Dynamics of ligand-protein interactions - impact on drug discovery

Outi Katriina Kämäräinen

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

School of Molecular and Cellular Biology

Astbury Centre for Structural Molecular Biology

June, 2017

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement

Kauas on pitkä matka.

Old Finnish saying

Acknowledgements

There are numerous people that, one way or another, contributed to this work over the years, to which I would like to express my gratitude.

Firstly, I would like to say a very big thank you to my supervisors Sarah Harris and Anastasia Zhuravleva for their support and encouragement. It has been a great pleasure to discuss science. You have been great role models and a source of constant motivation.

Special thanks to Geoff Holdgate, who encouraged me along the way and showed the ins and outs of ITC. Steve Homans, who initially came up with the project proposal and for showing interest in the progress over the years. David Reha for the help with the quantum mechanical calculations. Alan Cooper for providing a copy of the 'Hitchhikers guide to thermodynamics'. Richard Henchman for the insightful discussions about the effect of water.

I would also like to thank the Physics IT support for their help during various minor computer crisis, as well as the Leeds University Supercomputing and NMR facility members. I wish to acknowledge EPSRC and AstraZeneca for the funding for the project and National High Performance Computing Facility and Hartree centre for the supercomputing time received and the National Institute of Medical Research for the use of the NMR magnets.

I would also like to thank my fellow lab mates, both in Physics and Biology, for the stimulating discussions and for the many cups of tea enjoyed with or without bacon sandwiches.

Last but not least, I would like to thank my family and friends for their love and support over the years.

Abstract

Introducing a new drug to market is a lengthy and expensive process (typically 10-15 years and $1.7 billion). Better understanding of how and why a drug molecule binds to a target and what changes in the atomistic structure and chemistry could improve the binding affinity and shorten the process. In addition to structure-based approaches, the role of thermodynamics and molecular motions in binding selectivity and efficiency have attracted increasing attention. Whilst calorimetric methods can quantify total free energy and entropy change, it is difficult to estimate contributions from the different components of entropy, one of the largest unknowns being the magnitude of the configurational entropy. Molecular dynamics (MD) simulations of the drug and target protein can provide more details of the different atomistic movements contributing to the total entropy change, thus potentially providing valuable clues for lead optimisation.

In this study we use the well characterised N-terminal domain of the Hsp90 chaperone protein as a model system to study the changes in conformational flexibility (configurational entropy) upon binding of small molecule inhibitors using MD simulations, NMR and ITC. We show that the two inhibitors studied cause different changes in the protein dynamics. These effects were seen with NMR relaxation dispersion methods and with MD but the dynamic changes however are not reflected in the global ITC parameters. Here the water is assumed to have a dominating effect in the overall entropy change.

However, as some Hsp90 clients have been shown to preferentially interact with only one conformation of the protein, we propose that the changes seen with NMR and MD could be of interest for drug design. Manipulating the dynamics by small molecules could favour interaction with a subset of client proteins, without affecting the interaction of others, all together providing specificity and potentially allowing to design an 'ideal' drug that only prevents the folding of 'bad' cancer related proteins without affecting Hsp90 functions in the normal cells. As the MD simulations also reflect these dynamic changes, we propose that simulations could be also used as a screening tool for selecting which inhibitors could be taken for further development in the lab.

Table of Contents

List of Tables

List of Figures

Abbreviations

ΔG°	change in free energy
ΔH°	change in enthalpy
ΔS°	change in entropy
ε	extinction coefficient
BAT	Bond Angle Torsion
BMRB	Biological Magnetic Resonance Bank
CPMG	Carr-Purcell-Meiboom-Gill
CPU	Central processing unit
Cryo-EM	Cryo-electron microscopy
CSP	Chemical shift perturbation
CTD	C-terminal domain
GAFF	General AMBER force field
GPU	Graphical processing unit
Hsp90	Heat shock protein 90
HSQC	Heteronuclear single quantum coherence
ITC	Isothermal titration calorimetry
MD	Molecular dynamics
NTD	N-terminal domain
PC	Principal component
PCA	Principal component analysis
PDB	Protein data bank
PME	Particle mesh Ewald summation
QHA	Quasi-harmonic approximation
RESP	Restrained electrostatic potential
RMSD	Root mean square deviation
SASA	Solvent accessible surface area
TROSY	Transverse relaxation optimized spectroscopy
VdW	Van der Waals

Chapter 1
Introduction

1.1 Background

Molecular recognition is key to all cellular processes, from signal transduction to DNA replication, and it has been studied widely over the last few decades. Thanks to advances in X-ray crystallography, Nuclear Magnetic Resonance spectroscopy (NMR), and electron microscopy techniques, we have a wealth information on different biomolecular structures, accompanied by functional details. However, proteins are not simply static objects within a cell, but are in constant motion and, as a result, they can adopt and populate ensembles of different conformations. It is accepted that these conformational dynamics do not just present minor fluctuations around an average native structure. Instead, the changes to the structure, and also to the equilibrium population of conformations that a protein adopts, can affect the function and choice of binding partners [1-3]. Understanding these "wiggles and jiggles" is the key to both understanding life at the molecular level and how it could be manipulated by pharmaceuticals. Here, the ability to fine tune the equilibrium population by designing small molecule inhibitors that only interact with a particular subset would clearly be of much interest.

In this work, we set out to explore the dynamic landscape of proteins further and to specifically investigate how small molecules may change protein dynamics. Below, we describe the underlying thermodynamics that is vital for understanding molecular recognition. This is followed by a review of the methods used to access the thermodynamic information, as well as a description of the model system used for this study.

1.2 Thermodynamics and drug binding

The process of a drug binding to a protein has traditionally been described by lock and key mechanisms, where the shape and chemical composition of a

ligand determines whether interaction with the protein will take place. Whilst it is true that both shape and bonds do have an effect to binding, changes in the structure of binding partners also play a role. Although the connection between thermodynamics and structure is not well understood, it is better to use thermodynamics to describe the binding interaction, since this also takes into account the flexibility of the system [4, 5].

If thermodynamic parameters are used to describe the binding interaction, the change in the free energy (ΔG) dictates whether an interaction will take place, *i.e.* whether a drug will bind to the protein target or not. Here, the more negative the free energy change, more readily the reaction will happen. The overall free energy change is made up from contributions from chemical interactions that arise from for example hydrogen bonding, and ionic interactions between the drug and the target. These changes are described by the enthalpy (H) term. The second contributor to the total free energy change is the entropy (S) term, which describes changes in the dynamics of the system. Both the enthalpy and entropy terms include contributions from water, the target protein and the drug molecule, as shown in Figure 1.1 [6].
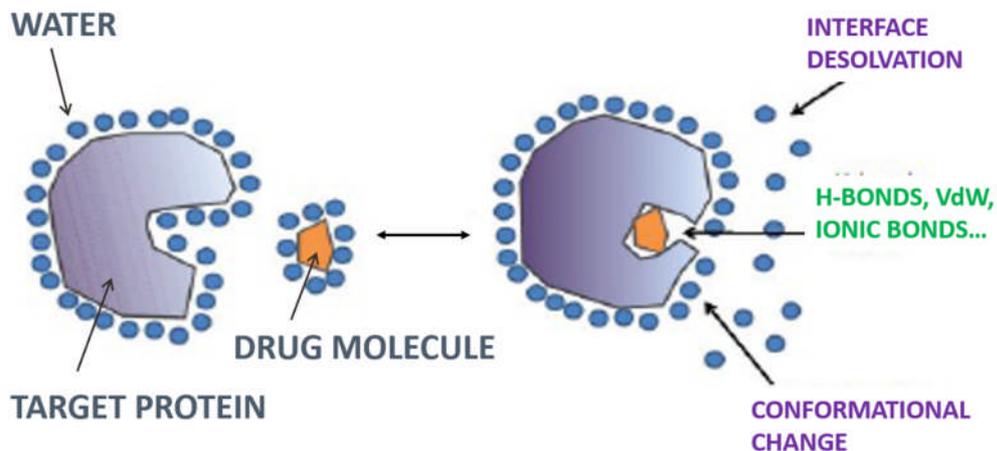
Figure 1.1: Thermodynamics plays a key role when the ligand binds to the target protein. To determine the overall binding free energy, contributions from water molecules (interface desolvation) need to be taken into account, as well as the chemical interactions between the protein and the drug, that are encompassed in the enthalphy (ΔH) term. Changes in the shape of both the protein and the drug also play a role in the overall binding free energy change, and these are described by the entropy (ΔS) term.

The free energy change (ΔG°) upon drug binding can thus be described by Gibb's free energy formula, where ΔH° is enthalpy, ΔS° entropy and T temperature:

$$\Delta G° = \Delta H° - T\Delta S \qquad\qquad (eq.1)$$

Alternatively, free energy can also be calculated in the following way if the association constant (K_a) is known:

$$\Delta G° = -RT \ln K_a \qquad\qquad (eq.2)$$

where R is the gas constant and T temperature.

It should be noted that in both cases, the changes in free energy (as well as and enthalpy and entropy) calculated by equation 1 are relative to standard conditions, as shown by the ° superscript. This means that any changes observed are relative to those measured using standard conditions, which are 1 atm pressure, a temperature of 298 K, and 1 M reactant (protein and ligand) concentrations [7]. Thus, any positive or negative changes per se lack physical significance, because if different conditions were chosen, the zero-point would change.

Whilst the changes in free energy will determine the binding affinity, for drug design, the thermodynamic descriptions that include the changes in the enthalpy and entropy are of more interest. The entropy and enthalpy terms help to understand why the ligand-protein interaction happens, and how the ligand could be manipulated to make it a 'better' drug molecule. The enthalpy term reflects the specificity and strength of the molecular interactions between the ligand and target. These include hydrogen bonds, electrostatic and van der Waals interactions. The change in enthalpy also includes contributions from water, as the bonds between the solvent (water molecules) and the solute (ligand and protein) may be broken or formed upon complex formation [8]. The entropy term, which reflects the changes in dynamics, similarly include contributions from the solvent, as well as from the protein and ligand. The solvent reorganisation, where water molecules are released to bulk solution from the ligand and protein surfaces upon binding, is often a major contributor to the overall entropy change. The solute contribution to the entropy change is a combined effect from changes to translational and

rotational motions (*i.e.* the overall tumbling) of the molecules, as well as changes in the intra-molecular structure resulting from bond stretching and angle bending [9].

A lot of effort has been made to understand and assign values to the different parameters that govern the thermodynamics of drug and ligand binding over the years as well as assigning values to the parameters [10]. These include attempts to isolate and estimate the individual contributions functional groups and H-bonds make to the binding energy. For example, H-bonds have been calculated to contribute between 2-10 kcal/mol to the binding energy, depending on the functional groups involved in the binding and whether they are buried or solvent exposed [11]. Functional group contributions to drug-receptor interactions have similarly been estimated. In general the more electronegative the group, the stronger the binding energy; for example an OH group has been estimated to contribute between 2.5-4 kca/mol to the free energy of binding compared to 3.2-4 kcal/mol for C=O group or 0.8-1.8 kcal/mol for nitrogen [12, 13]. Similarly, energetic penalties resulting in the loss of the rotational and translational degrees of freedom have been estimated [10, 14, 15]. These estimates have subsequently been used to build models to estimate the overall free energy upon binding. For this, each physical process, from restriction of rotations to burial of hydrophobic groups, has been assigned a value and a factor by which they increase the binding energy [16]. However, estimation of the free energy is not as simple as adding up the various constituents. There are many examples, where attempts to increase affinity by focusing on the the enthalpy component, for example by addition of H-bond motif, carry an entropic loss due to structuring of both the ligand and the binding site [17]. On the other hand, some binding interactions have been demonstrated to be more than sum of their parts. For example, studies with thrombin ligands found that addition of a H-bond (enthalpic effect) and increasing the size of a hydrophobic group (thus affecting entropy), had in fact a co-operative effect. Here the affinity gain from the two modifications to the ligand was larger than either on its own [18]. This additive affect has also been seen with fragment based drug design, where the affinity of a larger molecule (where the individual fragments are linked) can be greater than the combined affinity of either fragment. The additive affect in the case of thrombin

was suggested to depend on changes in the dynamic properties of the molecule [19].

Generally, the estimations of binding energy as described above do not take into account any changes to the target protein. In the past, the enthalpy term describing the chemical interactions has received more attention in drug design efforts. However, in recent years the static 'lock and key' mechanism used to describe drug and protein association has been adapted to include the changes in the conformations of the target protein as well as yielding the induced fit and conformational selection models. In the induced fit model a ligand binds to an inactive form of the protein and the interactions between ligand and protein upon binding induce the protein to change to an active form. The conformational selection model assumes that the protein exists in active and inactive forms and the ligand binding to the active form subsequently shifts the existing equilibria [20]. To understand these protein dynamics, the entropy term, which describes changes in the different conformational states that a system can adopt, is the key [21]. This will be the main focus point for this study.

1.3 Experimental approaches to probe the protein dynamics

It has been said that 'Heat does not come in many colours' [22]. Whilst all the thermodynamic parameters (ΔG, ΔH and ΔS) can be determined analytically with techniques such isothermal titration calorimetry (ITC), the global entropy value obtained by ITC includes contributions from both solvation and the changes in the macromolecular structure. However, understanding what the global value means, in terms of the dynamics of individual components of the binding system that determine configurational entropy at the molecular level, is not possible. This is due to the fact that it is not possible to distinguish what percentage of the change is due to the protein, or due to the drug ligand, and also what effects to the structure of the protein these changes may have.

Apart from ITC, many biophysical techniques can be used to get clues about thermodynamics of binding and contributions from conformational re-arrangements, but like ITC, most provide limited resolution either in terms of

detail of dynamic timescales or lack of atomistic detail. Techniques, such as surface plasmon resonance, allow the determination of the binding kinetics, but lack in structural detail [23]. In principle, hydrogen-deuterium exchange mass spectrometry allows the determination of both binding kinetics and some details of the structure of the protein. However, the structural information obtained from this technique is of very low resolution, and falls into the same category of 'rough shape and size' as the structures obtained via neutron scattering, or small angle X-ray scattering [24-26]. With correctly chosen labelling, or with the aid of mutational studies, fluorescent microscopy techniques can give information about structural re-arrangements that occur upon binding. However, the structural detail is limited to those areas adjacent to the fluorescent labels [27]. Cryo-electron microscopy (cryo-EM) and X-ray crystallography on the other hand allow determination of atomic structures, and can provide some detail of the dynamics. Cryo-EM structures have been used to give details of the dynamics, even back when the structures were 'fuzzy'. This is because the populations of the different conformations can be extracted from large datasets of imaged particles. Also, given the recent major advances in the resolution of cryo-EM structures, it has become possible to get atomistic detail of the structures, although the technique is suitable only for large protein systems [28]. X-ray crystal structures of proteins on the other hand are by definition rigid, but dynamic regions can still be inferred by using the per atom calculated B-factors that contain information about the thermal fluctuation of the atom in question. A small value for the B-factor describes rigid atoms, larger numbers suggest some fluctuations [29]. However, the dynamic information, in terms of timescales, obtained via the cryo-EM and X-ray crystallography techniques, is not very detailed. To fully understand the ligand-drug interaction, detailed information of both the atomistic structures and the dynamics including the timescales they occur are needed.

Protein motions are thought to occur in timescales spanning from picoseconds to seconds. This is to account for the bond vibrations to protein domain movements and for protein folding, as shown in Figure 1.2. To understand the dynamics of proteins, techniques which can span these timescales and provide atomic resolution are needed. Here one should look into nuclear

magnetic resonance spectroscopy (NMR) and all atom molecular dynamic simulations (MD).
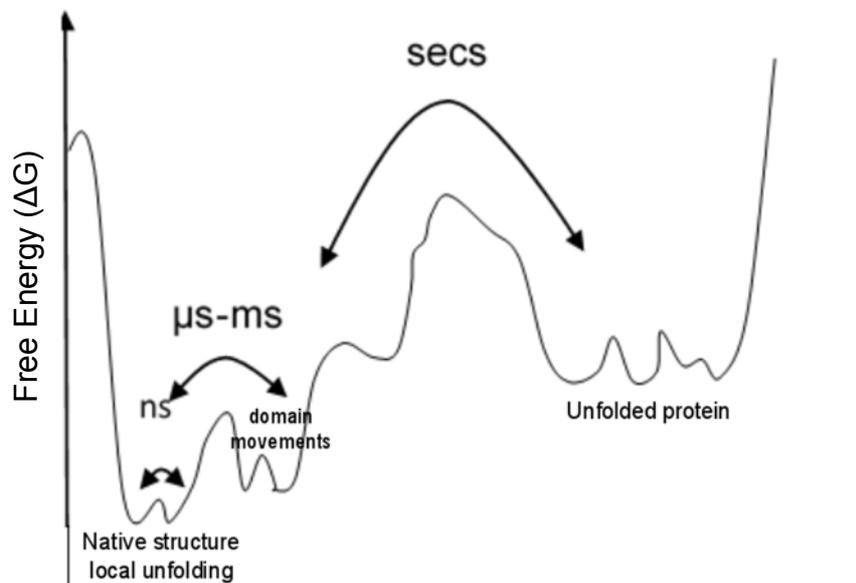
Figure 1.2: Protein motions cover the local bond vibrations to larger scale domain re-arrangements and these events take place in different timescales from nanoseconds to seconds. The changes in structure require different magnitudes of free energy change and can be imaged as peaks and troughs in the free energy landscape. Local bond vibrations can be thought to reside in single energy well, whereas for larger domain movements or protein folding/unfolding events which require larger energy changes similarly require more time to cross energy barriers (modified from [30]).

Both NMR and MD can provide information on different timescales. For classical MD simulations, the lower time limit is related to picosecond bond vibrations and whilst using special computer infrastructure millisecond simulation times have been achieved, normally the simulation times are currently in realm of hundreds of nanoseconds [31]. Nevertheless, with longer simulation times and use of replica simulations, the energy landscape can be explored adequately to observe dynamics resulting from ligand-protein interactions. NMR on the other hand is very powerful technique, as with the right choice of isotope atom labelling and pulse sequence, the dynamics ranging from the picosecond bond vibrations to protein folding that can be measured in milliseconds can be observed, as seen in Figure 1.3 [30]. NMR experiments can also measure slower processes from seconds to hours.

As both the atomistic structure and details of timescales are needed to further the understanding of dynamics and configurational entropy of macromolecules, we will use MD simulations and NMR techniques to study the protein and ligand dynamics. The pharmaceutically interesting protein, heat shock protein 90 (Hsp90), which is an active cancer drug target is used as a model system to be studied by NMR and MD techniques. Further, ITC measurements will be conducted to get an idea of global *vs* local entropy effects.

Figure 1.3: Atomistic scale techniques for probing protein dynamics at different timescales. Protein motions range from local bond vibrations to global domain re-arrangements and folding events. These different motions happen in various timescales, from nanoseconds to seconds. To probe these motions, NMR is a very powerful technique as different NMR techniques can examine the very fast, as well as the slow, events. Whilst different computer simulation techniques can access similar timescales, coarse graining approaches are required to measure the slower time scales.

1.3.1 Dynamic Hsp90 is used as model system to probe the role of protein dynamics in ligand and drug interactions

The molecular chaperone Hsp90 is one of the most abundant proteins in a cell. It is involved in maturation of large number of client proteins during the later stages of protein folding [32]. To date, hundreds of diverse clients of Hsp90 have been identified, ranging from kinases to nuclear receptors. As many of the clients are oncoproteins, Hsp90 is also an active cancer drug target [33, 34].

Given that Hsp90 is vital for cell function, it is perhaps slightly surprising to find it as a drug target. The pharmaceutical industry took note of the chaperone during the 1990s after a natural product geldanamycin was first found to revert cancerous phenotype in cell screens. Follow up studies using pull down assays with geldanamycin found that it interacted with a 90 kDa protein, which turned out to be Hsp90 [35]. After nearly two decades, there is a wealth of Hsp90 data, including over 200 structures deposited into the protein data bank on human Hsp90. A recent review found 280 published Hsp90 ligands, with thirteen inhibitors, that are undergoing clinical trials [36-38].

Hsp90 is a highly dynamic multi-domain protein made up of a N-terminal nucleotide binding domain (NTD), a middle domain, which has been implied to play a role in substrate and co-chaperone interactions, and a C-terminal domain (CTD) which is used for dimerisation. Hsp90 forms homo-dimers and fluctuates between wide variety of structurally distinct states [39]. Some of the early crystal structures defined the open and closed states of Hsp90, where the conformation change was regulated by ATP binding and hydrolysis. Here, in the apo-form, the Hsp90 adopts a V-shaped open conformation, where the dimer connections are formed via the C-terminal domains (Figure 1.4) [39]. Upon ATP binding to the NTD, the lid segment closes the binding pocket and this slowly leads to the whole chaperone changing from an open V-shape to a twisted closed conformation, where the two N-terminals also make contact upon strand exchange. The ATP binding also has an effect on the NTD and middle domain interface. Here, a highly conserved arginine residue of the catalytic loop of middle domain interacts with the ATP bound to the NTD and thus the middle domain contributes to the ATP hydrolysis. After ATP hydrolysis, and subsequent ADP release, the Hsp90 returns to the original open conformation [39-42]. The open and closed forms as defined by early crystal structures are not the only conformations the Hsp90 adopts. SAXS and EM studies have demonstrated that ATP binding and hydrolysis only shift the equilibria between a pre-existing set of conformational states, which range from open to closed structures with various intermediates and open structure where the subunits are much further apart [41, 43]. Interestingly, although the dynamic cycle between the open and closed forms of Hsp90 has been well

established, mutational studies of the full length protein with non-functional nucleotide binding domain found that both open and closed forms were also populated in the absence of the nucleotide. This suggests that the chaperone population exists in a dynamic conformational equilibrium [44].
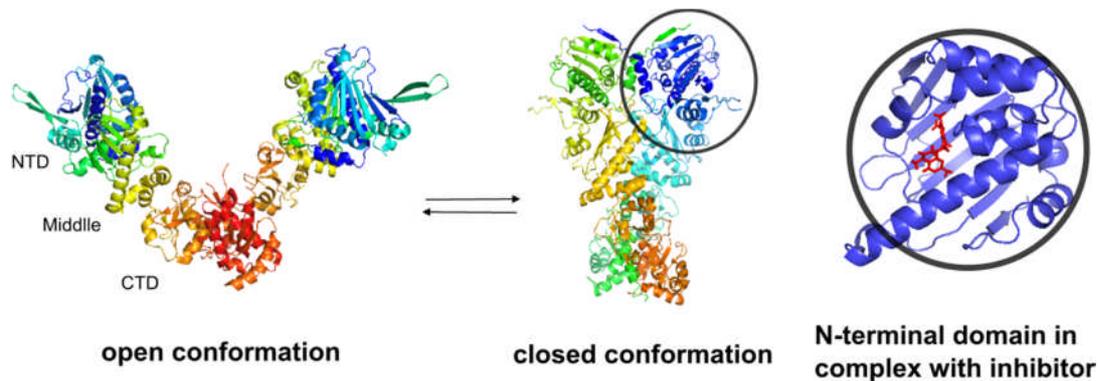
Figure 1.4: Hsp90 undergoes large conformational re-arrangements upon ATP binding and hydrolysis from V-shaped open conformation (*E.coli* Hsp90 structure – PDB:2IOQ) to tightly packed closed conformation (Yeast Hsp90 – PDB:2CG9). The chaperone is made up of three domains, the N-terminal domain (NTD), the middle domain (Middle) and the C-terminal domain (CTD). The N-terminal domain, where the inhibitors bind to, is depicted in blue on the right of the figure.

The highly dynamic cycle that Hsp90 undergoes has an effect on both the co-chaperone interactions as well as client binding. The Hsp90 co-chaperones have been shown to interact with, and stabilise, different conformational states during the ATP driven open-closed cycle [45]. Recent studies have also suggested that different pools of Hsp90 may be involved in different functions or pathways. Here, pull down assays have shown that the Hsp90 inhibitors, whilst binding to the same N-terminal pocket, only interact with a small portion of the total cellular Hsp90. The inhibitors also cause different cellular phenotypes to occur, suggesting that they have different biological effects [42, 46-48]. For client protein interactions, no common sequence or motif has been identified responsible for the interaction with the Hsp90. The current belief is that the different conformational states of Hsp90 are associated with different co-chaperone complexes, and that these interact with different clients. The chaperone conformation will have different effects on the client

folding/misfolding [39, 42, 49]. This conformational equilibrium is interesting for the pharmaceutical industry; if particular clients interact with only a certain conformation of Hsp90, it may be possible to selectively inhibit only a certain pool of cellular proteins. This could help with toxicity issues associated with the Hsp90 inhibitors.

Hsp90 has been a cancer drug target for over a decade, and most drugs target the N-terminal nucleotide binding domain. Two early inhibitors that have been the subject of many studies were geldanamycin and radicicol. A large number of inhibitors from different chemical classes have been added to the growing list of Hsp90 inhibitors [37]. Both computational and NMR studies have shown that nucleotide (ADP/ATP) binding causes changes in the N-terminal domain dynamics of the chaperone, which are likely to be linked to the open-closed states of the full length protein [50, 51]. It has also been demonstrated that the changes in the N-terminal domain observed upon ligand binding also affect the conformation of the middle and C-terminal domains in the full length protein [52]. To examine whether small inhibitors also cause differences to the chaperone dynamics, rather than merely inhibit the function by blocking the ATP hydrolysis step, we selected two different ligands from a panel of Hsp90 inhibitors to carry out biophysical characterisations using NMR and computer simulations. The two ligands selected for the studies were a large 17-DMAG molecule [37], which is a more soluble derivative of geldanamycin, and a small nucleotide mimic GVK0153 (see Figure 2.1 in Chapter 2 for structures). The N-terminal domain of the human Hsp90 was used for the studies, as its size is still amenable to both NMR methods and longer timescale MD simulations. For the inhibitors, both molecules studied have been removed from active drug research due to issues with *in vivo* toxicity. Since the Hsp90 is used here purely as a model system to study dynamics, rather than to develop new inhibitors, these side effects do not matter for this study.

1.3.2 Isothermal titration calorimetry

NMR and molecular dynamic simulations are used in this study to estimate configurational entropy change. We also carry out isothermal titration calorimetry (ITC) experiments, that allow the direct measurement of the changes in enthalpy (H), as well as stoichiometry (n) and affinity (K_a) for biomolecular processes in aqueous solution [53] [22]. These parameters allow the calculation of free energy, and entropy values, using equations 1 and 2

(shown in Section 1.2 and below). Thus, all thermodynamic parameters can be obtained from a single experiment.

$$\Delta G = -RT \ln K_a = \Delta H - T\Delta S \qquad \text{(eq. 1 \& 2)}$$

The ITC instrument has two cells, a reference cell filled with the assay buffer and a sample cell containing the macromolecule of interest to which the ligand is going to be titrated. A schematic of the instrument is shown in Figure 1.5. The reference and sample cells are connected together with thermo-coupled circuits that detect any temperature differences between the cells caused by the binding reaction of sample cell upon ligand titration. Depending on whether the reaction in the sample cell is endo- or exothermic, the resulting sample cell heat change causes power to either be applied, or reduced, to maintain identical temperature between the two cells [54]. The changes in the feedback power applied to the sample cell is measured, and recorded. This gives the total heat change per injection, which can be plotted as a power-*vs*-time plot to obtain the values of ΔH° and K_a.

This image has been removed by the author of this thesis for copyright reasons

Figure 1.5: Schematic of Isothermal calorimetry instrument. The sample cell, into which the ligand is titrated, is thermo-coupled to a reference cell. Depending on the reaction, power is applied or withdrawn from the sample cell to keep the temperature constant between the reference and sample cells. Figure adapted from [53].
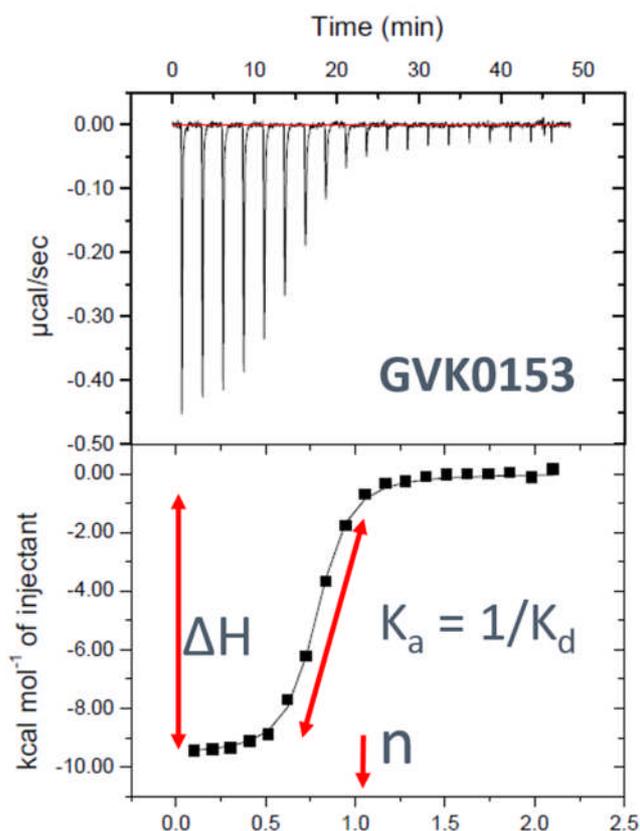
Figure 1.6: Example of data from an ITC experiment, where GVK0153 inhibitor was titrated into Hsp90. The top panel shows the power applied to the sample cell to maintain constant temperature between the reference and sample cells, thus reporting on heat changes. The bottom panel shows the integrals of the peaks, together with a line of best fit that is used to estimate ΔH, K_d and stoichiometry (n).

As seen on the top panel of Figure 1.6, the changes in power are generally large during the first few injections due to the excess of macromolecules, and these maximum peaks correspond to enthalpy of binding ($\Delta H°$). The changes in the heat become smaller with further titrations as the binding sites are being filled. Eventually, a proportion of the titrated ligand will remain free in the solution. The slope of the middle section of the plot, between the initial large changes and final small changes in heat, can be used to estimate K_d. This is the dissociation constant, and the inverse of K_a. Finally, if the concentrations of the macromolecule and the ligand are known accurately, the stoichiometry (n) can be determined from the graph [53].

Whilst all the thermodynamic parameters of interest can be determined by ITC, the values obtained are global values and conformational entropy

contributions cannot be separated from the effects of water and rotational and translational components of the macromolecules. Also, ITC experiments are sensitive only to certain affinity windows. Very strong interactions between the ligand and target produce too steep a curve without sufficiently many points to determine the K_a (*i.e.* slope). *Vice versa*, very weak interactions produce flat curves, which make slope determination similarly imprecise. Nevertheless, ITC is a popular method, and the global entropy values can give guidance as to whether the magnitude of the configurational entropy obtained from other techniques is the overall driving force of the binding reaction.

1.3.3 Nuclear Magnetic Resonance Spectroscopy (NMR)

It is close to four decades since the first protein structures were determined by NMR. The technique has since been developed with many other applications in mind. This includes NMR being used as a tool to study structural ensembles, to identify ligand binding and protein-protein interaction sites and to probe conformational dynamics, as well as binding affinity to name a few applications [55, 56]. Proteins are not rigid, but populate different conformational states. What makes NMR a particularly powerful technique is the possibility of obtaining not only information about the structures a macromolecule adopts in solution, but also the possibility of examining the dynamics that govern these structural fluctuations. This includes obtaining the global parameters for µs-ms exchange processes, such as the exchange rate constants (k_{ex}), chemical shift differences between different states, as well as populations to characterise protein dynamics. NMR also allows the characterisation of local (site or residue-specific) ps-ns dynamics. Here NMR experiments provide information about rigidity (by using order parameters, S^2) of individual bonds in the protein and the time scale of these motions [55].

NMR techniques were thus used for this study to probe the dynamic behaviour of the Hsp90-NTD, when it is in complex with different inhibitors. This included obtaining the assignments for the ligand bound Hsp90-NTD, and using chemical shift perturbation (CSP) analysis to look into changes in structure and/or dynamics of the protein upon ligand binding. As CSP analysis only indicates that some changes in protein structure or/and dynamics has occurred, but we cannot say what exactly happened, we also used relaxation

dispersion methods to look into protein dynamics. These methods are described below.

1.3.3.1 NMR protein 2D and 3D experiments and assignments

Protein NMR relies on monitoring magnetically active nuclei (typically these include ^{1}H, ^{15}N and ^{13}C for biological macromolecules). After a labelled protein has been produced for the study, the NMR experiments often start with assignment of the protein backbone atoms.

Here, the 2D heteronuclear single quantum coherence (HSQC), or transverse relaxation optimized spectroscopy (TROSY), on backbone amides is often the first multidimensional experiment run, as it can inform on spectral quality [57]. The HSQC experiment is generally used for smaller proteins, while the TROSY is used for larger systems (14-20+ kDa), and in higher magnetic fields. Although some side chains can also be seen (Trp, Asn and Gln), NH TROSY experiments mostly report on the backbone amides, and generally one peak can be seen for each amino acid backbone residue, although there are exceptions, such as Proline residues. If the assignments of the apo protein are available, and the peak positions in the amide 2D spectra of the protein in the presence and absence of different ligands are similar enough to transfer the existing peak assignments between the spectra, NH TROSY is often used as a sole experiment for finger printing the binding interactions using ^{15}N labelled protein [58].

For protein backbone assignments, 3D triple-resonance experiments HNCACB and HNCOCACB are used [59]. These link the backbone atoms with their neighbours. As apo Hsp90-NTD assignments were available, here we used 'an incomplete set' of experiments - TROSY versions of triple-resonance experiments such as HNCA, HNcaCO and HNCOca as well as HNCO, that are often used to assign larger proteins. The HNCA/HNCOca and HNCO/HNcaCO 3D experiments link the neighbouring carbons (Figure 1.7). In the HNCA experiment, the magnetisation is passed from H to N and then to C_α and $C_{\alpha-1}$ via the heteronuclear one-bond or two-bond J-coupling between the N and C_α or $C_{\alpha-1}$. Then the magnetisation is passed back again to HN for detection, as shown in Figure 1.7. In the resulting 3D HNCA spectra, each backbone HN bond is represented by two peaks. These have the same

chemical shifts in ^1H and ^{15}N dimensions, and different chemical shifts in the ^{13}C dimension, one corresponding to C_α and the other to $C_{\alpha-1}$, which is normally the weaker of the two peaks. Similarly, for HNCOca, the magnetisation is passed from H to N and to CO and to C_α and back again for detection. The resulting 3D HNCOca is similar to the HNCA spectra, but a peak is only observed from the $C_{\alpha-1}$ residue in the carbon dimension. HNcaCO and HNCO form a similar pairing to HNCA/HNCOca, where for HNCO, a single peak in the CO dimension is seen from 'i-1' residue, and in the HNcaCO spectra the carbon dimension has two peaks, one for the residue 'i' and the other from 'i-1'. For the HNCO experiment, the magnetisation is passed from H to N and then to the CO, via N-CO J-coupling and on the carbon dimension on the HNCO spectra, the CO from the 'i-1' residue is seen. For HNcaCO, the magnetisation is transferred via H to N, then to C_α and finally to CO via C_α-CO J-coupling, and back again for detection. As the amide nitrogen is coupled both to its own C_α and that of the previous residue, for each backbone NH bond, two carbonyl groups CO_i and CO_{i-1} are seen, where the CO_i peak is normally more intense.

The information from the HNCA/HNcaCO and HNCO/HNCOca experiments can be used to connect a residue with its preceding neighbour (*i.e.* 'i' with the previous 'i-1'). Identification of the C_a or CO and preceding C_{a-1} or CO_{i-1} peaks allows the building of strips of connected residues, as shown in the bottom panel of Figure 1.7. As some residues, such as Glycine, have very typical C_a chemical shifts, these can be used to 'anchor' the strips to a particular set of amino acid sequence in the protein structure [60]. Whilst the HNcaCO and HNCOca are needed to link the residues together to build strips of linked residues for assignment, the HNCA and HNCO spectra are also useful, especially for weaker peaks, as they are more sensitive experiments.

Even if protein assignments are available, triple resonance experiments are often required to verify peak assignments and/or identify peaks for residues affected by ligand binding, as discussed in the next section. Additionally, 3D experiments (particularly HNCO) help to resolve overlapped peaks in the 2D amide spectra. In this project, TROSY, HNCA, HNCO and HNcaCO experiments are used to verify previously published assignments of Hsp90-NTD (BMRB:7003), and assign the ligand bound Hsp90-NTD spectra [61].

HNCA
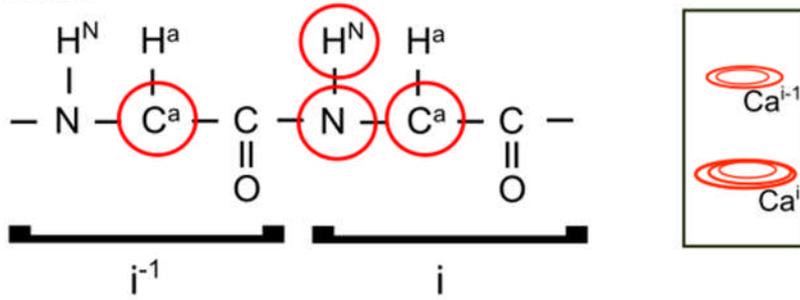
HNCO
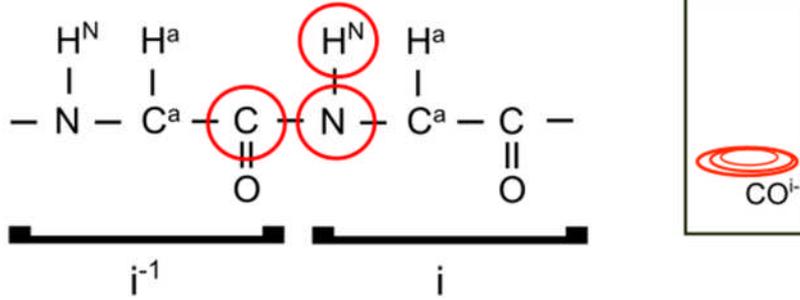
HNCACO
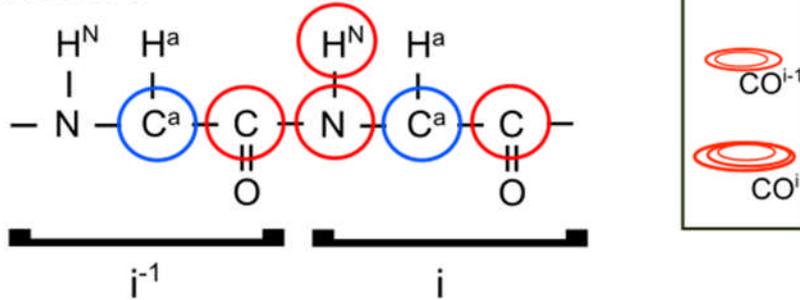
Sequential assingment using C^a shifts
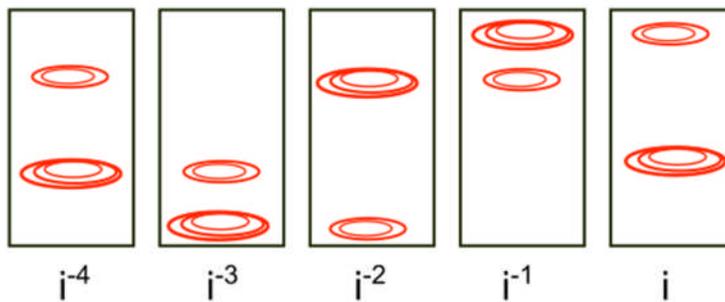
Figure 1.7: Schematic overview of the heteronuclear NMR experiments HNCA, HNCO and HNcaCO commonly used for the assignments of backbone resonances. Circles indicate the correlations that are recorded in the given experiment. Red circles represent nuclei for which chemical shift is recorded and blue circles are nuclei for which no chemical shift is recorded, but that are used for coherence transfer. The strips on the right of the figure show expected peaks and intensities in the carbon dimension of the spectra. The bottom panel of the figure shows an example of how the C_α carbon shifts can be used to link up the residues in a protein structure.

1.3.3.2 Chemical shift perturbation analysis

Chemical shifts are resonance frequencies of a magnetically active nuclei which are influenced by the nearby electrons. Thus in a protein molecule, where atoms are connected through bonds and space, chemical shifts differ even for residues with the same amino acid type. This results in differences in peak positions in the spectra. As chemical shifts are very sensitive to the electronic environment of a nucleus, even very small changes in the chemical environment can be detected. Measurements of changes in chemical shifts (*i.e.* chemical shift perturbations or CSPs) are commonly used to study protein-ligand binding. CSP measurements provide a valuable tool to characterise structural and dynamic changes in macromolecules, for example due to interactions with small molecule ligands.

The CSP analysis is used to identify site-specific differences between two protein states, for example, between the apo-protein and protein-ligand complex. Depending on the protein size, and complexity of the system under study, 2D or 3D NMR spectra are collected for the two states (apo and ligand bound), and peak positions for individual residues are analysed. If a ligand does not bind, no changes in chemical shifts are observed in NMR spectra. Detecting changes in peak positions upon ligand binding allows localisation of ligand-binding sites. This is because the protein residues involved in direct binding are likely to have changes in their peak positions. If ligand binding causes long-range structural and/or dynamic changes in the protein, perturbations in the magnetic environment of nuclei result in changes in peak positions of residues that are not directly involved in ligand interactions. If the crystal structure of the protein is known, and its NMR spectra assigned, the

observed chemical shift perturbations can be mapped onto the crystal structure. This allows us to identify residues that are directly affected by binding, as well as sites with ligand-induced long-range conformational changes [57].

Usually for small ligand binding, 2D TROSY or HSQC spectra in the apo and ligand bound form are recorded to detect chemical shift perturbations for the protein backbone amides, *i.e.* ^{1}H and ^{15}N chemical shifts. For each residue (or NH bond), the total chemical shift change ($\Delta\delta_{tot}$) is calculated using the following equation [62]:

$$\Delta\delta_{tot} = \sqrt{(\Delta\delta_H)^2 + (0.154\Delta\delta_N)^2} \qquad\qquad (eq.3)$$

where $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shift differences for ^{1}H and ^{15}N.

For larger, more complex systems, 3D HNCO spectra, that provide ^{15}N, ^{1}HN and ^{13}CO chemical shifts are often used in CSP calculations instead of 2D TROSY/HSQC to resolve peak overlap.

To analyse the chemical shift perturbations, the observed chemical shifts ($\Delta\delta$) are quantified to separate 'genuine changes' from experimental noise before mapping the changes onto the protein structure. Here a commonly used method is to calculate the standard deviation for all the observed chemical shift changes. Next, all the residues that have a very large $\Delta\delta$ (*e.g.* greater than 2 or 3 standard deviations) are excluded from the dataset to avoid biasing the distribution. Then the standard deviation is recalculated. The genuine changes can then be defined as those residues whose observed chemical shift is larger than the standard deviation [57].

1.3.3.3 NMR relaxation dispersion experiments to study µs-ms protein dynamics

The micro- to millisecond timescale dynamics are often relevant for small molecule binding events [63, 64]. The current view of proteins is that they exist in an ensemble of different states. In turn, ligand binding can shift the equilibrium of this conformational ensemble [65]. Thus, understanding the dynamics that govern the exchange between the states, as well as the structural information about the states, is needed to better understand the underlying biomolecular recognition events.

Provided that there is a difference in chemical shifts ($\Delta\delta$) between the states and exchange between states is on the micro- to millisecond timescale, different peak patterns can be seen in the spectra, depending on the exchange regime (k_{ex}) between the states (Figure 1.8).

$$k_{ex} = k_{AB} + k_{BA} \qquad \text{(eq. 4)}$$

where k_{AB} is the exchange rate constant from state A to B and k_{BA} the exchange rate constant from B to A.

A single peak is observed when the exchange rate constant between the states is significantly faster than the chemical shift difference ($\Delta\delta$) between the peaks corresponding to individual states ($k_{ex}>>\Delta\delta$) (Figure 1.8). In this case, the observed peak position can be used to calculate the populations of individual states using the following equation (provided here for state A):

$$p_A = (\delta_{obs} - \delta_A)/(\delta_B - \delta_A) \qquad \text{(eq. 5)}$$

Here p_A is the population of state A, δ_{obs} the chemical shift for the observed peak and δ_A, δ_B are the chemical shifts for states A and B respectively.

At the other end of spectrum, where the rate of exchange between two states is a lot slower than the difference in chemical shifts ($k_{ex}<<\Delta\delta$) between the states, the spectra shows two peaks. For each peak, its height is proportional to the population of the corresponding state (Figure 1.8). Between the fast and slow exchange rates falls the intermediate range, where the exchange rate constant is roughly equal to the difference in chemical shifts between the two states ($k_{ex}\sim\Delta\delta$). In this case, interconversion between state A and state B 'interferes' with the chemical shifts, resulting in 'exchange broadening' of the observed NMR peak.

In general, these differences in the peak pattern and peak width can be studied using line shape analysis. Line shape analysis can be used to estimate exchange rate constants and the population of individual states. However, for more careful quantitative analysis, relaxation dispersion measurements using the Carr-Purcell-Meiboom-Gill (CPMG) method are commonly used [30, 55].
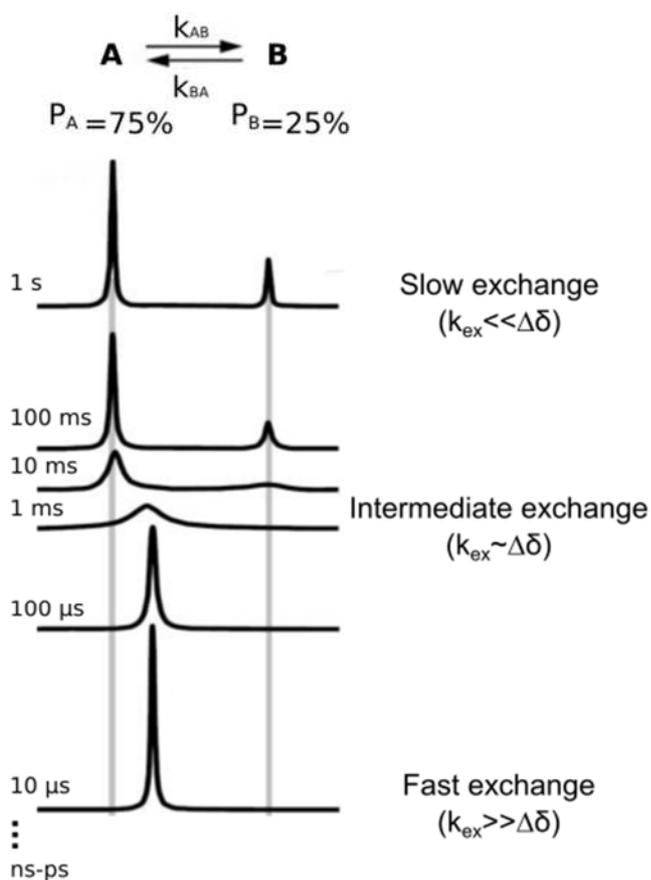
Figure 1.8: Overview of NMR dynamics experiments using the classic two-site chemical exchange example. If the exchange between the two states (A and B) is slow, two peaks are observed where the height distribution reflects the population of each state. For fast exchange, only a single peak is visible with its the position based on the population. Intermediate exchange, where the exchange between the two states is roughly equal to their chemical shift difference, results in a broad peak. (Figure modified from [55]).

The CPMG approach allows re-focusing the broad peaks caused by intermediate exchange. The basic idea behind these experiments is to quantify the effective line width of peaks as a function of number of refocusing pulses. Without refocusing, the line widths are broad due to dephasing of magnetisation caused by stochastically changing between two states (states A and B). In the absence of refocusing pulses, state A changes to state B

randomly, resulting in defocusing magnetisation in the XY plane. This causes a broadening of the resulting NMR peak (Figure 1.9, top). To re-focus magnetisation in XY plane, a variable number of 180° re-focusing pulses in the same delay period can be applied. As the refocusing pulses become more frequent, *i.e.* there is less time between 180° pulses, there is less of a chance for the state A to change to state B. The CPMG sequence can thus refocus magnetisation, and lead to sharper peaks (Figure 1.9). An elegant analogy of the CPMG experiment has also been suggested by Lewis Kay [66]. Here runners and walkers are used to describe the spins of the two different states (A and B), while differences in their speeds can be used to describe differences in chemical shifts between A and B. As states A and B can interchange, runners can start walking and walkers can start running. At the beginning of experiment runners and walkers are set off to run a lap around a field. In the absence of exchange (or if an exchange is very slow), there will be two populations of finishers; one group crosses the finish line first (runners) and another later (walkers). In this scenario, there will be two peaks in the NMR spectrum (one for state A and one for state B). If, on the other hand, the exchange rate constant is very fast (all runners became walkers many times over and *vice versa*), all the runners and walkers will finish together as one group, and here one peak is observed in the NMR spectrum. In the case of the intermediate exchange rate, some runners slow down to a walk once or twice, others five times and so on, during the lap of the field. As a result, all the runners and walkers finish at different times, and this would lead to a broad peak in the NMR spectrum. To refocus this group of runners and walkers, a CPMG experiment can be used. Here, the runners and walkers are again set off around the field, but this time, if a whistle is blown (representing the 180° pulse), all the runners and walkers will turn around. If there is no exchange (*i.e.* runners will keep running and walkers walking), two peaks are seen on the spectrum. If runners start randomly walking and walkers running, as in the case of intermediate exchange, if the whistle is blown, some of the runners may start walking back to start line and thus will not cover the same distance as they did on the way out. Here, not everyone will cross the start line at the same time. This would again lead to a broader peak in the NMR spectrum. However, if a larger number of whistles are blown to keep turning everyone

around again and again, the group should stay much less dispersed. As the period between whistles gets shorter, there is less time for runners to start walking and then running again. The distances covered running or walking are shorter and the participants are spread across a far smaller part of the field. This means that they are more likely to cross the finish line at the same time, equating to a sharper peak on the NMR spectra [66].
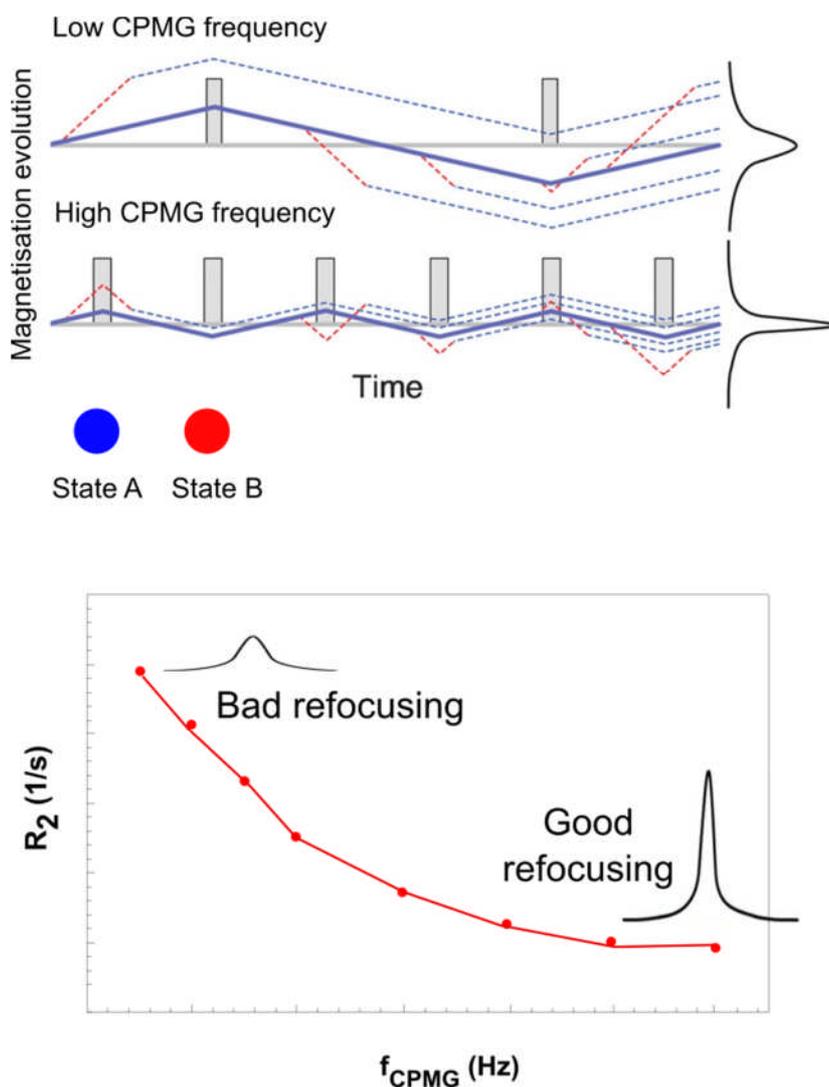
Figure 1.9: Schematic of a CPMG relaxation dispersion experiment peak re-focusing. If the two states of the system, A (in blue) and B (in red), interchange roughly at the same rate as their chemical shift difference, a broad peak is seen in spectra. The exchange between states A and B is shown as dashed red or blue line. The use of an increasing number of refocusing pulses (grey boxes) results in peak refocusing, as a peak from only one state is observed, as seen on the right panel. The bottom panel shows simulated data from CPMG experiments, where the peak is refocused when the CPMG frequency is increased. (Figure modified from [67] and [66])

Whilst CPMG experiments probe only µs-ms time scales, they have become a valued tool in studying ligand-protein binding, folding pathways and enzyme reactions [68-70]. To obtain exchange parameters such as the interconversion rate (k_{ex}) and the population of each state, as well as the chemical shift differences between two interconverting states, relaxation dispersion experiments need to be recorded at multiple magnetic fields. These parameters provide information on global and local dynamics and the thermodynamics of the exchange process [66]. Moreover, this approach allows characterisations of minor populations (smaller than 1%) [71]. In turn, information about chemical shift differences between the major and minor conformations can be used to obtain insights into structure and dynamics of the minor (populated less than 1%) state, which is not possible by any other experimental method.

From the CPMG data, the transverse relaxation rates (R2$_{eff}$) at different CPMG field strengths can be calculated from peak intensities using the following equation [72, 73]:

$$R2_{eff} = -\frac{1}{T_{CPMG}} \ln \frac{I}{I_0} \qquad \text{(eq. 6)}$$

Here, T$_{CPMG}$ is the constant time used for the CPMG experiment, I the peak intensity and I_0 the reference peak intensity.

In this work the relaxation dispersion method is used to study the differences in µs-ms dynamics between the different Hsp90:inhibitor systems, and to allow comparison, and validation, of computer simulation data.

1.3.4 Molecular dynamics simulations

Molecular dynamics simulations (MD) are a tool to describe, and visualise, the movement of atoms making up biomolecules using classical Newtonian physics. MD has provided insight into molecular movements since the early simulations were completed in the 1970s [74]. The early simulations, which included work on the 58 amino acid long bovine pancreatic trypsin inhibitor, were only a few picoseconds in length. With the advances in supercomputing,

larger systems, and even millisecond simulations, using specifically designed computer architecture have been achieved [74, 75].

As this study uses standard 'flavour' MD simulations, that have been widely adopted as a tool in biological research, a brief overview of the underlying principles for MD is provided. This is followed by a review of the different methods of obtaining entropy values from simulation data.

1.3.4.1 MD simulations - main principles

To run a molecular dynamics simulation, a co-ordinate file (typically from X-ray or NMR structures, but possibly also a homology model) for the system of interest is used as a starting structure. Next, a set of parameters to describe atom types, bond length and charges are applied to the structure. After this, the forces acting on each atom are calculated, based on classical mechanics, and subsequently each atom is moved based on the force acting upon it. A snapshot of the new system is taken, and the process is repeated over and over again to get a trajectory file of the molecular movements [76].

For MD simulations, the forces acting on the atoms are described by a force field, which contains terms for bonded and non-bonded interactions between the atoms, as shown in equation 7 [77]. The bonded interactions include terms for bonds, angles and dihedrals, to describe bond stretching, bending and twisting, as shown in the first three sums of the equation 7. The final term sums up the non-bonded interactions arising from van der Waals and electrostatic forces.

$$E_{total} = \sum_{bonds} k_r \, (r - r_{eq})^2 + \sum_{angles} k_\theta \, (\theta - \theta_{eq})^2 \; + \sum_{dihedrals} \frac{V_n}{2} \, [1 + \cos(n\emptyset - \gamma)]$$
$$+ \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

(eq.7)

The energy term for the bonds is modelled using two spheres connected by a spring, to which a different stiffness (*k*) is applied depending on the atoms involved in the bond. The letter *r* denotes the distance between the two atoms connected by the bond, and this term describes the bond stretch compared to the equilibrium position (see Figure 1.10). For angles, *k* is the force constant

for the angle bending, whose value depends on the types of atoms. The force constant depends on how large or small the angle (θ) between the atoms is in comparison to the equilibrium position. The dihedrals represent a torsional term describing the rotation about the chemical bond. This is described by a sinusoidal function that shows the energy differences between staggered and eclipsed conformations. This is the difference in energy of the atoms when they are furthest from, or closest to, one another due to bond rotation. As the bond could rotate around 360°, the same conformations can be reached at multiple times during the cycle ($n\emptyset$). The γ term accounts for shifts in the maxima and mimima positions. The dihedral term often includes terms for improper torsions, to maintain a particular geometry.
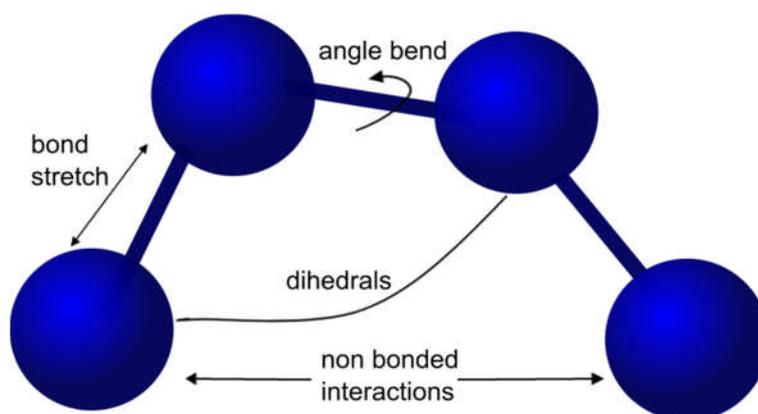
Figure 1.10: Different force field parameters shown at the atomistic level. The bond stretch, angle bend and dihedral terms describe interactions between connected atoms, whereas van der Waals and electrostatic terms are included in the non-bonded interactions.

The non-bonded terms, that are given in the last part of equation 7, are calculated for Van der Waals (VdW) interactions using the Lennard Jones 6-12 potential. When the atoms are sufficiently close to one another, there is an attractive force ($\frac{B_{ij}}{R_{ij}^{6}}$). The attraction term becomes weaker as the distance between the atoms increases. When the atoms are too close, there is an additional repulsive force ($\frac{A_{ij}}{R_{ij}^{12}}$). The electrostatic interactions are modelled by Coulombic laws, using atom-centred point charges [78].

Whilst the bonded terms are easy to handle in potential energy calculations, issues can arise with how great a distance the non-bonded interactions stretch to. In principle, all non-bonded atoms should be considered for the calculations, but in practice, this slows down the computations too much. To speed up the calculations, short range forces (*i.e.* VdW) are normally treated with a cut off value. For long range forces, *i.e.* electrostatics, more sophisticated algorithms, such as the particle mesh Ewald (PME) algorithm, are normally adopted. Additionally, simulations are normally run in periodic boundary conditions, where the simulation box is virtually surrounded by identical boxes. Here, if the molecule drifts to the edge of the box during the simulation, it does not simply disappear into space, but re-appears on the opposite side of the box (see Figure 1.11).
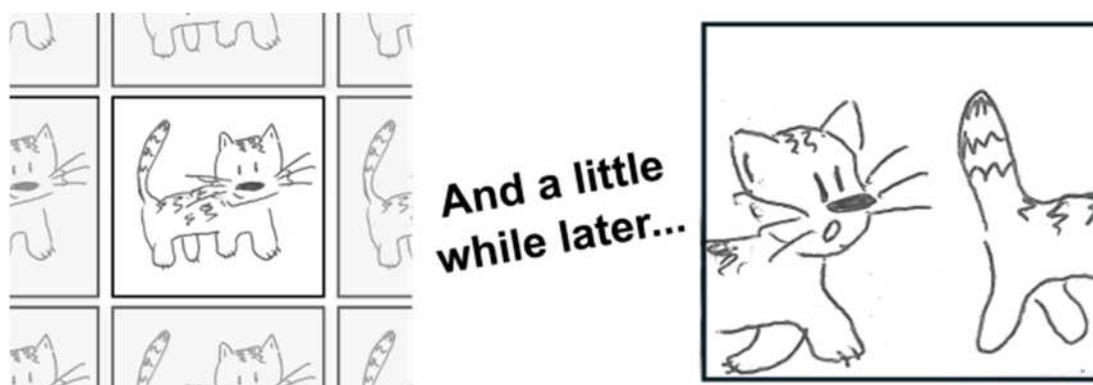
Figure 1.11: Virtual Periodic Boundary box. Simulations are normally run with 'imaginary' boxes surrounding the actual simulation box as shown on the left (greyed out cats for the virtual box *vs* black and white cat for the simulation box). This prevents the drifting of the molecules into space when they end up at the edge of the simulation box. The molecule that leaves the box on the right, simply re-appears on the left of the box, as depicted in the figure on right.

For MD runs to produce realistic atomic movements, the energy terms described above need to be parametrised based on the types of atoms. Collectively, these parameters are called a force field. The parameters are derived from *ab initio* calculations and from a range of experimental data to reproduce peptide and dihedral angles and protein crystal structures. These parameters include, for example, the radius of atoms, partial charges to

calculate the non-bonded terms, bond lengths, stiffness of the bonds, angles, and dihedral parameters [76]. There are few commonly used force fields, including those from Amber and CHARMM, that have been heavily tested by the simulation community. These force fields have subsequently improved over the years, in an attempt to make sure that they produce realistic movements [79] [80]. Whilst the commonly used force fields for biological macromolecules work reasonably well for protein and nucleic acid simulations, they have limited parameters for organic molecules. This is mainly due to difficulties with the accurate parametrisation of the significantly larger chemical space that small molecules cover, compared to the 20 standard amino acids and 5 different bases for DNA and RNA molecules.

To partly get around this issue of small molecule parametrisations, Amber developers have written a General Amber Force Field (GAFF) to enable more realistic simulations of organic molecule (*i.e.* drug) and protein or nucleic acid interactions [81]. The description of atomic charges is the key to understanding the chemical reactivity and physical properties of the molecules. The atomic charges are often calculated separately, using quantum-chemical *ab initio* methods, to obtain more accurate charges. These values are fed back into Amber to produce library files that make up specific force field parameters for the small molecule in question [82, 83].

1.3.4.2 Running a MD simulation

To run a simulation, first the forces acting on a system are calculated using classical mechanics formulas. Then the atoms are moved, proportional to the force acting on them. This two-step process is repeated over and over again. For the force calculations, each atom is considered as a single point and the forces (F) acting upon the atom can thus be calculated by:

$$\vec{F}_i = m_i \vec{a}_i \qquad\qquad (eq.8)$$

where m_i is the mass of atom *i* and a_i the acceleration. The force could alternatively be described by the gradient of potential energy (E_{total} from equation 7):

$$\vec{F}_i = -\nabla_i E_{total} \qquad\qquad (eq.9)$$

where ∇_i is the three dimensional derivative operator giving details of the x,y and z positions. This describes the force acting on atom *i*.

Combining these two equations allows us to compute the acceleration of the atom:

$$\vec{a}_i = -\frac{1}{m_i}\nabla_i E_{total} \qquad\qquad\qquad\qquad \text{(eq. 10)}$$

Equation 10 can be numerically integrated, via a choice of different algorithms, to get the position and velocity of the atom. Thus with the initial input of the position of the atoms, and known initial velocities and acceleration, the calculation can be repeated as long as necessary, to obtain a trajectory [78]. Whilst there are several slightly different implementations of how to integrate the above equation, this is not the crucial factor. Rather, what is important is to decide the length of the time step, given that a balance must be struck between the time taken to obtain a trajectory of sufficient length, and the accuracy of the calculations. The time step for the integration cannot be longer than the fastest motion in the system, which is the bond-stretching vibrations, typically in the order of 10 fs, and thus a routinely used time step for simulations is 2 fs [84].

Finally, as biological molecules in a cell are in an aqueous environment, to get a realistic picture of what is happening to the molecules, the MD simulations are normally carried out in water. Here there are several choices of water model, which range from explicit solvent (*i.e.* the water is modelled as individual water molecules) to implicit solvent model, where a force presenting bulk solution is applied to the surface of the biological molecule. The explicit water model is generally accepted to provide a more realistic description of water, although there is a large computational cost due to the sheer number of atoms required. Implicit models on the other hand, which describe the effect of water by simply adding forces, are often used when simulating very big systems (>10,000 atoms). Here, addition of explicit water molecules would simply make the system too large [85]. The calculation times are similar to explicit water simulations, but the motions of the molecules in simulation are accelerated. This can be explained by analogy - it is faster to run along the beach than in waist deep water. Improving the accuracy of the modelling of water is an active field of research in its own right, and a topic for another

thesis. For this study, the priority is to select the most suitable water model for the force field used. The chosen water model should match the water model which was used for the force field parametrisation and validation, for example, Amber ff99SB force field was parametrised using the three point TIP3P water model [77].

1.3.4.3 Configurational entropy calculations from molecular dynamics trajectories

MD simulations provide information on the flexibility and structural heterogeneity of macromolecules. When MD simulations are run for long enough to adequately sample the conformational space, the trajectory can be used to calculate the number of microstates (Ω) a system can adopt, together with the probability (p_i) of finding the system in a particular microstate. This information can be used to estimate entropy. Unsurprisingly, attempts to estimate the conformational entropy of macromolecules using MD simulations have been made for nearly as long as there has been simulation data available [86, 87].

The entropy calculations are based on the Boltzmann's formula, where the knowledge of the number of microstates (Ω), together with their probabilities, enables the calculation of the system's entropy (S) [88]:

$$S = k \log \Omega \qquad \text{(where k is the Boltzmann constant)} \qquad \text{(eq.11)}$$

There are two commonly used methods to calculate configurational entropies from simulation data; the quasi-harmonic approximation and the histogram method. Both methods are based on the above Boltzmann's formula, but they differ in the way the shape of the underlying energy landscape is approximated. The quasi-harmonic approach assumes that the fluctuations of the shape of the molecule during a simulation fit into a single large energy well. The macromolecule oscillates harmonically in this well, and a Gaussian-like distribution can be fitted to describe the well to estimate the entropy (Figure 1.12). In contrast, the histogram methods divide the conformations observed from the trajectory into different microstates that reside in different energy wells. Here, the size and steepness of the wells describes the likelihood of the molecule been found in each state. This gives the probabilities

that are needed for the entropy calculations [89]. In practice, the quasi-harmonic analysis works well for molecules that do not undergo large conformational fluctuations, whereas the histogram methods can be used for flexible molecules as well. However, for the histogram methods, issues do arise regarding how to define what are different microstates versus fluctuations within one state. Often an internal bond angle torsion (BAT) coordinate system is used for the histogram methods, rather than the Cartesian co-ordinates used to run the simulation. This is because it is easier to define limits using bond angle fluctuations. These two methods are further described below.

Quasi harmonic
approximation energy
landscape

Histogram based method's
energy landscape
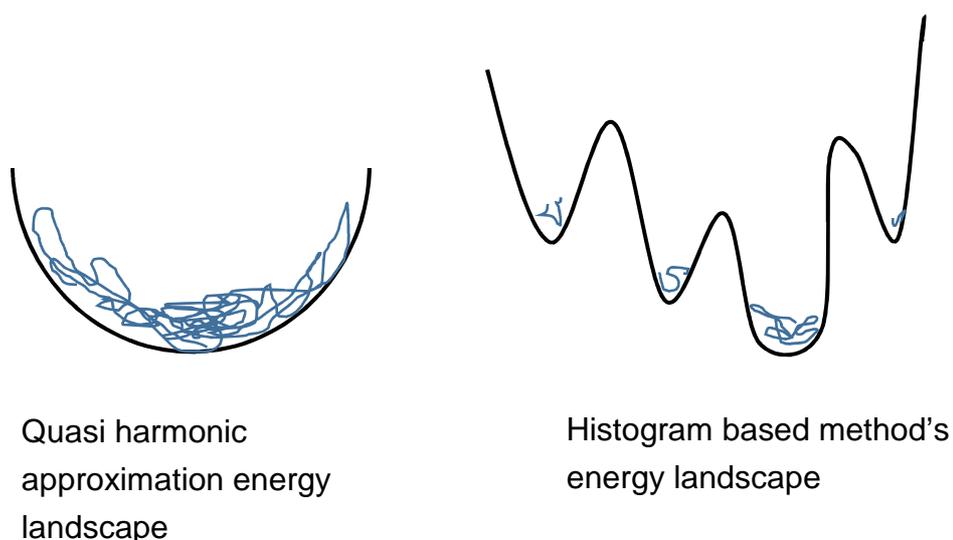
Figure 1.12: The quasi harmonic approximation and histogram methods for entropy calculations both assume different underlying potential energy landscapes. The quasi harmonic approach assumes a single energy well, as shown on left, in which the macromolecule oscillates. The histogram methods' energy landscape is depicted on right. Here the macromolecule is assumed to occupy several energy wells and the shape and steepness of the energy well gives the probability of it being occupied.

For the quasi-harmonic approach, the configurational entropy is generally assumed to originate only from the internal motions of the molecule, rather

than the overall tumbling of the molecule (*i.e.* rotational and translational motions). Moreover, the Cartesian co-ordinate system is normally used for the entropy calculations, as this allows the bond rotations and vibrations to be added together in a linear fashion. For the calculations, the fluctuations in atomic positions are converted into a mass weighted co-variance matrix that described the fluctuations of each Cartesian coordinate relative to its average position during the simulation. The co-variance matrix can then be diagonilised to obtain eigenvectors and eigenvalues, where the largest eigenvectors describe the dominant movements of the system. The total entropy (S) can then be calculated using the following formula [89, 90]:

$$S = \frac{1}{2}k \sum_{i}^{3N-6} \ln(1 + \frac{kTe^2}{\hbar^2}\gamma_i)$$
(eq.12)

Here the γ_i are the eigenvalues of the mass weighted co-variance matrix. The last logarithmic form is a correction term suggested by Schlitter, to reduce "numerical noise" from the very smallest eigenvalues that lie outside classical regime [90].

Quasi-harmonic approximation for conformational entropy is widely used. Although it has been found to overestimate entropy, due to the assumption that the internal vibrations are not correlated, the method has been demonstrated to give reasonable estimates of the relative entropies for both DNA and protein-ligand systems [14, 91]. The method has also been added as a 'standard tool' to many molecular dynamics software packages. The advantage of QHA is that it can help to rationalise the underlying biological movements that lead to entropy changes. This is because it is based on principal component analysis (PCA) with its use of eigenvectors and eigenvalues. The eigenvectors can be used to provide insights into the underlying biological system; the largest eigenvectors show where the major conformational changes are taking place and the eigenvalue describe the amplitude of the movement [92, 93].

One issue with the quasi harmonic approximation is that it breaks down if the underlying movements are anharmonic. The method works remarkably well for DNA, which is more rigid due to the ordered hydrogen bonds linking the bases [91, 92]. However, issues can arise with using the method to study more flexible and intrinsically disordered systems, which can be very anharmonic.

Also, due to the use of Cartesian co-ordinates, the trajectory needs to be first re-imaged, (*i.e.* all the simulation frames are aligned with a reference structure) to remove any rotational and translational movements. This again may present issues for flexible molecules.

Next, we describe in greater detail the histogram method for estimating configurational entropy from MD simulations. In this method, the microstates that are sampled during the simulation are divided into different bins to obtain a probability density function of the states. This can be used to calculate the entropy. Here, the probability density in a bin (p_i) is calculated first:

$$p_i = \frac{n_i}{n}\frac{1}{\Delta}$$
(eq.13)

Where the Δ is the size of bin and n_i is the number of times the macromolecule is found in the particular microstate during the simulation, compared to the total number of samples n. This information can then be added to the formula below to get the entropy:

$$S = -k\Delta \sum_{i=1}^{bins} p_i \ln p_i$$

(eq.14)

As explained earlier, how to define a microstate is often an issue with the histogram method. Often the trajectory data are converted to an internal co-ordinate system, where the 'bond-angle-torsion' (BAT) angles are used to define the bins. The problem is that the BAT co-ordinates only consider the bond twisting motions and ignores the contributions from bond stretching. Defining averages is also not easy with BAT co-ordinates, due to periodic movements of the bond rotations.

The advantage of the histogram method compared to the quasi-harmonic approximation is that it does not break down with anharmonic systems. However, the entropy value is often either over or under estimated, due to the difficulty of defining the bin sizes. It is not easy to estimate what constitutes a different microstate (and thus a different bin) as opposed to minor fluctuations within a state (and thus within the bin). There have been attempts to define the optimum dihedral angle to be used when defining the microstates, but the suggested values vary from 1° to 5° [94, 95]. In complex macromolecules,

correlated motions can play a large role. The configurational entropies may be overestimated, if this is not corrected by selecting only a subset of the dihedral angles of the macromolecule to be included in the entropy computations. Consequently, the histogram method tends to work relatively well for smaller molecules, but issues often arise with larger and more complex macromolecules, where correlated motion is an issue.

Whilst the quasi-harmonic and the histogram methods differ in how the underlying energy landscape and the probability distribution of the microstates is approximated, they can provide comparable entropy estimates, at least for calculations done on small molecules. An example of this was a study by Gilson's group, who tested their histogram method based the Mining Minima algorithm against the quasi-harmonic approximation on Ampenavir drug binding. The computed configurational entropy values were less than 1 kcal/mol apart, (calculated as 11.6 kcal/mol with the Mining Minima histogram method, compared to 12.3 kcal/mol by quasi-harmonic approximation) [14]. For larger molecules, comparative entropy values obtained using quasi-harmonic analysis and the histogram method have also been demonstrated. A general issues with all the currently used methods for computing entropy is convergence. When further MD data are processed, the entropy value increases. This highlights the issue with using finite simulation data to sample the nearly infinite energy landscape of complex molecules [96].

More recently, new methods for obtaining entropies from MD simulations have been suggested. One of these methods is using atomic forces instead of atomic positions for the calculations. This circumvents anharmonicity issues, but so far it has had only limited use in examining the entropy of water. It remains to be seen if it will be adopted for biomolecules [97]. Attention has also focused on accelerated MD techniques that improve the sampling of the conformational space. A study using an accelerated MD run, followed by per residue re-weighted dihedral entropy calculations, demonstrated that the 'accelerated MD entropies' corresponded better with the entropy values calculated from NMR dihedral analysis, compared to those from conventional MD simulations [98]. Despite recent developments, one of the main problems that remains is comparing the calculated entropy values to actual experimental data. Only NMR relaxation techniques allow the estimation of

configurational entropy at the atomistic level, rather than having to make comparisons with global entropy changes. Here, fast nanosecond methyl sidechain and amide backbone bond vector motions as measured by NMR have been used to calculate order parameters, which show how flexible or rigid the particular bond vector is [55, 99]. These order parameters can also be calculated from MD simulations and have been used as a proxy for global changes in protein dynamics to calculate changes in entropy [100-102]. The fast motions have been shown to be relevant to the overall entropy change, as work with thermo- and mesophile enzymes showed that the fast time scale motions can increase in some parts of a protein upon 'rigidification' of some other distal part of the enzyme [103]. Despite these advances, the configurational entropy calculations using NMR data also rely on similar assumptions to computations based on MD simulation data, where for example unharmonicity and correlated motions remain an issue. Indeed, whilst a number of studies have used the nuclear relaxation rate derived order parameters as a proxy for entropy, there are concerns on how well the limited number of measurements of internal motions characterise the overall protein dynamics, as often only amide bond vectors or methyl side chain motions are used [104]. Further, the side chain order parameters require different fit parameters depending on the residue type before they can be used for entropy estimation. These parameters also seem to be somewhat system specific. Issues arise also from correlated motions, as well as the effects on global flexibility affecting the general rotational term. To date the best results of configurational entropy values have been quoted to have been achieved using a rigid CAP protein [105].

1.4 Aims of project

The main aim of the project is to gain further insight into how protein dynamics is influenced by small molecule binding using the Hsp90 chaperone as a model system.

The interaction between protein and ligand are guided by thermodynamics, which can be examined by ITC and other methods that measure the overall binding affinity. However, examining the changes in configurational entropy,

which report on structural dynamics, could be of more interest. This is because it could enable conformational selection of only a certain pool of Hsp90, and consequently a particular downstream pathway. The binding of a ligand would further shift the conformational ensemble towards a certain ligand stabilised conformation. To examine the configurational entropy, the effect of small inhibitor binding to the Hsp90 protein was examined by NMR and MD techniques. These atomistic scale results were compared to the global values obtained from ITC. The overall aim of the study is not method development, but to gain further understanding what different thermodynamic signatures mean in the atomic level for protein dynamics.

The first chapter provides background on thermodynamics, the model system and the main techniques used to probe dynamics. The second chapter focuses in detailed description of the materials and methods used as well as describing method optimisation and validation experiments. Chapter 3 describes the results from experimental (NMR and ITC) and computational studies (MD simulations) of the effect small molecule binding has on Hsp90 dynamics. The results from the different techniques are compared and we try to rationalise the global entropy change in the light of protein conformational fluctuations. Chapter 4 focuses on configurational entropy calculations using MD simulation data. Here, both the quasi-harmonic analysis and the histogram methods are used to calculate entropies, with emphasis on highlighting issues with the methods. A further issue relating to the entropy calculations using MD simulation data relates to whether sufficient conformational space has been explored by the molecule during the simulation. We also try to answer the question of when 'enough' space has been explored by representing the analysis of a large number of replica simulations. Finally, Chapter 5 contains the overall conclusions of the studies.

Chapter 2

Materials, methods and method optimisation results

This 'Materials and Methods' chapter describes both the experimental and computational methods and how the materials were prepared. Results from validation studies and optimisation trials are also presented.

2.1 Materials and reagents

2.1.1 Small ligand inhibitors

The chemical structures of the initial panel of six Hsp90 small molecule inhibitors are shown in Figure 2.1. The panel included three commercially available inhibitors of Hsp90; geldanamycin and one of its derivatives 17-(dimethylaminoethylamino)-17-demethoxygeldanamycin (17-DMAG) as well as radicicol. These three compounds were obtained from Cayman Chemicals. The other three compounds, CNF2024, GVK0153 and GVK0161 were received from AstraZeneca. All ligands were dissolved in dimethyl sulfoxide (DMSO) to make 100 mM stock solutions.

2.1.2 Hsp90-NTD DNA template

The pET28a-6His-TEV-Hsp90(D9-E246) plasmid encoding for the human Hsp90-NTD was obtained from AstraZeneca. The plasmid has a kanamycin resistance gene for selection. The size of the Hsp90-NTD including the His-tag and thrombin cleavage site is around 26 kDa. The sequence of the Hsp90-NTD(D9-E246) used is shown below.

DQPMEEEEVETFAFQAEIAQLMSLIINTFYSNKEIFLRELISNSSDALDKIRYE
SLTDPSKLDSGKELHINLIPNKQDRTLTIVDTGIGMTKADLINNLGTIAKSGTK
AFMEALQAGADISMIGQFGVGFYSAYLVAEKVTVITKHNDDEQYAWESSA
GGSFTVRTDTGEPMGRGTKVILHLKEDQTEYLEERRIKEIVKKHSQFIGYPI
TLFVEKERDKEVSDDEAE

2.1.3. Chemicals

Standard chemicals were obtained from Sigma Aldrich and Fisher Scientific unless otherwise stated.

2.1.4 Heavy isotopes

Per-deuterated glucose, ^{13}C glucose, ^{15}N ammonium chloride, deuterium oxide (D_2O) and Celtone complete medium D as well as Celtone complete medium ^{13}C, ^{15}N were purchased from Cambridge Isotope Laboratories.
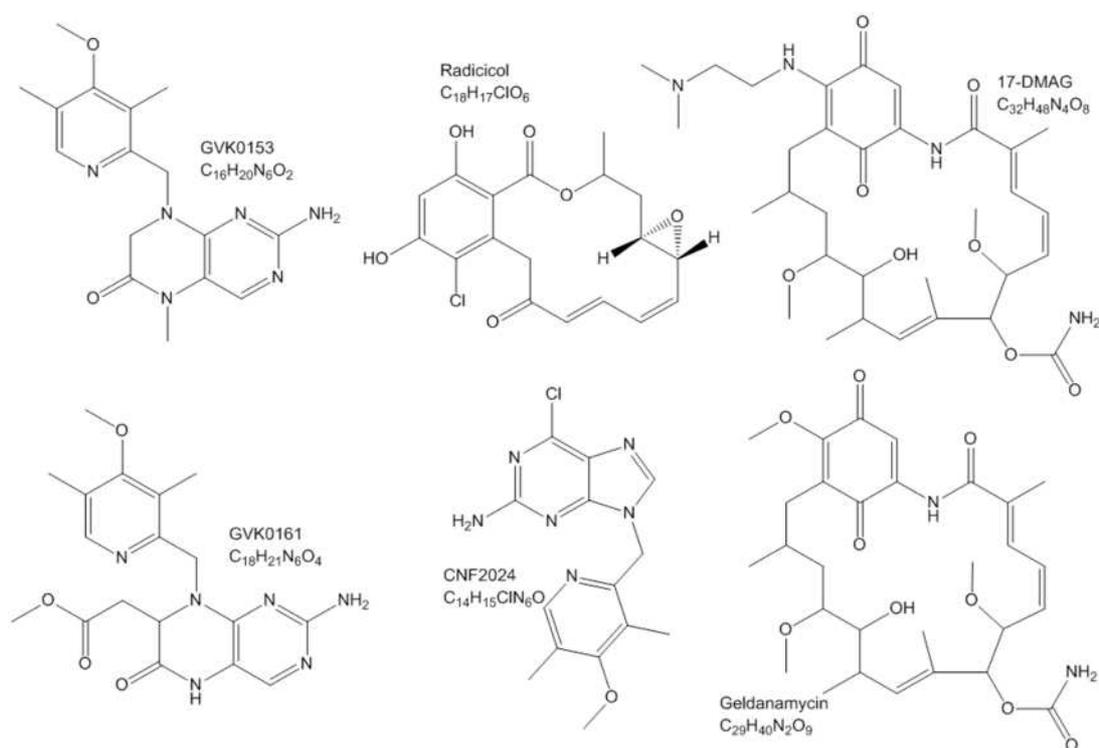
Figure 2.1: Chemical structures of the small molecule inhibitors used in this study. The inhibitors included three nucleotide mimic compounds, GVK0153, GVK0161 and CNF2024, shown on the left and middle of the second row. The other three compounds were geldanamycin and its derivative 17-DMAG as well as radicicol, which are shown on the right and middle of the top row.

2.2 General analytical methods

2.2.1 Quantifying plasmid DNA concentration

The plasmid DNA concentration was measured by using absorbance at 260 nm with the NanoDrop using the standard manufacturer's protocol [106].

2.2.2 Quantifying protein concentration

To measure protein concentration, the samples were diluted in 6 M guanidine HCl in 1:100 dilution prior to measuring the absorbance at 280 nm (Abs_{280}). The protein concentration was calculated using the Beer-Lambert law ($A = \varepsilon c l$), where absorption (A) depends on the extinction co-efficient (ε), concentration (c) and the length of the path that light travels through (l), which in this case is 1 cm. The value used for ε for Hsp90-NTD was 15,930 of M^{-1} cm^{-1} which was obtained from the Expasy server. The algorithm used by Expasy to obtain the value for ε calculates the molar extinction coefficients at 280 nm for tyrosine, tryptophan and cysteine residues, and afterwards sums these up [107].

2.2.3 SDS-polyacrylamide gel electrophoresis

To prepare samples for gel electrophoresis, 5 µl of 4XSB buffer (900 µl 4x Laemmli sample buffer (BioRad) and 100 µl 2-mercaptoethanol) was added to 10 µl of sample. After mixing, the samples were put on a heat block (99°C) for 2 minutes before loading. The samples were loaded on to precast BioRad MiniProtean TGX Precast Any kD gels. BioRad Precision Plus Protein Unstained Standard was used as the molecular weight marker. The gels were run at 200 V/400 mA for approximately thirty minutes. After this, the gels were stained for half an hour with BioRad Bio-Safe Coomassie G250 stain and de-stained with water. The His-tagged Hsp90-NTD band was expected to be around 26 kDa.

For samples that came from un-lysed bacterial cultures, 0.3 OD/ml of culture was taken. The cells were centrifuged after which the pellet was resuspended in 30 µl 8 M urea and 15 µl 4XSB buffer, prior to heating the sample briefly and loading the sample to gel as described above.

2.2.4 Preparation of agar plates and culture media

2.2.4.1 LB-agar plates

To make LB-agar plates for bacterial cultures, 25 g/l of Luria-Bertani media (10 g tryptone, 5 g yeast extract and 10 g NaCl for one litre) was mixed with 15 g/l of agar and dissolved in distilled H_2O. The media was next sterilised by autoclave (20 minutes at 121°C) and let to cool to 50°C before 25 µg/ml of kanamycin was added. After this LB-agar solution was poured onto plates to set. The plates were stored in 4°C.

2.2.4.2 M9 media

For M9 media preparation, 6.5 g Na_2HPO_4, 3.0 g KH_2PO_4, 0.5 g NaCl, 1 g NH_4Cl, 2 g D-glucose, 120 mg $MgSO_4$, 11 mg $CaCl_2$, 10 mg biotin and 10 mg thiamine were weighed and dissolved into 1000 ml of H_2O or D_2O. If isotope labelling was required, ammonia and glucose were replaced by labelled ammonia or glucose ($^{15}NH_4Cl$ and ^{13}C-glucose or per-deuterated glucose) as needed. Here, for the relaxation dispersion experiments, per-deuterated glucose and $^{15}NH_4Cl$ were used in order to obtain protein with 100% ^2H and ^{15}N labelling. The M9 media for expression of Hsp90-NTD used for assignments on the other hand were prepared using ^{13}C glucose and $^{15}NH_4Cl$ to get *ca.* 70-80% ^2H, ^{13}C, ^{15}N protein. After this, the media was filter sterilised. Finally, kanamycin and 10 ml of LB was added to make up the M9 culture media. LB was replaced by 10 ml Celtone complete labelled medium if ^2H, ^{13}C, ^{15}N isotope labelling was required.

2.3 Protein expression and purification

2.3.1 Transformation

To transform the plasmid into bacterial cells, 1 µl of pET28-Hsp90-NTD plasmid was added to a 50 µl aliquot of competent *E.coli* BL21 (DE3) cells. The tube containing the DNA and the bacterial cells was left to incubate on ice for 30 minutes. Following the incubation, the tube was placed on a 42°C heat block for 45 seconds, after which 500 µl of LB broth was added and the cells were placed in a shaker (250 rpm) to grow for 1 hour at 37°C. The cells

were centrifuged briefly and the cell pellet re-suspended to 100 µl of LB broth before plating them onto LB-Agar plates containing kanamycin antibiotic for selecting the plasmid containing cells.

2.3.2 Preparation of DNA stock

To prepare plasmid stock for Hsp90-NTD expression, the pET28a-6His-TEV-Hsp90(D9-E246) was transfected into *E.coli* DH5α cells following the heat shock method described above. A colony from the plate was picked to grow in 5 ml LB plus kanamycin overnight at 30°C. The following morning, the cells were harvested by centrifugation and DNA extracted using the Invitrogen Purelink Quick plasmid miniprep kit for plasmid DNA extraction according to manufacturer's instructions.

2.3.3 Protein expression trials

To optimise the expression protocol for the Hsp90-NTD using the M9 media, a small scale protein expression trial was carried out. For the trial, the effects of different expression temperatures (37°C, 30°C and 20°C) and IPTG amounts (0.1 mM and 1 mM) were tested.

First, a transformed *E.coli* BL21 (DE3) colony containing the pET28a-Hsp90-NTD plasmid was picked from the LB-Agar plates and added into 5 ml of LB/kanamycin media. The bacteria were left to grow in an incubator for 4 hours (37°C, 250 rpm). After this, 1 ml of the starter culture was taken to seed 25 ml cultures. The cells were further grown as above, until the, optical density of the culture measured at 600 nm (OD_{600}) was around 1.00. After this, the expression of the plasmid encoded Hsp90-NTD was induced with either 0.1 mM or 1 mM IPTG and the culture was left to grow at either at 37°C for 4 hours or at 30°C or at 20°C overnight. Samples from the cultures were taken before and after the induction and these were run in SDS-PAGE gel, as specified in Section 2.2.3.

2.3.4 Large scale protein expression (un-labelled)

To express the Hsp90-NTD protein, a transformed *E.coli* BL21 (DE3) colony containing the pET28a-Hsp90-NTD plasmid was picked from the LB-agar plates and added into 5 ml of LB/kanamycin media. The bacteria were left to

grow in an incubator for four hours (37°C, 250 rpm). After this, about 2 ml of the starter culture was taken to seed 50 ml of M9 media in a baffled flask. The cells were further grown as above until the optical density of the culture measured at 600 nm (OD_{600}) was around 1.00. Next, the culture was transferred to 500 ml M9 media to grow further until the OD_{600} again reached 1.00. After this, protein expression was induced by addition of 1 mM IPTG to a final concentration and the culture was left to grow overnight at 20°C.

The following day, the cells were harvested by centrifugation (5000 rpm for 20 min at 4°C). After this the samples were kept on ice to prevent any proteolysis and the pellet was re-suspended in 35 ml of Buffer A (40 mM HEPES pH 8, 300 mM NaCl and 5 mM DTT). The pellet was then flash frozen at -80°C until purification.

2.3.5 Expression of labelled protein

The above protocol was used to produce labelled protein, except that the starter culture was grown overnight at 30°C to allow for the slower bacterial growth in deuterium oxide.

2.3.6 Protein purification

To purify the His-tagged Hsp90-NTD, the cell pellet was defrosted. After this 20 mg/ml lysozyme and a protease inhibitor P8849 tablet (Roche) were added to the sample. The tube was briefly mixed, and the cells were left to incubate on ice for thirty minutes. After incubation, the sample was sonicated using a six seconds on, six seconds off cycle for ten cycles, after which the lysed cells were then centrifuged for 20 minutes at 20,000 rpm.

After centrifugation, the soluble fraction containing the His-tagged Hsp90-NTD was loaded onto a HisTrap HP column (GE Health Care) using the ÄKTA FPLC purification system. The unbound fraction was collected after which the column was washed first with five column volumes of Buffer A containing 10 mM imidazole followed by a second wash again with five column volumes of Buffer A containing 20 mM imidazole. Finally, the His-tagged protein was eluted with Buffer A containing 250 mM imidazole and the fractions containing the Hsp90-NTD were collected.

2.3.7 Buffer exchange

Following the purification, the fractions containing the purified Hsp90-NTD were pooled and dialysed against 500 ml of 20 mM sodium phosphate, pH 7.5 buffer (20 mM Na_2HPO_4 and 20 mM NaH_2PO_4) (NaPi buffer) for around twenty four hours at 4°C with two changes of the dialysis buffer.

If the protein sample required concentrating, this was done using a 10 K Amicon centrifuge tube until the desired concentration (350 μM for NMR samples) was reached.

2.3.8 NMR sample preparation

The samples for NMR measurements contained additionally 1% AEBSF hydrochloride protease inhibitor (from Biochemica) and 5% D_2O to allow the lock signal setting that stabilizes the magnetic field strength of the NMR instrument.

2.4 Isothermal titration calorimetry

The isothermal titration calorimetry (ITC) experiments were performed using a MicroCal iTC-200 instrument at 25°C. For the measurements, 20 µM Hsp90-NTD and 200 µM GVK0153 ligand, 400 µM 17-DMAG or 300 µM of CNF2024, geldanamycin, GVK1061 or radicicol were prepared in 20 mM sodium phosphate buffer (20 mM Na_2HPO_4, 20 mM NaH_2PO_4, pH 7.5) assay buffer with matched 2% DMSO concentration in both the protein and ligand samples.

For the experiments, 200 µl of 20 µM Hsp90-NTD protein solution was placed in the ITC instrument sample cell and the concentrated ligand solution in the 40 µl syringe. Nineteen injections of 2 µl of ligands were added into the protein sample and the mixing/equilibration time between samples injections was set to 150 seconds.

The ITC data were processed using an automated algorithm in Nitpic software to correct the base line [108]. The titration data was fitted using one site model and a non-linear least squares curve-fitting Levenberg and Marquardt algorithm with the Origin 7 Software from MicroCal.

2.5 Heat capacity measurements

ITC measurements were also used to determine the heat capacities (C_p) for the GVK0153 and 17-DMAG Hsp90-NTD complexes. The ITC measurements were conducted, as specified above, using the automatic MicroCal Auto-iTC200, except for the ligand concentration which was set to 300 µM. The measurements were repeated at 7°C, 14°C, 25°C, 32°C and 37°C for both GVK0153 and 17-DMAG ligands. The data were analysed as described above and the enthalpy values from the different measurements were plotted against the temperature to obtain value for the heat capacity (C_p) from the slope of the enthalpy *versus* temperature plot.

2.6 NMR methods

2.6.1 NMR experiments - general conditions

NMR experiments on the Hsp90 protein were carried out using 600 MHz, 750 MHz and 950 MHz Bruker spectrometers equipped with TCI-cryoprobes. All spectra were recorded at 25°C using the spectrometers either in the NMR facility in Leeds or the National Institute of Medical Research in Mill Hill, London. For NMR experiments, we used 350 µM Hsp90-NTD protein and 700 µM inhibitors and the samples were prepared in 20 mM Sodium Phosphate, pH 7.5 buffer (20 mM Na_2HPO_4 and 20 mM NaH_2PO_4) with 1% AEBSF hydrochloride protease inhibitor (from Biochemica) and 5% D_2O.

2.6.2 TROSY and HNCA, HNCO, HNcaCO

The backbone assignments of human Hsp90-NTD apo form have been published previously [61]. To transfer these assignments we used a set of triple resonance experiments: NH Transverse Relaxation Optimized Spectroscopy (TROSY) and a TROSY version of HNCA [109, 110]. The spectra were recorded with a Bruker 600 MHz for the GVK0153:Hsp90-NTD and 17-DMAG:Hsp90 complexes; the Bruker Avance III HD 950 MHz was used for the apo sample. For the TROSY version of HNCO and HNcaCO [110], the Bruker 950 MHz magnet was used for the GVK0153:Hsp90-NTD

complex and the 750 MHz magnet for the apo and 17-DMAG:Hsp90-NTD samples.

2.6.3 NMR spectra processing

All spectra were processed with NMRPipe [111] and analysed using Cara [112] and Analysis [113] software.

2.6.4 NMR - chemical shift perturbations

To identify the residues that experience perturbations between apo and ligand-bound forms, pairwise comparison of chemical shifts was performed. For each residue, the change in peak positions was calculated using the formula:

$$\Delta\delta_{tot} = \sqrt{(\Delta\delta_H)^2 + (0.154\Delta\delta_N)^2} \qquad [62] \qquad \text{(eq. 3)}$$

Here the $\Delta\delta$H and $\Delta\delta$N are [1]HN and ^{15}N chemical shift differences between apo and ligand bound forms and 0.154 is the weighting factor for [15]N. The changes in CSP were classified as: significant (>0.06 ppm and/or 0.6 ppm for amide ^{1}H and ^{15}N atoms) or large (larger than 0.3 ppm for $\Delta\delta_{tot}$).

2.6.5 NMR relaxation dispersion experiments

The ^{15}N relaxation dispersion experiments were acquired at 600 and 950 MHz fields for inhibitor:Hsp90-NTD complexes and 950 MHz only for the apo protein. The samples were ^{15}N labelled and deuterated with saturated molar excess of ligands. We used the NH TROSY version of the Carr-Purcell-Meiboom-Gill pulse sequence [114] with the 180° refocusing pulses set at 2, 4, 6, 8, 12, 16, 20 and 24 repeats (corresponding with CPMG field strengths of 50, 100, 150, 200, 300, 400, 500 and 600 Hz) with the CPMG pulses implemented in a constant time which was set to 40 ms.

For data analysis, only unambiguous residues, *i.e.* peaks that were not overlapped by any other, were included. Next, the transverse relaxation rate (R2$_{eff}$) at different CPMG field strengths was calculated from peak intensities using the following equation [72, 73]:

$$R2_{eff} = -\frac{1}{T_{CPMG}} \ln\frac{I}{I_0} \qquad \text{(eq. 15)}$$

Here, T_{CPMG} is the constant time used for the CPMG experiment, I the peak intensity and I_0 the reference peak intensity. The relaxation rate ($R2_{eff}$) was next plotted as a function of the CPMG field strength to identify residues with exchange broadening.

Uncertainties were estimated from duplicate or triplicate measurements. The residues with observable micro- to millisecond dynamics were classified using the criteria that overall standard deviation of the peak intensities had to be 0.02 or larger, demonstrating the refocusing of the peak width upon increased CPMG field strength with error less than 5% between the repeat measurements.

2.6.6 NMR experiments on ligands

NMR spectra of the ligands were recorded using a 500 MHz Varian Inova spectrometer. For the ligands, 10, 50 and 200 µM concentrations were prepared in aqueous buffer (10 mM KPi, 50 mM KCl, 0.02% NaN_3, 5 mM $MgCl_2$ and 5 mM DTT in D_2O). One dimensional (1D) hydrogen spectra were obtained at 298 K and 318 K using the presat pulse sequence to suppress the water signal.

2.7 Crystal structures of Hsp90-NTD and ligand complexes

We used the previously solved X-ray crystal structures of the Hsp90-NTD in complex with 17-DMAG (PDB:1OSF) [115] and GVK0153 ligands (unpublished data from AstraZeneca). Both structures are of reasonably high refinement (1.75 Å for the 1OSF and 2.07 Å for the GVK0153 bound structure) thus allowing the use of the co-ordinates for the molecular dynamics (MD) simulations.

Ligand and protein contacts were analysed by the LigPlot programme using the crystal structures [116]. The crystal structures (apo Hsp90-NTD and Hsp90-NTD in complex with GVK0153 or 17-DMAG) were compared using structural alignment algorithm in VMD software. Root mean square deviation values between the structures were calculated, as well as the solvent accessible and hydrophilic and hydrophobic surface areas with standard tools

in VMD [117]. For surface area calculations, we used a standard 1.4 Å cut off value.

2.8 Molecular dynamics simulations

2.8.1 Molecular dynamics packages used

The MD simulations on apo Hsp90-NTD, 17-DMAG:Hsp90-NTD and GVK0153:Hsp90-NTD complexes were performed using GROMACS (v 4.6.5) with the Amber ff99SB force field and TIP3P water model [118, 119]. The ligand only simulations were run using Amber 14.

2.8.2 Supercomputing resources

The simulations were run on various supercomputers, including the ARC1, ARC2 and Polaris machines in the Leeds High Performance Computing facility. Computing time was also obtained from the national ARCHER supercomputer as well as the BlueGene in the Hartree centre. Simulations of the Hsp90-inhibitor system and of free ligands were run using graphical processing units (GPUs) with Amber's pmemd GPU optimised code [120].

2.8.3 Generating files for MD runs

Before the MD runs can take place, the co-ordinate files also need accompanying parameters to describe the system. As the preparation of the files can be complex especially when small molecules are also included in the simulation, the workflow in Figure 2.4 shows an overview of the different steps taken for MD file preparation and these different steps are described in more detail below.

- Xray structures used as starting structures to obtain co-ordinates for protein and ligands for simulations
- Same 'protein' template for all simulations (1YET.pdb):
  - VMD used to align Hsp90 backbone from 17-DMAG:Hsp90 and GVK0153:Hsp90 crystal structures & 'cut and paste' ligand co-ordinates to '1YET.pdb' file.

Crystal structures

- Prepare parameter files for protein and ligand
  - Ligand: charges calculated with Gaussian-03, Amber Antechamber and Amber Generalised Force Field (GAFF) for other parameters
  - Protein: parameters generated using standard Amber force field (ff99SB)
  - 'xleap' in Amber to combine ligand and protein topology files
  - Generate simulation box, add water and ions with either Amber 'xleap' or Gromacs 'grompp'

Parametrisation

- Minimise structures
- Thermalise and equilibrate
- MD production runs

MD simulations

Figure 2.2: Workflow for the production of the MD simulation files starting from crystal structures, parametrisation to MD production runs.

2.8.4 Ligand parametrisation

Whilst the Amber force field contains the parameters for the protein, small ligands need to be separately parametrised including assigning atom types and partial charges before they can be used in the MD runs. For this, the initial co-ordinates for the ligands were obtained from the X-ray structures (1OSF.pdb and un-published GVK0153:Hsp90NTD structure file from AstraZeneca). Hydrogen atoms were added to the crystal structures using the Chimera software [121].

2.8.4.1 Generalised Amber force field

As the Amber ff99SB force field used for the protein does not contain all parameters for small organic molecules, Amber tools package 'Antechamber' was used to generate the force field parameters for the ligands. The Antechamber uses general Amber force field (GAFF), which contains extra atom types and the corresponding parameters for bonds and angles to cover small organic molecules. The parameters for the ligands were thus generated

by the Antechamber using the GAFF force field with extra input of files containing partial charges for the ligands derived from Gaussian quantum mechanical calculations as described below [81].

2.8.4.2 Calculating partial charges for the ligands

Whilst the Amber GAFF force field contains general parameters for bonds, angles and dihedrals, the charges for atoms need to be calculated separately. Here, the partial charges for the ligands were derived by using quantum mechanical calculations using the Hartree Fock (HF) level of theory and the 6-31G* basis set, which allows the calculation of the electron densities around the molecule. The restrained electrostatic potential (RESP) can be constructed from these data and fed back into Amber to use in the parametrisation. The HF/6-31G* has also been used to derive the ff99SB Amber force field parameters and thus the ligand charges and rest of the force field parameters will be compatible [119].

For the partial charge derivation, first the ligand (.pdb) files were converted to xyz format that is compatible with the Gaussian-03 using Open Babel [122]. After this, the Gaussian-03 package was used to first optimise the geometry of the ligands using B3LYP/6-31G*, which is a faster method to calculate charges based on density functional theorem to calculate point energy of each atom to allow finding a optimise geometry at local energy minimum [123]. The geometry optimisation results were checked with tools in the Molden programme after which the final configuration was entered back to Gaussian-03 for calculation of partial charges using HF/6-31G* [124].

2.8.5 Preparing protein structure files

To ensure that the different simulations (*i.e.* apo, 17-DMAG:Hsp90-NTD and GVK0153:Hsp90-NTD) could be compared, *i.e.* that any differences observed were only due to ligand and not due to any differences in protein structures, the protein co-ordinates from a single Hsp90-NTD crystal structure file were used for all simulations. For this, the 1YET.pdb file was chosen to be used for the protein co-ordinates. This structure is one of the first human Hsp90-NTD structures in complex with geldanamycin [125]. For the apo protein, the geldanamycin ligand co-ordinates were simply deleted from the file. For the

two different ligands, a 'docking' approach was used. To 'dock' the ligands to the Hsp90-NTD structure from the 1YET.pdb, the geldanamycin ligand was removed by cutting out the co-ordinates corresponding to the small molecule. Next, both the "apo-1YET" structure and the structures of 17-DMAG:Hsp90-NTD (PDB:1OSF) [115] and GVK0153:Hsp90-NTD (unpublished data) were opened with VMD and the protein structures were next aligned using the structural alignment tool within the MultiSeq extension of VMD [117]. After alignment, the new coordinates of the aligned structures were saved and the ligand co-ordinates simply cut and pasted to the end of the 1YET-Hsp90-NTD co-ordinate file and the atom numbers were corrected. The resulting structure was visually inspected for correct location of the ligand.

2.8.6 Generating the starting structures for MD

Once the co-ordinate files were ready, the xleap tool in Amber was used to generate the necessary files for the MD runs. Here the crystal structure co-ordinates were read in together with the GAFF force field and library files for the ligands as generated before and the ff99SB force field for protein to allow the topology file generation for the complex. Next, the systems were neutralised by addition of sodium counter ions and merged into periodic box surrounded by ~45,000 water molecules. The systems were then minimised, thermalised and equilibrated using standard protocol of initial 4 step structure minimisation followed by 8 equilibration steps where the temperature is increased in a step-wise fashion and non-water atoms are initially constrained with the constraining forces slowly moved over number of steps [126]. The final equilibrated structures were used to run 500 ns unrestrained MD simulations with periodic boundary conditions at constant pressure (1 atm) and temperature (298 K). Shake was used to constrain all bonds involving hydrogen atoms with time step of 2 fs for integration of Newton's equations and a 12.0 Å cut off was used for non-bonded interactions. For replica simulations, the same starting structures were used but the values for initial velocity were assigned using the random seed generator in Gromacs. The input files for the energy minimisation (em.mdp) and the equilibration steps (md1.mdp, md2.mdp and md3.mdp) as well as the production MD input are shown in Appendix.

For ligand only simulations, the parametrisation and MD starting file preparation was carried as above with the simulations run using Amber's GPU accelerated code.

2.8.7 Post-simulation trajectory file editing

After the simulation runs had completed, the solute (water and ions) were stripped from the files and the protein was re-imaged (*i.e.* centre of mass aligned to the centre of the simulation box) to remove the diffusion of the complex around the simulation box. The frames for the first 15 ns corresponding to the equilibration were removed using standard tools within Gromacs and Amber. There were issues with Gromacs imaging commands with the ligand-protein files as only the protein was centred and the ligand seemed to 'jump out' of the protein binding pocket when the trajectory was viewed. To get around these imaging issues, the trajectories were loaded into the VMD programme and the 'Align' option in the 'RMSD Trajectory Tool' within VMD was used to re-image the Gromacs trajectories to centre both the ligand and protein thus 'get the ligand to stick in the protein pocket'. The re-imaged trajectories were then saved to Amber file format with VMD [117].

2.8.8 Simulation data analysis

For the simulation data analysis, root mean square fluctuation (RMSF) of the backbone C_α residues were calculated for each simulation. Next principal component analysis was used to gain insight for the main motions of simulated systems. For principal component analyses, which enables one to find out the dominant molecular motions from the trajectory, we used the PCAzip compression and analysis toolkit [93]. The PCAzip calculates the average atom co-ordinates from the MD trajectory to get a co-variance matrix which is diagonalised to get a set of eigenvectors with corresponding eigenvalues to get the direction and amplitudes for the main motions in the simulation. These files can also be converted to short animations to visualise the largest motions and extreme structures of the simulated system.

We also calculated the entropies of the simulated systems using the Schlitter's quasi harmonic method [90]. For this, an in-house code 'X2S' to produce the mass weighted co-variance matrix of atom positions that allows the calculation

of maximum entropy was used. For the entropy calculations and the PCA analysis the N- and C-terminal loops as well as the C-terminal part of flexible helix 2 were removed to leave the globular core, to able the analysis of the smaller amplitude movements of the protein.

2.9 Method optimisation and validation results

2.9.1 Protein production trials

Prior to any large-scale protein expression, expression protocol optimisation trials were carried out. For this, small 25 ml cultures were grown as specified in section 2.3.3, after which the expression of the plasmid encoded Hsp90-NTD was induced with either 0.1 mM or 1 mM IPTG. The cultures were left to grow at either 37°C for four hours or 30°C or 20°C overnight (see Table 2.1 for summary of the different conditions used). The protein yield from the different temperatures and IPTG amounts was analysed by running a SDS-PAGE gel and comparing the band intensities (Figure 2.3).

The 1 mM IPTG induction followed by 20°C overnight culture produced the highest yield of the Hsp90-NTD, as seen in the induced band intensities in SDS gels (Figure 2.3). This condition was thus adopted for the large scale protein expression.

2.9.2 Protein expression yields

The yields of Hsp90-NTD from 500 ml culture in per-deuterated M9 media were typically 1 ml of 300 μM (~7.2 mg) protein and about twice the amount for bacteria grown in M9 media prepared in H_2O. The protein was purified as specified in Section 2.3.5. Figure 2.4 shows results of a typical purification with the over expressed Hsp90-NTD on lane four, that contains fractions eluted by 250 mM imidazole.

Table 2.1 Protein optimisation trial conditions

Temperature after induction (°C)	Amount of IPTG (mM)	Growth time after induction
37	1	4 h
30	1	Overnight
30	0.1	Overnight
20	1	Overnight
20	0.1	Overnight
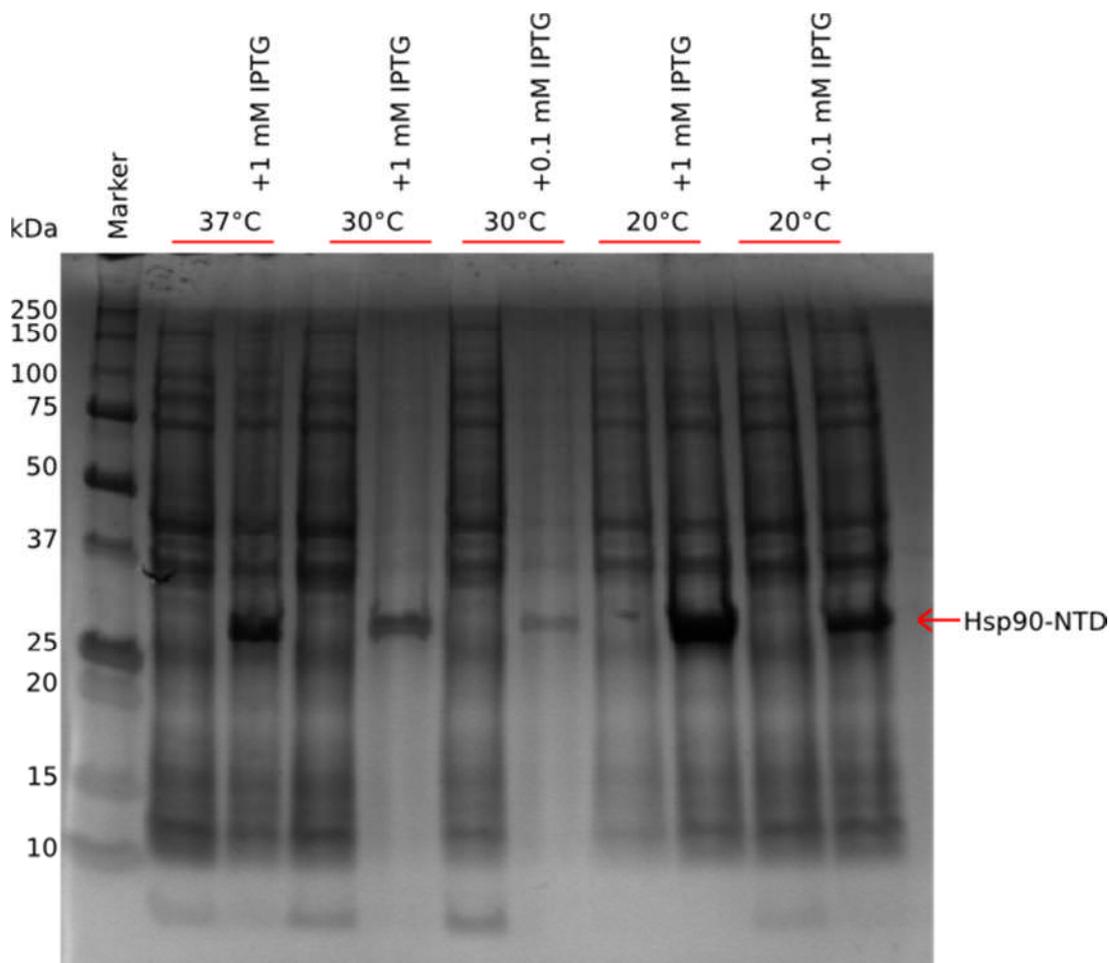
Figure 2.3: Small scale expression trials using M9 media. 25 ml cultures of *E.coli* containing the Hsp90-NTD containing plasmid were grown in M9 media until OD_{600} was 1.00. The expression of the protein was next induced either with 1 mM or 0.1 mM IPTG (see Table 2.1) and the culture were left to grow for 4 hours or overnight at different temperatures as specified on the sample lane labels. The band around 26 kDa present in the induced cultures (labelled with '+ IPTG'), corresponding to the Hsp90-NTD is indicated.
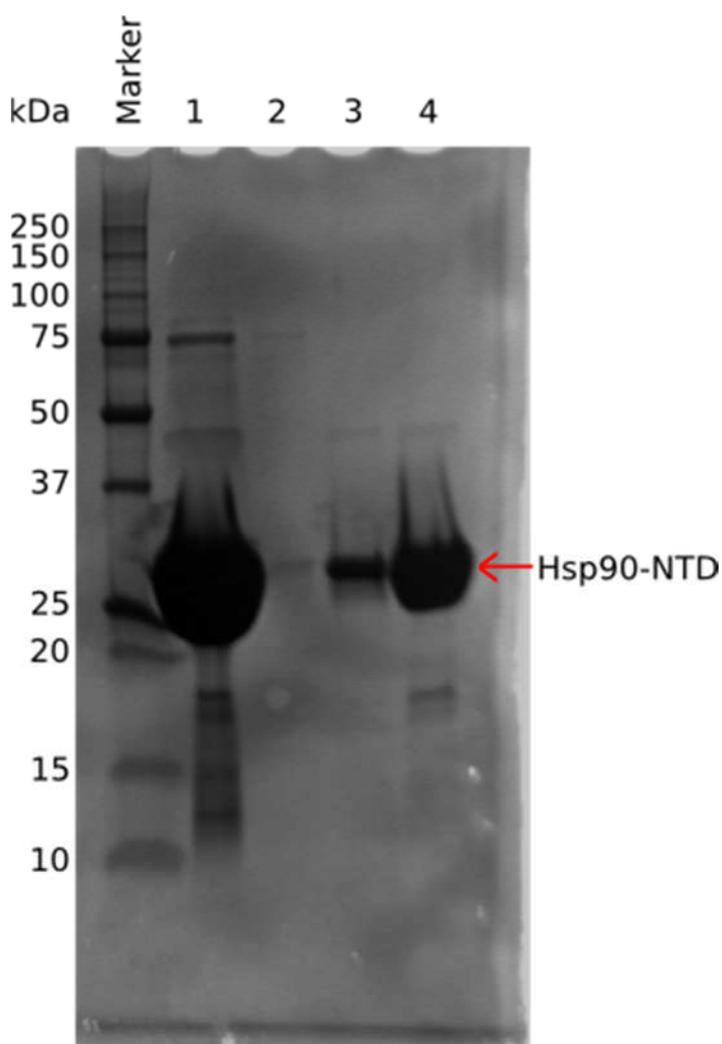
Figure 2.4: Purification of the His-tagged Hps90-NTD using affinity chromatography. The SDS-page gel shows results of a typical purification using the HisTrap HP column. Lane 1 contains the total cellular protein (*i.e.* sample that was loaded onto the column), lanes 2 and 3 show washes with increasing Imidazole concentration (10 mM and 20 mM respectively). Lane 4 the eluted fractions (250 mM Imidazole wash) which contains the majority of the over expressed Hsp90 as indicated.

The Hsp90-NTD was stable in the presence of protease inhibitors at room temperature for several days. This was tested by the 2D NMR NH TROSY spectra that were recorded at the start and end of the several day long NMR experiments. Both spectra were identical and did not show any sign of presence of unfolded protein peaks (data not shown). The oligomerisation

state of the Hsp90-NTD in the high concentration required for NMR (350 µM) was not measured, we assumed that the protein was monomeric.

2.9.3 Inhibitor selection for in depth studies

We were interested in whether different inhibitors binding to Hsp90-NTD have any effect on protein dynamics. As both in-depth NMR studies and MD simulations require considerable time investment for each different protein-ligand system, it was decided to concentrate on two different ligands. We describe here how the two ligands were selected.

A panel of small molecules was used for initial studies. The main criteria for selecting the compounds for the initial panel included the availability of crystal structures of the human Hsp90 in complex with the ligand, to allow for molecular dynamics studies. Additionally, we wanted to choose compounds that have similar binding affinities but different entropic contributions to binding. Finally, the compound of interest should be soluble enough in H_2O buffer in the absence of Hsp90. The initial set of six compounds (shown in Figure 2.1) included geldanamycin [127], its derivative 17-DMAG [128], radicicol [129] as well as three new nucleotide mimic compounds obtained from AstraZeneca, named; CNF2024, GVK0153 and GVK0161.

Early studies on yeast Hsp90 demonstrated that the geldanamycin and radicicol have very different entropic contributions to the Hsp90-ligand binding signatures [130]. No prior thermodynamic data was available for the new AstraZeneca compounds.

We conducted ITC experiments on all the compounds and recorded 2D protein NH TROSY amide spectra of Hsp90-NTD in complex with the inhibitors. Based on the initial biophysical characterisations using NMR and ITC, 17-DMAG and GVK0153 were selected for detailed NMR and MD analysis. The 17-DMAG molecule was chosen based on ITC studies, which showed the binding to be entropically driven (results from ITC are shown in Chapter 3). From the other ligands in the initial panel, GVK0153 and GVK0161 from the AstraZeneca compound library were shown to have a small entropic penalty upon binding, thus differing from the 17-DMAG. The CNF2024 and radicicol bound to Hsp90-NTD too tightly to be measured accurately with ITC,

as the nanomolar affinities suggested by the data resulted in a slope that is too steep for accurate K_a determination.

The 2D protein NH TROSY amide spectra of Hsp90-NTD in complex with the different compounds were all well resolved. The spectra for the AstraZeneca compounds were very similar, with only a few shifts in the protein backbone peak. The GVK0153 ligand had a few different shifts in the peaks, compared to GVK0161 and CNF2024 spectra, which could be of interest. This was the reason for selecting the GVK0153 compound for further studies.

2.9.4 MD simulation benchmarking and optimisation

Long simulation times and multiple replica simulations are needed in order to explore the conformational space adequately for entropy determination. Therefore we optimised the molecular dynamics runs for trajectory data generation speed. In the past, the Amber platform has been used by the Harris group. However, since the Gromacs software has been coded to optimise the data output, it was decided to conduct a benchmarking study to test the output speed of Amber *versus* Gromacs platforms. For this, short simulations of the apo-Hsp90 were prepared as specified in Section 2.8. These simulations were run on both the Amber and Gromacs engines using the ARC2 computer with a different number of central processing units (CPUs). The Hartree centre's BlueGene computer's performance was also compared to the ARC2.

As the supercomputing resources are, in general, shared between different users, there is a balance to be struck between requesting a large number of computer cores and the queuing time for the simulations. Generally, less cores equals shorter queue times. MD runs were thus submitted on ARC2 requesting 16, 32, 64 or 128 CPUs. These runs had queue times ranging from few hours to few days, which was quite acceptable for our purposes. Both Amber and Gromacs packages were used in the benchmarking study. From this benchmarking, the best Amber run produced 4.5 ns/day on 32 CPUs (Amber runs on a larger number of CPUs were slower). The best Gromacs runs were obtained using 128 CPUs, achieving 30.4 ns/day. For production runs, it was noticed that on ARC2, the best performance was achieved when all the CPUs used for the calculation were on the same node, rather than using any free CPUs in different nodes. When sole use of all CPUs in a node was

requested, this dramatically reduced the time spent on communication between the CPUs whilst performing parallel calculations. Here, the performance increased from 30 ns/day to 100 ns/day when calculations were carried out on a node. For Amber, a similar increase in simulation time was achieved, when all the calculations were done in one node with the top simulation speed around 20 ns/day. Table 2.2 shows a comparison of the simulation data speed with optimised conditions. The BlueGene architecture on the other hand was a lot slower with 50 ns/day obtained using four times as many CPUs.

As the benchmarking study showed that Gromacs runs were nearly five times faster than Amber, Gromacs use was adopted for the study. The simulation speed gain, which is normally the bottle neck in computational studies, far out weighted the time required for converting the Amber files to Gromacs format and back to Amber for analysis.

Finally, we obtained some graphical processing units (GPUs) and tested these with the GPU optimised Amber code. Here, the GPU speed was comparable to the ARC2 supercomputer, where 2 GPUs produced the same speed as 128 CPUs on the ARC2. As the GPUs were for the sole use of the Harris group, there was no queuing time. This makes the new GPU technology very promising in terms of output speeds, whilst being very economical in comparison to supercomputing.

Table 2.2: Optimised production MD simulation times

Computer	MD engine	Simulation speed (ns/day)
ARC2	Gromacs / CPU	100 ns/day (128 CPUs)
ARC2	Amber / CPU	20 ns/day (128 CPUs)
GPU	Amber / GPU optimised	35 ns/day (2 GPUs)
BlueGene	Gromacs / CPU	50 ns/day (512 CPUs)

2.9.5 Molecular dynamics simulations - replica data

In an attempt to cover as much conformational space as possible for entropy calculations, around 80 replica simulations were run. The total amount of

simulation time was 40 µs from these multiple replica simulations. Table 2.3 shows the number of replicas run for each different system.

Some of the simulations were sampled every 2 fs and these data were used for in depth analysis of the dynamics of Hsp90-NTD. A larger number of replica simulations were also ran with the aim of obtaining an estimate of how large a conformational space a single simulation explores. This would enable us to have some idea of the number of replicas, and length of simulations required to get a realistic representation of the conformational space. Due to the large number of simulations and thus data storage requirements, data from these replica simulations were saved every 10 fs.

Table 2.3: Summary of simulation data and total trajectory time

Complex	Simulation length	Number of replicas	Total simulation time
apo Hsp90-NTD[1]	600 ns	3	~1.5 µs
Hsp90-NTD:17-DMAG[1]	550 ns	3	~1.5 µs
Hsp90-NTD:GVK0153[1]	550 ns	3	~1.5 µs
Hsp90-NTD:17-DMAGs*	500 ns	50	25 µs
Hsp90-NTD:GVK0153*	200 ns	25	5 µs
Apo Hsp90-NTD*	75 ns	10	750 ns

[1]These simulations were sampled every 2 fs and were used for in depth analysis of differences in Hsp90-NTD dynamics of different complexes
*These simulations were sampled with every 10 fs and used to estimate how much conformational space is sampled with each replica

Chapter 3
Characterisation of Hsp90 and inhibitor dynamics

The chaperone Hsp90 is a highly dynamic protein and it has been suggested that the changes in the structure of the chaperone may influence the choice of binding partners [48]. To investigate the role that dynamics play in Hsp90-inhibitor interactions, we analysed experimental ITC and NMR data as well as performing a computational analysis based on MD simulations.

ITC is often seen as the gold standard method for determining Gibb's free energy, enthalpy and entropy of binding, as well as the binding (and dissociation) constant [53]. However, ITC only provides data on 'global' changes that include effects from protein, ligand and solvation. To understand the structural mechanisms of binding, one needs to know the individual contributions from the protein, the small molecule ligand and the solvent. These need to be complemented with information about the atomistic changes in protein and ligand structure and dynamics. We employed NMR and MD data to obtain these atomistic details.

The results from different NMR techniques and MD simulations that report on protein flexibility are presented and compared. These atomistic scale results are also compared with the ITC data. The apparent discrepancy between the overall entropy change upon binding (as reported by ITC) and the changes in configurational entropy is discussed. The latter changes were indirectly measured using changes in protein dynamics with NMR and MD techniques. To attempt to rationalise this discrepancy, the effect of the possible changes in the solvation structure is discussed.

3.1 Analysis of the inhibitor:Hsp90-NTD crystal structures

Crystal structures of both GVK0153 and 17-DMAG in complex with human Hsp90-NTD have been solved and these show that both molecules bind to the N-terminal ATP/ADP binding pocket ([115], unpublished data from AstraZeneca). To better understand, and interpret, the thermodynamic data, we performed a preliminary analysis of these protein structures by root mean

square difference (RMSD) of the backbone atomic coordinates, by examining the ligand-protein contacts, as well as by looking at the surface areas of the protein molecules.

The general structure of the human Hsp90-NTD is globular, with a number of alpha-helices surrounding a central cavity, where the ATP/ADP and inhibitors bind (Figure 3.1). Beta-sheets make up the bottom of the binding pocket [125]. One of the edges of the binding pocket is a so called 'lid region', which is made up of alpha helices 4 and 5. This region has been shown to be highly dynamic and changes conformation upon ATP/ADP binding [50].

The backbone RMSD between the apo protein and the 17-DMAG and GVK0153 Hsp90-NTD complexes was calculated by the VMD programme [117]. The Hsp90-NTD:17-DMAG (PDB:1OSF) and :GVK1053 (unpublished) complexes show very minor differences, with backbone RMSD of 0.75 Å between the two structures. This RMSD difference is very similar to that between the apo Hsp90-NTD (PDB:1YES) and the 17-DMAG:Hsp90-NTD and GVK0153:Hsp90-NTD structures, which are 0.73 Å and 0.75 Å respectively. The main difference between the 17-DMAG and GVK0153 bound proteins is the position of the loop 4 (L4), which connects the helices 4 and 5 within the lid region. The loop is orientated towards the binding pocket in the GVK0153 bound structure and away from the pocket in the 17-DMAG:Hsp90 complex (see Figure 3.1). As the 17-DMAG compound is about twice the size of the GVK0153 compound (molecular weights of the two compounds are 617 g/mol and 312 g/mol respectively). The orientation of the loop is likely caused by the larger area the 17-DMAG occupies within the binding pocket. Interestingly, this difference in the loop structure has also been observed in the apo Hsp90-NTD 'open' and 'closed' structures (PDB: 1YES and 1YER respectively). Here, the loop 4 orientation is either towards the binding pocket for the 'closed' form or away from it for the 'open' form [125]. The loop orientation in the GVK0153 structure is similar to the apo 'closed' structure and the larger 17-DMAG protein is similar to the loop in the apo 'open' structure. In contrast, when Hsp90 is in complex with either ADP or ATP, the loop is in the 'open' position (PDB: 1BYQ for the ADP bound structure and 1TC0 for the Hsp90 ER paralogue in complex with ATP) [131, 132].

The contacts between the ligands and protein were analysed using the LigPlot software, which plots H-bonds and hydrophobic contacts [116]. Both ligands interact with the same seven amino acids (Asn51, Ala55, Asp93, Met98, Leu107, Phe138 and Thr184) within the binding pocket (Figure 3.2). The ligands also make a number of 'unique' contacts with the protein. These are: Ser52, Lys58, Asp54, Ile96, Asp 102, Asn106, Lys112, Gly135, Val136 and Gly137 for the 17-DMAG ligand, and Leu103, Tyr139, Val150 and Trp162 for the GVK0153 ligand. These unique contacts are located at the bottom of the binding pocket for the GVK0153 ligand and around the top of the pocket for the 17-DMAG inhibitor. Looking at the X-ray structures of the protein-ligand complexes, it is evident that the GVK0153 ligand inserts deeply into the pocket, whereas the larger 17-DMAG ligand sits closer to the top of the pocket. The LigPlot analysis also suggests that for both ligands, most of the interactions are via hydrophobic contacts. There is only one hydrogen bond between the GVK0153 ligand and the Asp93 residue of the protein (Figure 3.3). The 17-DMAG ligand on the other hand makes four H-bonds; with residues Lys58, Asp93, Phe138 and Thr184 (Figure 3.3.). The smaller GVK0153 ligand has fewer contacts - to only 11 amino acids, compared to 17 different amino acid contacts made by the larger 17-DMAG ligand.

The contacts between Hsp90-NTD and its natural substrate ADP were also analysed using LigPlot (PDB:1BYQ) [131] to allow comparisons. Here, the nucleotide and the protein make the same contacts that were also shared between both 17-DMAG and GVK0153 ligands; *i.e.* Asn51, Ala55, Asp93, Met98, Leu107, Phe183 and Thr184, as shown in the Figure 3.2. There are several additional interactions between ADP and Hsp90-NTD, that include the residues Asn106, Val136 and Gly137. These are a subset of the contacts that were unique to the 17-DMAG ligand. These residues are located on the top of the binding pocket and indeed, ADP does not insert as deeply into the pocket as the GVK0153 ligand. The ADP molecule forms three hydrogen bonds with the protein; to Asn51, Phe138 and Asp93. There are no crystal structures of human Hsp90 in complex with ATP, only of yeast Hsp90 or human endoplasmic reticulum Hsp90 paralogue. These crystal structures, whilst similar in terms of structure and sequence, are not identical to the human

cytosolic protein. Thus the ligand-protein contact analysis, which would have allowed direct comparison, could not be done.

The solvent accessible surface areas (SASA), which can been used to estimate the solvation contribution to the global thermodynamics, were calculated for the ligand:Hsp90-NTD complexes (PDB:1OSF and unpublished AstraZeneca structure) as well as the apo protein 'open' and 'closed' structures (PDBs: 1YES and 1YER respectively) [133, 134]. As expected from the RMSD analysis (which suggested high similarity between all the crystal structures), the total surface areas are also similar, ranging from 10,197 $Å^2$ to 10,868 $Å^2$. The overall surface areas of the ligand bound structures are slightly larger than that of the apo structures; a difference of 560-670 $Å^2$ for the apo and GVK0153 bound structures and 220-340 $Å^2$ difference between the apo and 17-DMAG:Hsp90-NTD complex. There is also a small difference between the surface areas of the two ligand bound structures, where the GVK0153 bound structure is ~350 $Å^2$ larger (see Table 3.1). The two apo Hsp90-NTD crystal structures also differ slightly in their overall surface areas. Recall that the two apo states are defined as 'open' and 'closed'. The overall surface area of the 'closed' structure is slightly larger. Similarly, the GVK0153 bound structure surface area is larger than the 17-DMAG bound protein. As both 'closed' structures (apo Hsp90-NTD and GVK0153 bound protein) have slightly larger surface areas, it is likely that the loop 4 orientation towards the pocket increases the overall surface area.

The surface areas were divided further into hydrophobic and hydrophilic for additional analysis. The nature of the surface area can affect the surrounding water structure and thus play a role in entropy [135]. The hydrophobic surface area of the GVK0153 bound protein is slightly larger than the 17-DMAG:Hsp90-NTD structure (2,732 $Å^2$ for 17-DMAG bound protein compared to 3,010 $Å^2$ for the GVK0153:Hsp90 complex). The hydrophilic surface areas of the ligand bound structures are more similar (~60 $Å^2$ difference). The hydrophobic surface areas of the apo "open" and "closed" structures on the other hand differ only by 10 $Å^2$. The apo protein hydrophilic surface areas follow a similar trend to the ligand bound protein, where the 'closed' loop 4 orientation results in a slightly larger hydrophilic surface area. Although there are some differences in the overall, and hydrophobic, surface areas between

the crystal structures, these are all very small (6-10%). They are unlikely to have a large effect on the binding energy. However, the data from the surface area analysis is used later in this chapter in an attempt to rationalise the role that water plays in the overall binding reactions.
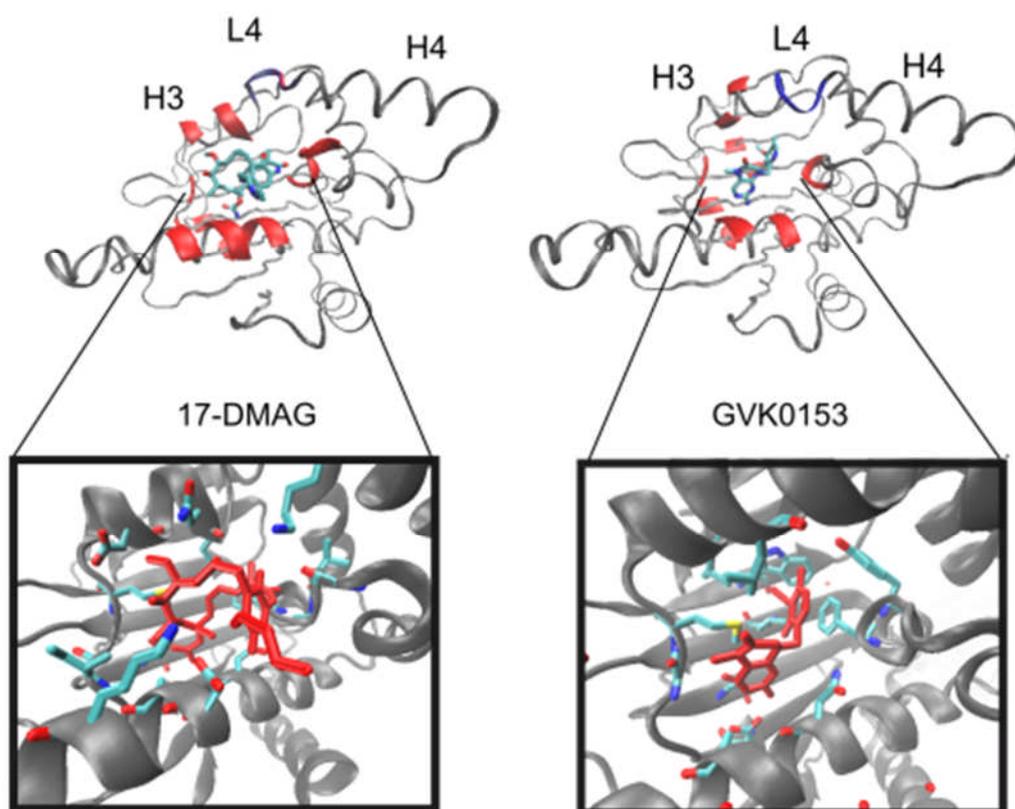
Figure 3.1: Crystal structures of Hsp90-NTD in complex with 17-DMAG and GVK0153. Both inhibitors bind to the same nucleotide binding pocket, that is located in the N-terminus of the protein, as can be seen in the top panel. The residues that are involved in the binding interactions, as predicted by LigPlot software, are coloured red in each protein structure on top panel. It can be seen that the larger 17-DMAG ligand makes more extensive contacts with the protein [116]. The loop 4 (L4), which is coloured in blue in the top structures, also has different orientation in the two different ligand bound complexes. When bound to the 17-DMAG, the loop is away from the binding pocket (*i.e.* 'open') and in contrast, the loop points towards the pocket in the GVK0153:Hsp90 structure (PDB: 1OSF for the 17-DMAG [115]; (structure from AstraZeneca, data not published). The bottom panel shows a close up of the amino acid contacts each ligand makes. It can be seen that the amino acids involved in the binding of the smaller ligand, on the right, are located towards the bottom of the binding pocket. The 17-DMAG contacts are mostly on the top edge of the pocket.

To summarise, the analysis of the crystal structures suggests that there are very minor structural differences between the ligand bound complexes. The crystal structure analysis data presented here is used later on in this chapter to interpret MD simulation data and the solvent accessible surface area estimations can help with interpretation of thermodynamic data from ITC.
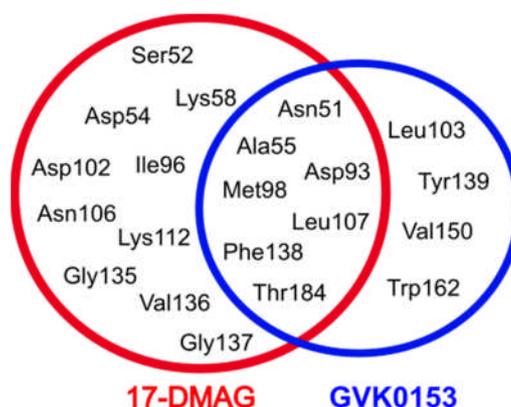
Figure 3.2: Summary of the residues involved in binding interactions between 17-DMAG and GVK0153 respectively, with the Hsp90-NTD, as predicted by LigPlot [116].

Figure 3.3: Polar contacts shown as dashed green line between (A) GVK0153 and Hsp90-NTD and (B) 17-DMAG and Hsp90-NTD as predicted by LigPlot [116].
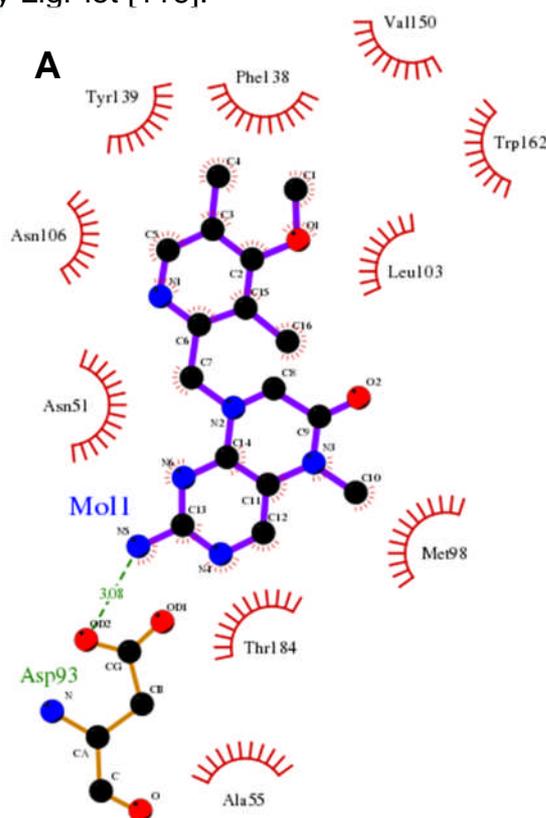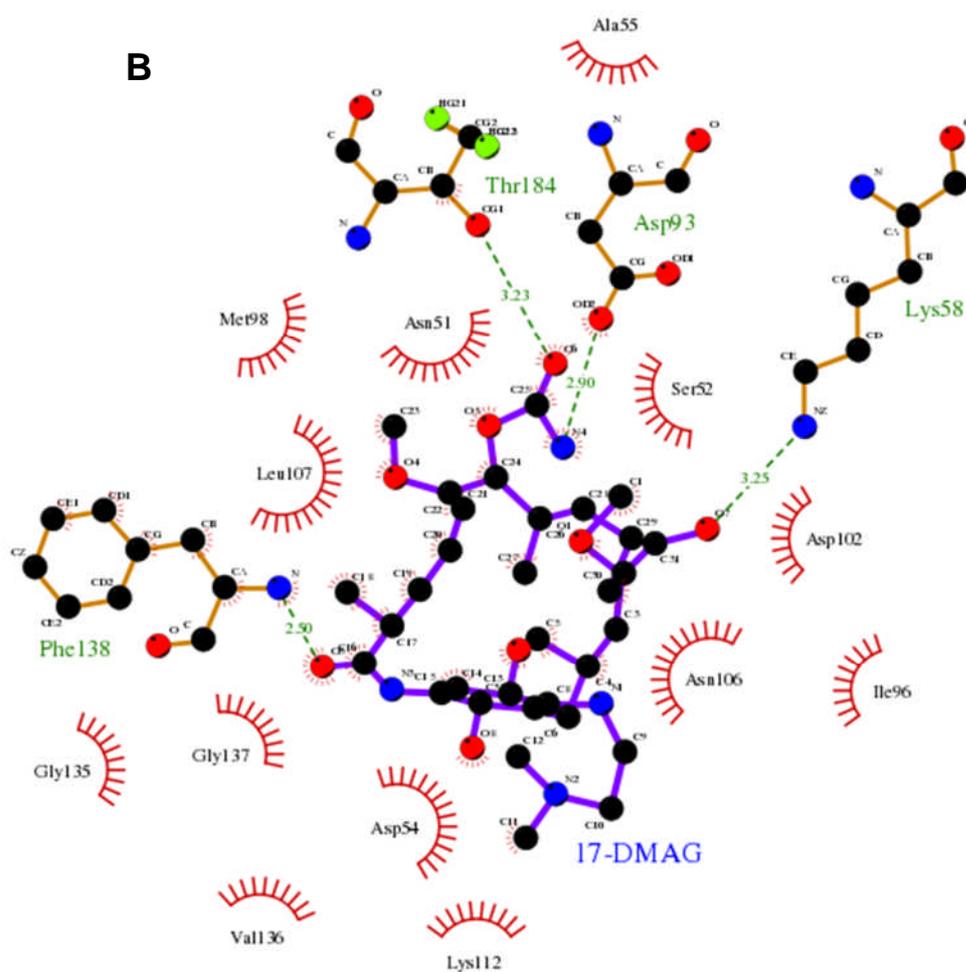
Table 3.1: The protein surface areas (solvent accessible, hydrophobic and hydrophilic) of the different crystal structures

Crystal structure	Total surface area (SASA) (Å²)	Hydrophobic surface area (Å²)	Hydrophilic surface area (Å²)
apo Hsp90-NTD "open" (PDB:1YES)	10,197	2,654	7,542
apo Hsp90-NTD "closed" (PDB:1YER)	10,311	2,644	7,667
17-DMAG:Hsp90-NTD (PDB:1OSF)	10,532	2,732	7,800
GVK0153:Hsp90-NTD (un-published)	10,868	3,010	7,859

3.2 ITC experiments on 17-DMAG and GVK0153 binding to Hsp90

Isothermal titration calorimetry (ITC) allows the determination of the thermodynamic parameters for ligand binding (dissociation and association constants as well as changes in Gibb's free energy, enthalpy and entropy) from a single experiment. ITC is a widely used technique in drug discovery [53]. In this project, ITC was used to study the binding of 17-DMAG and GVK0153 inhibitors to Hsp90-NTD. These measurements can be used to gain insight into any differences between the ligands, and to obtain global thermodynamic values. These values can be compared with those obtained from computer simulations.

The ITC data showed that both ligands bind to Hsp90-NTD, as the isotherms demonstrated that the Hsp90-NTD binding site became saturated upon titration of a ligand (Figure 3.4). The dissociation constant for both ligands was in the nanomolar range; 20 nM for the GVK0153 ligand and 54 nM for the 17-DMAG ligand. The free energy change for the binding of the smaller GVK0153 ligand was -9.14 kcal/mol; the enthalpy change was -9.71 kcal/mol and the $T\Delta S°$ term was 0.625 kcal/mol (Table 3.2). The $\Delta G°$ for the 17-DMAG ligand on the other hand was -8.57 kcal/mol. The entropy term ($T\Delta S°$) was the larger of the two contributors to the overall free energy change with -5.66 kcal/mol, compared to the enthalpy change of -2.91 kcal/mol for the 17-DMAG ligand (Table 3.2).

Table 3.2: ITC binding data for 17-DMAG and GVK0153

	$\Delta G°$ (kcal mol^{-1})	$\Delta H°$ (kcal mol^{-1})	$-T\Delta S°$ (kcal mol^{-1})	N
17-DMAG: Hsp90-NTD	-8.566±0.165	-2.906±0.110	-5.661±0.990	2.073±0.19
GVK0153: Hsp90-NTD	-9.140±0.100	-9.766±0.160	0.625±0.180	0.759±0.04

(± errors are standard deviations between repeats)
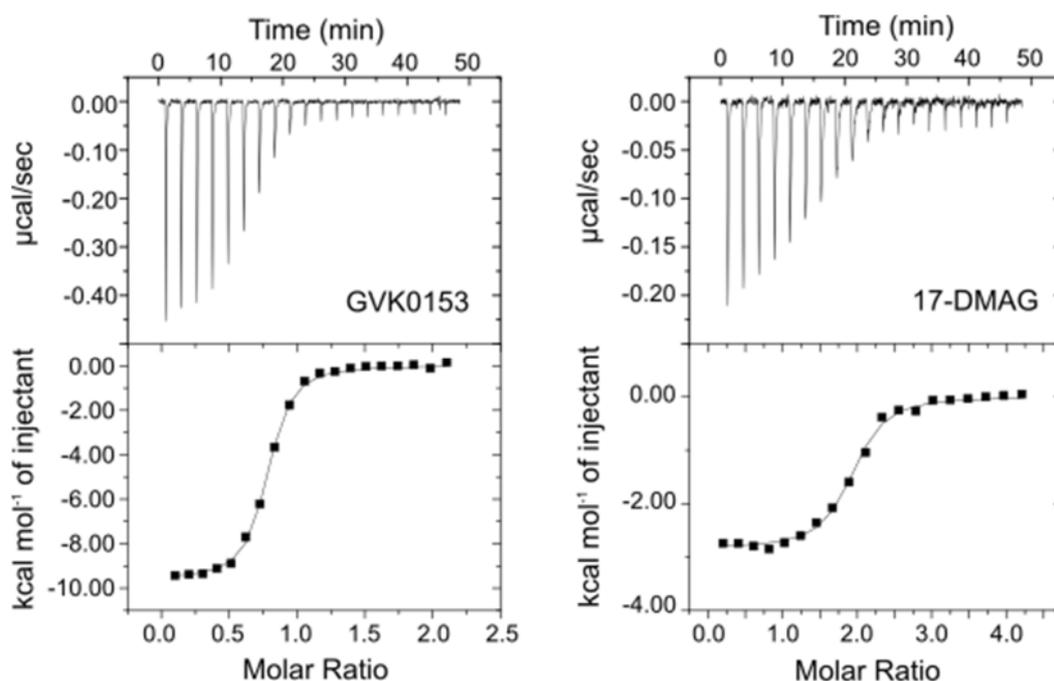
Figure 3.4: Isotherms from the ITC experiments for the 17-DMAG and GVK0153 ligands. Isothermal titration calorimetry experiments were carried out to determine the thermodynamic parameters for the binding of GVK0153 (left panel) and 17-DMAG (right panel) to Hsp90 N-terminal domain.

The titrations of both inhibitors were carried out in triplicate. In general, the ITC data were highly re-producible, as seen from Figure 3.5. The largest standard deviation was 159 cal/mol in the enthalpy values between the repeats, accounting for less than 2% of the total enthalpy value.

Whilst the ITC data were very reproducible, looking at the isotherms in Figure 3.4, it can be seen that the stoichiometry (n) is differs from the expected 1:1 binding constant for the 17-DMAG ligand. The stoichiometry value depends on the accurate estimation of both the ligand and protein concentrations, as well as the amount of active protein. However, the shape of the isotherm ('flat curve' indicative of a small heat change) observed for the 17-DMAG is typical for entropy driven binding. Additionally, crystal structures and other published data suggest that there is only a single binding site for the 17-DMAG ligand. Thus, we assume that the ligand and/or protein concentrations were inaccurate, rather than the ligand interacting with two sites. This error in stoichiometry does not change the thermodynamic parameters, and the

overall binding signature where entropy contribution is the major component to the overall binding free energy is still valid. This is because the heat change upon titration of ligand is not affected by stoichiometry value. Similarly, a large entropic contribution (~10 kcal/mol) to the binding affinity has also been observed with 17-DMAG binding to the full-length Hsp90 [136].
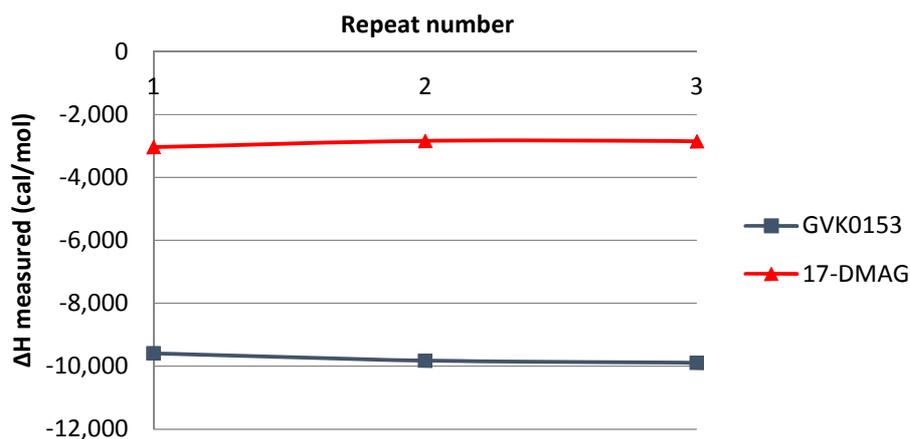
Figure 3.5: The repeat ITC measurements gave very consistent enthalpy (ΔH°) values, as seen from the graphs. The standard deviation for ΔH° values between the repeats was 110 cal/mol for the 17-DMAG and 159 cal/mol for GVK0153.

Now it is time to compare the two ligands. From the ITC data of Hsp90-NTD:GVK0153 and Hsp90-NTD:17DMAG, it was apparent that whilst the affinities are of a similar micromolar magnitude (dissociation constant K_d values were 20±0.02 µM for GVK0153 and 54±0.09 µM for the 17-DMAG), the two systems have different entropic contributions to the overall Gibb's free energy change (Figure 3.6 and Table 3.2). The binding of the smaller GVK0153 ligand is entirely enthalpy driven. In contrast, for the larger 17-DMAG ligand, the entropy term is the dominant contributor to the overall binding energy.
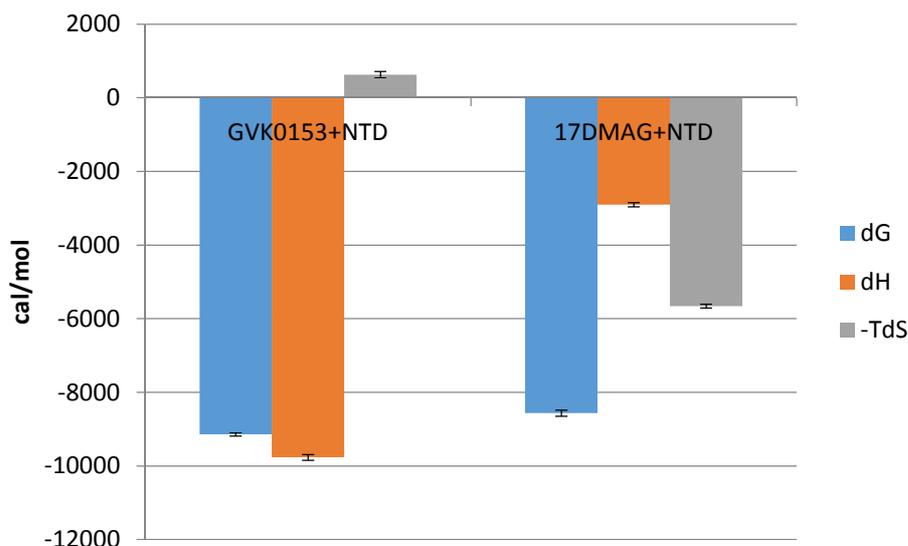
Figure 3.6: The binding signatures of GVK0153 and 17DMAG. Both inhibitors bind to the Hsp90 chaperone with similar affinity. The Gibb's free energy change (ΔG) is around 9,100 cal/mol for the GVK0153 ligand and 8,500 cal/mol for the 17-DMAG. However, the enthalpic and entropic contributions to the binding affinity differ between the ligands. The error bars show the standard deviations from repeat measurements.

Our ITC data suggest that 17-DMAG and GVK0153 have different thermodynamic binding signatures. This means that the ligands are likely to have different binding mechanisms. However, it is not possible to obtain mechanistic insights from these global ITC parameters. The affinity (K_d) and Gibb's free energy (ΔG) terms are straightforward to understand in terms of binding in drug design. Normally compounds with tighter binding are selected for further development. The enthalpy and entropy terms on the other hand are not as straight forward. Negative enthalpy values are caused by bond formations resulting in the release of heat. The net ΔH° effect, as observed by ITC, is a result of a large number of bonds being formed and broken [135]. The entropy change is similarly complex. A negative value for the –TΔS° term (*i.e.* when ΔS° is positive) reflects an increase in the disorder of the system. This can arise from increase in the ligand and/or protein flexibility, or from the release of water molecules to bulk solution. Thus, enthalpy and entropy terms are difficult to interpret as it is not possible to separate the contributions from the protein, the ligand and changes in solvation from the global ITC

parameters. Indeed, the large difference of ~6,200 calories observed in entropic binding signatures between 17-DMAG and GVK0153 (Figure 3.6, Table 3.2), could arise from changes to the protein and ligand dynamics upon complex formation, from solvent re-organisation due to the release of tightly bound water to bulk solvent, or from changes to the burial of hydrophobic residues [8].

In conclusion, the ITC data suggest that the binding of the larger 17-DMAG and smaller GVK0153 ligands to Hsp90-NTD have different balance of driving forces. 17-DMAG binding has larger entropic contribution, whereas GVK1053 binding is entirely enthalpy driven. These differences in the enthalpy and entropy contributions to the overall binding signatures are not unique to Hsp90 ligands, but have been observed in many different systems [137, 138]. Whilst the global values suggest differences in their binding mechanism, the structural and molecular details cannot be obtained from the thermodynamic data alone. To obtain mechanistic insights into the binding of these two inhibitors, we used NMR and MD experiments, that allow the investigation of the interactions in atomistic detail.

3.3 NMR experiments on Hsp90-NTD complexes

Chemical shift perturbation analysis was used to characterise conformational changes in Hsp90-NTD, when in complex with the 17-DMAG and GVK0153 ligands. To characterise the changes in µs-ms dynamics upon ligand binding, we employed CPMG relaxation dispersion experiments. As a preliminary step, protein backbone assignments, in the presence and absence of inhibitors, were obtained using previously published assignments for the apo Hsp90-NTD as a starting point.

3.3.1 Hsp90-NTD backbone resonance assignments

The backbone assignments for apo and inhibitor bound Hsp90-NTD were obtained using a TROSY versions of 3D triple resonance HNCA, HNCO and HNcaCO experiments [109, 110]. These experiments were described in Chapter 1.

NMR backbone assignments of the apo state of the human Hsp90-NTD have been obtained previously (BMRB code 7003) [61]. These assignments were transferred to our apo Hsp90-NTD sample, which was prepared using similar experimental conditions. We recorded a set of triple resonance spectra for the apo Hsp90-NTD; all of them were well resolved. For the majority of the residues, peak positions in our apo protein spectra matched the published assignments. The original assignments included amide peaks for 152 out of 205 non-proline residues (there are total of 209 residues in the Hsp90-NTD construct). Assignments were transferred for 145 residues. We could not transfer assignments for seven residues (Ala27, Lys100, Leu103, Asn105, Asn106, Met119, Asp157), most likely due to significant line-broadening observed in the NMR spectra for these residues. Additionally, we identified nine further residues (Thr65, Leu76, Ala166, Gly167, Thr176, Glu200, Gln212, Phe213, Ile214). In total, we obtained assignments for 74% of the apo Hsp90-NTD.

The sets of 2D HN TROSY, 3D HNCA, HNCO and HNcaCO experiments were performed for the 17-DMAG:Hsp90-NTD and GVK0153:Hsp90-NTD complexes. The 2D amide TROSY spectra of 17-DMAG and GVK0153 bound Hsp90-NTD are overall similar to the spectrum of the apo protein (Figure 3.7). Thus, the apo-protein assignments were used as a starting point to assign the 17-DMAG and GVK0153 bound Hsp90-NTDs. For the residues, which have similar or identical peak positions in the spectra of the apo and the ligand bound protein, backbone assignments were transferred from the apo protein. These assignments were confirmed using HNCA spectra. Approximately 30% of Hsp90-NTD residues experienced significant changes in peak positions upon 17-DMAG or GVK0153 binding. The positions of the majority of the shifted peaks differed between the two ligands (Figure 3.8). The majority of shifted peaks were assigned using a set of 3D triple resonance experiments. In total, we obtained assignments for 72% of the Hsp90-NTD:17-DMAG complex and 64% for the Hsp90-NTD:GVK0153 complex. Figure 3.9 shows the Hsp90-NTD structures, where the assigned *versus* un-assigned regions for the two protein-inhibitor complexes, as well as for the apo-Hsp90-NTD are highlighted.

Figure 3.7: Superposition of the NH TROSY spectra of apo (black contours) and 17-DMAG:Hsp90-NTD (red contours). Whilst the majority of the peaks overlap between the apo and the 17-DMAG bound spectra, there are also significant shifts in the peak positions for certain residues (see Figure 3.7 below). This indicates structural changes. The cross peaks are labelled for the apo state. The assignments for the crowded area in the middle of spectra are shown in the insert underneath. Here, only the apo peaks are shown for clarity. The 17-DMAG spectra was recorded at 600 MHz magnet and 950 MHz magnet was used for the apo sample.

Figure 3.8: A zoomed in view of the NH TROSY spectra shown in Figure 3.6. The red contours correspond to 17-DMAG:Hsp90-NTD, blue contours to GVK0153:Hsp90-NTD and black contours with labelled peaks to apo Hsp90-NTD. The peaks, which shift their position upon ligand binding, move differently in direction and in magnitude, depending on the ligand. For example, the position of the blue and red peaks next to residues G137 and T171 are different. The red peaks corresponding to 17-DMAG bound protein move upwards and the blue peaks corresponding to the GVK0153 bound protein move downwards. The 17-DMAG and GVK0153 bound spectra was recorded at 600 MHz magnet and 950 MHz magnet was used for the apo sample.

apo Hsp90-NTD   Hsp90-NTD:17-DMAG   Hsp90-NTD:GVK0153
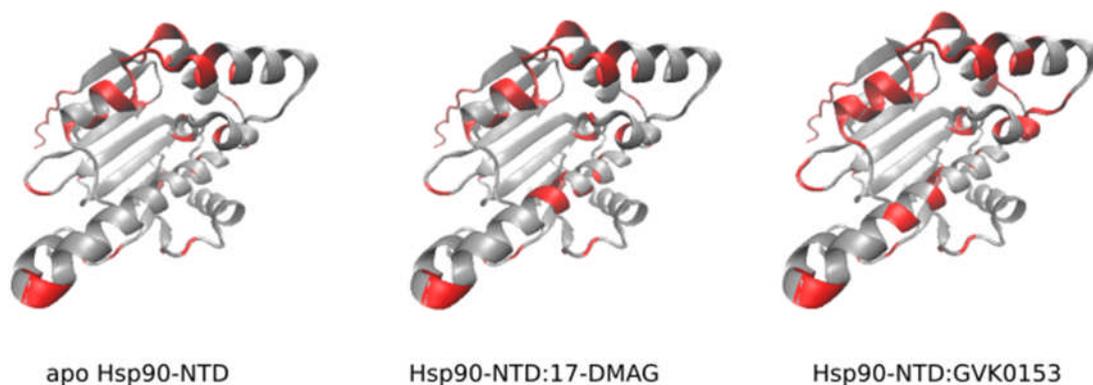
Figure 3.9: The Hsp90-structures, showing the amino acids for which we have obtained NMR assignments. The regions of the Hsp90-NTD that were assigned are shown in blue. The grey colour indicates the residues where no assignments were obtained.

3.3.2 Chemical shift perturbation analysis

To identify residues directly affected by ligand binding, we carried out chemical shift perturbation analysis (CSP). This also identifies possible ligand-induced long-range perturbations in the protein conformation [57]. The chemical shift difference of Hsp90-NTD in the presence, and absence, of its inhibitors enables the identification of residues involved in the binding interaction. Additionally, long range perturbations may also be observed in the residues that are located far from the binding site. For this study, the crystal structures of Hsp90-NTD were used in combination with the CSP data to identify the residues involved directly in the binding interactions. The crystal structures also helped to identify the residues with long-range perturbations located away from the binding site.

For the CSP analysis, the changes in peak positions upon ligand binding were calculated as described in Chapter 1, using the following equation:

$$\Delta\delta_{tot} = \sqrt{(\Delta\delta_H)^2 + (0.154\Delta\delta_N)^2)}$$
(eq. 3)

To define statistically significant changes we used the following criteria: residues with significant CSPs should have at least one of their amide chemical shifts (*i.e.* δ_H or δ_N) larger than 0.06 or 0.6 ppm respectively. The 0.06 and 0.6 ppm values for proton and nitrogen chemical shifts correspond

to two standard deviations, calculated from all CSPs observed upon ligand binding [57, 139]. We also defined residues that had undergone large changes as those with $\Delta\delta_{tot} > 0.3$ ppm. An example of the significant and non-significant changes in the peak positions is shown in Figure 3.9.
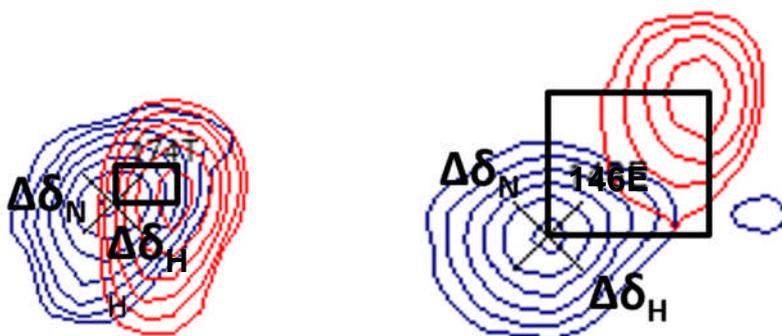
Figure 3.10: Examples of the peaks, for a 'non-significant CSP change' (left hand peaks) and significant CSP change (right hand peaks). The significant changes were classified as either $\Delta\delta_{tot}$ of larger than 0.3 ppm, or δ_H or δ_N larger than two corresponding standard deviations; *i.e.* 0.06 or 0.6 ppm respectively.

Using CSP analysis, significant and large changes were observed for 39 of the 149 assigned residues upon binding of the 17-DMAG ligand. For the GVK0153 ligand, this figure was changes in 50 of the 133 assigned residues (Figure 3.10). As expected, most of the significant changes were observed in the residues directly involved with the inhibitor binding (residues that are coloured red the crystal structures in Figure 3.1). These residues were defined as those that were within 5 Å of the ligand in the X-ray crystal structures [140]. About half of the residues that underwent significant chemical shift perturbations were classified as long range perturbations, *e.g.* CSPs observed for the residues located at least 5 Å away from the corresponding ligand. These changes are coloured in blue in the crystal structures in Figure 3.11. These CPSs indicate structural and/or dynamic changes in residues that are not directly involved in the ligand interactions.

Similar long range chemical shift perturbations upon ADP/ATP binding to Hsp90-NTD have also been observed in yeast Hsp90 [141] and also in human Hsp90-NTD [50], as shown in the bottom panel of Figure 3.11. Interestingly,

some of these changes occurred in the alpha helices that make up the 'lid' region of the protein (Ala111-Val136) (grey panel in Figure 3.11). This region is known to be highly dynamic; no residues could be assigned in the apo Hsp90-NTD spectra for the start of the lid region (residues 111-116). This is often suggestive of flexible residues that are involved in microsecond to millisecond motions, and thus too broad to be observed in NMR spectra. The B-factors of the X-ray structures that can indicate atomic vibrations are also higher for the two alpha helices making up the lid region [61, 125]. A previous NMR study looking into Hsp90-NTD structure, when the chaperone is in complex with ADP or ATP-mimic (non-hydrolysable AMPPNP), found that the lid region structures differed in these complexes. The residues making up the lid region were absent on the AMPPNP bound spectra, which the authors took to be indicative of microsecond dynamics. These peaks were observed for the ADP bound structure. This suggests that the ADP bound protein does not undergo microsecond timescale dynamic changes [50]. From our data, this seems to case for the 17-DMAG and GVK0153 complexes. As can be seen from Figure 3.11, the smaller GVK0153 inhibitor causes more shifts within the residues making up the lid region (highlighted with grey background in Figure 3.11), compared to the 17-DMAG inhibitor. For the GVK0153 ligand, seven peaks disappeared upon binding, suggesting large changes in conformation and dynamics. An additional two residues showed large perturbations. In comparison, only three residues showed large perturbations in the 17-DMAG:Hsp90-NTD complex within the same (Ala111-Val136) segment. This suggests that the binding of the smaller inhibitor causes larger changes to the protein structure, or dynamics, within the lid region.
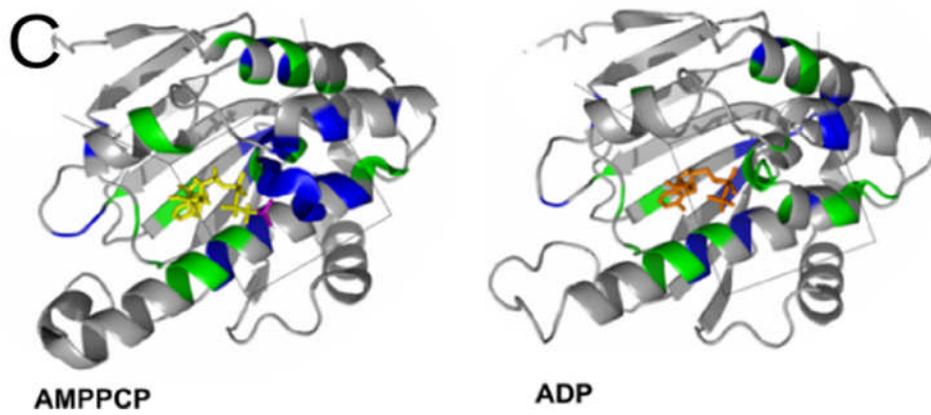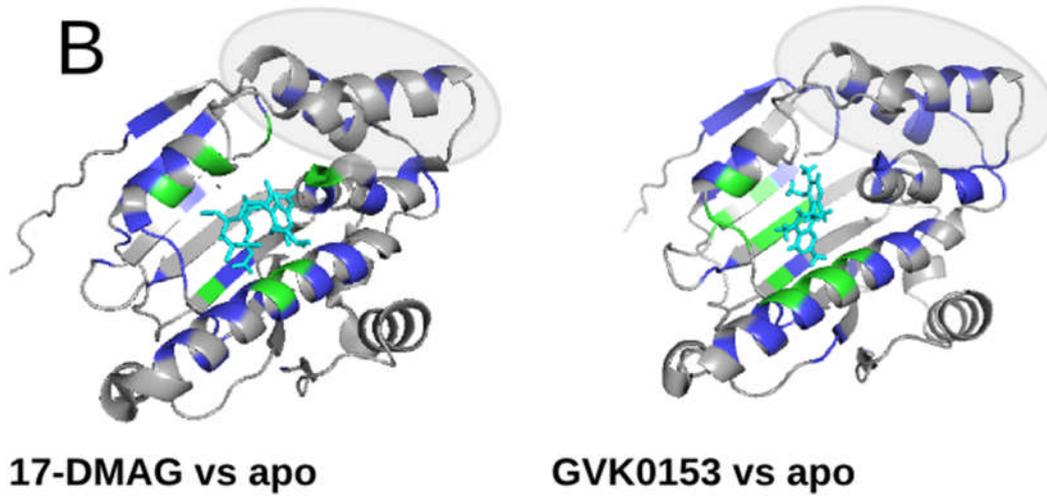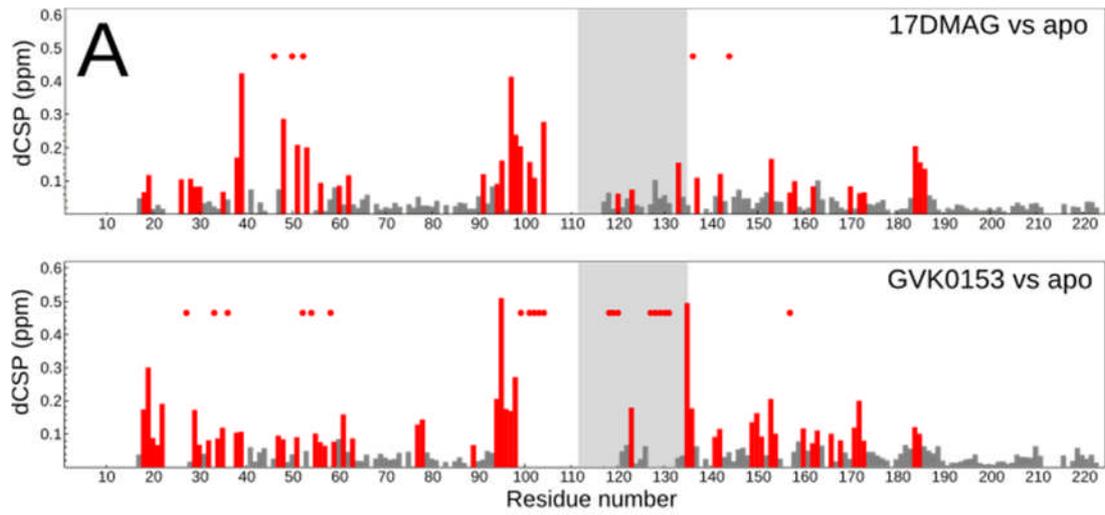
Figure 3.11: Amide backbone chemical shift perturbation analysis (CSP) show changes in the peak positions both for residues making up the binding pocket, as well as long range perturbations. (A) The histograms of the CSP changes show per residue perturbations, where the red bars are the residues with large ($\Delta\delta_{tot}$>0.3 ppm) or significant changes (where at least one of the shifts, *i.e.* δ_H or δ_N is larger than two corresponding chemical-shift errors; *i.e.* 0.06 or 0.6 ppm respectively). The red circles indicate residues that disappeared upon ligand binding. CSPs for individual residues were calculated using the differences in the chemical shifts for backbone amide ^{1}H ($\Delta\delta_H$) and ^{15}N ($\Delta\delta_N$) using the equation $\Delta\delta_{tot} = \sqrt{(\Delta\delta_H)^2 + (0.154\Delta\delta_N)^2}$. The lid region (Ala111-Gly135) is highlighted in grey in the histograms. This area is also circled in grey in the 17-DMAG and GVK0153 bound structures. (B) The crystal structures in panel A show the CSP changes upon GVK0153 and 17-DMAG inhibitor binding. Here, the green colour indicates the residues where significant CSP changes were observed and that are involved in ligand contacts. The blue coloured regions are those were significant CSP changes were observed in residues that are not involved directly in the ligand interactions (further than 5 Å of the ligand). (C) In comparison to small molecule inhibitor induced CSP changes, panel C shows the CSP changes reported upon AMPPCP or ADP binding. Here the changes are similarly observed both in the residues that are in direct contact with the ligand (green colouring), as well as those not involved in direct ligand interactions (blue coloured residues). The bottom AMPPNP and ADP bound structures were adapted from [50].

3.3.3 Relaxation dispersion measurements

Are the differences in the CSP profiles caused by changes in protein millisecond-microsecond dynamics? To further examine the differences in CSP pattern observed in the ligand binding signatures, we used ^{15}N relaxation dispersion measurements (Carr-Purcell-Meiboom-Gill, CPMG [55]) on the apo Hsp90-NTD and the protein in complex with the two ligands. This method can report on protein dynamics (*i.e.* structural changes) in the micro- to milliseond timescale as explained in Chapter 1. Intermediate or slow time scales were thought to be relevant for the Hsp90, as broad peaks were observed in the Hsp90-NTD spectra. Also, previous studies looking into the differences in ADP or ATP complexed chaperone suggested that slower timescale motions, rather than fast pico- to nanosecond ones, are relevant to the Hsp90-NTD dynamics due to line broadening [50].

In our study, ^{15}N CPMG relaxation dispersion profiles were measure for apo Hsp90-NTD and the protein complexed with the two ligands. A total of 97 residues for the GVK0153 bound protein and 109 residues for the 17-DMAG

bound and apo Hsp90 were used for our analysis. All these residues have non-crowded peaks and thus, any change in peak intensities can be unambiguously analysed. Relaxation dispersion profiles (Figure 3.12) indicative of micro- to millisecond dynamics were observed for 17 residues, located in multiple regions of the Hsp90 nucleotide binding domain. Examples of the dispersion data plots are shown in Figure 3.13. The dispersion data plots for all residues with micro- to millisecond dynamics are shown in Appendix.

Micro- to millisecond timescale changes in structure were observed in both apo and ligand bound proteins around the histidine 154, which is located in a loop region in the N-terminal domain and middle domain interface of the Hsp90 protein. Similarly, micro- to millisecond dynamics were observed for all protein states for residues Glu158 and Asn155, which are close to the His154. Additionally, for GVK0153-Hsp90-NTD complex, μs-ms dynamics was also observed for Asp175 that is located close in space to His154 in the GVK0153:Hsp90-NTD spectra. The peaks for this residue were located in over-crowded regions of the 17-DMAG and apo spectra, so for these states data are not available. These micro- to millisecond dynamics around His154 are likely to be reporting on the different protonation states of the His154 imidazole ring at pH 7.5, where depending on the orientation of the ring, the H-bond pattern with surrounding residues is likely to differ.

Micro- to millisecond dynamics around His154 (His154, Asn155 and Glu153) were observed for all the different protein states (apo and ligand-bound). This suggests that the experimental conditions (including pH and temperature) in the different samples were the same or very similar. Consequently, any additional changes in μs-ms dynamics observed between apo and ligand-bound states of Hsp90-NTD should report on ligand effects and are not caused by small variations in pH, experimental temperature, protein concentration, *etc.* between different samples.

In addition to His154 protonation, our analysis revealed several other regions with micro- to millisecond dynamics (Figure 3.12). Moreover, we also observed changes in the micro- to millisecond dynamics between different ligand-bound states. One of these regions includes residues within and

around the helices making up the lid region (Ala111-Gly135) (Figure 3.12, residues around Ile34). These near-lid residues showing µs-ms dynamics in the apo and GVK0153 bound state, but not in the 17-DMAG bound state. These residues include Ile34, Ile43, Gly135, Val136, Ser164 and Ala166. Residues Ile26, Ile128 and Ser129 were excluded from the analysis because CPMG dispersion data were available only for the apo state, while no data were available for either ligand complex (see Table 3.3).
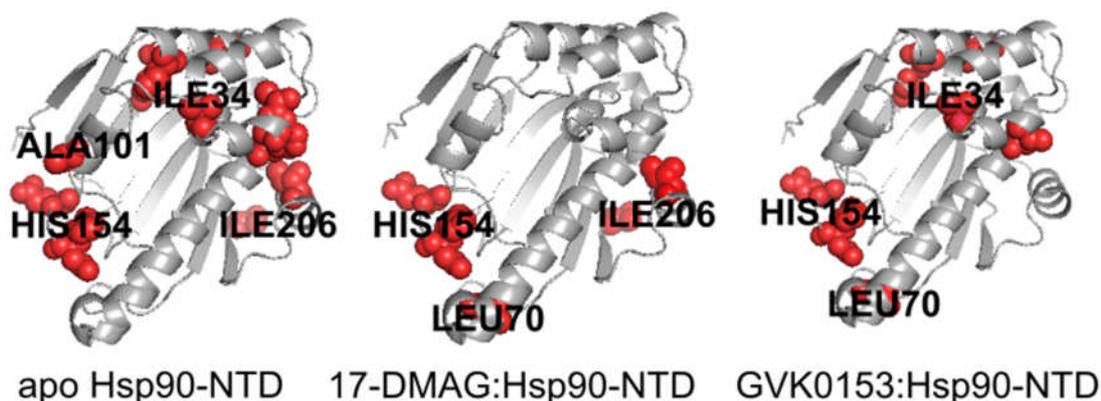
Figure 3.12: Hsp90-NTD structures, where red spheres are highlighting the residues for which relaxation dispersion profiles indicative of µs-ms dynamics, were observed. All the complexes show µs-ms dynamics for and around the His154 residue. The apo and GVK0153 bound proteins additionally show dynamics around the lid region (residue Ile34 and those surrounding it). These dynamics are absent from the 17-DMAG bound structure. There are further two regions, where intermediate timescale dynamics were observed; around Leu70 for the ligand bound structures and the C-terminal residues (Ile206) for the 17-DMAG and apo structures. No data were available for the Leu70 residue for the apo structure and C-terminal residues for the GVK0153 bound protein.

These findings suggest that the lid became more rigid on the micro-to millisecond timescale upon binding of the larger 17-DMAG ligand. Interestingly, previous crystallographic data (Table 3.1) suggests that the larger 17-DMAG ligand makes a contact with the glycine 135 and valine 136 [115]. No lid contacts with the smaller GVK0153 were observed (unpublished X-ray structure data from AstraZeneca). Thus, we hypothesise that the Val136 interaction in the 17-DMAG:Hsp90-NTD complex stabilises the whole lid and the surrounding region.

In addition to the residues located near His154 and the lid; residue Leu70 showed micro- to millisecond dynamics for both ligand bound forms (Figure

3.12). Unfortunately, for the apo protein, the peak corresponding to Leu70 was located in a crowded area and CPMG data for the apo state could not be
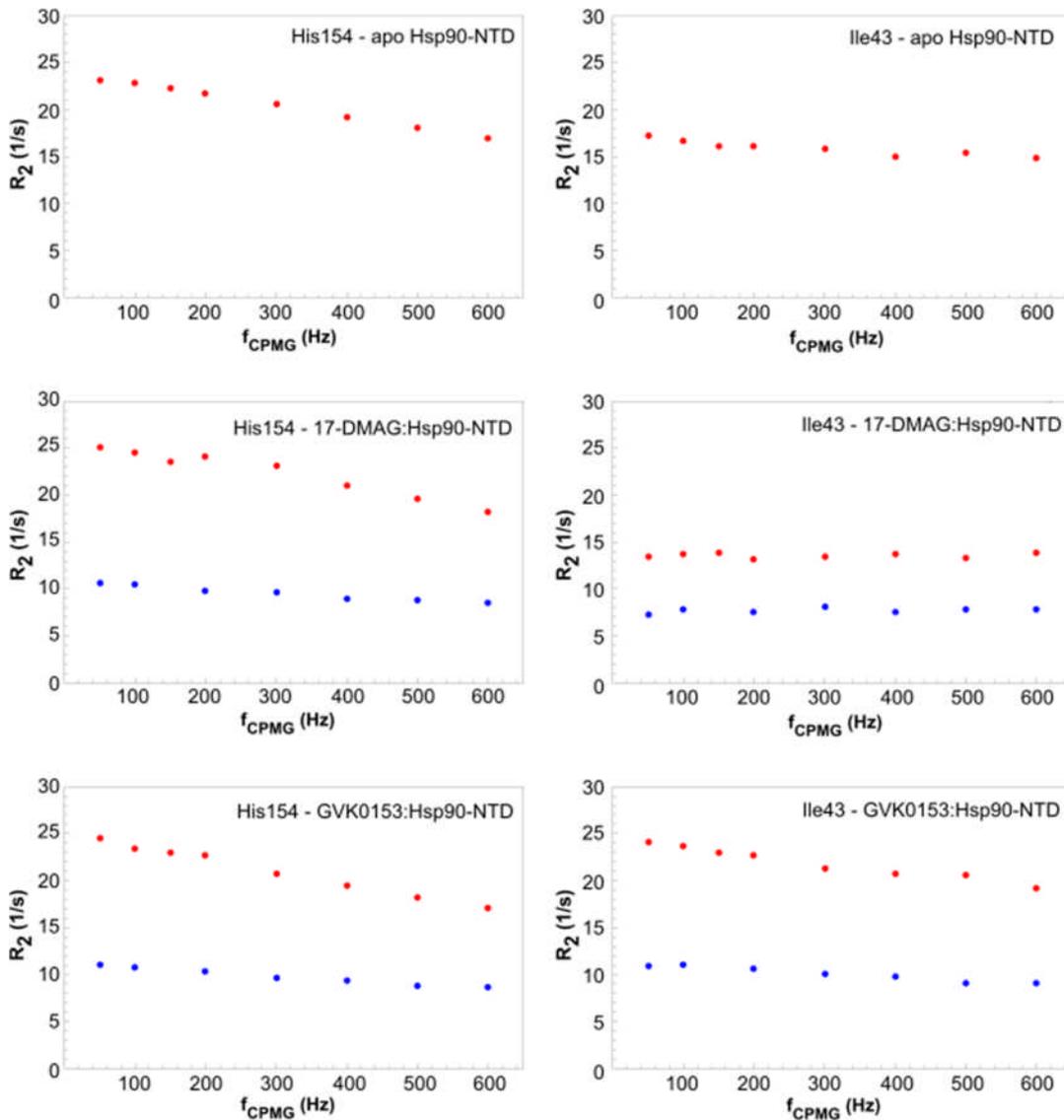
Figure 3.13: Examples of the CPMG relaxation dispersion profiles. The His154 showed relaxation dispersion profiles indicative of µs-ms dynamics for all complexes (left panel of graphs). In comparison, for the Ile43 shown on the right, relaxation dispersion profiles indicative of µs-ms dynamics were observed only for the apo and GVK0153 complex. The 17-DMAG bound protein shows a flat profile for this residue. The red data points correspond to data obtained from the 950MHz magnet and the blue ones to data from the 600 MHz one.

analysed. Leu70 is located at the end of a long alpha helix (helix 2). The C-terminal end of the helix 2 has no contacts with other parts of the Hsp90-NTD,

and thus, enhanced flexibility is expected for Leu70. In line with these CPMG results, our molecular dynamics simulations also show enhanced flexibility for this region (see Section 3.6.2 below).

Dynamics on a µs-ms timescale were also observed for Ala101 in the absence of ligands. No µs-ms dynamics were observed in the presence of 17-DMAG. CPMG data were not available for the GVK0153 bound protein, as no Ala101 peak was found in NMR spectra for GVK0153 bound Hsp90-NTD. Ala101 is located relatively close to His154 and the non-flat dispersion profile for Ala101 may be due to His154 protonation/deprotonation. Alternatively, Ala101 dynamics may be linked to micro- to millisecond dynamics around the lid area, as Ala101 is located at the edge of the flexible region making up the lid [125].

Finally, residues Ile206 and Val222, located near the C-terminus, showed non-flat relaxation dispersion for the apo and 17-DMAG bound proteins. Unfortunately, these peaks overlap in the GVK0153 bound spectra and were not included in the analysis. The observed micro- to millisecond dynamics for these residues may be due to the fluctuation of the C-terminal part of the NTD.

In summary, the analysis of CPMG data revealed several regions in Hsp90-NTD that are affected by micro- to millisecond dynamics. These include the functionally important lid region. We found that for the smaller GVK0153 ligand and apo protein there are lid fluctuations on the micro- to millisecond timescale. In contrast, the larger 17-DMAG, which interacts directly with the lid residues Gly135 and Val136, halts lid fluctuations.

84

Table 3.3: Summary of the residues with observable relaxation dispersion profiles.

Residue number	apo Hsp90-NTD	17-DMAG:Hsp90-NTD	GVK0153:Hsp90-NTD
Ile26	+	*no data*	*no data*
Ile34	+	-	+
Ile43	+	-	+
Leu70	*no data*	+	+
Ala101	+	-	*no data*
Ile128	+	-	*no data*
Ser129	+	-	*no data*
Gly135	-	-	+
Val136	+	*no data*	+
His154	+	+	+
Asn155	+	+	+
Glu158	+	+	+
Ser164	+	-	+
Ala166	+	-	+
Asp175	*no data*	*no data*	+
Ile206	+	+	*no data*
Val222	+	+	*no data*

(+) Relaxation dispersion observed (*i.e.* μs-ms dynamics) and (-) no relaxation dispersion (*i.e.* no μs-ms dynamics)

3.4 Molecular dynamics simulations and flexibility

What kind of motions does the Hsp90 protein undergo? NMR is a powerful technique, reporting on dynamic processes in atomistic detail. The results of the NMR experiments tell us that something is happening, but not what. To study the possible causes of the motions and changes in the structure

observed by the NMR techniques, we ran 500 ns full atom molecular dynamics simulations of the three systems (apo protein, and Hsp90-NTD in complex with 17-DMAG or GVK0153) in triplicates.

3.4.1 Measuring flexibility using root mean square fluctuations

To start the MD data analysis, root mean square fluctuations (RMSF) of the backbone C_a positions were calculated. The RMSF is used as a convenient and quick method to quantify differences in protein flexibility between the apo Hsp90 and the ligand bound protein. The calculation of RMSF of the protein structures takes a matter of minutes, compared to a few days of computation time for more complex principal component analysis. The fluctuations in the C_a backbone atoms were calculated separately for each replica trajectory. These data were combined to calculate the average fluctuation of a complex with standard errors. The RMSF data were plotted as histograms, that show per residue fluctuations for each system. These histograms are shown in Figure 3.14. From the histograms it can be seen that the N-terminal and C-terminal coil regions, as well as various loop regions within the protein, were highly flexible, as expected for non-structured elements. Pairwise comparisons of the RMSF values demonstrate that the apo structure is more flexible than either of the Hsp90-ligand complexes studied. These data suggest that the ligand-protein interactions stabilise the protein. Interestingly, there were differences in the backbone flexibility when the protein was in complex with the 17-DMAG compared to GVK0153 bound structure. Here, the 17-DMAG complex showed slightly higher fluctuations around two coil regions (end of the long helix two and coil connecting helix 4 and 5). In comparison, the residues making up the alpha helices in the lid region (around residues 110 to 135) were more flexible in the GVK0153 bound protein. These differences are highlighted in the Hsp90-NTD crystal structure in Figure 3.15. In the structure, the red coloured areas refer to the regions where the 17-DMAG bound protein was more flexible. The blue coloured region correspond to the areas of greater flexibility in the GVK0153 bound protein. The differences observed in the lid region using the RMSF data agree with the NMR relaxation dispersion measurements. In these, the 17-DMAG bound protein did not have any µs-ms dynamics around this region but the GVK0153:Hsp90-NTD complex did.
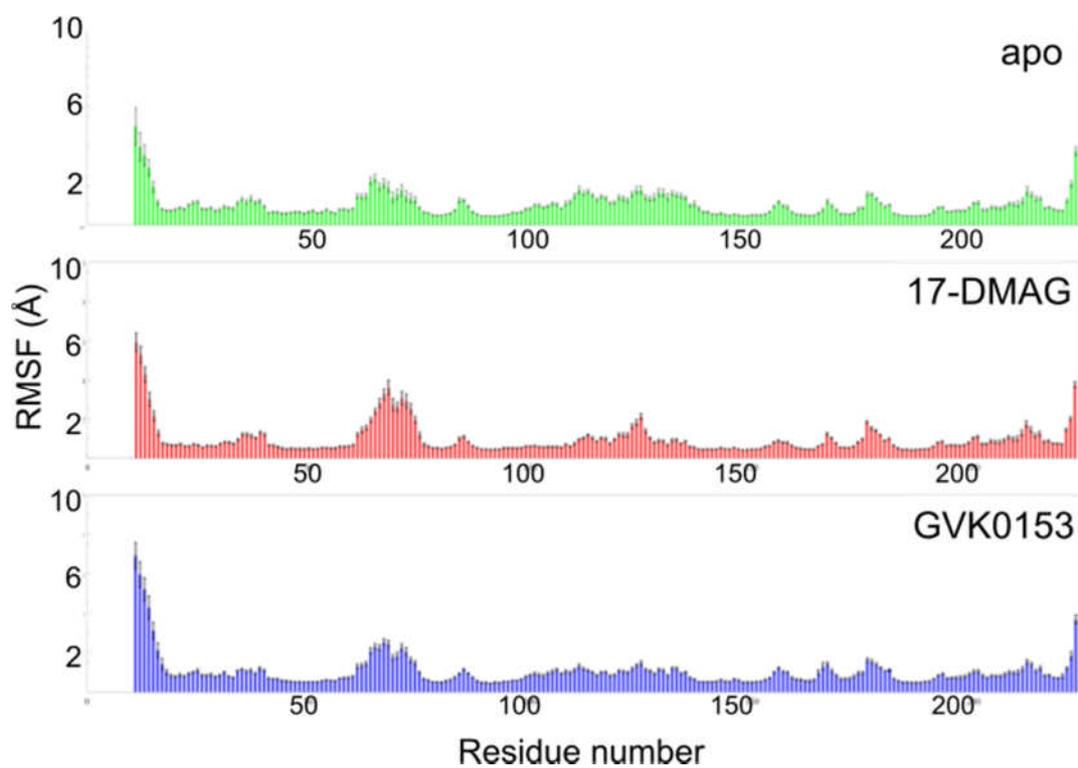
Figure 3.14: RMSF of the C_a backbone atoms calculated from MD simulations. The most flexible areas, seen as large bars in the graphs, correspond to the N- and C-terminal residues as well as residues making up loops and random coils.
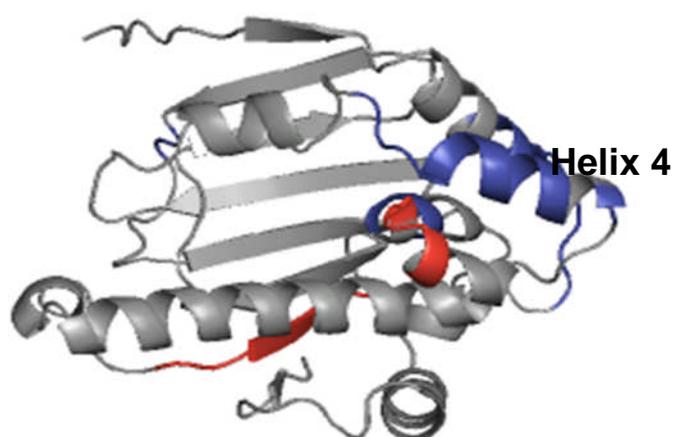
Figure 3.15: Hsp90-NTD structure coloured to highlight the regions of greater flexibility, as observed in the pairwise comparison of the C_a backbone of GVK0153 and 17-DMAG bound Hsp90-NTD by RMSF analysis. The regions coloured as red correspond to the more flexible parts observed in the 17-DMAG:Hsp90-NTD complex. The blue coloured regions were more flexible in the GVK0153 bound protein.

3.4.2 Principal component analysis

The RMSF analysis showed that there were differences in the backbone C_a flexibility between the apo and the ligand bound Hsp90-NTD complexes. To analyse these movements further, we used principal component analysis (PCA). For PCA, the Cartesian coordinates of atom positions from simulations are converted into eigenvectors. The largest eigenvector corresponds to the biggest movements. The power of the PCA method can be thought as "helping to see the wood from the trees". This is because the largest eigenvectors can be used to analyse concerted motions, such as movements of alpha helices, rather than attempting to make sense of the 'wiggles and jiggles' of individual atom positions of unprocessed MD trajectory data. The PCAzip tool used in this work also allows the generation of short animations illustrating the movement of the molecule along the eigenvector of choice [93]. These movies of the first few eigenvectors, that correspond to the motions with largest amplitude, can be used to identify flexible regions of the protein, and to visualise the motions.

As PCA reports on the largest fluctuations, removal of highly flexible regions, such as the N- and C-terminal 'tails' may help the data analysis. This is because the flexibility of these coils corresponds to thermal noise rather than

functional movement. They may mask smaller, but potentially biologically relevant motions. Therefore, both the N- and C-terminal 'tails' were removed from the analysis, as the RMSF data showed that the flexibility of these regions were several magnitudes larger than the rest of the protein. From initial PCA visualisations, it was also apparent that the C-terminal end of the long alpha helix 2 was very flexible. The bending of the helix 2 dominated the first few eigenvectors and 'hid' other movements. As this helix was highly flexible in all simulations, this region was also removed from data analysis. This left a globular core for further in-depth analysis.

The PCA analysis carried out on the globular core of Hsp90-NTD showed that the largest movement is that of helices 4 and 5, making up the lid region. We used the PCA animations to view these motions. The PCA animations portray the Brownian diffusive motions the protein undergoes during the simulations as symmetrical motions. Viewing these animations, the helices 4 and 5 seem to undergo concerted movements that are 'seesaw' like, *i.e.* periodical fluctuations back and forth. These fluctuations are larger in the apo Hsp90-NTD and in the smaller GVK0153 bound protein, compared to the 17-DMAG:Hsp90-NTD complex. These movements are shown in Figure 3.16, where the different frames of the Hsp90-NTD PCA animation have been superimposed. The PCA data ranks the apo and GVK0153 bound HSP90 as being more flexible than the 17-DMAG bound protein. Although the timescales between the MD simulations and NMR CPMG analysis differ (500 nanoseconds of simulation data compared to micro- to millisecond dynamics), this PCA data is in agreement with the NMR relaxation dispersion results. Here, the 17-DMAG:Hsp90-NTD complex did not show any µs-ms dynamics in the lid region. This is in stark contrast to the apo and the GVK0153 bound protein.
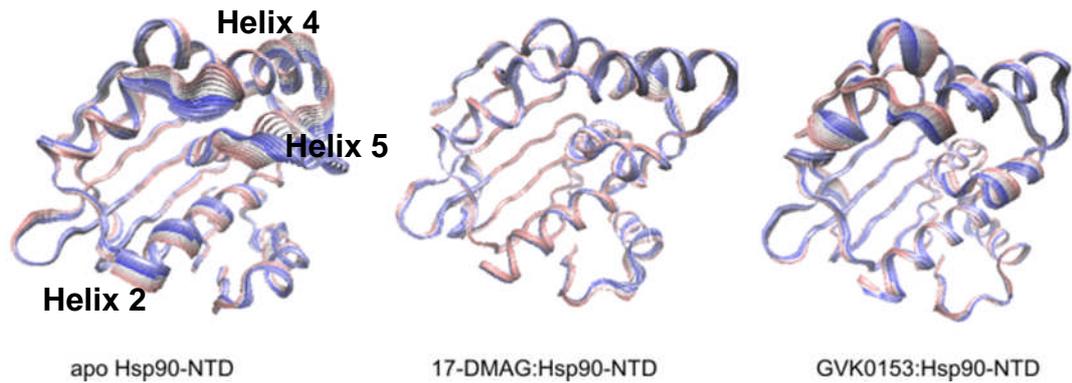
Figure 3.16: Hsp90-NTD structures showing the superimposed frames from PCA animations of the largest eigenvector movements for the apo and ligand bound Hsp90-NTD. The time steps are coloured from blue to red, and wider lines indicate larger movements. It can be seen that the apo protein movements are largest, and 17-DMAG:Hsp90-NTD complex is the least flexible.

3.4.3 Can crystal structures help to explain why the 17-DMAG bound protein is less flexible?

Both NMR and MD simulations suggested that the lid region of the 17-DMAG bound Hsp90 is less flexible, compared to the apo and GVK0153:Hsp90-NTD complex. To try to pinpoint the potential cause for the differences in flexibility, we returned to the crystal structure analysis. The contacts between the two ligands and protein were analysed using LigPlot, as explained in Section 3.1 and shown in Figure 3.2 and Figure 3.3 [116]. The LigPlot analysis shows that the 17-DMAG ligand interacts with the lid residues Gly135, Val136 and Gly137. On the other hand, the smaller GVK0153 ligand, which inserts deeper in the binding pocket does not interact with these or any other lid residues. In line with this observation, the NMR CSP data show that there are also large CSPs for these residues in the 17-DMAG bound Hsp90-NTD compared to the apo protein (Figure 3.11). As the helices around the lid region move as one, it is likely that the contacts between the 17-DMAG and the protein residues 135 to 137 prevent the movement by providing a counter force against the "seesaw" like motion that was seen in the apo and GVK0153 bound protein.

3.5 NMR experiments for characterisation of isolated ligands

As well as characterising the Hsp90-NTD protein dynamics, we had also hoped to analyse the changes in the ligand dynamics upon binding to the protein. This would have allowed a more comprehensive view of what is happening to the configurational entropy of the whole system upon complex formation. To initially examine ligand dynamics, we used NMR. One dimensional hydrogen spectra of the ligands in in aqueous buffer (10 mM KPi, 50 mM KCl, 0.02% NaN_3, 5 mM $MgCl_2$ and 5 mM DTT in D_2O) was recorded.

Our preliminary analysis of ^1H experiments on small molecule ligands suggest that the quality of NMR data is not enough for detailed dynamic analysis. For the larger 17-DMAG spectra, very poor signal to noise ratio was observed (Figure 3.17) because of line-broadening. This was most likely caused by unspecific aggregation of the isolated ligand in water, even though no aggregates were observed in sample tubes. The spectra quality did not improve with use of different temperatures or ligand concentrations (data not shown). Due to the poor ligand only spectra on the isolated 17-DMAG inhibitor, no further NMR experiments on the 17-DMAG ligand were possible. The smaller GVK053 ligand 1D spectra on the other hand looked promising for further studies. However, data from both ligands in isolation and in complex with the Hsp90-NTD would be required for the comparison of the two systems. As the 17-DMAG ligand was not suitable for further experiments, no further studies were conducted on either of the compounds.
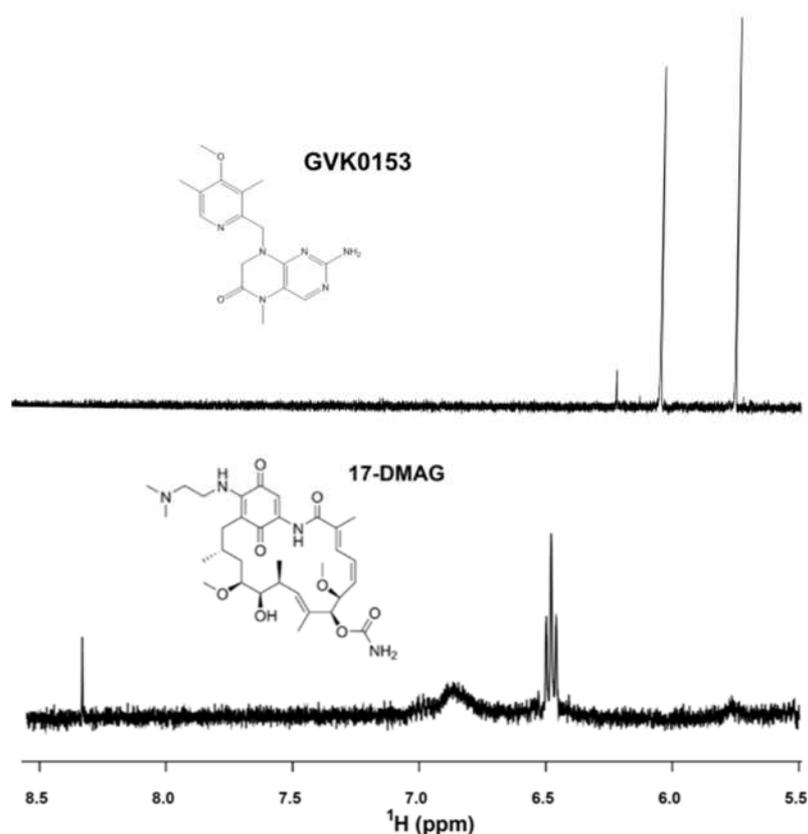
Figure 3.17: Example of NMR 1H spectra of ligands. The smaller GVK0153 compound spectra has sharp peaks as expected for small molecules (above spectra). The larger and more hydrophobic 17-DMAG spectra shown below has broad peaks. Thus this compound is not suitable for further NMR studies on ligand dynamics.

3.6 Comparing the ITC data and results from NMR and MD

Both the NMR and MD data suggest that the binding of the larger 17-DMAG ligand makes the Hsp90 protein less flexible compared to the apo and GVK0153 bound chaperone. The ITC data on the other hand showed that the 17-DMAG ligand binding has large entropic contribution. As entropy gain is normally associated with increased disorder, at a first glimpse, it seems that ITC results do not agree with the atomistic scale investigations.

Thermodynamic values from ITC provide information on the combined effect of changes in the protein, the ligand, as well as the water molecules. This

means that any comparison of the ITC results with the atomistic details of protein flexibility is not comparing 'like with like'. The total entropy change from ITC includes contribution from water, ligand and protein. Furthermore, the global entropy changes arise from rotational and translational movements, as well as internal changes reflected in the configurational entropy term, as shown by equations 16 and 17 below.

$$\Delta S_{total} = \Delta S_{water} + \Delta S_{protein} + \Delta S_{ligand} \qquad \text{(eq. 16)}$$

$$\Delta S_{(water/protein/ligand)} = \Delta S_{rot} + \Delta S_{trans} + \Delta S_{conf} \qquad \text{(eq. 17)}$$

Whilst it is not possible to separate the contributions, or to assign a magnitude that the protein, ligand or water make to the overall entropy change using ITC data alone, we tried to measure the heat capacities of the complexes. The structural information was also used to qualitatively assess the potential causes for the apparent discrepancy between the atomistic and ITC data.

3.6.1 Can heat capacity measurements help to solve solvation?

The heat capacity (C_p) value, can be used to aid the understanding the role of water in the binding reaction. To obtain heat capacities, ITC measurements are performed for a range of temperatures, to obtain enthalpy values as a function of temperature. These values can be plotted to obtain the heat capacity from the slope of the plot. A popular model, derived from protein folding and unfolding studies, links the changes in the heat capacity to changes in the polar and non-polar surface areas [135, 142]. Here, the burial of nonpolar residues during folding leads to a decrease in the ΔC_p value due to the displacement of water with the magnitude of heat capacity change being proportional to the change in the surface area [143]. Although this surface area model for explaining heat capacity data does not always work with the small magnitude changes observed upon ligand binding, perhaps as the model links the changes in thermodynamics exclusively to solvation effects, the method is still in use [135, 144]. Another technique relating the role of water contribution using ΔC_p measurements, is to repeat the experiment in "normal" (H_2O) and heavy water (D_2O). As the deuterium bonds are stronger than the hydrogen bonds, the difference in the heat capacities in otherwise identical systems should be solely linked to the change in the bond strengths.

This difference can be used to work out the contribution of water to the enthalpy change. Few studies have been completed using this method and it has been suggested that the water re-organisation accounts for 25-100% of the enthalpy change in binding reactions, depending on the system [145].

To obtain an experimental value that would allow comparisons with the qualitative estimations of the effect of water to ligand binding to Hsp90-NTD, ITC measurements at different temperatures were carried out to obtain the C_p values for the two ligands. These measurements were done initially in 'normal' H_2O buffer.

The ΔC_p for the GVK0153 binding to NTD was negative (-173 cal K^{-1} mol^{-1}), as can be seen from Figure 3.18, which is typical for small ligand binding reactions. Unfortunately, the change in heat capacity for the larger 17-DMAG could not be determined. This was because the changes in enthalpy values for the lower concentration of the 17-DMAG ligand used for the ΔC_p measurements (compared to the ITC measurements reported earlier) were too small to be accurately measured at all but the highest temperatures. Due to this, no comparisons could be carried out for potential magnitude of solvation changes upon binding of the two different ligands using ITC.
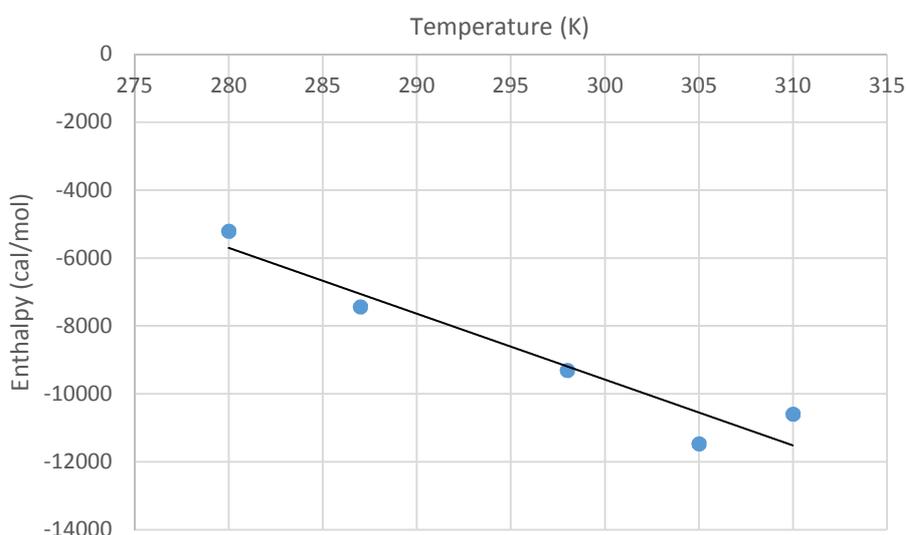
Figure 3.18: Heat capacity change for GVK0153 binding to Hsp90-NTD

3.6.2 Desolvation effect – general theory

As the heat capacity measurements did not manage to produce results to compare the role of water in the binding of the two ligands, we attempted to rationalise this from theoretical point of view.

The effect of water is often thought to be the dominating component in the overall entropy change [146]. This is partly explained by the vastly larger number of water molecules compared to atoms making up the protein and the ligand. In the simulations of ligand bound Hsp90-NTD that were run as a part of this study, the ligand-protein complex accounted for only ~4,400 atoms out of a total 45,000 atoms. However, only the water molecules that change state matter for the water entropy. These are the water molecules that were part of hydration shell surrounding the molecule which get released to bulk water upon binding, or *vice versa* bulk water molecules that becomes more ordered hydration water. Again, given the sheer number of water molecules in a simulation, the state of each individual water molecule is very difficult to measure during the cause of the simulation.

The effect of water on the binding reactions can be thought of in terms of surface areas. Here the size, the shape and polarity matter. The general assumption is that a water molecule in bulk solution has larger entropy compared to the hydration water, which is the water making up the solvation shell around the solute. The hydration water that surrounds the solute also interacts with the solute, and thus have reduced movement. With regards to surface areas, generally speaking, the bigger the area of solute (here a ligand and/or protein), the larger the entropic gain upon binding. This is because the water molecules surrounding the solute are released to the bulk solvent [147]. The nature of the solvent groups also matters. Both NMR and MD studies have demonstrated that water molecules around polar and nonpolar groups have different decay times (*i.e.* how long a water molecule is 'attached' to one place). This suggest that the solvation shell structures differ between the polar and nonpolar groups [148]. This effect can be explained by changes in the hydrogen bond network. Near to hydrophobic surfaces, water molecules form hydrogen bonds with the neighbouring water molecules, rather than with the

solute. This pulls water away from the surface of the solute. This model was justified by a recent molecular dynamics simulation study, which showed that the water molecules around a non-polar solute adopt more structured tetrahedron structures with slightly larger number of hydrogen bonds [149]. This would lead to a decrease in the entropy of water. Polar groups on the other hand participate in the hydrogen bonding with water and this causes water to adopt a different structure. MD studies have suggested that the decay times differ around polar and non-polar groups. The average decay time for a water molecule up to 3.2 Å away from a polar group is 100 ps, compared to 50 ps decay time of water within 4.5 Å of a hydrophobic group. [150]. This is because water around polar groups is bound into favoured positions, with hydrogen bonds between water and the polar surface, resulting in a decrease in entropy.

3.6.3 Desolvation of ligands upon binding

To enable the examination of the effect of desolvation of ligands upon binding, the surface areas of the ligands were analysed. Recall that the size and polarity of the solute surface areas can be related to the changes in the water structure. These changes in the water is one of the contributors to the overall entropy change upon binding.

The 17-DMAG ligand is nearly a third larger than the GVK0153 molecule; its solvent accessible surface area is 845.6 Å^2 compared to 508.5 Å^2 of the GVK0153 ligand. The larger 17-DMAG ligand is also more hydrophobic; the computationally predicted logP ('lipophilicity') and logS ('solubility') values for the 17-DMAG are 1.84 and -4.47, respectively, compared to logP of 1.00 and logS of -3.00 for the GVK0153 inhibitor [151]. The larger surface area and the greater hydrophobicity values suggest that there is a more favourable entropy gain with the larger ligand binding to the Hsp90-NTD due to the release of hydration water to bulk solution compared to smaller GVK0153 ligand. This release of water would contribute to the overall binding entropy, and agree with the entropically driven binding signature of the larger 17-DMAG ligand (more water being released by 17-DMAG desolvation compared to the desolvation of the smaller GVK0153 ligand).

Looking at the polarity of the two ligands, the larger 17-DMAG, whilst being more hydrophobic in general, has seven oxygen atoms that can act as hydrogen bond acceptors, compared to only one in the GVK0153. Hydrogen bonding between the water molecules and the solute, here the ligands, will increase the structure of the water. A simple model for analysing the magnitude of potential entropy change upon moving the ligand from solute to a binding pocket was recently suggested (Richard Henchman, personal communication). The release of a bonded water molecule to bulk solution upon ligand binding is estimated to be around -1.7 J K^{-1} mol^{-1} [152]. Hydrogen bond formation with the solute would have an effect of 5 surrounding water molecules [152, 153] . Based on these numbers, the entropy change, due to release of water, could be calculated using:

$$\Delta S_{water} = -1.7(n_{water} - 5n_{Ha}) \hspace{3cm} \text{(eq. 18)}$$

,where n_{water} is the number of waters in the hydration shell and n_{Ha} the number of H-bond acceptors. Using this method, the rough estimates for ΔS_{water} upon binding of 17-DMAG (*i.e.* due to release of solvent) is -142.8 J K^{-1} mol^{-1} and -108.8 J K^{-1} mol^{-1} for GVK0153 (109 water molecules, 5 acceptors for the 17-DMAG and 69 water molecules, 1 acceptor for GVK0153). These figures equate to ~10.2 kcal/mol ΔS_{water} for 17-DMAG and ~7.8 kcal/mol for the GVK0153 ligand at 25°C. The larger 17-DMAG ligand desolvation has a larger entropically favourable signature, agreeing with the ITC data.

In conclusion, looking at the surface areas and the hydrogen bonding, both effects would result in gain in entropy from release of structured water upon binding. This effect is larger for the 17-DMAG ligand, *i.e.* the desolvation of ligand is enhancing the entropic contribution to binding signature. The increase in entropy from desolvation of ligands is likely to counteract some of the loss of entropy due to 'stiffening' of the protein when in complex with the 17-DMAG, thus helping to rationalise some of the ITC data.

3.6.4 Protein solvation

The change in the Hsp90-NTD shape upon ligand binding, as suggested by NMR chemical shift perturbation analysis, will also have an effect on the surrounding water structure. The surface areas of the Hsp90-NTD crystal structures (apo protein and 17-DMAG and GVK0153 bound Hsp90) were

analysed in Section 3.1. This analysis showed that binding of either ligand increases the protein surface area, which would equate to entropic penalty, as the solvation shell is slightly larger. However, the 17-DMAG bound Hsp90-NTD structure is around 340 $Å^2$ smaller compared to the GVK0153 bound protein. This means that the entropic penalty is smaller for the 17-DMAG binding interaction. To try to quantify this difference, a diameter of a water molecule is approximately 2.75 $Å^2$. The difference between the two ligand bound structures would be maximum of 124 water molecules, assuming the water molecules are tightly packed [154]. Based on the surface areas, the entropic penalty due to increased protein solvent accessible surface area upon ligand binding would be larger for the GVK0153 bound structure, as it has larger surface area compared to the 17-DMAG bound protein. This implies that the 17-DMAG binding to Hsp90-NTD has a smaller entropic penalty compared to the binding of GVK0153. However, the estimations here do not take into account of release of water from the binding pocket of the apo protein upon ligand binding, which would counteract some of the entropic penalty of increased surface areas. 'Cavity water' has been described as being 'more structured' than hydration shell water, thus lead to larger entropic gain upon release [155].

To summarise, the ligand desolvation should cause a favourable entropy effect upon binding and the effect is larger for the 17-DMAG ligand. Similarly, the changes in the protein surface area are smaller with the 17-DMAG complex formation, compared to GVK01053 structure. It is also likely that there will be some favourable contribution to overall entropy from release of the water from the ligand binding pocket. Thus the combined effect of the desolvation is likely to counterbalance the configurational entropic penalty of the 17-DMAG binding that makes the protein less dynamic.

Finally, to get a feeling of the magnitude of the difference the water may play to the overall entropy values from ITC, we performed the following 'back of the envelope' calculation. The difference between the entropic signatures of the GVK0153 and 17-DMAG binding as measured by ITC was ~6,300 cal/mol, which seems quite large. The release of one solvent molecule to bulk solution has been estimated to contribute between 2 to 7 cal mol^{-1} K^{-1} [156]. When the entropic difference between the two ligands is converted to the same per

degree of Kelvin units, the difference of the entropic values between the two ligands is 21 cal mol^{-1} K^{-1} (6,286 cal/mol divided by 298.15 K). This difference would convert to three to ten more water molecules released to bulk solvent by the larger ligand. Additionally, a smaller increase in protein surface area for the 17-DMAG:Hsp90 complex compared to the GVK0153:Hsp90 complex should also contribute into the experimentally observed differences in the entropic contributions between the two ligands. Altogether, the apparent discrepancy in entropic signatures as measured by ITC compared to the changes in protein flexibility does not seem as alarming. These data suggest that for Hsp90-ligand binding, the solvation contributions into the entropy of binding are significantly larger than the contributions from the changes in protein dynamics. We believe that these results will help to develop better small molecule inhibitors for the chaperone as we now have a better understanding of the relationship between protein dynamics and the thermodynamic binding signatures.

3.7 Conclusions

The Hsp90 chaperone is known to be a dynamic protein. We have shown that the dynamics of the chaperone can be altered by small ligand inhibitors, but the effect of water is still the dominating factor in binding of ligands. For the protein dynamics, in particular the lid region dynamics were different when the protein was in complex with the larger 17-DMAG inhibitor, compared to the smaller GVK0153, with the larger inhibitor suppressing the dynamics of the lid region. The differences in the dynamics were seen both by NMR relaxation dispersion techniques and from simulation data, where molecular flexibility was analysed using RMSF and PCA. This is promising, as it can mean that it is possible to use simulations to predict dynamic behaviour of a drug target. This could inform which compound should be taken for further testing and development in the wet lab.

The global thermodynamic parameters, obtained from ITC experiments, however, suggest that for Hsp90, protein dynamics provide only a minor contribution to the binding free energy. On the contrary, effects from water molecules (*i.e.* solvation and desolvation of the ligand and protein upon

binding) overcome the protein dynamic contribution and drive the binding reaction. Indeed, the global thermodynamic parameters did not reflect the decreased dynamics (configurational entropy penalty) of the 17-DMAG bound protein, as the total entropy change as measured by ITC was favourable for the system. Thus, the overall effect was likely to be caused by changes in solvation structure that masked the changes in protein flexibility. This highlights the issue with the use of global entropy values, which include configurational changes as well as rotation and translation of all molecules in the system to validate the changes at the atomistic level in only one constituent of the system (here the protein configurational entropy). The end goal, whether it is to optimise the binding free energy, or to further understand the changes in flexibility of the molecule under study, will dictate the choice of methods.

The Hsp90 chaperone has been a notoriously difficult target for pharmaceutical development due to various toxic effects. This is presumably due to its role in helping the maturation of over a hundred different clients. Our results suggest that manipulating the configurational dynamics of the chaperone by small molecules could help to select only a certain subset of targets. However, the results presented here are preliminary and further studies are needed to see if the changes seen in NTD dynamics do have an effect in the context of full length protein. Further, it should be checked whether the dynamic changes caused by the inhibitors have any effect for co-chaperone recruitment. However, understanding the differences in dynamics caused by the different inhibitors may lead to a more targeted drug design approach.

Chapter 4
Multiple replica MD simulations and configurational entropy

The results presented in the previous chapter showed that the NTD of Hsp90 has different dynamics when it is in complex with different ligands. To understand what the differences in the dynamics mean, in terms of configurational entropies, this chapter focuses on configurational entropy calculations, and exploration of the conformational space, using molecular dynamics simulation data. For this, data from multiple replica simulations were used (50 replicas of 17-DMAG:Hsp90-NTD complex and 25 replicas for the GVK0153:Hsp90-NTD complex, totalling 25 μs and 5 μs of trajectory data respectively).

The configurational entropies can be used to quantify the changes in flexibility. We compute the configurational entropies from the different simulations using the quasi-harmonic approximation (QHA) approach to check on the robustness and suitability of this approach for the NTD of Hsp90 protein. The entropies are calculated using both the more frequently sampled simulation data of apo, 17-DMAG: and GVK0153:Hsp90-NTD systems (that were used in the data analysis in Chapter 3), and the large, not so frequently sampled, replica dataset. This allows us to explore one of the issues with configurational entropy calculations - the more data, the larger the entropy. We also attempt to answer the question "How much simulation data are needed to adequately sample the conformational space?".

4.1 Issues with configurational entropy calculations

Molecular dynamics simulations provide information on the flexibility and structural heterogeneity of macromolecules. When MD simulations are run for long enough to adequately sample the conformational space, the trajectory can be used to find the number of microstates (Ω) a system can adopt with the probability (p_i) of finding it in a particular microstate. This information can be used to estimate entropy [88]. Configurational entropy computations using simulation data have been attempted nearly as long as molecular dynamics

simulations of macromolecules have been run. However, there are still major issues relating to the entropy estimations from simulation data, some of which will be explored in this chapter.

One of the main issues in obtaining configurational entropy from simulations relates to sampling, where finite simulation data is used in an attempt to obtain an entropy for nearly infinite phase space. For this we use replica simulation data to try to answer the question of how much sampling is enough, even if full convergence cannot be achieved.

Secondly, whilst the molecular dynamics force fields have improved over the years, and shown to produce comparable data to experimental NMR results, the methods in use to calculate entropies have not undergone similar development [81, 89]. The two main methods currently in use for calculating configurational entropies from simulations are the quasi-harmonic approach (QHA) and the histogram based method, which were introduced in Chapter 1 [89]. Both of these methods have limitations. The histogram method, where the dihedral bond rotations are normally used to define different states, has been successfully applied to small molecules and more recently also to peptides and proteins [96, 157]. However, this approach tends to have convergence issues with complex systems. Furthermore, the choice of bin sizes, where the bins are used to define what constitutes a different state, rather than fluctuations within a state, remain problematic. Moreover, in the histogram method, traditionally only internal bond rotations are used as co-ordinates. This means that only rotational movements are considered, and any vibrational movements from bond stretching are ignored. As a result, the entropy values from the histogram method can be significantly underestimated [89].

On the other hand, the quasi-harmonic approach (QHA) uses Cartesian coordinates to calculate the fluctuations of the macromolecule in order to obtain the probability distribution required for entropy computations [89]. This approach takes into account both bond rotations and vibrations. However, the QHA method often overestimates entropies, as it is assumed that the different motions are independent. The main assumption behind QHA is that the molecule of interest resides in, and oscillates around, a single large energy

well, which can be described by a Gaussian-like distribution. As a result, the QHA does not work for flexible molecules, which have an energy landscape with multiple energy minima. Nevertheless, this approach has been extensively tested, and shown to work for larger systems such as proteins and DNA [91, 158] [90].

To further examine some of these issues with configurational entropy calculations using simulation data for relatively large dynamic protein systems, we ran multiple replica simulations of the NTD of Hsp90 model system. The overall aim is not to extensively test, and compare, the configurational entropy values obtained by the QHA method with other methods. Instead we want to test its suitability for Hsp90. The effect of simulation time and sampling of conformational space is also explored to see how it affects the convergence of entropy calculations.

4.2 Can the quasi-harmonic approach be used for Hsp90-NTD?

One of the key assumptions for the configurational entropy calculations using the quasi-harmonic approach is that the system resides in a single energy well, where it oscillates around the average structure. Whilst the harmonic oscillator model describes well the movements of more rigid molecules such as DNA, proteins are expected to have a rougher energy landscape containing multiple minima, and any larger structural re-arrangements will break the applicability of the model [159]. To test the suitability of the QHA entropy computation for the Hsp90-NTD system, configurational entropies of the 500 nanosecond long replica trajectories of apo Hsp90-NTD and 17-DMAG or GVK0153 complexed protein were computed. These data are the more frequently sampled set of trajectories, which was used for the analysis of protein flexibility in Chapter 3 (*i.e.* three ~500 ns long trajectories of each apo, 17-DMAG and GVK0153 bound Hsp90-NTD, see also Table 2.3).

As shown in Chapter 3, PCA analysis of the MD data revealed differences in the protein dynamics between the different Hsp90 conformations. Moreover, these computational results are in good agreement with our NMR analysis.

The differences in dynamics can be used to make a ranking of individual Hsp90 states, where the apo state was found to be most dynamic, followed by GVK0153 bound protein and finally, the most rigid 17-DMAG complex. As increased flexibility means higher entropy, the same ranking was expected of the configurational entropy values the different Hsp90 complexes. Next, we calculated the configurational entropies using the same trajectory data (three replica trajectories of each system), as used for the PCA analysis in Chapter 3.

The configurational entropies were calculated using the Schlitter's quasi-harmonic method, as described in Chapter 1, where the mass weighted co-variance matrix of atomic positions is diagonalised [90]. For the analysis, to allow direct comparison of the entropy values for the three systems, the ligands were 'cut out' from the 17-DMAG:Hsp90-NTD and GVK0153:Hsp90-NTD trajectories to match the atom numbers of apo Hsp90-NTD simulations. The entropies for each separate trajectory were computed both for the total 500 ns trajectory length, as well as for smaller windows. This is because the overall entropy has a hidden dependency on trajectory size; the longer the simulation is run, the more conformational space is explored. Consequently, the longer simulation times result in a larger entropy value, due to more microstates being populated. This means that entropy is increasing over the length of the simulation. However, for a smooth energy landscape (*i.e.* where there is only one deep minimum), it is possible to get around this problem by dividing the simulated data into smaller windows, from which the entropies are calculated separately [91]. The entropy values are then plotted as a function of the length of the sampling window to get the rate of increase over simulation time. The resulting curve can be fitted, using the equation below, to estimate the total configurational entropy value as it approaches a stable limit (S_∞):

$$S = S_\infty + \frac{A}{t^n} ,$$ (eq.19)

Here, A and n are fitting parameters and t is the size of the sampling window [91].

The MD data from the apo and ligand bound Hsp90 simulations were processed as described above (Figure 4.1). It can be seen that the entropy of

the apo protein is converging to the largest value (Figure 4.1, green data points). This is followed by the GVK0153 bound protein entropy (Figure 4.1, blue data points). The configurational entropy for the 17-DMAG:Hsp90-NTD complex is significantly less than the other two values (Figure 4.1, red data points). Moreover, the 17-DMAG:Hsp90-NTD complex has significantly smaller entropy values than that of the apo and GVK0153 bound Hsp90. This matches the ranking expected from the NMR relaxation dispersion experiments (Chapter 3.3.3) and from the MD trajectory PCA analysis (Chapter 3.4.2). This suggests that the quasi-harmonic approximation is suitable for entropy estimations for the Hsp90-NTD system.
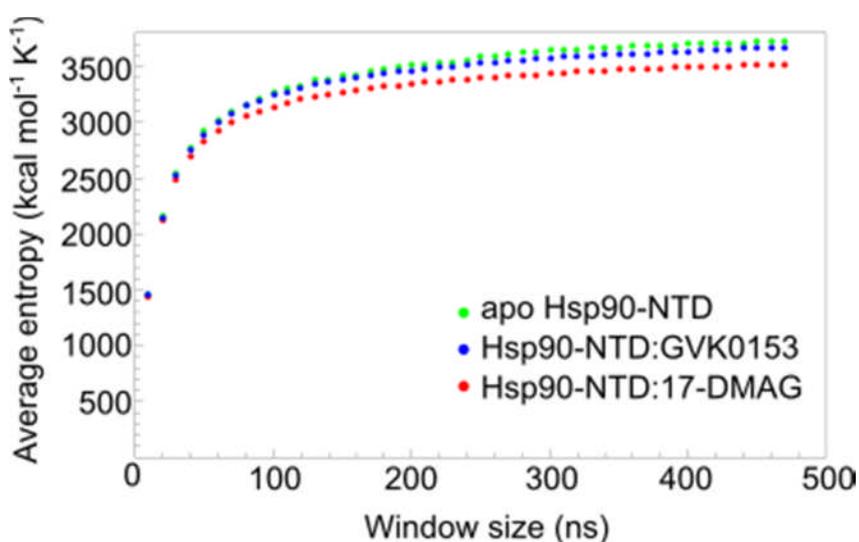
Figure 4.1: Configurational entropy values as calculated by quasi-harmonic analysis over the length of the sampling window, for the apo Hsp90-NTD and the Hsp90-NTD and inhibitor complexes.

Whilst the entropies calculated via quasi-harmonic approximation produce the expected ranking of configurational entropy values, as can be seen from the curves in Figure 4.1, the entropy values have not yet reached a stable limit, but are steadily increasing. Thus, to get a total entropy value, the data were fitted, as described above, using equation 19. Figure 4.2 shows an example of the fit of the 17-DMAG:Hsp90 trajectory data. The configurational entropies values obtained for each system were: 4,165±20 kcal/mol/K for the apo protein, 4,040±45 kcal/mol/K for the GVK0153 bound structure and 3,810±45 kcal/mol/K for the 17-DMAG:Hsp90-NTD complex. The entropies were calculated separately for each of the three replicas of each protein complex, and the errors in the entropy value were calculated as the standard deviation

between the different replicas. The ranking of the data from the graphs, and the estimated configurational entropy values suggest that the apo Hsp90-NTD and GVK0153 bound protein have similar entropies. This result is in agreement with the similar dynamics of the apo and GVK0153 bound Hsp90, as observed by the NMR relaxation dispersion methods. The more rigid 17-DMAG bound protein has a smaller configurational entropy, as expected.

Finally, to check whether the underlying assumption of the smooth energy landscape was true for Hsp90-NTD, the range of the entropy values computed from different windows along the trajectory were examined. For a smooth energy landscape, the QHA entropy should not vary based on the position of the sampling window. Instead, the data points should be interspersed with only a minor spread between the values calculated from different windows. For example, the entropy should not vary much between the first quarter and the last quarter of the simulation data. Figure 4.3 shows the configurational entropies, calculated using different position of the sampling window, for the apo protein. It can be seen that the entropy values are very similar, both within the same simulation (Figure 4.3, same colour points), and between different repeats (Figure 4.3, different colour points), as the points corresponding to different sets of data overlap. The maximum difference in the data points for ten nanosecond window size was 37.5 kcal/mol for the apo protein. This corresponds to 1% variation in the entropy value. The spread of values for the GVK0153 and 17-DMAG simulations were very similar to the apo Hsp90 entropies, 1% and 4% respectively. This relatively small spread of entropy values, calculated over the different windows of the whole trajectory, supports the assumption that the protein oscillates within one energy well, and does not undergo large conformational re-arrangements, at least during the simulation time (500 ns).
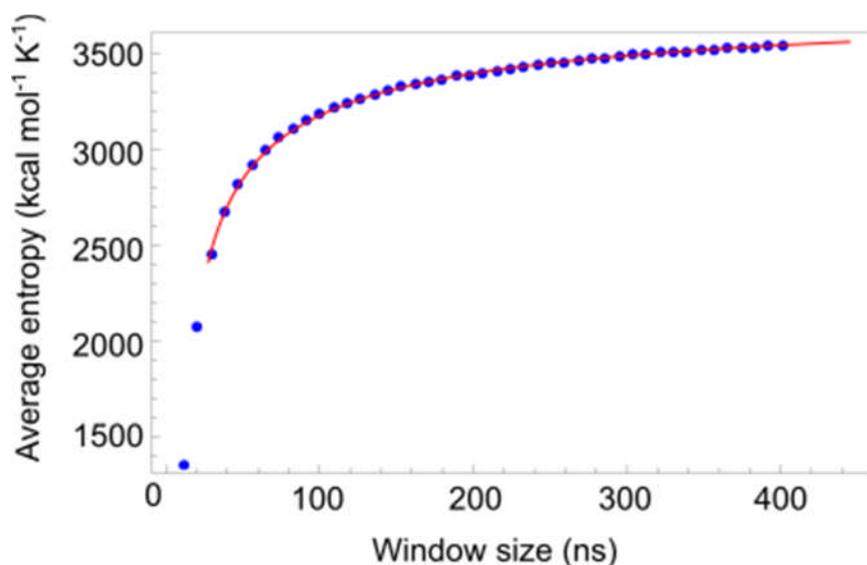
Figure 4.2: A fit of the average Schlitter entropy (red line) against window size (blue data points) for 17-DMAG:Hsp90-NTD.
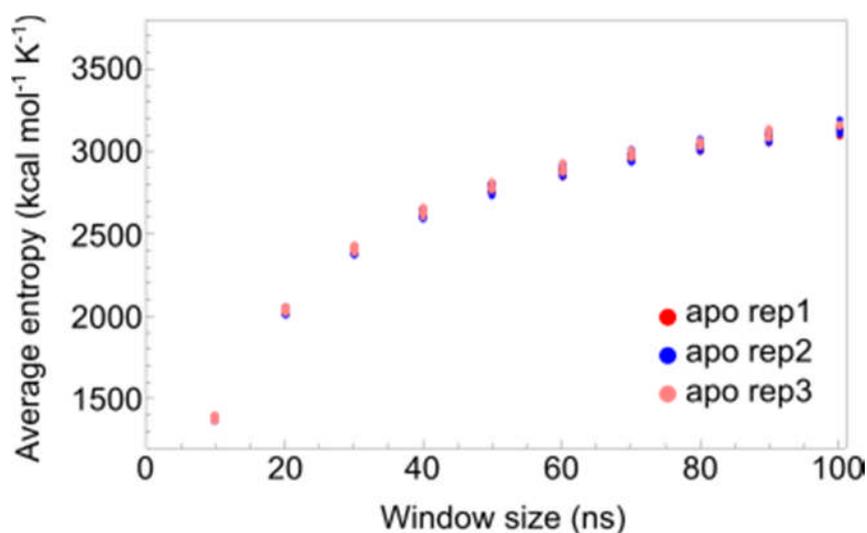
Figure 4.3: The spread of entropy values for the apo Hsp90-NTD simulations, as calculated for the short trajectory window sizes.

4.3 The size of conformational space

Hsp90-NTD configurational entropies calculated from 500 ns MD trajectories agreed surprisingly well with the ranking predicted from NMR experiments that characterise protein flexibility on μs-ms timescale. However, the entropy values were still increasing over the 500 nanosecond simulations, suggesting that the sampling of conformational space is not complete.

Adequate sampling of protein conformational space is one of the main issues of computing the configurational entropy from MD simulations. Indeed, the configurational entropy is a measure of the volume of the conformational space that the molecule can visit at a given temperature. The size of conformational space can be given as 3N-6 dimensions, where the N is the number of particles in the system [160]. To sample sufficient conformational space for entropy values to converge, the Hsp90-NTD system should sample at least 9660 dimensions. Each dimension has an undetermined number of configurations the system can adopt. This estimate of the number of dimensions is based on the assumption that each dimension is a perfect smooth well, which is unlikely to be true given the general roughness the energy landscape. Given the complexity of macromolecular landscape, even with the recent advances in computer hardware and software, that have enabled micro- and even millisecond simulation times to be reached, it is probable that simulation data still presents only partial exploration of the available space [31]. Indeed, a study conducted a few years ago suggested that molecular dynamics simulations may never reach an equilibrium, as the entropy from different replica simulations seemed to be increasing even after 70% of the conformational space was explored [96]. Similar conclusions have also been reached by other groups [161]. Nevertheless, whilst simulations may never produce a complete energy landscape, it is clear that the more space visited, the better the entropy estimations. Thus we set out to explore how much simulation data is required to sample the Hsp90 energy landscape.

4.3.1 Conformational space sampling by replica simulations

One of the issues with simulations is the time taken to collect sufficient data. It is generally thought that data from multiple replica simulations can represent the conformational landscape better than one long simulation, which can be trapped in a local energy minimum for long periods of time [162]. Data collection is also faster when simulations are run in replicate, as these can be run in parallel. Whilst accelerated simulation methods, such conformational flooding and meta-dynamics, have gained popularity over the last few years, the question remains as to how accurate the energy landscape produced is. This is due to the biases that are introduced to enhance the sampling [163]. The accuracy of sampling the relevant states that the system visits is vital for

calculations of thermodynamic parameters. At the same time, the 'standard' force fields have recently been shown to re-produce the structures and fluctuations of several molecules to the same accuracy as NMR data, as well as reproducing transient states in protein folding pathways [164] [165]. Thus, we ran fifty 500 nanosecond long replica simulations of the 17-DMAG:Hsp90-NTD system using 'standard' MD methods, in an attempt to answer the question "How many simulations should be run to get an adequate coverage of the conformational space?". Collecting this replica dataset took just under two calendar months using the supercomputing resources in Leeds University. In comparison, obtaining a single 25 µs long trajectory would take 250 days, assuming that there was no time spent queuing for the supercomputing resources.

The replica simulation data were converted to eigenvectors and eigenvalues using principal component analysis. This conversion of the data from Cartesian coordinates to eigenvector space considerably reduces the complexity of the data, as each vector component can be thought to represent one dimension in phase space. To further simplify the analysis of the volume of space explored by different simulations, the projections of the two largest eigenvectors (principal components 1 and 2, *i.e.* PC1 and PC2), which made up 25% of the total dynamics, were chosen to represent the conformational space. The projections of PC1 and PC2 produce a two dimensional volume to which standard geometrical analysis can be applied. It also makes the data easy to visualise. In contrast, to achieve eighty percent coverage of the movements seen during the simulations, the analysis would need to include the fifty largest eigenvectors.

To start the analysis into the conformational space exploration by the different replica simulations, PC1 *versus* PC2 volumes were plotted for each individual simulation. These plots were used to check that the different replicas indeed did explore different parts of the conformational space. The data for the individual PC1 *vs* PC2 plots were further broken into four separate time windows, that were plotted using different colours, to view the path each simulation processed along. A sample of the resulting graphs are shown in Figure 4.4. From the plots it can be seen that the different replicas do explore

different areas of the conformational space, as both the shape and the volume of the areas explored differ. Some replicas produced similar shaped areas, for example 'V-shaped' graphs were quite common, as shown in the third row in Figure 4.4. However, the path taken to produce the shape differs, as can be seen from the coloration of the frames making up the trajectories. Only a few trajectories produce a continuous area, the majority seemed to suddenly 'jump' from one conformation to another.
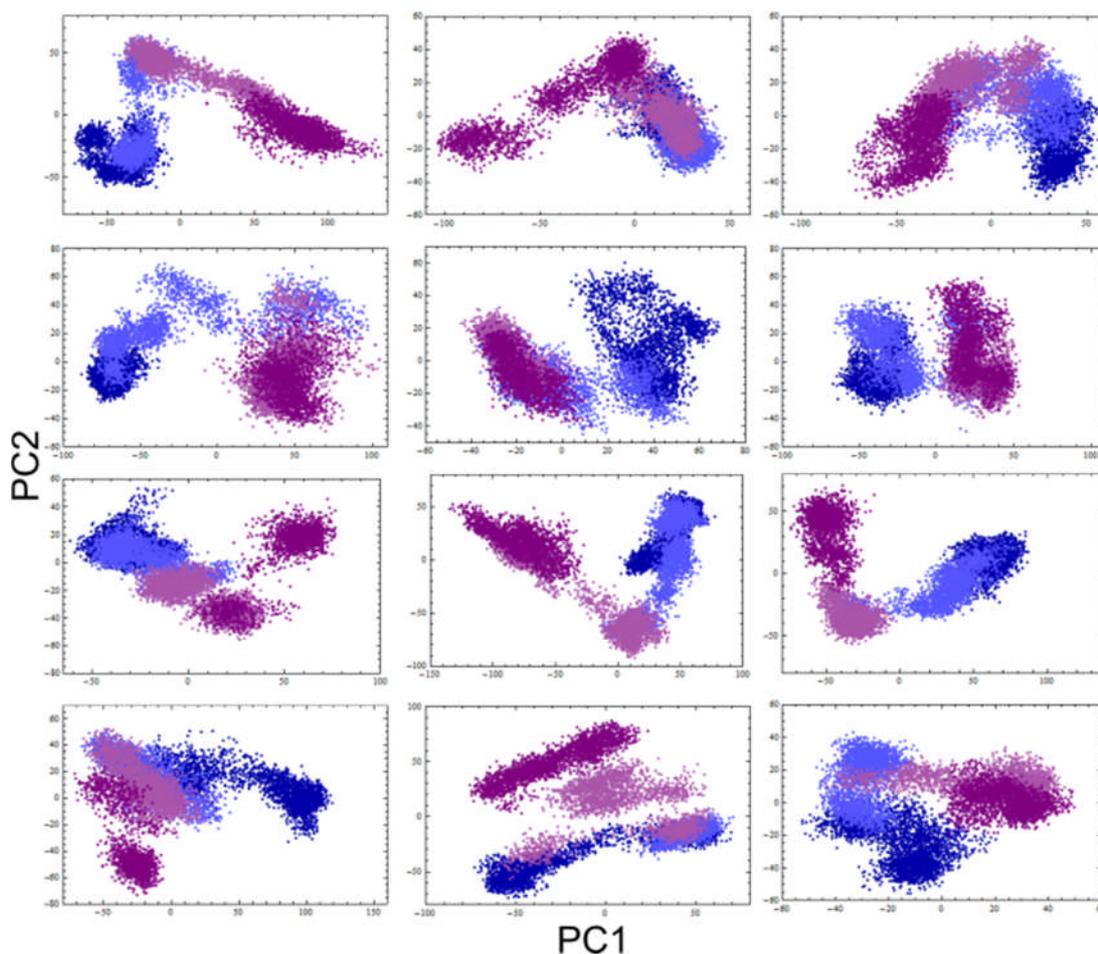
Figure 4.4: Volume plots of the first two eigenvector projections of the replica simulations. The different colours correspond to different hundred nanosecond windows of the trajectory. This shows the path (from dark blue, light blue, dark purple to light purple) the molecule explores the conformational space.

Whilst it was clear that each replica simulation visited slightly different areas of conformational space, it is not possible to tell the size of the overall space the simulations collectively explored. Although the largest movements making up the PC1 and PC2 occurred in similar areas of the protein, as seen in Figure

4.5, the dot products used to compare the similarities of the largest eigenvectors demonstrated that the eigenvectors were not identical between the replicas, even when the 'shape' of the area was similar. This is probably caused by the complex nature of the underlying energy landscape and 'softness' of proteins. Helix movements for example are a combination of many smaller twists that can take different directions. To get around this issue, the trajectories from the different replica simulations were joined together to form one long trajectory, for which the PCA analysis was completed. This ensured the largest eigenvectors describe the same movements.
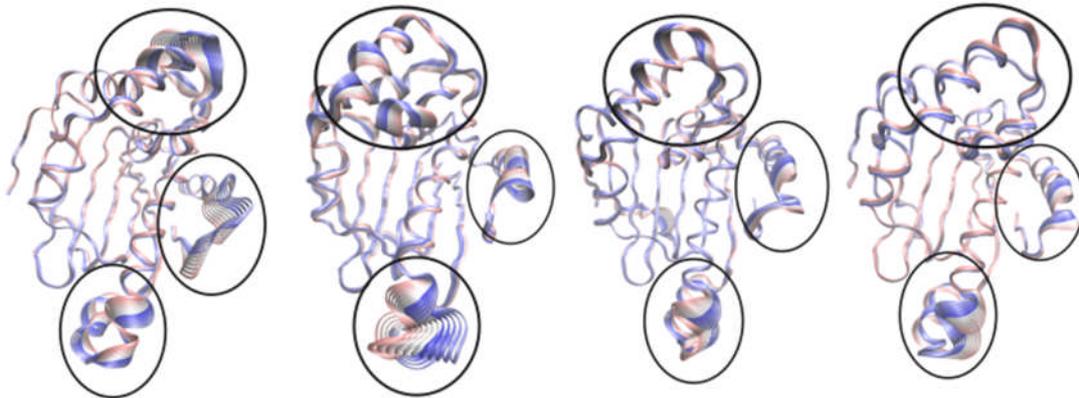
Figure 4.5: The animations of the largest eigenvectors for different replicas. Whilst there are movements observed in the same three areas as indicated by the circles, the magnitude and direction of the movement differs.

The projections of the two largest eigenvectors of the concatenated trajectory data produced a continuous 2D volume in the shape of an ellipse, as shown in Figure 4.6 (top panel on the left). The ellipse shaped volume clearly differs from the discontinuous areas the individual simulations explored, as shown in Figure 4.4. This confirms that, collectively, the fifty replicas making up a twenty five microseconds trajectory had explored a much larger part of the conformational space. For the analysis here, it was also assumed that the combined replica dataset had explored the total conformational space available (at least for the two largest dimensions examined), as no new space was visited after 22 microseconds.

Although replica simulations can be run in parallel, routinely running fifty replica simulations is computationally costly, at least from pharmaceutical industry point of view. The key point is to make sure enough conformational space is explored. Therefore, we wanted to estimate how many replicas need to be run in order to achieve coverage of the majority of the conformational space. For this, the dataset was divided into smaller parts (half, quarter, eight, sixteenth) and these partial datasets were superimposed on top of the total dataset, to see how much the explored space was reduced. From the plots in Figure 4.6, it can be seen that half of the data (25 replicas/12.5 µs) covers almost the same area as 50 replicas do (95%). Similarly, a quarter of the data (12.5 replicas/6.25 µs) also explores 86% of the conformational space. On the other hand, an eight of the data (6.25 replicas), whilst covering two thirds (78%) of the total space, has some discontinuity in the area covered. Finally, three replicas cover only 63% of the total space, and one 500 nanosecond long trajectory explores a fifth (22%) of the area. This analysis suggests that at least six replica simulations should be run to get 75% coverage of the conformational space, at least for the Hsp90-NTD system under study.

Full data set (50 replicas / 25 µs)	Half of data (25 replicas / 12.5 µs)	Quarter of data (12.5 replicas / 6.25 µs)
Eight of data (6.25 replicas / 3.1 µs)	Sixteenth of data (3.125 replicas / 1.5 µs)	1 replica (500 ns)
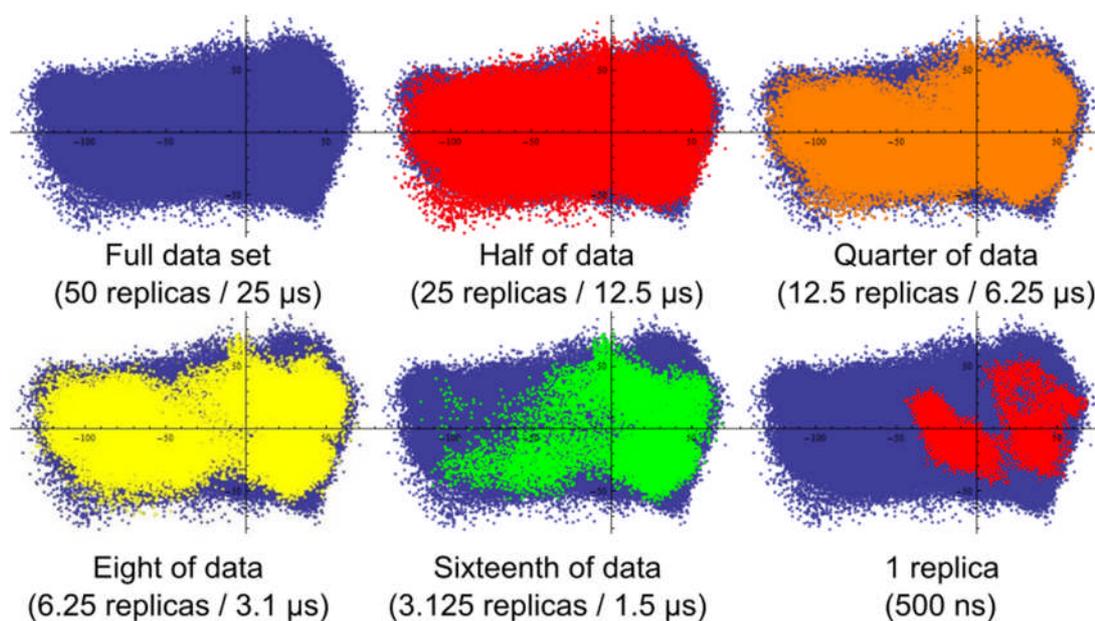
Figure 4.6: The PC1 and PC2 of the combined trajectory data were plotted to produce an area plot showing the space explored. The data was then halved, again and again, to try to estimate how many simulations would be enough to explore a similar area as the full dataset of fifty replicas.

4.3.2 What about the probability distributions?

The analysis of the conformational space exploration by replica simulations suggested that six replica simulations provide a reasonable coverage (>75%) of the total space. However, to get good data for entropy computations, it is not enough just to explore as much of conformational space as possible. The sampling of the explored space is also vital to get an accurate ensemble distribution. A heatmap was produced showing the probability distribution of the different areas of the conformational space visited by the simulations. The heatmap is shown in Figure 4.7, where white areas correspond to the most frequently sampled regions, followed by yellow, orange and purple, with the darker areas only explored very rarely. Comparing the heatmap of conformational space in Figure 4.7 to the total area plot shown in Figure 4.6, it is very clear that not all areas are sampled extensively. There is a clear tendency for the system to be found in a few 'hot spots'. These 'hot spots' were visited around 2800 times during the simulation, compared to only a few visits to the outer edges of the volume.

With the question still in mind; "How much simulation data are needed to explore 'enough' conformational space?", the combined replica dataset was divided into smaller parts. These were used to produce heatmaps of the conformational space explored, to see if the amount of data affected the probability distribution. As can be seen from the bottom panel of Figure 4.7, 25 replicas produce a similar probability distribution to the total 50 replica dataset. Some perturbations from the distribution are already seen when only a quarter of the data are used. This corresponding to around 6 microseconds of simulation data (12.5 replicas from the total dataset). These perturbations become larger as the dataset is reduced further, as seen on the lower panel of Figure 4.7. The heatmap analysis thus suggests that more than 12 replicas should be run to get a similar statistical distributions of the microstates to the full 50 replica dataset.
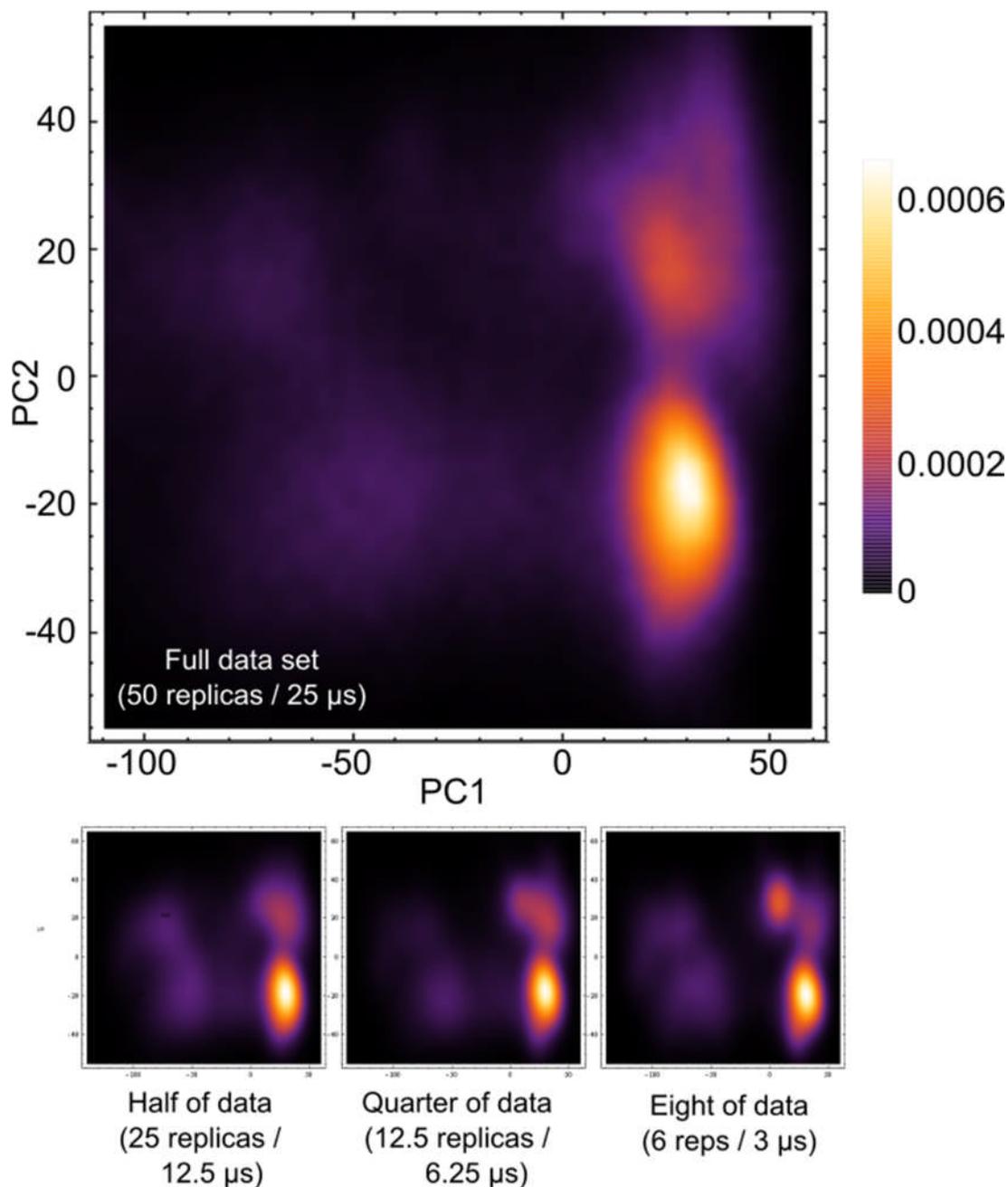
Figure 4.7: Probability distributions of the sampling of the conformational space by the 17-DMAG:Hsp90 replica trajectories. The large plot on the top panel of the figure shows the distribution for the full dataset of 25 µs. When half of the data (25 replicas/12.5 µs) is used to obtain the distribution, the shape of the data looks nearly identical to the full dataset. However, the distribution starts to skew when only a quarter or less of the data is used.

4.3.3 Is the sampling of conformational space system dependent?

To look into how the molecular constraints effect the size and sampling of conformational space, we also ran replica simulations of the GVK0153 bound protein. As the GVK0153:Hsp90-NTD system was shown to have different dynamics, and in particular larger entropy, compared to the 17-DMAG bound protein analysed in the previous section, comparing these two systems would allows us to examine the effect flexibility has on the sampling of conformational space. For this, we ran 25 replicas of 200 nanosecond long simulations of the Hsp90-NTD:GVK0153 complex.

The GVK0153 bound Hsp90-NTD replica dataset was concatenated with the 17-DMAG:Hsp90-NTD replica trajectories of the same amount (5 µs). The concatenated dataset was used to generate a combined set of eigenvectors to repeat the analysis described in Sections 4.3.1 and 4.3.2.

The resulting volume of space plots are shown in Figure 4.8. The area that the systems collectively visit is shown in purple; and the area visited by the GVK0153 bound protein correspond to the blue volume, and the red volume is for the 17-DMAG:Hsp90-NTD space. From the plots it can be seen that the more flexible Hsp90-NTD:GVK0153 system explores a much greater percentage of the combined space (91%) compared to the 17-DMAG bound protein, which visits only half of the combined space (55%).

The heatmaps shown in Figure 4.9 also demonstrate the differences in the space exploration by the two systems, where the GVK0153 bound protein clearly explores two different areas of space frequently, compared to the one clear 'hot spot' seen by the 17-DMAG bound Hsp90.

The overall sampling of the space differs between GVK0153 bound Hsp90-NTD compared to the protein in complex with 17-DMAG, as suggested by both the entropy calculations and the principal component analysis. Both the volume plots and heatmaps also reflect the more flexible nature of the GVK0153 bound structure. This means that PC (using heatmaps and/or volume plots) analysis could be used in conjunction with the entropy calculations to explore differences in the dynamics.

PC1 vs PC2
from combined 17-DMAG:
and GVK0153:Hsp90-NTD
trajectories

Blue = PC1 vs PC2
from GVK0153:Hsp90-NTD
trajectories

Red = PC1 vs PC2
from 17-DMAG:Hsp90-NTD
trajectories
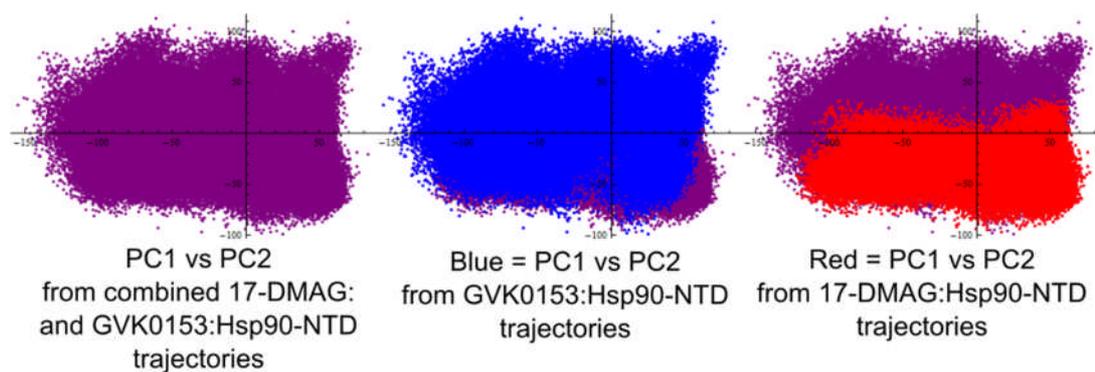
Figure 4.8: Two dimensional volume plots of the PC1 and PC2 projections of the combined 17-DMAG and GVK0153 datasets are shown in purple. The blue volume shows the area GVK0153 bound protein explored, and the red volume that of 17-DMAG:Hsp90 complex.
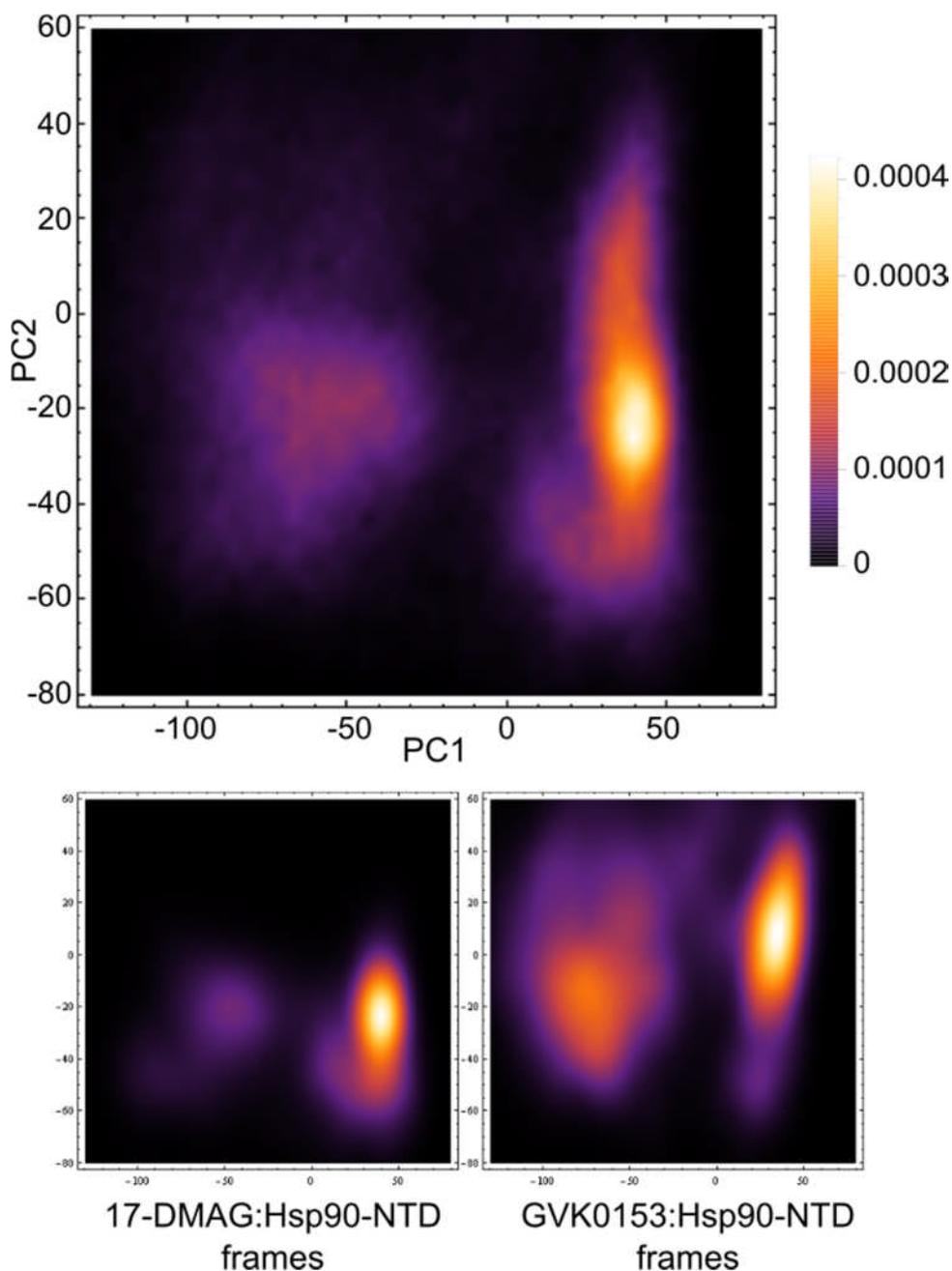
Figure 4.9: Probability distributions of the sampling of the conformational space by the concatenated 17-DMAG:Hsp90 and GVK0153:Hsp90 trajectories.

4.3.4 Convergence of configurational entropy

As configurational entropy has a hidden dependency on the length of the simulation dataset, the entropies were also calculated using the larger datasets. These values were compared to those computed in Section 4.2 from the 500 nanosecond more frequently sampled triplicate trajectories. The difference here was that for the more frequently sampled trajectories, we

calculated the entropy separately for each replica, in comparison for the larger dataset, where concatenated trajectories were used. For the larger dataset, the entropies were calculated from seven microseconds of data, as any larger dataset caused the software to run out of memory.

The values for configurational entropy obtained from the 500 nanosecond long trajectories were 4,040±45 kcal/mol/K for the GVK0153 bound structure and 3,810±45 kcal/mol/K for the 17-DMAG one (see section 4.2). In comparison, the configurational entropies were significantly higher when seven microseconds of data were used for the calculation and data fitting; 4,935 kcal/mol/K and 4,465 kcal/mol/K respectively (Figure 4.10). This difference in the entropy values is likely to be caused by differences in fitting with more data points available. The spread of the entropy values calculated for a 5 nanosecond window is very similar for the shorter 500 nanosecond and concatenated 7 microsecond datasets. The difference between the largest and smallest value calculated for the 5 nanosecond window was 130 kcal/mol/K with standard error of 15.3 kcal/mol/K for the concatenated dataset, compared to 168 kcal/mol/K with standard error of 6.7 kcal/mol/K for the 500 nanosecond long more frequently sampled repeats. Despite the difference in the configurational entropy values from the short and long trajectories, the ranking of entropy values matches, and both datasets show the differences in the flexibility between the two ligand bound proteins. The take home message is that whilst the actual entropy values do not seem to converge, the ranking of the configurational entropies does. Therefore, the entropy data should still be useful from the drug design point of view.
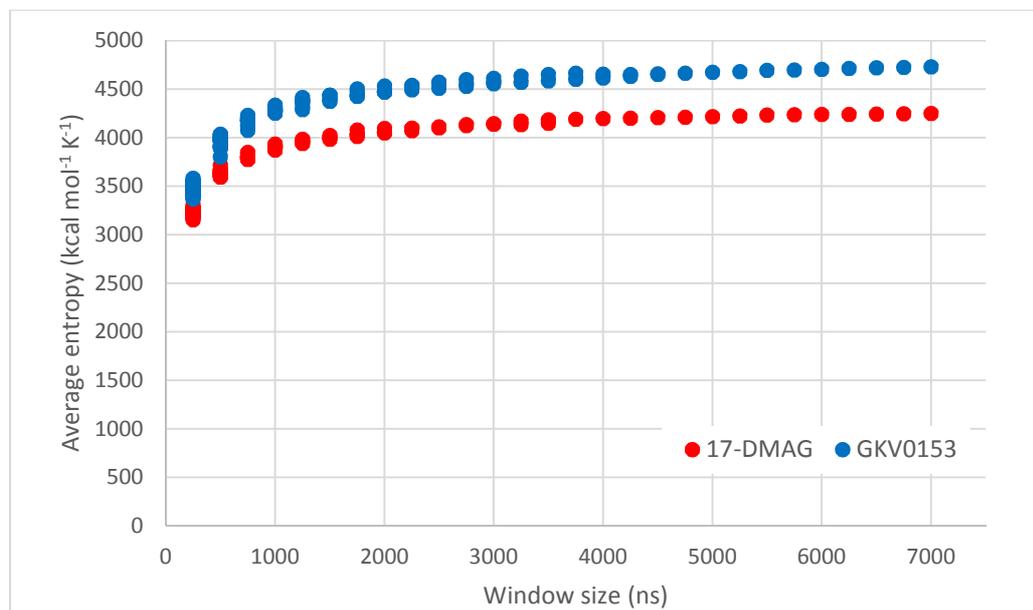
Figure 4.10: Configurational entropy values calculated by quasi-harmonic approach for the larger (5 µs) concatenated datasets.

4.4 'Naïve entropy' using the histogram based method

Since we have already calculated the probability distribution of states from the 2D energy landscape (in Section 4.3.2), we used these data to calculate the 'naïve' entropy using the histogram method. We call this 'naïve' because the entropy calculated is not the total configurational entropy of the molecule; it only takes the two largest movements of the system into account. One of the issues with the histogram method is the selection of a suitable window width for the bins. The above data allows to study this effect on overall configurational entropy.

For the calculations, the two dimensional area obtained from the two largest principal component projections of the 17-DMAG:Hsp90-NTD replica dataset was divided into smaller bins. Here, a bin corresponds to a microstate. The bins were then used to calculate the entropy using the 'standard' Boltzmann formula, where entropy is related to a weighted sum of the logarithms of each microstate:

$$S = -\kappa_B \sum_{states} p_i \ln p_i \qquad\qquad\qquad\qquad \text{(eq. 20)}$$

Here the κ_B is the Boltzmann constant and p_i is the probability of finding the system in microstate i [88]. The calculations were repeated using different bin widths, and an example of the differing bin sizes is shown in Figure 4.11.

The number of bins used to calculate probability distribution ranged from ten to ten thousand. The entropy values for the different bins are shown in Figure 4.12. The very coarse grain approach of using only 10 bins resulted in an entropy value three times smaller compared to when 10,000 bins were used. Figure 4.12 clearly shows that entropy value is increasing when smaller sampling windows are used. This increase is more pronounced when a very coarse sampling strategy is adopted. Whenever a decreasing window size is used (total bin number of 5,000 or greater), the entropy values start to tend towards a stable limit. This demonstrates the importance of selecting a sufficiently small sampling window. In principle, how many bins are used to calculate entropy, should not affect the comparison of differences in entropy between complexes. However, the selection of the bins is crucial if the absolute entropy is calculate as this value is dependent on the number of possible states (i.*e.* bins). For protein configurational entropy calculations, window widths ranging from 1 to 5 degree rotations around di-hedral angles have been suggested [89]. The issue remains on how to choose a suitably bin size; too large a bin results in coarse data that skews the entropy value, too small a bin is computationally expensive and may be biologically irrelevant.
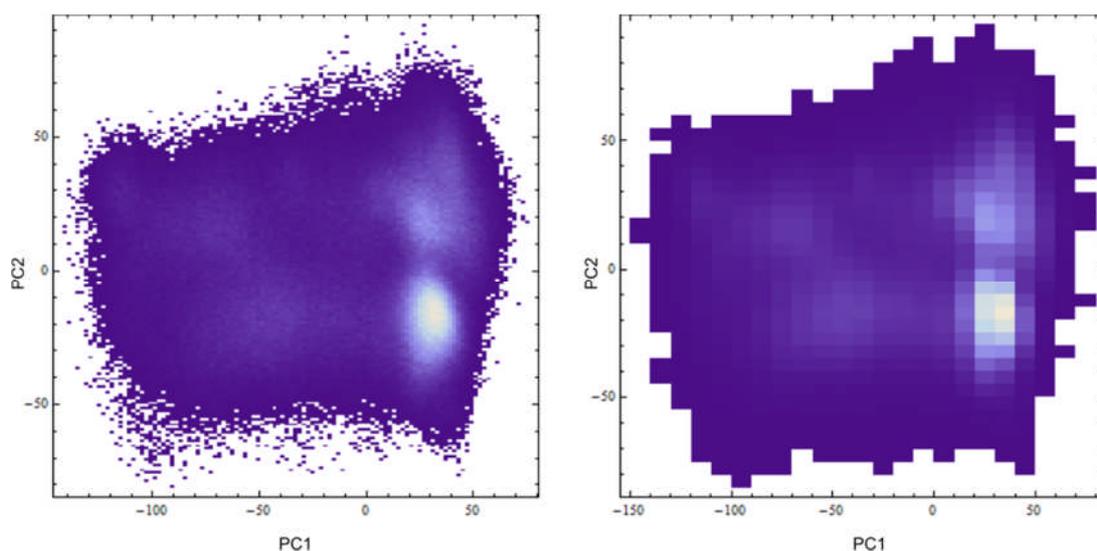
Figure 4.11: Illustration of the effect of bin sizes on conformational space sampling for the 17-DMAG:Hsp90-NTD dataset. The left hand image presents data where each PC divided into 100 bins is and the image on the right is divided into 25 bins.
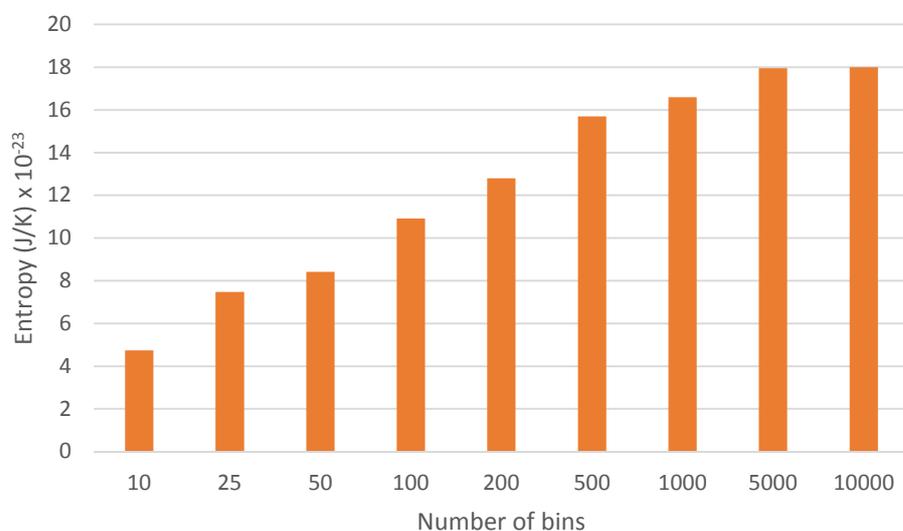
Figure 4.12: The configurational entropy values calculated by the histogram methods vary over three-fold for the same system when the bin size is increased from 10 to 10,000. This illustrating how critical it is to select a small enough sampling interval. It can be seen that as the number of bins is increasing, the entropy value starts to reach a stable limit.

4.5 Configurational entropy of ligands – un-harmonicity at play

Often the choice of the method used to calculate configurational entropies depends on the system. This final section illustrates this using simulation data for the two small ligand Hsp90 inhibitors, 17-DMAG and GVK0153. The ligand entropies were first calculated using the quasi-harmonic approach, as was done for the Hsp90-NTD protein in Section 4.2. The histogram method using dihedral angle calculations were also attempted for the smaller GVK0153 ligand. The ligand only simulation data were analysed by two undergraduate Physics students, Peter Adkins and Max Holme, for their final year project under my supervision, and it is these results that are presented here.

4.5.1 Ligand entropies calculated with the quasi-harmonic approach

MD trajectories were collected for one hundred nanosecond long simulations of the two inhibitors 17-DMAG and GVK0153. Given the smaller size of the molecules compared to the protein (44 atoms for GVK0153 and 92 for 17-DMAG versus ~3500 for Hsp90-NTD), this trajectory length was thought to be reasonable to obtain adequate sampling of conformational space.

The QHA was then used to calculate the configurational entropy as a function of sampling window for both molecules. The resulting entropy plots are shown in Figure 4.13. From the graphs it can be seen that the larger 17-DMAG molecule entropy seems to have stabilized over the 100 nanosecond simulations, as the entropy values have begun to reach a stable plateau over the larger time frames. The entropy calculations for the smaller GVK0153 molecule did not produce the expected logarithmic curve approaching a stable limit (right panel in Figure 4.13). Instead, the entropy seems to increase and decrease randomly as the simulation progresses.
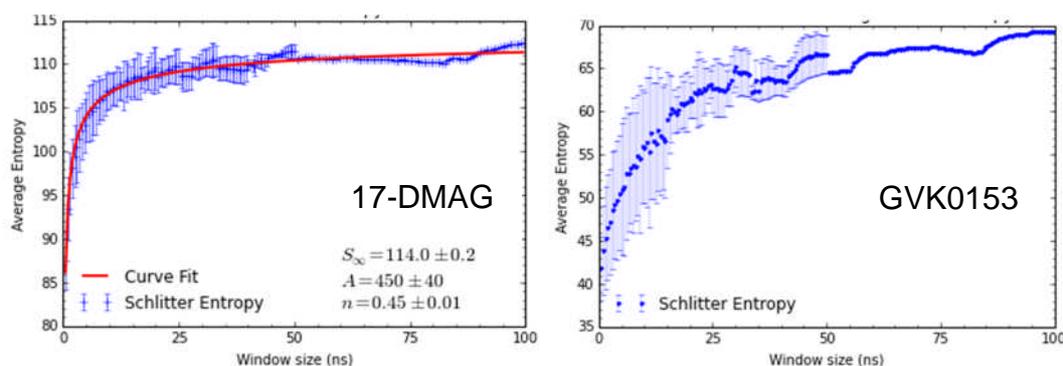
Figure 4.13: The configurational entropy of the 17-DMAG and GVK0153 ligands as calculated using the quasi-harmonic approach, and plotted against the sampling window length. The configurational entropy could be estimated by fitting the data as described for the larger 17-DMAG ligand, shown on the left. The data for the small GVK0153 molecule did not produce the expected logarithmic curve, thus the configurational entropy could not be estimated. (Graphs prepared by Peter Adkins)

For the 17-DMAG entropy, the spread of the entropy values over the smaller time windows, which are shown by blue error bars, has stabilised after fifty nanoseconds. For the GVK0153 on the other hand, the spread of points that present the average entropies along the same amount of sampling is very large as can be seen from the error bars. This suggests that the GVK0153 molecule is more flexible and the underlying energy landscape cannot be described as a single energy well.

The 17-DMAG molecule has a smooth enough energy landscape for the entropy to be calculated using the quasi harmonic approach. Therefore the data were fitted using $S = S_\infty + \frac{A}{t^n}$, as was done for the protein data in Section 4.2. The entropy for free 17-DMAG ligand was estimated to be 114 kcal/mol. In comparison, the configurational entropy of the protein bound 17-DMAG ligand was calculated to be 92 kcal/mol, suggesting a small penalty in configurational entropy upon binding.

To get an idea of the accuracy of the bound versus unbound 17-DMAG entropy results, a literature search of molecular dynamics and NMR studies was conducted on the 17-DMAG ligand. The 17-DMAG ligand is a derivative of geldanamycin and they share an identical 19-membered ansamycin ring (see Figure 2.1 in Chapter 2). This macroring has 14 single, and thus

rotatable, bonds. As the two structures share the same 'backbone', the published geldanamycin data was used as a model for 17-DMAG. The solid state structure of geldanamycin has been solved, and early crystallography work on the protein bound molecule suggested that the 'free' and 'bound' forms are remarkably different. The crystallography work suggested that the 'free' form was more open compared to the tight C-shape structure of the bound ligand (see Figure 4.14). The geldanamycin thus required *trans-cis* isomerisation upon binding to the Hsp90. Subsequently, the difference in the free and bound forms has been calculated using MMPBSA methods, which suggested that there is a -3.4 cal/mol/K to -5.4 cal/mol/K penalty upon binding of geldanamycin to Hsp90 [115]. However, a few years later, the solution structure of the inhibitor was solved. Here, the authors found that there are twelve different conformations of geldanamycin in the equilibrium population, each nearly equally likely to be occur. Hsp90 would simply selectively bind to a pre-populated state of geldanamycin. This is in contrast to the earlier studies suggesting that the inhibitor had to go through *trans-cis* isomerisation [166]. Studies on 17-DMAG also supported the view that the ansamycin ring does not undergo isomerisation upon binding [136]. Given the results from the literature, the difference we found in the configurational entropy between the free and bound 17-DMAG forms seems to be of a reasonable magnitude.

Free　　　　　　　　Bound
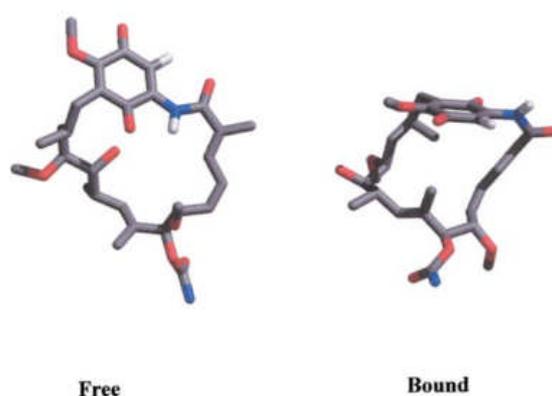
Figure 4.14: The two different conformations of geldanamycin as predicted from structural and simulation data (Figure from [115]).

Given the atypical data for the GVK0153 molecule, both trajectories were further analysed using root mean square deviation (RMSD) of the simulated structures relative to the starting structure. The RMSD data over time is shown in Figure 4.15. From the graphs it can be seen that the 17-DMAG molecule fluctuates around an average position. GVK0153 seems to undergo larger structural changes, as the graph of RMSD over time shows large 'jumps'. This RMSD plot of the GVK0153, showing large structural changes, explains the atypical configurational entropy graph, where the entropy was not only increasing over the duration of simulation, but also decreasing with large errors between the window sizes. The main assumption of the quasi-harmonic approximation for entropy calculation is that the molecule resides in a single large energy well, with small fluctuation around the average structure. From the RMSD data, the GVK0153 molecule clearly visits different states in multiple energy wells, thus does not satisfy the QHA assumption.
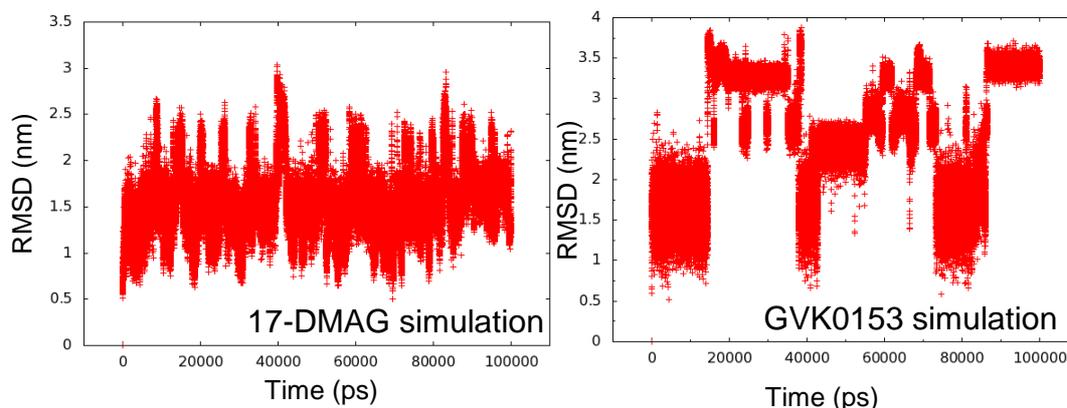
Figure 4.15: RMSD graphs of the simulations of the two ligands. The 17-DMAG ligand on the left fluctuations around similar average conformation. Larger conformational changes are seen for the GVK0153 ligand on the right. (Graphs prepared by Peter Adkins)

4.5.2 Dihedral angle analysis suggests three separate energy wells for GVK0153

To investigate the changes in the GVK0153 structure suggested by the QHA and RMSD data analysis, it was decided to analyse the data further by using

the dihedral angle fluctuations. This approach is similar to that used in entropy estimations by the histogram based method.

As can be seen from Figure 4.16, the GVK0153 molecule is made up of a purine scaffold that is connected to an aryl moiety by a CH_2 linker [167]. Both ring structures of the GVK0153 are rigid due to the double bonds, but the CH_2 linker as well as amide and methyl groups, highlighted in Figure 4.16, contain rotatable bonds. These rotatable bonds were selected for the dihedral angle fluctuation analysis.

The dihedral angle fluctuations over the simulation time were calculated using the Amber ptraj tool [80]. The bin sizes, that define what contributes a different conformation, was set to 10 degrees.

Based on the molecule structure, the methyl and the amide group dihedral angle fluctuations were expected to exhibit small fluctuations around an average position, rather than larger conformational re-arrangements. For the methyl group attached to the purine scaffold of GVK0153, the dihedral angle analysis showed three separate states around 60, 180 and 300 degrees, as seen in Figure 4.16. The size of the each peak was roughly the same, suggesting that the methyl group adopted three equal rotational states. A methyl group is made up of three equal hydrogens connected to a carbon. Given the bond geometry, these are expected to produce three interchangeable states. The three different states that were observed matched the expected distribution. The amide group dihedral angle analysis also produced the expected distribution of a single state, based on the chemical connectivity (data not shown).

In contrast to methyl and amide groups attached to the two different rings, any changes in the dihedral angle of the CH_2 linking the two rings, would result in large conformational changes in the GVK0153 molecule. From the dihedral angle analysis we observed three maxima centred around 65, 160 and 300 degrees. The probability distribution for the maxima was roughly 50%, 5% and 45% respectively. The least likely conformation is that of the two rings being stacked, as seen in the Figure 4.16. These different conformations of the molecule cause the underlying energy landscape to have multiple energy minima, rather than the single harmonic one, as is assumed by the QHA. This

explains why the configurational entropy calculations by the quasi harmonic approach did not converge.
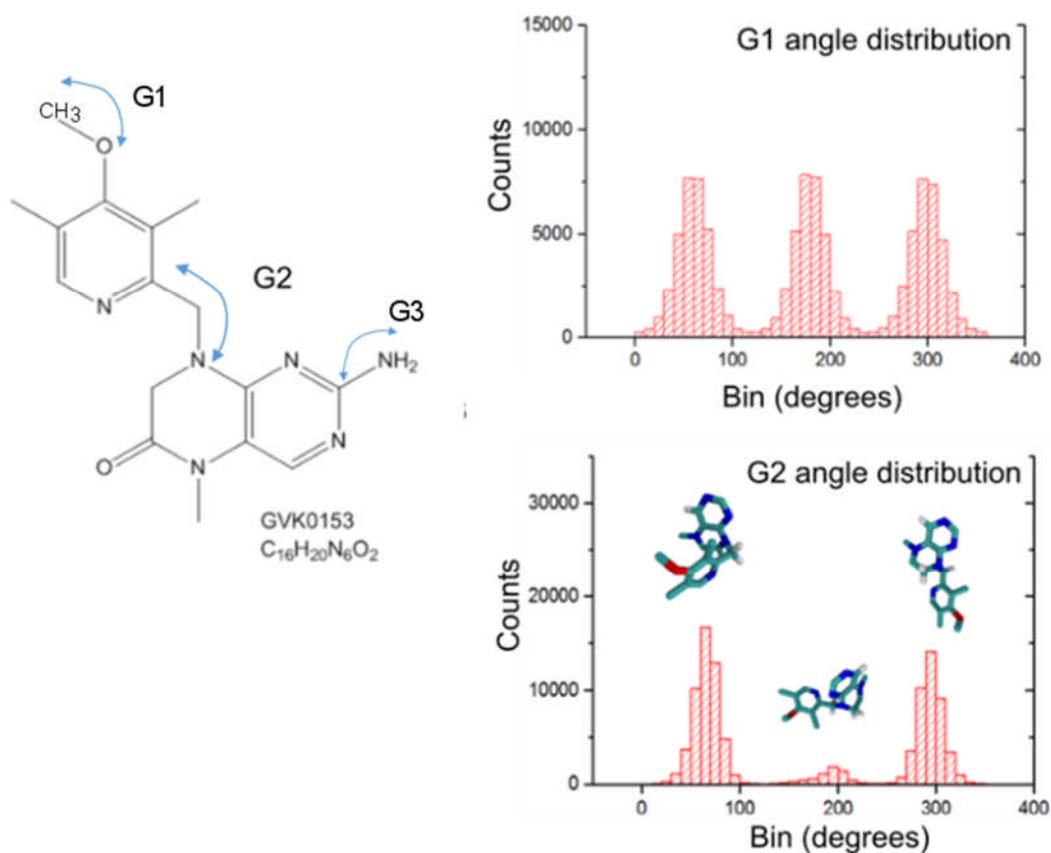
Figure 4.16: To look further into the structural fluctuations of the GVK0153 molecule, three different dihedral angles were selected for analysis as indicated on the left by the G1, G2 and G3 arrows. The right panel shows the probability distribution of the G1 and the G3 dihedrals. The resulting changes in the molecule conformations for the G3 dihedral is shown in the bottom figure. (The histogram distribution were calculated and plotted by Max Holmes)

4.6 Conclusions

In this chapter we explored the use of the quasi-harmonic approach method to compute configurational entropies of Hsp90-NTD in complex with inhibitors. We also explored the amount of simulation data required to get an accurate description of the underlying conformational space.

The results presented here show that the quasi-harmonic approach can be adopted with relative ease for both smaller and more complex systems, as long as the system does not undergo large conformational re-arrangements. The quasi-harmonic approximation provided the expected ranking of the configurational entropy values (S_{conf}) for the different Hsp90 systems. However, the calculated values may not reflect the exact entropy; when more simulation data were used for calculations, the entropy value increased. Despite this, the calculations can be used to produce 'entropy ranking'. This reflects the dynamics of the system, and this information can still be of use for drug design, if effects of molecule flexibility are of interest. Moreover, obtaining a value (here S_{conf}) to compare flexibilities of different systems will provide a more objective measure, rather than comparison of flexibilities of certain regions of the protein of interest, as in Chapter 3 using RMSF analysis and looking at PCA fluctuations in Chapter 3.

We also wanted to attempt to answer the question - how many simulations are needed to provide a "good coverage" of the conformational space (75% or greater)?. Here the results of the two largest eigenvector movements that were used to describe the conformational space suggested that at least 6 replica simulations that collectively provide several microseconds of data are needed. When the coverage of the underlying energy landscape is of interest, these results highlight the potential issue with the common practice of running three replica simulations. We have shown that data from three replica simulations cover under two-thirds the total conformational space. However, in Chapter 3, we found that the differences in the flexibility of the different Hsp90 complexes could be seen from analysis of only three replica simulations. The configurational entropies also provided the expected ranking when only 3 simulations were used for calculations, but the value was considerably larger when more replica data were used for the computations. The larger dataset is likely to provide a more accurate configurational entropy value and the question of how many replicas should be run will depend on what the data is used for; ranking of complexes or trying to obtain an accurate value for configurational entropy.

Finally, the method of choice for configurational entropy calculations does depend on the flexibility of the system under study. QHA cannot be used for

very flexible molecules, which occupy multiple energy minima. There are issues with the dihedral method relating to selection of the bin sizes and the complexity of the system it can describe, due to collective motions. This means that it may not be easy to compare the entropies of different systems, especially if they have been calculated by different methods. Thus it is clear that the currently used methods for entropy estimations still do not tell the whole story. It remains to be seen whether simply correction terms could be added to the existing methods to solve these issues, or whether a completely new way of looking into quantifying the flexibility of soft biomolecules is needed.

Chapter 5
Overall summary and conclusions

Introducing a new drug to market is a lengthy and expensive process, typically it takes 10-15 years and costs $1.7 billion [168]. The main problem is that only around 3% of the new drug projects manage to produce a successful drug molecule [169]. Whilst the majority of failures occur in the later stages of the process, during clinical trials, there is still room for improvement in the initial drug discovery stages. An overall success rate of ~35% is typical during the drug discovery process, problems leading to failure may arise due to issues with understanding the underlying biology, lack of lead molecules, poor potency or selectivity to name but a few [169]. Given this very high failure rate, are the methods used by the pharmaceutical industry to find new drugs suitable for the task? The current process of initial high throughput screening, followed by structural studies focuses on finding compounds that bind. These steps are followed by lead optimisation, where efforts are made to modify any promising lead compounds sometimes with help from structural data to further improve the binding affinity. Binding affinity, although providing a tool with which to rank the potential drug compounds, is a global value that includes changes in the systems flexibility, structure as well as changes in solvation. As proteins are dynamic rather than static structures, and dynamics plays an important role in binding affinity (in terms of the entropy), a better understanding of dynamics, and how they could be manipulated, could be one way of improving the process. The details of any changes in flexibility of the system cannot be readily accessed from the global affinity value, or the structural data.

In this study we used the Hsp90 chaperone, which is also a cancer medication target, to study the effects small molecules have on protein dynamics. We demonstrated that the structure and binding affinities, do not tell the whole story of what is happening to the protein upon complex formation. The X-ray crystal structures of 17-DMAG and GVK0153 bound Hsp90-NTD are nearly identical, and both ligands bind the protein with similar nanomolar affinity. However, a more in depth study into the protein dynamics showed that the larger 17-DMAG ligand suppressed the lid-region dynamics; this did not happen in the apo and the GVK0153 ligand bound structures. These

differences were seen both with NMR relaxation dispersion measurements and from analysis of MD simulation data.

From this study we have learned that dynamics also play a role in protein and ligand interactions. Further, the changes in dynamics may play a biologically relevant role. Hsp90 is known to be a highly dynamic protein with a large number of client proteins in whose maturation Hsp90 plays a role [39]. The current view of the Hsp90 chaperoning action is that not all cellular Hsp90 are the same, but the diverse client interactions are determined and regulated by different co-chaperones [39]. The Hsp90 co-chaperome is complex. Some of the co-chaperones are client specific (for example the Cdc37 is a kinase specific co-chaperone), whereas others are not [32, 170]. The co-chaperones play diverse roles ranging from client recruitment, regulating the Hsp90 ATPase activity or seem to 'trap' the Hsp90 complex in a certain conformation, to name but a few. The diverse co-chaperones also interact with different parts of Hsp90, some binding to the NTD, others to the CTD, and others contacting all three domains [32]. It has also been shown that different co-chaperone complexes are required for the client maturation. For example, the glucocorticoid receptor maturation. only requires Hop as well as the chaperones Hsp70 and Hsp90 for activity. However, a different steroid receptor, the progesterone receptor also requires the p23 and Hsp40 co-chaperones, in addition to Hsp70, Hsp90 and Hop for maturation [171, 172]. On top of the co-complex co-chaperone network, it has also been shown that different Hsp90 inhibitors interact with different cellular pools of Hsp90. For instance, geldanamycin and a purine analogue, will interact with different cellular pools of Hsp90 [47, 173]. H/D exchange mass spectrometry studies demonstrated that the binding of small molecule inhibitors and Cdc37 co-chaperone to the Hsp90 N-terminal domain, change the orientations of the N-, middle and C-terminal domains of the full length Hsp90, compared to the apo protein [52]. Given the large number of diverse Hsp90 clients, and the fact that there is no single co-chaperone, or client protein interactions site, it is likely that the conformational dynamics do play a role in chaperone's function. Indeed, it has been shown that some clients, such as HSF1, interact with only a single Hsp90 closed conformation, that is only present for a short time within the Hsp90 open-closed cycle. Other clients, such as HIF1α, interact with

multiple conformations [48]. The natural ligands, ADP and ATP, also change the N-terminal dynamics of Hsp90, suggesting that dynamics is likely to play a functional role in the chaperone activity [50]. Further understanding of the role that dynamics play in regulating Hsp90 activity is thus likely to be important, both for the underlying biology, and with regards to how it can be manipulated by small molecules. Here we have shown that 17-DMAG binding can lock the Hsp90-NTD into a more rigid conformation. This could favour interaction with a subset of client proteins, without affecting the interaction of others. All together this provides specificity. Potentially, this allows one to design an 'ideal' drug that only prevents the folding of 'bad' cancer related proteins, without affecting Hsp90 functions in normal cells.

Whilst our results suggest that manipulating the conformational dynamics of the Hsp90 chaperone by small molecules could help to select only a certain subset of targets, further studies are needed. As it is known that some clients interact only with a certain conformation of the Hsp90 using open and closed form mutants, follow up studies are needed to understand what effects the differing lid-region dynamics, that we found, play at cellular level [48]. The differences in dynamics may play a role in client recruitment, or effect the co-chaperone composition. Also, it would be interesting to study the effects of the two inhibitors have in the dynamics of the full length Hsp90. This would be particularly important as the changes in the N-terminal structure have been shown to have an effect on the NTD and middle domain interface [39]. Thus the full length chaperone may adopt different conformations when in complex by the two different ligands.

We used both atomistic level techniques and global thermodynamic measurements to study the dynamics of Hsp90. In contrast to the changes in protein flexibility seen by NMR and MD methods, the global thermodynamic parameters from ITC suggest that for Hsp90, protein dynamics provide only a minor contribution to the binding free energy. The effects from water molecules (*i.e.* solvation and desolvation of the ligand and protein upon binding) overcome the protein dynamics contribution, and drive the binding reaction. Although some studies have found good correlation between the ITC data and changes in protein flexibility, as measured by NMR, the effect of water has also been demonstrated to be the major driving force in the binding

of other drug molecules [50, 146]. Our results thus suggests that whilst understanding the underlying dynamics may be important to better manipulate the different cellular pools of Hsp90, further understanding of the role that water plays is also needed to optimise solvation energy. The end goal, whether it is to optimise the binding free energy, or to further understand the changes in flexibility of the molecule under study, will dictate the choice of methods.

Whilst MD data did reproduce the differences in the protein dynamics seen by NMR methods, the results obtained here were from a multi-disciplinary approach. The global thermodynamic parameters with differing entropic binding signatures suggested that there may be differences in the underlying dynamics. However, it turned out that the dynamic changes are masked by the effect of water in the binding signature. NMR studies on the protein-inhibitor complexes highlighted differences in the micro- to millisecond dynamics in the Hsp90-protein lid-area. This helped to pinpoint areas to focus on in the MD data analysis. Interestingly, the differences in the lid-region dynamics could be seen by PCA analysis, using 500 nanosecond long trajectories. Finally, the differences in the lid-area dynamics could be explained by analysing the ligand-protein contacts using X-ray crystal structures. It is likely that a similar multi-disciplinary approach is needed, if the study of differences in dynamics is to be incorporated into the drug discovery process. The simulations can be used to predict the dynamic behaviour of drug target. Crystal structures can then be used to pinpoint any residues or regions which may drive, or stabilise, the dynamic regions. These results could be used to decide, which compound should be taken for further testing and development in the wet lab. However, one should remember that optimising the target binding affinity is only one parameter that needs to be considered in the drug discovery process. Other parameters that need to be considered include oral bioavailability, the ADME/DMPK values (absorption, distribution, metabolism, and excretion/drug metabolism and pharmacokinetics), toxicology issues, patient demographics to name but a few [174]. Interestingly, it has been noted that optimisation of the enthalpic component of the drug-target binding by later generation statins and HIV-1 protease inhibitors lead to better pharmacokinetic and drug resistance profile

[174]. This suggests that it may be beneficial to keep in mind multiple parameters during each step optimisation to get a better end product.

Finally, we showed that calculations of the configurational entropies using MD data can be used to quantify the differences in the dynamics. This provides a 'value', which can be used in ranking molecules; this is similar to the currently used ranking based on binding affinity. However, it is difficult to understand the underlying biology from numbers alone. Viewing the main motions during the simulation is a far superior way to see what is happening, and to understand the underlying biology. This would be a similar method to the current use of visualising X-ray structures of the targets, for example to identify additional areas where the drug molecules could be extended in fragment based drug design in order to improve affinity. To investigate these protein dynamics better, it is more efficient to run multiple replica simulations. For replica simulations, based on our data, the commonly used three replica simulations explore just over half of the underlying energy landscape. We recommend that six or more replicas are run as this provides a much better description. However, how much simulation data is needed will ultimately depend on the question the simulation is seeking to answer, and of course, of the system under study. The recent advances in computer technology have speeded up the simulations considerably, The GPU technology has brought down the potential setup costs for 'faster' simulations considerably. With the continued improvements in the computing, it may be possible to run very large systems routinely, which may give a better picture of what is happening in the cell, rather than the reductionist approach that is currently used. Here, rather than simulating a part of the Hsp90-chaperone, a simulation of the full length protein, with the co-chaperone complex, may be possible in the future. This would allow us to see in 'real time' what effect manipulating the lid-region dynamics may have on the client protein selection.

List of References

1. Frauenfelder, H., S.G. Sligar, and P.G. Wolynes, *The energy landscapes and motions of proteins.* Science, 1991. **254**(5038): p. 1598-1603.
2. Zhuravleva, A. and L.M. Gierasch, *Substrate-binding domain conformational dynamics mediate Hsp70 allostery.* Proceedings of the National Academy of Sciences, 2015. **112**(22): p. E2865-E2873.
3. Eisenmesser, E.Z., et al., *Intrinsic dynamics of an enzyme underlies catalysis.* Nature, 2005. **438**(7064): p. 117-121.
4. Xu, Y., et al., *Induced-fit or preexisting equilibrium dynamics? Lessons from protein crystallography and MD simulations on acetylcholinesterase and implications for structure-based drug design.* Protein Science, 2008. **17**(4): p. 601-605.
5. Boehr, D.D., R. Nussinov, and P.E. Wright, *The role of dynamic conformational ensembles in biomolecular recognition.* Nature chemical biology, 2009. **5**(11): p. 789-796.
6. Chodera, J.D. and D.L. Mobley, *Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design.* Annual review of biophysics, 2013. **42**: p. 121.
7. Du, X., et al., *Insights into protein–ligand interactions: Mechanisms, models, and methods.* International journal of molecular sciences, 2016. **17**(2): p. 144.
8. Bronowska, A.K., *Thermodynamics of ligand-protein interactions: Implications for molecular design.* 2011: INTECH Open Access Publisher.
9. Gilson, M.K. and H.-X. Zhou, *Calculation of protein-ligand binding affinities.* Annual review of biophysics and biomolecular structure, 2007. **36**(1): p. 21.
10. Zhou, H.-X. and M.K. Gilson, *Theory of free energy and entropy in noncovalent binding.* Chemical reviews, 2009. **109**(9): p. 4092-4107.
11. Martin, T. and Z.S. Derewenda, *The name is bond—H bond.* Nature Structural & Molecular Biology, 1999. **6**(5): p. 403-406.
12. Andrews, P., D. Craik, and J. Martin, *Functional group contributions to drug-receptor interactions.* Journal of medicinal chemistry, 1984. **27**(12): p. 1648-1657.
13. Williams, D.H., et al., *Toward an estimation of binding constants in aqueous solution: studies of associations of vancomycin group antibiotics.* Proceedings of the National Academy of Sciences, 1993. **90**(4): p. 1172-1178.
14. Chia-en, A.C., W. Chen, and M.K. Gilson, *Ligand configurational entropy and protein binding.* Proceedings of the National Academy of Sciences, 2007. **104**(5): p. 1534-1539.
15. Searle, M.S. and D.H. Williams, *The cost of conformational order: entropy changes in molecular associations.* Journal of the American Chemical Society, 1992. **114**(27): p. 10690-10697.
16. Williams, D.H., et al., *Understanding noncovalent interactions: Ligand binding energy and catalytic efficiency from ligand-induced reductions*

in motion within receptors and enzymes. Angewandte Chemie International Edition, 2004. **43**(48): p. 6596-6616.

17. Freire, E., *A thermodynamic approach to the affinity optimization of drug candidates.* Chemical biology & drug design, 2009. **74**(5): p. 468-472.

18. Baum, B., et al., *More than a simple lipophilic contact: a detailed thermodynamic analysis of nonbasic residues in the S1 pocket of thrombin.* Journal of molecular biology, 2009. **390**(1): p. 56-69.

19. Klebe, G., *Applying thermodynamic profiling in lead finding and optimization.* Nature reviews. Drug discovery, 2015. **14**(2): p. 95.

20. Hammes, G.G., Y.-C. Chang, and T.G. Oas, *Conformational selection or induced fit: a flux description of reaction mechanism.* Proceedings of the National Academy of Sciences, 2009. **106**(33): p. 13737-13741.

21. Garbett, N.C. and J.B. Chaires, *Thermodynamic studies for drug design and screening.* Expert opinion on drug discovery, 2012. **7**(4): p. 299-314.

22. Cooper, A., et al., *Heat does not come in different colours: entropy–enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions.* Biophysical chemistry, 2001. **93**(2): p. 215-230.

23. Myszka, D.G., *Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors.* Current opinion in biotechnology, 1997. **8**(1): p. 50-57.

24. Engen, J.R., *Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS.* Analytical chemistry, 2009. **81**(19): p. 7870-7875.

25. Zaccai, G., *How soft is a protein? A protein dynamics force constant measured by neutron scattering.* Science, 2000. **288**(5471): p. 1604-1607.

26. Rambo, R.P. and J.A. Tainer, *Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering.* Current opinion in structural biology, 2010. **20**(1): p. 128-137.

27. Lippincott-Schwartz, J., E. Snapp, and A. Kenworthy, *Studying protein dynamics in living cells.* Nature Reviews Molecular Cell Biology, 2001. **2**(6): p. 444-456.

28. Gao, H., et al., *Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset.* Journal of structural biology, 2004. **147**(3): p. 283-290.

29. Frauenfelder, H., G.A. Petsko, and D. Tsernoglou, *Temperature-dependent X-ray diffraction as a probe of protein structural dynamics.* 1979.

30. Kleckner, I.R. and M.P. Foster, *An introduction to NMR-based approaches for measuring protein dynamics.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2011. **1814**(8): p. 942-968.

31. Shaw, D.E., et al., *Atomic-level characterization of the structural dynamics of proteins.* Science, 2010. **330**(6002): p. 341-346.

32. Röhl, A., J. Rohrberg, and J. Buchner, *The chaperone Hsp90: changing partners for demanding clients.* Trends in biochemical sciences, 2013. **38**(5): p. 253-262.

33. Taipale, M., et al., *Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition.* Cell, 2012. **150**(5): p. 987-1001.

34. Echeverría, P.C., et al., *An interaction network predicted from public data as a discovery tool: application to the Hsp90 molecular chaperone machine.* PloS one, 2011. **6**(10): p. e26044.

35. Whitesell, L., et al., *Inhibition of heat shock protein HSP90-pp60v-src heteroprotein complex formation by benzoquinone ansamycins: essential role for stress proteins in oncogenic transformation.* Proceedings of the National Academy of Sciences, 1994. **91**(18): p. 8324-8328.

36. Sidera, K. and E. Patsavoudi, *HSP90 inhibitors: current development and potential in cancer therapy.* Recent patents on anti-cancer drug discovery, 2014. **9**(1): p. 1-20.

37. Petrikaite, V. and D. Matulis, *Inhibitor Binding to Hsp90: A Review of Thermodynamic, Kinetic, Enzymatic, and Cellular Assays.* Current Protein and Peptide Science, 2014. **15**(3): p. 256-282.

38. Porter, J.R., C.C. Fritz, and K.M. Depew, *Discovery and development of Hsp90 inhibitors: a promising pathway for cancer therapy.* Current opinion in chemical biology, 2010. **14**(3): p. 412-420.

39. Krukenberg, K., Street, T., Lavery, L. and Agard, D., *Conformational dynamics of the molecular chaperone Hsp90.* Quarterly Reviews of Biophysics, 2011. **44**(2): p. 229-255.

40. Ali, M.M., et al., *Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex.* Nature, 2006. **440**(7087): p. 1013-1017.

41. Shiau, A.K., et al., *Structural analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements.* Cell, 2006. **127**(2): p. 329-340.

42. Li, J.a.B., J., *Structure, function and regulation of the Hsp90 machinery.* Biomed J, 2014. **36**: p. 106-117.

43. Krukenberg, K.A., et al., *Grp94, the endoplasmic reticulum Hsp90, has a similar solution conformation to cytosolic Hsp90 in the absence of nucleotide.* Protein Science, 2009. **18**(9): p. 1815-1827.

44. Hessling, M., Richter, K., Buchner, J., *Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90.* Nature Structural Molecular Biology, 2009. **16**: p. 287-293.

45. Li, J., J. Soroka, and J. Buchner, *The Hsp90 chaperone machinery: conformational dynamics and regulation by co-chaperones.* Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2012. **1823**(3): p. 624-635.

46. Shrestha, L., Patel, H. and Hiosis, G., *Chemical tools to investigate mechanisms associated wiht Hsp90 and Hsp70 in disease.* Cell Chemical Biology, 2016. **23**: p. 158-172.

47. Moulick, K., et al., *Affinity-based proteomics reveal cancer-specific networks coordinated by Hsp90.* Nature chemical biology, 2011. **7**(11): p. 818-826.

48. Prince, T.L., et al., *Client proteins and small molecule inhibitors display distinct binding preferences for constitutive and stress-*

induced Hsp90 isoforms and their conformationally restricted mutants. PloS one, 2015. **10**(10): p. e0141786.

49. Krukenberg, K.A., et al., *pH-dependent conformational changes in bacterial Hsp90 reveal a Grp94-like conformation at pH 6 that is highly active in suppression of citrate synthase aggregation.* Journal of molecular biology, 2009. **390**(2): p. 278-291.

50. Zhang, H., et al., *A dynamic view of ATP-coupled functioning cycle of Hsp90 N-terminal domain.* Scientific reports, 2015. **5**.

51. Colombo, G., et al., *Understanding ligand-based modulation of the Hsp90 molecular chaperone dynamics at atomic resolution.* Proceedings of the National Academy of Sciences, 2008. **105**(23): p. 7976-7981.

52. Phillips, J.J., et al., *Conformational dynamics of the molecular chaperone Hsp90 in complexes with a co-chaperone and anticancer drugs.* Journal of molecular biology, 2007. **372**(5): p. 1189-1203.

53. Holdgate, G.A. and W.H. Ward, *Measurements of binding thermodynamics in drug discovery.* Drug discovery today, 2005. **10**(22): p. 1543-1550.

54. Wiseman, T., et al., *Rapid measurement of binding constants and heats of binding using a new titration calorimeter.* Analytical biochemistry, 1989. **179**(1): p. 131-137.

55. Mittermaier, A.K. and L.E. Kay, *Observing biological dynamics at atomic resolution using NMR.* Trends in biochemical sciences, 2009. **34**(12): p. 601-611.

56. Wagner, G. and K. Wüthrich, *Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra: basic pancreatic trypsin inhibitor.* Journal of molecular biology, 1982. **155**(3): p. 347-366.

57. Williamson, M.P., *Using chemical shift perturbation to characterise ligand binding.* Progress in nuclear magnetic resonance spectroscopy, 2013. **73**: p. 1-16.

58. Coles, M., M. Heller, and H. Kessler, *NMR-based screening technologies.* Drug discovery today, 2003. **8**(17): p. 803-810.

59. Sattler, M., J. Schleucher, and C. Griesinger, *Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution.* Progress in nuclear magnetic resonance spectroscopy, 1999. **34**: p. 93-158.

60. Kay, L.E., et al., *Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins.* Journal of Magnetic Resonance (1969), 1990. **89**(3): p. 496-514.

61. Jacobs, D.M., et al., *NMR backbone assignment of the N-terminal domain of human HSP90.* Journal of biomolecular NMR, 2006. **36**: p. 52-52.

62. Tugarinov, V. and L.E. Kay, *Quantitative NMR studies of high molecular weight proteins: application to domain orientation and ligand binding in the 723 residue enzyme malate synthase G.* Journal of molecular biology, 2003. **327**(5): p. 1121-1133.

63. Zintsmaster, J.S., B.D. Wilson, and J.W. Peng, *Dynamics of ligand binding from 13C NMR relaxation dispersion at natural abundance.* Journal of the American Chemical Society, 2008. **130**(43): p. 14060-14061.

64. Grutsch, S., S. Brüschweiler, and M. Tollinger, *NMR methods to study dynamic allostery.* PLoS Comput Biol, 2016. **12**(3): p. e1004620.
65. Kim, E., et al., *A single-molecule dissection of ligand binding to a protein with intrinsic dynamics.* Nature chemical biology, 2013. **9**(5): p. 313-318.
66. Mittermaier, A. and L.E. Kay, *New tools provide new insights in NMR studies of protein dynamics.* Science, 2006. **312**(5771): p. 224-228.
67. Sára, T., et al., *Magnetic resonance access to transiently formed protein complexes.* ChemistryOpen, 2014. **3**(3): p. 115-123.
68. Bouvignies, G., et al., *Solution structure of a minor and transiently formed state of a T4 lysozyme mutant.* Nature, 2011. **477**(7362): p. 111-114.
69. Korzhnev, D.M., et al., *A transient and low-populated protein-folding intermediate at atomic resolution.* Science, 2010. **329**(5997): p. 1312-1316.
70. Whittier, S.K., A.C. Hengge, and J.P. Loria, *Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases.* Science, 2013. **341**(6148): p. 899-903.
71. Baldwin, A.J. and L.E. Kay, *NMR spectroscopy brings invisible protein states into focus.* Nature chemical biology, 2009. **5**(11): p. 808-814.
72. Tollinger, M., et al., *Slow dynamics in folded and unfolded states of an SH3 domain.* Journal of the American Chemical Society, 2001. **123**(46): p. 11341-11352.
73. Neudecker, P., P. Lundström, and L.E. Kay, *Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding.* Biophysical journal, 2009. **96**(6): p. 2045-2054.
74. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins.* Nature, 1977. **267**(5612): p. 585-590.
75. Shaw, D.E., et al. *Millisecond-scale molecular dynamics simulations on Anton.* in *Proceedings of the conference on high performance computing networking, storage and analysis.* 2009. IEEE.
76. Durrant, J.D. and J.A. McCammon, *Molecular dynamics simulations and drug discovery.* BMC biology, 2011. **9**(1): p. 1.
77. Cornell, W.D., et al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules.* Journal of the American Chemical Society, 1995. **117**(19): p. 5179-5197.
78. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins.* Chemical reviews, 2006. **106**(5): p. 1589-1615.
79. MacKerell, A.D., N. Banavali, and N. Foloppe, *Development and current status of the CHARMM force field for nucleic acids.* Biopolymers, 2000. **56**(4): p. 257-265.
80. Case, D., et al., *AMBER 12; University of California: San Francisco, 2012.* There is no corresponding record for this reference.
81. Wang, J., et al., *Development and testing of a general amber force field.* Journal of computational chemistry, 2004. **25**(9): p. 1157-1174.
82. Bayly, C.I., et al., *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model.* The Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
83. Jensen, F., *Introduction to computational chemistry.* 2013: John Wiley & Sons.

84. Feenstra, K.A., B. Hess, and H.J. Berendsen, *Improving e_ciency of large timescale molecular dynamics simulations of hydrogen-rich systems.* J Comput Chem, 1999. **20**: p. 786-798.

85. Zhou, R., *Free energy landscape of protein folding in water: explicit vs. implicit solvent.* Proteins: Structure, Function, and Bioinformatics, 2003. **53**(2): p. 148-161.

86. Karplus, M. and J.N. Kushick, *Method for estimating the configurational entropy of macromolecules.* Macromolecules, 1981. **14**(2): p. 325-332.

87. Edholm, O. and H. Berendsen, *Entropy estimation from simulations of non-diffusive systems.* Molecular Physics, 1984. **51**(4): p. 1011-1028.

88. Wereszczynski, J. and J.A. McCammon, *Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition.* Quarterly reviews of biophysics, 2012. **45**(01): p. 1-25.

89. Polyansky, A.A., R. Zubac, and B. Zagrovic, *Estimation of conformational entropy in protein–ligand interactions: a computational perspective.* Computational Drug Discovery and Design, 2012: p. 327-353.

90. Schlitter, J., *Estimation of absolute and relative entropies of macromolecules using the covariance matrix.* Chemical Physics Letters, 1993. **215**(6): p. 617-621.

91. Harris, S.A., et al., *Cooperativity in drug-DNA recognition: a molecular dynamics study.* Journal of the American Chemical Society, 2001. **123**(50): p. 12658-12663.

92. Andricioaei, I. and M. Karplus, *On the calculation of entropy from covariance matrices of the atomic fluctuations.* The Journal of Chemical Physics, 2001. **115**(14): p. 6289-6292.

93. Meyer, T., et al., *Essential dynamics: a tool for efficient trajectory compression and management.* Journal of Chemical Theory and Computation, 2006. **2**(2): p. 251-258.

94. Krishnan, M. and J.C. Smith, *Response of Small-Scale, Methyl Rotors to Protein− Ligand Association: A Simulation Analysis of Calmodulin− Peptide Binding.* Journal of the American Chemical Society, 2009. **131**(29): p. 10083-10091.

95. Diehl, C., et al., *Conformational entropy changes upon lactose binding to the carbohydrate recognition domain of galectin-3.* Journal of biomolecular NMR, 2009. **45**(1-2): p. 157-169.

96. Genheden, S. and U. Ryde, *Will molecular dynamics simulations of proteins ever reach equilibrium?* Physical Chemistry Chemical Physics, 2012. **14**(24): p. 8662-8677.

97. Hensen, U., F. Gräter, and R.H. Henchman, *Macromolecular Entropy Can Be Accurately Computed from Force.* Journal of Chemical Theory and Computation, 2014. **10**(11): p. 4777-4781.

98. Kamenik, A.S., et al., *Localization of Millisecond Dynamics: Dihedral Entropy from Accelerated MD.* Journal of Chemical Theory and Computation, 2016. **12**(8): p. 3449-3455.

99. Akke, M., R. Brueschweiler, and A.G. Palmer III, *NMR order parameters and free energy: an analytical approach and its application to cooperative calcium (2+) binding by calbindin D9k.*

Journal of the American Chemical Society, 1993. **115**(21): p. 9832-9833.

100. Wand, A.J., *The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation.* Current opinion in structural biology, 2013. **23**(1): p. 75-81.

101. Frederick, K.K., et al., *Conformational entropy in molecular recognition by proteins: A first example from the calmodulin system.* Nature, 2007. **448**(7151): p. 325.

102. Sharp, K.A., et al., *On the relationship between NMR-derived amide order parameters and protein backbone entropy changes.* Proteins: Structure, Function, and Bioinformatics, 2015. **83**(5): p. 922-930.

103. Henzler-Wildman, K.A., et al., *A hierarchy of timescales in protein dynamics is linked to enzyme catalysis.* Nature, 2007. **450**(7171): p. 913.

104. Diehl, C., et al., *Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3.* Journal of the American Chemical Society, 2010. **132**(41): p. 14577-14589.

105. Tzeng, S.-R. and C.G. Kalodimos, *Protein activity regulation by conformational entropy.* Nature, 2012. **488**(7410): p. 236-240.

106. Desjardins, P. and D. Conklin, *NanoDrop microvolume quantitation of nucleic acids.* JoVE (Journal of Visualized Experiments), 2010(45): p. e2565-e2565.

107. Gasteiger, E., et al., *Protein identification and analysis tools on the ExPASy server.* 2005: Springer.

108. Keller, S., et al., *High-precision isothermal titration calorimetry with automated peak-shape analysis.* Analytical chemistry, 2012. **84**(11): p. 5066-5073.

109. Solyom, Z., et al., *BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins.* Journal of biomolecular NMR, 2013. **55**(4): p. 311-321.

110. Favier, A. and B. Brutscher, *Recovering lost magnetization: polarization enhancement in biomolecular NMR.* Journal of biomolecular NMR, 2011. **49**(1): p. 9-15.

111. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* Journal of biomolecular NMR, 1995. **6**(3): p. 277-293.

112. Keller, R., *The computer aided resonance assignment tutorial*, 2004, Cantina Verlag, Goldau, Switzerland.

113. Vranken, W.F., et al., *The CCPN data model for NMR spectroscopy: development of a software pipeline.* Proteins: Structure, Function, and Bioinformatics, 2005. **59**(4): p. 687-696.

114. Vallurupalli, P., et al., *Measurement of bond vector orientations in invisible excited states of proteins.* Proceedings of the National Academy of Sciences, 2007. **104**(47): p. 18473-18477.

115. Jez, J.M., et al., *Crystal structure and molecular modeling of 17-DMAG in complex with human Hsp90.* Chemistry & biology, 2003. **10**(4): p. 361-368.

116. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.* Protein engineering, 1995. **8**(2): p. 127-134.

117. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* Journal of molecular graphics, 1996. **14**(1): p. 33-38.

118. Pronk, S., et al., *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.* Bioinformatics, 2013. **29**(7): p. 845-854.

119. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins: Structure, Function, and Bioinformatics, 2010. **78**(8): p. 1950-1958.

120. Götz, A.W., et al., *Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born.* Journal of chemical theory and computation, 2012. **8**(5): p. 1542-1555.

121. Pettersen, E.F., et al., *UCSF Chimera—a visualization system for exploratory research and analysis.* Journal of computational chemistry, 2004. **25**(13): p. 1605-1612.

122. O'Boyle, N.M., et al., *Open Babel: An open chemical toolbox.* Journal of cheminformatics, 2011. **3**(1): p. 1.

123. Frisch, A., *Gaussian 03.* 2004: Gaussian.

124. Schaftenaar, G. and J.H. Noordik, *Molden: a pre-and post-processing program for molecular and electronic structures.* Journal of computer-aided molecular design, 2000. **14**(2): p. 123-134.

125. Stebbins, C.E., Russo, A.A., Schneider, C., Rosen, N., Hartl, F.U. and Pavletich, N.P., *Crystal Structure of an Hsp90-Geldanamycin Complex: Targeting of a Protein Chaperone by an Antitumor Agent.* Cell, 1997. **89**: p. 239-250.

126. Shields, G.C., C.A. Laughton, and M. Orozco, *Molecular Dynamics Simulations of the d (T⊙ A⊙ T) Triple Helix.* Journal of the American Chemical Society, 1997. **119**(32): p. 7463-7469.

127. DeBoer, C., et al., *Geldanamycin, a new antibiotic.* The Journal of antibiotics, 1970. **23**(9): p. 442-447.

128. Schnur, R., et al., *erbB-2 oncogene inhibition by geldanamycin derivatives: synthesis, mechanism of action, and structure-activity relationships.* Journal of medicinal chemistry, 1995. **38**(19): p. 3813-3820.

129. Evans, G. and N. White, *Radicicolin and radicicol, two new antibiotics produced by Cylindrocarpon radicicola.* Transactions of the British mycological Society, 1966. **49**(4): p. 563IN7-576.

130. Roe, S.M., et al., *Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin.* Journal of medicinal chemistry, 1999. **42**(2): p. 260-266.

131. Obermann, W.M., et al., *In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis.* The Journal of cell biology, 1998. **143**(4): p. 901-910.

132. Immormino, R.M., et al., *Ligand-induced conformational shift in the N-terminal domain of GRP94, an Hsp90 chaperone.* Journal of Biological Chemistry, 2004. **279**(44): p. 46162-46171.

133. Ooi, T., et al., *Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides.* Proceedings of the National Academy of Sciences, 1987. **84**(10): p. 3086-3090.

134. Myers, J.K., C. Nick Pace, and J. Martin Scholtz, *Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding.* Protein Science, 1995. **4**(10): p. 2138-2148.

135. Holdgate, G.A., *Making cool drugs hot.* Biotechniques, 2001. **31**: p. 164-184.

136. Onuoha, S., et al., *Mechanistic studies on Hsp90 inhibition by ansamycin derivatives.* Journal of molecular biology, 2007. **372**(2): p. 287-297.

137. Olsson, T.S., et al., *The thermodynamics of protein–ligand interaction and solvation: insights for ligand design.* Journal of molecular biology, 2008. **384**(4): p. 1002-1017.

138. King, N.M., et al., *Extreme Entropy–Enthalpy Compensation in a Drug-Resistant Variant of HIV-1 Protease.* ACS chemical biology, 2012. **7**(9): p. 1536-1546.

139. Zhuravleva, A. and L.M. Gierasch, *Allosteric signal transmission in the nucleotide-binding domain of 70-kDa heat shock protein (Hsp70) molecular chaperones.* Proceedings of the National Academy of Sciences, 2011. **108**(17): p. 6987-6992.

140. Foster, M.P., C.A. McElroy, and C.D. Amero, *Solution NMR of large molecules and assemblies.* Biochemistry, 2007. **46**(2): p. 331-340.

141. Nilapwar, S., et al., *Structural–thermodynamic relationships of interactions in the N-terminal ATP-binding domain of Hsp90.* Journal of molecular biology, 2009. **392**(4): p. 923-936.

142. Spolar, R.S., J.-H. Ha, and M.T. Record, *Hydrophobic effect in protein folding and other noncovalent processes involving proteins.* Proceedings of the National Academy of Sciences, 1989. **86**(21): p. 8382-8385.

143. Cooper, A., *Protein heat capacity: an anomaly that maybe never was.* The Journal of Physical Chemistry Letters, 2010. **1**(22): p. 3298-3304.

144. Cooper, A., *Heat capacity effects in protein folding and ligand binding: a re-evaluation of the role of water in biomolecular thermodynamics.* Biophysical chemistry, 2005. **115**(2): p. 89-97.

145. Chervenak, M.C. and E.J. Toone, *A direct measure of the contribution of solvent reorganization to the enthalpy of binding.* Journal of the American Chemical Society, 1994. **116**(23): p. 10533-10539.

146. Ladbury, J.E., *Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design.* Chemistry & biology, 1996. **3**(12): p. 973-980.

147. Chandler, D., *Interfaces and the driving force of hydrophobic assembly.* Nature, 2005. **437**(7059): p. 640-647.

148. Karplus, P.A. and C. Faerman, *Ordered water in macromolecular structure.* Current Opinion in Structural Biology, 1994. **4**(5): p. 770-776.

149. Galamba, N., *Water's structure around hydrophobic solutes and the iceberg model.* The Journal of Physical Chemistry B, 2013. **117**(7): p. 2153-2159.

150. Levitt, M. and B.H. Park, *Water: now you see it, now you don't.* Structure, 1993. **1**(4): p. 223-226.

151. Tetko, I.V., et al., *Virtual computational chemistry laboratory–design and description.* Journal of computer-aided molecular design, 2005. **19**(6): p. 453-463.

152. Irudayam, S.J. and R.H. Henchman, *Solvation theory to provide a molecular interpretation of the hydrophobic entropy loss of noble-gas hydration.* Journal of Physics: Condensed Matter, 2010. **22**(28): p. 284108.

153. Irudayam, S.J. and R.H. Henchman, *Long-range hydrogen-bond structure in aqueous solutions and the vapor-water interface.* The Journal of chemical physics, 2012. **137**(3): p. 034508.

154. Zhang, Y. and Z. Xu, *Atomic radii of noble gas elements in condensed phases.* American Mineralogist, 1995. **80**(7-8): p. 670-675.

155. Huggins, D.J., *Quantifying the entropy of binding for water molecules in protein cavities by computing correlations.* Biophysical journal, 2015. **108**(4): p. 928-936.

156. Dunitz, J.D., *The entropic cost of bound water in crystals and biomolecules.* Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information, 1994. **264**(5159): p. 670-670.

157. Killian, B.J., J.Y. Kravitz, and M.K. Gilson, *Extraction of configurational entropy from molecular simulations via an expansion approximation.* The Journal of chemical physics, 2007. **127**(2): p. 024107.

158. Schäfer, H., et al., *Entropy calculations on the molten globule state of a protein: Side-chain entropies of α-lactalbumin.* Proteins: Structure, Function, and Bioinformatics, 2002. **46**(2): p. 215-224.

159. Kitao, A., S. Hayward, and N. Go, *Energy landscape of a native protein: jumping-among-minima model.* Proteins Structure Function and Genetics, 1998. **33**(4): p. 496-517.

160. Caves, L.S., J.D. Evanseck, and M. Karplus, *Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin.* Protein Science, 1998. **7**(3): p. 649-666.

161. Hu, X., et al., *The dynamics of single protein molecules is non-equilibrium and self-similar over thirteen decades in time.* Nature Physics, 2015.

162. Papaleo, E., *Integrating atomistic molecular dynamics simulations, experiments, and network analysis to study protein dynamics: strength in unity.* Frontiers in molecular biosciences, 2015. **2**.

163. Luitz, M., et al., *Exploring biomolecular dynamics and interactions using advanced sampling methods.* Journal of Physics: Condensed Matter, 2015. **27**(32): p. 323101.

164. Lindorff-Larsen, K., et al., *Systematic validation of protein force fields against experimental data.* PloS one, 2012. **7**(2): p. e32131.

165. Piana, S., K. Lindorff-Larsen, and D.E. Shaw, *Protein folding kinetics and thermodynamics from atomistic simulation.* Proceedings of the National Academy of Sciences, 2012. **109**(44): p. 17845-17850.

166. Thepchatri, P., et al., *Relationship among ligand conformations in solution, in the solid state, and at the Hsp90 binding site: geldanamycin and radicicol.* Journal of the American Chemical Society, 2007. **129**(11): p. 3127-3134.

167. Gilson, M.K. and K.K. Irikura, *Symmetry numbers for rigid, flexible, and fluxional molecules: theory and applications.* The Journal of Physical Chemistry B, 2010. **114**(49): p. 16304-16317.
168. Munos, B., *Lessons from 60 years of pharmaceutical innovation.* Nature reviews Drug discovery, 2009. **8**(12): p. 959-968.
169. Borhani, D.W. and D.E. Shaw, *The future of molecular dynamics simulations in drug discovery.* Journal of computer-aided molecular design, 2012. **26**(1): p. 15-26.
170. Roe, S.M., et al., *The mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50 cdc37.* Cell, 2004. **116**(1): p. 87-98.
171. Dittmar, K.D. and W.B. Pratt, *Folding of the Glucocorticoid Receptor by the Reconstituted hsp90-based Chaperone Machinery THE INITIAL hsp90· p60· hsp70-DEPENDENT STEP IS SUFFICIENT FOR CREATING THE STEROID BINDING CONFORMATION.* Journal of Biological Chemistry, 1997. **272**(20): p. 13047-13054.
172. Kosano, H., et al., *The assembly of progesterone receptor-hsp90 complexes using purified proteins.* Journal of Biological Chemistry, 1998. **273**(49): p. 32973-32979.
173. Beebe, K., et al., *Posttranslational modification and conformational state of heat shock protein 90 differentially affect binding of chemically diverse small molecule inhibitors.* Oncotarget, 2013. **4**(7): p. 1065-1074.
174. Freire, E., *Do enthalpy and entropy distinguish first in class from best in class?* Drug discovery today, 2008. **13**(19): p. 869-874.

Appendix 1 - NMR samples and experiments

Samples and experiments:

Sample 1: Hsp90-NTD (D9-E246) (^{2}H, ^{13}C, ^{15}N), 20 mM Na_2HPO_4, 20 mM NaH_2PO_4, 1% AEBSF, 5% D_2O

Sample 2 and 3: As sample 1 plus 750 µM ligands (either GVK0153 or 17-DMAG)

Conditions: pH 7.5; temperature: 25°C

Experiments:

^{1}H^{15}N-TROSY

TROSY-HNCA

TROSY-HNCO

TROSY-HNcaCO

Hsp90-NTD FASTA:

DQPMEEEEVETFAFQAEIAQLMSLIINTFYSNKEIFLRELISNSSDALDKIRYE
SLTDPSKLDSGKELHINLIPNKQDRTLTIVDTGIGMTKADLINNLGTIAKSGTK
AFMEALQAGADISMIGQFGVGFYSAYLVAEKVTVITKHNDDEQYAWESSA
GGSFTVRTDTGEPMGRGTKVILHLKEDQTEYLEERRIKEIVKKHSQFIGYPI
TLFVEKERDKEVSDDEAE

Appendix 2 – MD parameter files

em.mdp (energy minimisation)

```
; VARIOUS PREPROCESSING OPTIONS =
title           =
cpp             = /lib/cpp -traditional
include         =
define          = -DPOSRES

; RUN CONTROL PARAMETERS =
integrator      = steep
; start time and timestep in ps =
tinit           = 0
dt              = 0.001
nsteps          = 10000

; ENERGY MINIMIZATION OPTIONS =
emtol           = 0.00001
emstep          = 0.1
nstcgsteep      = 1000

; OPTIONS FOR ELECTROSTATICS AND VDW =
; Method for doing electrostatics =
coulombtype     = PME
rcoulomb        = 1.2
rlist           = 1.2
; Method for doing Van der Waals =
vdw-type        = Cut-off
rvdw            = 1.2
; Spacing for the PME/PPPM FFT grid =
fourierspacing  = 0.12
; FFT grid size, when a value is 0 fourierspacing will be used =
fourier_nx      = 0
fourier_ny      = 0
fourier_nz      = 0
; EWALD/PME/PPPM parameters =
pme_order       = 4
ewald_rtol      = 1e-05
ewald_geometry  = 3d
epsilon_surface = 0
optimize_fft    = no


***********************************
```

md1.mdp (equilibration)

```
; VARIOUS PREPROCESSING OPTIONS =
title           =
cpp             = /lib/cpp -traditional
include         =
define          = -DPOSRES
```

```
; RUN CONTROL PARAMETERS =
integrator = md        ; leap-frog integrator
nsteps    = 5000
dt        = 0.002          ; 2 fs

; Output control
nstxout    = 500            ; save coordinates every n ps
nstvout    = 500            ; save velocities every n ps
nstxtcout      = 500
nstfout        = 0
nstenergy  = 500            ; save energies every n ps
nstlog    = 1000
nstcomm    = 10
nstcalenergy = 10
xtc-grps   = Protein

; Bond parameters
continuation = no          ; Restarting after NVT
constraint_algorithm = lincs
constraints = all-bonds ; all bonds (even heavy atom-H bonds) constrained
lincs_iter = 1
lincs_order = 4        ;
; Neighborsearching
ns_type    = grid          ; search neighbouring grid cells
nstlist    = 10
rlist      = 1.2            ; short-range neighbour list cutoff (in nm)
rcoulomb   = 1.2            ; short-range electrostatic cutoff (in nm)
vdw-type       = Cut-off
rvdw      = 1.2            ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype = PME
pme_order  = 4
fourierspacing  = 0.12
epsilon-rf = 1
tcoupl     = V-rescale  ; modified Berendsen thermostat
tc-grps    = Protein Non-Protein
tau_t      = 0.1  0.1
ref_t      = 100  100
pcoupl     = Berendsen
pcoupltype = isotropic
tau_p      = 1.0
ref_p      = 1.0
compressibility = 4.5e-5
refcoord_scaling = com
; Periodic boundary conditions
pbc        = xyz            ; 3-D PBC
; Dispersion correction
DispCorr   = EnerPres
; Velocity generation
gen-vel    = yes
```

```
gen-temp   = 100
gen-seed   = 32926
```

****

md2.mdp (equilibration)
```
; VARIOUS PREPROCESSING OPTIONS =
title              =
cpp                = /lib/cpp -traditional
include            =
define             = -DPOSRES

; RUN CONTROL PARAMETERS =
integrator = md        ; leap-frog integrator
nsteps    = 5000
dt        = 0.002           ; 2 fs

; Output control
nstxout    = 500            ; save coordinates every n ps
nstvout    = 500            ; save velocities every n ps
nstxtcout      = 500
nstfout        = 0
nstenergy  = 500             ; save energies every n ps
nstlog     = 1000           ; update log file every n ps
nstcomm    = 10
nstcalenergy = 10
xtc-grps   = Protein

; Bond parameters
continuation = yes
constraint_algorithm = lincs
constraints = all-bonds ; all bonds (even heavy atom-H bonds) constrained
lincs_iter = 1

lincs_order = 4

; Neighborsearching
ns_type    = grid            ; search neighboring grid cells
nstlist    = 10
rlist      = 1.2             ; short-range neighborlist cutoff (in nm)
rcoulomb   = 1.2             ; short-range electrostatic cutoff (in nm)
vdw-type       = Cut-off
rvdw       = 1.2             ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype = PME
pme_order  = 4
fourierspacing  = 0.12
epsilon-rf = 1
; Temperature coupling is on
tcoupl     = V-rescale  ; modified Berendsen thermostat
tc-grps    = Protein Non-Protein
tau_t      = 0.1  0.1
ref_t      = 100  100
; Pressure coupling is on
```

```
pcoupl    = Berendsen
pcoupltype = isotropic
tau_p    = 1.0
ref_p    = 1.0
compressibility = 4.5e-5
refcoord_scaling = com
; Periodic boundary conditions
pbc      = xyz           ; 3-D PBC
; Dispersion correction
DispCorr  = EnerPres  ; account for cut-off vdW scheme
; Velocity generation
gen-vel   = no       ; Velocity generation is off
gen-temp  = 100
***********************************************
```

md3.mdp (equilibration)

```
; VARIOUS PREPROCESSING OPTIONS =
title             =
cpp               = /lib/cpp -traditional
include           =
define            = -DPOSRES

; RUN CONTROL PARAMETERS =
integrator = md       ; leap-frog integrator
nsteps    = 5000
dt        = 0.002         ; 2 fs

; Output control
nstxout    = 500          ; save coordinates every n ps
nstvout    = 500          ; save velocities every n ps
nstxtcout     = 500
nstfout      = 0
nstenergy  = 500           ; save energies every n ps
nstlog    = 1000          ; update log file every n ps
nstcomm    = 10
nstcalenergy = 10
xtc-grps   = Protein

; Bond parameters
continuation = yes
constraint_algorithm = lincs
constraints = all-bonds ; all bonds (even heavy atom-H bonds) constrained
lincs_iter = 1
lincs_order = 4
; Neighborsearching
ns_type   = grid           ; search neighboring grid cells
nstlist   = 10
rlist     = 1.2            ; short-range neighborlist cutoff (in nm)
rcoulomb   = 1.2              ; short-range electrostatic cutoff (in nm)
vdw-type      = Cut-off
rvdw       = 1.2            ; short-range van der Waals cutoff (in nm)
```

; Electrostatics
coulombtype = PME   ; Particle Mesh Ewald for long-range electrostat
ics
pme_order = 4   ; cubic interpolation
fourierspacing = 0.12   ; grid spacing for FFT
epsilon-rf = 1
; Temperature coupling is on
tcoupl   = V-rescale
tc-grps   = Protein Non-Protein
tau_t   = 0.1 0.1
ref_t   = 300   300
; Pressure coupling is on
pcoupl   = Berendsen
pcoupltype = isotropic
tau_p   = 1.0
ref_p   = 1.0
compressibility = 4.5e-5
refcoord_scaling = com
; Periodic boundary conditions
pbc   = xyz   ; 3-D PBC
; Dispersion correction
DispCorr   = EnerPres   ; account for cut-off vdW scheme
; Velocity generation
gen-vel   = no
gen-temp   = 100
****

md9.mdp (production MD)

; VARIOUS PREPROCESSING OPTIONS =
title   =
cpp   = /lib/cpp -traditional
include   =
define   = -DPOSRES

; RUN CONTROL PARAMETERS =
integrator = md   ; leap-frog integrator
nsteps   = 25000000   ; 2 * 25000000 = 50ns
dt   = 0.002   ; 2 fs

; Output control
nstxout   = 500   ; save coordinates every 1 ps
nstvout   = 500   ; save velocities every 1 ps
nstxtcout   = 500
nstfout   = 0
nstenergy = 500   ; save energies every 1 ps
nstlog   = 1000   ; update log file every 2 ps
nstcomm   = 10
nstcalcenergy   = 10
xtc-grps   = Protein

; Bond parameters

```
continuation = yes
constraint_algorithm = lincs
constraints = all-bonds
lincs_iter = 1
lincs_order = 4
; Neighborsearching
ns_type    = grid
nstlist    = 10        ; 20 fs
rlist      = 1.2              ; short-range neighborlist cutoff (in nm)
rcoulomb   = 1.2             ; short-range electrostatic cutoff (in nm)
vdw-type       = Cut-off
rvdw       = 1.2             ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype = PME            ; Particle Mesh Ewald for long-range
electrostatics
pme_order  = 4          ; cubic interpolation
fourierspacing  = 0.12       ; grid spacing for FFT
epsilon-rf = 1
; Temperature coupling is on
tcoupl     = V-rescale  ; modified Berendsen thermostat
tc-grps    = Protein Non-Protein  ; two coupling groups - more accurate
tau_t      = 0.1  0.1
ref_t      = 300  300   ; reference temperature, one for each group, in K
; Pressure coupling is on
pcoupl     = Berendsen
pcoupltype = isotropic  ; uniform scaling of box vectors
tau_p      = 1.0              ; time constant, in ps
ref_p      = 1.0             ; reference pressure, in bar
compressibility = 4.5e-5     ; isothermal compressibility of water, bar^-1
refcoord_scaling = com
; Periodic boundary conditions
pbc        = xyz            ; 3-D PBC
; Dispersion correction
DispCorr   = EnerPres   ; account for cut-off vdW scheme
; Velocity generation
gen_vel    = no        ; Velocity generation is off
gen-temp   = 100
```

Appendix 3 – CPMG data