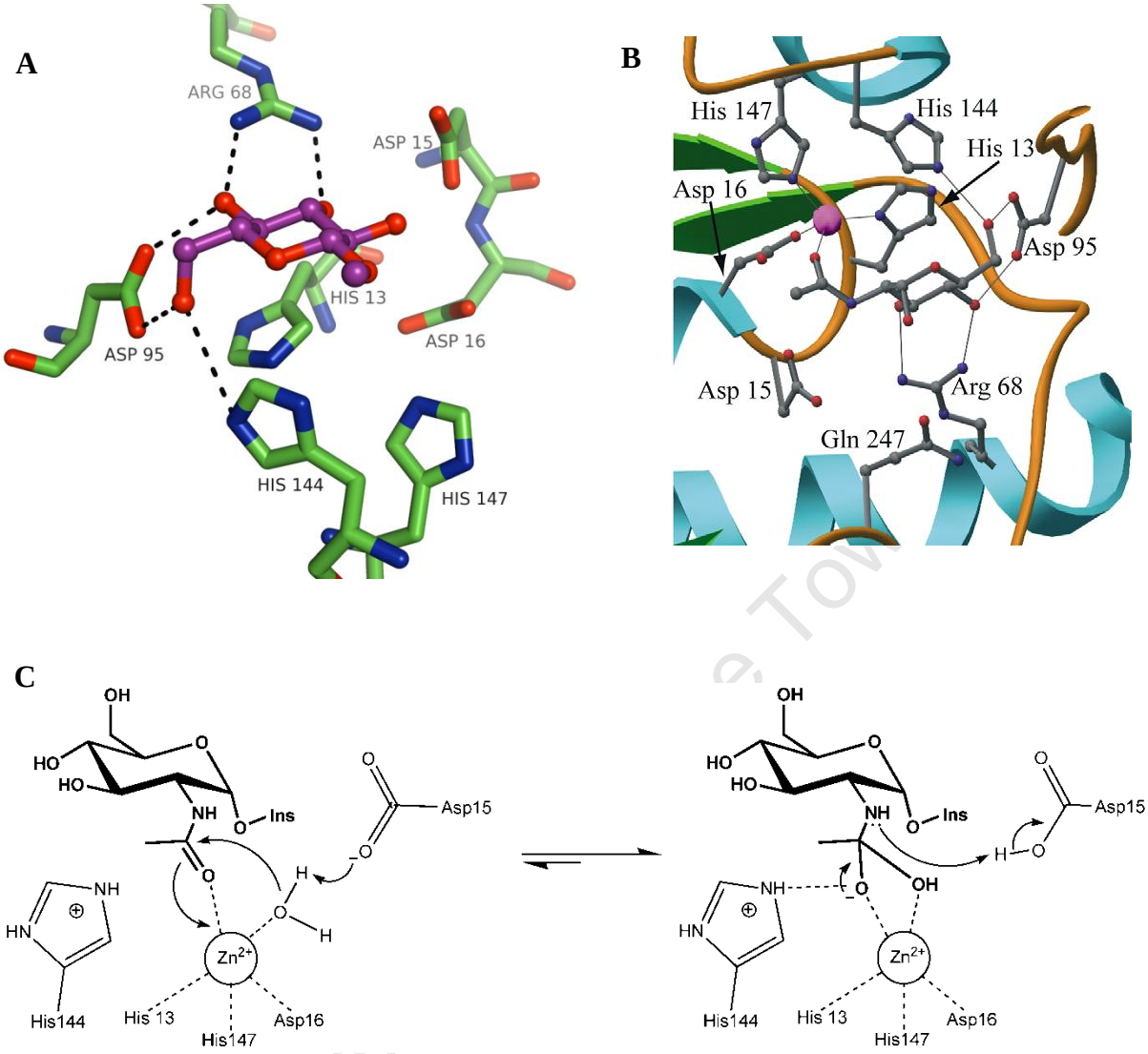


FIGURE 1-6 **(A) Binding orientation of the β-octyl glucoside co-crystallised near the zinc-binding site in 1Q74. (B) Binding of the GlcNAc fragment from the natural substrate. Binding is modelled using the BOG crystal structure and proposed catalytic mechanism. Illustration taken from Baker *et al.*[17] . (C) Proposed catalytic mechanism for deacetylase action of MshB. The mechanism was proposed based on structural similarity with the other metalloproteases. Figure adapted from Fierke *et al.* [18].**

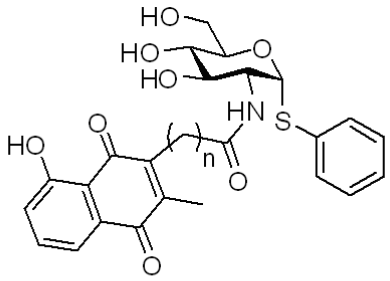
tetrahedral intermediate which is stabilised by zinc coordination and a purported interaction with the positively charged histidine residue. Finally, proton transfer from protonated Asp15 to the amine leaving-group facilitates the collapse of the tetrahedral intermediate with resultant cleavage of the C-N bond. In this mechanism, Asp15 acts as a single acid base pair. Hernick and Fierke have pointed out that the similarity to UDP-3-O-(R-3-hydroxymyristoyl)-*N*-

acetylglucosamine deacetylase makes it possible that MshB could function *via* a GABC pair with His144 transferring its proton to the amino sugar during the collapse of the tetrahedral intermediate [18]. Additional site mutagenesis and kinetics experiments are needed for clarity on the catalytic mechanism.

The binding of the natural substrate, GlcNAc-Ins, also remains uncertain. Binding of the GlcNAc fragment has been modelled based on the BOG crystal structure. As mentioned previously, this model was generated by building an acetyl group onto BOG so that it displayed coordination geometry with Zn^{2+} that was consistent with the proposed catalytic mechanism (Figure 1-6C). Subsequently, in a study of the inhibitory activity of bromotyrosine-derived compounds, GlcNAc-Ins was docked with MshB [20]. Bewley and co-workers obtained a low-energy cluster of solutions which had GlcNAc geometries that were close to the BOG model. However, no crystal structure has been reported in which an inhibitor or substrate analogue is coordinated to the MshB zinc ion. Furthermore, the above models do not describe interactions of the inositol and acetyl methyl moieties which are critical features for substrate recognition [5, 6]. A 70-fold decrease in rate of deacetylation is observed when the inositol was removed, and turnover rate of MSH, which has an AcCys at the glucose C-2 position, proceeds 330 times slower than GlcNAc-Ins [21]. The stereochemical configuration of the inositol also has a significant impact on selectivity and MshB is unable to turnover GlcNAc-L-Ins [22].

1.3. PLUMBAGIN CONJUGATES AS INHIBITORS OF MSHB

Recent investigtion into targeting the MSH redox pathway has led to the design and preparation of a set of subversive substrates of Mtr, **1a**-**1d** (Figure 1-7) [23]. A phenylthioglycoside was tethered to plumbagin in an effort to confer specificity of the naphthoquinone to redox cycling of Mtr over other NAD(P)H reductases. Although the compounds synthesised were much poorer substrates of Mtr than plumbagin itself, it was found that the series of substituted naphthoquinones potently inhibited MshB (Table 1-1). Mca was also inhibited, but at a lower



1a n = 2
1b n = 3
1c n = 4
1d n = 5

FIGURE 1-7 Series of compounds designed as subversive substrates of Mtr.

TABLE 1-1 Percentages of MshB and Mca inhibition by plumbagin conjugates.

Substrate	Mca	MshB
1a	28.8	57.4
1b	37.8	81.6
1c	23.2	81.4
1d	44.5	94.8

Percentage inhibition of the hydrolysis of the substrates MSmB and GlcNac-Ins by Mca and MshB respectively, was determined at substrate and inhibitor concentrations of 250μM each for Mca and 500μM each for mshB. Table excerpted from Gammon *et al.* [22].

level. Kinetic studies showed the inhibition by the series of naphthoquinones to be competitive and constants for **1a**, **1c** and **1d** were estimated to be to be 167 $\pm$ 15μM, 94 $\pm$ 11 μM and 16.8 $\pm$ 1.9μM, respectively. Two interesting observations arise from these results.

Firstly, the K_i values of the naphthoquinones indicate that their binding affinity increases with the length of the spacer. The inhibition constant, K_i, is the dissociation constant of the enzyme inhibitor complex, and is a measure of the strength of the enzyme-inhibitor complex: a high K_i indicates a weak binding and a low K_i indicates strong binding [19]. Preliminary docking studies were undertaken to understand how the spacer length improves the binding affinity of **1a-1d**. The conformations of **1a** and **1d** showed that the longer carbon chain of **1d** extends the hydrophobic

naphthoquinone group facilitating favourable nonpolar interactions with the dipeptide of Val 184 and Leu 185 [23].

Gammon *et al.* [23] also compared the inhibition constants with the K_M value of 1.695±0.151mM for phenyl-2-acetamido-2-deoxy-1-thio-α-D-glucopyranoside. The K_M value is numerically equal to the substrate concentration at which the rate of conversion is half of v_{max} and like the K_i a small value indicates a high substrate affinity. Such a large difference between the K_i and K_M values show that the naphthoquinone significantly improves that affinity to the active site.

It is pertinent to understand the MshB active site and explore inhibitor binding more comprehensively so that the potential drug lead of the substrate-based inhibitors can be rationally developed. Computational chemistry provides computer based methods for understanding and predicting the behaviour of such a molecular system.

1.4. OBJECTIVES

The primary objective of this thesis is to rationalise the trend observed for MshB inhibition by **1a**-**1d** (Figure 1-7). Since inhibition is competitive, relative free energies of binding were calculated to investigate the trend in terms of binding strength.

Neither substrate analogue nor inhibitor has been co-crystallised in the MshB active site. Therefore, the first objective of this thesis is to describe the binding of the natural substrate. This required prior characterisation of the active site. CASTp was used to identify the MshB binding cavity, and the chemical property of the cavity was evaluated by identifying the residues lining the surface and calculating the electrostatic surface potential (Chapter 4). Once this had been completed the natural substrate, GlcNAc-Ins, was docked and key interactions of the GlcNAc-MshB complex were elucidated using atomistic molecular dynamics (MD) simulations (Chapter 4).

The next objectives of this thesis are to identify the favourable conformations of the inhibitors in solution, and to simulate the inhibitor series bound to the MshB active site. In order to achieve this, a parameterised force field was required. The CHARMM-27-consistent force field employed here did not contain parameters for the plumbagin moiety or the dihedrals of the phenyl thioglycoside of the inhibitor series. These parameters were calculated from quantum mechanics (QM) calculations, before a Principal Component Analysis (PCA) was conducted on the dynamics of the inhibitors in solution (Chapter 5). The inhibitors were then docked with MshB and the complexes equilibrated using MD simulations.

The final objectives of this thesis are to validate the force field against experimental data and describe the inhibitor binding mode. For this purpose relative free energies of binding were computed, and the inhibitors ranked accordingly (Chapter 5). Key interactions were also identified from the MD trajectories (Chapter 5).

1.5. REFERENCES

1. http://www.who.int/mediacentre/factsheets/fs104/en/index.html, *Tuberculosis fact sheet*, World Health Organisation, accessed 8 February 2011.
2. D. G. Russell, *Nature Reviews Microbiology*, 2007, **5**, 39-47.
3. *Global Tuberculosis Control*, World Health Organisation Report, 2010.
4. *The Global MDR-TB & XDR-TB Response Plan 2007-2008*, World Health Organisation Report, 2007.
5. G. L. Newton, N. Buchmeier and R. C. Fahey, *Microbiology and Molecular Biology Reviews*, 2008, **72**, 471-494.
6. V. K. Jothivasan and C. J. Hamilton, *Natural Product Reports*, 2008, **25**, 1091-1117.
7. N. A. Buchmeier, G. L. Newton, T. Koledin and R. C. Fahey, *Molecular Microbiology*, 2003, **47**, 1723-1732.
8. N. A. Buchmeier, G. L. Newton and R. C. Fahey, *Journal of Bacteriology*, 2006, **188**, 6245-6252.

9. G. L. Newton and R. C. Fahey, *Archives of Microbiology*, 2002, **178**, 388-394.
10. N. Buchmeier and R. C. Fahey, *Fems Microbiology Letters*, 2006, **264**, 74.
11. D. Sareen, G. L. Newton, R. C. Fahey and N. A. Buchmeier, *Journal of Bacteriology*, 2003, **185**, 6736-6740.
12. M. Rawat, C. Johnson, V. Cadiz and Y. Av-Gay, *Biochemical and Biophysical Research Communications*, 2007, **363**, 71-76.
13. C. Vilcheze, Y. Av-Gay, R. Attarian, Z. Liu, M. H. Hazbon, R. Colangeli, B. Chen, W. Liu, D. Alland, J. C. Sacchettini and W. R. Jacobs, *Molecular Microbiology*, 2008, **69**, 1316-1329.
14. M. A. Kohanski, D. J. Dwyer, B. Hayete, C. A. Lawrence and J. J. Collins, *Cell*, 2007, **130**, 797-810.
15. C. D. den Hengst and M. J. Buttner, *Biochimica Et Biophysica Acta-General Subjects*, 2008, **1780**, 1201-1216.
16. J. T. Maynes, C. Garen, M. M. Cherney, G. Newton, D. Arad, Y. Av-Gay, R. C. Fahey and M. N. G. James, *Journal of Biological Chemistry*, 2003, **278**, 47166-47170.
17. A. A. McCarthy, N. A. Peterson, R. Knijff and E. N. Baker, *Journal of Molecular Biology*, 2004, **335**, 1131-1141.
18. M. Hernick and C. A. Fierke, *Archives of Biochemistry and Biophysics*, 2005, **433**, 71-84.
19. D. Voet, J. G. Voet and C. W. Pratt, *Fundamentals of Biochemistry*, John Wiley & Sons, Inc., New York, 1999.
20. B. B. Metaferia, B. J. Fetterolf, S. Shazad-Ul-Hussan, M. Moravec, J. A. Smith, S. Ray, M. T. Gutierrez-Lugo and C. A. Bewley, *Journal of Medicinal Chemistry*, 2007, **50**, 6326-6336.
21. G. L. Newton, M. Ko, P. L. Ta, Y. Av-Gay and R. C. Fahey, *Protein Expression and Purification*, 2006, **47**, 542-550.
22. G. M. Nicholas, L. L. Eckman, P. Kovac, S. Otero-Quintero and C. A. Bewley, *Bioorganic & Medicinal Chemistry*, 2003, **11**, 2641-2647.
23. D. W. Gammon, D. J. Steenkamp, V. Mavumengwana, M. J. Marakalala, T. T. Mudzunga, R. Hunter and M. Munyololo, *Bioorganic & Medicinal Chemistry*, 2010, **18**, 2501-2514.

CHAPTER 2

COMPUTER SIMULATION METHODS

The inhibition of MshB by a series of plumbagin conjugates has been described in Chapter 1. To develop this anti-tubercular lead it is important to understand the mode of inhibition. Although kinetic studies indicated that the inhibition was competitive, attempts at co-crystallisation have been unsuccessful. Therefore, the binding and mode of inhibition remain unknown.

Computational methods can be used to quantify the properties of a system at a microscopic scale and on a time scale that is otherwise inaccessible. Furthermore, if the model used is validated against experimentally determined properties, that model may be used to predict unknown or unmeasurable quantities. Simulation of a molecular system is achieved using Monte Carlo (MC) or Molecular Dynamics (MD) simulations to generate a set of configurations from which properties of the system are calculated by the application of statistical mechanics. MD simulations create a time evolution of systems by solving Newton's equations of motion, and have the advantage that time dependent properties can be calculated from the resultant trajectories. In order to solve Newton's equations of motion the forces acting on the particles of the system need to be derived from the potential energy. The energy landscape of molecular systems can be modelled using Quantum Mechanic (QM) or Molecular Mechanic (MM) methods. While solving the Schrödinger equation gives a complete description of systems at an atomic and molecular scale, such calculations are not feasible for large systems. MM ignores electronic motions and calculates the energy as a function of nuclear coordinates only. Models consist of spherical atoms connected by springs and the inter-atomic forces are described by simple mathematical functions.

In this thesis QM methods were used for parameterisation of a CHARMM27-consistent force field which was employed in the MD simulations of the inhibitor-water and inhibitor-MshB complexes. The resultant trajectories were analysed using complementary computational

approaches (discussed in Chapter 3) to quantify the stability and the free energy differences of the molecular systems. The free energy differences were compared with experimentally determined inhibition constants to ascertain the validity of the modelled binding.

2.1. ELECTRONIC STRUCTURES OF MOLECULES

2.1.1. QUANTUM MECHANICS OF PARTICLES

In quantum mechanics the wave-particle duality is reconciled, and the practical mathematics of waves are used to describe a particle distribution [1-3]. The particle is described mathematically using a wavefunction, Ψ, which is dependent on time and the position of a particle. The wavefunctions of the particle and its corresponding energy can be calculated by solving the time-dependent Schrödinger equation:

$$\left\{-\frac{\hbar^2}{2m}\nabla^2 + V\right\}\Psi(\boldsymbol{r},t) = i\hbar\frac{\partial\Psi(\boldsymbol{r},t)}{\partial t} \tag{2.1}$$

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

Equation (2.1) refers to a molecule of mass *m* which is moving through space (given by a position vector **r**) and time *t* under the influence of an external field *V*. This equation can be simplified using a number of postulates. When *V* is independent of time then the molecular wavefunction can be described in terms of a spatial wavefunction [1, 2]:

$$\Psi(\boldsymbol{r},t) = \Psi(\boldsymbol{r})\tau(t) \tag{2.2}$$

This enables solution of the time independent Schrödinger equation:

$$\hat{H}\Psi(\boldsymbol{r}) = E\Psi(\boldsymbol{r}) \tag{2.3}$$

$$\hat{H} = \left\{-\frac{\hbar^2}{2m}\nabla^2 + V\right\}$$

2.1.2. QUANTUM MECHANICS OF MOLECULES

Molecules can be described mathematically by a molecular wavefunction, which is dependent on time, and positions of the nuclei and electrons. The wavefunctions and their corresponding energies can then be calculated by solving the time-independent Schrödinger equation:

$$H\Psi(\boldsymbol{R}_1, \boldsymbol{R}_2 \ldots \boldsymbol{R}_N, \boldsymbol{r}_1, \boldsymbol{r}_2 \ldots \boldsymbol{r}_n) = E\Psi(\boldsymbol{R}_1, \boldsymbol{R}_2 \ldots \boldsymbol{R}_N, \boldsymbol{r}_1, \boldsymbol{r}_2 \ldots \boldsymbol{r}_n) \quad (2.4)$$

where $\boldsymbol{r}$ is the position vector of the n electrons and $\boldsymbol{R}$ the position vector describing the N nuclei. The Hamiltonian contains all operators that describe the kinetic and potential energy of the system. The exact Schrödinger equation can be solved for only a small number of simple systems for which boundary conditions have been applied [1-3]. However, a number of approximations can be made to make the mathematics of solving the Schrödinger equation for molecules tractable. One such approximation is the Born-Oppenheimer approximation, which postulates that since electrons are much lighter than nuclei, they can instantaneously respond to any changes in the relative positions of the nuclei. This allows separation of the total wavefunction into an electronic wavefunction and a nuclear wavefunction:

$$\Psi(\boldsymbol{R}_1, \boldsymbol{R}_2 \ldots \boldsymbol{R}_N, \boldsymbol{r}_1, \boldsymbol{r}_2 \ldots \boldsymbol{r}_n) = \Phi(\boldsymbol{R}_1, \boldsymbol{R}_2 \ldots \boldsymbol{R}_N)\psi(\boldsymbol{r}_1, \boldsymbol{r}_2 \ldots \boldsymbol{r}_n) \quad (2.5)$$

The Born-Oppenheimer approximation introduces negligible error with the exception of excited states in polyatomic molecules and ground states of cations [1, 2]. The positions of the nuclei can now be fixed and then only the electronic part of the Schrödinger equation needs be solved using the appropriate Hamiltonian. The non-relativistic Hamiltonian for the electronic part of the Schrödinger equation given in atomic units is:

$$\hat{\mathrm{H}} = \boldsymbol{h}_i + \boldsymbol{g}_{ij} \quad (2.6)$$

$$\boldsymbol{h}_i = -\frac{1}{2}\nabla_i^2 - \sum_I^N \frac{Z_I}{r_{Ii}}$$

$$\boldsymbol{g}_{ij} = \frac{1}{r_{ij}}$$

In Equation (2.6) i and j index the electrons, upper case I indexes the N nuclei, and r is a distance between the objects specified by the subscript. $\boldsymbol{h}_i$ is the single electron operator, and $\boldsymbol{g}_{ij}$ is the two

electron operator. To compensate for spin-dependent terms which are relativistic in origin it is necessary to introduce electron spin as a quantum effect [1]. Each electron has a spin quantum number of 1/2, and can occupy one of two possible states of alignment under an external magnetic field, either spin up (α) or spin down (β).

There are two primary quantum mechanical theories for the description of this electronic structure, valence-bond theory (VB theory) and molecular orbital theory (MO theory) [3]. In VB a covalent bond is formed between two atoms by overlap of their half-filled valence atomic orbitals. MO, on the other hand theorises that electrons do not belong to particular bonds but are spread throughout the entire molecule [3]. The latter approach has been more more fully developed [3] and is discussed here.

MOLECULAR ORBITAL THEORY: HARTREE-FOCK (HF) METHODS

Determinantal Wavefunction

Hartree proposed that the electronic wavefunction can be separated into a product of functions, or Molecular Orbitals (MO), that depend on only one electron.

$$\psi(\boldsymbol{r}_1, \boldsymbol{r}_2 \ldots \boldsymbol{r}_n) = \phi_1(\boldsymbol{r}_1)\phi_2(\boldsymbol{r}_2) \ldots \phi_n(\boldsymbol{r}_n) \tag{2.7}$$

For the Pauli principle to be satisfied the total electronic wavefunction must be antisymmetric [2, 4]. Thus, the sign of the wavefunction must change with the interchange of any two electron coordinates. Antisymmetric wavefunctions are built using the Slater determinant, which gives the sum of all possible permutations of electron pairs:

$$\psi(r_1, r_2 \ldots r_n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_{1(e_1)} & \phi_{2(e_1)} & \cdots & \phi_{n(e_1)} \\ \phi_{1(e_2)} & \phi_{2(e_2)} & \cdots & \phi_{n(e_2)} \\ \phi_{1(e_n)} & \phi_{2(e_n)} & \cdots & \phi_{n(e_n)} \end{vmatrix}, \tag{2.8}$$

where $\frac{1}{\sqrt{n!}}$ is the normalisation factor, e_i denotes electrons and the MOs, ϕ_i, are spinorbitals which are the product of a spatial orbital and a spin function (α, β).

The energy of the molecule is written as the expectation value of the one and two electron operators from Equation (2.6) in the MO basis:

$$E = \int \psi(\boldsymbol{r}_1, \boldsymbol{r}_2 \ldots \boldsymbol{r}_n)^* \hat{\mathrm{H}} \psi(\boldsymbol{r}_1, \boldsymbol{r}_2 \ldots \boldsymbol{r}_n) d\tau$$

$$= \langle \psi | \hat{\mathrm{H}} | \psi \rangle$$

$$= \sum_i^n \langle \phi_i | \boldsymbol{h}_i | \phi_i \rangle + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left(\langle \phi_j | \boldsymbol{J}_i | \phi_j \rangle - \langle \phi_j | \boldsymbol{K}_i | \phi_j \rangle \right) \quad (2.9)$$

$$\boldsymbol{J}_i | \phi_j(2) \rangle = \langle \phi_i(1) | \boldsymbol{g}_{12} | \phi_i(1) \rangle | \phi_j(2) \rangle$$

$$\boldsymbol{K}_i | \phi_j(2) \rangle = \langle \phi_i(1) | \boldsymbol{g}_{12} | \phi_j(1) \rangle | \phi_i(2) \rangle$$

The total energy can be written as:

$$E = \sum_i^n h_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left(J_{ij} - K_{ij} \right) \quad (2.10)$$

In Equation (2.10) the one electron integral h_i describes the kinetic and potential energy of the electron moving in the field of the bare nuclei. The two-electron integral J_{ij} is the Coulomb integral representing the classical repulsion between two electrons, and K_{ij} is the exchange integral.

Many of the systems considered in molecular modelling are in their ground states [5]. These molecules are said to have closed shell configurations where the valence shell is completely filled. Thus, MOs are doubly occupied, with one electron spin up and the other spin down. In such a system each spatial orbital is restricted to two electrons and the total energy can be written:

$$E = 2 \sum_i^{n/2} h_i + \frac{1}{2} \sum_{i=1}^{n/2} \sum_{j=1}^{n/2} \left(2J_{ij} - K_{ij} \right) \quad (2.11)$$

Finding the Optimum Wavefunction: Hartree-Fock Equations

The optimum determinantal wavefunction, ψ, can be found using the variational principal. The theorem states that if an arbitrary wavefunction is used to calculate the energy, the value calculated is never less than the true energy [2, 6]. Thus, the best 'trial' wavefunction, ϕ_i, can be generated by minimising the molecular energy as a function of the MOs. Variation of the MOs must be done under the constraint that the MOs remain orthonormal [1]. Carrying out the constrained optimisation using the method of Lagrange multipliers, it can be shown that the optimum MOs must satisfy the Hartree-Fock (HF) equations:

$$\boldsymbol{F_i}\phi_i = \varepsilon_i\phi_i \tag{2.12}$$

$$\boldsymbol{F_i}\phi_i = \boldsymbol{h}_i + \sum_j^{n/2}\left(2\boldsymbol{J}_j - \boldsymbol{K}_j\right), \tag{2.13}$$

where ε_i is the MO energy and **F** is the Fock operator. The MOs are termed canonical MOs and they are obtained by choosing a unitary transformation which makes the matrix of ε diagonal (i.e. $\varepsilon_{ij}=0$ and $\varepsilon_{ii}=\varepsilon_i$) [2]. Since the Fock operator depends on the occupied MOs of all the electrons, iterative methods must be used for determining the Fock orbitals. This procedure is known as a self-consistent field (SCF) calculation.

There are two important points regarding the results of HF SCF calculations. Firstly the canonical MOs are not unique since a unitary transformation applied to the occupied MOs does not change the total wavefunction [2]. One set of optimised MOs can be used to generate other sets, such as localised MOs, through linear combinations. The second note is that the repulsion between an electron and all other electrons is calculated using a single Slater determinant [2, 5, 7]. Thus, the repulsion is not exact and the HF method is also referred to as a Mean Field approximation. Nevertheless HF methods often give geometries and relative energies for equilibrium structures that are in good agreement with experiments [2].

The Basis Set Approximation

In solving the HF equations it is practical to express the unknown MOs in terms of a set of known functions, or Atomic Orbitals (AO), χ_α.

$$\phi_i = \sum_{\alpha}^{M} c_{\alpha i}\chi_\alpha \tag{2.14}$$

where α annotates the M spinorbitals. As the number of basis functions increase, the accuracy of the MOs improves until the HF limit is reached [2, 5]. Exponential functions $\chi \propto e^{(-\alpha r)}$ located on the nucleus are well suited for electronic structure calculations, and are known to be exact solutions for the hydrogen atom [2]. However, they are computationally expensive. Gaussian functions $\chi \propto e^{(-\alpha r^2)}$ are more efficient for calculating two-electron integrals because a product of two Gaussians located at two different postions can be written as a single Gaussian [2].

Roothaan-Hall Equations

Expansion of the Hartee-Fock equations in the atomic orbital basis yields:

$$\boldsymbol{F_i}\sum_{\alpha}^{M} c_{\alpha i}\chi_\alpha = \varepsilon_i \sum_{\alpha}^{M} c_{\alpha i}\chi_\alpha \tag{2.15}$$

where $\boldsymbol{F}_\mathrm{i}$ is the Fock operator. When Equation (2.15) is multiplied on both sides by a specific basis function and integrated the Roothaan-Hall equations are obtained:

$$\sum_{\alpha}^{M} c_{\alpha i}\langle\chi_\beta|\boldsymbol{F_i}|\chi_\alpha\rangle = \varepsilon_i \sum_{\alpha}^{M} c_{\alpha i}\,\langle\chi_\beta|\chi_\alpha\rangle \tag{2.16}$$

$$\sum_{\alpha}^{M} c_{\alpha i}F_{\alpha\beta} = \varepsilon_i \sum_{\alpha}^{M} c_{\alpha i}\,S_{ij} \tag{2.17}$$

Since there are M basis functions, there will be M such equations which can be collected in matrix notation.

$$\boldsymbol{FC} = \boldsymbol{SC\varepsilon} \tag{2.18}$$

F is the Fock matrix for which each element contains the one electron integrals, and a sum of the two-electron integrals over the occupied MOs:

$$\boldsymbol{F}_{\alpha\beta} = \langle \chi_\alpha | \boldsymbol{h} | \chi_\beta \rangle + \sum_{j}^{occ.\ MO} \langle \chi_\alpha | \boldsymbol{J}_j - \boldsymbol{K}_j | \chi_\beta \rangle$$

$$= \langle \chi_\alpha | \boldsymbol{h} | \chi_\beta \rangle + \sum_{\gamma}^{AO} \sum_{\varepsilon}^{AO} D_{\alpha\beta} \left(\langle \chi_\alpha \chi_\gamma | \boldsymbol{g} | \chi_\beta \chi_\delta \rangle - \langle \chi_\alpha \chi_\gamma | \boldsymbol{g} | \chi_\delta \chi_\beta \rangle \right) \quad (2.19)$$

$$D_{\alpha\beta} = \sum_{j}^{occ.\ MO} c_{\gamma j}\, c_{\delta j}$$

S is the matrix of overlap integrals. **C** is a *M* x *M* matrix of the $c_{\alpha i}$ coeffients. Each column of **C** is an expansion of ϕ_i in the terms of the atomic orbitals χ_α. **ε** is the *M* x *M* matrix of orbital energies.

The matrix notation is advantageous because a simple algorithm can be used for solving Equation (2.18) [1, 2, 5]. An initial guess is made for the coefficients and used to generate the Fock matrix which is then diagonalised to generate a new set of coefficients. These coefficients are in turn used to generate a new Fock matrix and the process repeated until the coefficients used to generate the Fock matrix are equal to the coefficients generated from the diagonalisation.

RESTRICTED AND UNRESTRICTED HARTREE-FOCK

The closed shell system described above represents a special case of Restricted Hartree-Fock (RHF), which uses combinations of singly and doubly occupied molecular orbitals. RHF makes the restriction that for doubly occupied molecular orbitals each spatial orbital should have two electrons, one spin up and the other spin down [2, 5]. Alternatively, if different spatial orbital are allowed for the two electrons in an orbital the trial function is known as an Unrestricted Hartree-Fock (UHF) wavefunction. The UHF treatment is an extension of the Roothaan-Hall equations and two sets of trial wavefunctions are optimised using the Pople-Nesbet calculations [8], one for spin up electrons and one for spin down electrons.

SEMI-EMPIRICAL METHODS

Full solution of the HF equations where all the necessary integrals are calculated from a given basis set are known as *ab initio* HF methods. Computational cost can be minimised by reducing the number of two-electron integrals. Most semi-empirical methods achieve this by employing

two approximations [2]. In the first approximation only the valence electrons are treated explicitly using a minimum number of basis functions required for the electrons of a neutral atom. The second approximation is known as the Zero Differential Overlap and neglects the overlap between pairs of different orbitals which depend on the same electron coordinates. Some or all of the neglected integrals are modelled using parameters estimated from spectroscopic or physical data [2, 3]. Semi-empirical methods include Neglect of Diatomic Differential Overlap Approximation (NDDO), Intermediate Neglect of Differential Overlap Approximation (INDO) and Complete Neglect of Differential Overlap Approximation (CNDO). These methods are used when the systems are large and too expensive to model using *ab initio* methods. In this dissertation QM methods were used to obtain a reference data set for parameterisation of a MM force field (described later in Chapter 5). This was done for model compounds that were sufficiently small for treatment with post Hartree-Fock *ab initio* methods.

ELECTRON-CORRELATION: POST HARTREE-FOCK METHODS

HF is a Mean Field theory that assumes that electrons move in an average potential of the other electrons. The instantaneous position of an electron is not influenced by the presence of a neighbouring electron and so electron correlation is not accounted for [5]. The correlation energy is defined as the difference between the HF energy and the exact energy. Post HF methods improve on the SCF method described above by adding electron correlation. These methods include configuration interaction (CI), coupled cluster (CC) and Møller–Plesset perturbation (MP2, MP3 etc.) theory methods.

DENSITY FUNCTIONAL THEORY

Density functional theory (DFT) utilises the Hohenberg-Kohn theorem [9] which states that the potential of a ground state system is a functional of its electron density:

$$E_{elec} = E[\rho(\boldsymbol{r})] \tag{2.20}$$

Kohn and Sham [10] proposed that the functional take the form

$$E[\rho(\boldsymbol{r})] = E_{KE}(\rho) + E_C(\rho) + E_H(\rho) + E_{XC}(\rho) \tag{2.21}$$

$E_{KE}(\rho)$ is the kinetic energy, $E_C(\rho)$ is the electron-nuclear interaction term, $E_H(\rho)$ is the electron-electron Coulombic energy, and $E_{XC}(\rho)$ contains the exchange and correlation terms. These terms are functions of the electron density. The approximate density is constructed from a set of one-electron orbitals:

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{n} |\psi_i(\boldsymbol{r})|^2 \qquad (2.22)$$

Since the electron density obeys the variational theorem the orbitals constructing the electron density can be obtained by optimising Equation (2.21) with respect to the density in a self consistent manner. All the ground state properties can then be calculated from the electron density.

One advantage of DFT is that the energy of the molecule includes the electron correlation, although there is no way of deriving this term and its dependence on ρ(r) has to be approximated [5]. The primary advantage though, is that unlike the electronic wavefunction which is dependent on 3*n* variables (the x, y and z coordinates of each of the *n* electron), the total electron density is dependent on three spatial coordinates [11]. This is because the electron density is dependent on the integral of the orbitals squared, which is independent of the number of electrons [2]. As a result DFT calculations are much more computationally efficient than the wavefunction approach.

APPLICATIONS OF MOLECULAR ORBITAL THEORY

Once the molecular wavefunction or electron density has been solved the electronic charge distribution can be explored, which has important implications on reactivity [5]. Furthermore, using the Born-Oppenheimer approximation the potential energy surfaces of molecules can be created by determining the electrostatic energy of a molecule while varying the positions of the nuclei. A time evolution of the system can be generated by perturbing the nuclei at finite time intervals according to the laws of Newtonian dynamics, where the force acting on the nuclei is derived from the quantum mechanical potential. However, when systems are large it is necessary to define the potential energy of the system using a MM force field.

2.2. MOLECULAR MECHANICS

Molecular mechanics (MM) is also based on the Born-Oppenheimer approximation and models energy as a function of nuclear coordinates [5]. MM methods model the energy landscape using a force field which consists of a functional form describing the molecular potential and parameters used by that potential function. The potential function is based on simple models of intra- and intermolecular interactions. All force fields contain terms for bond stretching, angle bending, torsional rotation and electrostatic interactions. More sophisticated force fields will contain additional terms [2]. The parameters used by the potential function are normally specific to atom types, which identify the hybridisation and local environment of the atom.

Since the electronic effects are ignored in MM this treatment of molecules is more efficient than QM. A result of this simplification is that the primary function of force fields is often to predict structural properties, although they can be specifically developed to predict desired properties [5].

2.2.1. FORCE FIELDS

THE POTENTIAL FUNCTION

The current research used the CHARMM potential function [12, 13]. This function is the sum over the individual inter- and intra-molecular interactions and takes the form:

$$U(\boldsymbol{r}) = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\varphi\big(1 + cos(n\varphi - \delta)\big)$$
$$+ \sum_{impropers} K_\omega(\omega - \omega_0)^2 + \sum_{residues} U_{CMAP}(\varphi, \psi) + \sum_{Urey-Bradley} K_{UB}(S - S_0)^2$$
$$+ \sum_{non-bonded\ \ pairs} \left\{ \varepsilon_{ij}^{min} \left[\left(\frac{R_{ij}^{min}}{\boldsymbol{r}_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{min}}{\boldsymbol{r}_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon \boldsymbol{r}_{ij}} \right\} \qquad (2.23)$$

INTRAMOLECULAR INTERACTIONS

The first three terms in Equation (2.23) model the standard intermolecular interactions: bond stretching, angle bending and torsional rotation. These terms describe the interaction energy as a function of the distance from an equilibrium value which is given in the parameter set. The equilibrium value is annotated with a zero subscript and the respective force constants, also defined in the parameter set, are K_b, K_θ, K_φ. The functional forms are illustrated in Figure 2-1. The bonding and angle terms are modelled by Hooke's law, while the torsional rotation is modelled using cosine series expansion.

The CHARMM potential function contains three additional terms. An out-of-plane bending term is used to maintain planar geometries and stereochemistry of chiral centres. The Urey-Bradley function is a cross-term to account for coupling between angles and the length of adjacent bonds. This coupling models the observation that as an angle decreases the adjacent bonds stretch to reduce the interactions between the 1,3 atoms [2]. Finally CMAP is the backbone torsional correction term [13].

INTERMOLECULAR INTERACTIONS

The last component of a system's energy function is the pairwise potential for non-bonded interactions. This term is the addition of coulomb electrostatic and van der Waals (vdW) interactions, the functional forms of which are summarised in Figure 2-2. Although, these interactions can be affected by the presence of a third, fourth or more molecules, the many-body effects can be parameterised into the pairwise model [5]. In this case the model uses 'effective' pairwise potentials which are not the true interaction energy between two isolated particles.

Electrostatic Interactions

The observed electrostatic interactions between molecules arise from the unequal distribution of charge due to differences in electronegativity of the constituent atoms [2]. In the CHARMM potential the electrostatic interaction is modelled by arranging partial atomic charges on the nuclei. Multipole energies are then the sum of the interaction of point charges, described by Coulomb's law.

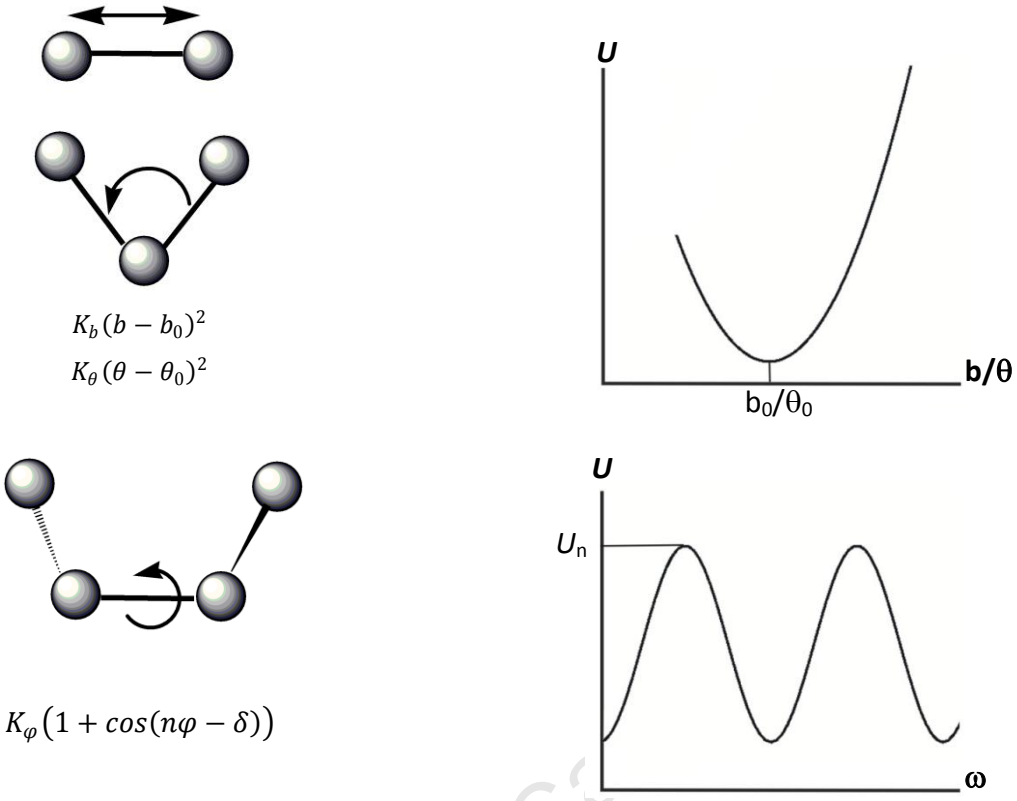


FIGURE 2-1 Harmonic and cosine potential functions describing intermolecular terms present in all force fields. U_n is known as the barrier height for torsional rotation.

Van Der Waals Interactions

Electrostatic interactions do not account for the total non-bonded interaction and in classical physics the vdW forces are those which account for the deviation of a real gas from ideal gas behaviour. VdW interactions result from long-range attractive interactions and short-range exchange-repulsive forces between atom pairs [2]. The attractive interactions result from dispersion forces between instantaneous dipoles arising from fluctuations in the electron cloud. The repulsive contribution arise from short-range nuclear repulsion of partially shield nuclei. The nuclei become partly de-shielded at short-range since exchange forbids electrons to occupy the same region of inter-nuclear space. In the CHARMM potential function vdW interactions are modelled using a Lennard-Jones potential.

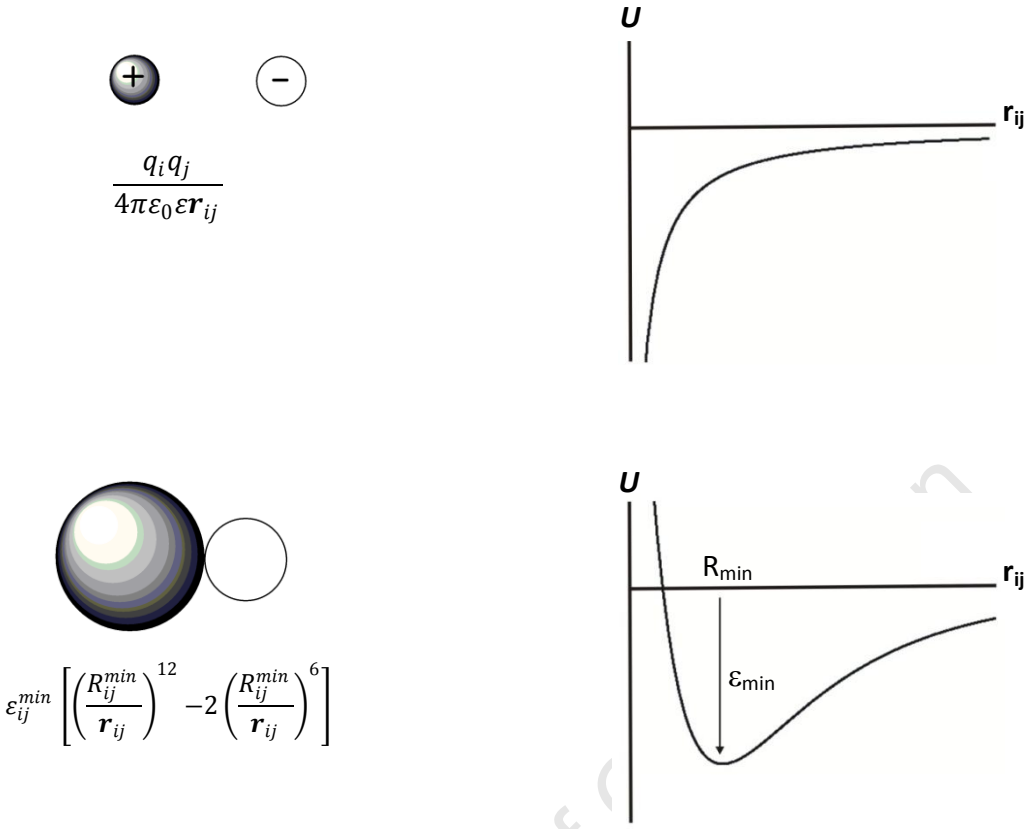


FIGURE 2-2 Coulomb and Lennard-Jones potential functions used to describe non-bonded terms.

Potential Truncation

The non-bonded interactions are calculated for all atom pairs in the molecular system so that the summation is proportional to the square of the number of atoms in the system. Since the other parts of the calculation are only proportional to the number of atoms in the system, calculating non-bonded energies is the most time consuming part of a molecular dynamics simulation [2, 5]. This expense can be minimised by truncating the pair potential for the non-bonded terms. Thus, non-bonded interactions are only calculated from neighbour atoms lying within a specified cutoff distance. The assumption is that the forces are negligible at the cutoff boundary because the vdW and electrostatic potentials tend to zero as the distance between the atom pairs tends to infinity. This is appropriate for the short-ranged vdW interactions since the Lennard-Jones potential has a r^{-6} distance dependence. For a system with bare charges or with correlated dipoles, such as an α-helix, the Coulomb term is proportional to r^{-1} [14]. Discontinuity of the electrostatic interactions causes fluctuations in the molecular potential near the cutoffs [2, 5, 14]. One way to reduce this

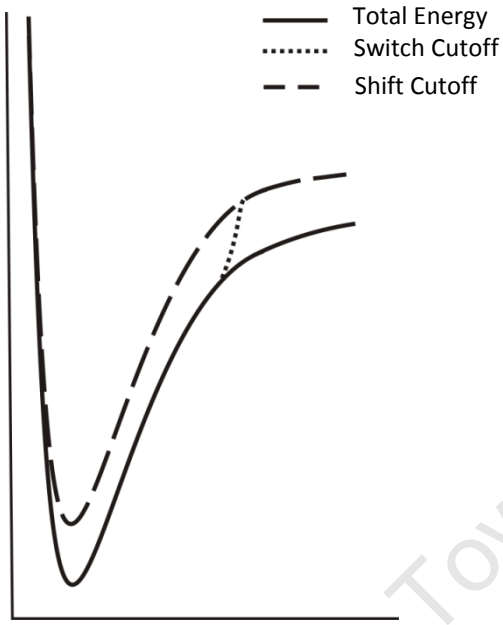


FIGURE 2-3 Illustration of the methods used for smoothing the non-bonded interaction potential to zero at a specified non-bonded cutoff distance.

noise is to use group-based cutoffs. In these methods the distance between neutral groups of atoms are calculated and pairwise interactions are only calculated between the atoms of two groups if the groups fall within a cutoff distance. Since the electrostatic interaction between two neutral groups is proportional to r^{-3}, the range of the interaction is reduced considerably [2, 5, 14]. Another way to avoid discontinuities in the potential near the cutoff value is to apply a switching or shifting function to the non-bonded potential [15]. In the former method the entire non-bonded potential energy surface is shifted so that the interaction potential is zero at the cutoff distance (Figure 2-3). On the other hand, switching leaves the potential unaltered until a first cutoff value. Between this first cutoff value and the last cutoff value the interaction potential is tapered to zero. Beyond the practical importance of cutoff radii, truncating the pairwise potential also eliminates long range order which is undesirable in a simulation of bulk solvent.

Non-Bonded Neighbour Lists

Computing all atom pair distances to evaluate whether they fall within the cutoff is still time consuming [5]. On the assumption that an atom's neighbours do not change significantly over 10 or 20 MD steps, a non-bonded neighbour list can be used to compute the interactions between

previously identified neighbours [5]. This list stores atoms within the cutoff distance, together with all atoms that are slightly further away. This ensures that only the distances between the central atom and the atoms in its neighbour list have to be computed. The frequency of updating the list should be slow enough to be computationally efficient but fast enough to ensure accuracy.

The Ewald Summation

When using periodic boundary conditions (Section 2.2.4), the electrostatic interaction between a charge and all its periodic images can be accurately computed using the Ewald summation [16, 17]. For a box defined as (n_xL, n_yL, n_zL), where n is its position at a cubic lattice point and L is the length of the cube, the charge-charge contribution to the potential energy due to all pairs of charges can be calculated:

$$U_{coulomb} = \frac{1}{2}\sum_{|n|}^{\infty}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\varepsilon_0|\boldsymbol{r}_{ij}+n|} \quad (2.24)$$

where $\boldsymbol{r}_{ij}$ is the minimum distance between charges i and j. Equation (2.24) converges slowly and is a conditionally convergent series (i.e. depends on the order in which its terms are considered). The summation was reformulated by Ewald [16] as the sum of three terms which each converge more rapidly than Equation (2.24):

$$U_{coulomb} = U_{real} + U_{fourier} + U_{self} \quad (2.25)$$

In the real space summation, U_{real}, each charge is considered to be surrounded by a neutralising charge distribution of equal magnitude but of opposite sign. The interactions of a set of distributions added to cancel the initial neutralising Gaussians are calculated in reciprocal space, $U_{fourier}$. Finally a self-term, U_{self}, cancels the interaction of each of the introduced artificial counter-charges with itself.

PARAMETERISATION

With a potential energy function in hand, an appropriate parameter set is required. Parameters are chosen to best reproduce a data set which describes a system's properties of interest. Since force fields are most commonly used to model structurally related properties, the data set often includes

geometries and relative conformational energies (vibrational frequencies are also common) [2, 5]. Ideally the data set is obtained from experimental values, but when this information is not available, the results of QM calculations are used. For large systems, such as that considered in this work, it is not practical to calculate the reference data for the whole system. Therefore, the system is divided into suitable fragments. Parameters are then optimised using model compounds representative of these fragments.

Parameters for the model compounds are coupled and parameterisation demands an iterative procedure. Intramolecular interactions, vibrational profile and intermolecular interactions are optimised in turn and the cycle repeated until satisfactory agreement with the reference data set is obtained for each stage of the cycle [5]. Within each stage there are two methods for obtaining the parameters for the required terms in the potential function. The first method is by trial and error, in which parameters are gradually refined to get a better fit to data. The second method expresses an error as the sum of squares of differences between observed and calculated values for the set of properties, and the error is then minimised in an iterative procedure. When modelling a set of compounds not included in the existing force field, new parameters must be added. The initial parameters can be estimated from analogous compounds already present and then refined. An important consideration is the partial charge assignment which characterises the electrostatic interactions.

CHARGE ASSIGNMENT

For small molecules, partial charges can be calculated to exactly reproduce experimental electric moments if the geometry is known. Some force fields assign partial charges by fitting solvent-solute interactions to QM calculated energies. A third option is to fit the partial charges to the electrostatic potential of the model compound [18]. The electrostatic potential is defined as the force acting on a unit positive charge placed at a point, and can be determined from the molecular wavefunction [5]. A grid of points outside the vdW radius of atoms constituting the molecule can be generated and the potential at each point calculated. Partial charges can then be fitted to reproduce the electrostatic potential at the grid points. In the CHELPG [19] method utilised in the current work this is achieved using the method of Lagrange multipliers with the condition that the sum of the point charges is equal to the molecular charge.

2.2.2. OPTIMISATION

The energy landscape described by the force field can be explored by optimising the energy as a function of the nuclear coordinates. The resultant stationary points are where the first derivative is zero. Important stationary points are the energy minima where the second derivatives are positive. There are three classes of minimisation algorithms which can be used to find the local minimum.

Steepest Descent Methods

Steepest descent (SD) methods calculate the gradient vector, which points in the direction of greatest energy increase, and the energy is lowered by stepping in the opposite direction. If the energy increases along this line search, an approximate minimum is determined and a new gradient is calculated. Subsequent line searches must be perpendicular to each other which results in the path oscillating around the minimum path. Furthermore as the minimum is approached the rate of convergence slows [2, 5]. This method is often used to quickly relax the starting coordinates or if more advanced algorithms fail.

Conjugate Gradient Methods

Conjugate gradient (CG) methods are similar to the SD methods except they perform each search along a line which is constructed so that it is conjugate to the previous search direction. During minimisation this results in a path which oscillates around the minimum path to a lesser extent than SD methods [2, 5].

Newton-Raphson Methods

Newton-Raphson (NR) methods use first and second derivatives to find the minimum of a potential function. The advantage of the NR methods is that the convergence is second-order near a stationary point. However, minimisation can become unstable far from a minimum. Furthermore, if one of the second derivative values of the Hessian matrix is not positive definite then NR methods will move in a direction of increasing energy [2, 5].

2.2.3. MOLECULAR DYNAMICS

CLASSICAL NEWTONIAN DYNAMICS

Molecular dynamics (MD) simulates 'real' dynamics of a system. The simulation generates successive configurations of the system by integrating Newton's laws of motion. The result is a trajectory, or a collection of coordinates and dynamical properties (velocities, accelerations, etc) in phase space that are connected in time. Thus, MD allows one to follow the structural and energetic properties of a system over time and the analysis of representative structures throughout the trajectory can yield a variety of information. Time-dependent properties can be calculated as correlation coefficients, and thermodynamic relationships are calculated from ensemble averages. Generating the trajectory requires solving the differential equations embodied in Newton's second law:

$$\frac{d^2x_i(t)}{dt^2} = \frac{\boldsymbol{F}_i\{x(t)\}}{m_i} \tag{2.26}$$

Equation (2.26) describes the motion of a particle of mass m_i in one dimension x_i experiencing a force F_{xi}. The forces acting on each particle are calculated from the differential of the system's force field potential. Solving this second order differential equation for every particle in the system, we can calculate the next configuration at any future time from an initial configuration. This deterministic method is continued to obtain the dynamic behaviour of the system.

INTEGRATING THE EQUATION OF MOTION

When solving Newton's second law of motion for a system, the force acting on each particle will change as the particles move. This many-body problem cannot be solved analytically and the equations of motion embodied in Equation (2.26) are integrated using a finite difference method [5]. The integration is broken down into many small time intervals *dt*. The forces acting on each particle is calculated at a time *t* as a function of the coordinates. The accelerations of the particles are then determined from the forces, and used to calculate the positions and velocities at a time *t* + *dt* [5]. The force, and thus, acceleration is assumed to be constant during the time step.

There are many algorithms for integrating equations of motion using finite difference methods. These algorithms assume that the positions and dynamic variables can be approximated as Taylor series expansion [2, 5]:

$$r(t+dt) = r(t) + dtv(t) + \frac{1}{2}dt^2a(t) + \frac{1}{6}dt^3b(t) + \frac{1}{24}dt^4c(t)\cdots \quad (2.27)$$

$$v(t+dt) = r(t) + dta(t) + \frac{1}{2}dt^2b(t) + \frac{1}{6}dt^3c(t)\cdots \quad (2.28)$$

$$a(t+dt) = r(t) + dtb(t) + \frac{1}{2}dt^2c(t)\cdots \quad (2.29)$$

where v is the velocity and a is the acceleration. CHARMM uses the Verlet [20], Leap-frog [21] and Velocity Verlet [22] algorithms to numerically solve Netwton's equations of motions for all atoms of the system.

2.2.4. SIMULATION CONDITIONS

SOLVATION

CHARMM all-atom force fields, such as that used in this dissertation, were designed for simulation with explicit waters [12, 13]. Water models can be broken into three types: simple interaction site water models, flexible models and polarisable water models [5]. In simple interaction site models each water molecule is defined by three to five interaction sites distributed over a rigid geometry. The interaction between molecules is described using pairwise Coulombic and Lennard-Jones expressions. Developed models include TIP3P [23], SPC/E [24], TIP4P [23] and TIP5P [25]. TIP3P has been used in this thesis since CHARMM force fields have been parameterised with this model [13]. The TIP3P model has three charges, one centred on the oxygen nucleus and two centred on the hydrogen nuclei. A single vdW interaction site is specified on the oxygen.

BOUNDARY CONDITIONS

The focus of this thesis is the dynamics of inhibitor enzyme complexes, and more specifically the active site of the enzyme. Therefore, it is not necessary to explicitly simulate the behaviour of solvent molecules far away from the enzyme. Boundary conditions are used to model a small number of particles far away from the boundary as if they were in bulk fluid.

PERIODIC BOUNDARY CONDITIONS (PBC)

In periodic boundary conditions the particles are placed inside a unit cell which is then replicated in all directions. The shape of the cell must be such that the central cell fills all of space by translation operations in three dimensions. Shapes that fulfil this criterion include the cube, the hexagonal prism, the truncated octahedron and the rhombic dodecahedron. The cubic cell is the simplest, and is illustrated using a 2D example in Figure 2-4. As can be seen, the number of particles within the central box is held constant by replacing any particle leaving the box with an image particle that enters from the opposite side. Although the cubic cell is the simplest, cells which have a more spherical geometry, such as the truncated octahedron and rhombic dodecahedron, will require fewer particles and are more appropriate for the simulations of spherical molecules. Thus, the cube and truncated octahedron have been used most widely [5]. Calculation of the potential is performed using the minimum image convention, in which the interactions of an atom are restricted to only the nearest image of the other atoms. The size of the unit cell is an important consideration for the simulation of non-bonded interactions. The box should be large enough so that the solute does not interact with itself and that no water molecules interact with the solute twice. In order to achieve this, the non-bonded cutoff (discussed in section 2.2.1) should be no greater than half the length of the shortest side [5].

STOCHASTIC BOUNDARY CONDITIONS

When dealing with macromolecules, solvation in a sufficiently large unit cell for PBC requires a large number of water molecules to be included in the simulation. In stochastic boundary methods the solute of interest is solvated in a thin film of explicit water molecules, and the effect of the bulk solvent is included as average interactions. Such methods facilitate the efficient simulation of enzyme active sites. A sphere of explicit water molecules with radius $\mathbf{R_2}$ is centred on the active site (Figure 2-5, explicit water molecules are not shown). The enzyme and water

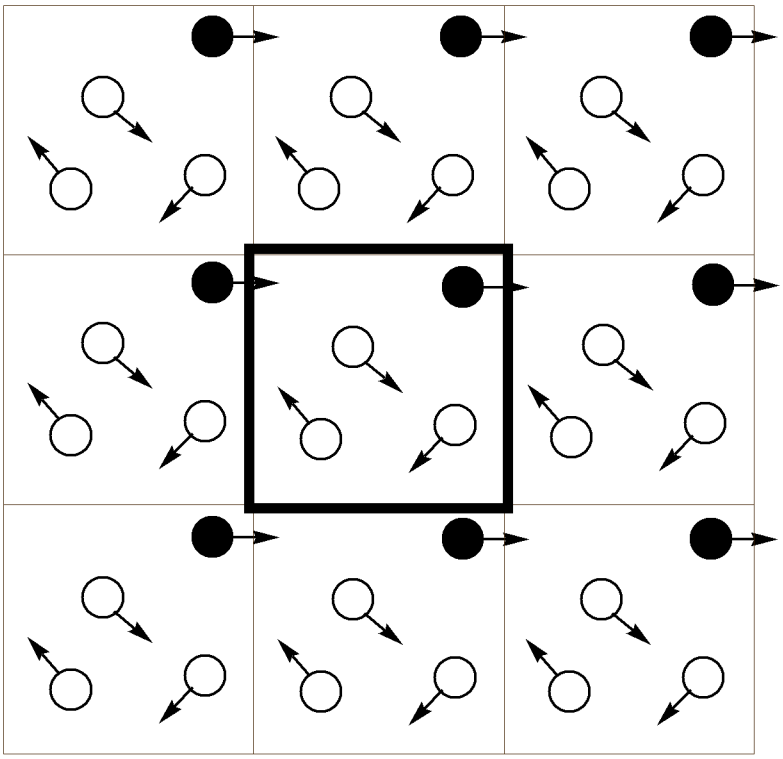


FIGURE 2-4 Illustration of periodic boundary conditions. The central cell has a bold border and is translated to fill the 2D space with image cells. Any atom which leaves the cell (shown by the black atom) is replaced by its image from the opposite side. In 3D the central box would have 26 nearest neighbours Illustration adapted from Leach [5].

molecules are then divided into three regions. All residues with an atom within $\mathbf{R_1}$, the reaction region, are subject to Newtonian dynamics, while all residues with an atom outside $\mathbf{R_2}$ are eliminated by keeping their atoms fixed. Atoms of residues that fall between the reaction and reservoir regions belong to the buffer region. The atoms here evolve in time according to an adapted form of the Langevin equation:

$$m_i \frac{d^2 x_i(t)}{dt^2} = \boldsymbol{F}_i\{x_i(t)\} - m_i \Omega_i^2 \left[x_i(t) - x_i^{ref}\right] - \gamma_i \frac{dx_i(t)}{dt} m_i + \boldsymbol{R}_i(t) \qquad (2.30)$$

In Equation (2.30), $\boldsymbol{F}_{\mathrm{i}}$ is the normal force experienced by the atoms in Newtonain dynamics, and $\Omega_i^2\left[x_i(t) - x_i^{ref}\right]$ are boundary forces applied to solute atoms derived from atomic mean-square fluctuations. These boundary forces are gradually scaled over the buffer region so that they are weakest at the boundary of the free moving reaction region (< $\mathbf{R_1}$) and strongest near the

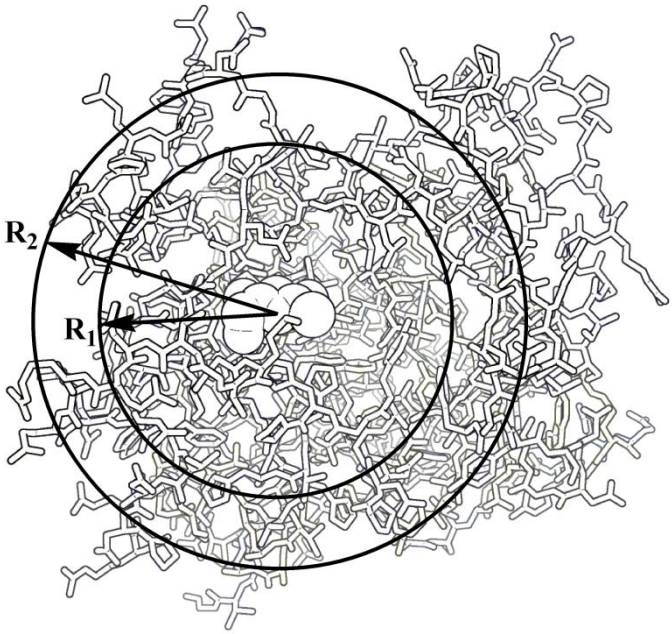


FIGURE 2-5 Illustration showing how stochastic boundary conditions are used to study the dynamics of enzyme-ligand complexes. This illustration does not display the explicit waters comprising the sphere of radius R_2.

boundary of the static reservoir region (> $\mathbf{R_2}$). Explicit water molecules are allowed to diffuse between the reaction and reservoir regions, and are acted upon by a deformable boundary potential which maintains the correct average distribution of water molecules and prevents water from escaping into the vacuum. The remaining forces in Equation (2.30) originate from Langevin dynamics and simulate the effect of bulk water. $\boldsymbol{R}_i$ are the forces of random collisions associated with energy and temperature, while $\gamma_i \frac{dx_i(t)}{dt} m_i$ simulates the frictional drag and removes energy from the system[2, 5].

ENSEMBLES

Statistical mechanics uses ensembles to calculate macroscopic properties of a system, where an ensemble is a collection of microscopic replications belonging to the particular thermodynamic state [26]. The observed value of the property is replaced by an ensemble average:

$$A_{ave} = \langle A \rangle = \iint d\boldsymbol{p}^N d\boldsymbol{r}^N \, A(\boldsymbol{p}^N, \boldsymbol{r}^N) \rho(\boldsymbol{p}^N, \boldsymbol{r}^N) \tag{2.32}$$

where the brackets indicate the ensemble average and ρ is the probability density of the ensemble. The ensemble generated by MD methods, discussed above, consists of configurations connected in time. From this trajectory the time average of a property can be calculated, which in accordance with the ergodic hypothesis is equal to the ensemble average:

$$\langle A \rangle = \langle A \rangle_{time} = \frac{1}{M} \sum_{t=1}^{M} A(\boldsymbol{p}^N, \boldsymbol{r}^N) \tag{2.33}$$

This equality holds only if the MD simulation sufficiently samples the phase space (i.e. generates enough representative conformations of the system). Due to the formulation of molecular force field, MD most often samples the low energy conformations of the system and it is important to run simulations for long enough to be able to compare time averages to experimentally relevant information.

A further consideration when comparing computational and experimental results is the choice of ensemble. The simulation ensemble is characterised by the parameters of the system held constant during the dynamics (Table 2-1). Conditions of constant number of particles, pressure and volume (NPT), or constant number of particles, volume and temperature (NVT), correlate most closely to experimental condition [27]. During simulations the temperature can be kept constant by rescaling velocities, coupling to a heat bath or using the Langevin equations of motion. The canonical (NVT) ensemble was used in the production simulations in this project and the temperature was kept constant using a method based on the Nose-Hoover thermostat, whereby the system is coupled to a heat bath. The free energies calculated in the canonical ensemble correspond to the Helmholz free energy.

TABLE 2-1 Constant variables in different ensembles, and corresponding equilibrium states.

N	P	V	T	E	μ	Equilibrium	Name
x		x	x			A is maximum	Canonical
x		x		x		S has maximum	Micro-canonical
x	x		x			G has maximum	Isothermal-isobaric
		x	x		x	(PV) has maximum	Grand canonical

N: Number of particles; P: Pressure; V: Volume; T: Temperature; E: Energy; μ: Chemical potential; A: Helmholtz free energy; S: Entropy; G: Gibbs free energy. Excerpt from Jensen [2].

2.2.5. SIMULATION PROCEDURE

A typical MD simulation is conducted as follows:

Initial coordinates

Initial coordinates for proteins can often be obtained from structures resolved using X-ray crystallography or Nuclear Magnetic Resonance. Alternatively a theoretical structure obtained from molecular modelling can be used. The choice of initial structure is important and it is best to choose a configuration close to the targeted simulation state [2].

Minimising the starting structure

Some parts of the initial structure may contain high-energy interactions. There is a risk that in the course of the simulation the potential energy of these parts may be transformed into a high level of kinetic energy at a localised spot, known as a 'hot spot'. It is therefore important to remove high-energy interactions by performing an energy minimisation [2].

Assignment of initial velocities

Initial velocities are assigned to all atoms to reach the desired temperature. Assignment is often made according to a Gaussian distribution of velocities.

Heating and equilibration dynamics

The purpose of the heating and equilibration dynamics is to enable the crystal structure to evolve to reach equilibrium. The kinetic energy of the all the atoms is gradually increased by scaling or

reassigning their velocities at regular intervals until the system reaches room temperature. Equilibration is continued until a set of properties become stable. These properties include the thermodynamic properties of energy, temperature and pressure. The root mean squared deviation (RMSD) is a structural property that is a useful metric for equilibration. RMSD of a system of N particles can be calculated using:

$$\Delta r^2(t) = \frac{1}{N}\sum_{i=1}^{N}[\boldsymbol{r}_i(t) - \boldsymbol{r}_i(0)]^2 \tag{2.34}$$

For a solid, the RMSD should oscillate about a mean value when at equilibrium [2, 5].

Production dynamics

The equilibrated structure is used as the starting point and the time evolution of the system corresponding to the desired ensemble is simulated. The length of the simulation is dependent on the measurements to be made and required phase space sampling.

2.2.6. STRUCTURAL ANALYSIS

MD simulations of inhibitor-enzyme and substrate-enzyme complexes aim to describe the structure and stability of the molecular systems. The hydrogen bonding and hydrophobic interactions play an important role in binding of these complexes [28].

Hydrogen bonding is an interaction of the type D-H···A, where a hydrogen is shared between donor and acceptor atoms. The D-H bond length is consistent with a covalent bond while the H···A bond length is shorter than the normal vdW contact distant. Typically the donor and acceptor groups are electronegative groups, and the hydrogen bond has both a covalent and electrostatic nature. The interaction also has a directional preference and the optimal orientation is linear [28]. The strength and geometry hydrogen bonds differ with the chemical properties of the donor and acceptor groups. Since the bonding is an electronic effect, characterisation of these experimentally observed interactions are only possible using QM methods. In MM force fields,

such as the CHARMM-27 force field used here, the hydrogen bond interactions are parameterised into the Coulomb and Lennard-Jones potentials of the atoms, and a definition is required for simulation analyses. Structural properties are often used for this purpose. In the current work hydrogen bond analysis was done on an H···A distance criterion of 2.4Å, and an angle criterion of 120°.

The hydrophobic effect is generally considered to be an entropic effect [5, 28]. Non-polar solutes are not able to form hydrogen bonds with water and in solution will disrupt the hydrogen bond network of water. As a result waters become locally ordered around the solute in a cage-like structure in order to maximise hydrogen bonding with the bulk water. This local order reduces entropy. When non-polar solutes associate water molecules are liberated and the non-polar surface area is reduced thus increasing the entropy. The favourable entropy change in the solvent due to the association of the non-polar solutes is dominant over the unfavourable enthalpic and entropic contributions in the solute. The association of non-polar residues due to the hydrophobic effect are known as hydrophobic interactions. Since the CHARMM force field models atoms explicitly there is no energy term for hydrophobic interactions. In classical molecular dynamics the hydrophobic effect is simulated using parameterised non-bonded terms. These electrostatic and vdW terms give rise to favourable interactions between polar moieties and favourable interactions between non-polar moieties. A suitable definition for identifying favourable hydrophobic interactions is required. One metric that can be used is the solvent exposed surface area of the non-polar residues. Alternatively, an interaction can be evaluated as hydrophobic in nature if the distance between two non-polar residues approaches their vdW contact distance and no water molecules lie between them.

2.3. REFERENCES

1. S. M. Bachrach, *Computational Organic Chemistry*, Wiley-Interscience, Hoboken, 2007.
2. F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons Ltd, Chinchester, 1999.

3. P. Atkins and J. d. Paulo, *Atkins's Physical Chemistry*, 7th edn., Oxford University Press, Oxford, 2002.
4. G. H. Grant and W. G. Richards, *Computational Chemistry*, Oxford University Press, Oxford, 1995.
5. A. R. Leach, *Molecular Modelling: Principles and Applications*, Addison Wesley Longman Limited, Harlow, 1996.
6. D. B. Cook, *Handbook of Computational Chemistry*, Dover Publications Inc., New York, 2005.
7. N. A. Buchmeier, G. L. Newton and R. C. Fahey, *Journal of Bacteriology*, 2006, **188**, 6245-6252.
8. J. A. Pople and R. K. Nesbet, *Journal of Chemical Physics*, 1954, **22**, 571.
9. P. Hohenberg and W. Kohn, *Physical Review B*, 1964, **136**, B864.
10. W. Kohn and L. J. Sham, *Physical Review*, 1965, **140**, 1133.
11. M. Orio, D. A. Pantazis and F. Neese, *Photosynthesis Research*, 2009, **102**, 443-453.
12. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *Journal of Computational Chemistry*, 1983, **4**, 187-217.
13. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *Journal of Computational Chemistry*, 2009, **30**, 1545-1614.
14. W. F. Van Gunsteren and H. J. C. Berendsen, *Angewandte Chemie International Edition in English*, 1990, **29**, 992-1023.
15. P. J. Steinbach and B. R. Brooks, *Journal of Computational Chemistry*, 1994, **15**, 667-683.
16. P. P. Ewald, *Annalen der Physik*, 1921, **369**.
17. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *Journal of Chemical Physics*, 1995, **103**, 8577.

18. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. MacKerell, *Journal of Computational Chemistry*, **31**, 671-690.
19. C. M. Breneman and K. B. Wiberg, *Journal of Computational Chemistry*, 1990, **11**, 361-373.
20. L. Verlet, *Physical Review*, 1967, **159**, 98.
21. R. W. Hockney, *Journal in Computational Physics*, 1970, **9**, 136.
22. W. C. Swope, *Journal of Chemical Physics*, 1982, **76**, 637.
23. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *Journal of Chemical Physics*, 1983, **79**, 926-935.
24. J. Guenot and P. A. Kollman, *Protein Science*, 1992, **1**, 1185-1205.
25. M. W. Mahoney and W. L. Jorgensen, *Journal of Chemical Physics*, 2000, **112**, 8910-8922.
26. D. A. McQuarrie, *Statistical Mechanics*, University Science Books, California, 2000.
27. S. A. Adcock and J. A. McCammon, *Chemical Reviews*, 2006, **106**, 1589-1615.
28. A. Fersht, *Structure and Mechanism in Protein Science*, W. H. Freeman and Company, New York, 1999.

CHAPTER 3

FREE ENERGY AND CONFORMATIONAL ANALYSIS TECHNIQUES

3.1. FREE ENERGY CALCULATIONS

Free energy is an entropic property and is defined as the total amount of energy in a system that can be converted to do work [1]. A change in a system is spontaneous if the resulting change in free energy is negative. Therefore evaluating the free energy can give valuable information on the conformational preferences of molecules and favourable binding of host-guest complexes.

The Helmholtz free energy, corresponding to the canonical ensemble, can be calculated exactly from an ensemble using Equation (3.1):

$$A = k_b T ln\left(\iint d\boldsymbol{p}^N d\boldsymbol{r}^N e^{\frac{+\mathcal{H}(\boldsymbol{p}^N,\boldsymbol{r}^N)}{k_B T}} \rho(\boldsymbol{p}^N,\boldsymbol{r}^N)\right) \tag{3.1}$$

Higher energy configurations play an important role in calculating the free energy [2]. However, MD simulations are formulated to sample the low energy conformations of a system and will not adequately sample important high energy regions. Therefore, using a MD-generated ensemble to calculate the free energy of the system will result in poorly converged and inaccurate results. On the other hand, free energy differences can be readily calculated. They are a useful tool in examining the binding of ligand-enzyme complexes [2, 3].

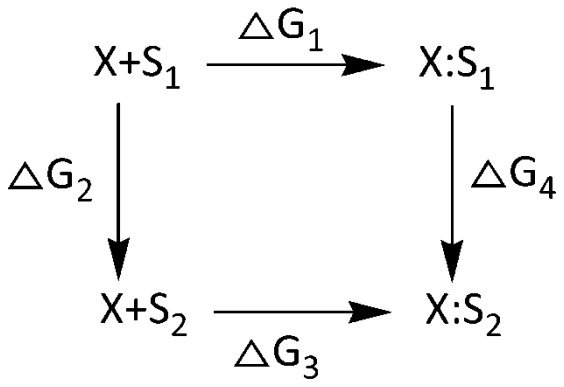


FIGURE 3-1 Closed thermodynamic cycle. ΔG_1 and ΔG_3 are the free energies of association of macromolecule X with S_1 and S_2 respectively. Figure adapted from Grant and Richards [4].

3.1.1. THERMODYNAMIC CYCLE FOR THE BINDING OF HOST-GUEST COMPLEXES

Since free energy is a state function, a closed thermodynamic cycle can be derived for association of two different molecules S_1 and S_2 with a macromolecule X (Figure 3-1). There is no requirement that the free energy change must take place over a physical pathway. Therefore, by 'mutating' ligand **S_1** to ligand **S_2** both in solution and at the active site of an enzyme, the relative free energy of association, $\Delta\Delta G_{ass}$, can be calculated without having to simulate both the ligand binding as well as de-solvation of the ligands and the binding site [4]:

$$\Delta\Delta G_{ass} = \Delta G_1 - \Delta G_3 = \Delta G_2 - \Delta G_4 \quad (3.2)$$

In Equation (3.2), ΔG_1 and ΔG_3 are the absolute free energies of association of X with **S_1** and **S_2** respectively, and cannot be calculated with good accuracy ΔG2 and ΔG4 are computationally feasible and are used to calculate relative free energy of association. The above scheme provides an efficient pathway for applying 'computational alchemy' to investigate the specific contributions to ligand binding. A functional group of the ligand or a protein amino acid can be mutated and the resulting change in free energy evaluated. A negative free energy change indicates a more favourable binding of the product system and the opposite is true for a positive free energy change.

3.1.2. METHODS FOR CALCULATING RELATIVE FREE ENERGY OF ASSOCIATION

There are three methods with which to calculate the free energy difference of two states: Free Energy Perturbation (FEP), Slow Growth (SG) and Thermodynamic Integration (TI) [2, 5-7].

FREE ENERGY PERTURBATION

The relationship between the Helmholtz free energy and the partition function *Q* is given in Equation (3.3):

$$\Delta A = A_Y - A_X = -k_B T ln \frac{Q_Y}{Q_X} \tag{3.3}$$

Zwanzig showed that the free energy difference can be obtained using an ensemble average over the representative configurations of the initial state of the system [8]:

$$\Delta A = -k_B T ln \left\langle e^{\frac{-(\mathcal{H}_Y - \mathcal{H}_X)}{k_B T}} \right\rangle_X \tag{3.4}$$

where $\mathcal{H}_X$ and $\mathcal{H}_Y$ are the Hamiltonians for the initial state X and the final state Y respectively. To calculate the free energy difference a series of configurations for X is generated. The ensemble average is then calculated using the energy calculated for each configuration of X using the potential parameters of Y [2].

If X and Y do not overlap in phase space the phase space of Y will not be adequately sampled when simulating X. Such a scenario arises when $|\mathcal{H}_Y - \mathcal{H}_X| \gg k_B T$ and the calculated relative free energy difference will be inaccurate [2]. This problem can be overcome using intermediate states defined by a parameter λ, which couples the initial and final states. The Hamiltonian of the intermediate state is then defined by a proportion of the initial state and proportion of the final state:

$$\mathcal{H}_\lambda = (1 - \lambda)Y + (\lambda)X \tag{3.5}$$

To perform an FEP calculation a number of intermediate states, λ_i, are specified moving from 0 (corresponding to X) and 1 (corresponding to Y). For each intermediate state the system is equilibrated using the force field parameters appropriate to λ_i. A production stage dynamics is simulated and the difference in free energy calculated as:

$$\Delta A(\lambda_i \rightarrow \lambda_{i+1}) = -k_B T ln \left\langle e^{\frac{-(\mathcal{H}_{i+1}-\mathcal{H}_i)}{k_B T}} \right\rangle \tag{3.6}$$

The total free energy is then the sum of the free energy changes.

Similarly, the free energy difference $\Delta A(\lambda_i \rightarrow \lambda_{i-1})$ can be calculated. In the double-wide sampling technique free energy differences are calculated in both directions from λ_i. This is more efficient since twice as many free energy differences are obtained from a single simulation [2].

SLOW GROWTH

In the slow growth method the energy function is changed continuously over the simulation by incrementing/decrementing λ. The free energy difference is then given by:

$$\Delta A = \sum_{i=1;\,\lambda=0}^{i=N_{step};\lambda=1} (\mathcal{H}_{i+1} - \mathcal{H}_i) \tag{3.7}$$

The assumptions are that if the increments are small enough the system will remain at equilibrium, and furthermore that the ensemble average can be approximated by a single data collection point. In practice a large number of windows are required (50 000-100 000) and the method should be applied with caution as the implicit assumptions detailed above have not yet been proven [5].

THERMODYNAMIC INTEGRATION

The free energy can also be defined as a continuous function of λ :

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \frac{\partial A(\lambda)}{\partial \lambda} d\lambda \tag{3.8}$$

and by substituting

$$A(\lambda) = -k_B T ln Q(\lambda) \tag{3.9}$$

Equation (3.10) can be derived:

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \mathcal{H}(\boldsymbol{p}^N, \boldsymbol{r}^N)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \tag{3.10}$$

where λ is a coupling parameter such that λ = 0 corresponds to the initial state and λ = 1 corresponds to the final state. TI approximates the integral in Equation (3.10). A discrete number of λ_i values are chosen and a series of simulations corresponding to those points is performed. The ensemble average with respect to λ is calculated. The integral is approximated by numerical methods such as the trapezoidal integration method for which Equation (3.10) becomes:

$$\Delta A \approx \sum_{0}^{M} \left(\left\langle \frac{\partial \mathcal{H}(\boldsymbol{p}^N, \boldsymbol{r}^N)}{\partial \lambda} \right\rangle_{\lambda_i} + \left\langle \frac{\partial \mathcal{H}(\boldsymbol{p}^N, \boldsymbol{r}^N)}{\partial \lambda} \right\rangle_{\lambda_{i+1}} \right) \times \left(\frac{\lambda_{i+1} - \lambda_i}{2} \right) \tag{3.11}$$

Although TI makes use of intermediates defined in terms of a coupling parameter, λ_i, unlike FEP, TI is not a windowing procedure. In FEP the formulation of the free energy difference is composed of exact terms and the λ-spacing is dependent on sufficient phase space sampling. Intermediate λ points are interpolated using double-wide sampling [5]. The free energy calculation using TI, on the other hand, requires the analytical integration of Equation (3.10) and the λ spacing is limited by the need for sufficient discrete points to approximate the integral accurately. Therefore, with infinite sampling both formulae are equivalent, however the convergence properties differ and errors can often be reduced using TI [5].

3.1.3. IMPLEMENTATION

SINGLE AND DUAL-TOPOLOGY SIMULATIONS

The formulations introduced above are implemented using either the single-topology [9] or dual-topology [10] methods. The two methods will give different results for the free energy difference, but the double free energy difference will be the same [2, 5, 6].

In single-topology methods every atom in the initial state has a counterpart in the final state [11, 12]. This is illustrated in Figure 3-2A, which considers the transformation of ethane to methanol. Since the number of atoms in the initial and final states is different, dummy atoms (D) are introduced for atoms which exist in one state and have no counterpart in the other. The dummy atoms have no non-bonded energy terms, but they are connected to the rest of the system through bonded interactions. The energy of the Hamiltonian for the system becomes:

$$\mathcal{H}(\boldsymbol{r},\lambda) = \mathcal{H}_0(\boldsymbol{r}) + (1-\lambda)^N \mathcal{H}_{reactant}\ (\boldsymbol{r}) + (\lambda)^N \mathcal{H}_{product}\ (\boldsymbol{r}) \tag{3.12}$$

Where $\mathcal{H}_0$ is the part of the potential energy that does not change, $\mathcal{H}_{reactant}$ contains the energy terms unique to the initial state, and $\mathcal{H}_{product}$ contains the energy terms unique to the final state.

In the dual-topology method the parts which are not the same in the initial and final state are defined simultaneously. In the hybrid shown in Figure 3-2B both the hydroxyl and C_2 methyl fragments are bonded to C_1. The Hamiltonian for the initial and final states therefore involves different coordinates:

$$\mathcal{H}\left(\boldsymbol{r},\boldsymbol{r}_{reactant},\boldsymbol{r}_{product},\lambda\right) = \mathcal{H}_0(\boldsymbol{r}) + (1-\lambda)^N \mathcal{H}_{reactant}\ (\boldsymbol{r},\boldsymbol{r}_{reactant}) + (\lambda)^N \mathcal{H}_{product}\ \left(\boldsymbol{r},\boldsymbol{r}_{product}\right) \tag{3.13}$$

Atoms for which only the charge changes in going from reactant to product are said to have a co-located charge. During the perturbation the electrostatic interactions of the 'colo' atom are treated as perturbation interactions, and the atom is otherwise treated as an environmental atom [12]. Therefore, the electrostatic interactions for which the 'colo' atom is assigned the reactant charge are included in H_R, while electrostatic interactions of the 'colo' atom assigned with the

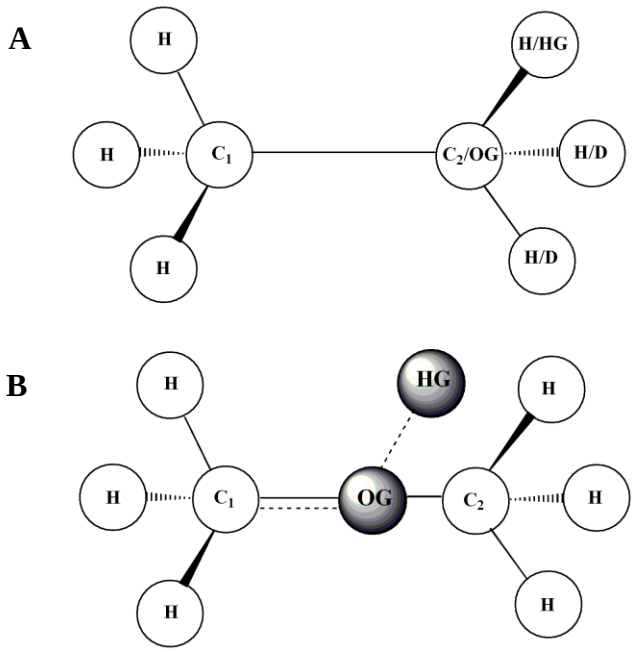


FIGURE 3-2 (A) The hybrid ethane/methanol solute used in the single-topology simulations (B) The ethane/methanol hybrid used in the dual-topology simulations. Figure adapted from Jorgensen and Boresch [11].

product charge are included in H_P. Reactant and product atoms do not see each other in that no non-bonded interactions are computed between the two groups. This thesis employs the TSM module in CHARMM which utilises the dual-topology methodology.

TREATMENT OF BONDED INTERACTIONS

In the dual-topology method, if the bonded interactions for the parts of the system undergoing transformation are scaled, certain bonds and angles are broken/formed and the end state corresponds to an ideal gas atom end state [11]. In the example given in Figure 3-2B the initial state will consist of the ethane molecule and ideal gas atoms of O and H, while the final state will consist of the methanol molecule and ideal gas atoms of C_2 and 3 × H. Difficulties are encountered in breaking or making bonds since there will not be adequate sampling in simulations of finite length [11, 13].

The problems can be avoided by using a hybrid potential energy function that does not scale the bond and angle terms. This will obviously apply only when the transformations does not involve

covalent bond association or dissociation. The end state for the free energy simulations will then correspond to ideal gas molecules in which the molecular fragment remains bonded to the rest of the system while all the other interactions of the fragment are scaled to zero. In the example given in Figure 3-2 the initial state will consist of the ethane and methanol molecules, however, environmental atoms will only interact with the ethane molecule at the initial state and only the methanol molecule at the final state. It has been shown that the artifacts of the simulation arising from not scaling the bonded terms do not affect the double free energy difference [11, 13].

However, such a treatment of the bonded interactions is not appropriate for this thesis. Transformation along the inhibitor series in the direction **1a**-**1d** (Figure 1-7 in Chapter 1) involves addition of a methylene group to the alkyl linker. If the bond and angle interactions for the reactant linker and product linker are treated in full, the linker of the hybrid molecule will assume a conformation which is a compromise between the shorter and longer linker length. Therefore, as λ goes from 0 to 1, the linker will not extend and differences in the interactions of the plumbagin moiety due to extension of the linker will not be simulated for the initial and final states. An alternative strategy is to scale the bonded terms but to limit the value of λ to a value at which a bond or angle is considered broken. Boresch and Karplus have suggested that the limits of 0.01 and 0.99 are appropriate for typical values for force constants of bonded terms [11, 13].

EXAMPLE: RELATIVE FREE ENERGY OF SOLVATION FOR ETHANE AND METHANOL

The difference in free energy of solvation between ethane and methanol is calculated below to illustrate the protocol for free energy calculations:

First, the starting and end topologies for the simulation were chosen. The TSM module in CHARMM was used with the dual-topology method. As discussed both the methyl and hydroxyl fragments will exist simultaneously in the hybrid (Figure 3-2B). The chemical transformation at hand allowed for the internal bond and angle parameters of the hybrid to be neglected as perturbation interactions. Therefore the endpoints were ideal gas molecules.

The CHARMM General Force Field (CGENFF) [14] was used to model the ethane and methanol molecules in solution. The starting conformations of the hybrid's methyl and hydroxyl fragments were generated using the internal coordinates present in the CGENFF topology file. These are minimum energy conformations in vacuum for each of the fragments in relation to the C_1 methyl group.

The atoms were divided into atom groups in order to calculate the classical Hamiltonian for the hybrid:

$$\mathcal{H}(\boldsymbol{r}, \boldsymbol{r}_{ethane}, \boldsymbol{r}_{methanol}, \lambda) = \mathcal{H}_0(\boldsymbol{r}) + (1-\lambda)^N \mathcal{H}_{ethane}(\boldsymbol{r}, \boldsymbol{r}_{ethane}) + (\lambda)^N \mathcal{H}_{methanol}(\boldsymbol{r}, \boldsymbol{r}_{methanol}) \quad (3.14)$$

The atoms of the C_2 methyl group were designated as reactant atoms and the atoms from the hydroxyl fragment were designated as product atoms. The C_1 atom is a 'colo' atom with a reactant charge of –0.27 and a product charge of –0.04. The remaining hybrid and solvent atoms were treated as environmental atoms.

The coupling parameter, λ, was assigned values of 0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 and 1. Non linear λ scaling was used to modify the reactant and product Hamiltonians. In this case, the derivatives due to the product atoms are zero at λ = 0 and the hydroxyl atoms were deleted so that only ethane molecules interacted with the solvent. Likewise, the ethane atoms were deleted at λ = 1. The co-located charge on C_1 was changed to –0.04 as λ was perturbed from 0 to1.

For each λ value the system was equilibrated for 100ps using stochastic boundary conditions. The equilibrated structures were then used for a further 100ps production. The free energy files written out during the production were post-processed using TI. The transformation in vacuum gave ΔA = 4.20kcal/mol, while the transformation in water gave ΔA = –3.27kcal/mol giving a ΔΔG of –7.47kcal. The negative ΔΔA correctly indicates that the solvation of the polar methanol is more favourable, and corresponds to the experimental value for difference in free energy of –6.93kcal/mol. To test for convergence the reverse simulation was conducted. Here the hydroxyl

atoms were designated as reactant atoms, the C_2 methyl atoms were designated as product atoms, and the charge on C_1 was perturbed from –0.04 to –0.27. For the reverse direction $\Delta\Delta A$ = 7.54kcal/mol was obtained and is convergent with the forward calculation.

3.1.4. CONSIDERATIONS

The primary sources of error in free energy calculations are the sampling of phase space and the accuracy of the Hamiltonian [2, 3, 5-7].

The sampling issue refers to the adequate coverage of phase space to give reliable results. There is no set protocol to ensure sufficient sampling of all the relevant conformations. For proteins a X-ray structure for the macromolecule is chosen for which it is assumed that a simulation of 10s to 100s of ps on this structure would retain its essential structure and any mutation would involve a small enough structural change that it occurs in a practical timeframe [3]. Convergence can be evaluated for simulations of different length, transformations in both directions and calculations using different methods (FEP, TI or SG) [3, 5].

The parameters and functional form of the force field, as well as the implementation of the molecular mechanical model must also accurately represent the system. Inaccuracies in force fields mean that conformations from longer simulations may drift from the crystal structure and calculated free energies may be less accurate than for shorter simulations [3].

3.2. CONFORMATIONAL ANALYSIS

3.2.1. SIMULATED ANNEALING

The inhibitor series investigated here contains a pseudo-disaccharide in the form of the 1-phenyl-2-acetamido-α-D-glycopyranoside fragment. Conformation analysis of disacharides must account for the large number of possible combinations of rotational states for the exocyclic functional groups. In such a system with many degrees of freedom, there are multiple minima which prevent simple energy minimisation calculations from surmounting local barriers. However, starting the minimisation at each of the possible combinations of side groups for a particular φ, ψ point on the conformational energy map is computationally expensive [17].

Simulated annealing can be used to efficiently solve the multivariate optimisation problem [17, 18]. This method is equivalent to crystal melting and refreezing into a configurational minimum. At high temperatures the system is able to pass over any energy barriers separating local minima barriers. As the temperature falls, the lower energy states become more probable in accordance with the Boltzmann distribution. At absolute zero the system should occupy the lowest energy state, or global minimum (to ensure the solution is in fact the global minimum, the system would have to come to thermal equilibrium at an infinite number of temperature steps). It has been shown that the application of simulated annealing to the conformational analysis of disaccharides yields the same results as the computationally expensive static molecular mechanics techniques mentioned above [17].

3.2.2. CLUSTER ANALYSIS

A molecular dynamics trajectory contains a set of atomic coordinates which are dependent variables. Multivariate analysis can be used to reduce the dimensionality of the data and group conformations based on measured characteristics. In this thesis, clustering analysis was used to

explore the conformational space of the inhibitor series of interest. A non-hierarchical clustering algorithm, ART-2' [15, 16], was used in CHARMM.

A trajectory is created using MD and the conformation, *j*, of each frame is described as a vector of *N* parameters:

$$x_j = [x_{1j}, \ldots, x_{Nj}] \tag{3.15}$$

The conformations are then clustered according to a cutoff radius in an initial learning phase and subsequent refining phases.

The learning phase is initialised by assigning the first vector, x_j, as the cluster centre of cluster α, c_α. The Euclidean distance between the cluster centre (c_α) and the next conformation k is evaluated:

$$d_{ik} = [(x_k - c_\alpha)^T (x_k - c_\alpha)]^{\frac{1}{2}} \tag{3.16}$$

If the distance is within a specified threshold distance conformation *k* is assigned to cluster α and the cluster centre is recalculated as the average conformation of its M members:

$$c_\alpha = \frac{1}{M} \sum_{x_j \in \alpha} x_j \tag{3.17}$$

If the distance falls outside the specified threshold distance, conformation *k* is assigned as the cluster centre for cluster β. This process is repeated until all conformations have been analysed, and the cluster centres are recalculated as each new member is added. In the refining phase the cluster centres are only updated until all conformations have been read in and assigned. The choice of the cutoff radius is important since if too big a radius is chosen there will be overlap between clusters, but if too small a radius is chosen artificial clusters may form.

3.3. REFERENCES

1. P. Atkins and J. d. Paulo, *Atkins's Physical Chemistry*, 7th edn., Oxford University Press, Oxford, 2002.
2. A. R. Leach, *Molecular Modelling: Principles and Applications*, Addison Wesley Longman Limited, Harlow, 1996.
3. P. Kollman, *Chemical Reviews*, 1993, **93**, 2395-2417.
4. G. H. Grant and W. G. Richards, *Computational Chemistry*, Oxford University Press, Oxford, 1995.
5. D. A. Pearlman and B. G. Rao, in *Encyclopedia of Computational Chemistry*, eds. P. v. R. Schleyer, N. L. Allinger, J. G. T. Clark, P. A. Kollman, I. H. F. Schaefer and R. P. Schreiner, John Wiley & Sons, New York, Editon edn., 1998, pp. 1036-1061.
6. W. L. Jorgensen, in *Encyclopedia of Computational Chemistry*, eds. P. v. R. Schleyer, N. L. Allinger, J. G. T. Clark, P. A. Kollman, I. H. F. Schaefer and R. P. Schreiner, John Wiley & Sons, New York, Editon edn., 1998, p. 1061.
7. A. E. Mark, in *Encyclopedia of Computational Chemistry*, eds. P. v. R. Schleyer, N. L. Allinger, J. G. T. Clark, P. A. Kollman, I. H. F. Schaefer and R. P. Schreiner, John Wiley & Sons, New York, Editon edn., 1998, p. 1070.
8. R. W. Zwanzig, *The Journal of Chemical Physics*, 1954, **22**, 1420-1426.
9. J. Gao, K. Kuczera, B. Tidor and M. Karplus, *Science*, 1989, **244**, 1069-1072.
10. D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel and P. Kollman, *Computer Physics Communications*, 1995, **91**, 1-41.
11. S. Boresch and M. Karplus, *Journal of Physical Chemistry A*, 1999, **103**, 119-136.
12. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *Journal of Computational Chemistry*, 2009, **30**, 1545-1614.
13. S. Boresch and M. Karplus, *Journal of Physical Chemistry A*, 1999, **103**, 103-118.

14. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. MacKerell, *Journal of Computational Chemistry*, **31**, 671-690.
15. G. A. Carpenter and S. Grossberg, *Applied Optics*, 1987, **26**, 4919-4930.
16. M. E. Karpen, D. J. Tobias and C. L. Brooks, *Biochemistry*, 1993, **32**, 412-420.
17. K. J. Naidoo and J. W. Brady, *Chemical Physics*, 1997, **224**, 263-273.
18. R. B. Best, G. E. Jackson and K. J. Naidoo, *Journal of Physical Chemistry B*, 2001, **105**, 4742-4751.

CHAPTER 4

MSHB ACTIVE SITE AND NATURAL SUBSTRATE BINDING

4.1 INTRODUCTION

The structure of 1-D-*myo*-inosityl-2-acetamido-2-deoxy-α-D-glucopyranoside Deacetylase (GlcNAc-Ins Deacetyalse, MshB), and the catalytic mechanism for deactylation of GlcNAc-Ins (Figure 4-1A) have been presented in detail in Chapter 1. As discussed, the binding mode of the natural substrate and the proposed mechanism (Figure 4-1C) remain uncertain. The binding mode of the 2-acetamido glucose unit has been modeled by Baker and coworkers [1] (Figure 4-1B) based on the orientation a β-octylglucoside (BOG) molecule, which was adventitiously co-crystallised in the putative active site of their MshB crystal structure. In addition docking studies have also been used to explore the binding of GlcNAc-Ins [2, 3], but there is no report in the literature of the dynamic behaviour of MshB. Therefore, in view of the primary aim of this project, to describe the binding of the inhibitor series **1a-d** (Figure 1-7, Chapter 1), it is important to delineate the MshB active site and model the binding of GlcNAc-Ins.

There are two crystal structures available for MshB, pdb structures 1Q74 [4] and 1Q7T [1]. Crystal structure 1Q7T, resolved at 1.9Å by Baker *et al.*, contains two independent protein molecules (Chain A and Chain B). Chain A contains the co-crysallised BOG molecule used to model the binding of GlcNAc, and is also the most complete structure. Whereas 1Q74 is missing coordinates for residues 1-2, 100-103, 164-169 and 300-303, 1Q7T Chain A is only missing residues 165-170 only. For the above reasons, the 1Q7T structure was used in the current research.

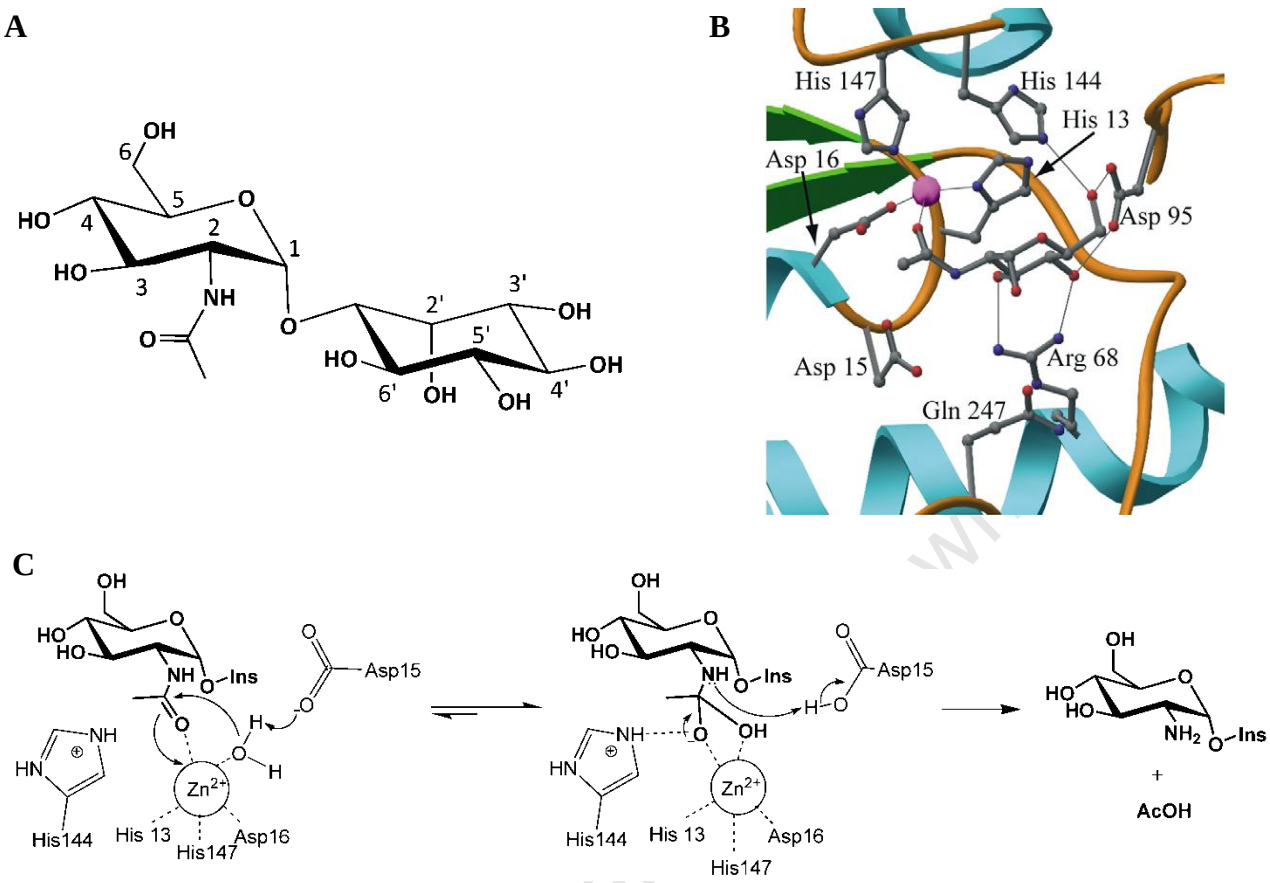


FIGURE 4-1 **(A) Natural substrate with carbon atoms numbered. (B) Proposed binding of the GlcNAc moiety of the natural substrate modeled on the binding of BOG. The model shows coordination of the acetyl oxygen to the zinc ion . The binding of the glucose moiety is by mediated hydrogen bonds between the oxygens at C-3 and C-4, and Arg68. Asp95 forms hydrogen bonds with the oxygens at C-4 and C-6. Figure taken from Baker *et al.* [1]. (C) Proposed mechanism for the catalytic deacetylation of GlcNAc-Ins. Figure adapted from Fahey *et al.* [5].**

4.2 METHODOLOGY

4.2.1 PROTEIN PREPARATION

STRUCTURE REFINEMENT

The missing residues of MshB, 165-170, form a loop between helices 6 and 7 and is situated on the surface of the protein some 25Å away from the recognised zinc binding pocket (Figure 1-5, Chapter 1). The loop was repaired by building the missing residues in Maestro [6] and minimising the loop using the OPLS 2005 force field while keeping all resolved atoms fixed. The zinc binding site was modeled on the 1Q74 holoenzyme crystal structure. To achieve this, 1Q74 chain A was aligned to 1Q7T chain A and the coordinates of the zinc, together with its bound waters, were added to 1Q7T. The His13 side chain was then flipped so that $N^{\delta 1}$ approached within 2.2Å of the metal ion. Finally the expression tag residues of 1Q7T were deleted.

PROTONATION

Protons were added to the repaired structure to simulate the protein at a pH of 7. The protein was analysed using pKaTool [7], which yielded a titration curve and a calculated pKa for each residue. When the pKa of an ionisable group is well below the current pH, the amino acid will most likely exist in its deprotonated state. Similarly amino acids with a pKa greater than the current pH are likely to be protonated. The calculated pKa's for the 23 Asp and 14 Glu residues of the protein were well below 7 suggesting deprotonated states at that pH value. The calculated pKa's for the 2 Lys, 20 Arg and 8 Tyr residues suggested the normal protonated states. For the His residues which had pKa values close to 7, protonation states were assigned by examining the titration curves from the pKaTool calculations and looking at the environment of each residue. This analysis indicated that His27, His60 and His144 should be protonated. However, the above protonation scheme gave a total charge of -12, and it was decided for more accurate simulation of the electrostatic interactions in CHARMM to protonate His170, His149, His131 and His36 which were more than 15Å from the suggested active site. The final charge of MshB was then -8.

4.2.2 BINDING POCKET DETECTION

Potential MshB binding sites for inhibitor series **1a-1d** were located using CASTp [8]. CASTp is a webserver that uses alpha shape and discrete flow theory to locate and measures protein pockets and pocket mouth openings, as well as cavities. A single cavity was detected and a molecular surface of the atoms lining the binding site was calculated in PyMol [9] using van der Waals (vdW) radii from the CHARMM27 force field [10, 11]. The electrostatic surface potential was calculated using Adaptive Poisson-Boltzmann Solver (APBS) [12] which solves the Poisson-Boltzmann equation numerically.

4.2.3 DOCKING

DOCKING PROCEDURE

Since the binding mode of the natural substrate is unknown, the previously modeled binding orientation of the 2-acetamido glucose unit (Figure 4-1B) was used as a template for *de novo* docking. Docking was carried out using Glide [13-16]. A receptor grid (inner-box size 10Å x 10Å x 10Å; outer-box size 35Å x 35Å x 35Å) was generated for MshB. The zinc-coordinated waters were excluded to allow sampling of possible coordination geometries for the acetyl oxygen and zinc. The grid was centered inside the cavity identified by CASTp and a metal site was specified at the Zn^{2+} ion. The grid was then used in standard precision docking calculations, in which the MshB protein structure was held rigid during the docking. Poses had to satisfy a favourable metal interaction between Zn^{2+} and the GlcNAc-Ins amide oxygen. The docking poses were evaluated by the empirical scoring function GlideScore, and poses were regarded as duplicates if they had an RMS of less than 1.5Å and a maximum atom displacement of less than 2.3Å. The highest ranked pose with a binding mode resembling that modeled by Baker *et al.* was chosen for subsequent dynamics simulations.

4.2.4 MOLECULAR DYNAMICS SIMULATIONS

The CHARMM33b2 program [17, 18] was used for all macromolecular simulations. The empirical energy function employed by CHARMM is shown below:

$$U(\boldsymbol{r}) = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\varphi\big(1 + cos(n\varphi - \delta)\big)$$
$$+ \sum_{impropers} K_\omega(\omega - \omega_0)^2 + \sum_{residues} U_{CMAP}(\varphi, \psi) + \sum_{Urey-Bradley} K_{UB}(S - S_0)^2$$
$$+ \sum_{non-bonded\ pairs} \left\{ \varepsilon_{ij}^{min} \left[\left(\frac{R_{ij}^{min}}{\boldsymbol{r}_{ij}} \right)^{12} - 2\left(\frac{R_{ij}^{min}}{\boldsymbol{r}_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon \boldsymbol{r}_{ij}} \right\} \quad (4.1)$$

The bonded and non-bonded terms were calculated using the CHARMM27 force field parameters [10, 11] for protein atoms and CSFF[19] force field parameters for GlcNAc-Ins atoms. The zinc ion was modeled using recent experimentally consistent parameters developed by Naidoo *et al.* [20]. A 14Å cut-off radius was applied in the generation of the pairlist, and a 12Å cut-off radius was applied on an atom by atom basis for non-bonded energy calculations. The long range interactions were truncated using a shifting function.

For stochastic boundary dynamics, a spherical boundary force was applied to the surface of a 29Å water sphere to maintain a consistent water density of 1.0g/dm^3. A 6Å thick outer shell was selected as the buffer region, wherein langevin friction coefficients of 62.0 ps^{-1} and 200 ps^{-1} were applied to water oxygen atoms and protein heavy atoms respectively. Stochastic boundary forces were applied to protein heavy atoms in the buffer region, and multiplied by a screening function in 13 discrete intervals. The stochastic boundary forces were 2.31 kcal.mol^{-1}.Å^{-2} for carbon and nitrogen atoms, 2.17kcal.mol^{-1}.Å^{-2} for oxygen atoms, and 1.30kcal.mol^{-1}.Å^{-2} for sidechain heavy atoms. All hydrogen bond lengths were kept fixed using the SHAKE algorithm [21]. The simulation protocol for the enzyme substrate complex is described below.

SIMULATION SETUP

The coordinates of the protein and ligand atoms were initialised in CHARMM. The waters of crystallisation, including one of the zinc bound waters, were read in and their hydrogens added using the hbuild utility. The complex was then carefully minimised in vacuum to reduce any steric clashes: (1) The protein atoms were fixed and the ligand, waters of crystallisation and protein hydrogens were minimised for 100 steps; (2) The protein backbone was fixed and a harmonic force of 30N was applied to the side chains and then progressively removed over three cycles of 200 minimisation steps; (3) A harmonic force of 30N was applied to the backbone atoms and then progressively removed over three cycles of 200 minimisation steps; (4) The whole system was minimised without any constraints for 200 steps. All minimisations were performed using the Adopted Basis Newton-Raphson (ABNR) method with a step size of 0.2. After the initial minimisation, the complex was hydrated in a 29Å sphere of TIP3P [22] water molecules to a density of 1.21g/dm^3. The hydrated complex was minimised in four steps: (1) Protein heavy atoms were fixed while the ligand, water and protein hydrogen atoms were minimised for 1000 steps; (2) The protein backbone atoms were fixed while the rest of the system was minimised for 1000 steps; (3) Finally the whole system was minimised without any constraints for 1000 steps.

HEATING

After minimization the solvated system was slowly heated from 138.15K to 298.15K. A 1ps verlet dynamics was run, with velocities initially assigned to a Gaussian distribution at 138.15K. The velocities were then scaled every 0.1ps, using a single factor for all atoms, until the system reached 148.15K. Thereafter, fifteen cycles of 5ps langevin dynamics was run using the leapfrog verlet algorithm. During each cycle the velocities were scaled to increase the system's temperature by 10K in the first 2ps, and equilibration was run at that temperature for the next 3ps. A stepsize of 1fs was used and the non-bonded lists were updated every 20 steps. After heating to 298.15K, the system was re-solvated to a density of 1.23g/dm^3 in order to fill any vacuum spaces near the boundary created by diffusion of waters into the protein.

DYNAMICS

Atom velocities were reassigned at 298.15K according to a Gaussian distribution, and, after a 1ps verlet dynamics, the system was subjected to 5ns of langevin dynamics using the leapfrog verlet integrator. A step size of 1fs was used and the non-bonded lists were updated every 20 steps. Coordinates were saved every 0.2ps.

4.3 PHYSIOCHEMICAL PROPERTIES OF THE MSHB BINDING SITE

4.3.1 GEOMETRY OF THE BINDING SITE

MshB is a monomer consisting of 303 amino acids that are folded into a single α/β domain (Section 1.2). The loops and excursions originating from the C-terminal ends of the β strands form the putative active site as well as a three-stranded antiparallel β-sheet subdomain. Analysis of MshB using CASTp revealed a large binding site (Figure 4-2). The site has been divided into geometric regions in relation to the putative active site, which is positioned near the center. The active site cavity, annotated Region A, is composed of the BOG-binding residues as well as the zinc binding pocket. On one side the active site narrows at Gly97 and Asp95 to form Region B, a tunnel between the loop connecting helices α_2 and α_4, and the loop linking β-strand β_2 and helix α_4. On the opposite side of the active site, Tyr142 lies across the pocket partially closing off the active site from a large, buried cavity. This cavity, coloured purple, sits between the mixed β-sheet of the α/β domain and the anti-parallel β-sheet comprising the subdomain. In the crystal structure resolved by James and coworkers [4] Tyr142 faces away from the binding pocket towards the protein's environment.

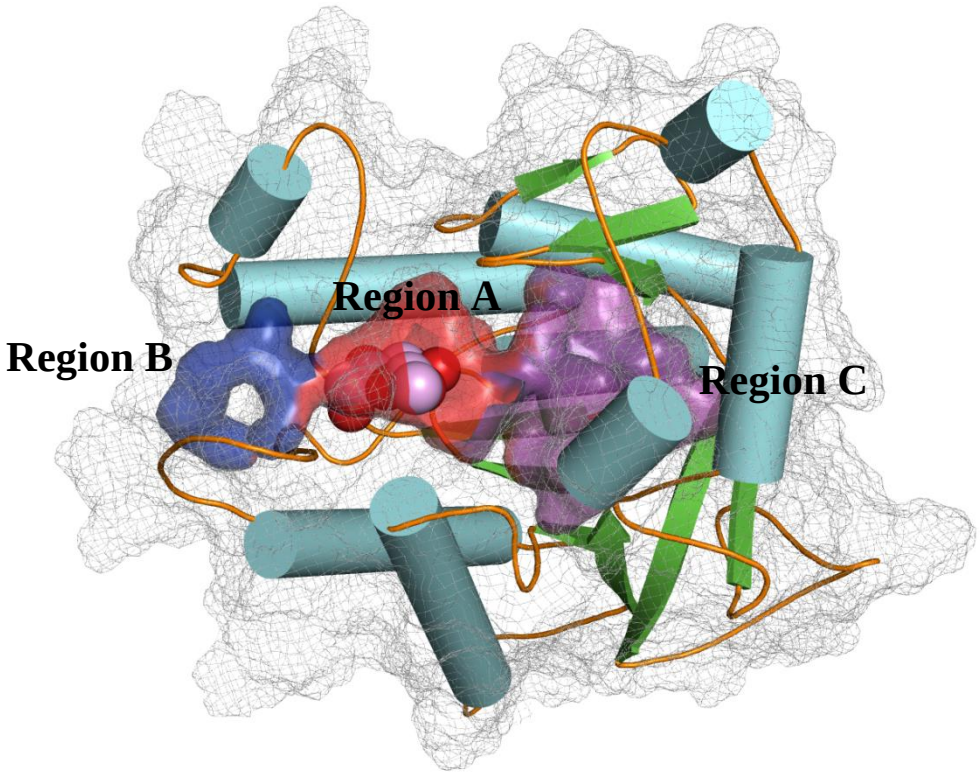


FIGURE 4-2 Molecular surface of the binding site detected by CASTp. The putative active site is shown in red and encloses a vdW representation of the BOG molecule co-crystallised in the 1Q7T structure.

4.3.2 CHEMICAL PROPERTIES OF THE BINDING SITE

The residues forming the MshB binding site are summarised in Table 4-1. The amino acids are grouped according to the region in which they are found and the polarity of their side chain. Region A, the active site cavity, is composed primarily of residues with polar side chains. As previously discussed in Section 1.2, Asp16, His13 and His147 form the zinc binding pocket, while Arg68, Asp15, Asp95 and His144 are proposed to mediate binding of the glucose ring in the natural substrate [1]. With this proposed glucose binding mode the α-linked *myo*-inositol group at C-1 of the natural substrate would sit at the mouth of the binding pocket and the amide group would point towards Region C. Glu47 Gln247, Ser260 and Asn261 are all positioned at the roof of the Region A as depicted, and are potential hydrogen bonding partners for *myo*-inositol. Of the non-polar residues, Met98 helps form the mouth of the binding site on the left hand side of Region A, while Val248 and Leu259 are situated next to Region C. The latter are in close proximity to the position of the non-polar acetyl hydrogens.

TABLE 4-1 Decomposition of binding site residues according to region and property of their side chain.

Region	Side Chain Property	Residue
A	**Charged Polar**	Arg 68 Asp 15, 95, 16 His 144 Glu 47
	Uncharged Polar	His 13, 147 Gln 247 Ser 260 Asn 261
	Non-polar	Met 98 Val 248 Leu 259
B	**Charged Polar**	Asp 137 Lys 237 Glu 271
	Uncharged Polar	Ser 20 Tyr 136, 142, 218 thr 183 His 272
	Non-polar	Leu 19, 185, 190, 241, 263, 265, 268 Val 184 Pro 209 Ile 214 Phe 216 Ala 257, 189 Pro 266 Ile 267 Gly 186
C	**Charged Polar**	Glu 45 Asp 59, 62, 102 Arg 94, 104, 108
	Uncharged Polar	Thr 57, 101 Gln 103 Ser 105
	Non-polar	Ala 58

Region B consists of almost exclusively polar residues (Table 4-1) with a large number of negatively charged sidechains. Region C on the other hand has many non-polar residues lining its surface. Figure 4-3C shows the contribution of non-polar residues to the molecular surface of the MshB binding site.

The electrostatic potential calculated for MshB using APBS [12] was plotted onto the molecular surfaces of the protein and binding site (Figure 4-3). Region A and Region B are mostly electronegative with some areas of positive electronegativity due to contributions from Zn^{2+}, Arg68 and His144. Region C shows a much more neutral electronegativity profile, which is indicative of its non-polar, hydrophobic nature.

4.4 NATURAL SUBSTRATE BINDING

GlcNAc-Ins was docked into the active site of MshB, and the solvated system was subjected to 5ns molecular dynamics (MD) simulation. The resulting trajectory was analysed to investigate the binding of the natural substrate. The root-mean-squared-deviation (RMSD) of amino acids within 8Å of GlcNAc-Ins was plotted against time, once the translational and rotational movement of the macromolecule had been removed (Figure 4-4). The plot shows the active site settles quickly and can be seen to be stable from 2ns onwards.

The final nanosecond of the trajectory was used to analyse the enzyme-substrate binding interactions. The interaction energy (IE) was calculated for residues of increasing distance from the inhibitor in order to identify residues which are important for binding. The non-bonded interactions were then calculated for these residues every 0.2ns, and the average values are reported in Table 4-2. Hydrogen bonding was detected using an in house code based on a distance cutoff of 3.5Å and an angle cutoff of 120^{o}. Since the CHARMM force field models atoms explicitly, there is no energy term for hydrophobic interactions. Rather, the hydrophobic effect is simulated using parameterised non-bonded terms. These electrostatic and vdW terms

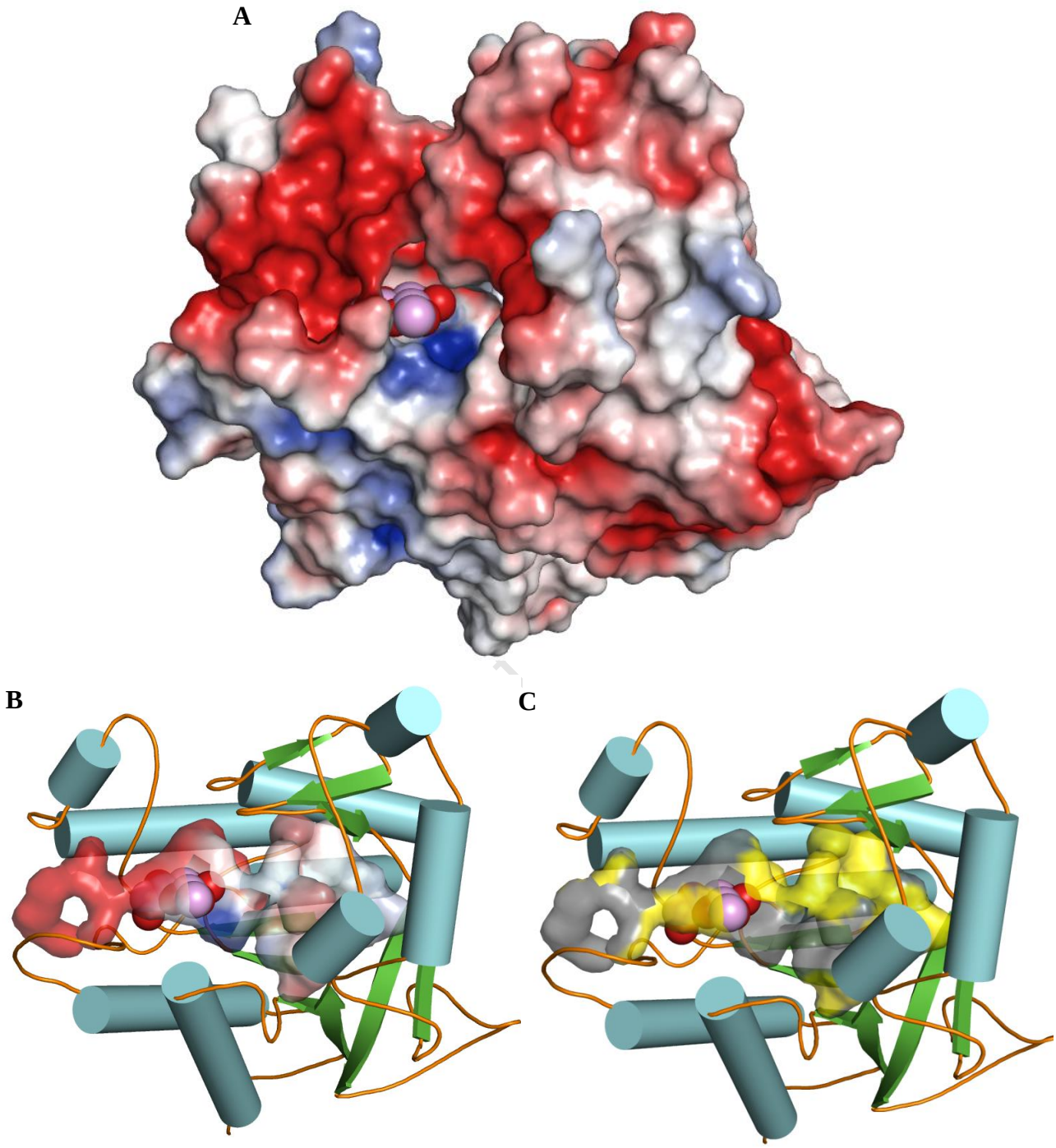


FIGURE 4-3 **(A) Electrostatic potential mapped on to the molecular surface of MshB. Blue shows areas of positive electronegativity, and red shows areas of negative electronegativity. The RWB color scheme covers the range from -3 to 3kT.e^{-1}. (B) Electrostatic potential mapped onto the molecular surface of the MshB binding site. (C) Molecular surface of the binding pocket for which the contribution to the surface by hydrophobic resdues is coloured yellow.**

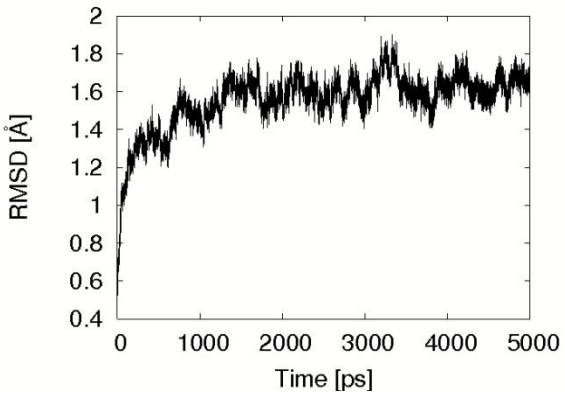


FIGURE 4-4 RMSD plot for all amino acid residues which have an atom within 8Å from GlcNAc-Ins.

give rise to favourable interactions between moieties of the same polarity. Therefore, hydrophobic interactions were evaluated based on an inter-residue distance in conjunction with the vdW interaction energy.

Figure 4-5 summarises the binding simulated between the natural substrate and MshB (GlcNAc-Ins is illustrated with relevant atom labels in Figure 4-1A). The IE of binding is dominated by the electrostatic interaction between the acetamido oxygen and zinc ion (−93.27kcal/mol). The C-2 acetamido methyl group is in close proximity to the non-polar residue Leu259, which has an IE of −1.66kcal/mol with the natural substrate. Figure 4-6 shows that the distance between the center of mass (COM) of the Leu259 side chain and the methyl group fluctuates minimally around its optimal distance. This distance was calculated by moving the Leu259 away from the methyl group in vacuum and locating the energy minimum. Furthermore, there are no water molecules between the two groups, which are seen to have favourable hydrophobic interactions.

Recognition of the glucose moiety is mediated by Arg68 (−16.48kcal/mol) and Asp95 (−22.79kcal/mol) through hydrogen bonds. The oxygen from the C-3 hydroxyl group is a hydrogen bond acceptor to Arg68 $N^{\eta 1}$, while the hydroxyl hydrogen displays hydrogen bonding geometry, for a significant proportion of the trajectory, with a water molecule bridging the proposed catalytic residue Asp15. The oxygen of the hydroxyl group at C-4 is also hydrogen bonded to Arg68 $N^{\eta 1}$ and the hydroxyl hydrogen displays a hydrogen bond with Asp95 $O^{\delta 2}$. $O^{\delta 1}$ of Asp95 is hydrogen bonded to the hydroxyl hydrogen at C-6. Contrary to the binding model

proposed by Baker *et al.*, His144 fulfills the geometric criteria for hydrogen bonding with the C-6 hydroxyl oxygen only sparsely over the 1ns trajectory.

The *myo*-inositol moiety sits at the entrance of the binding pocket exposing the hydroxyls to non-specific interactions with the bulk water. However, the oxygen of the hydroxyl group at C-2′ is hydrogen bonded to a water molecule bridging Glu47 and the hydroxyl hydrogen displays hydrogen bond geometry with Ser260 for part of the trajectory. The oxygen from the hydroxyl at C-5′ displays consistent hydrogen bonding to Asn261 over the trajectory. Contacts between the non-polar hydrogens of Met98 and the three axial hydrogens on one face of the D-*myo*-inositol satisfied the distance criteria for hydrophobic interactions (Figure 4-6). Gammon *et al.* [2] have reported that in a study of the substrate specificity of MshB, a benzyl group was found to be the most suitable analogue for D-*myo*-inositol. It is envisaged that the non-polar benzyl moiety forms hydrophobic interactions with Met98 in a similar fashion to the interactions reported here for D-*myo*-inositol. This would explain the suitability of the benzyl group as a substrate analogue.

TABLE 4-2 Decomposition of the IEs between GlcNAc-Ins and MshB residues. The sum of the reported energies are given (Total), along with the IEs calculated for the entire inhibitor with MshB. All energies are reported in kcal/mol.

Residue	Total	VDW	ELEC
Zn	-93.27	6.43	-99.70
Leu259	-1.66	-0.65	-1.01
Arg68	-16.48	0.42	-16.90
Asp95	-22.79	3.14	-25.93
Asp15	-3.71	-1.04	-2.67
His144	-9.08	-0.46	-8.62
Met98	-2.10	-3.45	1.35
Ser260	-1.07	-0.92	-0.15
Asn261	-4.02	-1.46	-2.57
Total	**-154.18**	**2.00**	**-156.19**
MshB	**-140.47**	**-14.32**	**-126.15**

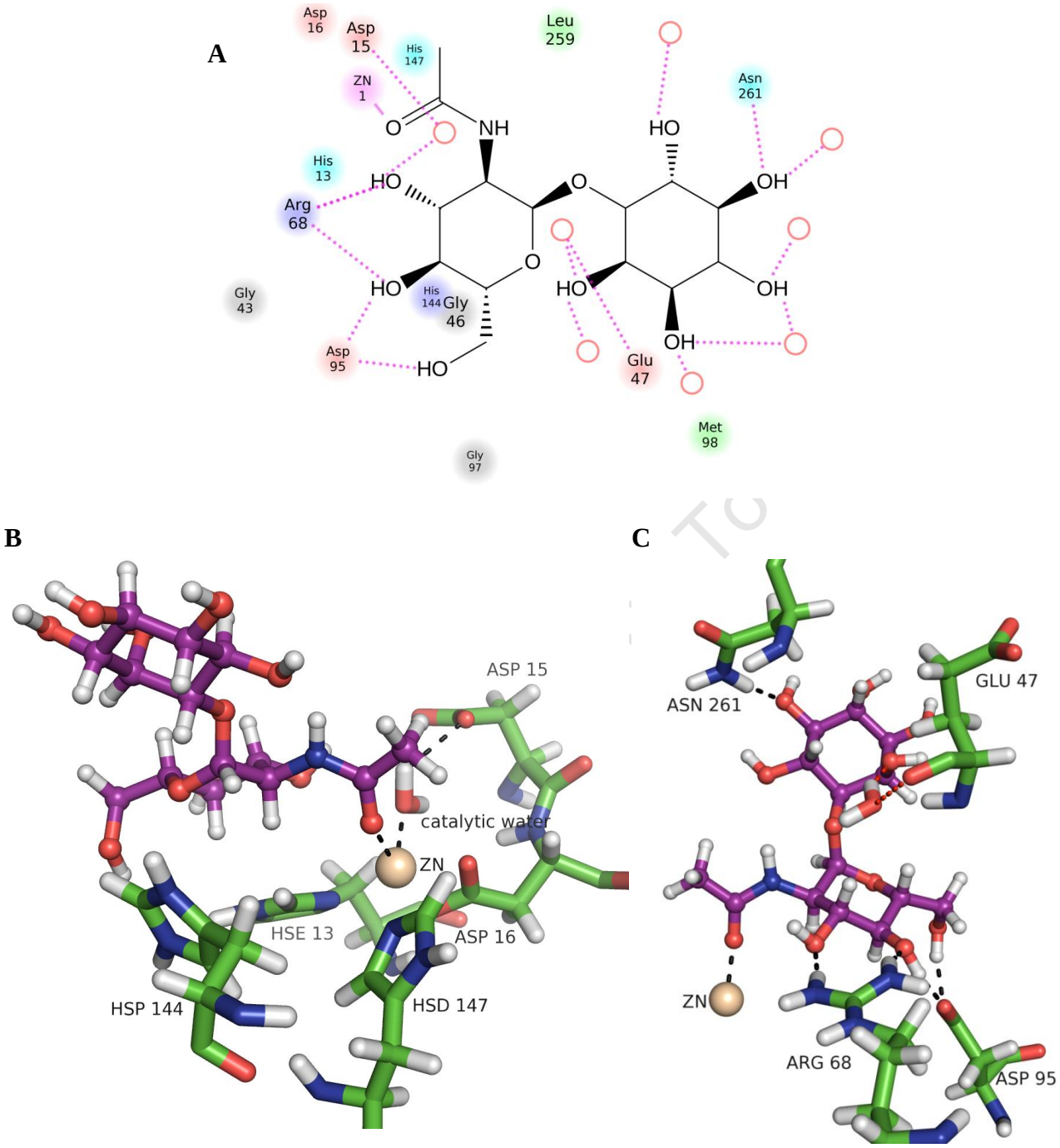


FIGURE 4-5 **(A) 2D diagram showing the binding mode of the natural substrate after 5ns of MD simulation. (B) 3D illustration of the natural substrate binding after 5ns MD showing the zinc binding pocket and catalytic residues. The catalytic water is labelled and its oxygen is in a suitable position for attack on the acetylamido carbonyl. (C) Hydrogen bond network mediating the binding of GlcNAc-Ins. The water-bridged interaction between Glu47 and the hydroxyl oxygen at C-2′ is shown in red.**

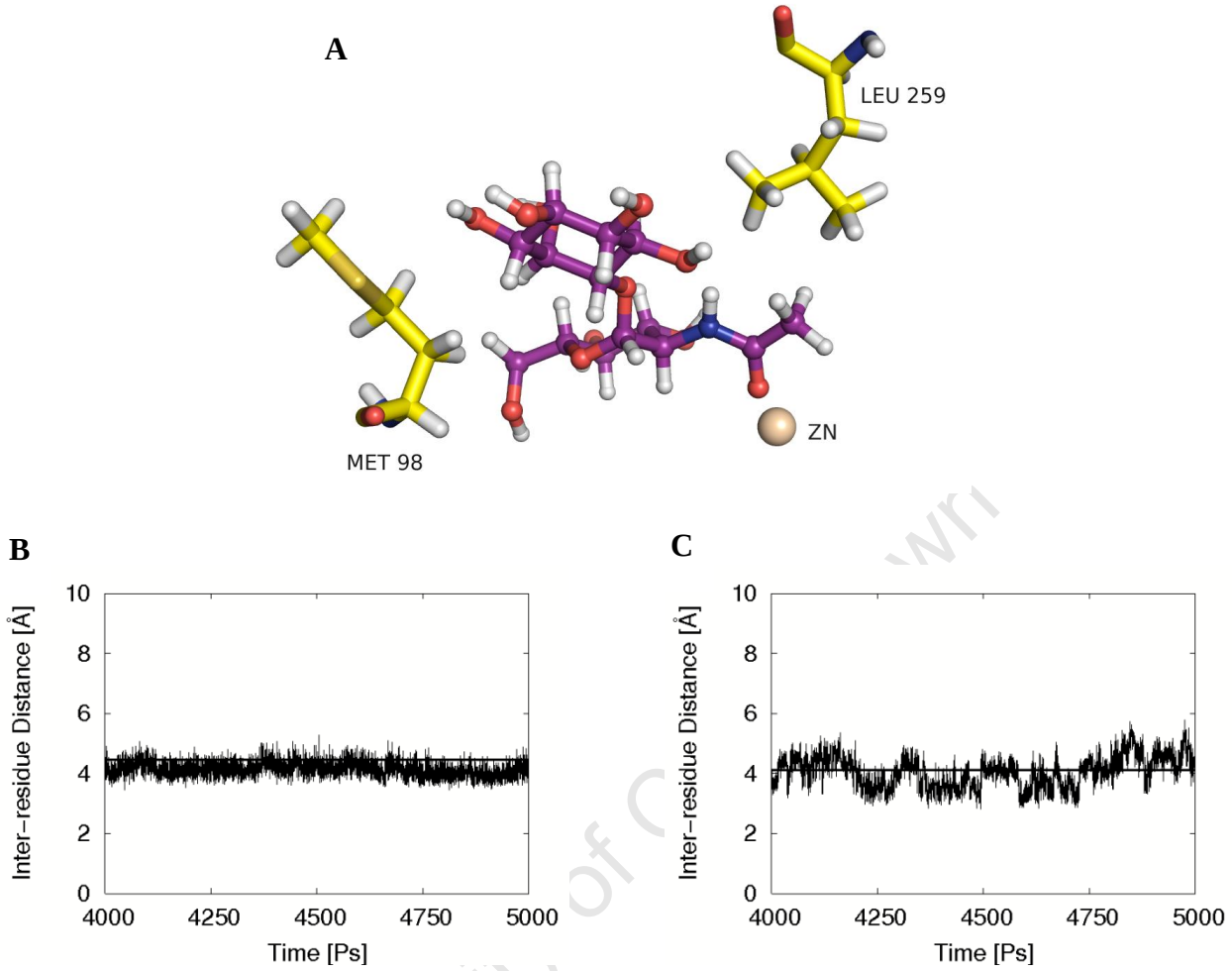


FIGURE 4-6 **(A) Hydrophobic interactions involving Met98 and Leu259. (B) Time series for the distance between the COM of Met98 side chain atoms and the COM of non-polar hydrogens from the face of the *myo*-inositol closest to Met98. (C) Time series for the distance between the COM of non-polar hydrogens from Leu259 and the COM of the C-2 acetamido methyl group. Horizontal lines show the optimal inter-residue distance calculated from interaction energy scans.**

4.5 PLUMBAGIN CONJUGATES AS MSHB INHIBITORS

The current research aims to investigate the inhibition of substituted naphthoquinones in relation to binding strength. The series of inhibitors under investigation consists of a methyl derivative of plumbagin linked at the C-2 position of a phenyl thioglycoside. Since the 2-acylamido-α-D-glucose unit is identical to that present in the natural substrate, it is envisaged that MshB will

bind this moiety selectively and in a similar orientation to GlcNAc-Ins. In this orientation the naphthoquinone derivative and alkyl linker would be directed towards the cavity adjacent to the active site, which has a complementary hydrophobic character (Figure 4-3). The C-1 phenylthio group, like *myo*-inositol, would sit at the mouth of the binding pocket in close proximity to the non-polar Met98.

4.6 REFERENCES

1. A. A. McCarthy, N. A. Peterson, R. Knijff and E. N. Baker, *Journal of Molecular Biology*, 2004, **335**, 1131-1141.
2. D. W. Gammon, D. J. Steenkamp, V. Mavumengwana, M. J. Marakalala, T. T. Mudzunga, R. Hunter and M. Munyololo, *Bioorganic & Medicinal Chemistry*, **18**, 2501-2514.
3. B. B. Metaferia, B. J. Fetterolf, S. Shazad-Ul-Hussan, M. Moravec, J. A. Smith, S. Ray, M. T. Gutierrez-Lugo and C. A. Bewley, *Journal of Medicinal Chemistry*, 2007, **50**, 6326-6336.
4. J. T. Maynes, C. Garen, M. M. Cherney, G. Newton, D. Arad, Y. Av-Gay, R. C. Fahey and M. N. G. James, *Journal of Biological Chemistry*, 2003, **278**, 47166-47170.
5. G. L. Newton and R. C. Fahey, *Archives of Microbiology*, 2002, **178**, 388-394.
6. Maestro, version 9.1, Schrödinger, LLC, New York, NY, 2010.
7. J. E. Nielsen and G. Vriend, *Proteins-Structure Function and Genetics*, 2001, **43**, 403-412.
8. J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz and J. Liang, *Nucleic Acids Research*, 2006, **34**, W116-W118.
9. The PyMOL Molecular Graphics System, Version 1.3r3pre, Schrödinger, LLC.
10. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux,

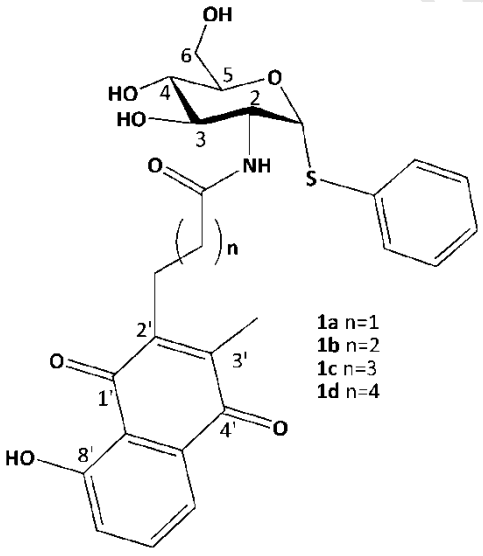
M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *Journal of Physical Chemistry B*, 1998, **102**, 3586-3616.

11. A. D. Mackerell, M. Feig and C. L. Brooks, *Journal of Computational Chemistry*, 2004, **25**, 1400-1415.
12. N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *Proceedings of the National Academy of Sciences of the United States of America*, 2001, **98**, 10037-10041.
13. Glide, version 5.6, Schrödinger, LLC, New York, NY, 2010.
14. R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *Journal of Medicinal Chemistry*, 2004, **47**, 1739-1749.
15. R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *Journal of Medicinal Chemistry*, 2006, **49**, 6177-6196.
16. T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *Journal of Medicinal Chemistry*, 2004, **47**, 1750-1759.
17. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *Journal of Computational Chemistry*, 1983, **4**, 187-217.
18. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *Journal of Computational Chemistry*, 2009, **30**, 1545-1614.
19. M. Kuttel, J. W. Brady and K. J. Naidoo, *Journal of Computational Chemistry*, 2002, **23**, 1236-1243.
20. R. P. Matthews and K. J. Naidoo, *Journal of Physical Chemistry B*, **114**, 7286-7293.
21. J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *Journal of Computational Physics*, 1977, **23**, 327-341.
22. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *Journal of Chemical Physics*, 1983, **79**, 926-935.

CHAPTER 5

RATIONALISING THE INHIBITION OF MSHB BY A SERIES OF PLUMBAGIN CONJUGATES

A series of inhibitors, constructed from plumbagin tethered *via* 2 to 5 methylene carbons and an amide linkage to phenyl-2-deoxy-2-amino-1-thio-α-D-glucopyranoside (GlcNAc-SPh, Figure 5-1), have been shown to inhibit MshB [1]. Kinetic studies revealed the inhibition to be competitive and inhibition constants were calculated for **1a**, **1c** and **1d**. Since the inhibition constant, K_i, is the dissociation constant of the enzyme-inhibitor complex, it is a measure of its strength: a high K_i indicates weak binding and a low K_i indicates strong binding. The K_i values of the plumbagin derivatives indicate that their binding affinity increases with the length of the spacer. This chapter describes the use of free energy calculations to calculate the relative binding strengths of the inhibitors. Molecular dynamics (MD) trajectories of the inhibitors bound to the MshB active site are then used to rationalise the trend in inhibition.



Substrate	Percentage MshB Inhibition	Calculated K_i (µM)
1a	57.4%	167±15
1b	81.6%	-
1c	81.4%	94±11
1d	94.8%	16.8±19

FIGURE 5-1 (A) Structure of inhibitor series with carbon atoms number. (B) Table reporting the percentage inhibition of MshB at 500µM, and calculated K_i value for each of the inhibitors.

5.1. METHODOLOGY

5.1.1. PARAMETERISATION

The GlcNAc-SPh moiety was modelled using the CSFF force field [2]. The phenyl, alkyl linker and plumbagin functional groups were modelled using the CGENFF force field [3], and missing parameters were estimated from analogous bonds, angles and dihedrals. However, there were no templates for the thioglycoside dihedral or quinone charges, and these terms required parameterisation.

PLUMBAGIN CHARGES

Brooks *et al.* have investigated how different charge assignment schemes for small molecules in conjunction with the CHARMM22 force field impact the quality of computed binding affinities [4]. It was found that fitting partial charges to the electrostatic potential yielded the best results. Thus, in the current research a 5-hydroxy-2,3-dimethyl-1,4-naphthoquinone fragment was optimized at the MP2/6-31g(d) level of theory using Guassian03 [5]. Partial charges were then fit to the electrostatic potential using the CHELPG algorithm [6] (Figure 5-2). The charges of the quinone and hydroxyl moiety, including the ipso carbon, were estimated directly from the CHELPG charges and then uniformly offset such that this group of atoms had a total charge of 0. The remaining atoms were assigned charges according to the CGENFF force field guidelines [3]. Thus, aliphatic hydrogen atoms were assigned a charge of +0.09 and the aromatic C-H groups not adjacent to a heteroatom were assigned a charge of -0.115 and +0.115 on the C and H atoms respectively. To validate the charge assignment the electrostatic potential from the QM minimised structure was compared to electrostatic potential generated from point partial charges using an in-house code (Figure 5-3).

DIHEDRALS IN THE THIOGLYCOSIDE LINKAGE

Dihedral parameters for the thioglycoside linkage (Figure 5-4) were absent in the CGENFF and CHARMM27 force fields. The φ and ψ parameters were first approximated using CSFF-consistent parameters for sulphur linked saccharides [7]. To validate the parameters a relaxed dihedral scan of GlcN-SPh was simulated using GuassianO3 at the HF/6-31g(d) level of theory. Coordinates at the stationary point determined for each (φ,ψ) combination were read into CHARMM and minimised. The energies from the QM and MM simulations were used to plot comparative adiabatic maps (Figure 5-4). Use of the CSFF-consistent parameters yielded an energy profile with an extra minima at φ=50. However, when Amber-consistent parameters [8] were used for φ, an adiabatic map similar to that for the HF calculation was obtained.

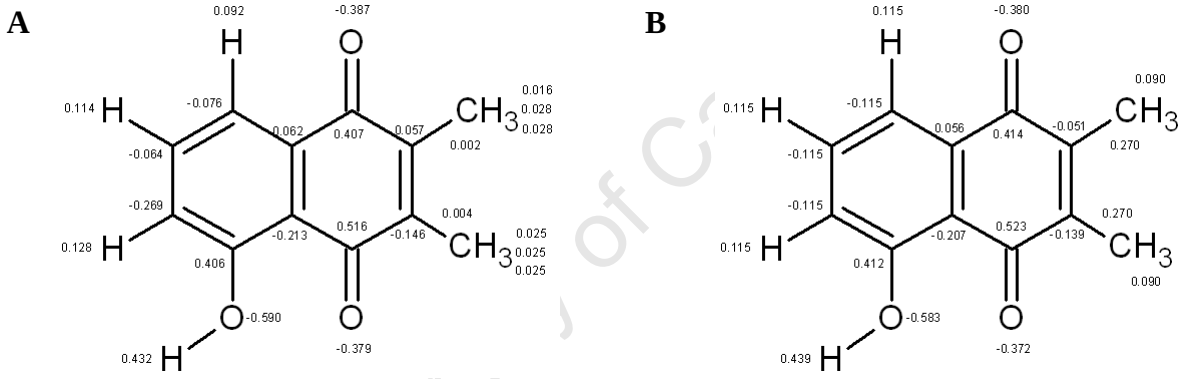


FIGURE 5-2 (A) Charges calculated for the 5-hydroxy-2,3-dimethyl-1,4-naphthoquinone fragment using the CHELPG algorithm and (B) final charge assignment used in the current research.

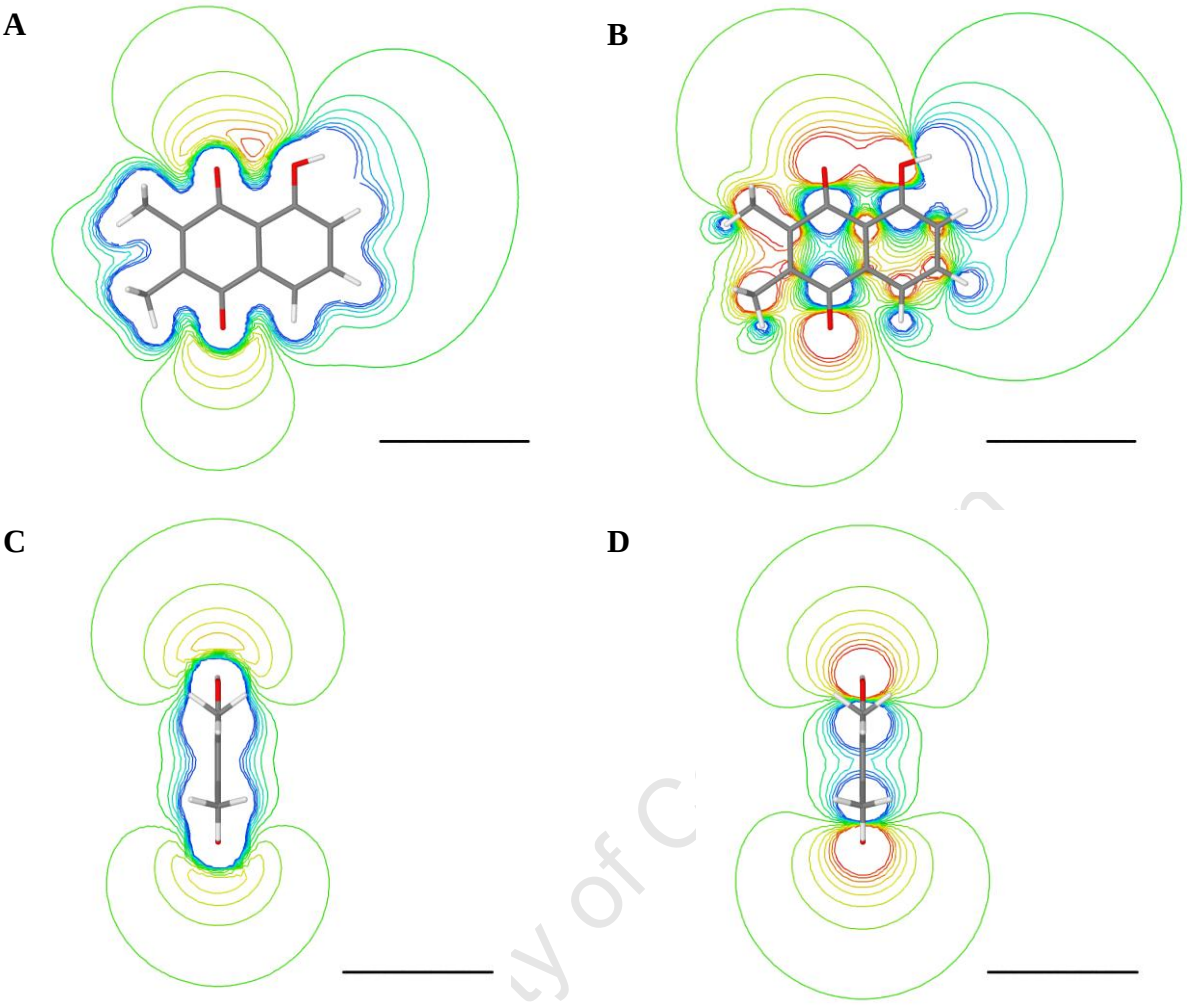
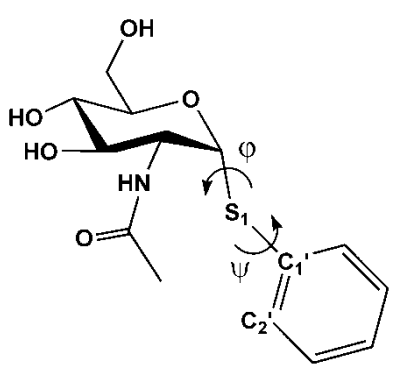
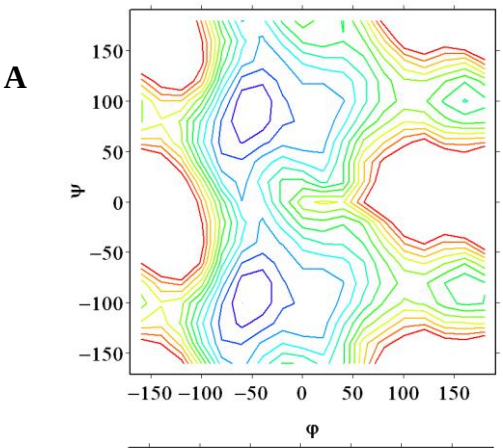
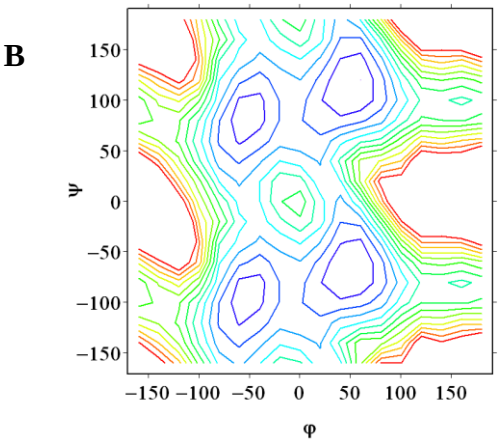
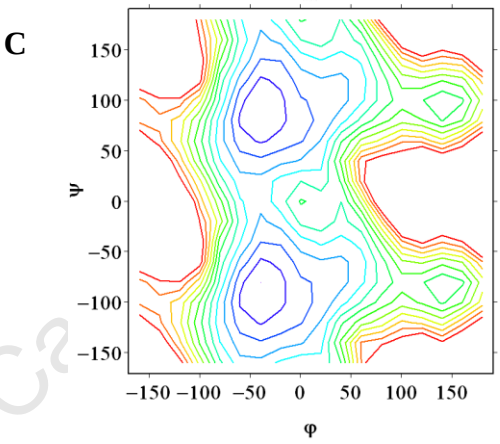


FIGURE 5-3 **2D contour plots for the QM (A & C) and MM (B & D) generated electrostatic potentials for 5-hydroxy-2,3-dimethyl-1,4-naphthoquinone: (A & B) Electrostatic potential in the plane of the aromatic groups and (C&D) Electrostatic potential perpendicular to the plane of the aromatic groups and through the quinone oxygens.**









Dihedral	K	n	δ
O5-C1-S1-C1′	0.1095	1	0
	-0.6860	2	0
	0.7820	3	0
C2-C1-S1-C1′	-0.7257	1	0
	-0.1997	2	0
	0.3473	3	0
H1-C1-S1-C1′	0.1854	3	0
C2′/C3′-C1′-S-C1	-0.7257	1	0
	-0.1997	2	0
	0.3691	3	0
C2′-C1′-S1-C1	-0.7257	1	0
	-0.1997	2	0
	0.3473	3	0

Dihedral	K	n	δ
O5-C1-S1-C1′	1.4200	1	-9.7
	1.2700	2	-27.8
	0.2000	3	-32.1
C2-C1-S1-C1′	0.8100	1	-22.8
	-0.2100	2	-83.8
	0.1800	3	-77.5
	0.4600	4	-77.5
H1-C1-S1-C1′	0.1853	3	0
C2′/C3′-C1′-S-C1	-0.7257	1	0
	-0.1997	2	0
	0.3691	3	0
C2′-C1′-S1-C1	-0.7257	1	0
	-0.1997	2	0
	0.3473	3	0

Figure 5-4 **Parameterisation of the thioglycoside dihedrals. (A) QM adiabatic map. (B) MM adiabatic map using CSFF-consistent parameters. (C) MM adiabatic map using Amber parameters for φ and CSFF-consistent parameters for ψ. Contours are plotted from 0 to 12kcal/mol at intervals of 1kcal/mol. The parameters for B and C are given in the tables below each adiabatic map**

5.1.2. STRUCTURAL STUDY OF INHIBITORS IN SOLUTION

SIMULATED ANNEALING

The phenyl 2-acetylamido-1-thioglycoside unit was subjected to simulated annealing analysis, according to the procedure described by Naidoo *et al.* [9]. Thus, an energy surface was generated in three stages:

The energy landscape of the pseudo-disaccharide was divided into 361 grid points. Grid points were defined by each combination of the (φ_i,ψ_i) dihedral angles resulting from their rotation through 360˚ in intervals of 20˚. At each (φ, ψ) combination the GlcNAc-SPh fragment was subjected to a three step simulated annealing procedure. First energy was supplied to the system in order to overcome any barriers separating local minima. This was achieved by heating the pseudo-disaccharide from 200K to 900K in intervals of 100K every 1ps over a period of 9ps. Next the molecule was cooled to 300K over 20ps, in decrements of 60K every 2ps forcing it into lower energy regions. These low energy regions were explored by dynamic quenching, where the temperature was dropped to 250K after 2ps and then to 200K after a further 2ps, before the system was equilibrated at 200K for 11ps. Finally the coordinates resulting from this annealing process were minimised with 1000 steps of conjugated gradient minimisation, while constraining the glucose ring to a $^4C_1$ chair and the thioglycosidic dihedrals to the (φ, ψ) combination of the relevant grid point. The final combination of exocyclic dihedral angles of the glucose ring was used in the starting geometry for each new next grid point.

Five low energy conformations identified from the results of the above calculation were refined by repeating the simulated annealing procedure 35 times to obtain a normal distribution of means. Initial conditions were different at every point since the velocities were assigned using a newly generated set of random numbers. For each calculation starting geometries were continually replaced with the lowest energy geometry obtained from the preceding runs in order to increase the probability of locating a global minimum at that φ, ψ point.

A smoothed adiabatic map was finally obtained by evaluating six geometries at each of the 361 grid points describing the energy landscape of GlcNAc-SPh. The geometries were: the

conformation at that point from step 1, and five new geometries constructed by superimposing onto this initial conformation the five different sets of exocyclic dihedrals obtained from the refinement procedure.

SOLUTION DYNAMICS

Inhibitors were solvated in a 50Å^3 water box of TIP3P [10] water molecules such that the water density was 1g/dm^3. Each solute was placed in the centre of the box, and all water molecules closer than 2.2Å from the solute heavy atoms were deleted. The system was slowly heated from 48K to 298K over 25ps using the leapfrog verlet integrator, and equilibrated at a constant pressure of 1atm for a further 250ps, using Nose-Hoover thermostat to control the temperature. A 14Å cut-off radius was applied in the generation of the pairlist, and a 12Å cut-off radius was applied on an atom by atom basis for non-bonded energy calculations. The electrostatic interactions were treated using the Ewald summation, and the long range van der Waals (vdW) interactions were truncated using a shifting function. Each of the systems were then subjected to a further 5ns of MD using the Leapfrog Verlet integrator. The simulations were conducted under conditions of constant volume and temperature, and the systems were considered to be at equilibrium when the temperature was stable at 298K, and the kinetic energy fluctuated minimally around a steady value.

5.1.3. SIMULATION OF INHIBITORS DOCKED WITH MSHB

DOCKING

The inhibitors were docked with the MshB crystal structure repaired in Chapter 4. Docking was carried out using Glide [11-14]. A receptor grid (inner-box size 10Å x 10Å x 10Å; outer-box size 35Å x 35Å x 35Å) was generated for MshB without the zinc-coordinated waters. The grid was centred inside the binding cavity identified by CASTp and Asp95 $O^{\delta 2}$ was specified for use in subsequent steps. Standard precision docking calculations were run with the MshB protein structure held rigid. The charges and vdW radii of atoms belonging to Tyr142, which lies across the pocket in the 1Q7T crystal structure, were scaled by a factor of 0.2. The glucose C-4 hydroxyl

hydrogen was constrained to within 2.7Å of Asp95 $O^{\delta 2}$. The top twenty distinct poses were reported, where poses were regarded as duplicates if they had an RMS of less than 0.5Å and a maximum atom displacement of less than 1.3Å. The docking poses were evaluated by the empirical scoring function GlideScore, and the highest ranked pose with a binding mode of the sugar resembling that modelled by Baker et. al [15] (and discussed in Chapter 4) was chosen for subsequent dynamics simulations.

DYNAMICS

The docked complexes were solvated, minimised, heated and then subjected to 5ns stochastic boundary dynamics using the same procedure described for the natural substrate in Section 4.2.4.

5.1.4. FREE ENERGY SIMULATIONS

The methodology of free energy calculations is given in Section 3.1. Calculation of the relative free energy of solvation for ethane and methanol is presented in Section 3.1.3 to aid the reader in understanding the calculation process using thermodynamic integration (TI).

Dual-topology free energy simulations were run in the TSM module in CHARMM [16, 17]. Calculations were run for hybrids in both directions, from **1a** to **1d** and from **1d** to **1a**. Inhibitor hybrids for water free energy calculations were constructed from the most populated cluster centres found in the principal component analysis of the inhibitor trajectories in water, and hybrids for protein free energy calculations were constructed from the equilibrated enzyme-inhibitor complexes. Inhibitor hybrids were constructed by building additional methylene atoms onto the structure of the initial inhibitor, so that the alkyl linkers of the reactant and product inhibitors existed simultaneously. To reduce errors a minimum number of atoms were mutated. Thus, a single methylene group was inserted/deleted at the C-2′ position of the plumbagin (Figure 5-1). The enzyme hybrid and water hybrid systems were then partitioned into environmental atoms, reactant atoms (linker atoms belonging exclusively to the reactant) and product atoms (linker atoms belonging exclusively to product atoms), where energy terms between the reactant

and product atoms were ignored. Internal bond, angle and dihedral terms of the alkyl linkers were treated as perturbation terms, since it is important that the linker chain extends fully.

Free energy perturbation (FEP) methods were initially used, but large changes in ΔA at the end-points were observed indicating insufficient phase space sampling. This was seen to be the result of product/reactant atoms that were only weakly bonded to the hybrid since the bonded terms of the alkyl linker were scaled. This was overcome using TI and non-linear λ scaling. (Dependence of ΔA on λ for the transformation of **1d**-**1c** using both FEP and TI is illustrated in Appendix A. As noted in Section 3.1.2, in FEP, intermediate λ points are interpolated using double-wide sampling and the λ-spacing is dependent on sufficient phase space sampling. On the other hand λ-spacing in TI is dependent on sufficient discrete points to approximate the integral accurately). For each transformation, free energy data at the initial and final states and from λ= 0.01 to 0.99 was analysed using thermodynamic integration. λ points were spaced at intervals of 0.025 except at the endpoints (0.01 to 0.1 and 0.9 to 0.99) where the values were spaced at intervals of 0.005. In the protein simulations each intermediate was equilibrated for 400ps followed by 100ps of data collection, and in water simulations each intermediate was equilibrated for 200ps to ensure the inhibitor did not approach the Langevin boundary.

5.2. STRUCTURAL STUDY OF HYDRATED INHIBITORS

Simulated annealing (Section 3.2.1) was used to describe the energy landscape of the pseudo-disaccharide GlcNAc-SPh with respect to the two thioglycoside dihedrals φ (C2-C1-S1-C1′) and ψ (C1-S1-C1′-C2′). The resulting adiabatic map of phenyl 2-acetamido-1- thioglycoside showed low energy conformations at φ=-40 (Figure 5-5). Within this valley there are two minima at ψ -100 and 100, separated by a barrier of 2kcal/mol. For simulations of **1a**, **1b**, **1c** and **1d** in solution, inhibitors were initialised with φ=-40 and ψ=100, and the alkyl linker projecting linearly from the C-2 amido group. The inhibitors were solvated in water and subjected to 2.5ns of molecular dynamics. Analysis showed that the thioglycoside linkage populated both of the low

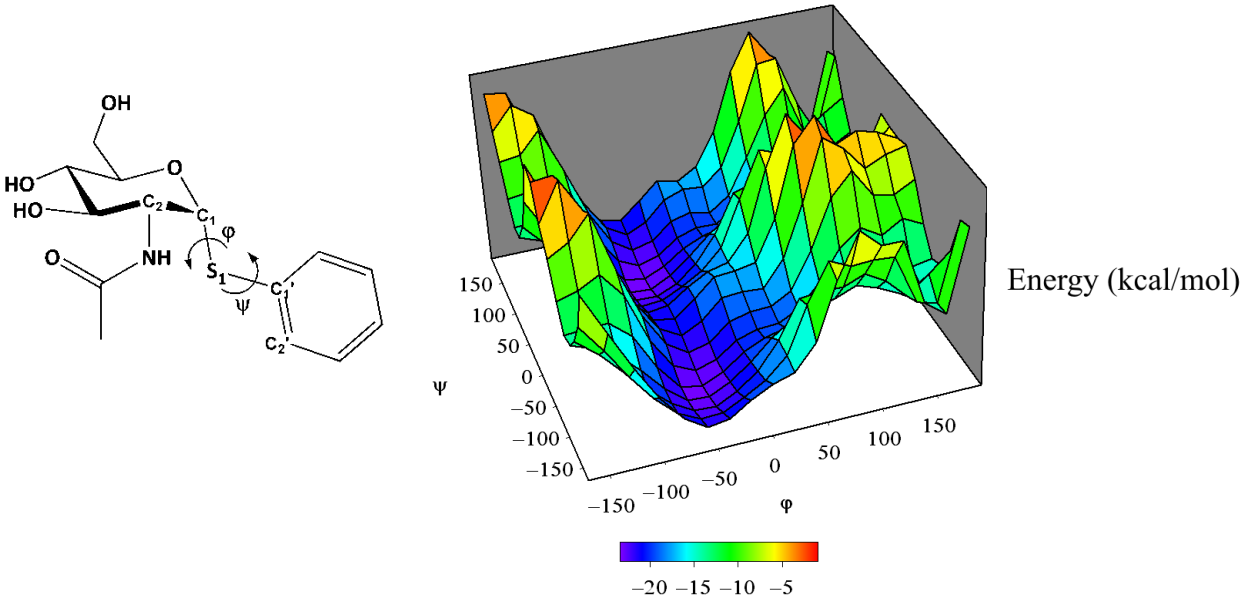


FIGURE 5-5 Adiabatic map obtained from the simulated annealing of the phenyl 2-acetamido thioglycoside fragment shown alongside.

energy (φ,ψ) conformations (Figure B-1, Appendix B). The trajectories were then analysed with principal component analysis (PCA), using the ART-2' algorithm [18, 19] (Section 3.2.2). The linker dihedrals were used as the principal component (Figure 5-6) and a cluster threshold was chosen to give the least cluster overlap, while at the same time not artificially separating clusters. The results are summarised in Table 5-1. Although there is little variation in the dihedral connecting the linker to the naphthoquinone, the rest of the linker shows a greater freedom. Figure 5-7 shows the conformations of the most populated cluster centres from each simulation. Interestingly, these cluster centres have their plumbagin moiety proximal to the phenylthio and/or linker groups. This is seen to be a result of the hydrophobic effect, in which the two hydrophobic moieties try to maximise their surface contact and in so doing release waters to the bulk solution and increase the entropy of the system. In **1a** and **1b** the phenyl ring interacts with the linker and plumbagin methyl groups, but in **1c** and **1d** a longer linker results in the aromatic hydrogens of plumbagin approaching the phenylthio group. To further investigate the conformation of each inhibitor in solution, **1b, 1c** and **1d** were built with their plumbagin moiety sitting under the phenylthio moiety and subjected to 5ns of molecular dynamics. Two starting conformations (annotated i and ii) were used for the simulation of each inhibitor, such that each face of the

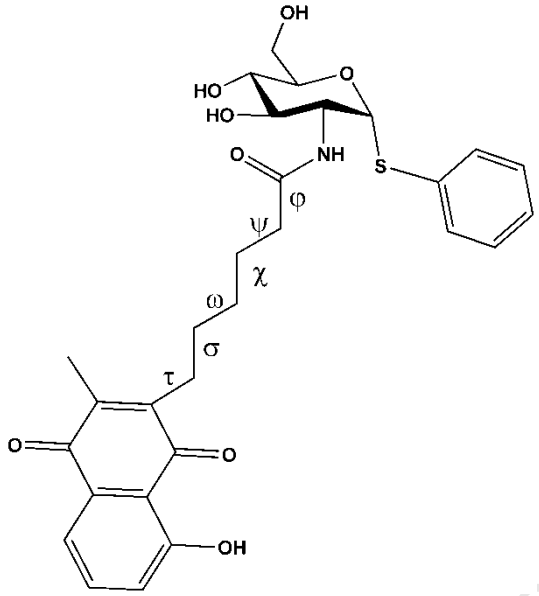


FIGURE 5-6 1d linker dihedrals used as the principal component in the first PCA analysis.

planar plumbagin was presented to the phenylthio group. The conformations from the trajectories were analysed using PCA, with the distance between the centres of mass (COM) of the phenylthio and naphthoquinone moieties as the criteria for clustering. The clusters are summarised in Table 5-2. With the exception of **1b**(ii), the most populated clusters are described by centres which have the smallest distance between their phenylthio and naphthoquinone groups, and typically the number of members decreases as the bond distance of the cluster centre increases. The conformations for the centres of the most populated clusters are presented in Figure 5-8. The cluster centres from the analysis of the two **1b** simulations are very similar, with the plumbagin methyl group and the phenyl ring in close proximity. The cluster centres of **1c** and **1d** show geometry indicative of favourable intramolecular hydrophobic interactions between the phenylthio and plumbagin aromatic groups. For each of these two inhibitors, different MD starting conformations resulted in two distinct relative conformations of the phenylthio and plumbagin groups at the cluster centres. The most populated cluster centre for **1a** in the first PCA, and the most popular cluster centres for **1b**, **1c**(i), **1c**(ii), **1d**(i) and **1d**(ii) in the second PCA, were used for the construction of hybrids employed in the free energy calculations.

TABLE 5-1 Clusters arising from inhibitor trajectories in solution, using linker dihedrals as the principal component.

Sim	Cluster	#members	std. dev.	φ	ψ	χ	ω	σ	τ
1a	1	7076	49.90	128.8	-170.44	-35.01	-	-	-
	2	6496	53.00	124.92	171.3	-25.07	-	-	-
	3	5072	48.50	-137.39	171.36	-42.09	-	-	-
	4	5356	46.30	-128.4	-171.28	-51.27	-	-	-
1b	1	9048	98.50	35.81	-155.6	-10.47	88.9	-	-
	2	14952	96.40	44.08	117.77	-6.14	88.48	-	-
1c	1	18451	109.00	105.71	-24.01	-26.6	14.48	-75.06	-
	2	5549	122.00	-136.38	9.33	-26.98	-29.29	-47.94	-
	3	8443	84.70	-76.91	100.59	43.63	146.41	84.32	-
1d	1	3736	62.00	22.15	111.24	155.18	52.85	160.47	-91.12
	2	6649	75.60	58.39	-17.58	124.25	52.84	-171.33	-88.31
	3	2971	91.80	42.04	124.56	-165.29	50.39	33.05	-90.99
	4	3299	97.00	61	-119.14	-158.41	-4.82	15.13	-89.86
	5	7345	48.90	92.46	-96.95	98.31	64.5	165.34	-86.25

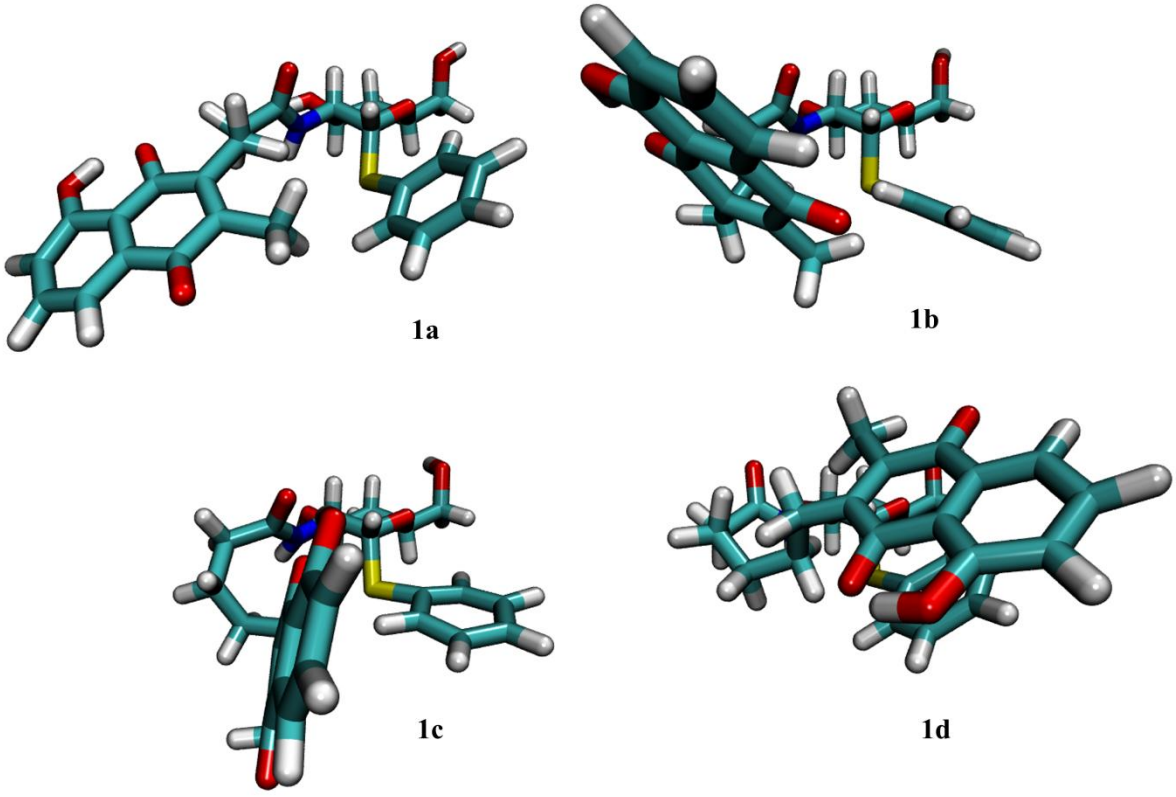
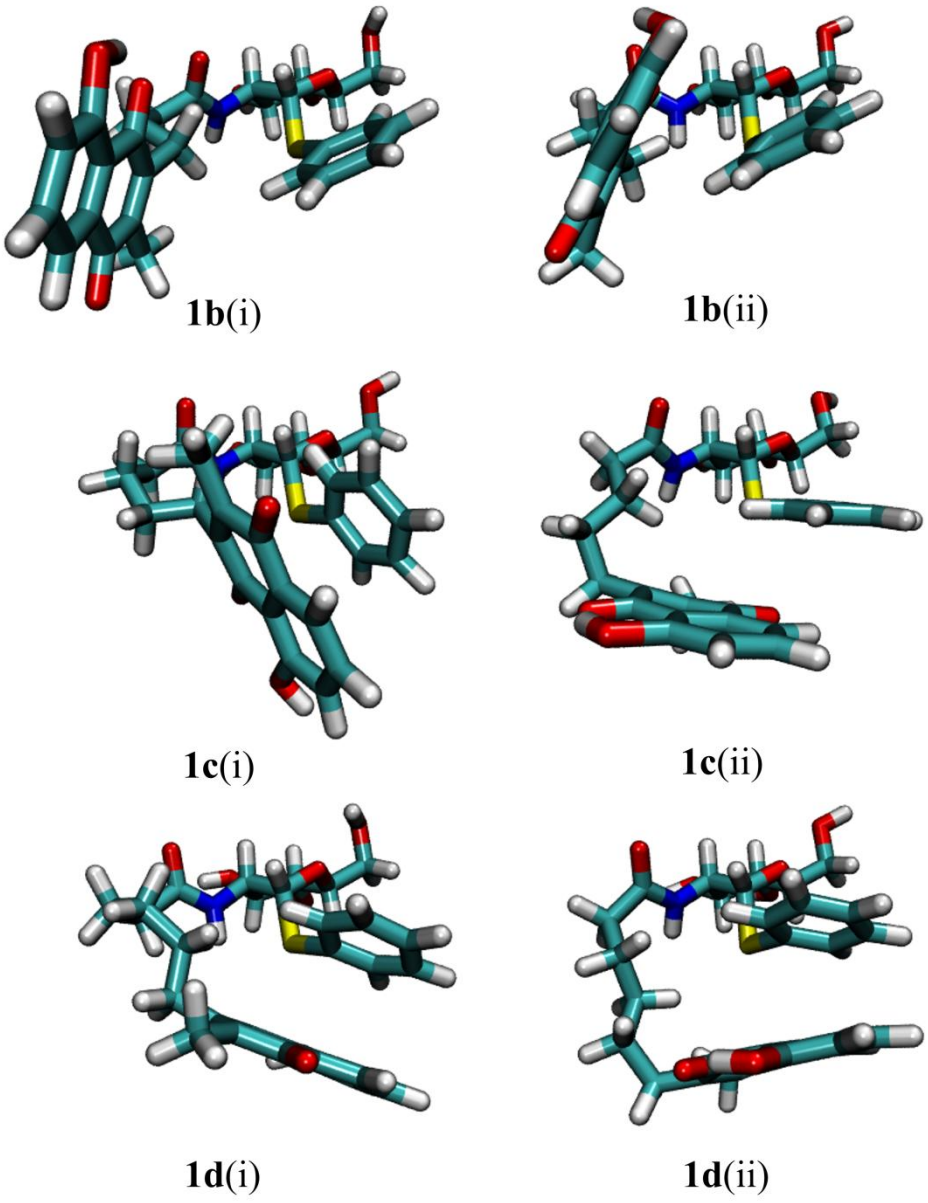


FIGURE 5-7 Cluster centre conformations of the largest clusters presented in Table 5-1.

TABLE 5-2 Clusters arising from PCA analysis of the inhibitor trajectories in solution. The distance between the COM of their phenylthio and plumbagin moieties was used as the principal component.

Simulation	Cluster	# members	std. dev.	d
1b(i)	1	21354	0.80	5.58
	2	15588	0.94	8.00
	3	13058	0.92	12.20
1b(ii)	1	14260	0.36	6.00
	2	13802	0.39	4.65
	3	13197	0.39	7.15
	4	5101	0.59	8.86
	5	2640	0.86	11.31
1c(i)	1	16495	0.17	3.83
	2	13072	0.18	4.38
	3	8590	0.22	5.06
	4	4264	0.28	6.87
	5	1863	0.56	7.90
	6	5716	0.26	5.93
1c(ii)	1	25424	0.36	4.19
	2	10845	0.52	5.69
	3	6953	0.57	7.88
	4	3486	0.67	9.82
	5	3292	0.80	12.41
1d(i)	1	9413	0.34	4.85
	2	24793	0.24	3.92
	3	4838	0.44	6.34
	4	7789	0.40	7.77
	5	3167	0.69	9.23
1d(ii)	1	27621	0.41	4.14
	2	5787	0.71	6.83
	3	5858	0.64	8.97
	4	7668	0.63	11.20
	5	3066	0.98	13.53



.

FIGURE 5-8 Cluster centre conformations of the largest clusters presented in Table 5-2.

5.3. ENZYME INHIBITOR COMPLEXES

5.3.1. DOCKING AND MOLECULAR DYNAMICS

Since a structure for **1a**, **1b**, **1c** or **1d** (Figure 5-9) co-crystallised with MshB has not been reported, *de novo* docking had to be carried out. The Sitemap [20] utility in Maestro was used to generate hydrophobic, hydrogen bond acceptor and hydrogen bond donor surfaces for the binding cavity (Figure 5-10). The natural substrate from previous docking studies is also shown. It is clear from Figure 5-10A that, if the binding orientation of the inhibitor glucose moieties is similar to that modelled for the natural substrate, the large hydrophobic cavity adjacent to the active site will be suitably located for binding the C-2 acylamido group. Within this cavity the carbonyl groups of Ser20 and Thr183, as well as the hydroxyl group of Thr183, form a favourable environment for hydrogen bond donating groups. The roof of the cavity is lined with a number of leucine groups (Leu259, Leu263, Leu265 and Leu19) which give rise to a hydrophobic region shown in yellow. Hydrogen bond accepting regions include the space around Lys237, Ser20 and His272.

In order to be consistent with the orientation of the co-crystallised BOG in the 1Q7T structure[3] [15], docking was carried out with the constraint that the oxygen at the inhibitor C-4 position must be hydrogen bonded to Arg68 and Asp95. The vdW radii and charges for Tyr142 were scaled since it lay across the cavity, which was not observed in 1Q7T chain B or any of the 1Q74 molecules. The docking results showed that the longer chains of **1c** and **1d** allowed their plumbagin hydroxyl group to approach within hydrogen bonding distance of the Thr183 hydroxyl moeity. The docking conformation of **1d** (Figure 5-10B), also shows the naphthoquinone oxygen at C-1’ close to the Ser20 hydroxyl group, and the methyl group at C-3′ in close proximity to the previously identified hydrophobic region. This orientation of the naphthoquinone, in comparison to a conformation in which the quinone is flipped by 180˚, appears complimentary to the MshB

[3] The assumption made here is that MshB binds specifically to the glucose unit of the inhibitor, the same moiety proposed to regulate the binding mode of the natural substrate (Chapter 4). There is evidence in the literature of the binding of the plumbagin orthohydroxycarbonyl functional group to metal ions (For an example of coordination of plumbagin to Cu(II) refer to [21]). Thus, an alternative binding mode may considered in which plumbagin coordinates to the MshB Zn^{2+}. However, there is no experimental data on the inhibitory activity of plumbagin, and due to the limited scope of the MSc dissertation this possibility was not further investigated.

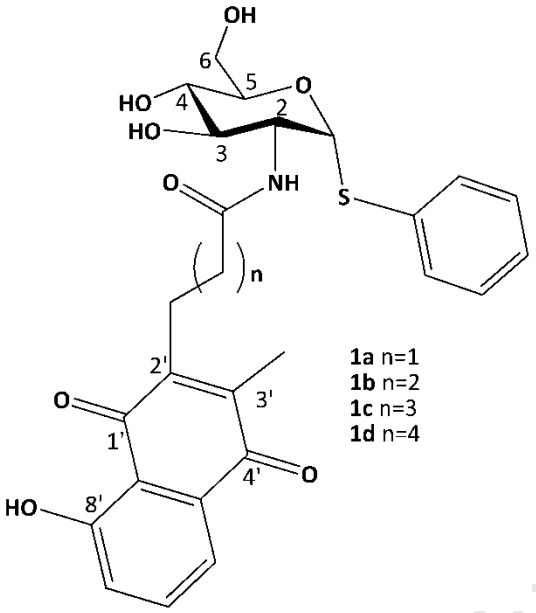


FIGURE 5-9 Diagram of the current inhibitor series with carbon atoms numbered.

binding pocket, and is preserved in the top ranked docked poses for **1a** and **1b**, even though they do not have long enough alkyl linkers to form hydrogen bonding with Thr183. The docking conformations with the best Glide score were solvated in a TIP3P water sphere, minimised and heated before they were subjected to 5ns stochastic boundary dynamics. In order to confirm that the aforementioned naphthoquinone orientation gave the best binding, the plumbagin group was flipped by 180˚ and these conformations were also subjected to MD simulations. IEs for all inhibitors were lowest for the inhibitors with the original naphthoquinone orientation.

5.3.2. TRAJECTORY ANALYSIS

The MD trajectories were analysed using a series of techniques. RMSD analysis was conducted to determine when the systems were at equilibrium. The time series for the RMSD of amino acids within 8Å of the inhibitors are given in Appendix C. In all systems the MshB active site is seen to be at equilibrium from 1ns, at which time the amino acid residues fluctuate about stable positions.

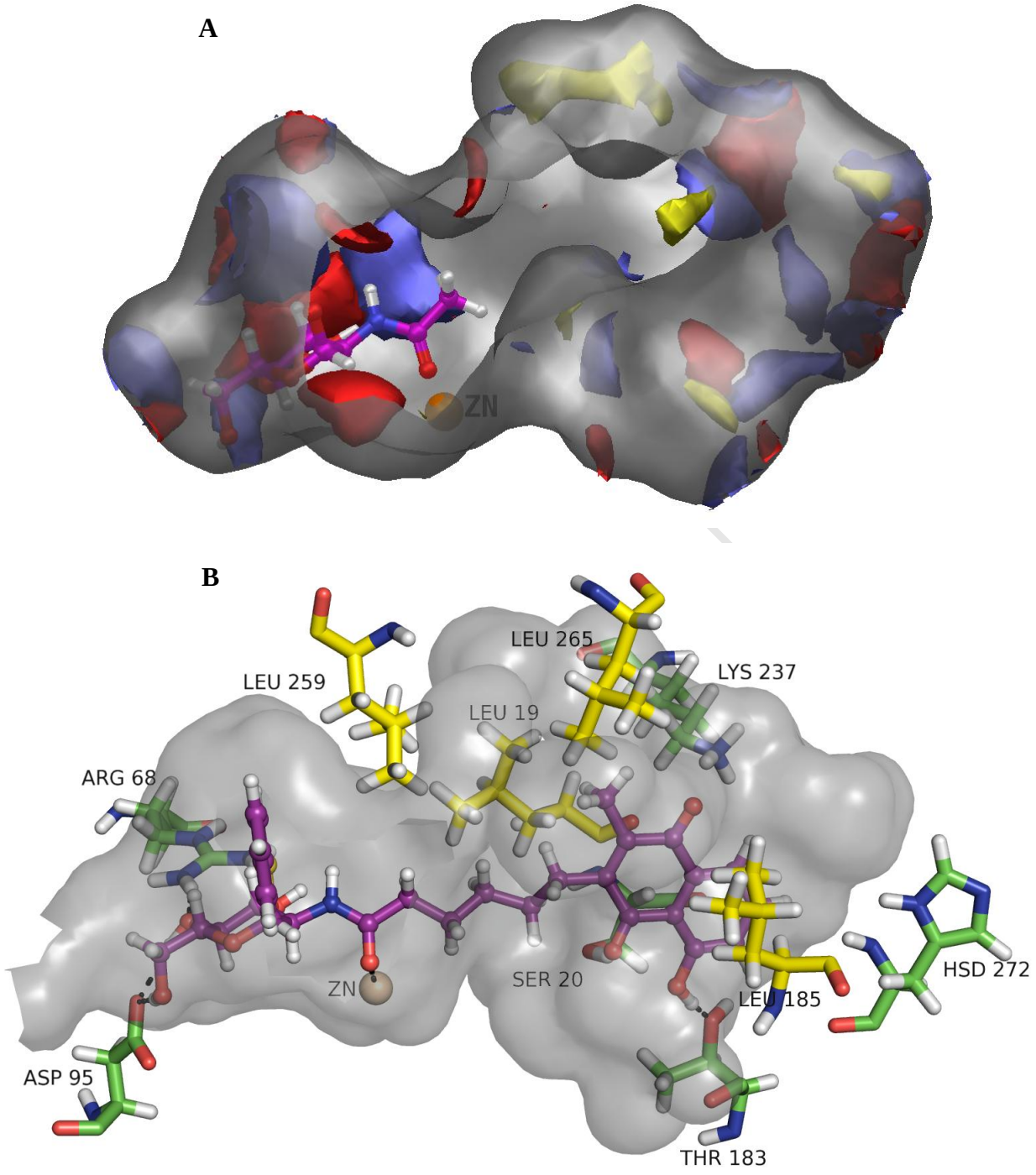


FIGURE 5-10 (A) Surfaces produced by the Sitemap utility in Maestro. Hydrophobic surface at an isovalue of -1.5kcal/mol is shown in yellow, and hydrogen bond donor and acceptor sites are shown at an isovalue of -10kcal/mol in red and purple respectively; (B) Docked orientation of 1d.

Analysis of inhibitor binding was conducted on the final 1ns of the 5ns trajectory. Key amino acids were identified by computing their average interaction energies (IEs) with the respective inhibitors. The residues have been grouped into amino acids forming the active site cavity (Table 5-3) and those lining the adjacent hydrophobic cavity (Table 5-4). Table 5-5 reports the sum of the energies presented in Table 5-3 and Table 5-4, and also gives the IEs calculated between the inhibitors and all protein residues, as well as between the inhibitors and all residues in the system. The time series of the IEs with all residues in the system are given in Figure 5-11. The van der Waals (vdW) and electrostatic components of the interactions were also delineated to determine the nature of interactions.

Hydrogen bonding was defined using an angle criterion of 120˚ and a distance criterion of 3.5Å. As in Chapter 4, hydrophobic interactions were assessed according to an interaction specific distance criterion. For each residue type a representative orientation with respect to the relevant inhibitor moiety was chosen from the coordinates of the MshB-**1d** complex after 5ns of equilibration. An IE scan was then conducted by moving the residue away from the inhibitor moiety in vacuum, and the optimal inter-residue distance was calculated to be at the energy minima. Contacts between non-polar inhibitor moieties and non-polar amino acids were then evaluated to be hydrophobic in nature if there were no waters between them, and the two groups were stable at or close to the optimal distance calculated above.

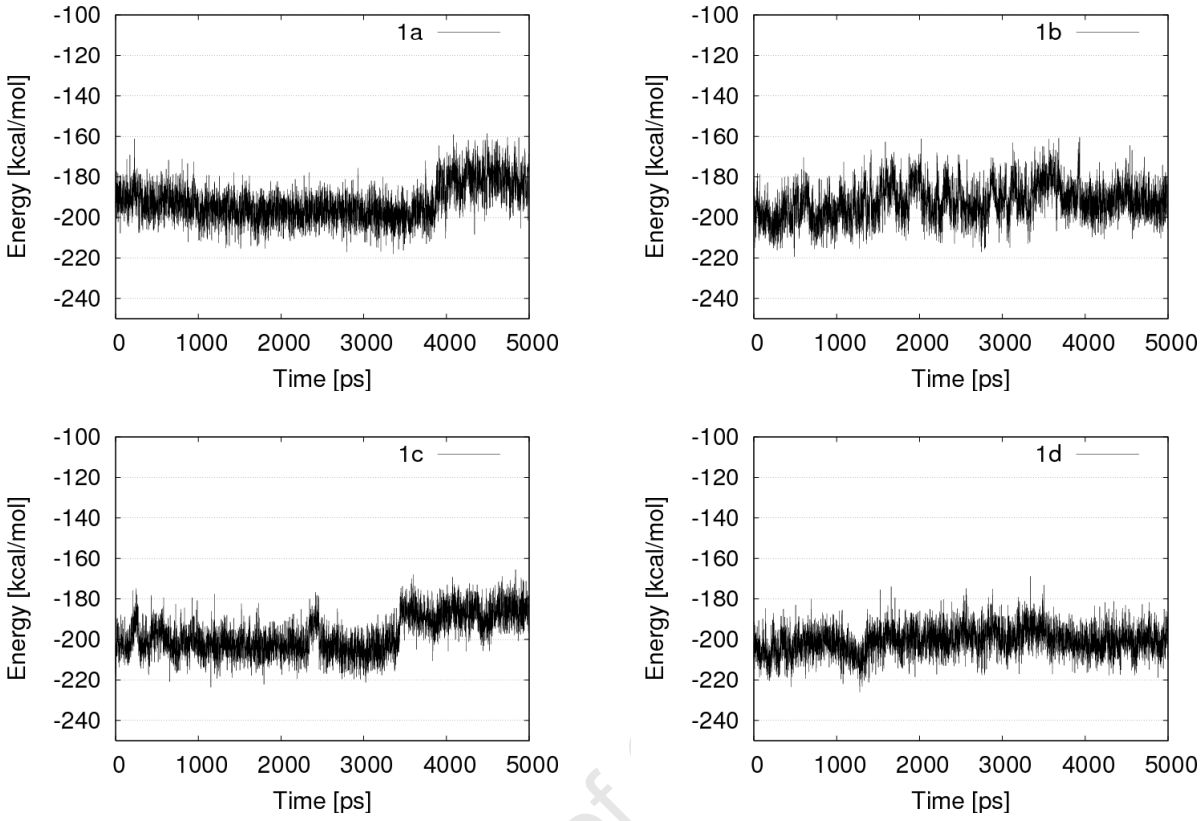


FIGURE 5-11 Time series for average IEs with all atoms.

TABLE 5-3 Average IEs for amino acid residues lining the ative site. All Energies are calculated in kcal/mol and only interactions with an energy less than -1kcal/mol or greater than 1 kcal/mol are reported. The sums of the interactions are also given (Total).

Residue	1a			1b			1c			1d		
	Total	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**
Asp15	**-7.47**	-1.59	-5.88	**-4.50**	-0.96	-3.54	**-2.64**	-1.11	-1.53	**-4.05**	-1.26	-2.79
Asp16	**15.77**	-1.48	17.25	**18.84**	-1.56	20.40	**16.63**	-1.58	18.21	**18.71**	-1.53	20.25
Glu44	**-0.98**	-0.27	-0.71	**-0.98**	-0.32	-0.66	**-0.99**	-0.29	-0.70	**-1.06**	-0.33	-0.73
Glu45	**-2.54**	-0.96	-1.57	**-2.00**	-1.05	-0.95	**-2.36**	-1.34	-1.02	**-3.06**	-1.27	-1.80
Gly46	**-2.79**	-1.18	-1.61	**-1.82**	-0.60	-1.22	**-2.33**	-0.92	-1.42	**-2.71**	-1.00	-1.72
Glu47	**-2.92**	-1.74	-1.18	**-4.22**	-2.25	-1.97	**-3.21**	-1.52	-1.69	**-4.74**	-2.89	-1.85
Arg68	**-16.77**	0.89	-17.66	**-14.65**	-0.27	-14.38	**-17.01**	0.92	-17.93	**-18.15**	0.54	-18.69
Asp95	**-22.14**	2.70	-24.83	**-24.20**	3.33	-27.53	**-21.85**	2.35	-24.20	**-23.03**	3.51	-26.54
Met98	**-0.18**	-0.52	0.35	**-1.25**	-1.45	0.20	**-1.08**	-1.19	0.10	**-3.61**	-3.24	-0.37
His144	**-11.11**	-3.15	-7.96	**-7.86**	-2.25	-5.61	**-11.79**	-2.68	-9.11	**-1.94**	-0.87	-1.07
His147	**5.67**	-1.00	6.67	**2.27**	-1.35	3.62	**3.17**	-2.08	5.25	**4.51**	-1.49	6.00
Glu213	**-0.47**	-0.02	-0.45	**-2.41**	-0.32	-2.09	**-3.56**	-0.46	-3.09	**-1.34**	-0.07	-1.27
Gln247	**-1.21**	-0.95	-0.26	**-1.38**	-0.87	-0.51	**-1.40**	-0.55	-0.85	**-1.17**	-0.96	-0.21
Asn261	**0.04**	-0.25	0.29	**0.02**	-0.20	0.23	**-4.11**	-3.09	-1.02	**-0.11**	-0.42	0.30
Zn	**-91.55**	6.54	-98.09	**-97.49**	5.81	-103.30	**-91.63**	6.36	-97.99	**-98.18**	6.23	-104.42
Total	-138.64	-2.98	-135.66	-141.63	-4.32	-137.32	-144.16	-7.17	-136.99	-139.93	-5.02	-134.91

TABLE 5-4 **Average IEs for amino acid residues lining the cavity adjacent to the putative active site. All Energies are calculated in kcal/mol and only interactions with an energy less than -1kcal/mol or greater than 1 kcal/mol are reported. The sums of the interactions are also given (Total).**

Residue	1a			1b			1c			1d		
	Total	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**
Leu19	**-4.87**	-4.85	-0.02	**-2.03**	-1.99	-0.04	**-2.54**	-3.05	0.51	**-2.09**	-3.52	1.43
Ser20	**-2.55**	-0.88	-1.67	**-0.50**	-0.45	-0.06	**-1.21**	-1.81	0.60	**-3.56**	-3.78	0.22
Tyr136	**-0.98**	-0.21	-0.77	**-0.35**	-0.77	0.43	**-1.57**	-1.65	0.08	**-0.87**	-1.46	0.59
Asp137	**0.32**	-0.01	0.32	**-4.38**	-0.12	-4.26	**-1.16**	-0.19	-0.98	**-4.80**	-0.32	-4.48
Gly140	**0.39**	-0.06	0.45	**-3.06**	0.04	-3.09	**1.64**	-0.47	2.11	**1.30**	-0.40	1.70
Tyr142	**-6.02**	-4.97	-1.05	**-4.30**	-3.89	-0.41	**-4.52**	-3.85	-0.67	**-3.97**	-2.59	-1.38
Thr183	**-0.06**	-0.07	0.00	**-0.78**	-0.12	-0.66	**-1.05**	-0.63	-0.42	**-4.04**	-1.38	-2.66
Val184	**-0.15**	-0.04	-0.11	**0.14**	-0.10	0.24	**-0.36**	-0.39	0.04	**-1.83**	-1.47	-0.36
Leu185	**-1.26**	-1.28	0.03	**-0.60**	-0.67	0.06	**-3.46**	-3.38	-0.07	**-4.42**	-4.02	-0.40
Lys237	**1.34**	-0.09	1.42	**-0.06**	-0.01	-0.05	**-1.78**	-0.06	-1.72	**-2.59**	-0.17	-2.41
Leu259	**-1.45**	-1.76	0.31	**-1.16**	-1.65	0.49	**-0.93**	-1.53	0.60	**-0.52**	-1.02	0.50
Leu265	**-1.04**	-1.20	0.16	**-1.07**	-0.88	-0.19	**-0.86**	-0.92	0.06	**-0.80**	-0.78	-0.03
Leu268	**-0.44**	-0.70	0.25	**-0.20**	-0.18	-0.02	**-1.40**	-1.04	-0.36	**-1.44**	-1.06	-0.38
Total	**-16.79**	-16.10	-0.69	**-18.33**	-10.78	-7.55	**-19.20**	-18.98	-0.22	**-29.63**	-21.97	-7.65

TABLE 5-5 **The sum of total IEs reported in Table 5-3 and Table 5-4 (Total) is compared to IEs between MshB and the inhibitors (MshB). IEs with all atoms in the system are also reported (All). All Energies are calculated in kcal/mol and K_i values are given in brackets.**

Residue	**1a** (167 ± 15μM)			**1b** (94 ± 11μM)			**1c**			**1d** (16.8 ± 1.9μM)		
	Total	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**
Totals	**-155.43**	-19.08	-136.34	**-159.97**	-15.10	-144.87	**-163.36**	-26.15	-137.21	**-169.56**	-26.99	-142.56
MshB	**-155.74**	-27.84	-127.89	**-157.43**	-23.43	-134	**-162.95**	-36.44	-126.52	**-166.68**	-37.07	-129.61
All	**-181.23**	-38.52	-142.71	**-192.16**	-39.4	-152.77	**-186.65**	-47.29	-139.35	**-201.27**	-48.57	-152.7

BINDING MODE

Figure 5-9 gives the structures of the inhibitor series with relevant atom labels. Their binding is summarised in Figure 5-12 to Figure 5-18, which give a snapshot of the binding modes after 5ns stochastic boundary dynamics. All inhibitors displayed a glucose binding orientation resembling that of the natural substrate model (Chapter 4). Hydrogen bonds and metal interactions dominate inhibitor binding. The hydroxyl groups at C-3 and C-4 show hydrogen bond geometries with Arg68 $N^{\eta 1}$ and $N^{\eta 2}$ respectively. Hydrogen bonding is also observed between the C-6 hydroxyl hydrogen and Asp95 $O^{\delta 1}$, and between the C-4 hydroxyl hydrogen and Asp95 $O^{\delta 2}$. Together with the carbonyl-zinc interaction, these electrostatic interactions are the most favourable ones (Tables 3, 4 and 5). With the above glucose orientation the phenylthio group is directed towards the entrance of the active site, and the C-2 acylamido side chain extends from the zinc coordinated carbonyl into the adjacent hydrophobic cavity.

Interactions with MshB residues that play a key role in binding **1a**, **1b**, **1c** and **1d** are described in more detail below. For the characterisation of hydrophobic interactions, the trajectory was inspected to see that water molecules did not lie between the inhibitor and the respective residues. The inter-residue distances were also calculated and are given as time series in Figure D-1 to Figure D-16 in Appendix D. Average energies are reported for the entire inhibitor unless specified.

BINDING OF 1A

The hydrogen bonding and electrostatic interactions are summarised in Figure 5-12, and the hydrophobic interactions are illustrated in Figure 5-14. As discussed above interaction of the 2-acetamido glucose with MshB is dominated by hydrogen bonding and electrostatic interactions. However, binding is also mediated by a large number of hydrophobic interactions. The phenylthio and linker groups form a hydrophobic pocket with Leu259 (vdW −1.76kcal/mol, total −1.45kcal/mol). The time series of the COM separation between the Glu47 non-polar atoms and the phenylthio atoms (Figure D-1, Appendix D), shows that the two methylene groups from the glutamate are within hydrophobic contact distance of **1a**. Glu45 and Glu46 are in close proximity

to the non-polar hydrogens at C-5 and C-6, and have favourable vdW and electrostatic interactions with the inhibitor (vdW –2.54kcal/mol, total –2.79kcal/mol). His144 makes an important contribution to the IE with strong vdW and electrostatic interactions (total –11.11kcal/mol). However, hydrogen bond analysis showed His144 $N^{\varepsilon 2}$ fulfilled hydrogen bond criteria with the C-6 hydroxyl hydrogen periodically over the trajectory.

The plumbagin moiety sits in a cavity adjacent to the active site with its aromatic atoms lying against a hydrophobic surface composed of Tyr142, Leu185, Leu265 and Leu268 (Figure 5-14). Tyr142 (vdW –4.97kcal/mol, total –6.02kcal/mol) and Leu19, which is located on the opposite face of plumbagin (vdW –4.85, total –4.87), make a significant enthalpic contribution to binding. Hydrogen bond analysis showed that the plumbagin hydroxyl forms temporary hydrogens bonds with Ser20 (total –2.55kcal/mol) and a water molecule bridging Asp16. Non-specific hydrogen bonding was also detected between the naphthoquinone oxygens and water along the final 1ns of the trajectory.

BINDING OF 1B

The hydrogen bonding and electrostatic interactions are summarised in Figure 5-13, and the hydrophobic interactions are illustrated in Figure 5-14. In **1b** the phenylthio moiety forms favourable interactions with Met98 (vdW -1.45kcal/mol, total, -1.25kcal/mol). A time series shows that their relative positions fluctuate about the optimal distance calculated from energy scans in vacuum (Figure D-5, Appendix D). As in **1a**, Glu45, Gly46 and Gly47 show favourable interactions with the glucose and phenylthio moieties with Gly47 providing a significant enthalpic contribution to binding (total -4.22kcal/mol). In addition the phenylthio group displays an electrostatic interaction of -2.09kcal/mol (total -2.41kcal/mol) with Glu213.

The longer linker of **1b** extends the plumbagin group further into the cavity than in **1a**. As a result Leu265 and Tyr142 (vdW –3.89kcal/mol, total –4.30kcal/mol), together with Leu19 (vdW –1.99kcal/mol, total –2.03kcal/mol), now surround the alkyl linker. While Tyr142 is shown in close proximity to the methyl group at C-3′, analysis of the MD trajectory revealed that over the last 1ns, the non-polar atoms of this residue also approach the vdW contact distance of the

plumbagin aromatic atoms (Figure D-7 and Figure D-8, Appendix D). Although not illustrated in the 2D representation of the last time step (Figure 5-15), hydrogen bonding analysis showed that the plumbagin hydroxyl oxygen is hydrogen bonded to His147 $N^{\delta 1}$. However, the favourable hydrogen bonding is offset by a clash in partial charges to give a net positive IE. Gly140 makes an important enthalpic contribution to binding (total −3.06kcal/mol). Hydrogen bonding analysis showed that the C-8′ hydroxyl hydrogen alternatively forms and breaks hydrogen bonds with Gly140 O, and a water bridging the protein. The hydroxyl group at C-8 also forms favourable electrostatic interactions with Asp137 (ELEC −4.26kcal/mol, total −4.28kcal/mol). The naphthoquinone oxygens at C-1 and C-4 form hydrogen bonds with the surrounding water molecules over the last 1ns of the trajectory

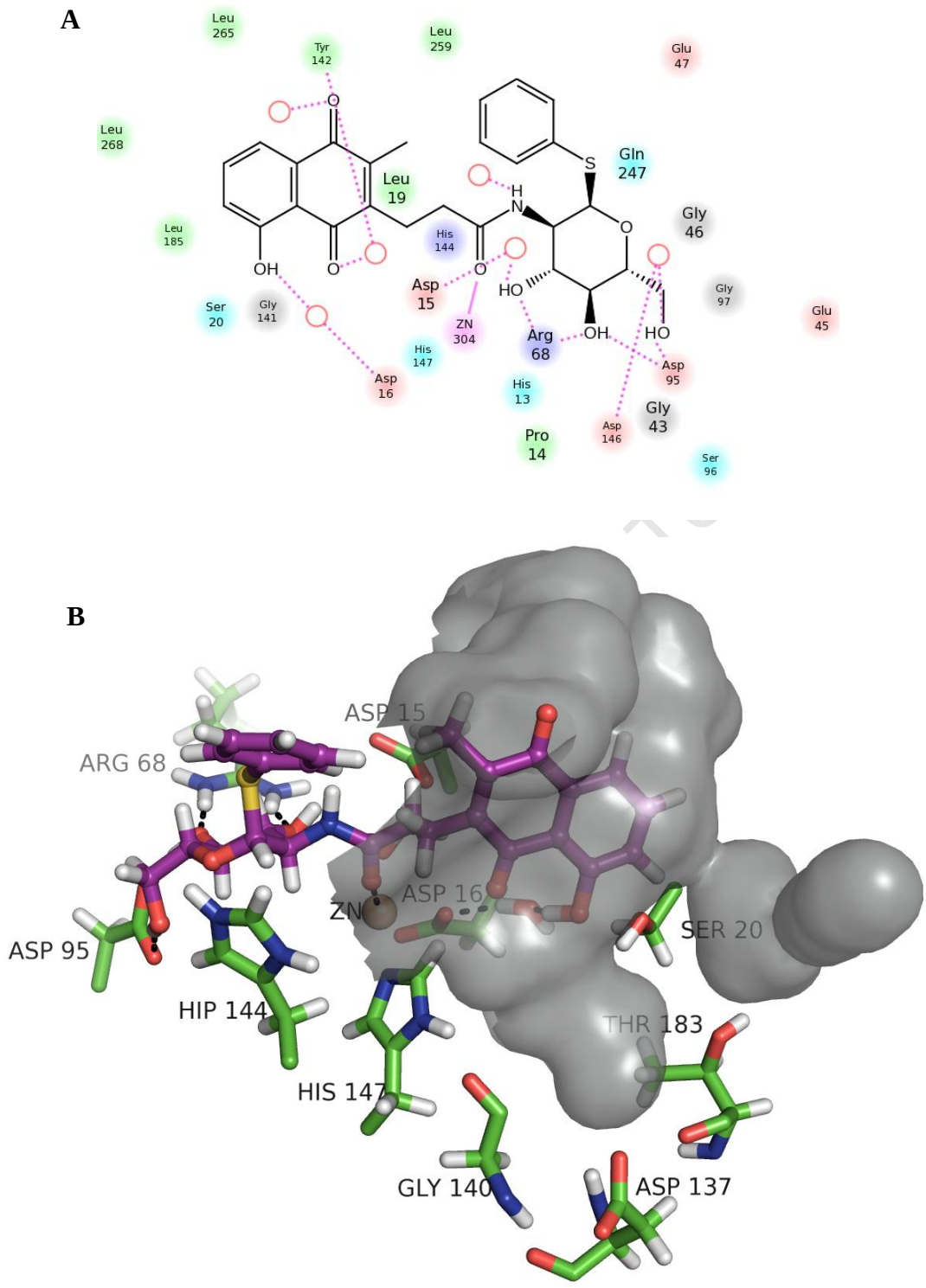


FIGURE 5-12 **(A)A 2D snapshot summarising the binding of 1a with MshB. Hydrophobic atoms are shown in green, polar amino acids are shown in cyan and charged amino acids are shown in red. The size of the residue indicates its relative position. (B) A 3D snapshot showing residues which have important electrostatic and hydrogen bond interactions with the inhibitor series. The cavity adjacent to the putative active site is shown as a gray surface.**

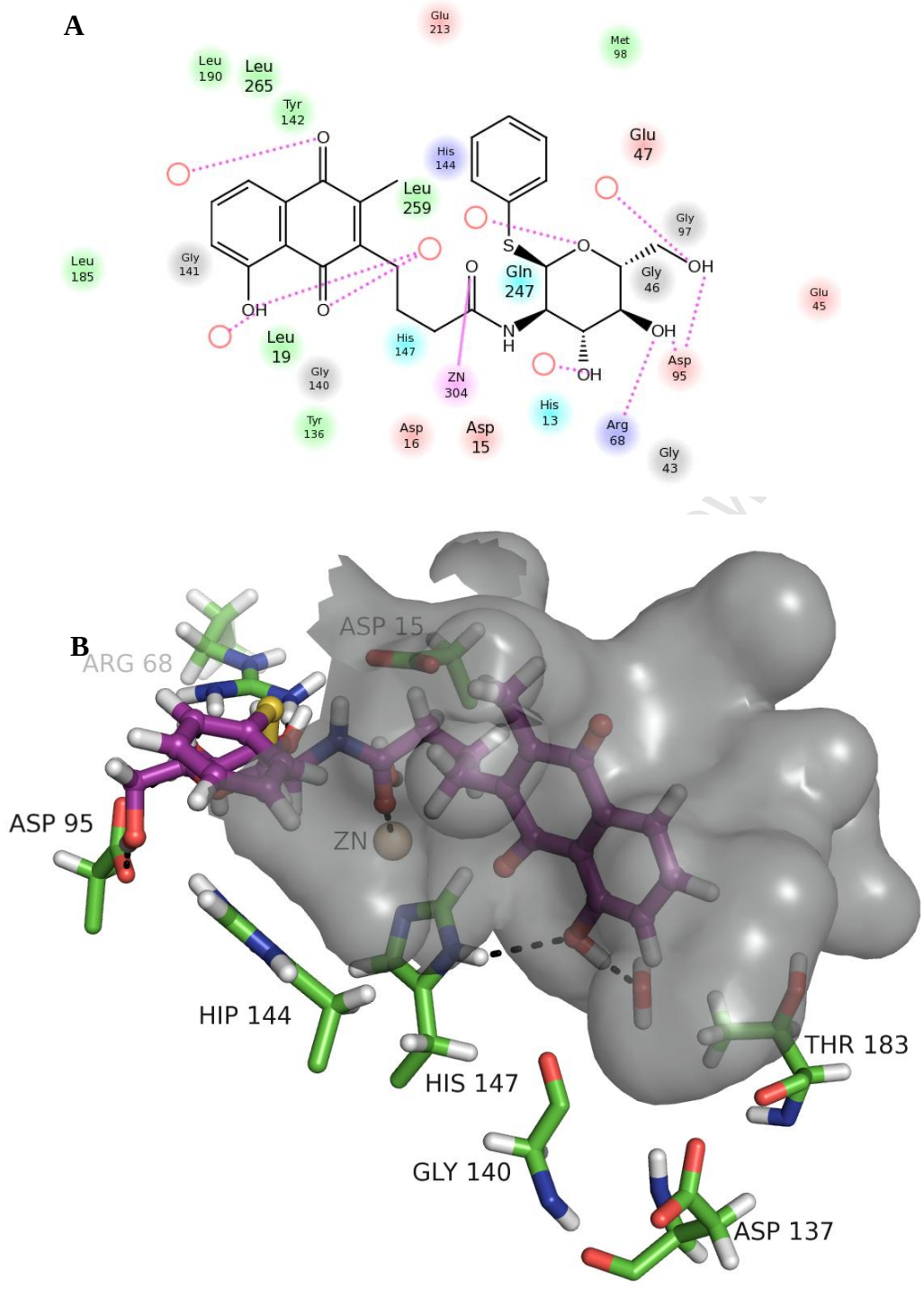


FIGURE 5-13 **(A)A 2D snapshot summarising the binding of 1b with MshB. Hydrophobic atoms are shown in green, polar amino acids are shown in cyan and charged amino acids are shown in red. The size of the residue indicates its relative position. (B) A 3D snapshot showing residues which have important electrostatic and hydrogen bond interactions with the inhibitor series. The cavity adjacent to the putative active site is shown as a gray surface.**

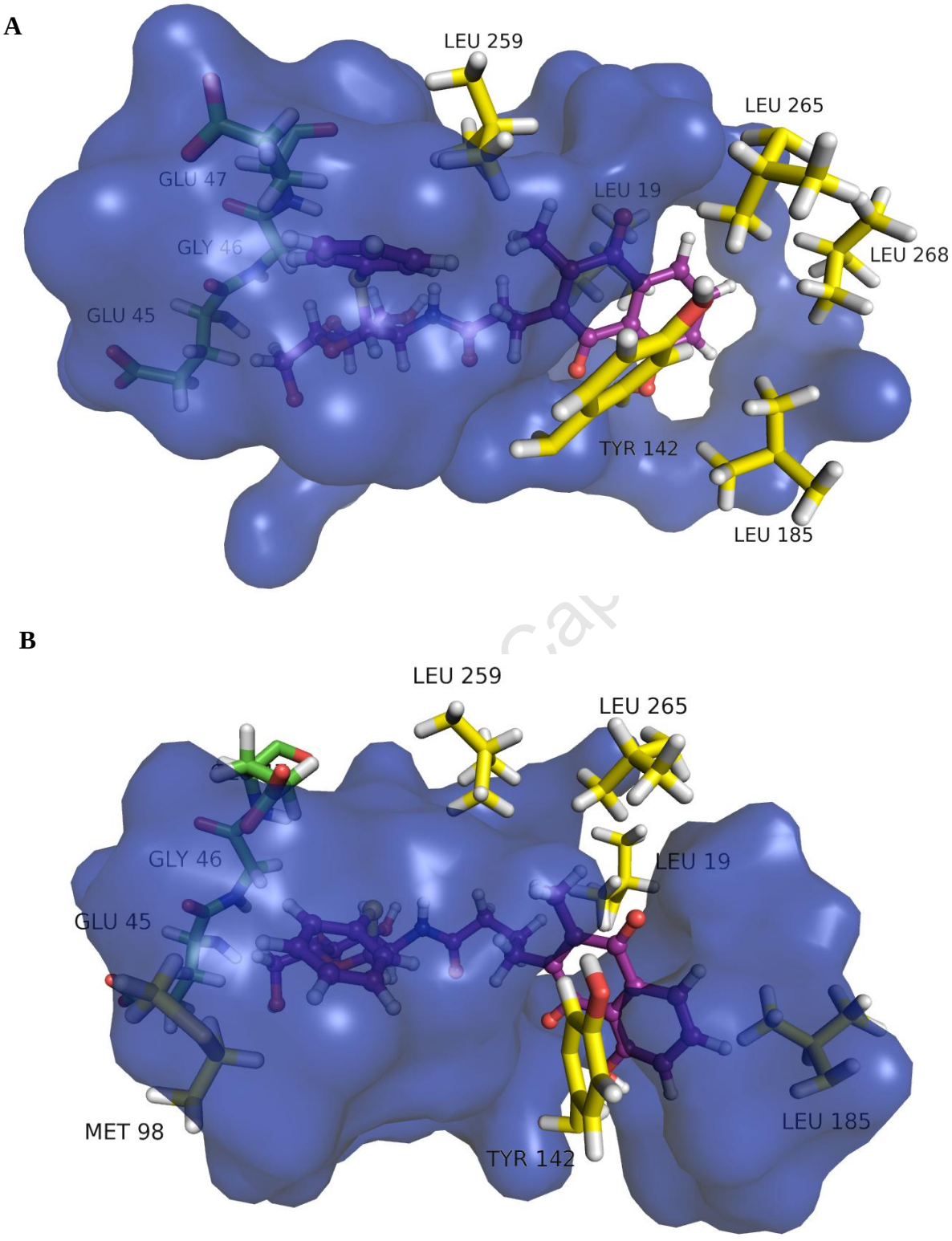


FIGURE 5-14 **3D snapshot showing the hydrophobic interactions of (A)1a and (B) 1b. Carbon atoms of the hydrophobic residues are shown in yellow and the water surrounding 1a is shown as a blue surface.**

BINDING OF 1C

The hydrogen bonding and electrostatic interactions are summarised in Figure 5-15, and the hydrophobic interactions are illustrated in Figure 5-16. The interactions mediating **1c** binding are very similar to the MshB-**1b** complex. Met98 (vdW -1.19kcal/mol, total -1.08kcal/mol), His144 (total −11.79kcal/mol), Glu45 (total -2.36kcal/mol), Gly46 (total −2.33kcal/mol), Glu47 (vdW −1.34kcal/mol, total −3.21kcal/mol) and Glu213 (total −3.56kcal/mol) are important in binding the phenyl thioglycoside moiety (Figure 5-16). Asn261 shows favourable vdW interactions with the inhibitor (vdW −3.09kcal/mol, total −4.11kcal/mol), and together with Glu47, makes a hydrophobic pocket with the phenylthio group of the inhibitor. The distance between **1c** and the non-polar atoms of Glu47 and Asn261 fluctuate around a steady value close to the optimal inter-residue distance in vacuum (Figure D-9, Appendix D). Hydrogen bond analysis showed unstable hydrogen bonding between His144 and the C-6′ hydroxyl oxygen (total −11.79kcal/mol) over the final 1ns of the trajectory.

Leu19 (vdW -3.05kcal/mol, total -2.54 kcal/mol), Leu259, Leu263, Leu265 and Leu268 form a hydrophobic pocket with the linker and plumbagin methyl groups. Extension of plumbagin into the cavity allows for favourable hydrophobic interactions with Tyr142 (vdW −3.85kcal/mol, total −4.52kcal/mol) and Leu185 (vdW -3.38kcal/mol, total −3.46kcal/mol). On the opposite plumbagin face, Tyr136 forms hydrophobic contacts with the aromatic atoms (vdW −1.65kcal/mol, −1.57kcal/mol) and the positively charged Lys237 residue has an electrostatic IE of -1.72kcal/mol (total −1.78kcal/mol). The hydroxyl oxygen at C-8′ shows hydrogen bond geometry with His147 $H^{\delta 1}$, but again clashes in the atomic charges result in a net positive IE. Analysis of the MD trajectory showed that the C-8′ hydroxyl hydrogen is initially hydrogen bonded to a water molecule bridging Thr183 and Asp137. However, the hydrogen bond breaks and is replaced by an intramolecular hydrogen bond to the quinone oxygen at C-1′. The break of the aforementioned bonding results in an increase in IE with MshB (Figure D-17, Appendix D).

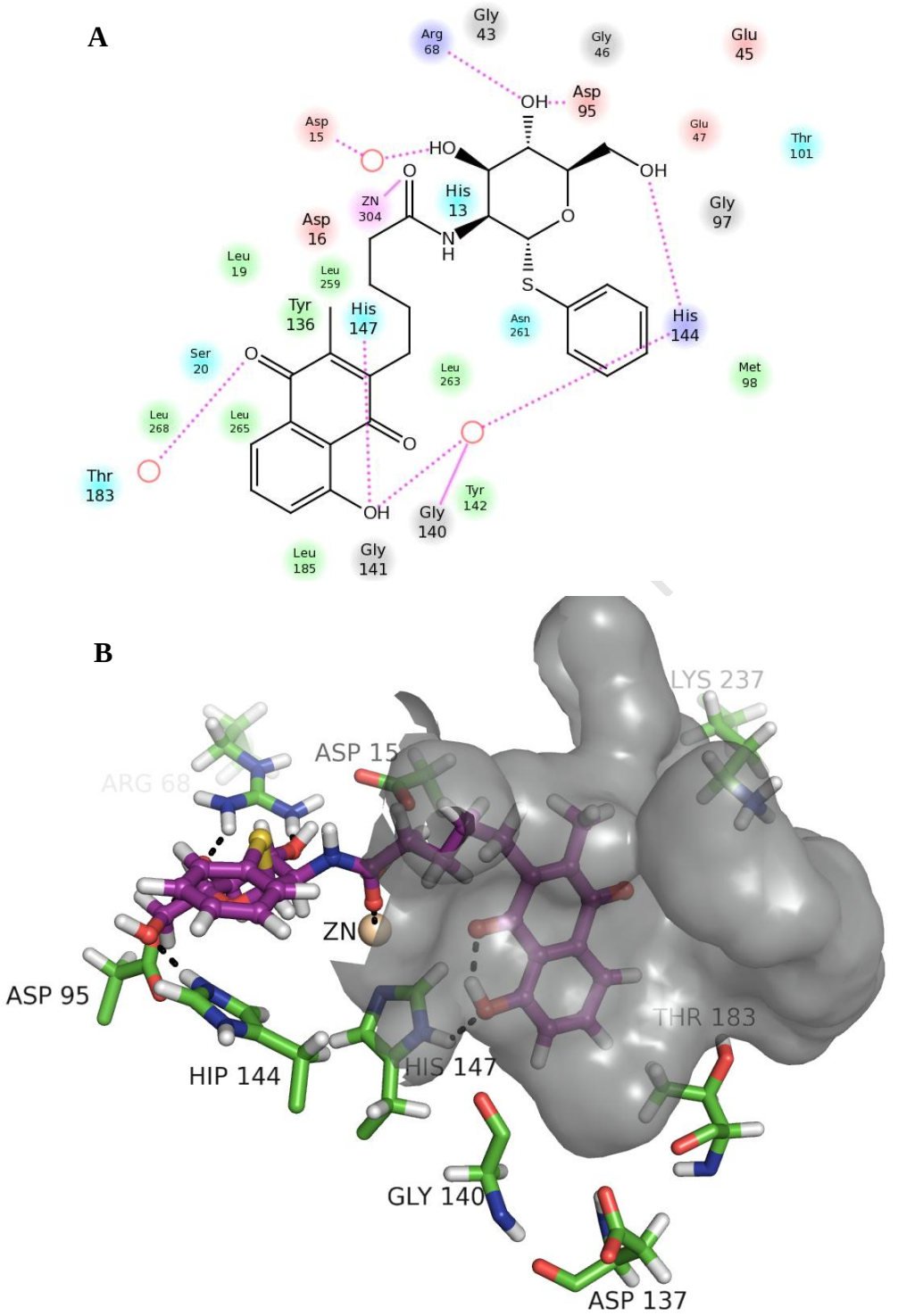


FIGURE 5-15 **(A) A 2D snapshot summarising the binding of 1c with MshB. Hydrophobic atoms are shown in green, polar amino acids are shown in cyan and charged amino acids are shown in red. The size of the residue indicates its relative position. (B) A 3D snapshot showing residues which have important electrostatic and hydrogen bond interactions with the inhibitor series. The cavity adjacent to the putative active site is shown as a gray surface.**

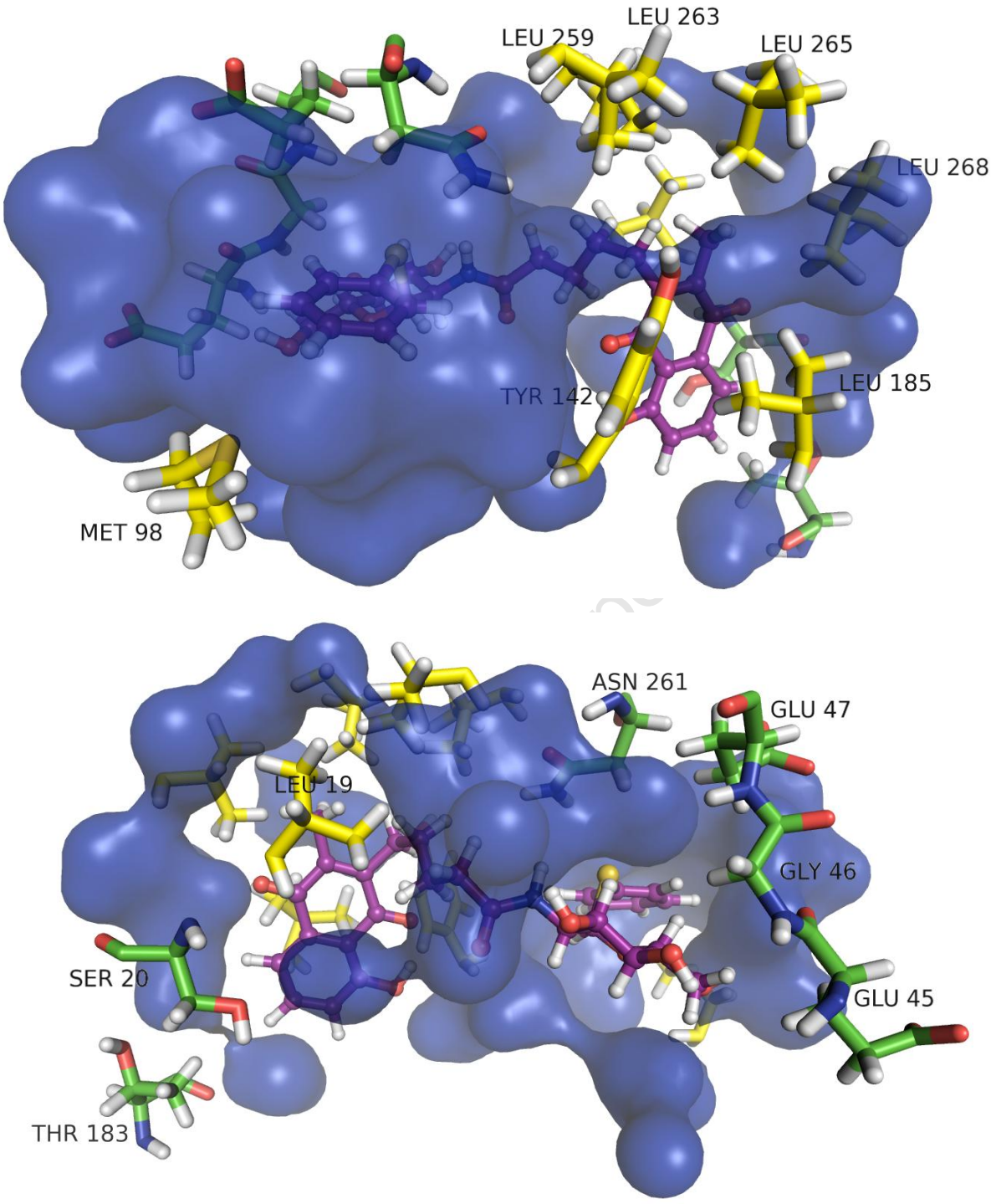


FIGURE 5-16 3D snapshots showing the hydrophobic interactions of 1c. Carbon atoms of the hydrophobic residues are shown in yellow and the water surrounding 1c is shown as a blue surface.

BINDING OF 1D

The hydrogen bonding and electrostatic interactions are summarised in Figure 5-17, and the hydrophobic interactions are illustrated in Figure 5-18. Met98 lies directly below the phenylthio group and has a strong IE with **1d** (vdW –3.61kcal/mol). The tripeptide Glu45 (–3.06kcal/mol), Glu46 (–2.71kcal/mol) and Glu47 (vdW –2.89kcal/mol, total –4.74kcal/mol) make significant enthalpic contributions to binding in a manner similar to that already described for **1a**, **1b** and **1c**.

Leu265, Leu268 and Leu269 make hydrophobic contacts with the C-3′ methyl and/or linker groups. Tyr142 (vdW –2.59kcal/mol, total –3.97kcal/mol) and Leu185 (vdW –4.02kcal/mol, total –4.42kcal/mol) form a complementary surface for the plumbagin and linker moieties. At the opposite plumbagin face, Tyr136 forms hydrophobic contacts with the aromatic atoms of the plumbagin moiety. Ser20 makes favourable vdW and electrostatic interactions with the plumbagin (total –3.56kcal/mol), and the positively charged Lys237 interacts electrostatically with **1b** (ELEC –2.41kcal/mol, total –2.59kcal/mol). Finally, the 5-membered alkyl linker extends the plumbagin sufficiently so that it forms a stable hydrogen bond with a water molecule bridging Thr183 and Asp 137. An IE of –4.67kcal/mol was computed for the interaction between the C-8′ hydroxyl group and the bridging water molecule

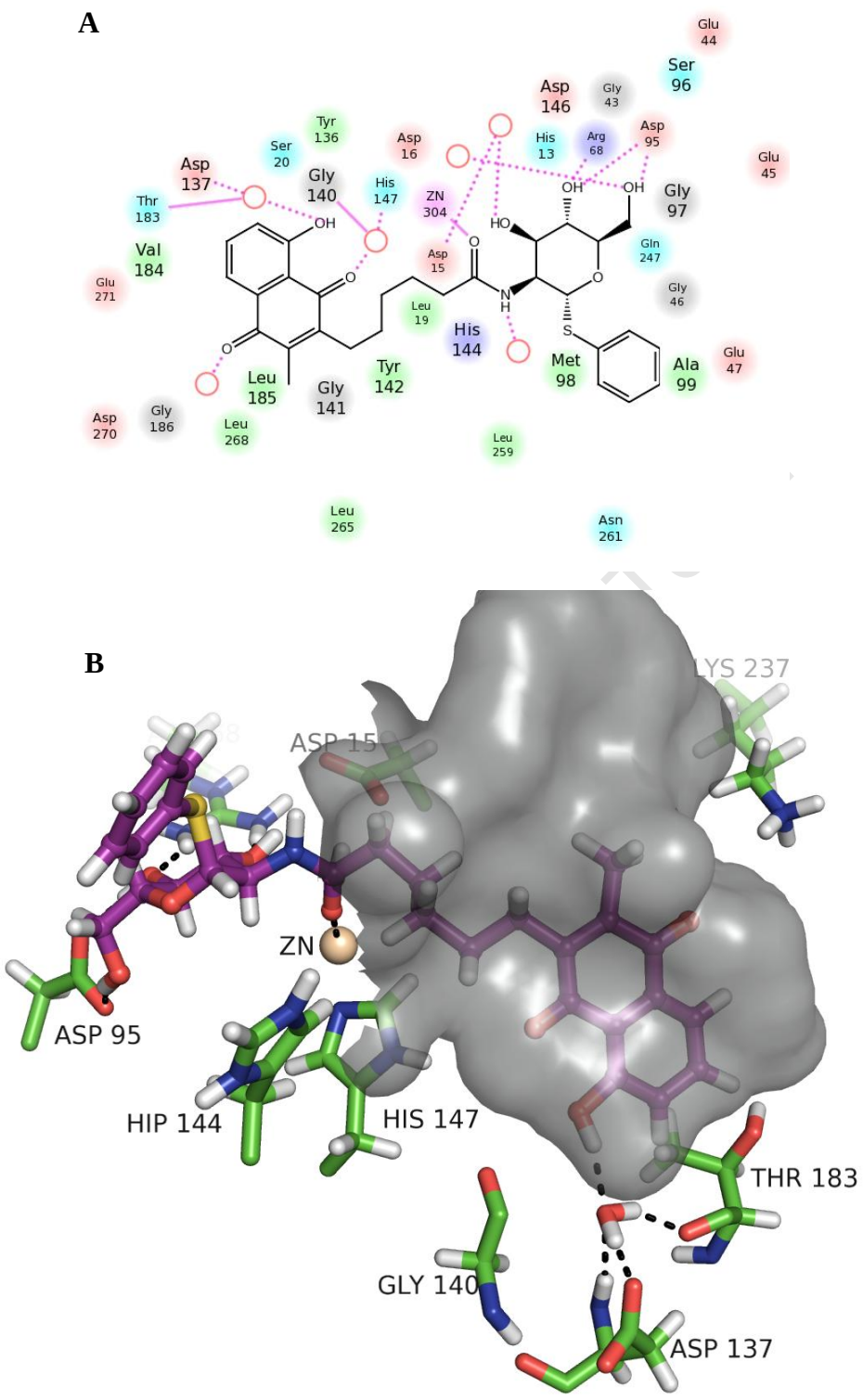


FIGURE 5-17 **(A) A 2D snapshot summarising the binding of 1b with MshB. Hydrophobic atoms are shown in green, polar amino acids are shown in cyan and charged amino acids are shown in red. The size of the residue indicates its relative position. (B) A 3D snapshot showing residues which have important electrostatic and hydrogen bond interactions with the inhibitor series. The cavity adjacent to the putative active site is shown as a gray surface**

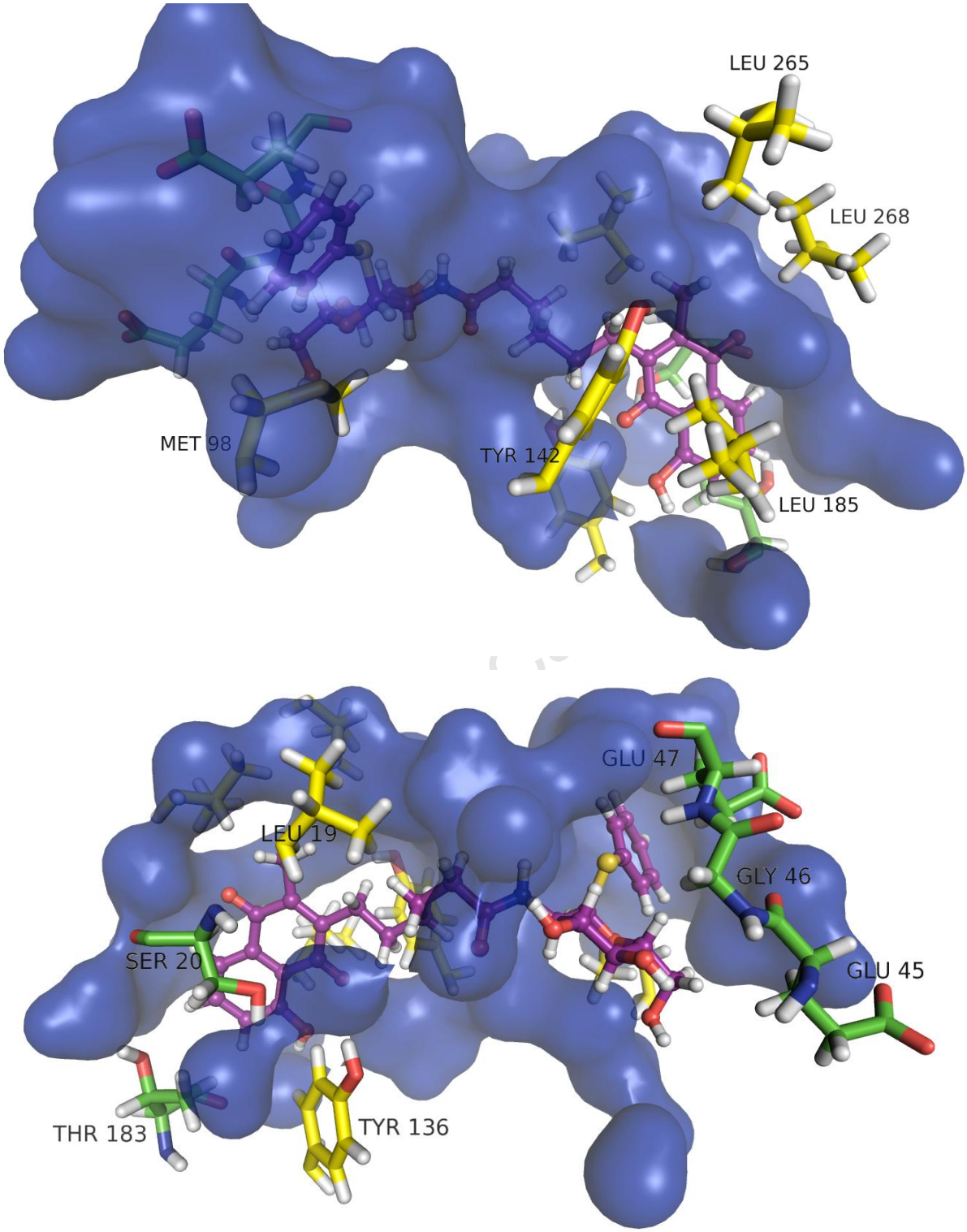


FIGURE 5-18 3D snapshots showing the hydrophobic interactions of 1d. Carbon atoms of the hydrophobic residues are shown in yellow and the water surrounding 1d is shown as a blue surface.

COMPARISON OF BINDING

It may be expected that the phenyl thioglycoside moiety displays similar IEs along the inhibitor series, since an identical moiety is binding specifically to MshB in each instance. However, the IEs between the phenyl thioglycoside and MshB (Table 5-6) showed a 9.9kcal/mol difference for **1b** and **1c.** This difference can be explained in part by binding variation with Met98 and His144 described above. The binding of Met98 is particularly flexible. **1d** displays the most favourable interactions with Me98, which is positioned directly under the phenylthio group of the inhibitor, thereby optimising hydrophobic surface contact with the aromatic ring. The IE is weaker in the **1b** and **1c** complexes, both of which have Met98 orientated against the edge of the phenylthio group. In the MshB-**1a** complex**,** Met98 is far from the phenylthio group and the IE is negligible. In the improved interaction geometry of the **1d** complex, the methionine occupies the space inhabited by His144 in **1a**, **1c** and **1d** (Figure D-18, Appendix D). As a result His144 is located further from the phenylthio group with a concomitant weakened interaction when compared to the shorter linked inhibitors. Hydrogen bonding analysis detected hydrogen bonds between the C-6 hydroxyl oxygen and His144 for **1a** and **1c**, which have the strongest IEs with the histidine. Further significant variation in IE is observed in the glucose binding residues Asn261 (Figure D-18, Appendix D), Arg68 and Asp95. Zn^{2+} displays improved interaction with **1c** and **1d** because of electrostatic interactions with the plumbagin (−7.86kcal/mol for **1c** and −2.57kcal/mol for **1d**). The variation in phenyl thioglycoside binding suggests that the active site formed by the loops between secondary structures (Section 1.2.1) is flexible.

The variation in phenyl thioglycoside binding makes it difficult to establish whether increasing the alkyl spacer improves binding of the linker and plumbagin. The unnormalised IE of the linker with MshB increases with the number of methylene groups. The plumbagin moiety on the other hand does not follow the experimentally observed trend. **1d** and **1b** have the lowest IEs followed by **1a** and then **1b** (Table 5-6). The longer linker does improve the IE with certain amino acids. Closer proximity of plumbagin to Leu185 improves the IE of their hydrophobic interaction. Extension of the naphthoquinone group also improves electrostatic interactions with Thr183 and Lys237. A comparison of the distance and energy time series of **1a**-**1d** with these residues can be found in Figure D-19 (Appendix D). In **1d** the longer linker also facilitates hydrogen bonding between the C-8′ hydroxyl group and a water molecule that bridges Thr183 and Asp137. This

TABLE 5-6 **Average IEs for inhibitor moieties with MshB atoms. Energies are all reported in kcal/mol. The sum of the IEs are given (Total) along with the IE calculated for MshB with the entire inhibitor (MshB).**

Inhibitor Moeity	1a			1b			1c			1d		
	Total	**VDW**	**Elec**	**Total**	**VDW**	**Elec**	**Total**	**VDW**	**Elec**	**Total**	**VDW**	**Elec**
Amidoglucose	**-122.13**	-2.75	-119.37	**-113.49**	-2.04	-111.45	**-118.82**	-2.46	-116.36	**-115.36**	-1.45	-113.91
Phenylthio	**-8.03**	-5.90	-2.13	**-9.27**	-6.08	-3.19	**-13.89**	-8.31	-5.59	**-9.24**	-7.83	-1.42
Linker	**-4.92**	-3.00	-1.92	**-7.60**	-3.73	-3.86	**-9.15**	-5.78	-3.37	**-9.47**	-6.50	-2.97
Plumbagin	**-20.66**	-16.19	-4.47	**-27.07**	-11.58	-15.49	**-21.09**	-19.89	-1.20	**-32.61**	-21.30	-11.31
Total	**-155.74**	-27.84	-127.89	**-157.43**	-23.43	-134.00	**-162.95**	-36.44	-126.52	**-166.68**	-37.07	-129.61
MshB	**-155.74**	-27.84	-127.89	**-157.43**	-23.43	-134.00	**-162.95**	-36.44	-126.52	**-166.68**	-37.07	-129.61

TABLE 5-7 **Average IEs for inhibitor moieties with all atoms. Energies are all reported in kcal/mol. The sum of the IEs are given (Total) along with the IE calculated for inhibitors with all the remaining atoms in the system (All).**

Inhibitor Moeity	1a			1b			1c			1d		
	Total	**VDW**	**Elec**	**Total**	**VDW**	**Elec**	**Total**	**VDW**	**Elec**	**Total**	**VDW**	**Elec**
Amidoglucose	**-123.92**	-0.72	-123.21	**-120.28**	-0.36	-119.92	**-127.31**	-1.96	-125.35	**-122.54**	-0.65	-121.89
Phenylthio	**-16.34**	-13.25	-3.09	**-16.39**	-13.67	-2.72	**-16.31**	-14.63	-1.67	**-17.08**	-14.77	-2.31
Linker	**-5.93**	-4.40	-1.53	**-8.30**	-5.72	-2.58	**-10.21**	-8.13	-2.08	**-12.95**	-10.38	-2.57
Plumbagin	**-35.04**	-20.16	-14.88	**-47.20**	-19.64	-27.55	**-32.83**	-22.57	-10.26	**-48.70**	-22.77	-25.94
Total	**-181.23**	-38.52	-142.71	**-192.16**	-39.40	-152.77	**-186.65**	-47.29	-139.35	**-201.27**	-48.57	-152.70
All	**-181.23**	-38.52	-142.71	**-192.16**	-39.40	-152.77	**-186.65**	-47.29	-139.35	**-201.27**	-48.57	-152.70

interaction is unstable in **1c** and does not exist in **1a** and **1b**. However, the above interactions do not underlie a trend in the IE. Hydrophobic interactions with Tyr142, Leu19 and Leu259 are weakened in the longer chained inhibitors. Electrostatic interactions with Asp137 are most favourable for **1b** and **1d** which have an outward facing C-8′ hydroxyl hydrogen and less favourable for **1a** and **1c**. Furthermore, the hydrogen bond formed by the C-8′ hydroxyl group with the MshB-bridging water molecule does not furnish the **1d** with a significantly improved IE. The plumbagin hydroxyl group of **1a** has an IE of -7.64kcal/mol with the surrounding waters, which is comparable to **1b** (-6.00kcal/mol) and **1c** (−12.72kcal/mol).

Therefore, on a purely enthalpic interpretation, the binding modelled here does not reproduce the experimental trend of increasing binding strength with increasing alkyl linker. Inhibitor IEs do not arise from specific interactions enabled by a longer linker. Rather, they arise from additive IEs of the respective moieties, for which only the linker region displays a trend. However, intramolecular interaction and the entropic contributions will play an important role in the binding, especially considering the hydrophobic nature of the linker and naphthoquinone moieties. Therefore, free energy calculations are required to determine which inhibitors bind most favourably and to validate the model presented above.

5.3.3. FREE ENERGY CALCULATIONS

The model of inhibitor binding presented here must be validated against experimental inhibition constants data. Relative free energies of inhibitor binding, $\Delta\Delta A_{binding}$, should mirror the trend observed in the K_i values. **1d** displays the strongest binding and, therefore, should have a more favourable absolute free energy of binding than **1c**. Consequently, $\Delta\Delta A_{binding}$ should be positive for the mutation of **1d** to **1c**. Likewise $\Delta\Delta A_{binding}$ should be positive for the mutation of **1c** to **1a**. Although no inhibition constant has been calculated for **1b**, it displays a MshB inhibition that is similar to **1c** (Table 1-1, Chapter 1), and $\Delta\Delta A_{binding}$ should be smaller than for the other two transformations.

TABLE 5-8 **Free energy results for the transformation in which a methylene group is removed. Roman numerals specify the hybrids constructed from different 1c and 1d water-equilibrated structures. K_i values for the respective inhibitors are given in brackets.**

Transformation	ΔΔA(i) (kcal/mol)	ΔΔA(ii) (kcal/mol)
1d (16.8 ± 1.9μM)-**1c** (94 ± 11μM)	-2.14	-1.06
1c (94 ± 11μM)-**1b**	0.93	1.29
1b-**1a** (167± 15μM)	-1.78	

TABLE 5-9 **Free Energy results for the transformation in which a methylene group is added. Roman numerals specify the hybrids constructed from different 1c water-equilibrated structures. K_i values for the respective inhibitors are given in brackets.**

Transformation	ΔΔA(i) (kcal/mol)	ΔΔA(ii) (kcal/mol)
1c (94 ± 11μM)-**1d** (16.8 ± 1.9μM)	1.15	1.20
1b-**1c** (94 ± 11μM)	-2.11	
1a (167± 15μM)-**1b**	3.14	

Relative free energies of binding were calculated using the parameterised force field, and protein coordinates described above. The results for the forward perturbations (**1d** to **1a**) are presented in Table 5-8 and the reverse perturbations (**1a** to **1d**) are presented in Table 5-9. PCA in Chapter 5.2 identified two favourable conformations for each of **1c** and **1d** in solution**.** These conformations, annotated by (i) and (ii), had the phenylthio group proximal to each face of the plumbagin**,** and were used to construct the hybrids for the **1d** to **1c**, **1c** to **1b** and **1c** to **1d** transformations in water. The signs of $\Delta\Delta_{binding}$ converge in the forward and reverse directions for all calculations. However, the results are not in accordance with the experimental data, and show that **1c** binds more favourably than **1d**, and **1a** has binds more favourably than **1b**. This could be a result of two sources of error in the free energy simulations. Firstly errors in the free energy calculations may arise from deficiencies in the force field. However, the CHARMM-27 force field is well established and missing parameters were carefully calculated from QM calculations. It is not envisaged that the deficiencies in the force field would result in such large deviation from the experimental data. A second source of error could be that the simulation length was too short and led to insufficient sampling of the conformational space.

However, beyond the simulation errors the analysis of the IEs above (Section 5.4.3) showed that the differences in inhibitor IEs do no arise from specific interactions enabled by a longer linker.

This observation prompted re-evaluation of the initial coordinates before errors in the free energy simulations were reduced.

5.4. RE-MODELLING INHIBITOR BINDING WITH MSHB

When reconsidering the initial coordinates, the plumbagin and alkyl linker geometry was examined, since GlcNAc binding was assumed to be the same as for the natural substrate. Analysis of solution dynamics showed significant interaction between the plumbagin and phenylthio groups in **1c** and **1d** (Chapter 5.2). Therefore, it was decided to re-dock these inhibitors so that their geometries were consistent with their favourable conformations in solution.

5.4.1. DOCKING AND MOLECULAR DYNAMICS

In Section 5.2, PCA was conducted on the MD simulation of the inhibitor series in solution. The distance between the COM of the phenylthio and plumbagin moieties was used as the principal component. The centres of the most populated clusters for **1c** and **1d** showed two distinct conformations, annotated (i) and (ii). These conformations had the phenylthio group proximal to each face of the plumbagin**.** The most populated cluster centres for **1b** showed interaction with the same plumbagin face despite starting from conformations with different relative plumbagin orientations**.** Therefore, the linker was built onto the GlcNAc fragment in the MshB active site, such that the linker dihedrals were the same as the favoured conformations from the solution dynamics of **1b**, **1c**(i), **1c**(ii), **1d**(i) and **1d**(ii) respectively (Section 5.2). The docked conformation of **1d**(ii) is compared with the previous starting conformation in Figure 5-19. The poses were refined in Glide [11] using the same grid generated for previous docking studies (Section 5.1.3). The systems were then solvated and subjected to 5ns stochastic boundary dynamics. The RMSD time series are given in Figure E-1 (Appendix E) and the average IEs for the final 1ns of the trajectory are summarised in Table 5-12 and Table 5-11. Time series of the

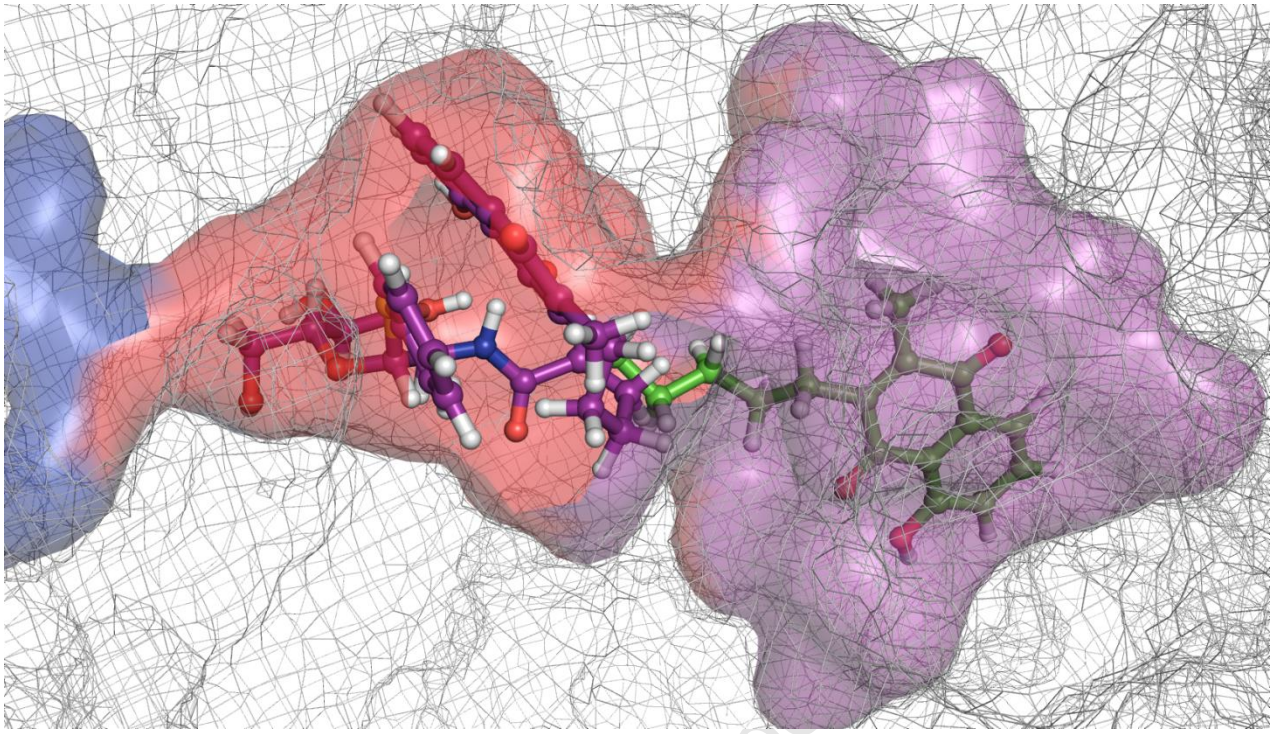


FIGURE 5-19 Re-docked conformation of 1d(ii) with carbon atoms show in purple. The previous orientation with the plumbagin inserted into the hydrophobic cavity of Region B is shown with green atoms.

interaction with all the other atoms in the system are given in Figure 5-20. **1c**(ii) and **1d**(ii), which have a relative orientation of their plumbagin and phenylthio groups that is similar to **1b**(ii), displayed lower IEs than their respective conformational isomers **1c**(i) and **1d**(i).

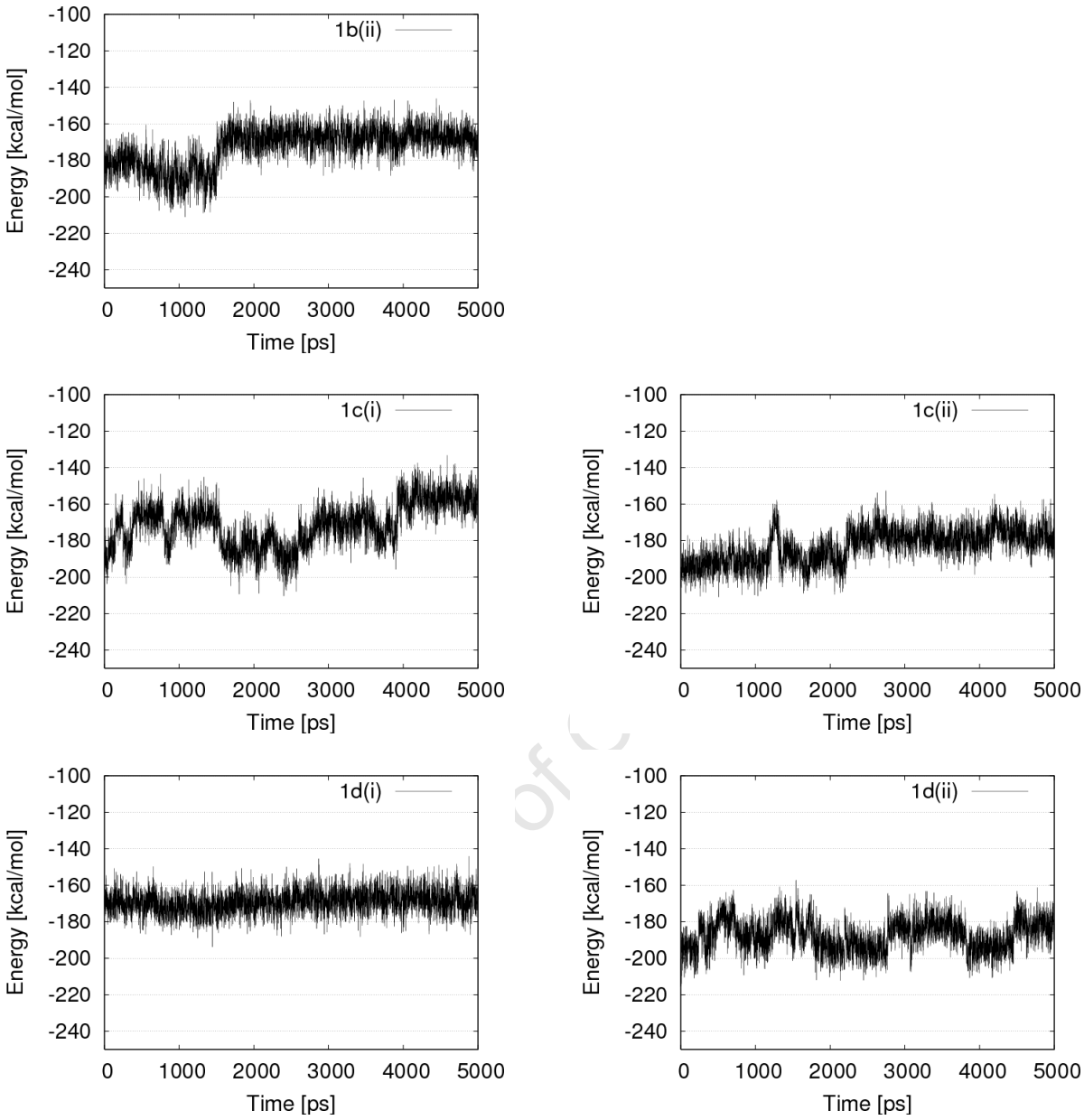


FIGURE 5-20 Time series for the IEs of inhibitors with all atoms in the system.

5.4.2. FREE ENERGY CALCULATIONS

1b, **1c**(ii) and **1d**(ii) were used as starting structures for new free energy calculations in the direction of **1d** to **1a**. (Reactions in the reverse direction to test for convergence have not been calculated). The results are given in table Table 5-10. These results compare much more favourably with the experimental K_i data. The relative free energies show that **1d** binds more favourably than **1c**, and **1c** bind more favourably than **1a.** The remaining mutation, **1c** to **1b**, has a smaller $\Delta\Delta A_{binding}$, which is also consistent with the percentage MshB inhibition at 500μM concentrations of the inhibitors (81.4% and 81.6% respectively).

5.4.3. TRAJECTORY ANALYSIS

Having achieved satisfactory agreement with experimental data, a preliminary analysis of the models was conducted to identify key interactions. Again the phenyl thioglycoside IEs are variable with a difference of 9.09kcal/mol between **1c** and **1d**. The unnormalised linker interactions as well as the plumbagin interactions increase with the number of linker atoms. The plumbagin MshB interactions of **1c** and **1d** are similar, but there is a difference of 4.17kcal/mol in their IEs with water.

TABLE 5-10 Free energy results for the transformation in which a methylene group is removed. K_i values for the respective inhibitors are given in brackets.

Transformation	ΔΔA
1d (16.8 ± 1.9 μM)-**1c** (94 ± 11 μM)	1.92
1c (94 ± 11 μM)-**1b**	0.88
1b-**1a** (167± 15 μM)	1.19

TABLE 5-11 Average IEs between inhibitor moieties and MshB residues. All energies are reported in kcal/mol, and the sum of the energies (Total) are compared with interaction of the entire inhibitor with MshB (MshB).

Moiety	**1b**(ii)			**1c**(ii)			**1d**(ii)		
	Total	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**
Amidoglucose	**-117.41**	-3.07	-114.34	**-108.12**	-3.74	-104.38	**-115.86**	-3.04	-112.82
Phenylthio	**-2.36**	-4.33	1.97	**-7.14**	-5.80	-1.34	**-8.49**	-7.31	-1.17
Linker	**-5.88**	-3.49	-2.39	**-7.34**	-4.73	-2.62	**-8.00**	-5.53	-2.47
Plumbagin	**-9.82**	-9.08	-0.75	**-19.03**	-13.73	-5.30	**-20.70**	-16.53	-4.17
Total	**-135.47**	-19.96	-115.51	**-141.63**	-28.00	-113.64	**-153.04**	-32.41	-120.63
MshB	**-135.47**	-19.96	-115.51	**-141.63**	-28.00	-113.64	**-153.04**	-32.41	-120.63

TABLE 5-12 Average IEs between inhibitor moieties and all remaining atoms in the system. All energies are reported in kcal/mol, and the sum of the energies (Total) are compared with interaction of the entire inhibitor with all atoms (All).

	1b(ii)			**1c**(ii)			**1d**(ii)		
	Total	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**	**Total**	**VDW**	**ELEC**
Amidoglucose	**-119.54**	-2.81	-116.73	**-121.78**	-2.60	-119.18	**-123.90**	-3.53	-120.37
Phenylthio	**-11.32**	-9.98	-1.34	**-14.12**	-11.88	-2.24	**-14.98**	-13.28	-1.70
Linker	**-7.50**	-5.51	-1.99	**-9.73**	-7.77	-1.96	**-11.67**	-9.88	-1.80
Plumbagin	**-28.50**	-18.43	-10.07	**-31.06**	-20.55	-10.50	**-36.94**	-22.20	-14.75
Total	**-166.86**	-36.73	-130.13	**-176.69**	-42.80	-133.89	**-187.49**	-48.89	-138.60
All	**-166.86**	-36.73	-130.13	**-176.69**	-42.80	-133.89	**-187.49**	-48.89	-138.60

BINDING MODES

The binding of the inhibitors are summarised in Figure 5-21 to Figure 5-23. In all instances the 2 -acylamido glucose binding is the same as previously modelled. Zn^{2+} coordinates the amide linker in accordance with the proposed catalytic mechanism. The C-3 hydroxyl oxygen is hydrogen bonded to Arg68, the 4-C hydroxyl group is hydrogen bonded to Arg68 and Asp95, and the C-6 hydroxyl oxygen is hydrogen bonded to Asp95. The binding of the hydrophobic alkyl linker and C-3′ methyl group is mediated by non-polar interactions with a hydrophobic pocket composed of Leu19, Leu259, Leu263, Leu265 and Pro209. Improved contact of these residues with a larger non-polar surface provided by longer linkers may contribute to more favourable binding reported in the free energy calculations.

The plumbagin moiety lies next to Tyr142 in all inhibitors and is close to Ile214 in **1c** and **1d**. In **1b,** the aromatic atoms of the plumbagin group lie parallel to the phenylthio group, and show a T-like geometry with the Tyr142 aromatic ring. In **1c** and **1d** the thioglycoside linkage has rotated to its second minima (adiabatic map is given in Figure 5-5) and the phenyl ring shows stacking orientation with His144. To explore differences in binding of these two residues, the distance between the COM of their side chains and the inhibitors were calculated over the last 1ns of the trajectory. Figure 5-24 correlates these distance time series with their associated IEs. It is observed that the phenylthio group lies far from His144 in **1b** and displays a weak IE. The inter-residue distance in **1c** and **1d** is much closer, and leads to a stable interaction in **1d**. However, fluctuation in the **1c**-histidine IE, shows that this interaction is not as stable. In the above geometry the plumbagin group of **1c** and **1d** maximises surface contact with Tyr142. Unfortunately, time did not allow for vector analysis of their relative orientations which is important for characterising π-stacking interactions. The COM distance and IE time series for the Tyr142-inhibitor interactions are reported in Figure 5-24, and show that the proximity of the two hydrophobic groups gives rise to a stronger IE in **1c** and **1d**. The more favourable interaction between His144 and Tyr142 contribute to an improved MshB interaction in **1c** and **1d**. It is suggested that these interaction are facilitated by a longer linker, which imparts more flexibility to the C-2 acylamido side chain. A more in depth analysis is required to confirm this hypothesis, and elucidate whether such interactions are important for the free energy of binding.

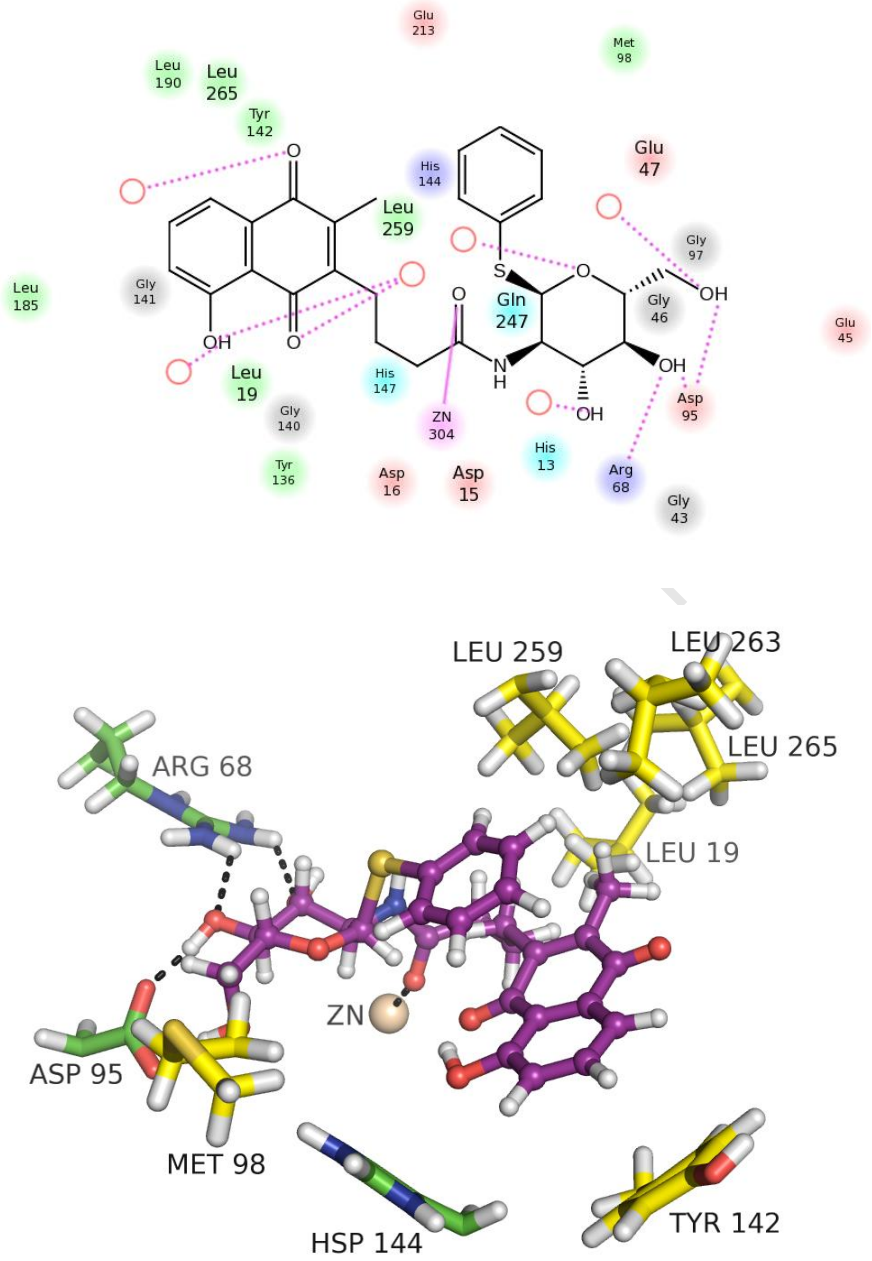


FIGURE 5-21 **(A) 2D illustration summarising the binding of 1b with MshB. (B) Snapshot showing important hydrogen bonding and hydrophobic interactions. Polar MshB residues are shown with green carbons, while hydrophobic residues are shown with yellow carbons.**

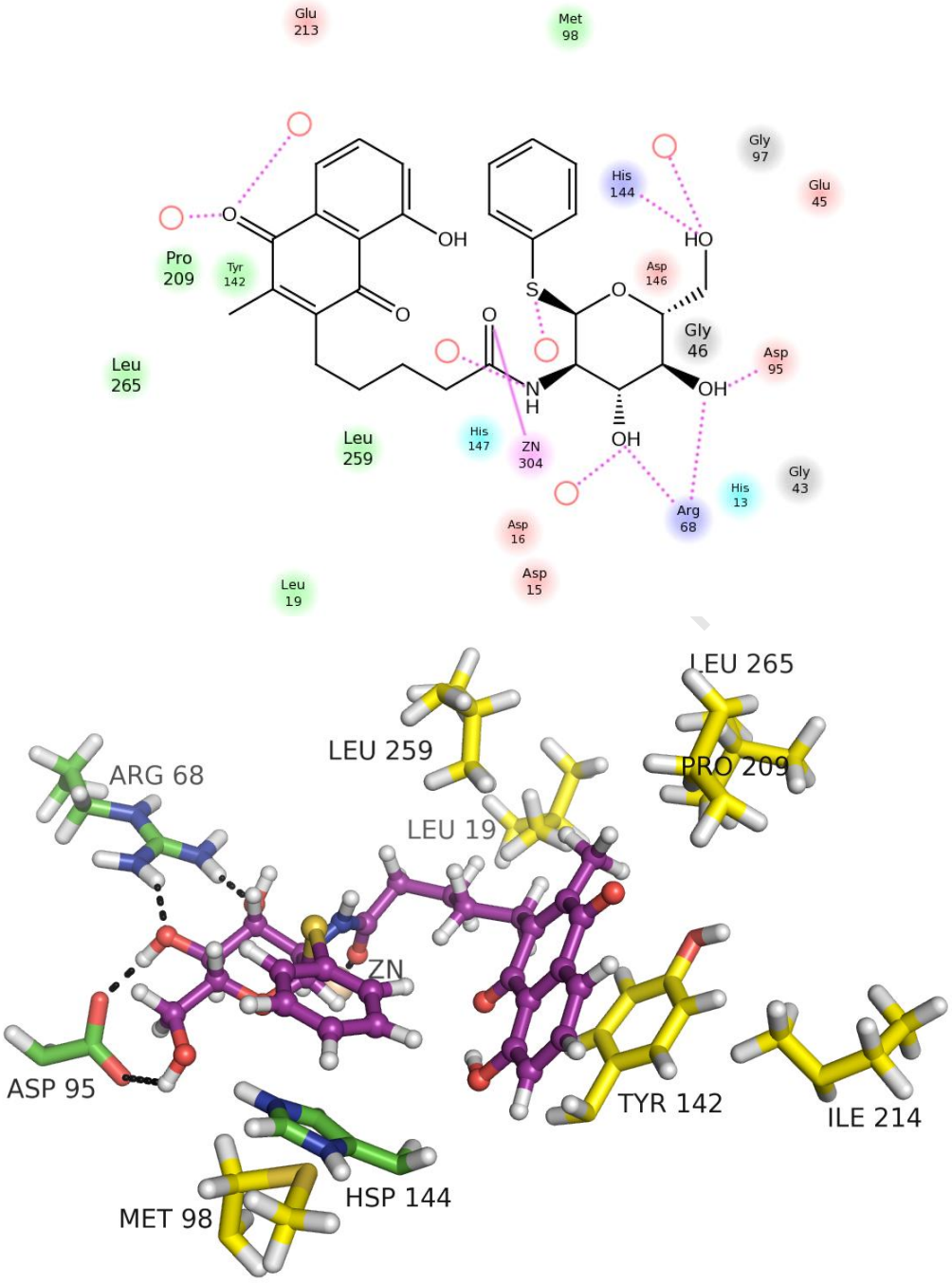


FIGURE 5-22 **(A) 2D illustration summarising the binding of 1c with MshB. (B) Snapshot showing importatnt hydrogen bonding and hydrophobic interactions. Polar MshB residues are shown with green carbons, while hydrophobic residues are shown with yellow carbons.**

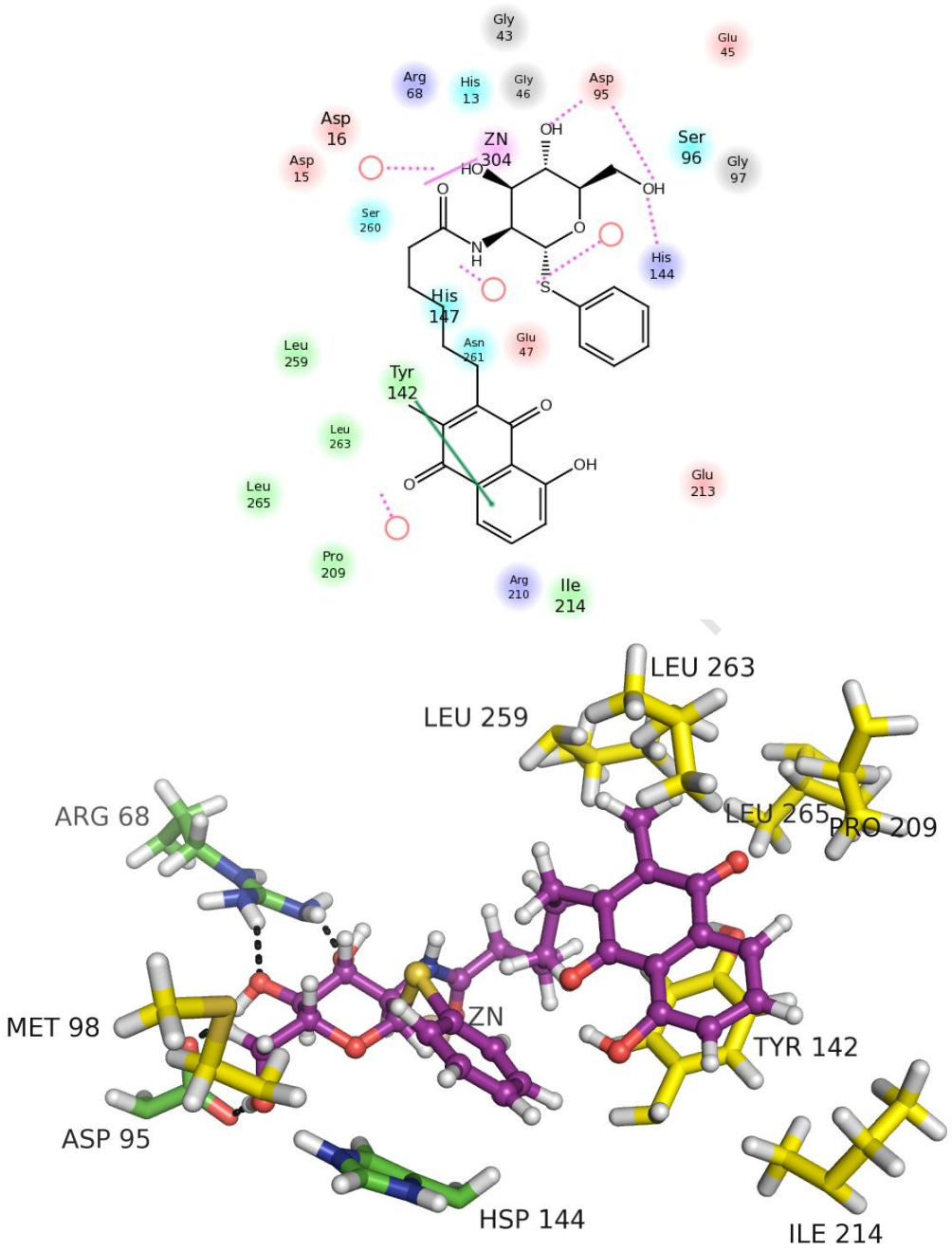


FIGURE 5-23 **(A) 2D illustration summarising the binding of 1d with MshB. (B) Snapshot showing important hydrogen bonding and hydrophobic interactions. Polar MshB residues are shown with green carbons, while hydrophobic residues are shown with yellow carbons.**

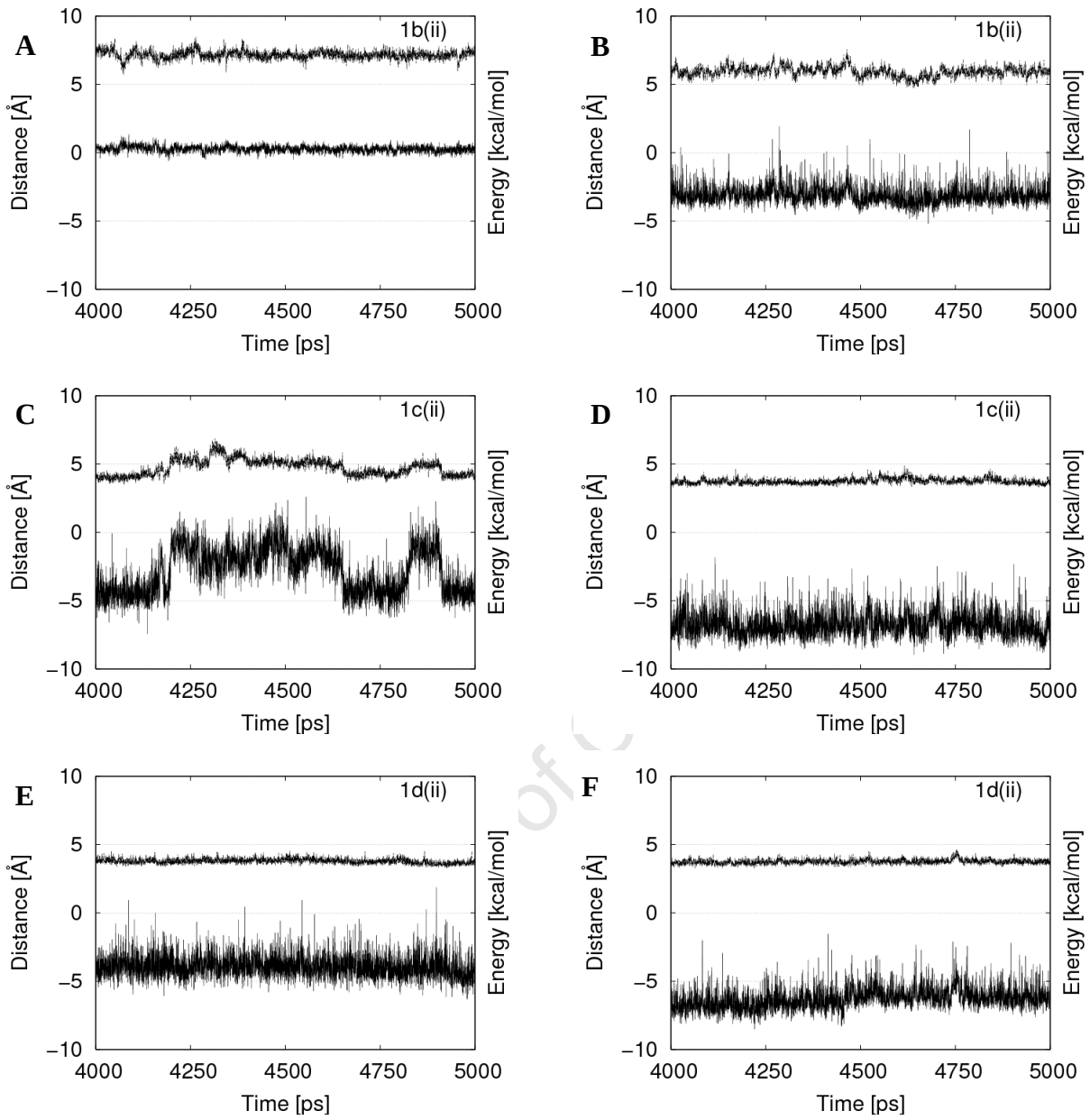


FIGURE 5-24 **(A, C & E) Time series for distance between the COM of the phenylthio group and His144 side chain. (B, D & F) IE time series for the Tyr142-inhibitor interactions.**

5.5. RATIONALISATION OF INHIBITION

It is not clear from the results of the MD simulations how the plumbagin conjugates inhibit MshB. MshB has known amidase activity with a range of substrates [22, 23], which is foreseen to proceed *via* the same proposed catalytic mechanism for the deacetylation of GlcNAc-Ins. The equilibrated structures discussed above have the C-2 amido carbonyl coordinated to zinc at the same coordination site as that modelled for the natural substrate. In this conformation the zinc-coordinated water lies close to the proposed general base Asp15, and the water molecule would be in a good position to attack the amide linkage after abstraction of its hydrogen. It could be that the inhibitors are acting as competing substrates rather than inhibiting the amidase activity of MshB. Difficulty in obtaining a co-crystallised structure with **1d** supports is in agreement with the hypothesis. It would therefore be pertinent to repeat the inhibition assays and test for the presence of GlcN-SPh.

5.6. REFERENCES

1. D. W. Gammon, D. J. Steenkamp, V. Mavumengwana, M. J. Marakalala, T. T. Mudzunga, R. Hunter and M. Munyololo, *Bioorganic & Medicinal Chemistry*, **18**, 2010, 2501-2514.
2. M. Kuttel, J. W. Brady and K. J. Naidoo, *Journal of Computational Chemistry*, 2002, **23**, 1236-1243.
3. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. MacKerell, *Journal of Computational Chemistry*, **31**, 671-690.
4. J. L. Knight and C. L. Brooks, *Journal of Chemical Theory and Computation*, 2009, **5**, 1680-1691.
5. M. J. Frisch; G. W, Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A., Jr. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam,

S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A Pople., *Gaussian 03*, revision C. 02, Gaussian Inc.:Wallingford, CT, 2004.

6. C. M. Breneman and K. B. Wiberg, *Journal of Computational Chemistry*, 1990, **11**, 361-373.
7. G. A. Venter, R. P. Matthews and K. J. Naidoo, *Molecular Simulation*, 2008, **34**, 391-401.
8. M. Saito and I. Okazaki, *Journal of Computational Chemistry*, 2009, **30**, 2656-2665.
9. K. J. Naidoo and J. W. Brady, *Chemical Physics*, 1997, **224**, 263-273.
10. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *Journal of Chemical Physics*, 1983, **79**, 926-935.
11. Glide, version 5.6, Schrödinger, LLC, New York, NY, 2010.
12. R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *Journal of Medicinal Chemistry*, 2004, **47**, 1739-1749.
13. R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *Journal of Medicinal Chemistry*, 2006, **49**, 6177-6196.
14. T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *Journal of Medicinal Chemistry*, 2004, **47**, 1750-1759.
15. A. A. McCarthy, N. A. Peterson, R. Knijff and E. N. Baker, *Journal of Molecular Biology*, 2004, **335**, 1131-1141.

16. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *Journal of Computational Chemistry*, 1983, **4**, 187-217.
17. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *Journal of Computational Chemistry*, 2009, **30**, 1545-1614.
18. M. E. Karpen, D. J. Tobias and C. L. Brooks, *Biochemistry*, 1993, **32**, 412-420.
19. G. A. Carpenter and S. Grossberg, *Applied Optics*, 1987, **26**, 4919-4930.
20. SiteMap, version 2.4, Schrodinger, LLC, New York, 2010.
21. Z. Chen, M. X. Tan, L. Liu, Y. Liu, H. Wang, B. Yang, Y. Peng, H. Liu, H. Liang and C. Orvig, *Dalton Transactions*, 2009, 10824-10833.
22. G. L. Newton and R. C. Fahey, *Archives of Microbiology*, 2002, **178**, 388-394.
23. V. K. Jothivasan and C. J. Hamilton, *Natural Product Reports*, 2008, **25**, 1091-1117.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

This research has provided insight into the competitive inhibition of MshB by substrate-based analogues synthesised by Gammon *et al.* [1]. Neither inhibitor nor substrate analogue has been co-crystallised with MshB, and the natural substrate binding is uncertain. Therefore, the binding site was characterised and showed a large hydrophobic cavity adjacent to the putative active site. The results of docking calculations were in accordance with previous models of the GlcNAc fragment, and with the proposed catalytic mechanism. Atomistic MD simulations of the docked pose elucidated hydrophobic interactions between D-*myo*-inositol and Met98, as well as between the C-2 acetamido group and Leu259.

The inhibitors were docked having the same glucose orientation as the natural substrate. The systems were equilibrated and used as starting points for free energy simulations. The inhibitors were ranked according to their relative free energy of binding, and the ranking was compared to experimental K_i values. The results indicated that the binding mode is not one which would be expected given the landscape of the MshB binding pocket.

Initial docking studies yielded conformations with C-2 acylamido side chain extending linearly from the glucose moiety, into the hydrophobic cavity adjacent to the putative active site. This appeared plausible due to the glucose orientation, and the complementary nature of the hydrophobic linker and plumbagin groups with the cavity. However, free energy results using the above conformations did not reproduce the experimental ranking. Analysis of the MshB-inhibitor interactions showed that, in these models, a longer alkyl spacer did not enable more favourable interactions. This prompted re-evaluation of the starting coordinates.

PCA studies of the inhibitor dynamics in solution had revealed hydrophobic interaction geometries between the phenylthio and plumbagin groups for the most populated cluster centres

of **1b**, **1c** and **1d**. These conformations were docked with the MshB active site and the free energy calculations were repeated. The ranking of the inhibitors based on the new relative free energies of binding were in agreement with the experimental ranking. Due to the timeframe of this dissertation free energy transformations were only run from **1d** to **1a.** It is necessary to run the calculations in the reverse direction, and for longer periods, in order to test for convergence. Preliminary analysis of the MD trajectories showed that the longer linker facilitated more flexible interactions with His144 and Tyr142. Future work requires a more detailed analysis of the enthalpic and entropic contributions to inhibitor binding, as well as investigation of the hydrophobic interactions between the plumbagin and phenylthio groups.

REFERENCES

1. D. W. Gammon, D. J. Steenkamp, V. Mavumengwana, M. J. Marakalala, T. T. Mudzunga, R. Hunter and M. Munyololo, *Bioorganic & Medicinal Chemistry*, 2010, **18**, 2501-2514.

APPENDIX

APPENDIX A: DEPENDENCE OF FREE ENERGY OF λ FOR TI AND FEP METHODS

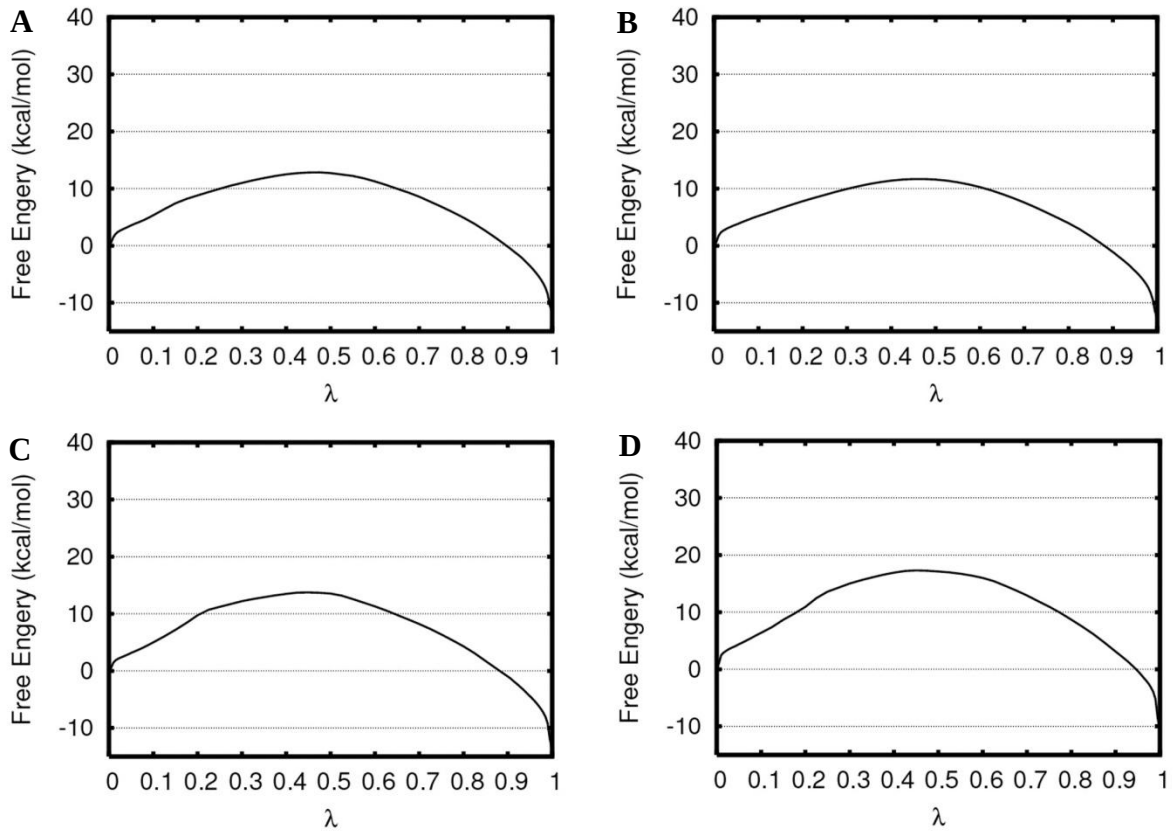


FIGURE A-1 **Dependence of free energy on λ for the transformation of 1d to 1c using TI. (A) 1d(i) to 1c in water. (B) 1d(ii) to 1c in water. (C) 1d to 1c at the MshB active site. (D) 1d(ii) to 1c at the MshB active site. The dependence of free energy on λ can be compared to FEP results in Figure A-2.**

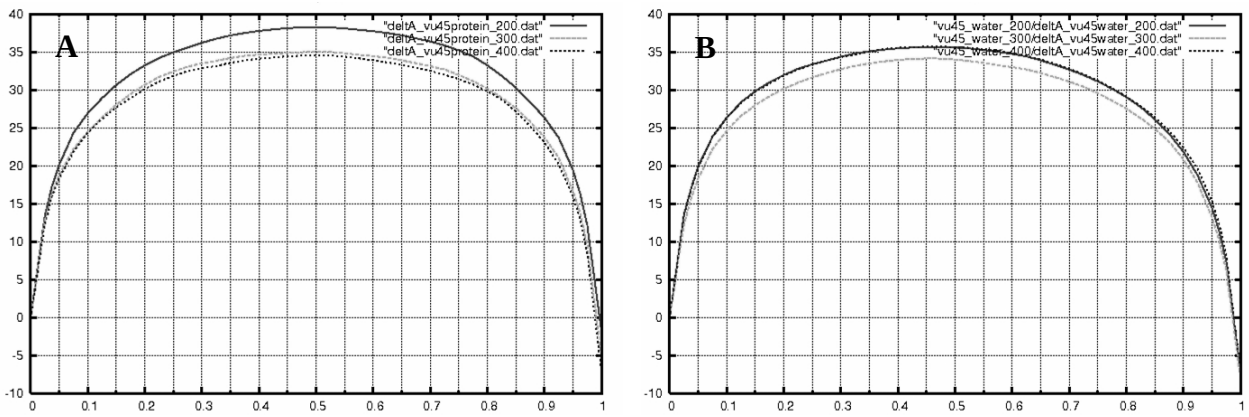


FIGURE A-2 **Dependence of free energy on λ for the transformation of 1d-1c using FEP. (A) 1d to 1c at the MshB active site. (B) 1d to 1c in water. Plots were taken from early work. Different curves on the plots represent results after different equilibration times. A large change in free energy is observed with change in λ at the end-points (0 to 0.1 and 0.9 to 1), which indicates insufficient phase space sampling.**

APPENDIX B: ANALYSIS OF Φ AND Ψ DURING INHIBITOR DYNAMICS IN SOLUTION

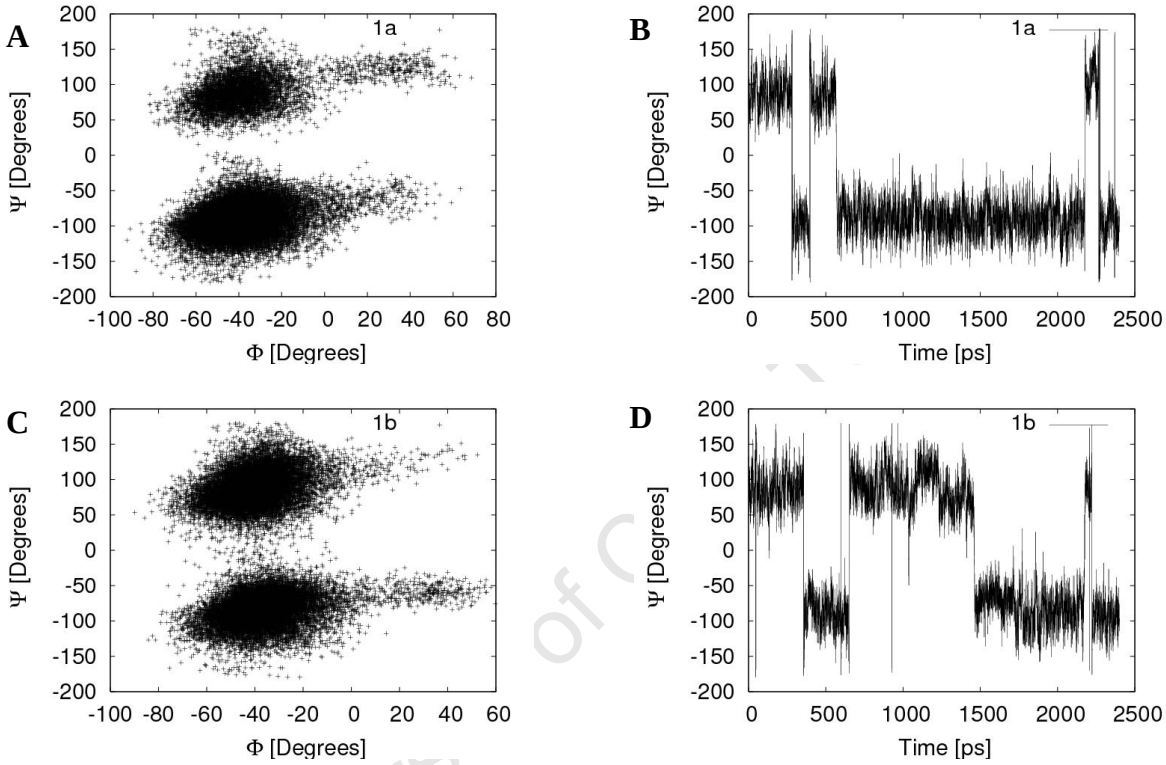


FIGURE B-1 (A & C) φ,ψ scatter plot for 1a and 1b over the 2.5ns trajectory. (B & D) ψ time series. The plots show that the thioglycoside linkage populates both minima identified on the adiabatic map.

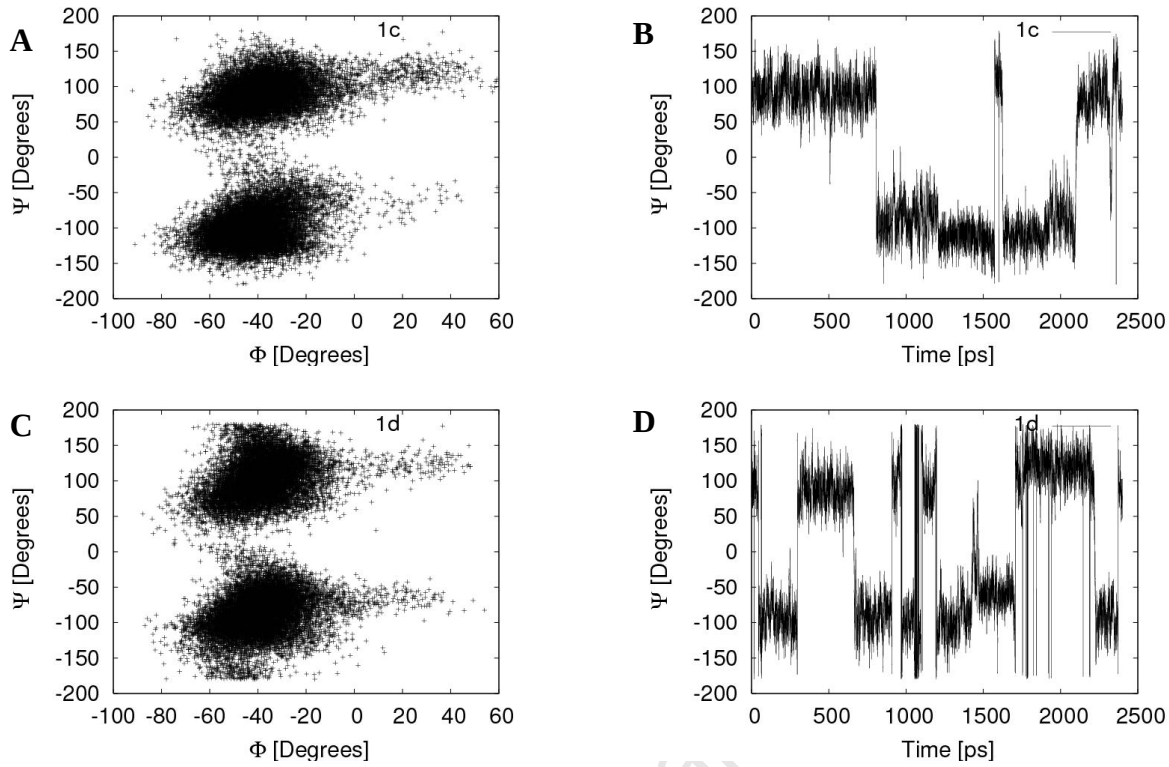


FIGURE B-2 (A & C) φ,ψ scatter plot for 1c and 1d over the 2.5ns trajectory. (B & D) ψ time series. The plots show that the thioglycoside linkage populates both minima identified on the adiabatic map.

APPENDIX C: RMSD TIME SERIES FOR MSHB-INHIBITOR COMPLEXES

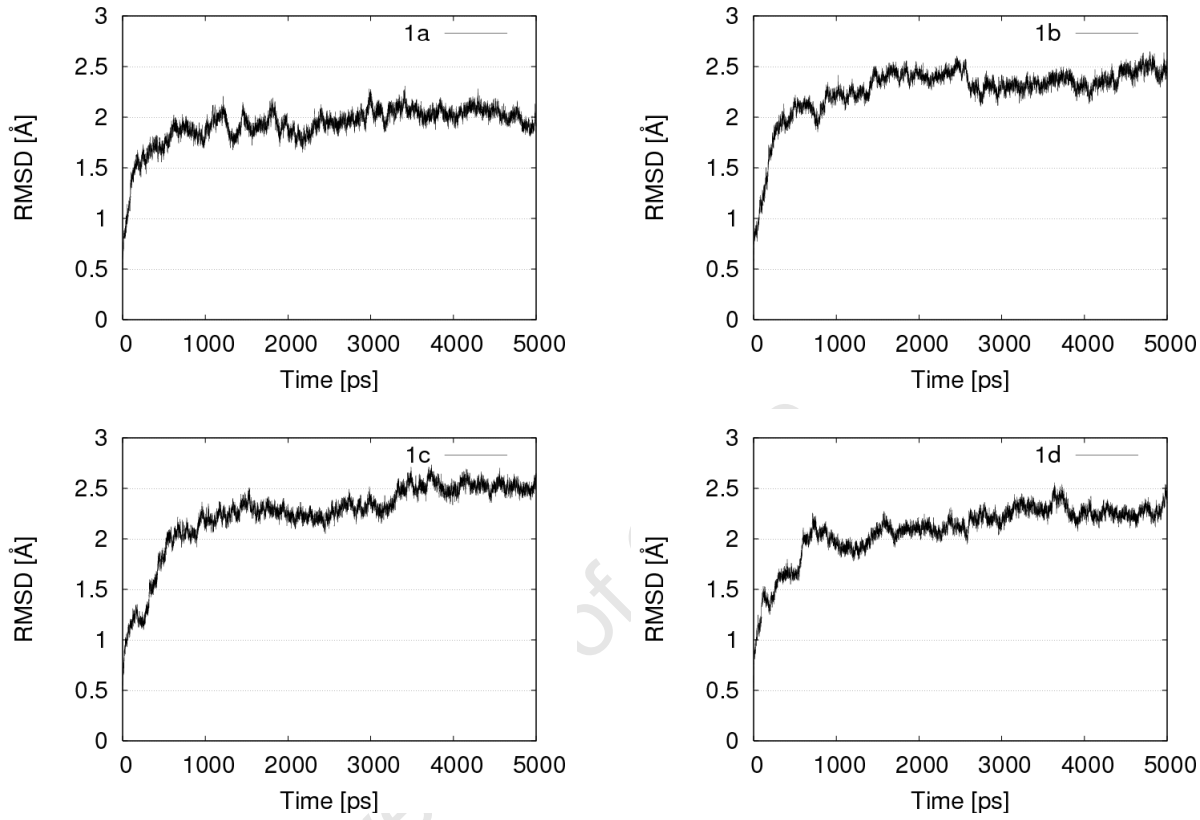


FIGURE C-1 RMSD time series for all amino acid residues which have an atom within 8Å from the inhibitor indicated in the legend.

APPENDIX D: TIME SERIES OF IES AND DISTANCES BETWEEN COMS FOR INTERACTION PAIRS IN THE MSHB-INHIBITOR COMPLEXES

1A

INTERACTIONS OF THE PHENYLTHIO GROUP

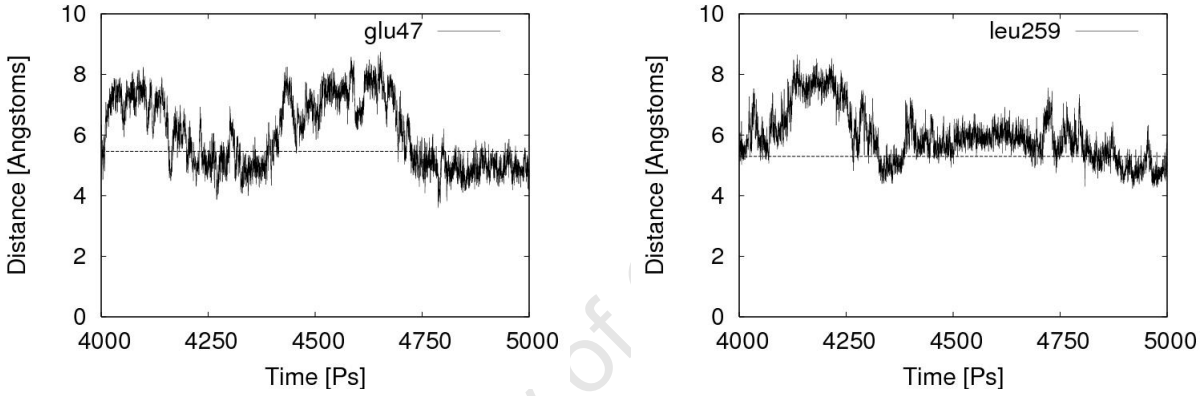


FIGURE D-1 Distance between COM of the phenylthio group of 1a and the amino acid side chains indicated in the legends. In the case of Glu47 the COM for the non-polar methylene atoms were calculated.

LINKER INTERACTIONS

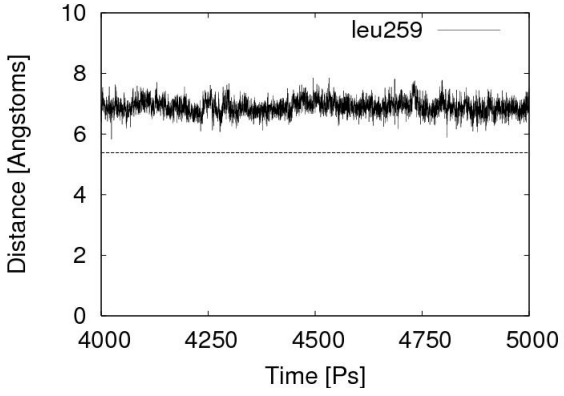


FIGURE D-2 Distance between COM of alkyl linker of 1a and the amino acid side chain indicated in the legend.

INTERACTION OF THE PLUMBAGIN AROMATIC ATOMS

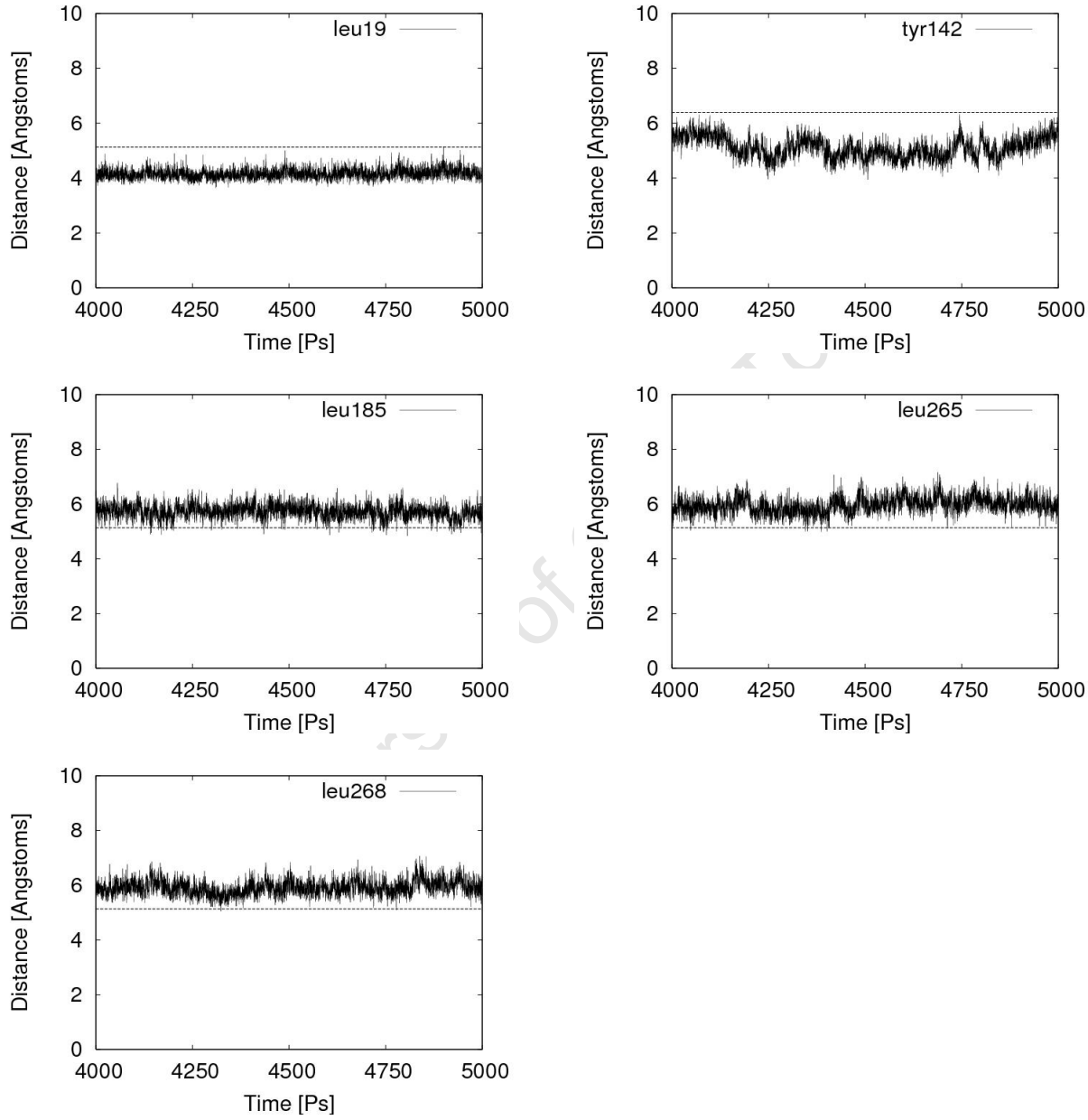


FIGURE D-3 Distance between COM of plumbagin aromatic atoms of 1a and the amino acid side chains indicated in the legends.

INTERACTION OF THE PLUMBAGIN METHYL GROUP

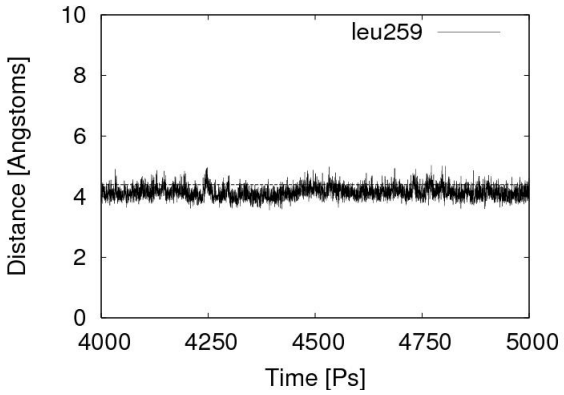


FIGURE D-4 Distance between COM of the plumbagin methyl of 1a and the amino acid side chain indicated in the legends.

1B

INTERACTIONS OF THE PHENYLTHIO GROUP

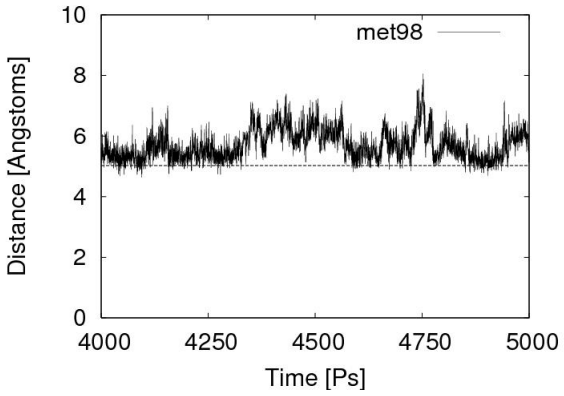


FIGURE D-5 Distance between COM of the phenylthio group of 1b and the amino acid side chain indicated in the legend.

LINKER INTERACTIONS

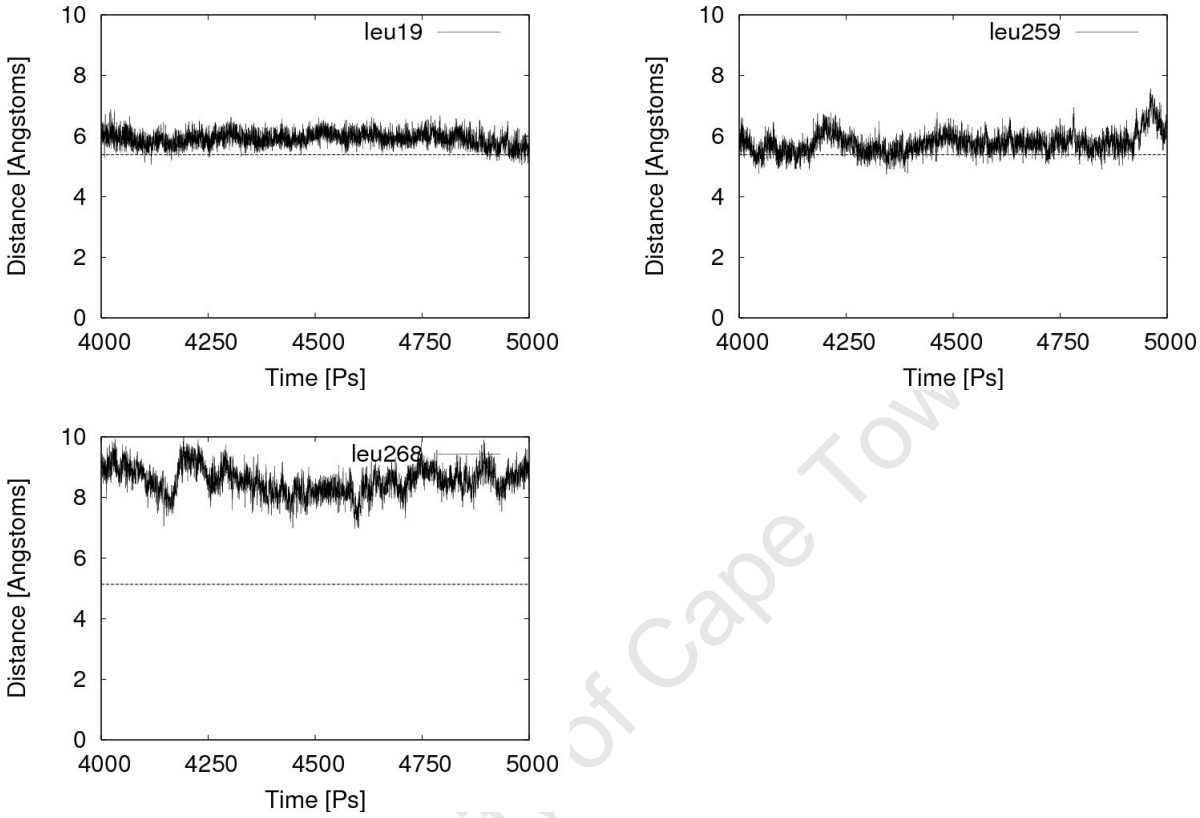


FIGURE D-6 Distance between COM of the alkyl linker of 1b and the amino acid side chains indicated in the legends.

INTERACTIONS OF THE PLUMBAGIN METHYL GROUP

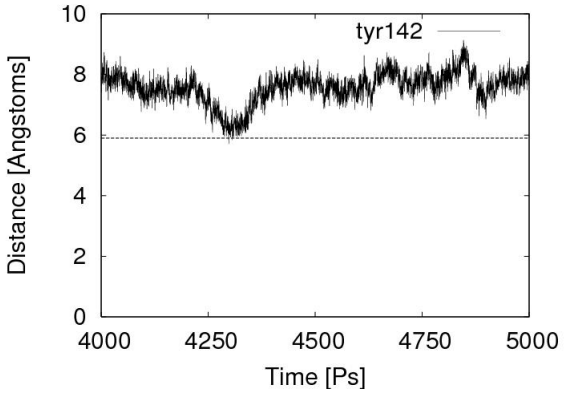


FIGURE D-7 Distance between COM of the plumbagin methyl group and the amino acid side chain indicated in the legend.

INTERACTIONS OF THE PLUMBAGIN AROMATIC ATOMS

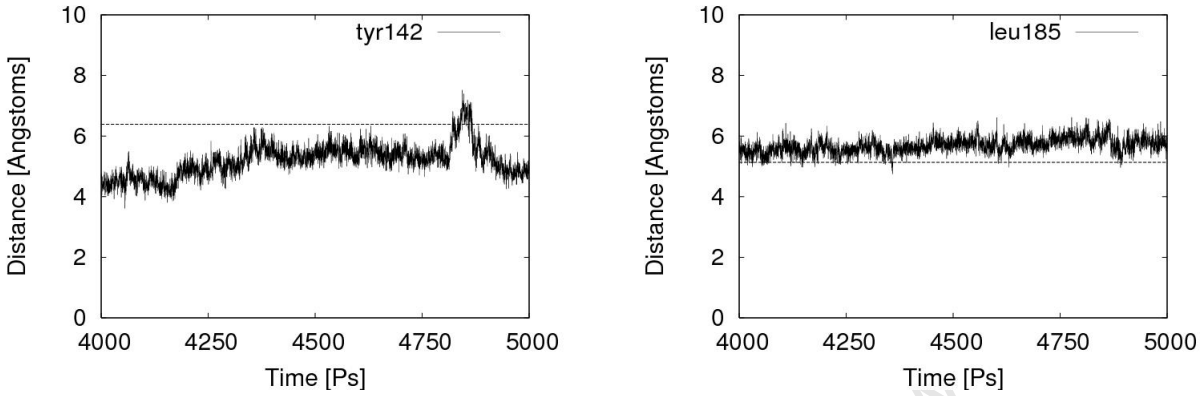


FIGURE D-8 Distance between COM of the plumbagin aromatic atoms of 1b and the amino acid side chains indicated in the legends.

1C

INTERACTIONS OF THE PHENYTHIO GROUP

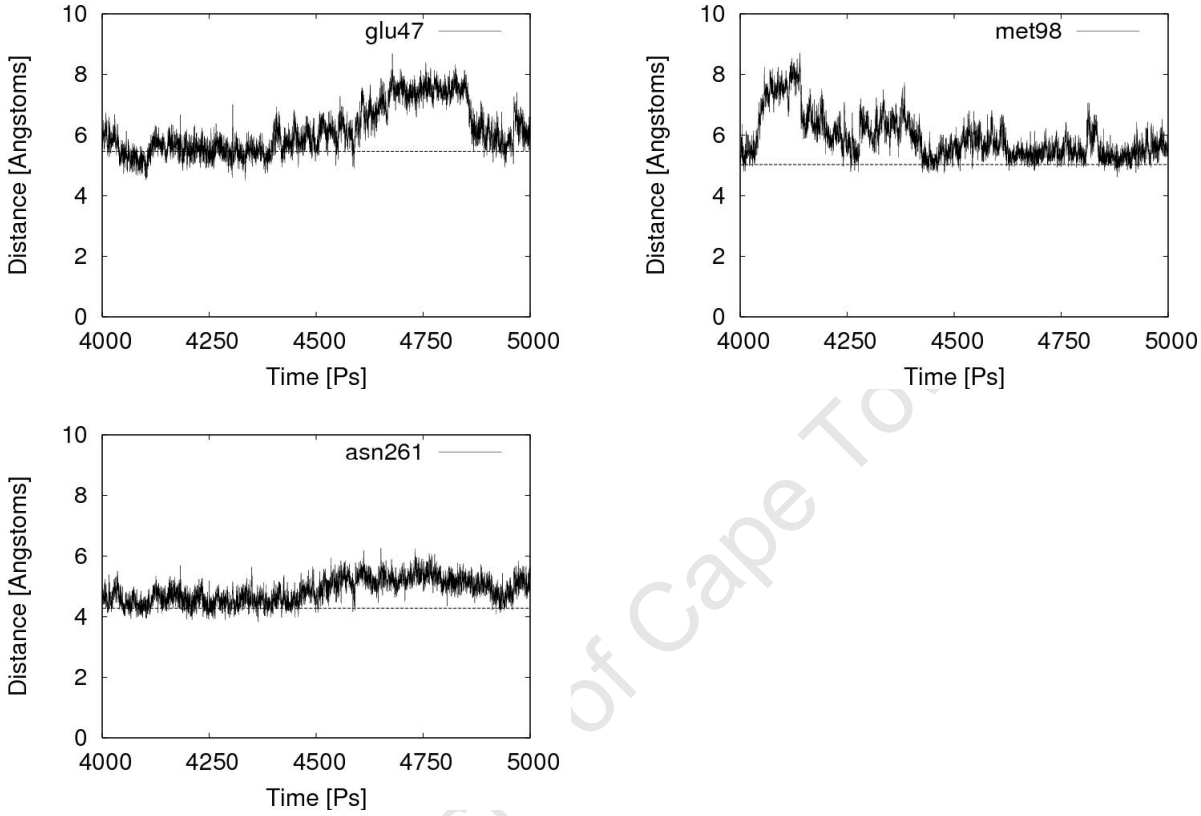


FIGURE D-9 Distance between COM of the phenylthio group of 1c and the amino acid sides chain indicated in the legends. In the case of Glu47 and Asn261 the COM for the non-polar methylene atoms were calculated.

LINKER INTERACTIONS

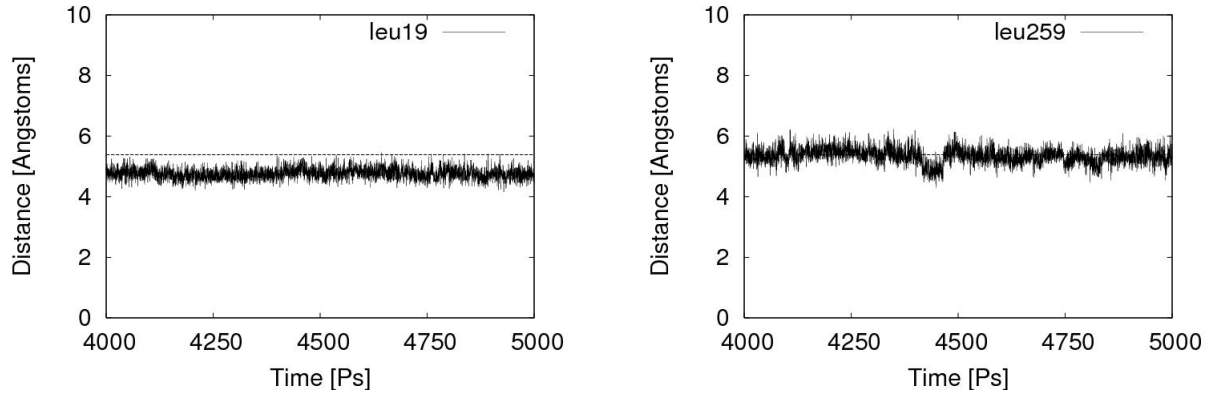


FIGURE D-10 Distance between COM of the alkyl linker of 1c and the amino acid side chains indicated in the legends.

INTERACTIONS OF THE PLUMBAGIN METHYL GROUP

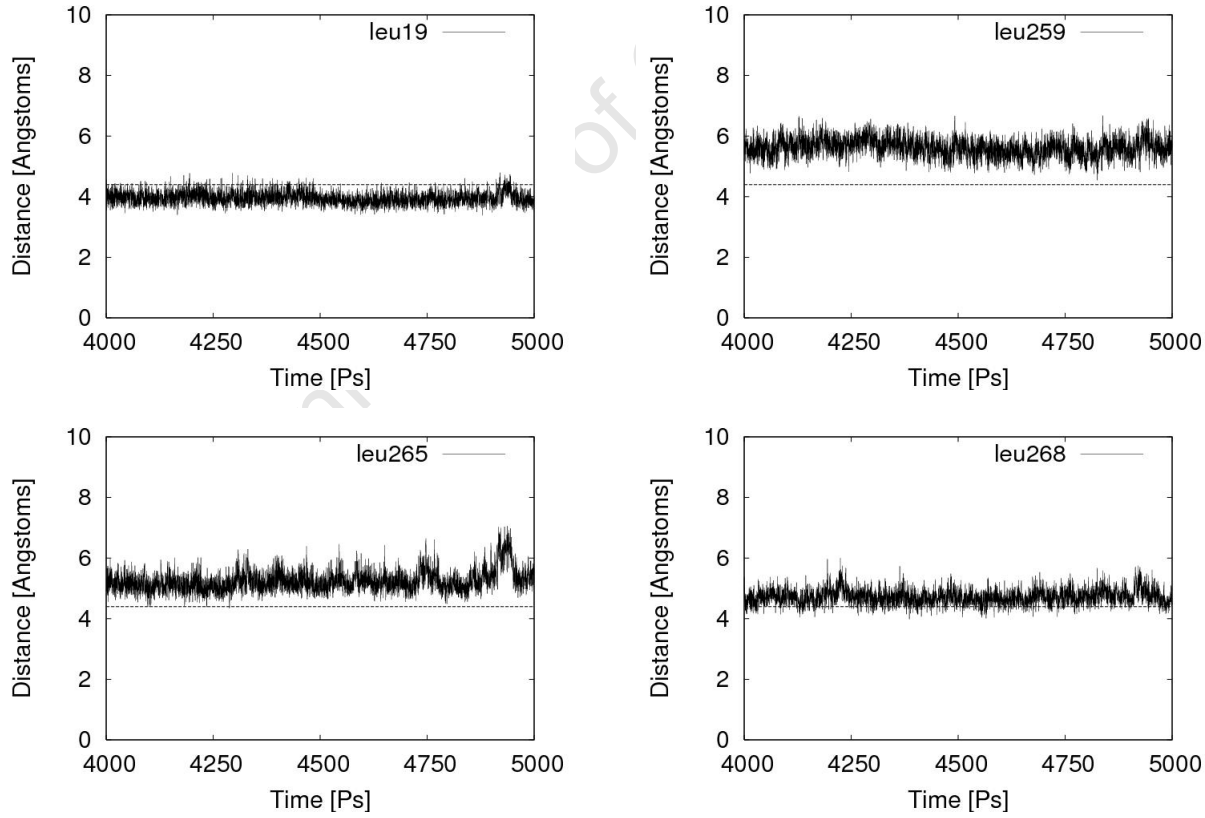


FIGURE D-11 Distance between COM of the plumbagin methyl group of 1c and the amino acid side chains indicated in the legends.

INTERACTIONS OF THE PLUMBAGIN AROMATIC ATOMS

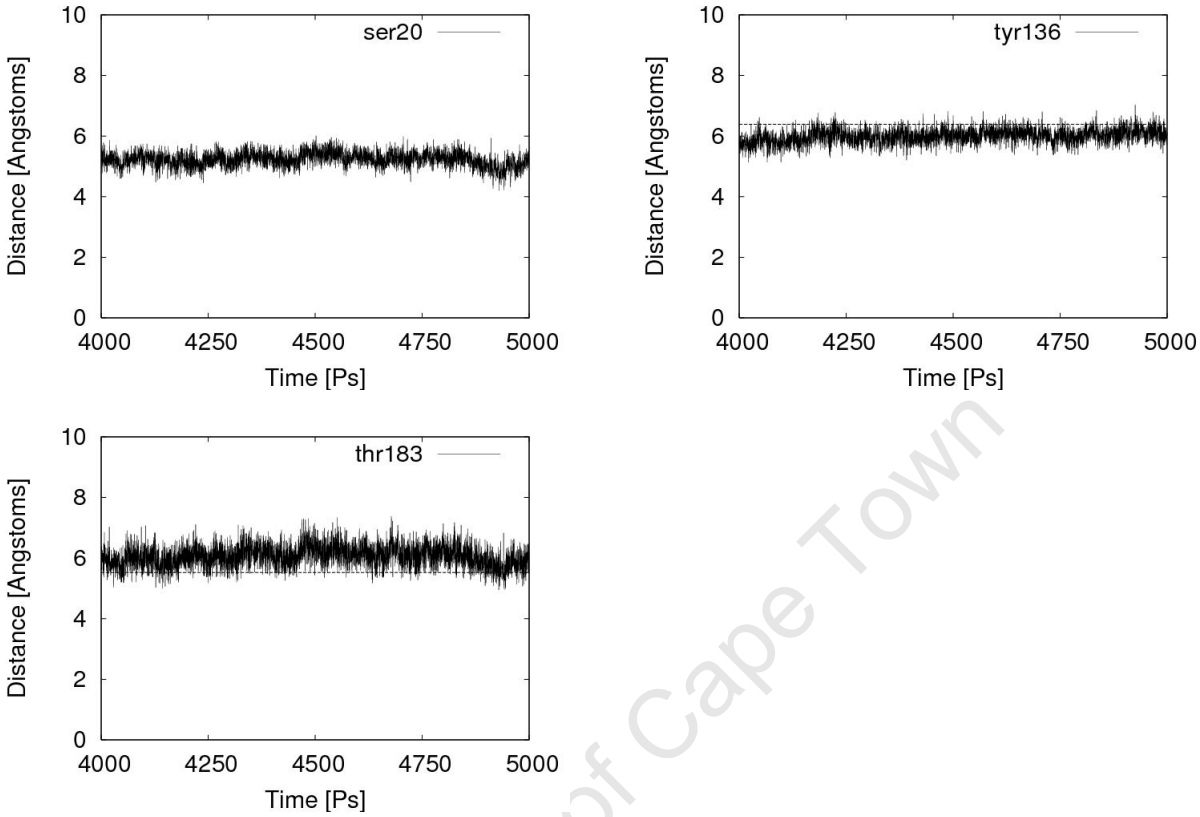


FIGURE D-12 Distance between COM of the plumbagin aromatic atoms of 1c and the amino acid sides chain indicated in the legends

1D

INTERACTIONS OF THE PHENYLTHIO GROUP

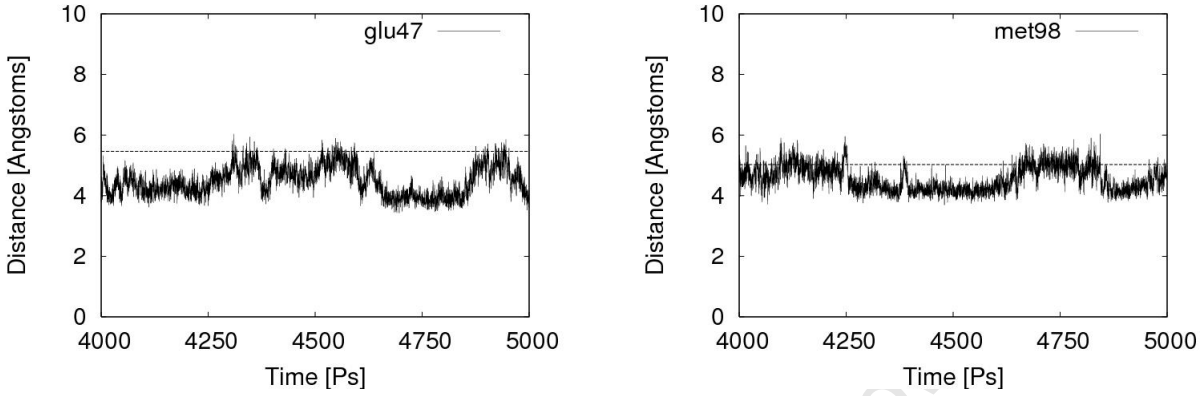


FIGURE D-13 Distance between COM of the phenylthio group of 1d and the amino acid side chains indicated in the legends. In the case of Glu47 the COM for the non-polar methylene atoms were calculated.

LINKER INTERACTIONS

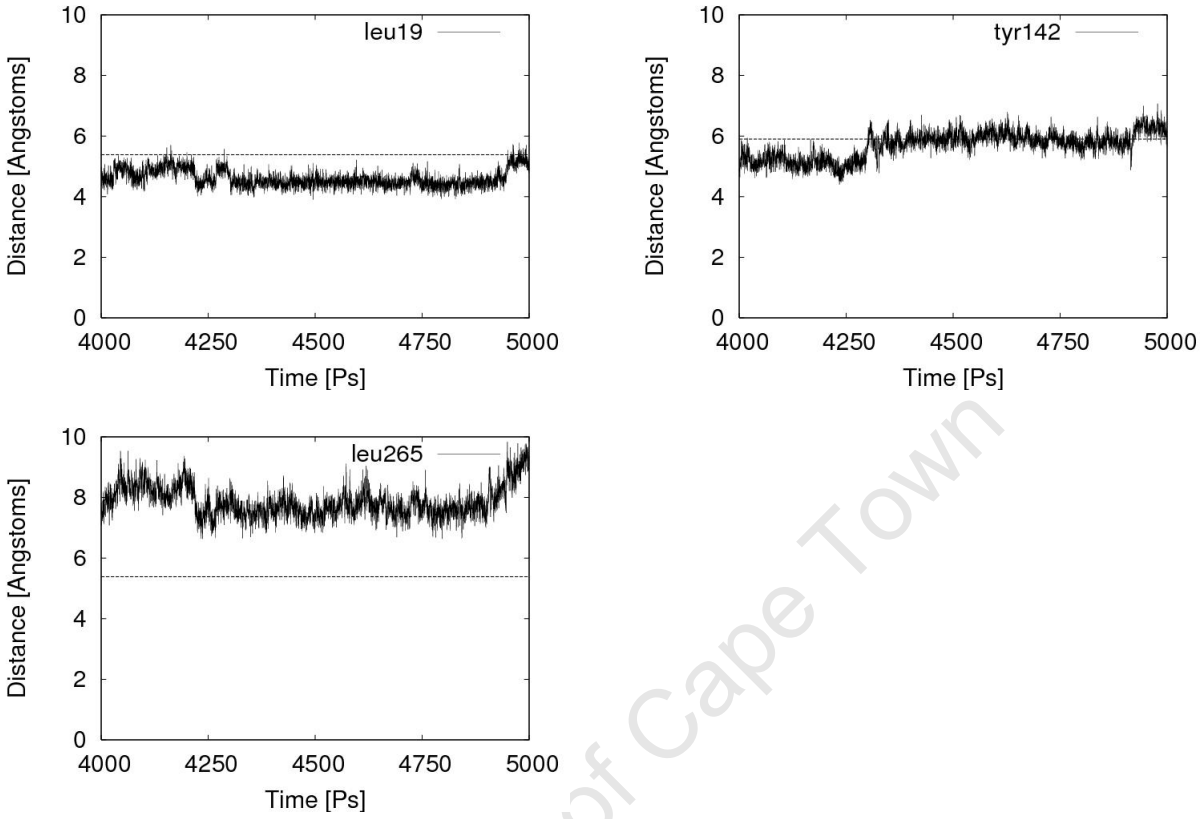


FIGURE D-14 Distance between COM of the alky linker of 1d and the amino acid sides chain indicated in the legends.

INTERACTION OF THE PLUMBAGIN METHYL GROUP

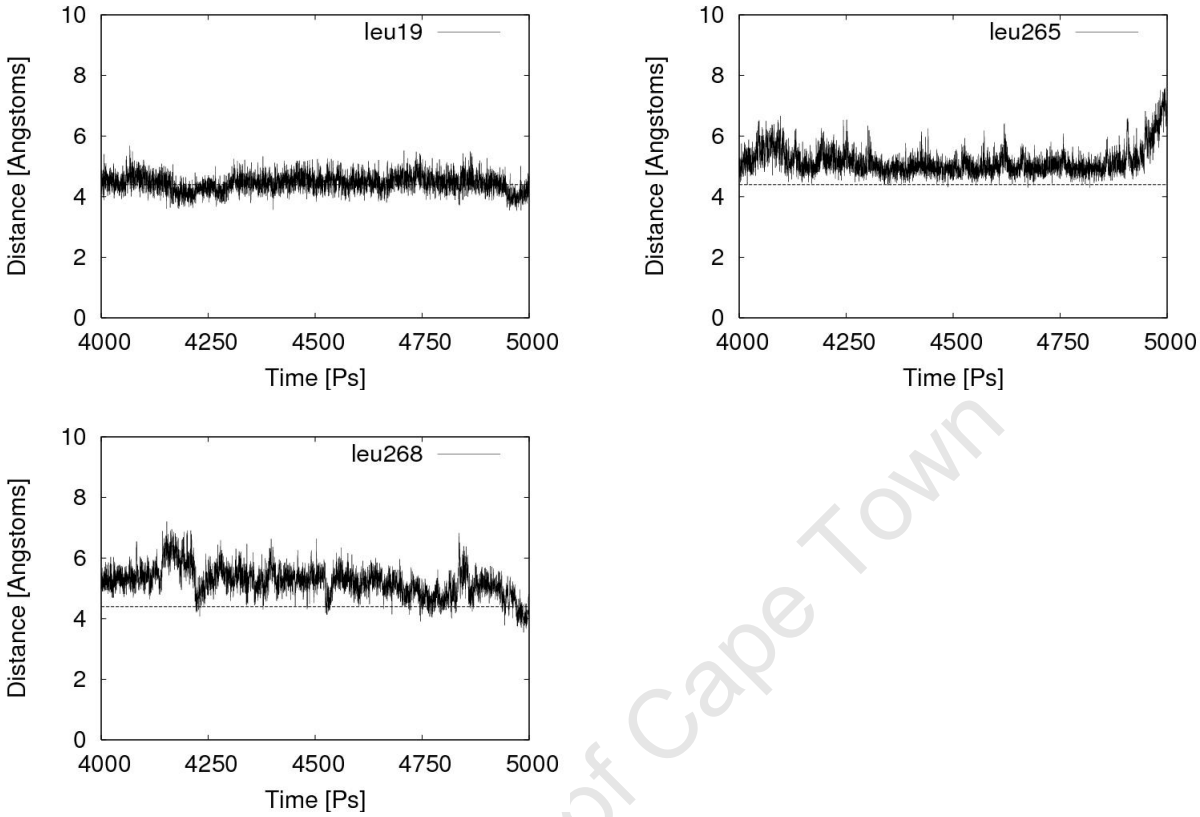


FIGURE D-15 Distance between COM of the plumbagin methyl group and the amino acid side chains indicated in the legends.

INTERACTIONS OF THE PLUMBAGIN AROMATIC ATOMS

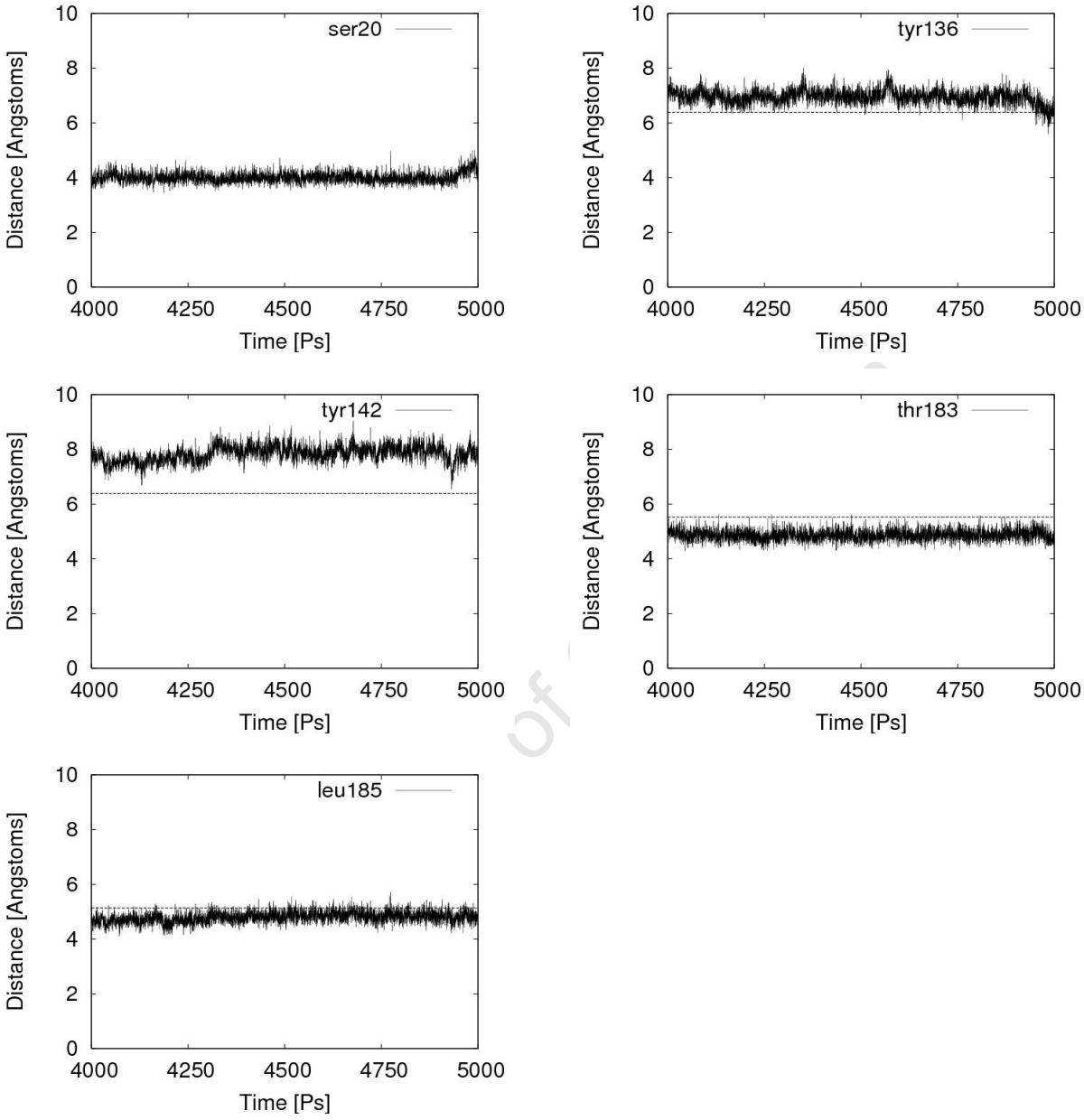


FIGURE D-16 Distance between COM of plumbagin aromatic atoms of 1d and the amino acid side chains indicated in the legends.

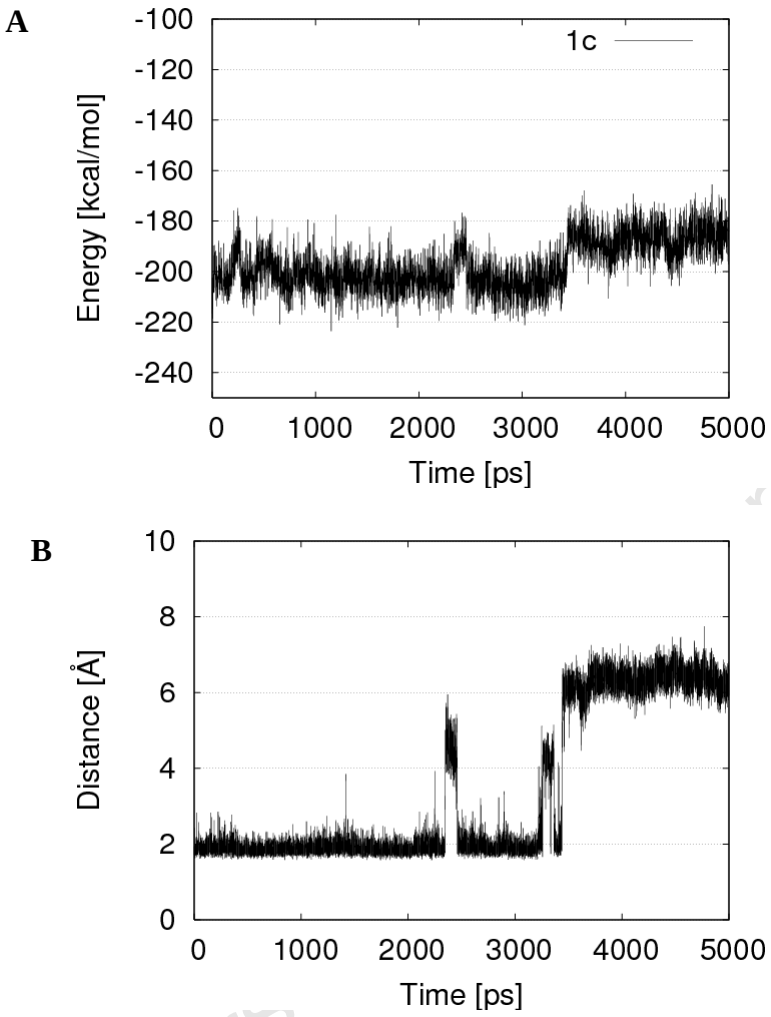


FIGURE D-17 (A) Total IE of 1c with all atoms in the system, and (B) distance time series for the 8′-C-hydroxyl hydrogen and the oxygen of the water molecule bridging Asp137 and Thr183. Comparison of the figures shows the correlation of decreased IE with increased distance.

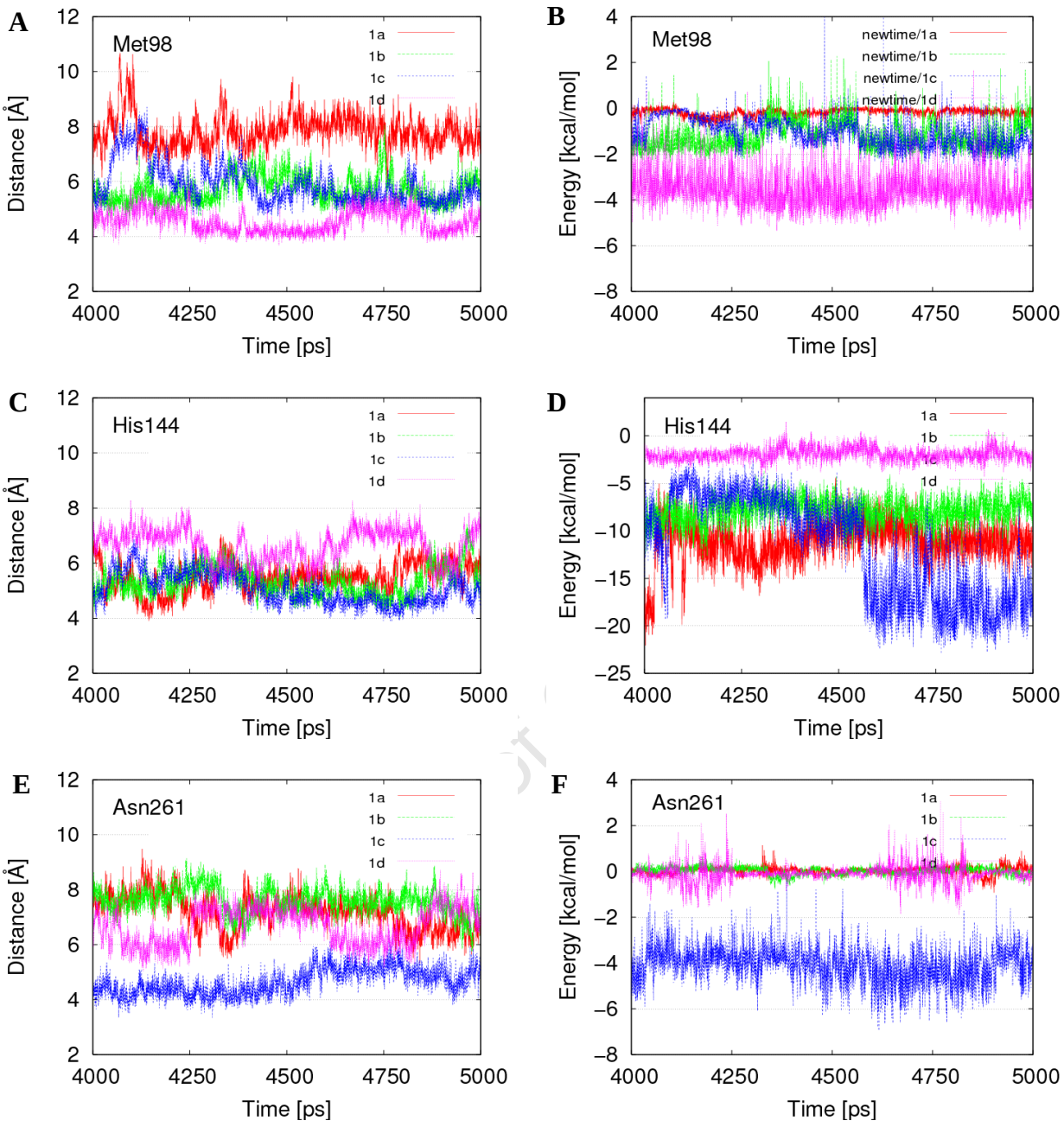


FIGURE D-18 **(A, C & E) Time series for the distance between the COM of the phenylthio groups of the inhibitors and the amino acid side chain specified. (B, D & F) IE of the amino acid residues with the inhibitors. Hydrogen bond analysis showed that the decrease in IE of 1c with His144 corresponded with the formation of a hydrogen bond between the C-6 hydroxyl hydrogen and His144.**

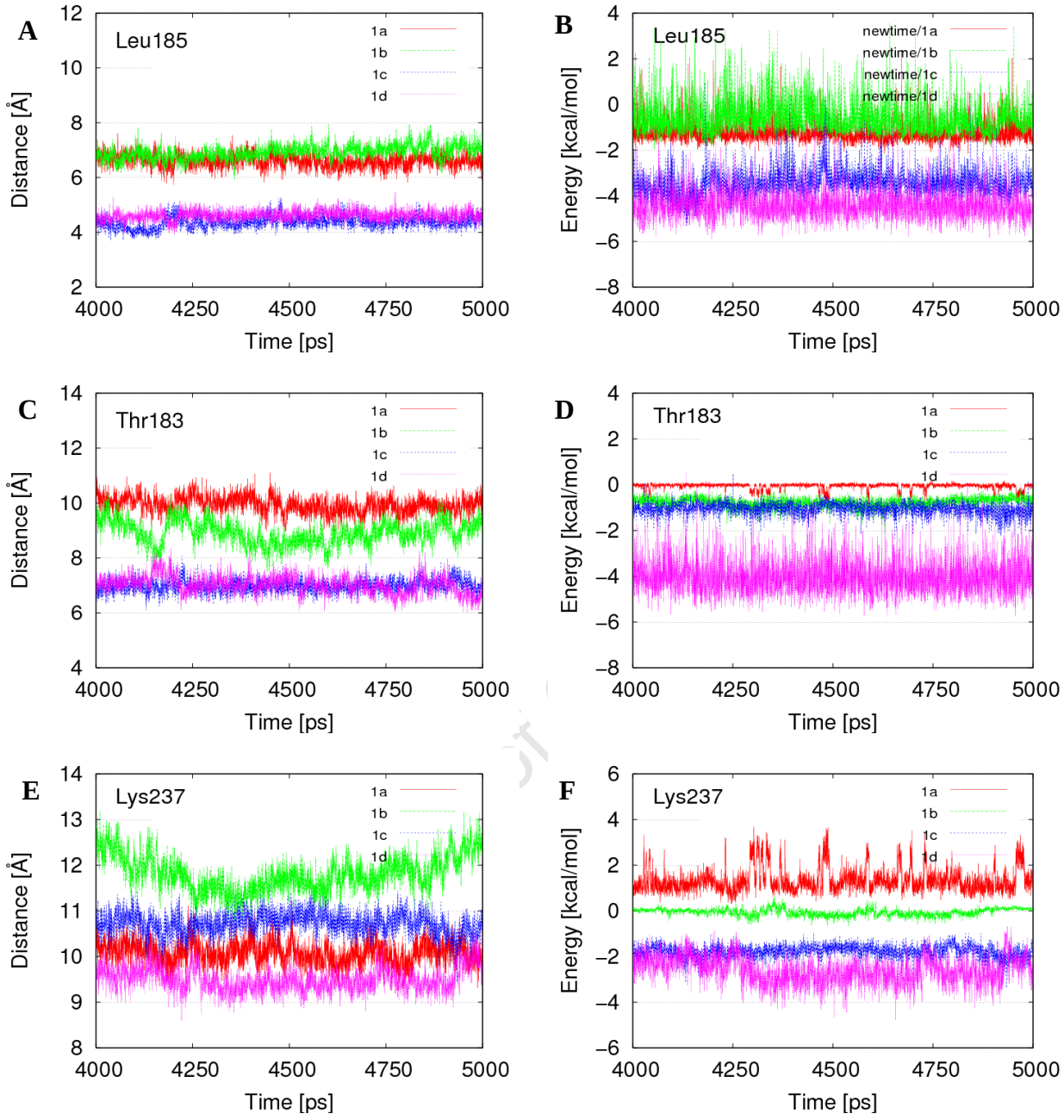


FIGURE D-19 (A, C & E) Time series for distance between the COM of the plumbagin greoups of the inhibitors and the amino acid side chain specified. (B, D & F) IE of the amino acid residues with the inhibitors.

APPENDIX E: RMSD TIME SERIES FOR REVISED MSHB-INHIBITOR CONFORMATIONS

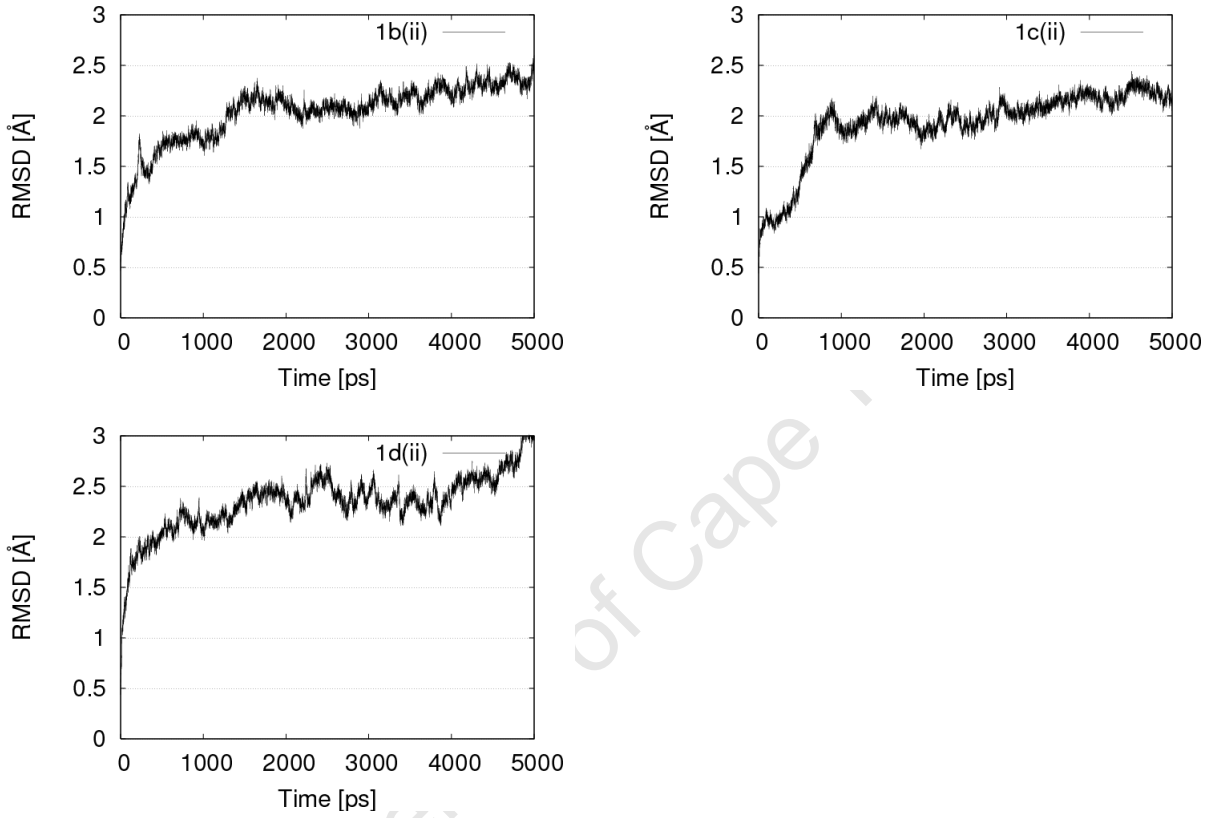


FIGURE E-1 RMSD time series for inhibitors with the revised conformation of binding. The RMSD was calculated for all amino acid residues which had an atom within 8Å from the respective inhibitor specified in the legend.