Dynamics and Thermodynamics of Protein-Ligand Interactions

Richard William Malham

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

The University of Leeds

Astbury Centre for Structural Molecular Biology

January 2012

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

Acknowledgements

To list everyone would be impossible. So many people have got me here, whether they have taught, helped, collaborated, conversed, fed, watered, housed or generally put up with me.

My primary supervisor Steve Homans, for his sanity-restoring powers and heroic proof-reading, let alone all the guidance throughout. My secondary supervisor Emanuele Paci, for his reassurances, guidance and his enthusiasm for a wide range of questions. Gary Thompson and Arnout Kalverda for proof reading, teaching me NMR and their sustained patience. Andrew Nightingale for his varied teaching and help with problem solving. Girish Tampi for continuously interesting discussions and provision of a labelled MUP sample for my experiments. Sue Matthews for her technician skills in MUP production. Cait Dennis for her inhuman tolerance of my continuous questions about her crystallography. Chi Trinh for teaching me crystallography, and collaborating on the rotor work. Martin Fascione, for his almost legendary patience, skill and humour in guiding me through the ups and downs of organic synthesis. Kasia Tych for discussions, and for designing a THz system that opened up a whole avenue of work. Phil Morrison, who proficiently continued my work when I had to finally start writing my thesis. Julie Roy and Charlie Laughton for being interested in MUP enough themselves to do QCM and long timescale MD. David Cramer, John Clements and Steve Martin of the University of Texas at Austin for much discussion and inspiration, as well as more direct help with the crystallographic analysis as mentioned in §3.2.2.4. All the members and honorary members of the lab formerly known as Homans; Neil, Timo, Andras, Tosh, Al, Eva, Wei-Feng Xue and George. From the Paci lab I'd especially like to thank the long-gone Zu Thur Yew for all our conversations and his MD tutelage.

I hope that one day I can truly repay the incredible hospitality of Ben Wolfson and Kelsey Forbes. John Dyer and Hannah Sargeant for becoming a home-from-home. The new friends I have made or sustained over the last four years, and whom I hope to see for many years to come, for making life excellent. Finally, and most importantly, my parents: for their love, and their unwavering dedication and support of my education.

Abstract

Complex networks of protein-ligand interactions underpin cellular function and communication. Disease can arise from disruption of these networks through the alteration of protein-ligand interaction affinities, for example by protein mutation or ligand modification. Understanding the mechanisms and principles that define affinity is therefore critical to both understanding and engineering biomolecular interactions, e.g. optimising drug molecules to interact effectively with their biomolecular targets. Thermodynamics reveals that affinity can be expressed in terms of the Gibbs free energy change upon interaction. In turn, this is composed of enthalpic and entropic terms, which can be thought of loosely as arising from structural and dynamic factors respectively. Though enthalpic terms can be estimated to a reasonable degree using structural data, a better understanding of entropic contributions from dynamic processes is required.

The mouse major urinary protein (MUP) has been successfully established as a model system to investigate the thermodynamics of protein-ligand interactions. This work uses MUP, and employs a wide range of biophysical techniques, to develop our understanding of the dynamic factors in the thermodynamics of protein-ligand interactions. Four factors are addressed. Protein solvation is addressed by investigating proposed entropic solvation of the MUP binding pocket, and the possibility of engineering a new binding profile through manipulation of sidechains and solvation in the binding pocket. Ligand conformational entropy is addressed by performing the first systematic assessment of the widely predicted, yet inconsistently observed, benefits of removing and restricting ligand bonds. The greatest entropic loss upon binding, that of ligand rotational and translational entropy, is addressed by assessing MD predictions of significant residual translation and rotational motion of IBMP bound to MUP. This is achieved by using a combination of NMR techniques. Finally, protein dynamics are addressed by undertaking a preliminary investigation of a potentially promising novel technique for probing site-specific changes in protein dynamics upon ligand binding.

Abbreviations

MUP	Major Urinary Protein
HBP	Histamine Binding Protein
IBMP	2-methoxy-3-isobutyl pyrazine
IPMP	2-methoxy-3-isopropyl-pyrazine
NPOME	N-acetylated phenylalanine methyl ester
LB	Luria broth
PBS	Phosphate buffered saline
MD	Molecular Dynamics
NMA	Normal mode analysis
ITC	Isothermal titration calorimetry
NMR	Nuclear Magnetic resonance
HSQC	Heteronuclear single-quantum correlation spectroscopy
NOESY	Nuclear overhauser effect spectroscopy
TOCSY	Total correlation spectroscopy
rdc	Residual dipolar coupling
NOE	Nuclear Overhauser effect
ESI-MS	Electrospray ionisation mass spectrometry
QCM	Quartz crystal microbalance
MSA	Molecular surface area
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
PDB	Protein data bank
vdW	van der Waals

Contents

Chapter 4 Ligand Rotational and Translational Entropy

Chapter 5 Protein Dynamics

List of Figures

Chapter 1 Introduction

Chapter 2 Protein Solvation

Chapter 3 Ligand Conformational Entropy

Chapter 4 Ligand Rotational and Translational Entropy

Chapter 5 Protein Dynamics

Appendix I Analysis of intermediates in IBMP organic synthesis (§4.2.1)

List of Tables

Chapter 2 Protein Solvation

Chapter 3 Ligand Conformational Entropy

Chapter 4 Ligand Rotational and Translational Entropy

Chapter 5 Protein Dynamics

Chapter 1

General Introduction

1.1 Biomolecular interactions and thermodynamics

1.1.1 Affinity underpins biology, and thermodynamics underpin affinity

In a reductionist view, biomolecular interactions are the root cause of all biological phenomena. Networks of molecular interactions, interference with which can cause diseases such as cancer, take place in a crowded and diverse cellular milieu. This fine line between health and disease is predicated on the exquisite specificity of molecular interactions, i.e. the affinity of two partners to interact.

Understanding the mechanisms and principles that define affinity is therefore critical to both understanding and engineering biomolecular interactions, for example optimising drug molecules to interact effectively with their biomolecular targets. Protein interactions with small molecules, henceforth referred to as ligands, are responsible for many of the key interaction networks, such as hormonal signalling or neurotransmission. A seemingly modest difference in the chemical structure of a protein or ligand can result in a great change to affinity, and effect a pronounced change in biological phenomenon.

Despite the increasing body of protein and protein-ligand structural data, its use in computational drug design, and the understanding of molecular interactions in general, is currently inefficient. Biomolecular interactions are governed by thermodynamics, wherein binding affinity, K_a, is dictated by the change in the standard free energy upon binding, ΔG°_b. This is in turn composed of enthalpic and entropic terms, ΔH°_b and $T\Delta S^\circ_b$, as demonstrated below. Protein, P, and ligand, L, can interact to form a protein-ligand complex, with the binding affinity represented by the ratio of free species to complex at equilibrium, as shown in Equations 1.1 and 1.2.

$$P + L \rightleftharpoons PL$$

Equation 1.1

$$K_a = \frac{[PL]}{[P][L]} = \frac{1}{K_d}$$

Equation 1.2

K_d, the inverse of the association constant, K_a, is also commonly used to represent affinity. It has units of concentration, whereby the lower the concentration, the tighter the binding. At a given temperature, affinity is represented thermodynamically by the change in standard Gibbs free energy of binding, $\Delta G°_b$, calculated from K_a using Equation 1.3.

$$\Delta G°_b = -RT \ln K_a$$

Equation 1.3

R is the gas constant of 8.314 J mol^{-1} K^{-1} and T is the absolute temperature. $\Delta G°_b$ is composed of two terms, the standard enthalpy change of binding, $\Delta H°_b$, and standard entropy change of binding, $T\Delta S°_b$. Spontaneous, i.e. favourable, processes are those where the free energy is minimised, such that the enthalpy change is negative and entropy change positive.

$$\Delta G°_b = \Delta H°_b - T\Delta S°_b$$

Equation 1.4

Understanding the mechanisms and principles that define affinity can therefore be achieved through understanding the physical factors that define the sign and magnitude of $\Delta H°_b$ and $T\Delta S°_b$. This requires a framework, including the contribution of solvent, for experimentally decomposing these contributions.

1.1.2 Dissecting affinity: thermodynamic decomposition using a perturbation approach

We can overcome the inherent complexity of unravelling binding thermodynamics, and analyse the structural basis of thermodynamic contributions more simply, by considering the differences observed when comparing very closely related systems.

This is called the perturbation approach. This could be the contribution of a single factor, such as a water molecule, a bond or a functional group. Such comparisons allow many thermodynamic contributions to binding to either cancel out, or become zero to first order. The thermodynamic values defining a biomolecular interaction are state functions, meaning that they are independent of the path followed by the interaction: only the initial and final states matter. As such we can represent the biomolecular interactions of two closely related ligands to the same protein using the following Born-Haber thermodynamic cycle.[1]

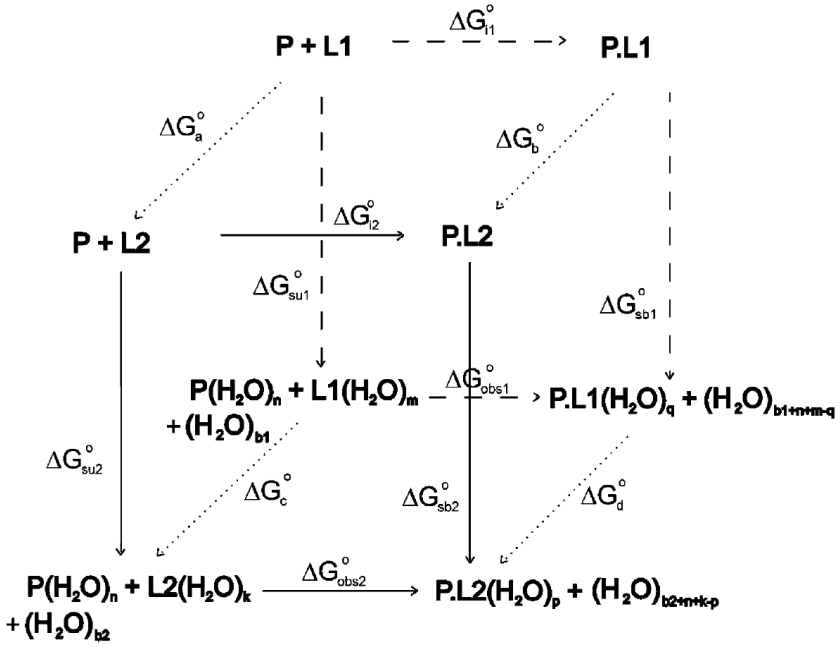


Figure 1.1 Born-Haber thermodynamic cycle for comparison of two ligands binding to the same protein. Reproduced from reference [1]. The back plane of the cube represents the binding of one ligand, L1, and the front plane corresponds to the binding of another ligand, L2. The top of the cube pertains to the difference in the interactions in the absence of solvent, termed 'intrinsic', and the bottom in the presence of solvent.

Each interaction can be described using the following equation, wherein ΔG°_b is referred to as the observed free energy of binding, ΔG°_{obs}

$$\Delta G^\circ_{obs} = \Delta G^\circ_i + [\Delta G^\circ_{sb} - \Delta G^\circ_{su}]$$

Equation 1.5

where subscripts represent b for binding, i for intrinsic, sb for complex solvation and su for the solvation of uncomplexed components. Intrinsic refers to the interaction in the absence of solvent. Therefore when comparing the difference in binding of two related ligands to the same protein, the difference in the observed free energy changes can be expressed as Equation 1.6.

$$\Delta\Delta G^{\circ}_{obs} = \Delta G^{\circ}_{obs2} - \Delta G^{\circ}_{obs1}$$
$$= [\Delta G^{\circ}_{i2} - \Delta G^{\circ}_{i1}] + ([\Delta G^{\circ}_{sb2} - \Delta G^{\circ}_{sb1}] - [\Delta G^{\circ}_{su2} - \Delta G^{\circ}_{su1}])$$

Equation 1.6

1 and 2 refer to the ligands being compared. Equation 1.6 can also be written for the enthalpy or entropy of binding in place of the free energy. This framework allows the physical contributions to be addressed separately as the terms on the right hand side of the equation; those arising from solvent behaviour and interactions, and those arising from factors internal to the protein or ligand. Using the same protein, the contribution from the solvation of the unbound protein cancels out. Likewise, using the same ligand, for example comparing binding to two mutants or versions of the same protein, the ligand desolvation term cancels out.

The observed thermodynamic parameters are obtained by isothermal titration calorimetry (ITC), described in §1.3.1. Using decomposition with a perturbation approach necessitates an appreciation of the main physical sources of intrinsic and solvent enthalpic and entropic contributions, and a range of techniques capable of measuring, or at least approximating, them.

1.1.3 Main sources of intrinsic and solvent enthalpy and entropy, and methods for their measurement or approximation

1.1.3.1 Intrinsic enthalpy

In chemical reactions, enthalpy arises from the making or breaking of covalent bonds. Alternatively in biomolecular interactions, enthalpy arises from the making or breaking

of non-covalent interactions. These most commonly include hydrogen bonds, dispersive forces such as vdW interactions and electrostatic interactions such as salt bridges. These are distance, permittivity and orientational dependent, and therefore their energy may fluctuate somewhat as the system undergoes thermal motion. Upon ligand binding, intrinsic enthalpy changes can arise from protein conformational change and the formation of protein-ligand interactions upon complexation. The enthalpic content of non-covalent interactions can be calculated to a close approximation using high resolution structural data from X-ray crystallography and NMR. Energies can be calculated using known relationships, such as the Lennard-Jones 12-6 or 10-12 potentials and Coulomb's Law. These formulae are used in molecular dynamics (MD) both to study the time evolution of biomolecular systems and in computational ligand docking calculations, §1.3.3.[2-4]

1.1.3.2 Intrinsic entropy

Entropic contributions arise from changes in disorder and therefore dynamics in the system. Unlike enthalpic changes, which are approximated relatively easily from structural data, entropic changes are more elusive.

The largest source of unfavourable intrinsic entropy upon protein-ligand binding is considered to be the loss of rotational and translational entropy of the ligand. Due to the size difference, complexation affects protein rotation and translation very little, whereas a ligand is presumed to lose almost all independent rotation and translation in assuming its bound position. Theoretical calculations predict that for most small organic molecules, this loss is on the order of -40 to -60 kJ mol^{-1}.[5] Residual ligand motion is usually predicted using MD simulations, or inferred by looking for high B-factors in x-ray crystal structures, which are an approximate indicator of disorder.

Intrinsic entropy changes can also arise from ligand conformational dynamics. Upon assuming its bound position, ligand bond rotations that allowed sampling of multiple conformational states in solution are restricted, resulting in an unfavourable intrinsic entropy contribution. This has been proposed as costing approximately 5 to 6 kJ mol^{-1} per restricted bond.[5] Many attempts have been made to experimentally observe this contribution by observing the effects of ligand constraints on binding affinity, yet have led to inconsistent results.[6,7]

Finally, changes in protein conformational dynamics, e.g. torsional and

librational degrees of freedom, are known to occur on ligand binding. Flexibility can again be inferred by high B-factors in an x-ray crystal structure. Nuclear magnetic resonance (NMR) structural calculations, which result in an ensemble of structures that fit the solution state experimental data, inherently capture aspects of conformational dynamics.[8] However, differences in NMR relaxation measurements upon ligand binding allow the approximation of site-specific changes in entropy across the protein, as outlined in §1.3.2.6.

1.1.3.3 Solvent enthalpy and entropy

Solvent water contributions to binding enthalpy and entropy arise from reorganisation of solvent-solvent and solvent-solute interactions upon binding. This mainly takes the form of desolvating the ligand and protein to allow their interaction.

Water molecules can form hydrogen bonds to themselves, and ligands or protein surfaces that are polar. This is enthalpically favourable, again in a distance, permittivity and orientationally dependent manner. Non-polar ligands or protein surfaces do not provide this hydrogen bonding opportunity, therefore water is believed to order at these surfaces to maximise the distance and orientation dependence of inter-solvent hydrogen bonds.[9] Though improving enthalpy, this process decreases entropy. Hence, the desolvation of non-polar ligands is enthalpically unfavourable and entropically favourable. Ligand desolvation enthalpies and entropies can potentially be experimentally measured using air-solvent partition equilibria experiments.[10] Unfortunately, these experiments are not practicable for some ligands, because sufficient volatility is required to obtain a measurable concentration in gas phase. However, an additive technique for calculating these parameters was developed using an extensive pool of published experimental data, and gives close experimental agreement for hydrocarbons at 298 K and 1 atm.[11]

Some non-polar protein surfaces are pockets whose solvating waters are not contiguous with bulk water. Consequently, there may be an insufficient number of water molecules to form an ordered but enthalpically favourable hydrogen bond network. In such cases, entropic solvation has been proposed, whereby the solvating water molecules, without any electrostatic interactions to restrict them, have higher entropy than bulk water.[12-15] Enthalpies and entropies of protein desolvation are notoriously difficult to calculate. A few theoretical and experimental estimates exist,

and these reveal no clear overall pattern, with different thermodynamics being observed dependant on the individual system.[16,17]

1.2 The Major Urinary Protein (MUP): a model system

The decomposition and perturbation approach is only tractable when a protein fulfils two main criteria. Firstly that protein or ligand structure can be perturbed, and binding is still observed. Therefore the protein must be relatively promiscuous. Secondly, the protein must be amenable to experimentation using all the biophysical techniques required to observe the various intrinsic and solvation contributions; ITC, NMR, MD and X-ray crystallography. Medically-targeted proteins, in which ligand affinity optimisation might reap a direct benefit, do not usually meet the first criterion due to their biological function usually necessitating a lack of, or only minimal, promiscuity.

The first quantitative decomposition of binding thermodynamics for a single protein-ligand interaction was achieved using the mouse major urinary protein (MUP) as a model system.[18] MUP is a 20 kDa pheromone binding protein from mouse urine that has an eight-stranded beta barrel plus alpha helix structure typical of its lipocalin superfamily. Various MUP isoforms are produced naturally in mice. The homogenous isoform produced recombinantly and used in this thesis is MUP-I, which will be referred to throughout as MUP.[19] As part of its function, MUP binds a wide range of small hydrophobic ligands in its internal hydrophobic pocket. This promiscuity is possible because interactions are dominated by weak, non-polar, non-directional interactions that appear to scale with hydrophobic surface area.[1,20] Tyr 120 is the single H-bond donor in the binding pocket. The structure of MUP and an example of some ligands so far investigated are displayed in Figures 1.2 and 1.3 respectively. MUP is amenable to all the required techniques, and therefore many of these interactions have been analysed using NMR, MD, X-ray crystallography and ITC.[1,18-26]

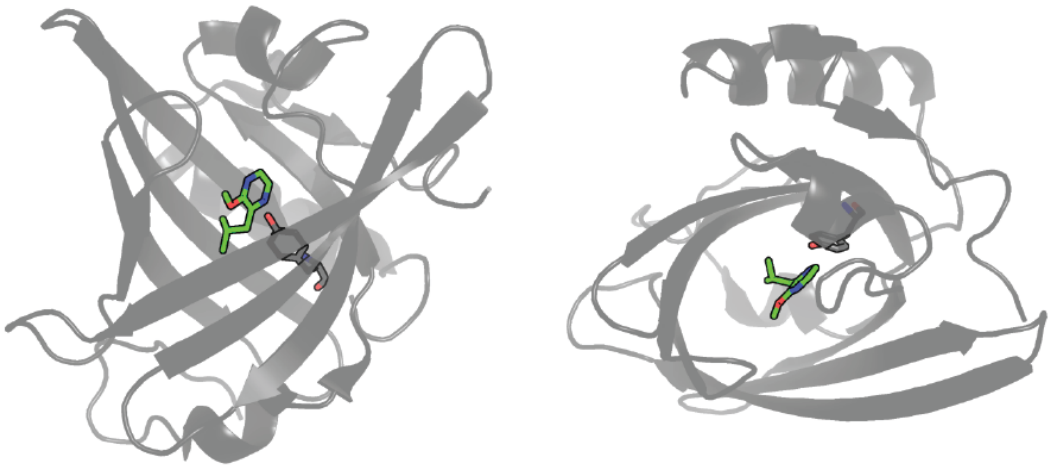


Figure 1.2 MUP structure (PDB 1QY1). IBMP is shown bound in the pocket (green and blue sticks), and the sidechain of Tyr120 is shown (grey sticks).

These analyses have revealed four key characteristics of MUP as a ligand binding system. Firstly, binding is observed for a wide range of ligand structures with variable associated affinities. The ligands shown in Figure 1.3 bind with dissociation constants that range across one order of magnitude: 0.3 to 2.9 μM.[1,23] A wider range of accessible affinities was recently demonstrated with the recent report of K_d = 32 nM, for the binding of N-phenyl-naphthylamine to MUP.[20]

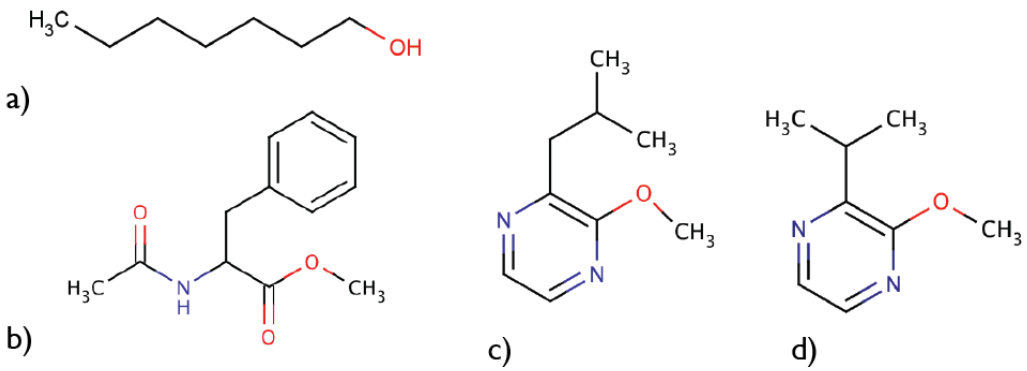


Figure 1.3 Examples of MUP ligands. a) 1-heptanol. b) N-acetylated phenylalanine methyl ester (NPOME). c) 2-methoxy-3-isobutyl pyrazine (IBMP). d) 2-methoxy-3-isopropyl pyrazine (IPMP).

A second characteristic of MUP as a ligand binding system is that interactions of the hydrophobic MUP pocket and hydrophobic ligands invert the thermodynamic signature

of the classical hydrophobic effect. The classical hydrophobic effect is believed to result from the ordering of polar solvent around non-polar groups and surfaces to maximise enthalpic H-bond interactions with neighbouring solvent molecules.[9] Desolvation of non-polar groups removes this ordering, therefore resulting in an entropically favourable process.[27] Though the hydrophobic effect is widely regarded as being dominated by entropy, the biophysical description above is still not universally accepted, due to little experimental evidence of such solvent ordering. Despite the hydrophobic nature of the MUP pocket, a non-classical enthalpy driven hydrophobic effect has been consistently observed. This is due to a sub-bulk water density, ~0.2 gm cm^{-3}, in the occluded MUP binding pocket, as observed by MD and X-ray crystallography.[21] Therefore there is minimal order to disorder gain by ejecting these molecules upon binding. Moreover, it has been suggested that MUP pocket water molecules may actually lose entropy upon release to bulk, meaning the apo pocket is 'entropically solvated'.[15,21] Incoming ligands form better enthalpic contacts with the pocket than the displaced water, resulting in enthalpically favourable binding. This was demonstrated by generating a Y120F MUP mutant which has no water in the pocket, as observed by MD and X-ray crystallography. Binding of IBMP was observed to be less enthalpically favourable with Y120F MUP than MUP, due to lack of the enthalpically favourable water ejection. Furthermore, the enthalpic non-equivalence of water-protein and ligand-protein dispersive forces as a driver for association was observed as a surface area dependent enthalpic benefit across a panel of alcohols.[1]

NMR relaxation measurements performed before and after IBMP binding revealed site-specific changes in dynamics and entropy, calculated as outlined in §1.3.2.6.[23] Overall these result in zero entropy change within error, but with decreasing flexibility in some areas being offset by increasing flexibility at others, Figure 1.4. Consequently, a mechanism was proposed whereby affinity is improved by redistribution of protein conformational entropy upon binding.[23]

The acquisition of x-ray crystal structures of mutant MUP, such as Y120F, demonstrate that the protein is very structurally resilient to mutations of binding pocket residues: RMSD values are small between mutants, with no change to overall structure even with the introduction of ionisable residues to the binding pocket (unpublished, see Figure 2.4). Throughout this thesis, RMSD values between structures are calculated using the Pymol 'align' command, with the default iterative outlier removal disabled to avoid artificially low RMSD values being returned.

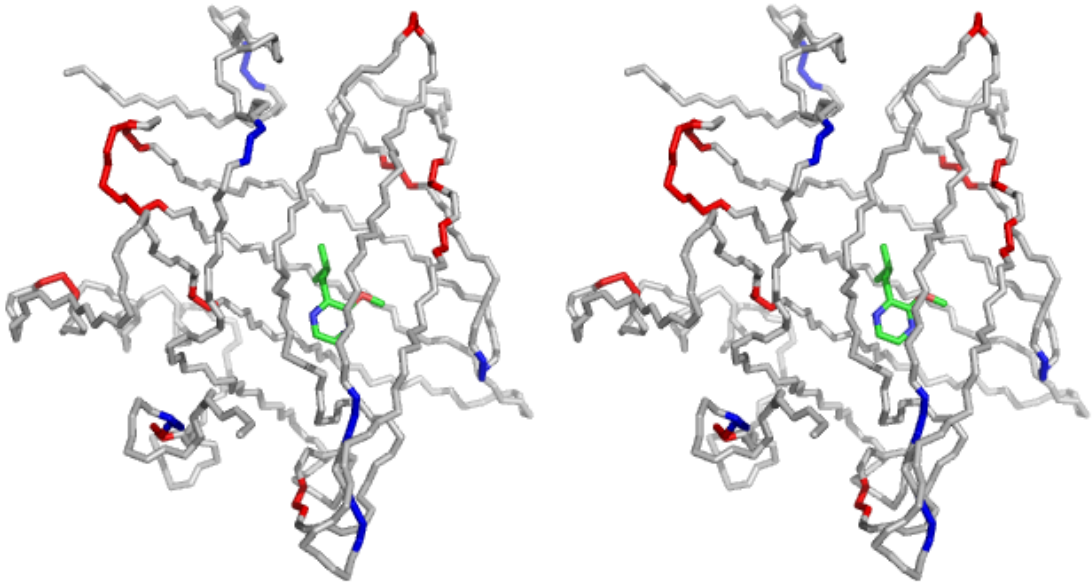


Figure 1.4 Stereo view of changes in MUP protein backbone (N-H) dynamics upon binding IBMP, as observed by NMR relaxation measurements. Reproduced from reference [28]. IBMP is shown in the pocket (green sticks). Contributions to protein conformational entropy upon binding are indicated as blue for positive and red for negative changes.

1.3 Techniques employed in this thesis

1.3.1 Isothermal titration calorimetry (ITC)

Expulsion or uptake of heat is a function of most biomolecular interactions. Direct measurement of the heat change associated with an interaction is possible by using ITC. It is the only technique that allows detection of all the thermodynamic parameters of binding: ΔG°_b, , ΔH°_b, and $T\Delta S^\circ_b$. These parameters observed by ITC represent the total contributions from all sources, and are henceforth interchangeably referred to as the observed parameters, with subscript obs. K_a, K_d, ΔG°_b and ΔG°_{obs} are henceforth used as representations of affinity.

The ITC instrument contains two cells, manufactured using an inert alloy with high thermal conductivity, whose temperatures are controlled by a feedback driven electric heater in a highly sensitive thermocouple circuit, Figure 1.5. An

adiabatic jacket surrounds the cells. The experiment takes the form of titrating precise aliquots of a ligand solution of known concentration into a protein solution of known concentration in the sample cell. The integrated power input required to return this cell to thermal equivalence with the reference cell is monitored, giving the heat change of the interaction. To ensure equilibration of the interacting solution between injections the syringe tip functions as a constantly rotating stirrer. Though the whole titration process is computer controlled, sample and instrument preparation are critical to successful experiments. Thorough degassing of solutions is necessary to minimise the chances of spontaneous bubble formation, a process that will change the heat of the system and result in aberrant data. Cleaning the sample cell between runs is also necessary to ensure no contamination from previous runs, with a strong alkali (1M NaOH) being the most potent cleaning agent available.

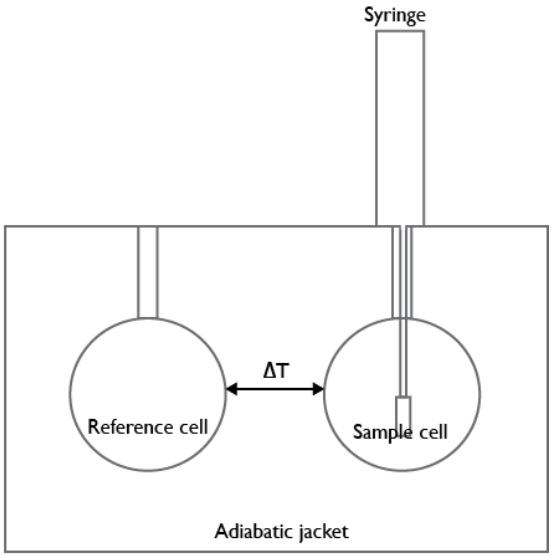


Figure 1.5 Isothermal titration calorimeter schematic. Precise aliquots of a ligand solution of known concentration are titrated into a protein solution of known concentration, and mixed by rotation of the syringe needle. Power to the sample cell is modulated to maintain equivalent temperature to the reference cell. The integral of the power is the heat change of the interaction.

The plot of molar ratio (ligand:protein) versus the integrals of the power fluctuations for each injection is the isotherm. Thermodynamic parameters of binding are extracted through least-squares fitting of an appropriate model to the isotherm. The Wiseman model[29], given in Equation 1.8 for the 1:1 binding of ligand X to protein

M, describes the following for a given injection: the change in heat, dQ, normalised to the total number of moles ligand added thus far, X_{tot}, in terms of the volume of the sample cell, V_0, and the absolute ratio of ligand and protein concentrations, X_r. M_{tot} is the total moles of protein in the sample cell.

$$X_r = X_{tot}/M_{tot}$$

Equation 1.7

$$\frac{dQ}{dX_{tot}} = \Delta H^\circ_b V_0 \left(\frac{1}{2} + \frac{1 - X_r - r}{\sqrt[2]{(1 + X_r + r)^2 - 4X_r}}\right)$$

Equation 1.8

Fitting gives the enthalpy of binding, ΔH°_b, and the value r, from which K_a is calculated according to Equation 1.9.

$$1/r = M_{tot}K_a = c$$

Equation 1.9

It is worth noting that ΔH°_b is temperature dependent, and for some interactions may be zero within the physiological and ITC accessible range, resulting in no ITC signal. Therefore for some interactions, it may be necessary to repeat the experiment at alternative temperatures before concluding that no binding occurs.

The stoichiometry of the interaction, n, is also reported, Figure 1.6. A deviation from n=1 for interactions that should have a 1:1 stoichiometry may have one of the following causes: the one site model is not appropriate; solutions are impure, containing other interacting entities; protein is unfolded, therefore the effective concentration is lower than calculated; or protein and ligand concentrations are inaccurately calculated. Therefore large, > 0.1, deviations from n=1 must be investigated to ensure that impurities or unfolding are not the cause. Presence of either would invalidate the experiments due to potential alternative interactions contributing to the observed data. It has been adequately demonstrated that MUP has a single binding site that can only accommodate one molecule.[1,21-23] The single exception was the accommodation of two pentanol molecules in the pocket, due to its small size.[1] All ITC solutions were filtered, and NMR allows estimation of both the purity of the solution and the folding state of the protein. As solutions were pure and

MUP fully folded, stoichiometric deviations can only arise due to inaccurate protein and ligand concentration measurements. Due to the critical influence of these concentrations on the extracted parameters, it is essential to ensure their accuracy, for example by empirical calculation of spectrophotometric extinction coefficients.

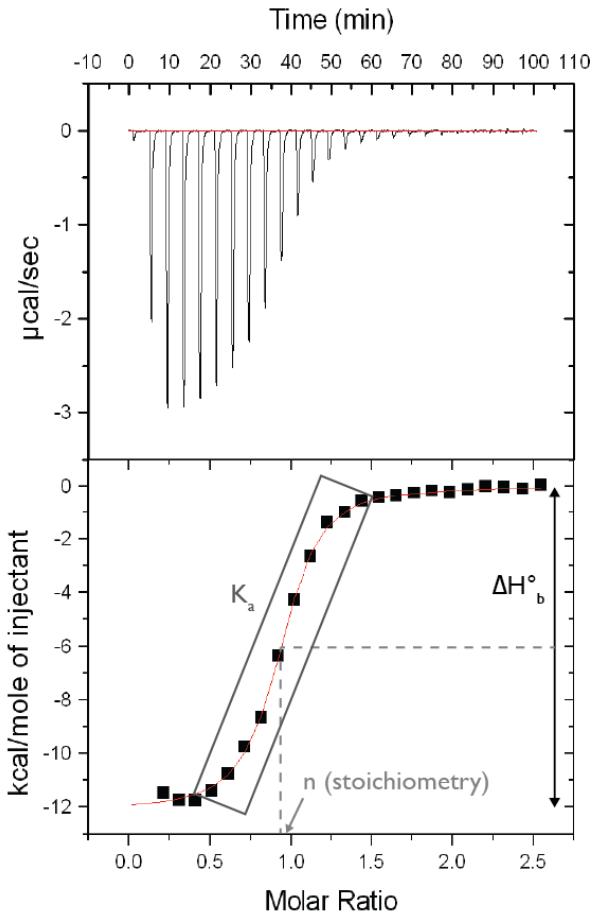


Figure 1.6 ITC isotherm and interpretation. (top) Differential power to the sample cell as a function of time. Each spike in power is due to a single titration injection, during which a finite amount of heat is released, followed by the cell temperature returning to baseline. The return to baseline, i.e. equivalent temperature to the reference cell, is achieved through the action of the thermocouple controlled feedback circuit. (bottom) Integrated peaks from the top half as black squares. The initial point is removed from the integrated data to account for equilibration of protein and ligand solutions at the syringe tip, before fitting the one-site Wiseman model (red line) to the isotherm. ΔH°_b is the distance between the two plateaus of the curve. The interaction stoichiometry (n) is observed as the molar ratio value where the fit curve is exactly halfway between the plateaus. K_a, and subsequently ΔG°_b, are defined by the curvature of the transition, captured by the r value in Equation 1.8.

The instrument was typically described as allowing measurement of interactions with K_a = 10^3 to 10^8 M^{-1}. Turnbull and Daranas showed this range can be extended to encompass lower affinity systems, down to K_a = 10^2 M^{-1}, if the ligand concentration is sufficiently higher than K_d.[30] Nonetheless, very weak interactions can be investigated using displacement ITC, wherein the binding thermodynamics of a weakly interacting molecule are obtained by observing its displacement by a stronger competitor for which the thermodynamic binding parameters have been determined.[31,32] The function used for least squares fitting of the isotherm in displacement ITC is described by Sigurskjold in reference [33], and is displayed below, wherein strong ligand, A, displaces weaker ligand, B, from protein, P.

$$Q_i = V_0 [P]_0 \left(\Delta H_A \left[x_{PA,i} - f_i x_{PA,i-1} \right] + \Delta H_B \left[x_{PB,i} - f_i x_{PB,i-1} \right] \right) + q_d$$

Equation 1.10

$[P]_0$ is protein concentration, V_0 is sample cell volume, ΔH are enthalpies of binding, and the terms in square brackets incorporate the change in mole fractions, molar ratios, c values (Equation 1.9) and sample volume with each titration. q_d is the heat of dilution upon injection of ligand.

1.3.2 Nuclear magnetic resonance (NMR)

Nuclear magnetic resonance (NMR) spectroscopy provides both temporally and spatially resolved information regarding protein structure and dynamics on timescales between picoseconds to microseconds, and with some types of experiments, even hours. This resolution, coupled with the experiments taking place in solution, contrasts with X-ray crystallography, from which higher resolution structures of entire proteins are obtained at the expense of introducing non-physiological temperatures (110 K) and crystal packing constraints on protein behaviour.

A wide range of NMR experiments has evolved for analysis of protein dynamics, each pertaining to motion averaged over different timescales. Relaxation experiments are the most prominently used, and measure protein dynamics at per-residue resolution on the ps-ns timescale, §1.3.2.6.[34] Residual dipolar couplings, which measure the relative orientation of interatomic vectors of a molecule with respect to the applied magnetic field in a partially aligned sample, are well established in structural

determination and refinement, §1.3.2.3.[35] Their use is now being extended to dynamics, and they can probe dynamics on the nanosecond-millisecond timescale, for example the observation of long-range correlated motions in allosteric mechanisms.[36] NMR experiments and derived values either measured or utilised in this thesis are summarised below.

1.3.2.1 Basics of NMR

Some atomic nuclei possess a property called spin, which is a form of angular momentum. They possess a magnetic moment due to the presence of both charge and spin. The proportionality of the spin and magnetism is defined by the gyromagnetic ratio, γ, which is a constant for a particular nuclear isotope. Therefore the nucleus has a magnetic dipole moment, meaning it both produces a magnetic field and is affected by a magnetic field. The overall spin of NMR active nuclei can be ½ or an integer multiple of ½, depending on the balance of protons and neutrons in the nucleus. Spin ½ nuclei, such as ^{1}H, have only two possible orientations for their magnetic moment, the energies of which are equivalent in the absence of a magnetic field. However, when placed into a magnetic field, the energies are non-equivalent, such that the nuclei are distributed between the two energy levels according to the Boltzmann distribution. The marginal energy difference results in a slightly larger population of nuclei in the ground state. This allows for the excitation of these nuclei into the higher energy state through the absorption of a photon of equivalent energy to the difference between states. The energy difference and therefore frequency of the absorbed photon is defined by the gyromagnetic ratio and the strength of the magnetic field at the nucleus. The electrons of an atom provide a nuclear shielding effect: they result in a difference between the applied magnetic field and that felt at each nucleus. Therefore, for each nucleus the immediate chemical environment modulates the frequency of the absorbed photon. The shift of this frequency from a reference resonance frequency is the chemical shift, the basic parameter measured by NMR spectroscopy. The chemical shift crucially provides the capacity to observe individual nuclei, and allows us to gain exquisitely sensitive information on their local chemical environment. There is typically a requirement for spin ½ nuclei in NMR such that naturally occurring ^{12}C and ^{14}N atoms in biological samples (spin 0 and spin 1 respectively) are replaced with their spin ½ stable isotopes, equivalent to ^{13}C and ^{15}N, for effective measurement of spectra.

1.3.2.2 Relaxation

When at equilibrium in a magnetic field, slightly more magnetic moments align with the field than against it, and this defines the ground state. Thus the net magnetisation vector is parallel to the field, which is described as the longitudinal bulk magnetisation. The main B_0 field is defined as the z axis, giving two orthogonal axes describing the xy plane. When excited, the vector precesses around the B_0 field at the Larmor precession frequency, defined as the product of the gyromagnetic ratio and the applied external field. The magnitude of the bulk magnetisation vector is dependent upon the concentration of spins, the external field strength and the gyromagnetic ratio. Motion of the bulk magnetisation vector away from equilibrium is achieved by applying radiofrequency pulses. The vector can be detected when moved away from equilibrium onto the xy plane. The process by which the bulk magnetisation vector returns to its equilibrium value is termed relaxation.

There are two major forms of relaxation: longitudinal and transverse. Longitudinal relaxation, T_1, occurs when spins excited onto the z axis return to equilibrium due to vibrations and rotations within the sample, which result in a changing magnetic field local at a nucleus. A nuclear magnetic moment will relax back to equilibrium if components of the local field are equivalent to to its Larmor precession frequency. Transverse relaxation, T_2, arises from any process that decreases the phase coherence of the excited spins in the xy plane.

1.3.2.3 Couplings

Interactions between groups of NMR active nuclei due to their spin state manifest as splittings of the peaks in NMR spectra, and are known as scalar couplings. The value of a coupling is defined in Hz.

Indirect interactions between bonded nuclei, which are mediated by the bonding electrons, are known as scalar or J-couplings. The values of J-couplings are proportional to the product of the gyromagnetic ratios of the nuclei involved. Couplings are well defined for bonds, and are used to transfer magnetisation selectively between nuclei in many solution state biological NMR experiments, underpinning two dimensional spectra such as the heteronuclear single quantum correlation (HSQC) experiment. In a HSQC experiment, proton magnetisation is

selectively transferred to a directly bonded heteronucleus, using the one bond J-coupling. After the chemical shift is evolved, the magnetisation is finally transferred back again to the proton for detection. Overall this allows signals overlapped in the proton dimension to be resolved in the heteronuclear dimension.

Easily measurable J-couplings typically extend over one to three bonds, and in case of the 3-bond ^{3}J-coupling the value is dependent on the torsional angle of the bonds. The ^{3}J-coupling can therefore be used to define local geometry. Karplus described an empirical relationship for the torsion angle dependence of the ^{3}J-coupling, Equation 1.11. J(Φ) is the coupling and Φ is the torsional angle. A, B and C are empirical parameters that depend on the atoms involved.[37] Accurate parameterisation of the relationship for a given system is therefore necessary: multiple efforts have been made to perform this for protein systems.[38,39]

$$J(\Phi) = A\cos^2\Phi + B\cos\Phi + C$$

Equation 1.11

Another form of coupling is the through-space interaction of magnetic dipoles, referred to as dipolar coupling. The coupling, D, is dependent upon the magnetic properties of the two nuclei, κ, their distance, r, and the orientation of the internuclear vector with respect to the external magnetic field, θ.

$$D = \frac{\kappa}{r^3}\left(\cos^2\theta - \frac{1}{3}\right)$$

Equation 1.12

κ is a function of the gyromagnetic ratios of the two nuclei, γ, and some fundamental constants.

$$\kappa = -\frac{3}{8\pi^2}\gamma_I\gamma_S\mu_0\hbar$$

Equation 1.13

In solution, molecules tumble isotropically due to Brownian motion and ($\cos^2\theta$ - 1/3) averages to zero, so the overall coupling is lost. The distance dependent effect of

dipolar coupling is exploited as the basis of the nuclear Overhauser effect (NOE), §1.3.2.5.

When molecules are partially aligned with respect to the magnetic field, i.e. not tumbling isotropically, part of the angular dependence of the dipolar coupling is again observable as an addition to the observed J-coupling. When measuring couplings for directly bonded atoms, the distance dependence is removed due to the average bond length, leaving the primary variable as θ. Thus the difference between couplings in isotropic and partially aligned samples, namely the residual dipolar coupling, rdc, contains an angular dependence of the internuclear bond vector with respect to average alignment with the magnetic field, which contains structural and dynamic information.[40] Rdc values are averaged over the nanosecond to millisecond timeframe. A principal obstacle to performing rdc experiments is finding an alignment medium for a given system. Fortunately several systems have been devised that align with the magnetic field and induce partial alignment through steric restriction or electrostatic interactions with the protein, such as polyacrylamide gels, bicelles or bacteriophage solutions.[35] Five rdc values from a known rigid unit containing non-colinear vectors spanning the three dimensional space can be used to define an alignment tensor. The alignment tensor is a description of the average alignment with the magnetic field of the molecular coordinate frame, such as that found in a pdb file.[41] The alignment tensor is described by 5 parameters, three are Euler angles, which define the rotation from the molecular frame to the orthogonal tensor frame, and the other two parameters describe the degree of alignment: A_a is the magnitude of alignment along the z axis, and A_r is the rhombicity of the xy axes. The overall RDC can be described by Equation 1.14.[42]

$$D_{ij} = A_a \left(3\cos^2\theta - 1\right) + A_r \frac{3}{2} \left(\sin^2\theta \cos 2\Phi\right)$$

Equation 1.14

A_a and A_r are the axial and rhombic components of the alignment tensor. θ and Φ are angles relating the internuclear vector to the alignment tensor.

1.3.2.4 COSY and TOCSY

Correlation spectroscopy (COSY) is a technique which generates homonuclear or heteronuclear spectra of 2 or more dimensions. In homonuclear correlation spectra the diagonal peaks are those that appear in a 1d spectrum, whereas cross peaks represent J-couplings between nuclei, and are therefore observed for nuclei within typically 2-3 bonds of each other. The horizontal and vertical intersections of crosspeaks with the diagonal reveal the shifts of coupled spins. These spectra are commonly used to assign the chemical structure of molecules. Total correlation spectroscopy (TOCSY) gives spectra that are superset of COSY, but transfers magnetisation over all bonds within a given spin system. In crowded spectra, this can help to reveal the peaks belonging to a particular spin system and improve assignment of that spin system.

For a protein, HCCH TOCSY and ^{13}C TOCSY HSQC experiments involve magnetisation transfer from all sidechain protons to sidechain carbons in a ^{13}C labelled sample. The signals are then transferred between connected ^{13}C nuclei, and finally are transferred back to sidechain protons or carbons for detection. Because all hydrogens attached to carbons in the same side chain appear as cross peaks, these experiments allow assignment of amino acid sidechains.

1.3.2.5 NOESY

When there is a dipolar interaction between two spins close in space, as described in §1.3.2.3, they can exchange magnetisation. This transfer is referred to as cross-relaxation, and forms the basis of the NOE, which has an r^{-6} dependence on internuclear distance, assuming that the dipolar coupling is the main mechanism of relaxation. Protons are primarily used to observe NOEs, because their larger gyromagnetic ratio and magnetic moment leads to a more effective magnetisation transfer. Typically for most proteins, if r is less than about 6 Å the cross-relaxation can be observed. As per COSY and TOCSY, in a NOESY experiment cross-peaks appear in the spectrum, in this case representing a through-space connection. In this thesis, ‘NOE’ is henceforth used interchangeably with ‘NOE crosspeak’. The relative strength of NOE crosspeaks therefore approximately indicates the relative internuclear

distances. Combining HSQC and NOESY experiments correlates the crosspeaks to assigned resonances, improving the structural interpretation of NOEs.

NOEs require a sufficient mixing time to build up and become observable. However, at longer mixing times, multi spin interactions, so called spin diffusion, result in NOEs being observed between spatially-distant protons due to multi-step magnetisation transfer through networks of close protons. This generates crosspeaks that cannot be directly interpreted in terms of internuclear distance. Spin diffusion occurs faster for larger molecules, due to an inverse dependence on the global correlation time. A mixing time of 120 milliseconds is typically used for proteins as a compromise, to minimise spin diffusion whilst maximising useable NOE intensity.

1.3.2.6 Relaxation-derived dynamics

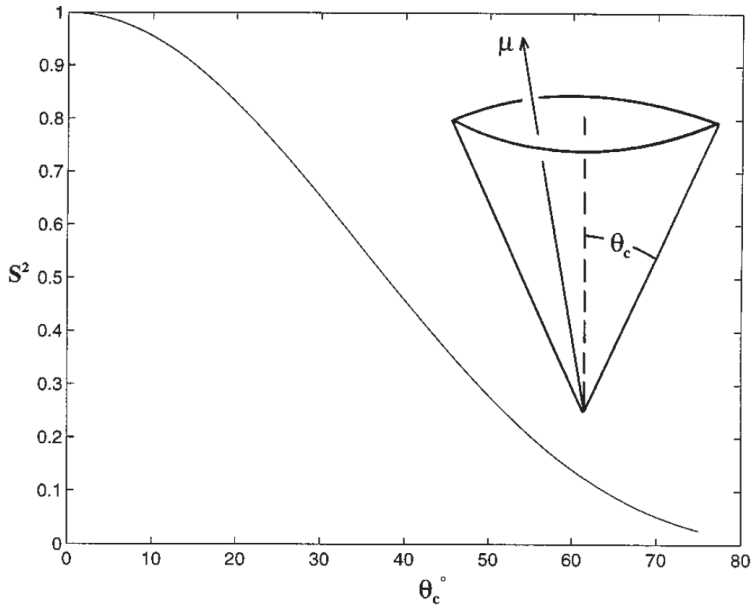


Figure 1.7 The 'diffusion in a cone' motional model commonly used for physically interpreting backbone amide S^2 parameters obtained by model-free analysis of NMR relaxation data. Here the S^2 describes an effective 'cone of diffusion'. As S^2 decreases bond vector motion increases, represented by a larger cone of diffusion. Figure taken from reference [34].

'Model free' analysis of NMR relaxation parameters, so called because it only specifies a timescale and amplitude, is the main source of NMR-derived information about site-specific dynamics on the ps-ns timescale.[43,44] Relaxation parameters can be acquired for backbone amide or sidechain methyl bond vectors. The analysis results in a number of parameters for each site at which relaxation parameters were measured. The most commonly cited is the square of the generalised order parameter, S^2, which can be related to the degree of motional restriction and therefore bond vector dynamics. Multiple physical models are available for interpretation of the S^2. A model used commonly for backbone amide S^2 values is that of diffusion in a cone, Figure 1.7. As the S^2 decreases, the effective 'cone of diffusion' of the bond vector increases, represented by θ_c°. Therefore an S^2 of 1 describes a rigid bond vector attached to the molecular frame, whereas a value of 0 would describe a bond vector tumbling isotropically.

As the distribution of bond vector orientations described by the order parameter will be related to disorder and therefore entropy, a method for obtaining an approximate upper estimate of conformational entropy change between two states was devised by Yang and Kay.[45] This is calculated using the generalised order parameter before and after binding, S_{apo} and S_{holo} respectively, and the gas constant R.

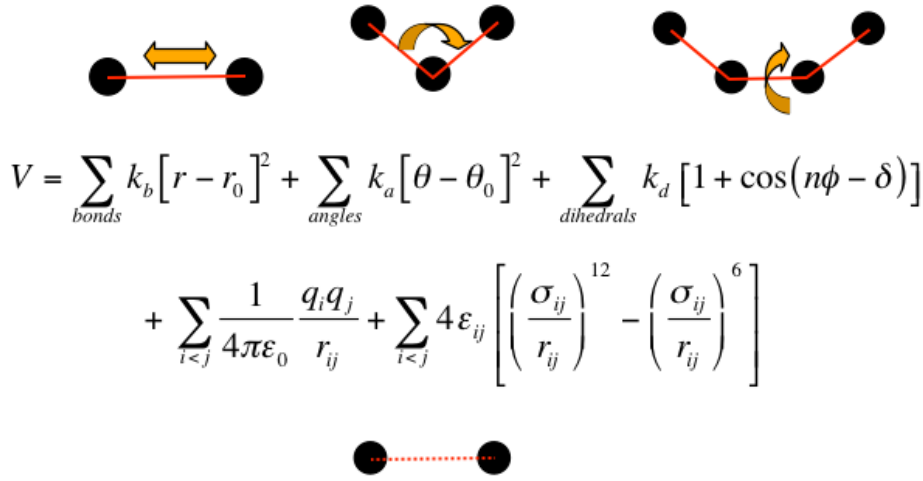
$$S_p = \ln\left[\frac{\left(3-\sqrt{(1+8S_{holo})}\right)}{\left(3-\sqrt{(1+8S_{apo})}\right)}\right]R$$

Equation 1.14

This approach currently offers the best and most resolved method for measuring conformational fluctuations/disorder in proteins at an atomistic level, and therefore approximating their associated entropy. The parameters apply only to the bond vector analysed, which assumes independent motion for each vector, and captures motions only on the ps-ns timescale. Nonetheless, the common use of these values to represent flexibility changes and calculate entropy changes is supported by examples of close agreement with independent measurements.[46]

1.3.3 Molecular Dynamics (MD)

Molecular dynamics (MD) solves Newton's equations of motion for an atomic resolution system, generating a time evolution of a system as a sequence of atomic positions and velocities termed the trajectory. The energy of the system configuration sampled at each step is calculated using a potential energy function.



$$V = \sum_{bonds} k_b\left[r - r_0\right]^2 + \sum_{angles} k_a\left[\theta - \theta_0\right]^2 + \sum_{dihedrals} k_d\left[1 + \cos\left(n\phi - \delta\right)\right]$$

$$+ \sum_{i<j} \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} + \sum_{i<j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

Figure 1.8 CHARMM MD potential energy function. The energy of the system is calculated using bonded and non-bonded terms. k values are force constants and subscript 0 refers to equilibrium values generated by parameterisation. All terms with equilibrium values are harmonic functions. Figure provided by Dr Emanuele Paci (personal communication).

The potential energy function represents a compromise between the speed and accuracy of the calculation, as terms that could be accurately calculated quantum mechanically at greater computational expense are approximated into classical forms. Contributions from both bonded and nonbonded terms are summed over a number of terms to calculate the potential energy. These terms include bond lengths and angles, dihedral angles, and van der Waals and electrostatic interactions. For nonbonded terms, the first is a Coulombic term for point charges, and the second a Lennard Jones term for calculation of vdW attraction and atomic repulsion. The various K values are force constants, and the use of subscript 0 is to indicate the equilibrium value of the

relevant parameter. All bonded terms are evaluated as harmonic functions, with the energy minimum at the equilibrium value.

Equilibrium values are contained in the parameter files generated for each molecule. The generation of parameter files, namely parameterisation, can be a lengthy process involving quantum mechanical calculations: fortunately efforts are being made to parameterise molecules on a faster timescale.[47] Combined, the potential energy function and parameters are referred to as the forcefield.

MD trajectories are maximally temporally and spatially resolved, presenting an ideal method for investigating the structure and dynamics of biomolecules. Almost any parameter or experimental observable is potentially calculable from a trajectory using MD software itself, or external trajectory analysis tools such as wordom for CHARMM[48]. Some observables that can be calculated are NMR S^2, radius of gyration, average structures, and rmsd or rmsf values between frames of the trajectory. Assuming sufficent accuracy of the forcefield, as simulation length tends toward the experimental timescale an increasingly representative set of system configurations should be sampled. Accordingly, extending the practicable limit of simulation timescales has received much attention.[49]

However, the forcefield is an approximation, optimised for the native states of proteins. The requirement for accurate sampling of non-native states, for example in protein folding simulations, resulted in recent modifications of the backbone torsional energy term in both the major MD packages, CHARMM and AMBER.[50-53] Further approximations are available that reduce the number of particles in the system and thus the computational cost of MD. These include implicit solvent models that replace the need to include explicit solvent atoms by representing solvent as a continuous medium,[54] and coarse-grained protein models wherein entire amino acid residues are represented as single particles with averaged characteristics.[55] Therefore MD of all types contain degrees of approximation, and for every new purpose, new approximation or new timescale, need to be corroborated against experimental measurements to be used with confidence.

Fortunately, average parameters extracted from short timescale all-atom MD have been robustly corroborated by NMR, the only experimental technique that can measure at atomic resolution on timescales easily accessible by MD, i.e. picosecond-nanosecond.[50]

1.4 Aims and scope of this thesis

Though techniques exist for the computational estimation of ΔH°_{b} to a reasonable degree[56], a better understanding of dynamic entropic contributions to binding thermodynamics is required for truly rational manipulation and optimisation of interactions.

This thesis addresses four important questions regarding the dynamics and thermodynamics of protein-ligand interactions, using MUP as a model system and employing a wide range of biophysical techniques. Some work directly utilises the perturbation-decomposition approach described herein. Other work capitalises on avenues of inquiry that have arisen due to the wealth of data regarding protein-ligand interactions in MUP.

Chapter 2 is an investigation into the proposed entropic solvation of the MUP binding pocket, and the possibility of engineering a new binding profile through manipulation of sidechains and solvation in the binding pocket. Chapter 3 is the first systematic assessment of the widely predicted and presumed benefits of minimising ligand conformational entropy loss by removing and restricting ligand bonds. The effects of such modifications on intrinsic entropy are considered across a wide panel of ligands. Chapter 4 investigates the biggest entropic loss, that of ligand rotational and translational entropy, an under-investigated question addressed for the first time using a combination of NMR approaches to assess MD predictions of significant residual translation and rotational motion when bound. Finally, Chapter 5 constitutes a preliminary investigation into a potentially promising novel technique for probing site-specific changes in protein dynamics upon ligand binding.

Chapter 2

Protein Solvation

This chapter is divided into two parts, a and b. a is an investigation into the proposed entropic solvation of the MUP binding pocket. b is an investigation into the possibility of engineering a new binding profile through manipulation of sidechains and solvation in the binding pocket. The two sections are summarised together in §2.4.

2a.1 Introduction

2a.1.1 A slippery understanding of protein desolvation contributions to binding thermodynamics

Almost all biomolecular interactions take place in an aqueous environment, and the contribution of water to molecular interactions is critical. In fact their contribution can be the thermodynamic driver for a particular interaction or the folding of a protein, as described by the classical hydrophobic effect, §1.2.

Estimation of protein desolvation thermodynamic contributions to biomolecular interactions is therefore critical. Despite many theoretical predictions of the cost of trapping/releasing waters at binding interfaces, most notably those of Dunitz, who calculated entropic limits of 0 to 29 kJ mol^{-1} for release of a single water molecule to bulk (and free energy limits of 0 to ~8 kJ mol^{-1})[57], little thermodynamically-detailed experimental data are available.[16] MD and NMR studies have demonstrated that changes in the size, shape and chemical nature of a protein's binding cavity greatly alter its solvation thermodynamics. Denisov *et al* used NMR relaxation-dispersion experiments to estimate the entropy of seven bound water molecules in a single protein pocket. The estimates span a wide range, suggesting that solvation of non-polar pockets may be entropically driven.[12] This is further supported by other NMR experiments, which suggest large hydrophobic cavities have crystallographically unobservable waters that may have significant dynamics[13]. Furthermore, MD studies have shown the protein desolvation, though always accompanied by enthalpic benefits, can also result in an entropic penalty.[14,15,17]

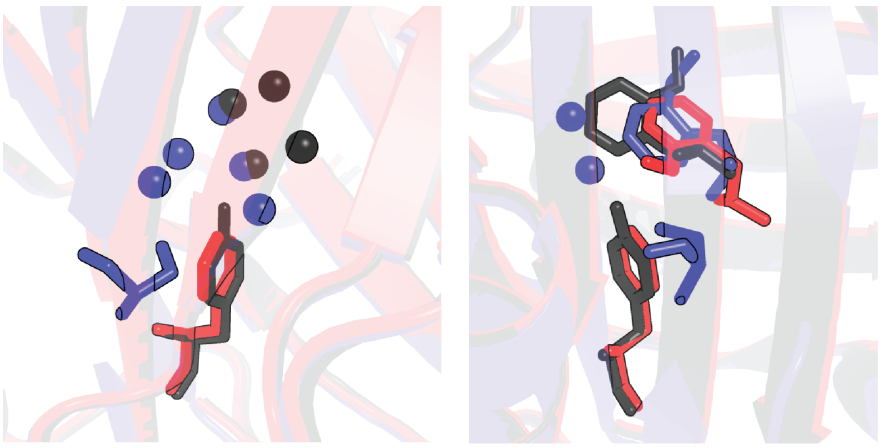
2a.1.2 The elusive solvation thermodynamics of MUP's sub-optimally hydrated and large non-polar cavity

A non-classical hydrophobic effect is observed in MUP interactions.[1,21] This has been attributed to very low density solvation of the unbound MUP pocket as observed by high residency waters in crystal structures: compared to the standard hydrophobic effect model, this results in reduced entropy gain and reduced enthalpy loss from pocket desolvation.

Relying on changes in crystallographically observed water to interpret binding thermodynamics is problematic. Firstly, structures must be resolved < 2 Å to ensure observation of water. Secondly, attributive error can arise during structure generation: water molecules may be incorrectly modelled into experimental electron density. Thirdly, as mentioned in §2a.1.1, crystallographically unobservable water may also play a significant thermodynamic role. The first issue can be resolved by using crystal structures of sufficiently high resolution. The second and third issues can be addressed by employing independent techniques to support the crystallographically derived observations.

Experimental quantitation of MUP pocket desolvation upon ligand binding was recently attempted using Quartz Crystal Microbalance (QCM) experiments performed by Miss Julie Roy at the University of Nottingham. QCM calculates the change in mass of a substrate deposited upon the surface of a resonating quartz crystal, using the change in resonant frequency upon binding. Though QCM is not established as a technique for distinguishing between masses of proteins with differential water content, it is sufficiently sensitive, Figure 2.1c. The mass change was calculated for three versions of the protein, MUP, Y120F MUP and A103S MUP, binding to IBMP. Apo and holo crystal structures were obtained at below 2 Å resolution, Figure 2.1. These two MUP mutants were generated to adjust the number of crystallographically observable water in the apo binding pocket from the four observed for MUP. Y120F removes the single H-bond donor from the pocket, resulting in no water being crystallographically observable in the pocket.[21] A103S introduces an additional H-bond donor, resulting in an additional water molecule in the binding pocket being crystallographically observable, unpublished. The changes in mass upon IBMP association from QCM, when attributed to water molecules, agree with the crystallographic observations other than for the MUP-IBMP interaction, Figure 2.1.

These observations provide support to using the patterns of water observed crystallographically to assess protein desolvation upon ligand binding.



a) b)

c)

# of water displaced	MUP	A103S MUP	Y120F MUP
Crystallography	4 (4→0)	3 (5→2)	0 (0→0)
QCM	1.9 ± 0.3	3.1 ± 0.4	0.0 ± 0.5

Figure 2.1 MUP (grey), A103S MUP (blue) and Y120F MUP (red) binding IBMP. A) and b) are crystal structures of the apo and holo protein respectively, showing the sidechains of residues 103 and 120. c) changes in pocket water upon binding: comparison of crystallography and QCM (personal communication, Julie Roy).

MD analysis of water occupancy of the MUP pocket was also performed. Four water molecules were observed by MD, as observed crystallographically, corresponding to a density one fifth that of bulk, ~0.2 g cm^{-3}.[21] In conjunction with enthalpic data, these MD results supported "the absence of ordered water in the binding site", suggesting dynamic water may be present.[21] Therefore these data corroborate the crystallographic water, yet also indicate dynamic, crystallographically unobservable, water. The presence of two such water molecules in the MUP-IBMP complex may explain the QCM-crystallography discrepancy for this interaction. No experimental approach yet exists to accurately quantify the thermodynamics of dynamic water molecules. An assumption that the change in their contribution upon binding is zero or minimal remains a caveat.

Thermodynamic decomposition of MUP makes this assumption, and includes only the change in crystallographically observable water molecules in analysis of differential protein solvation upon ligand binding. Previous decomposition estimates of the thermodynamic cost of ejecting a water molecule from the MUP pocket upon ligand binding are shown in Table 2.1, with their associated caveats.

Given the many caveats, no reliable estimate of MUP desolvation entropy has yet been made. It has been assumed that the release of water from the cavity upon ligand binding would be accompanied by an entropic benefit, from the classical understanding that water has less entropy in the bound than bulk phase. The studies mentioned in §2a.1.1 challenge this understanding: one of these studies addresses MUP directly, proposing entropically unfavourable desolvation.[15] Syme recently proposed the same as a possible source of the negative heat capacity change associated with MUP interactions.[28]

A better estimate of MUP desolvation entropy would answer these proposals and enable comprehensive thermodynamic decomposition in MUP.

Interaction	$T\Delta S°$ (kJ/mol)	Caveats / not direct measurement
MUP-IBMP *vs* A103S MUP-IBMP[58]	-5.8	Interactions differ by one water ejected. Assumes no other differences between two interactions
MUP-IPMP[18]	0.1	Entropy assumes ligand is 'frozen' upon binding. Enthalpy assumes solute:solute enthalpy extrapolated from alcohol work[1]

Table 2.1 Estimates of the entropic cost of ejecting a single water molecule from the MUP pocket, with caveats. The second entry was reported as a value for the loss of four water molecules, which was divided by four to generate the value in this table.

2a.1.3 Opportunity to measure the entropic cost of ejecting a single water molecule from the MUP pocket

Crystal structures (unpublished) of N-acetylated phenylalanine methyl ester (NPOME) bound to MUP and A103S MUP were solved to a resolution ≤ 2 Å. Their analysis revealed an opportunity to obtain a better-estimated value for the ejection of water

from the MUP pocket, Figure 2.2. The protein, ligand and solvent in the two complexes overlay almost exactly, RMSD = 0.37 Å, containing the same number of binding pocket water molecules. However, because the apo protein pockets contain a different number of water, the difference between the two interactions has two contributions: i) the ejection of one extra water molecule in A103S MUP-NPOME and ii) a potential change in S103 sidechain entropy upon binding. In decomposition terms, the ligand desolvation terms cancel, and the intrinsic terms almost cancel. Obtaining and comparing ITC data for the two interactions will reveal the binding contribution of these two factors.

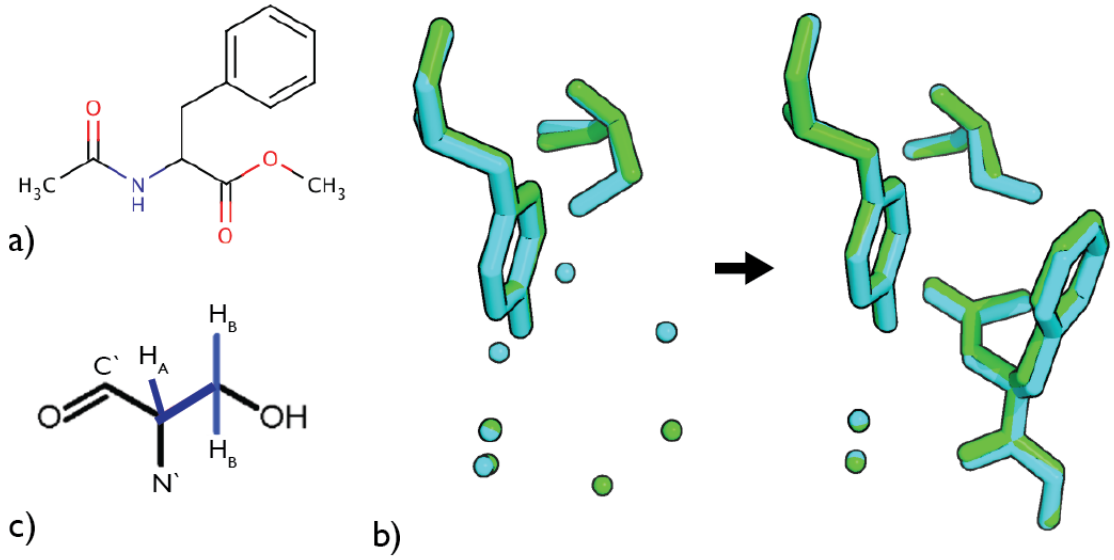


Figure 2.2 MUP (green) and A103S MUP (cyan) bound to NPOME. a) structure of NPOME. b) pocket before (apo) and after (holo) ligand binding, sidechain shown are Tyr120 and residue 103. c) serine schematic: N'-H_β, C'-H_β, and H_α-H_β ^3J couplings are torsion-angle dependent, revealing rotamer populations.

The A→S mutation involves breaking methyl group symmetry and introducing three distinguishable sidechain conformations, potentially with a resulting entropic change upon ligand binding. The crystal structure shows the S103 hydroxyl has a different average orientation before and after ligand binding, Figure 2.2b. However, this does not reveal whether rotamer dynamics and thus entropy, represented by the relative populations of rotamers, change. Experimentally measuring the relative populations of

the rotameric states before and after binding could permit an entropy estimate. This could then be subtracted from the ITC observed differences, resulting in a more rigorous estimate for the entropic cost of water ejection from the MUP pocket. An assumption of zero entropy change can be made if the relative populations of S103 rotameric states are unaffected by ligand binding. These populations can be estimated experimentally by ascertaining sidechain torsional angles using amino-acid sidechain parameterised Karplus relationships[39] with Ser103 N'-H_β, C'-H_β, and H_α-H_β ^3J couplings measured for a ^{13}C-^{15}N labelled A103S MUP sample, Figure 2.2c.

2a.1.4 Work undertaken

A better estimate of the entropic cost of water ejection from the MUP binding site will aid future decomposition. It will also address the possibility of entropic solvation, i.e. bound waters having higher entropy than bulk water. This was sought using isothermal titration calorimetry to observe the difference in binding thermodynamics of MUP-NPOME and A103S MUP-NPOME. To investigate and account for potential changes in rotamer sampling of S103 upon ligand binding, NMR experiments to measure N'-H_β, C'-H_β, and H_α-H_β ^3J couplings for a ^{13}C-^{15}N labelled A103S MUP sample were attempted. Samples for QCM were provided to Miss Julie Roy. Extinction coefficients were calculated for MUP, A103S MUP and the NPOME ligand, to ensure accurate ITC-derived data.

2a.2 Materials and Methods

2a.2.1 Protein Expression and Purification

2a.2.1.1 Standard procedure for MUP

E.coli strain SG13009 containing the MUP gene, with a hexa-histidine tag, had been generated previously and stored in a glycerol broth.[23] A single colony was picked from an agar plate containing 1 mg/mL carbenicillin and grown overnight at 37 °C with vortexing in 100 mL LB medium containing 1 mg/mL carbenicillin. 10 mL of this was

used to seed 1 L of LB medium (in a 2 L flask), and the culture left to grow at 37 °C with shaking at 220 rpm. Once the OD_{600} of the culture reached 0.6 to 0.8, isopropyl-thiogalactosidase (IPTG) was added to 1 mM to induce gene expression. After 6 hours incubation at 37 °C with shaking at 220 rpm, the cultures were centrifuged at 4 °C, 5000 g for 10 minutes to harvest the cells, which were frozen overnight. An equivalent process was also performed using a 30 L fermenter facility to provide greater quantities of protein. The cell pellet was resuspended in 5 mL phosphate-buffered saline (PBS, pH 7.4), containing 0.16 mg/mL lysozyme, per gram cell pellet and shaken at room temperature for 20 minutes. Deoxycholic acid was subsequently added (4 mg/g pellet) and the solution was incubated at 37 °C for 30 minutes, before addition of DNase I (10 μg/g pellet) and $MgCl^2$ (to 5 mM) and further incubation at 37 °C for 20 minutes with vortexing. The solution was centrifuged at 4 °C and 8400 g for 20 minutes. The MUP was separated first by Ni-NTA affinity chromatography (Qiagen) and then size exclusion chromatography using Sephacryl or Superdex medium (Sigma and GE Healthcare respectively). An ethanol precipitation step was included to remove any endogenous ligands in the binding pocket by adding two volumes ice cold ethanol to 1 volume MUP (PBS pH 7.4) and incubating for 2 hours at 4 °C, before centrifuging at 500 g. The pellet was lyophilised and then dialysed against water overnight before further lyophilisation to generate a protein stock.

2a.2.1.2 Modification of procedure for ^{13}C-^{15}N labelled protein.

The same procedure as described in §2a.2.1.1 was used apart from the following changes. M9 minimal medium containing ^{15}N ammonium chloride and ^{13}C glucose (Cambridge Isotopes) was substituted for LB medium. Due to lower populations in minimal medium compared to LB, lower induction limits and longer incubation/induction times were used.

2a.2.2 Extinction coefficient (ε) calculations

2a.2.2.1 MUP

5.05 mg of lyophilised MUP, M_W = 20359 Da, was reconstituted into 2 mL PBS (pH 7.4) to an approximate concentration of 124 μM. The sample was split into 2 x 1 mL aliquots. One was diluted with PBS pH 7.4, the other to 6 M Guanidine Hydrochloride (GdnHCl). Serial dilutions were performed for each and a small amount of detergent (Brij 35, Aldrich) added to each sample, to avoid cuvette wall adhesion. Neither GdnHCl nor the detergent absorb at 280 nm. The resulting samples were measured using two independent spectrophotometers, a CE1021 (Cecil Instruments) and a Genesys 6 (Thermo Spectronic). This is because the CE1021 has been historically used for MUP measurements in the lab, yet gives different readings to the more modern Genesys 6.

In order to obtain an accurate sample concentration, 200 μL of the most concentrated MUP PBS sample was sent for quantitative amino acid analysis. In this technique, the sample is boiled in HCl under vacuum for 24 hours and the resulting hydrolysed amino acids are reacted with ninhydrin, separated using ion exchange and measured using UV/VIS absorbance (ALTA Biosciences, University of Birmingham, UK).[59] This type of analysis does not reliably measure the concentration of Cys, Gln, Asn Trp, Ser or Thr. It was assumed that the sample sent for analysis was uncontaminated, which was supported by the observation that the distribution of reliable amino acid concentrations closely matched their percentage composition in MUP. The concentration of the sample sent for analysis was calculated as the mean of the back-calculated concentration estimates from each reliable amino acid. These agreed with the expected concentration from weight only, further supporting the purity of the sample. Accurate concentrations of each sample in the series were calculated, accounting for the dilution by detergent. The discrepancy of the accurate concentrations and those calculated using the original extinction coefficient (10650 M^{-1} cm^{-1}) was used to generate an improved ε_{280}. ε_{280} values for both samples and spectrophotometers are shown in Table 2.2.

For the Cecil CE1021 spectrophotometer, subsequently used for all MUP quantifications, the original ε_{280} consistently calculated concentrations 0.755 that of the accurate concentration, in agreement with previously consistently observed ITC stoichiometries. Therefore the improved ε_{280} is 14105 M^{-1} cm^{-1} (=10650/0.755).

Spectrophotometric analysis of the GdnHCl (unfolded) sample using the Genesys 6 generated an ε_{280} equivalent to that calculated using the online ProtParam tool (http://www.expasy.ch/tools/protparam.html)[60,61]: 11538 M^{-1} cm^{-1} empirically compared to 11522 M^{-1} cm^{-1} computationally.

The agreement of the unfolded protein in 6 M GdnHCl with the calculated value, Table 2.2, provides a potential route for faster calculation of ε_{280nm} for MUP mutants. No MUP mutants change residues that absorb at 280nm, meaning that an ε_{280nm} = 11538 M^{-1} cm^{-1} on a sample in 6 M GdnHCl using the Genesys 6 spectrophotomer can be assumed to report the accurate concentration of protein in the sample, assuming no degradation of the sample. If an equivalent PBS pH 7.4 sample is also prepared, this concentration can be used to back-calculate an ε_{280nm} for the Cecil CE1021.

	Spectrophotometer	
Buffer	Cecil CE1021	Genesys 6
PBS, pH 7.4	14105 ± 209 (1.5%)	12227 ± 71 (0.5%)
6 M GdnHCl	13067 ± 16 (0.15%)	11538 ± 63 (0.5%)

Table 2.2 MUP extinction coefficients calculated from amino acid analysis, §2a.2.2.1. The ProtParam reported a value of 11522 M^{-1} cm^{-1}.

2a.2.2.2 N-acetylated Phenylalanine methyl ester (NPOME)

A spectrophotometric UV absorbance measurement (Genesys 6 spectrophotometer) revealed 257.5 nm as an absorbance maximum for NPOME. Sonication is necessary to dissolve crystalline NPOME. 3.8 mg NPOME (Sigma-Aldrich Rare Chemicals Library), M_W = 221 Da, was dissolved in 50 mL PBS (pH 7.4) by sonication. In order to track the dissolving of NPOME with sonication, $A_{257.5}$ was measured in 10-minute intervals using the Genesys 6 spectrophotometer until a plateau was reached after 30 minutes ($A_{257.5}$ = 0.042). This was taken to represent all ligand having dissolved and therefore a concentration of 343.89 μM, resulting in a calculated $A_{257.5}$ of 122 M^{-1} cm^{-1}.

2a.2.3 Isothermal titration calorimetry

2a.2.3.1 MUP or A103S MUP vs NPOME

ITC experiments were performed on a MicroCal VP-ITC unit with a cell volume of 1.4109 mL at 298 K. MUP solution was prepared from freeze-dried protein in filtered and degassed PBS pH 7.4 containing 1 mM sodium azide. This solution was dialysed overnight at 4 °C, using 7 kDa molecular weight cutoff Snakeskin dialysis tubing (Thermo Scientific). Post-dialysis the MUP solution was filtered using a SS-10 ES10 mL syringe (Terumo) and 0.2 μM Minisart filter (Sartorius Stedim). MUP concentration was calculated using the CE1021 spectrophotometer (Cecil Instruments) using the extinction coefficient determined in §2a.2.2.1. All MUP solutions were stored at 4 °C, and underwent 30 minutes of pre-experiment degassing at approximately 23 °C under vacuum in the ThermoVac temperature-controlled vacuum chamber (MicroCal). The PBS dialysate was re-filtered using a 0.2 μM cellulose membrane filter (Whatman) and thoroughly degassed before being used to make ligand solutions. Crystalline NPOME (Sigma-Aldrich Rare Chemicals Library) was dissolved via sonication and solution concentration was measured on the Genesys 6 spectrophotometer (Thermo Spectronic) using the extinction coefficient calculated in §2a.2.2.2.

Experiments comprised a single injection of 2 μL after a 60 s initial delay, followed by 50 injections of 5 μL, at 240 s intervals using a 310 rpm stirring speed. The cell was cleaned between each run with 1 M NaOH, then extensively with both water containing 1 mM sodium azide and PBS dialysate. Between runs of different ligands the syringe was extensively flushed using both these solutions, but only PBS dialysate was used between titrations with the same ligand.

Data were analysed in Origin 5.0 (MicroCal). Blank experiments (ligand into dialysate) all showed consistent dilution heats and were fitted with a line of zero slope to get the average (blank) value. Blank values were subtracted, and the initial data point removed to allow for equilibration at the syringe tip, before data were fit to the standard one site model of the Wiseman isotherm.[62] Arithmetic means and errors were generated from replicate experiments. Error was defined as the standard error in the mean of the observed parameters. The error was propagated upon subtraction between the datasets.

2a.2.4 NMR experiments of ^{13}C-^{15}N A103S MUP

An NMR sample was generated containing ^{13}C-^{15}N A103S MUP, produced as detailed in §2a.2.1.2, in PBS pH 7.4 with 10% v/v D_2O. All NMR spectra were gathered using a Varian Inova 600 MHz NMR spectrometer (Varian Inc., CA, USA) at 298 K. All experiments use sequences from the Biopack pulse sequence library supplied by Varian Inc. ^{1}H 1D spectra were measured using the water experiment from the Biopack library. 1D spectra underwent phasing and baseline correction using iNMR software.[63] The ^{13}C-^{1}H and ^{15}N-^{1}H HSQC spectra were processed using NMRPipe.[64] In these latter cases, phase correction was performed manually for ^{1}H using NMRDraw (part of NMRPipe), and automatically by Biopack (Varian) for heteronuclei. Data are shown in §2a.5, an annex to this chapter.

2a.3 Results and Discussion

2a.3.1 Extinction Coefficient (ε) Calculations

ε were empirically calculated for the MUP proteins and the NPOME ligand, §2a.2.2 and §2b.2.2, Tables 2.2 and 2.4, to aid concentration measurements for ITC experiments, the relevance of which is explained in §1.3.1. The ε_{280nm} value for the folded A103S MUP, 11752 M^{-1} cm^{-1}, lies close to the theoretically calculated unfolded protein ε_{280nm} of 11522 M^{-1} cm^{-1}. MUP has a higher ε_{280nm} of 14105 M^{-1} cm^{-1}. The NPOME $\varepsilon_{257.5nm}$ was calculated as 122 M^{-1} cm^{-1}.

2a.3.2 NPOME binding to MUP and A103S MUP

2a.3.2.1 ITC

The difference in binding parameters between two interactions that differ in the expulsion of a single water molecule from the MUP pocket, NPOME binding to MUP and A103S MUP, was observed using ITC, Table 2.3. The difference in the free energy of binding, $\Delta\Delta G^{\circ}_{b}$, is unfavourable when expelling the extra water molecule, arising from a larger enthalpic penalty and a smaller entropic benefit. Accounting for error,

the difference in the entropy of binding, $T\Delta\Delta S°_b$, is definitely positive, and at least 0.9 kJ mol^{-1}. This is much more positive than the previous estimates of 0.1 and -5.8 kJ mol^{-1}, Table 2.1.

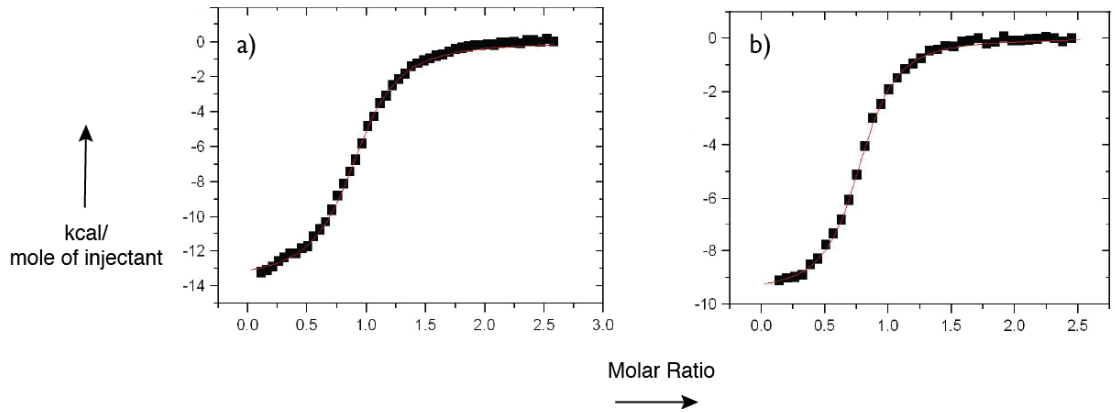


Figure 2.3 Representative ITC isotherms for NPOME binding to a) MUP and b) A103S MUP.

	$\Delta G°_b$	$\Delta H°_b$	$T\Delta S°_b$
MUP	-32.7 ± 0.1	-49.1 ± 1.4	-16.4 ± 1.3
A103S MUP	-30.2 ± 0.2	-42.7 ± 2.7	-12.5 ± 2.7
	$\Delta\Delta G°_b$	$\Delta\Delta H°_b$	$T\Delta\Delta S°_b$
MUP → A103S MUP	2.4 ± 0.2	6.3 ± 3.0	3.9 ± 3.0

Table 2.3 Thermodynamic parameters from ITC, 298 K: MUP and A103S MUP versus NPOME. All values are in kJ mol^{-1} and to 1.d.p. Observed parameters are arithmetic means from repeat experiments, errors are standard errors of the mean. Errors in the difference value, final row, are propagated from above.

The RMSD values for the crystal structures are low, indicating only a potential contribution from the A→S mutation as a non-solvent source of differential binding thermodynamics, as addressed in §2a.3.2.2. If the change in S103 sidechain entropy upon NPOME binding is between -6.9 and -0.9 kJ mol^{-1}, the entropic benefit of water ejection is zero within error. If more positive than -0.9 kJ mol^{-1}, water ejection has a positive entropic benefit. If rotating freely, i.e. with no energy barriers, a serine side

chain has 11.6 kJ mol^{-1} of entropy at 298 K.[65] This can be considered as an upper limit of entropy loss upon constraining a serine sidechain. The change for S103 upon the binding of NPOME, though unknown, is likely to be lower. Therefore experimental estimation of the change in sidechain entropy upon ligand binding is necessary for rigorous estimation of the magnitude and sign of the entropic cost of ejecting a water molecule from the MUP pocket.

2a.3.2.2 NMR

Experiments were started to investigate and account for potential changes in rotamer populations of S103 upon ligand binding. The goal was measurement of NMR N'-H_β, C'-H_β, and H_α-H_β ^3J couplings for a ^{13}C-^{15}N labelled A103S MUP sample before and after NPOME binding. However, due to a degraded sample, the experiments were aborted. Details are contained in §2a.5, an annex to this chapter.

2b.1 Introduction

2b.1.1 MUP as scaffold: mutating from hydrophobic to hydrophilic

Another member of the lipocalin superfamily is the Histamine Binding Protein (HBP), which is released in the saliva of the brown tick and sequesters host histamine to evade inflammation and detection.[66] The thermodynamics of this hydrophilic 'cousin' to MUP were recently investigated using panels of structurally related hydrophilic ligands.[28] The HBP pocket contains multiple polar sidechains, such that in the bound state a dicationic cognate histamine is thought to be stabilised by a network of hydrogen bonds and ionic interactions, PDB 1QFT.[66] Lipocalins have attracted attention beyond thermodynamic studies as structural scaffolds for biotechnological applications, due to their scavenging capacities and tolerance to mutation.[67,68] Furthermore, HBP's anti-inflammatory activity proves that lipocalins could be safely administered into the bloodstream.[69]

Engineering MUP for a novel capacity was attempted by Dr Caitriona Dennis, namely to enable histamine binding in a mode similar to related hydrophilic HBP. This was performed by introducing the ionisable sidechain Asp40 (D40) into the MUP pocket. Crystal structures, all obtained to a resolution of below 2.1 Å, revealed an increased apo pocket water content, accompanied by capacity to bind histamine and a partial incapacity to bind IBMP, Figure 2.4. All the crystals were grown and soaked in the same way, and are of the same space group and unit cell dimensions as for MUP. Therefore these relative differences in IBMP and histamine binding are not an artefact of crystal preparation: they may represent a successful change in binding profile through mutation of pocket sidechains.

Considering that large quantities of crystals (>50) can be produced with a small amount of material (<1 mg), crystallographic observations represent an attractive screening strategy for monitoring changes in binding profile as a result of MUP mutations. However, this approach must be validated by corroborating the putative changes in binding profile using ITC.

MUP crystallisation solution is pH 5.5 and 0.3 M salinity. MUP ITC is usually run at pH 7.4 and 0.15 M salinity. Therefore in attempting to partially recreate the crystalline conditions in ITC experiments, the salinity can be increased or the pH decreased: in this investigation the latter was attempted. Due to the potential risk of protein aggregation, increasing salinity was not attempted in this work.

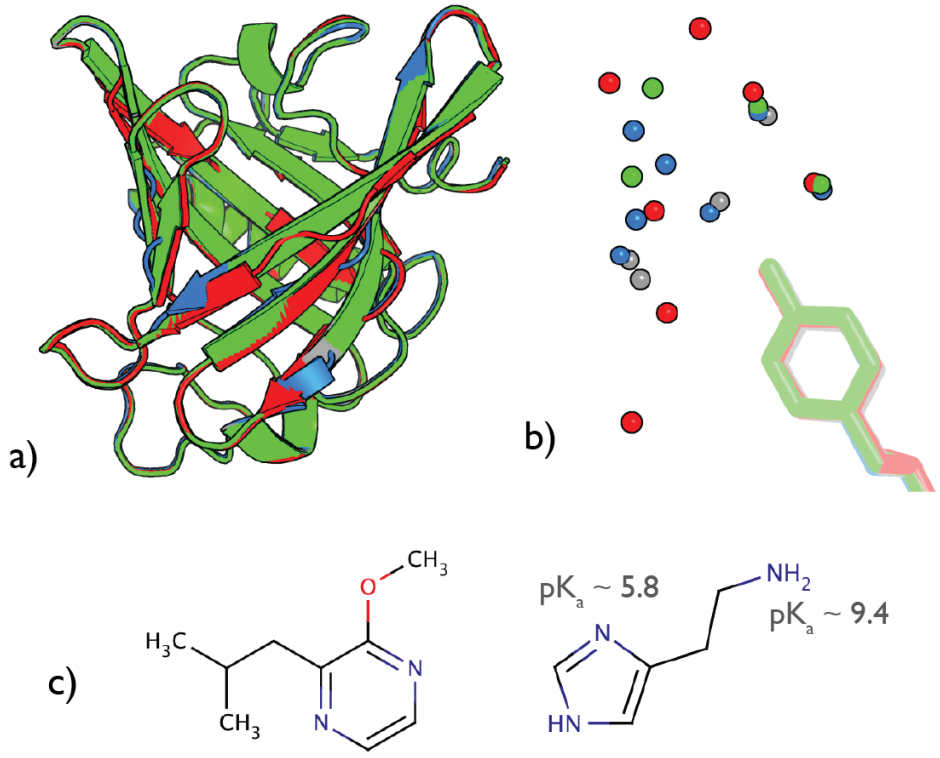


Figure 2.4 MUP mutants with increasing pocket water and crystallographically indicated shift in binding from IBMP to histamine. Mutant name (colour, # pocket water, ligands that bind to crystal [IB=IBMP, HA=Histamine]): MUP (green, 4: IB), A103S MUP (grey, 5: IB), A103S L40D MUP (blue, 7: IB, HA) and L40D MUP (red, 8: HA). a) overlay of all mutants, demonstrating close agreement of the backbone. b) increase in pocket water, showing Tyr120. c) structures of IBMP and histamine, the latter with associated pKa values for ionisable atoms.

The pK_a values for histamine dictate that it be mostly in dicationic form at pH 5.5, and therefore that the crystallographically observed interactions with MUP mutants involve salt bridges. However, pK_a values of ionisable amino acid sidechains sometimes vary

greatly when incorporated into a folded protein structure.[70] The pK_a value *in situ* is determined by the permittivity, electrostatic interactions and the availability of proton donors and acceptors. Furthermore, lower permittivity in protein interiors renders charge separation increasingly energetically costly and therefore unfeasible. The pK_a of Asp (D) in aqueous solution is 3.9, and a comprehensive meta-analysis of 139 measured Asp residues in 78 folded proteins revealed an average *in situ* Asp pK_a of 3.5 ± 1.2.[71] Nonetheless, computational techniques are herein applied to MUP to appreciate the potential risk of D40 sidechains being neutral at pH 5.5, and therefore incapable of forming salt bridge interactions with dicationic histamine. These techniques have been devised to estimate approximate *in situ* pK_a values using crystal structures of proteins, and report RMSD from experimentally-derived values of ~ 0.9 pK_a units, §2b.2.5.

2b.1.2 Work undertaken

Engineered lipocalins are promising for biotechnological applications. MUP and HBP are both lipocalins with hydrophobic (IBMP binding) and hydrophilic (histamine binding) binding profiles respectively. Histamine binding capacity, with an accompanying decrease in IBMP affinity, was putatively engineered in MUP by introduction of the ionisable sidechain D40 and increased solvent water content into the binding pocket. These observations were made using x-ray crystallography. Crystallography represents an attractive screening strategy for monitoring changes in binding profile as a result of MUP mutations, because large quantities of crystals can be produced with a small amount of protein.

To validate the crystallographic observations, ITC was performed at the crystalline pH of 5.5, on the panel of MUP variants shown in Figure 2.4 binding to both IBMP and histamine. Despite the crystallographically-observed histamine binding, there may be a risk of neutralising D40 at pH 5.5 due to its *in situ* pK_a being unknown. If the putative change in binding profile observed crystallographically is correct, then MUP and A103S MUP should not bind, or have distinctly lower affinity for, histamine. The same applies to L40D MUP-IBMP binding. All MUP variants used herein were expressed and purified, and checked by NMR. To ensure accurate ITC-derived data, extinction coefficients were calculated for all MUP variants used herein, and NMR was

used to calculate histamine concentration. An extinction coefficient for IBMP was previously calculated by Dr N. Shimokhina.[58]

2b.2 Materials and Methods

2b.2.1 Protein Expression and Purification

2b.2.1.1 Generation of MUP mutants

The MUP mutants A103S, A103S L40D and L40D were generated by Dr Caitriona Dennis and Ms Sue Matthews using QuikChange (Stratagene) site-directed mutagenesis, QIAprep spin miniprep kits (Qiagen) and XL1Blue competent cells (Stratagene) in accordance with a previously published procedure.[21] Dr Caitriona Dennis and Ms Sue Matthews performed much of the subsequent expression and purification for these proteins using the procedure detailed in §2a.2.1.1.

	Spectrophotometer / Buffer	
Protein	Genesys 6 (μM) / 6M GdnHCl	Cecil CE1021 (M^{-1} cm^{-1}) / PBS pH 5.5
MUP	293	14105
A103S MUP	195	11752
A103S L40D MUP	222	11612
L40D MUP	155	11844

Table 2.4 MUP absolute concentrations assuming an accurate unfolded ε_{280nm} for a protein sample in 6M GdnHCl using the Genesys 6 is 11522 M^{-1} cm^{-1}. Cecil CE1021 ε_{280nm} calculated for the same concentration solution in PBS pH 5.5.

2b.2.2 Extinction coefficient (ε) calculations

2b.2.2.1 MUP mutants

As indicated in §2a.2.2.1, assuming ε_{280nm} = 11522 M^{-1} cm^{-1} for unfolded 6M GdnHCl samples analysed with the Genesys 6 allows accurate calculation of MUP concentration. Samples derived from the same stock solution can then be used to measure ε_{280nm} for any solution conditions and spectrophotometer. ITC samples produced for experiments detailed in §2b.2.6.2 were used as stock solutions to produce samples in both 6M GdnHCl, unfolded, and PBS pH 5.5, folded. The GdnHCl samples were measured using the Genesys 6 spectrophotometer, and the PBS samples were measured using the Cecil CE1021. Accurate concentrations from the GdnHCl sample and the Cecil CE1021 PBS pH 5.5 ε_{280nm} were calculated, Table 2.4.

2b.2.3 ^{1}H 1D NMR experiments on MUP mutants

All NMR samples containing protein were measured with 10% v/v D_2O. All NMR spectra were gathered using a Varian Inova 500 MHz NMR spectrometer (Varian Inc., CA, USA) at 298 K, Figure 2.6. All experiments use sequences from the Biopack pulse sequence library supplied by Varian Inc. ^{1}H 1D spectra were measured using the water experiment from the Biopack library. 1D spectra were phased and baseline corrected using iNMR software.[63]

2b.2.4 Crystallography of MUP mutants

Crystallography was previously performed by Dr Caitriona Dennis, using previously defined MUP crystallisation conditions.[23]

2b.2.5 Theoretical pKa calculation for MUP mutants

Two online webservers, PROPKA and H++, were used (April 2010) to estimate the

pK_a of the Asp40 sidechain mutated into the MUP pocket.[72,73] Both methods use original crystal structures, and add missing hydrogen atoms before calculation.

2b.2.5.1 PROPKA

PROPKA is a structure-based heuristic 'fast empirical' method for predicting protein pK_a values that reports an RMSD from experimental values of 0.89.[73] The results are displayed in Table 2.5.

2b.2.5.2 H++

H++ uses a Poissson-Boltzmann approach to calculate the change in pK_a moving from known full solvation values to the position in the crystal structure. Adjustable parameters are salinity, external dielectric and internal dielectric. The latter two parameters refer to the dielectric constant (relative permittivity) at the surface and core of the protein respectively. An external dielectric constant of 80 was used, corresponding to water, and a salinity of 0.15 M, corresponding to the PBS buffer used for ITC experiments. The internal dielectric constant was set to both 4 and 15 and the range of values reported. The values reported as 'pK_(1/2)' by the webserver are displayed in Table 2.5. A comparison of predicted and experimental values was used to calculate an RMSD for H++, using only entries with predicted values between 0 and 14.[74] The calculated RMSD is 0.9 pK_a units.

2b.2.6 Isothermal titration calorimetry

2b.2.6.1 MUP mutants vs IBMP or Histamine

The same procedure as detailed in §2a.2.3.1 was performed, except as follows. ITC experiments were performed on a MicroCal MCS-ITC unit with a cell volume of 1.3047 mL at 298 K. The dialysis buffer was adjusted to pH 5.5 using HCl and filtered before dialysis using a 0.2 μM cellulose membrane filter. MUP concentrations, all ~ 100

μM, were calculated using the CE1021 spectrophotometer (Cecil Instruments) and the extinction coefficients determined in §2b.2.2.3. IBMP (Sigma Aldrich) concentration, ~ 1 mM, were calculated using ε_{220nm} = 4980 M^{-1} cm^{-1} (calculated in reference [58]) and the Genesys 6 spectrophotometer. Histamine concentration was estimated by weight, the solution adjusted to pH 5.5 using HCl, and then the solution concentration, ~ 1 mM, measured using ^{1}H NMR acquired using a Varian Inova 500 MHz spectrometer as follows. A tryptophan solution in PBS dialysate was made to a known concentration (~ 1 mM) using spectrophotometric measurement (Genesys 6 spectrophotometer, ε_{280nm} = 5502 M^{-1} cm^{-1}, taken from reference [75]). Histamine and tryptophan solutions were mixed in equal volume with 10% v/v D_2O, and histamine concentration was determined by measuring the ratio of ^{1}H peak integrals arising from each species, Figure 2.5. A 60 s delay was used between scans to ensure the system had returned fully to equilibrium and therefore avoid relaxation artefacts in the peak intensities.

Displacement ITC, described in §1.3.1, was performed by titrating IBMP into a solution containing the MUP and histamine samples mixed in equal volume, wherein concentrations of all constituents were calculated as above. Approximate concentrations were as follows; [MUP], 0.1 mM; [Histamine], 5 mM; [IBMP], 1 mM. Data were processed as per §2a.2.3.1, but instead of the Wiseman isotherm, the data were fit using the displacement model, §1.3.1 (files and instructions received in personal communication from Dr. Sigurskjold)[33], incorporating data from the IBMP titrations herein performed.

2b.3 Results and Discussion

2b.3.1 Extinction Coefficient (ε) Calculations

ε were empirically calculated for the MUP variants, §2a.2.2 and §2b.2.2, Tables 2.2 and 2.4, to aid concentration measurements for ITC experiments, the relevance of which is explained in §1.3.1. The ε_{280nm} values for the folded mutant protein, 11612 to 11844 M^{-1} cm^{-1}, lie close to the theoretically calculated unfolded protein ε_{280nm} of 11522 M^{-1} cm^{-1}. MUP has a higher ε_{280nm} of 14105 M^{-1} cm^{-1}. An ε_{220nm} of 4980 M^{-1} cm^{-1} was used for IBMP (calculated in reference [58]), and NMR was used to calculate histamine concentration, as detailed in §2b.2.6.

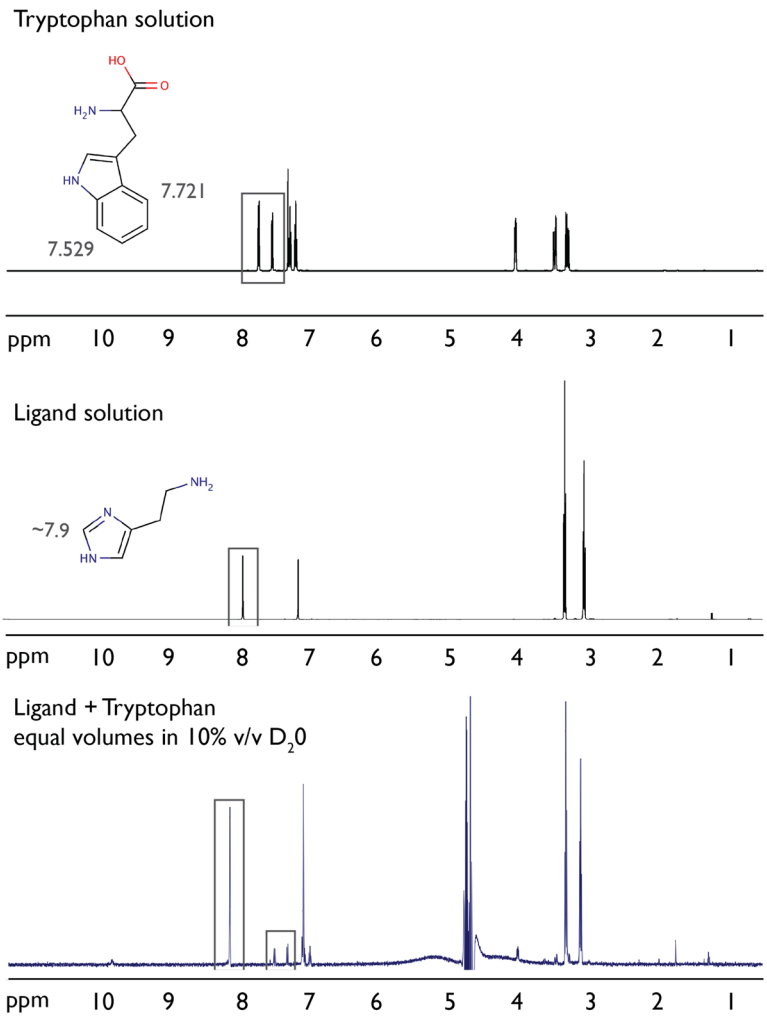


Figure 2.5 ^{1}H NMR measurements of histamine ligand concentration using a tryptophan solution. Molecular structures of tryptophan (top) and histamine (middle) are shown. ^{1}H resonance frequencies are indicated next to the relevant proton positions in the structures, and correspond to the boxed signals in the spectra. ^{1}H spectrum assignment for L-Tryptophan and ^{1}H spectrum for histamine were obtained from the Biological Magnetic Resonance Databank (BMRB).[76,77] The ratio of the indicated peak integrals from each species was used with the known tryptophan concentration to determine histamine concentration for ITC experiments. A 60 s delay was used between scans to ensure the system had returned fully to equilibrium and therefore avoid relaxation artefacts in the peak intensities.

2b.3.2 IBMP and Histamine binding to MUP mutants

2b.3.2.1 Theoretical pKa calculations

As already stated in §2b.1.1, the risk of D40 neutralisation at pH 5.5 was considered by using two webservers that utilise different calculation approaches, Table 2.5. Both estimates predict an increase of at least 1.5 pH units in the Asp pK_a due to the MUP pocket environment. This is much greater than the average *in situ* Asp pK_a of 3.5 ± 1.2 revealed by the meta-analysis mentioned in §2b.1.1, which represents a decrease of 0.4 pH units from the aqueous pK_a of 3.9. However, the same meta-analysis revealed a few Asp whose pKa increased upon folding by multiple pH units, demonstrating that such predicted increases have been observed in a few cases.

The empirical method PROPKA indicates that in both D40 versions of MUP, most of the D40 sidechains should be neutralised at pH 5.5. The H++ method, which calculates the *in situ* electrostatics by solving the Poisson-Boltzmann equation for each sidechain, reports a range of values depending on the internal dielectric constant variable. Nonetheless, at pH 5.5 most of the D40 sidechains in the sample are predicted to be neutralised. These predictions may guide further work, should the pH 5.5 ITC fail to reflect the trends seen in structures acquired from pH 5.5 crystals.

	A103S L40D MUP	L40D MUP
PROPKA	6.7	7.1
H++	5.4 to 10.4	5.4 to 8.8

Table 2.5 Predicted pK_a values for D40 in A103S L40D MUP and L40D MUP. The H++ row displays the range of values corresponding to a range of 4 to 15 in the value of the internal dielectric constant (relative permittivity).

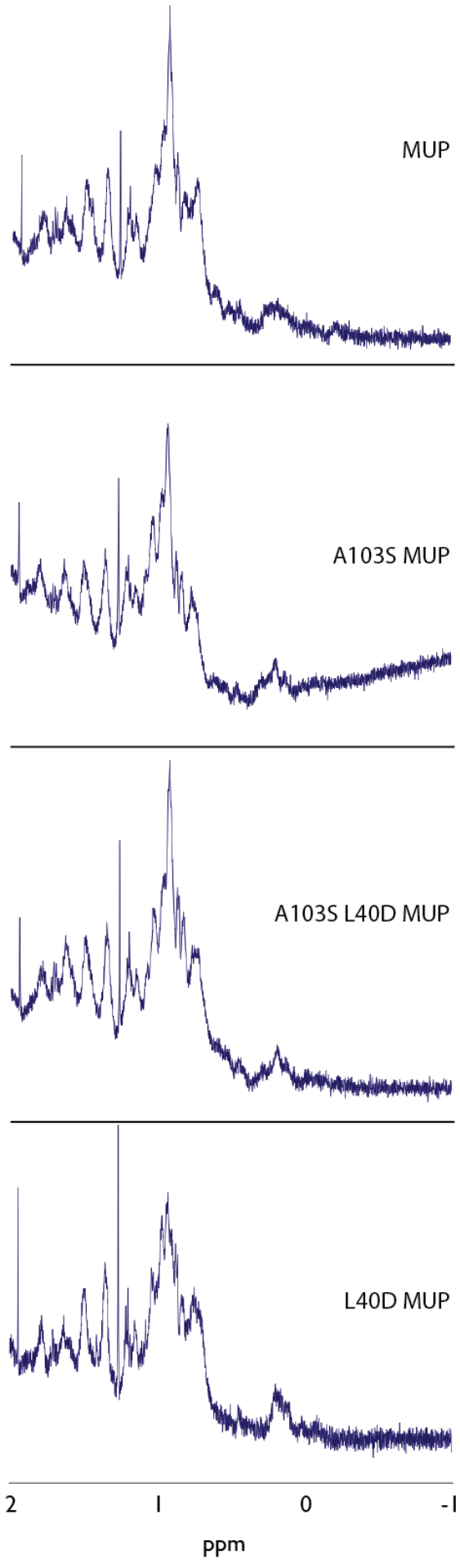


Figure 2.6 1D ^{1}H NMR spectra of MUP variants, PBS pH 7.4 with 10% v/v D_2O. The section of the spectrum between 2 and -1 ppm is shown for comparison. All the spectra are very similar, demonstrating that the mutations introduce no change from MUP observable by NMR. Therefore mutation does not appear to result in unfolding.

2b.3.2.2 NMR

To assess whether the mutations introduce any structural changes under ITC conditions, e.g. unfolding, ^{1}H spectra were acquired for all MUP variants to be investigated using ITC, Figure 2.6. These experiments used the same buffer as the ITC experiments, PBS pH 5.5 with 10% v/v D_2O. All the spectra are very similar, indicating that the mutations do not introduce any obvious difference from MUP as observable by ^{1}H NMR. Furthermore, the thermodynamic parameters for IBMP binding to MUP, Table 2.6, are sufficiently close to values at pH 7.4 to conclude that the protein is still folded and binding-competent upon titration to pH 5.5.[23] Thus all MUP mutants should be folded and binding competent in these solution conditions.

	$\Delta G°_b$	$\Delta H°_b$	$T\Delta S°_b$
	IBMP		
MUP	-36.3 ± 0.3	-42.5 ± 0.5	-6.2 ± 0.9
A103S MUP	-36.2 ± 0.3	-46.9 ± 1.2	-10.8 ± 1.5
A103S L40D MUP	-35.4 ± 0.1	-45.9 ± 0.8	-10.5 ± 0.7
L40D MUP	-35.4 ± 0.1	-37.7 ± 0.7	-2.3 ± 0.6
	Histamine		
MUP	-14.4	6.5	20.9
A103S MUP	-16.7 ± 1.1	12.7 ± 3.1	29.4 ± 4.3
A103S L40D MUP	-18.7 ± 0.1	8.3 ± 0.1	27.0 ± 1.1
L40D MUP	-17.8	21.8	39.6

Table 2.6 Thermodynamic parameters from ITC at pH 5.5, 298 K: MUP variants binding to IBMP and Histamine. All values are in kJ mol^{-1} and to 1.d.p. Observed parameters are arithmetic means from repeat experiments, errors are standard errors of the mean. No errors are reported for those values arising from a single experiment. The values for histamine binding refer to displacement ITC experiments as described in §2b.2.6.2.

2b.3.2.3 ITC

Table 2.6 details the thermodynamic parameters for MUP variants binding to IBMP and Histamine at pH 5.5, 298 K. Representative isotherms are displayed in Figure 2.7. There is little change in binding affinity for IBMP across the variants, 0.9 kJ mol^{-1}, despite the increasing number of pocket water molecules and the inability to obtain a crystal complex of IBMP bound to L40D. The histamine isotherms showed no evidence of binding after multiple titrations, despite a 50-fold excess of ligand (5 mM ligand and 0.1 mM protein), leading to abortion of the experiments. Displacement ITC was used to assess whether weak binding was occurring, as detailed in §2b.2.6.2. IBMP was titrated into a pre-mixed solution of MUP and histamine, and by using the IBMP binding parameters acquired herein, the histamine binding parameters were extracted. Due to limited sample, only single displacement titrations were performed for both MUP and L40D MUP, as shown by the lack of errors for these values in Table 2.6. This limits the interpretation of changes in binding profile across the series. Nonetheless, crystallographically, A103S MUP does not bind histamine whereas A103S L40D MUP does. However ITC reveals only a 2 kJ mol^{-1} difference in binding affinity between the two interactions.

The weak binding of histamine and continued binding of IBMP could be consistent with a neutralised D40, predicted by computational methods. Though large increases in Asp pK_a upon burial are very rare, they have been observed.[71] Alternatively, the D40 could still be charged. In this case, histamine affinity would be fundamentally weak and relatively unaffected by the presence of a charge in the pocket, despite the considerable enthalpic benefit of forming a salt bridge in a lower dielectric environment such as the MUP pocket. Additionally, the introduction of a charged D40 would have to not affect IBMP affinity. Consequently, these data appear to support the computational predictions that D40 is neutralised at pH 5.5. It is possible that the salinity, which affects pK_a values and differs between ITC and crystallography as detailed in §2b.1.1, needs to be adjusted to replicate the binding trends observed crystallographically. Alternatively, the trends may be reproducible at higher pH. Furthermore, the *in situ* D40 pK_a could be directly measured using NMR, by performing a pH titration on a ^{15}N-^{13}C labeled protein sample and monitoring the chemical shift of the D40 sidechain atoms.

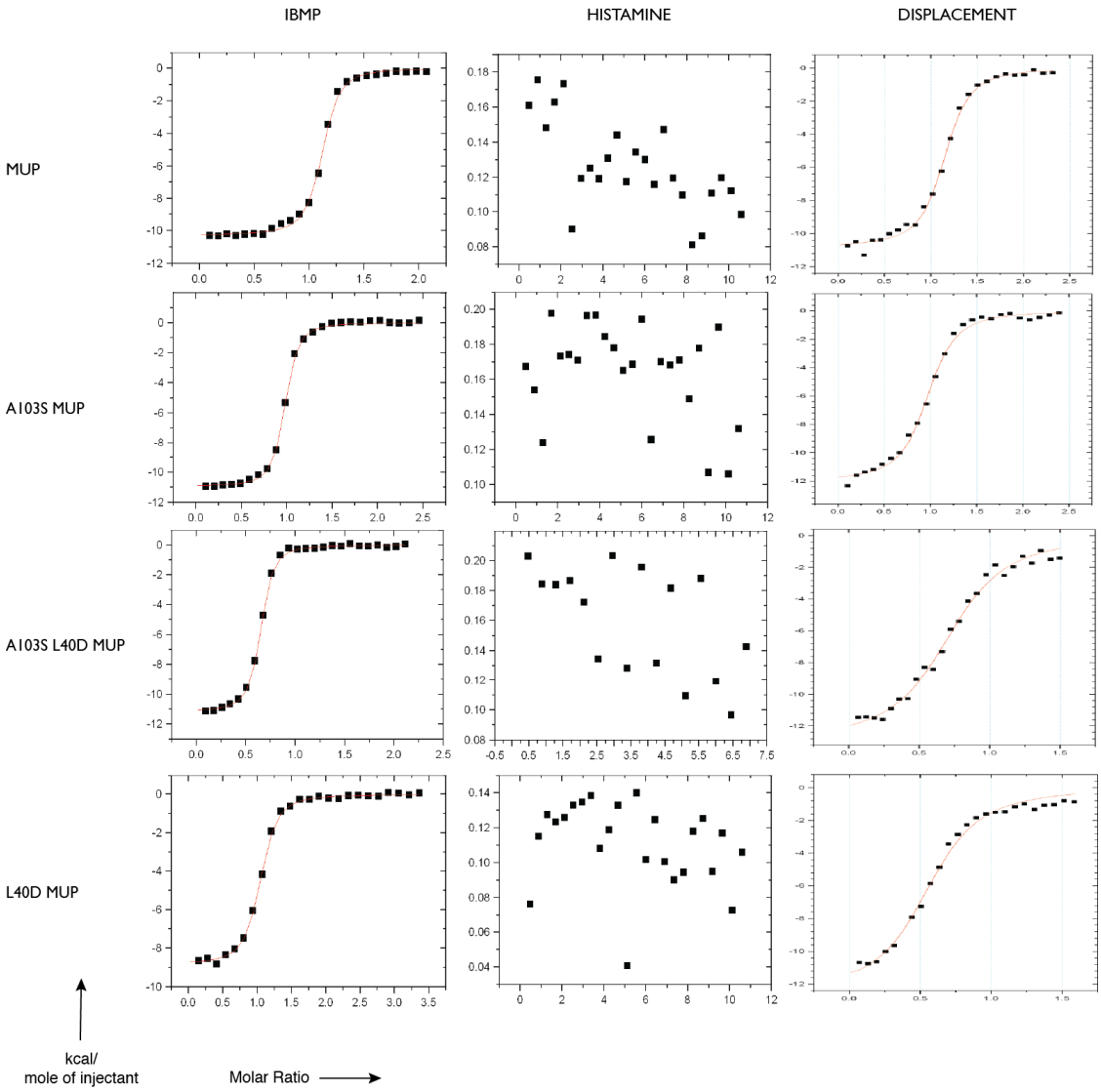


Figure 2.7 Representative ITC isotherms for MUP mutants binding to IBMP and histamine in PBS pH 5.5. The histamine isotherms show no evidence of binding. Subsequently, displacement ITC was used, wherein IBMP was titrated into a mixture of MUP and histamine. Details are in §2b.2.6.2.

2.4 Summary and Conclusions

A more reliable entropic estimate for the ejection of a single water molecule from the MUP pocket was sought, to aid future decomposition and investigate proposed entropic solvation. The difference in binding entropy of MUP-NPOME and A103S

MUP-NPOME was observed via ITC to be 3.9 ± 3.0 kJ mol^{-1}, therefore a positive value of at least 0.9 kJ mol^{-1}. Previous estimates suggested 0.1 or -5.8 kJ mol^{-1}.

In decomposition terms, ligand desolvation cancels between the interactions, and due to the close RMSD of the structures, 0.37 Å, the only difference in the intrinsic entropic contribution is presumed to be the possible change in S103 rotameric entropy upon ligand binding. Therefore, to make this estimate reflect only the ejection of an extra water molecule, the entropic contribution of the serine sidechain rotamer sampling needs to be accounted for. If rotating freely, a serine side chain has 11.6 kJ mol^{-1} of entropy at 298 K, the temperature at which ITC was performed.[65] Presuming the serine sidechain is not rotating freely in the pocket, any entropic difference arising from ligand binding is likely to be much lower than 11.6 kJ mol^{-1}. Therefore, if the sidechain entropy change upon NPOME binding is between -6.9 and -0.9 kJ mol^{-1}, the entropic benefit of water ejection is zero within error. If more positive than -0.9 kJ mol^{-1}, water ejection has a positive entropic benefit. S103 rotamer sampling measurements were sought using NMR, but a defective sample halted the experiments. Further work would include regenerating this sample.

Crystal structures indicated that the histamine binding of HBP was successfully engineered into MUP through mutating ionisable sidechains into the binding pocket and increasing pocket solvation. The variants containing L40D appear capable of binding histamine, which is charged at the crystal pH of 5.5, and one variant even seems to no longer be capable of binding IBMP.

Because large quantities of crystals can be produced with a small amount of protein, crystallography is an attractive screening strategy for monitoring mutation-derived changes in MUP binding profile. However, this approach requires validation. ITC was performed on the MUP variants binding IBMP and histamine, to assess whether the same trends in binding across the panel of MUP mutants are observed. To attempt the closest possible comparison between the techniques, the ITC solution conditions were changed to partially mimic the crystalline conditions. The pH was lowered to 5.5, but the salinity was not increased in this work due to the potential risk of aggregation. Computational methods predicted a rare high pK_a for the D40 residue, presenting a risk of neutralisation at pH 5.5 that would likely ablate the binding of charged histamine.

Contrary to the crystallographic indications, no significant changes in IBMP or histamine affinity were observed across the series. Histamine affinity was consistently very weak for all MUP mutants. These results indicate that the computational

predictions of D40 being neutralised at pH 5.5 are correct. Possibly the salinity needs to be adjusted to replicate the binding trends observed crystallographically, as it affects pK_a values and differs between ITC and crystallography as detailed in §2b.1.1. It is also possible that the crystallographically observable trends may be reproducible at higher pH. Furthermore, ongoing work could be to directly measure the *in situ* D40 pK_a using NMR, by monitoring the chemical shift of the D40 sidechain atoms whilst performing a pH titration on a ^{15}N-^{13}C labeled protein sample. In summary, this study has demonstrated that validating the use of crystals to screen interactions that are heavily pH dependent is non-trivial. The validity of the crystallographic observations in this case remains unresolved.

2a.5 Annex: NMR experiments on ^{15}N-^{13}C A103S MUP

In order to experimentally estimate the change in rotameric sampling of S103 upon ligand binding, a sample of ^{15}N-^{13}C A103S MUP was produced for NMR experiments as detailed in §2.2.1.2 and §2.2.7. Initial ^{15}N-^{1}H HSQC spectra revealed a lack of peaks for the A103S MUP sample compared to an equivalently labelled MUP sample, Figure 2.8. To assess whether this was due to an insufficiently folded sample, the A103S MUP was unfolded using ethanol and refolded in buffer, as per the final step of protein purification detailed in §2.2.1.1, and a ^{15}N-^{1}H HSQC spectrum again obtained. However, more peaks are absent from the A103S MUP spectrum after refolding the protein. A ^{13}C-^{1}H HSQC spectrum revealed a similar comparative lack of peaks, especially in the methyl region of the spectrum, where no methyl peaks are observed below 0.5 ppm, Figure 2.9a. To assess whether these spectra are characteristic of A103S MUP or due to a defective sample, a ^{1}H spectrum of unlabelled A103S MUP was acquired, Figure 2.9b. There are clearly defined peaks in the methyl region below 0.5 ppm, concluding that this particular ^{15}N-^{13}C A103S MUP sample is defective. However, many of the peaks are present in both 2D spectra, and are similarly dispersed, indicating the presence of secondary structure.[78,79] This is not consistent with proteolytic digestion, which would result in peptide fragments and lead to very narrowly dispersed peaks in 2D spectra.[80] It would appear that much of the 'core structure' of the protein is preserved, indicating only a minor defect. This was confirmed by electrospray ionisation mass spectral (ESI-MS) analysis of a small aliquot of the ^{15}N-^{13}C A103S MUP sample. Whereas the expected weight is 21650 Da[81], multiple peaks are present spanning a range of ~90-95% of the expected weight, Figure 2.10. Therefore it appears that this sample is defective because limited amounts of protein mass, 5-10%, have been removed. SDS-PAGE run on aliquots of this sample confirms this observation, Figure 2.11. If proteolytic degredation had occurred, no band would be visible at ~21 kDa, and multiple lower weight bands would appear. On the contrary, the banding pattern appears the same, consistent with the subtle differences observed via mass spectrometry. Further work would be required to define the nature of this defect/degredation, and this sample would need to be regenerated for these experiments to continue.

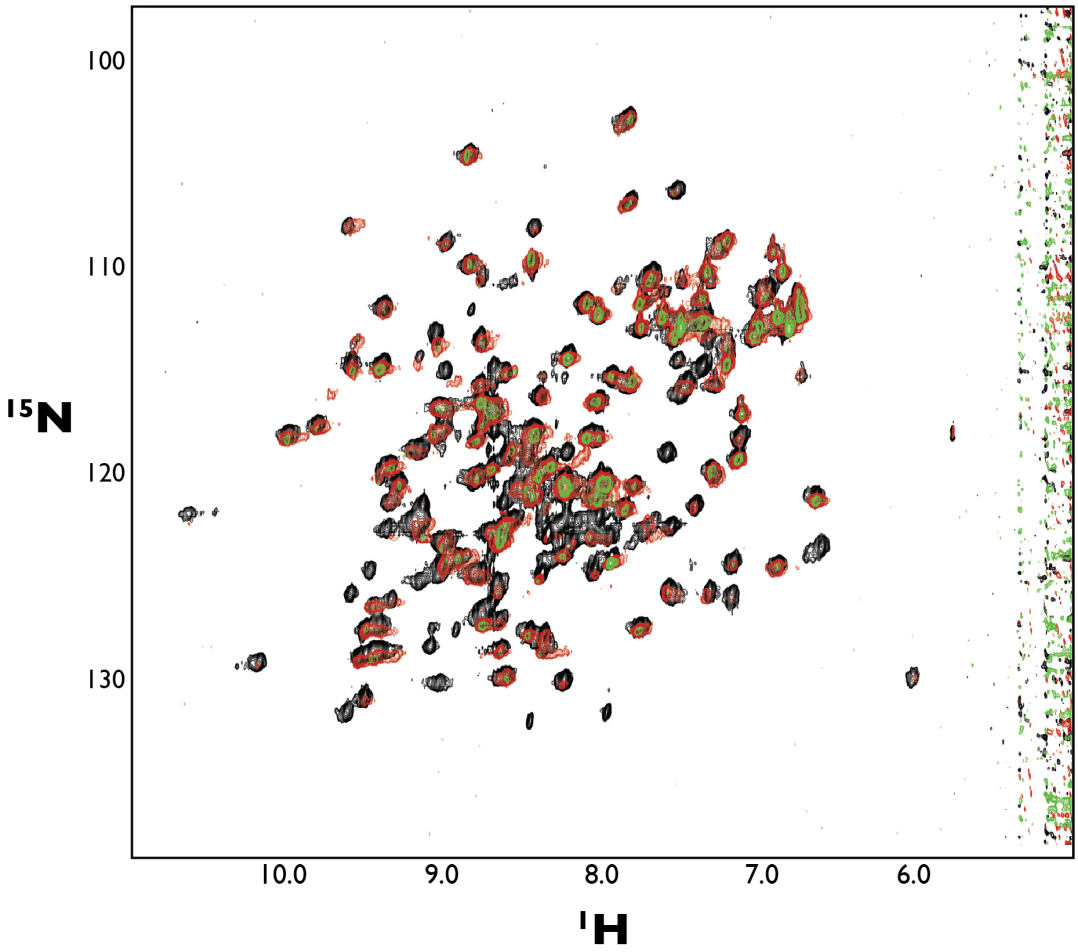


Figure 2.8 ^{15}N-^{1}H HSQC spectra of ^{15}N-^{13}C MUP (black), ^{15}N-^{13}C A103S MUP (red) and refolded ^{15}N-^{13}C A103S MUP (green). Many peaks observed in the MUP spectrum are absent in the A103S MUP spectra, especially after refolding the protein.

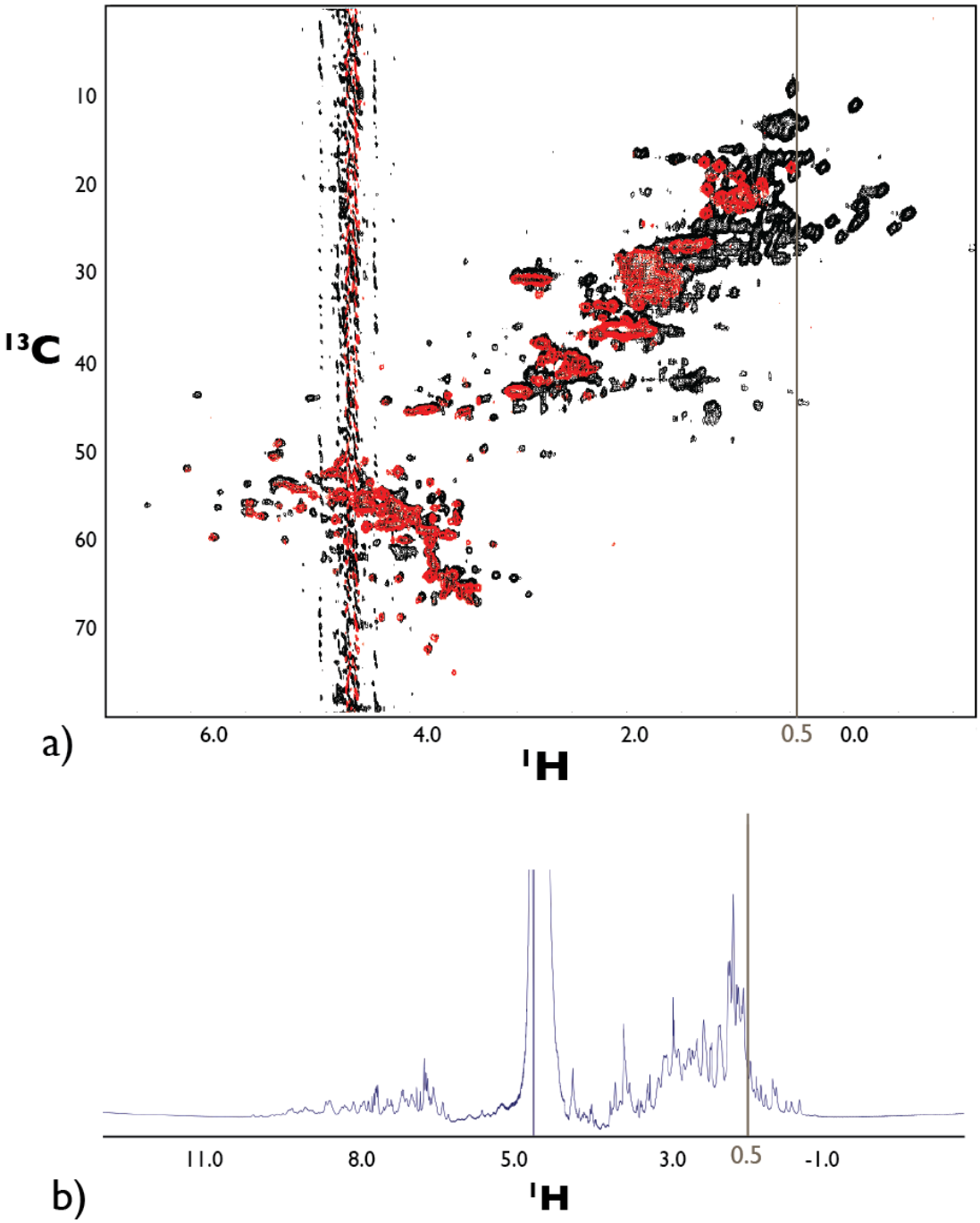


Figure 2.9 a) ^{13}C-^{1}H HSQC spectra of ^{15}N-^{13}C MUP (black) and ^{15}N-^{13}C A103S MUP (red). Many peaks observed in the MUP spectrum are absent in the A103S MUP spectra, especially in the methyl region, top right. No signals are observed lower than 0.5 ppm (^{1}H) for the A103S MUP sample. b) ^{1}H NMR spectrum of unlabelled A103S shows clear peaks below 0 ppm. Therefore this particular ^{15}N-^{13}C A103S MUP sample is defective, rather than the discrepancies observed here and in Figure 2.8 being characteristic of A103S MUP.

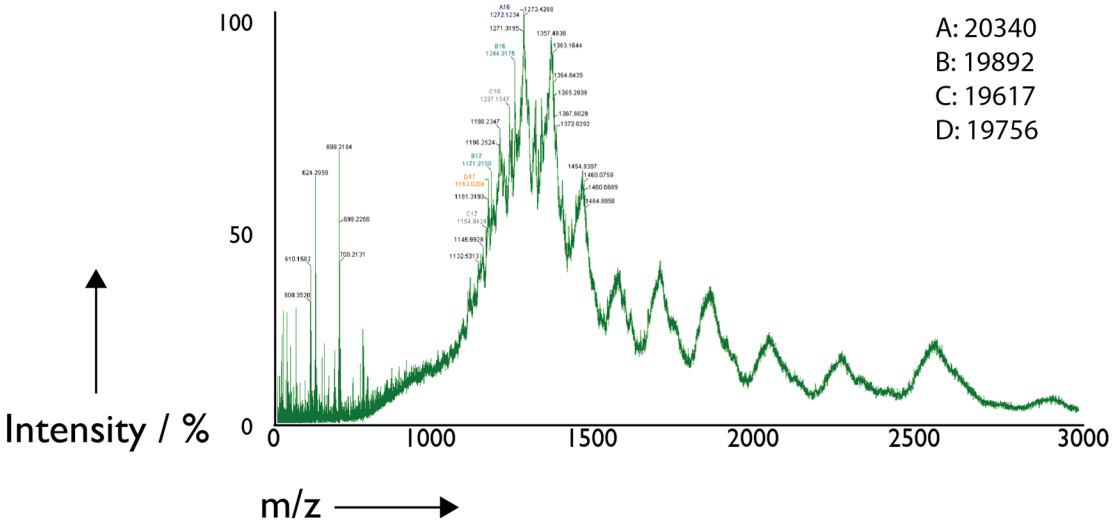


Figure 2.10 Mass spectrum (ESI-MS) of ^{15}N-^{13}C A103S MUP. The expected weight is 21650 Da.[81] Multiple peaks are present on the mass spectrum up to a weight of 20340 Da, indicating a partially degraded sample. 91-94% of expected weight.

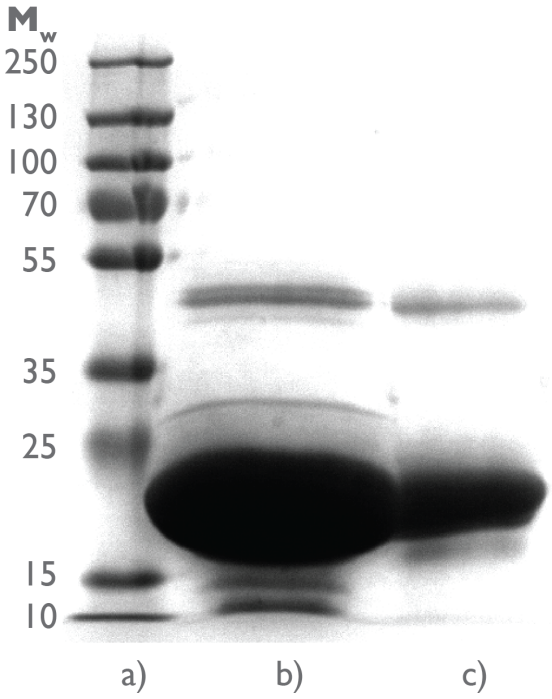


Figure 2.11 SDS-PAGE gel: ^{15}N-^{13}C A103S MUP sample and unlabelled A103S MUP. The NMR samples used in Figure 2.9 were run on a 12.5% polyacrylamide gel. a) molecular weight markers, b) unlabelled A103S MUP, c) ^{15}N-^{13}C A103S MUP sample. The main bands are seen at about the same molecular weight for both samples, corroborating the mass spectral indication of subtle degredation, Figure 2.10.

Chapter 3

Ligand conformational entropy

3.1 Introduction

3.1.1 A received wisdom based on insufficient data

Upon interacting with a protein, ligand flexibility is reduced. This incurs an entropic penalty that unfavourably contributes to the free energy of association, and therefore the affinity of the interaction. The most prominent loss, of rotational and translational entropy, is addressed in the following chapter. However, the second most prominent loss is that of conformational entropy, which arises from the restriction of bond rotation.

The entropic penalty associated with decreasing accessible ligand rotameric degrees of freedom upon protein binding has long been considered avoidable by employing chemical modifications that constrain or decrease the number of rotatable bonds (rotors). Values of reclaimable entropy between 2 and 6 kJ mol^{-1} per rotor have been suggested and supported by various studies.[5,82-85] Because a 5.7 kJ mol^{-1} change to the observed free energy of binding, $\Delta G°_{obs}$, corresponds to an order of magnitude difference in affinity, realizing these predictions is of significant interest to drug design. However, decreasing rotor numbers as a general principle has produced inconsistent results.[6,7] Despite a lack of consistent experimental data, these predictions have already had an impact upon important decisions in drug development; computational ligand docking scoring functions currently apply these values as context-independent additive per-rotor penalties.

Therefore these supposedly acheivable benefits remain to be experimentally evaluated for multiple systematic modifications in a protein-ligand system. It is also unknown whether any effects are significantly dependent upon the nature of the structural modifications employed. Decreasing the number of ligand rotors by one can be achieved in two ways, either restricting a rotor through introducing a double bond, or removing a rotor, e.g. shortening an alkyl chain by one methylene.

3.1.2 A tractable system is required to overcome major barriers for investigation

There are two related obstacles to observing the effect of multiple systematic chemical modifications in a single system.

Firstly, to quote Benfield *et al.*, "there are few cases where association constants are determined for a pair of constrained and flexible ligands having the same number and type of heavy atoms, the same functional groups, and the same number of hydrogen-bond donors and acceptors. Appropriate controls are thus generally absent".[86] Overcoming this requires a relatively promiscuous system tolerant to modification, necessary for multiple pairwise comparisons to be performed.

Secondly, many protein-ligand interactions are dominated by specific polar interactions that are distance and angle dependent. This renders explicit assessment of the entropic effects of structural modifications that constrain or remove rotors difficult. A system wherein protein-ligand interactions are dominated by weak, non-polar, non-directional interactions is necessary for an explicit assessment of the entropic effects of ligand structural modification.

MUP meets both of these requirements, as described in §1.2, due to its large apolar binding cavity. Ligand interactions with MUP are dominated by weak, non-polar, non-directional interactions. Combined with the size of the cavity, this allows MUP to be promiscuous, i.e. bind a wide range of ligands, allowing multiple pairwise comparisons to be performed.

3.1.3 Thermodynamic decomposition

Two parameters are most important in evaluating the effects upon binding of structural modifications that alter ligand conformational entropy. First is the change in binding affinity as represented by the change in the observed interaction free energy, $\Delta\Delta G^\circ_{obs}$, which is directly accessible by comparing isothermal titration calorimetry (ITC) data. Second is the contribution arising from changes in the ligand conformational entropy as a result of structural modification, which requires decomposition of the changes in observed thermodynamic parameters of binding measured by ITC. These parameters, $\Delta\Delta G^\circ_{obs}$, $\Delta\Delta H^\circ_{obs}$ and $T\Delta\Delta S^\circ_{obs}$, can be separated into contributions arising separately from differences in complex solvation, ligand

solvation, and 'intrinsic' factors.[1] 'Intrinsic' is defined as 'without effects of solvation', i.e. contributions from ligand and protein only, therefore containing the ligand conformational entropy change. This approach can be used equivalently for free energy, enthalpy and entropy, producing $\Delta G°_i$, $\Delta H°_i$ and $T\Delta S°_i$, as detailed in §1.1.

Decomposition has previously been performed for interactions of MUP[18], and the theory is detailed in §1.2. Briefly, the observed thermodynamic parameters can be decomposed by focusing on the differences in binding thermodynamics across a panel of similar ligands, using a Born-Haber cycle to represent and analyse pairs of interactions.[1] Giving an example in terms of entropy, assuming that the solvation of the two complexes is equivalent, the difference in intrinsic binding entropy ($T\Delta\Delta S°_i$) between two related ligands is composed only of two terms: the difference in observed binding entropy ($T\Delta\Delta S°_{obs}$), and the difference in ligand desolvation entropy, ($T\Delta\Delta S°_{sL}$), Equation 3.1.

$$T\Delta\Delta S°_i = [T\Delta S°_{i2} - T\Delta S°_{i1}] = [T\Delta S°_{obs2} - T\Delta S°_{obs1}] + [T\Delta S°_{sL2} - T\Delta S°_{sL1}]$$

Equation 3.1

$T\Delta S°_i$ is the intrinsic binding entropy, $T\Delta S°_{obs}$ is the observed binding entropy (measured using ITC), $T\Delta S°_{sL}$ is the ligand desolvation entropy, and the subscripts '2' and '1' refer to two closely related ligands that bind to a given protein.

3.1.4 Accounting for solvent and protein contributions in thermodynamic decomposition

When generating intrinsic thermodynamic parameters, decomposition should directly account for changes in ligand desolvation upon modification. Ligand desolvation entropy, $T\Delta S°_{sL}$, can be measured experimentally from air-solvent partition equilibria, although this is not practicable for all ligands. Fortunately, an additive technique for calculating desolvation parameters, derived from an extensive pool of published experimental data, gives excellent experimental agreement for various hydrocarbons at 298 K and 1 atm.[11] If protein-ligand complexes are differentially solvated, then the calculated intrinsic values contain contributions from this difference. Furthermore, to obtain ligand-specific contributions to intrinsic entropy, decomposition must also account for differential protein dynamics. The most rigorous experimental method

would be to calculate entropy values from differences in NMR relaxation parameters obtained before and after ligand binding.[87] NMR relaxation data acquired for MUP-ligand binding[23] demonstrate that the magnitude of error in the measurements, and subsequently computed entropy values, can be sufficiently high to render such time-consuming experiments of little value to decomposition in this system. Furthermore, the magnitude of differential protein dynamics contributions are expected to be small, because those previously measured were zero within error.[23] If not accounted for, Equation 3.1 can still be used, but the intrinsic values calculated will potentially contain contributions from these sources.

Lacking an empirical value for differential complex solvation or useful NMR relaxation data, crystal structures can be analysed to appreciate the potential contributions from these two factors to reported intrinsic parameters.

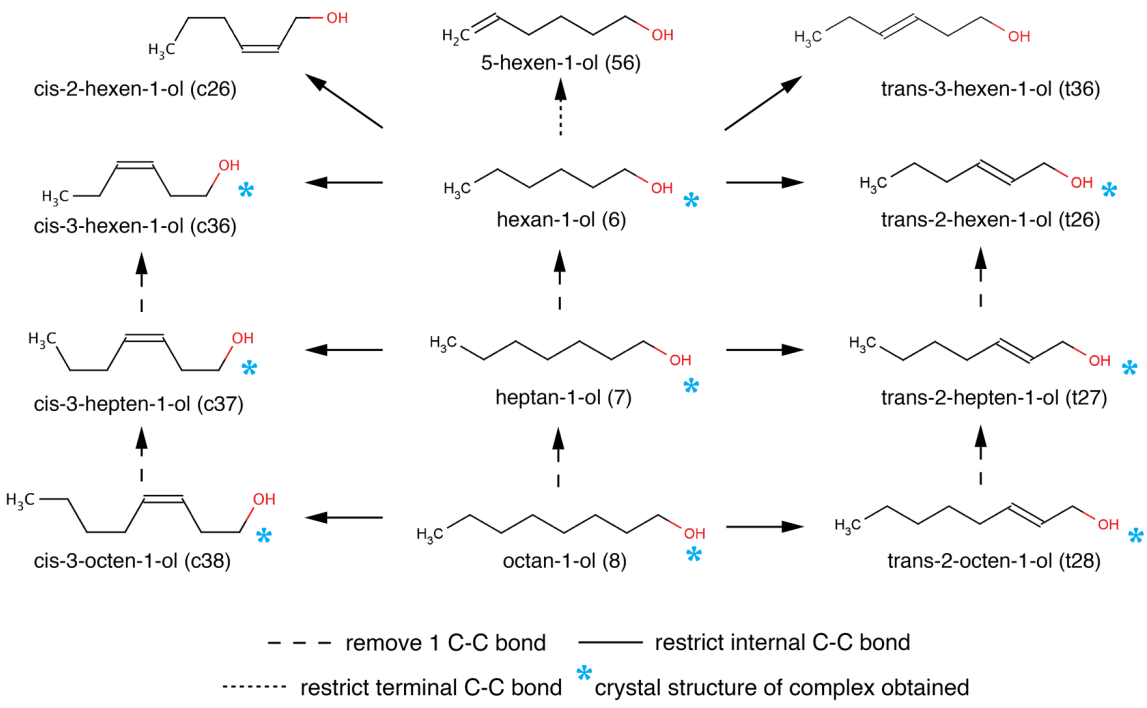


Figure 3.1 Saturated and unsaturated alcohols, including variation in cis/trans isomeric state and double bond position. Ligand abbreviations used throughout the chapter are shown in brackets. Pairwise comparisons, represented by arrows, describe the effect of restricting a rotor (introduction of a double bond) or removing a rotor (one less methylene group in the alkyl chain). Previously generated ITC[28] and crystal data[1] were used for 6, 7 and 8.

3.1.5 Work undertaken

Removing a single ligand rotor in a model hydrophobic binding system was investigated by comparing MUP binding within and between panels of saturated and unsaturated alcohols, Figure 3.1. Using this panel of ligands allows two methods for decreasing rotor numbers to be assessed: rotor removal through shortening the molecule by one methylene group, and rotor rigidification (restriction) through introduction of a double bond. Restriction modifications were also assessed in both cis and trans isomers, and with the double bond placed at different positions. Therefore this work evaluates a large number of systematic modifications in a protein-ligand system. Observed and intrinsic thermodynamics of binding were obtained using ITC and calculated desolvation parameters. Crystal structures were obtained and analysed to appreciate potential contributions from differential complex solvation and protein dynamics to the intrinsic thermodynamic parameters.

3.2 Material and Methods

3.2.1 Isothermal titration calorimetry

3.2.1.1 Experimental

MUP protein was expressed and purified as described in §2.2.1.1. ITC experiments were performed on a MicroCal MCS-ITC unit with a cell volume of 1.3047 mL at 298.15 K. MUP solution was prepared from freeze-dried protein in filtered and degassed PBS pH 7.4 containing 1 mM sodium azide (Fisher Scientific). This solution was dialysed overnight at 4 °C, using 7 kDa molecular weight cutoff Snakeskin dialysis tubing (Thermo Scientific). Post-dialysis the MUP solution was filtered using a SS-10 ES10 mL syringe (Terumo) and 0.2 μM Minisart filter (Sartorius Stedim). MUP concentration was calculated using ε_{280nm} = 14105 M^{-1} cm^{-1} using a CE1021 spectrophotometer (Cecil Instruments). All MUP solutions were stored at 4 °C, and underwent 30 minutes of pre-experiment degassing at approximately 23 °C under vacuum in a ThermoVac temperature-controlled vacuum chamber (MicroCal).

The PBS dialysate was re-filtered using a 0.2 μM cellulose membrane filter (Whatman) and thoroughly degassed before being used to make ligand solutions. Alkenols (95% +) were obtained from Alfa Aesar. Alkenol concentrations were estimated by weight and then concentrations measured using ^{1}H NMR spectra acquired using a Varian Inova 500 MHz spectrometer. A tryptophan solution in PBS dialysate was made to a known concentration (~ 1 mM) using spectrophotometric measurement (Genesys 6 spectrophotmeter, ε_{280nm} = 5502 M^{-1} cm^{-1}, taken from reference [75]). Alkenol and tryptophan solutions were mixed in equal volume and 10% v/v D_2O added. Alkenol concentration was determined by measuring the ratio of ^{1}H peak integrals arising from each species, as shown in Figure 3.2. A 60 s delay was used between scans to ensure the system had returned fully to equilibrium and therefore avoid relaxation artefacts in the peak intensities.

Experiments comprised a single injection of 2 μL after a 60 s initial delay, followed by 24 injections of 10 μL, at 240 s intervals using a 300 rpm stirring speed. The cell was cleaned between each run with 1 M NaOH, then extensively with both water containing 1 mM sodium azide and PBS dialysate. Between runs of different ligands the syringe was extensively flushed using both these solutions, but only PBS dialysate was used between titrations with the same ligand.

3.2.1.2 Data Analysis

Data were analysed in Origin 5.0 (MicroCal). Blank experiments (ligand into dialysate) all showed consistent dilution heats and were fit with a line of zero slope to get the average (blank) value. Blank values were subtracted, and the initial data point removed to allow for equilibration at the syringe tip, before data were fitted to the standard one site model of the Wiseman isotherm.[62] Representative isotherms for each ligand are shown in Figure 3.3. Arithmetic means and errors were generated for replicate experiments. $\Delta\Delta G^{\circ}_{obs}$, $\Delta\Delta H^{\circ}_{obs}$, $T\Delta\Delta S^{\circ}_{obs,}$ were calculated directly from ITC values using Equation 3.1. $\Delta\Delta G^{\circ}_{i}$, $\Delta\Delta H^{\circ}_{i}$ and $T\Delta\Delta S^{\circ}_{i}$ values were calculated with measured ITC values in Table 3.2 and calculated desolvation values using Equation 3.1. Group averages were calculated by type of modifications and alkenol isomer.

Error was defined as the standard error in the mean and calculated for $\Delta\Delta G^{\circ}_{obs}$ and $\Delta\Delta H^{\circ}_{obs}$, then propagated through the Gibbs function to obtain errors for $T\Delta\Delta S^{\circ}_{obs}$. The standard error of the mean requires that a standard deviation be

calculated, which is not meaningful for datasets of less than three measurements. A publication that investigated random ITC error detailed a method for generating errors applicable to such datasets, which was used for experiments with less than three replicates.[88] Errors in the calculated solvation parameters $\Delta G°_h$ and $\Delta H°_h$, were defined as the average of errors reported in Tables 1 and 2 of Plyasunov *et al.*,[11] excepting data for monoterpenes, alkynes and diynes. These errors were propagated through the Gibbs function to obtain errors for $T\Delta S°_h$. All errors were further propagated through Equation 3.1 to obtain errors for intrinsic and observed changes in thermodynamic parameters. Data are displayed in Tables 3.2 and 3.3.

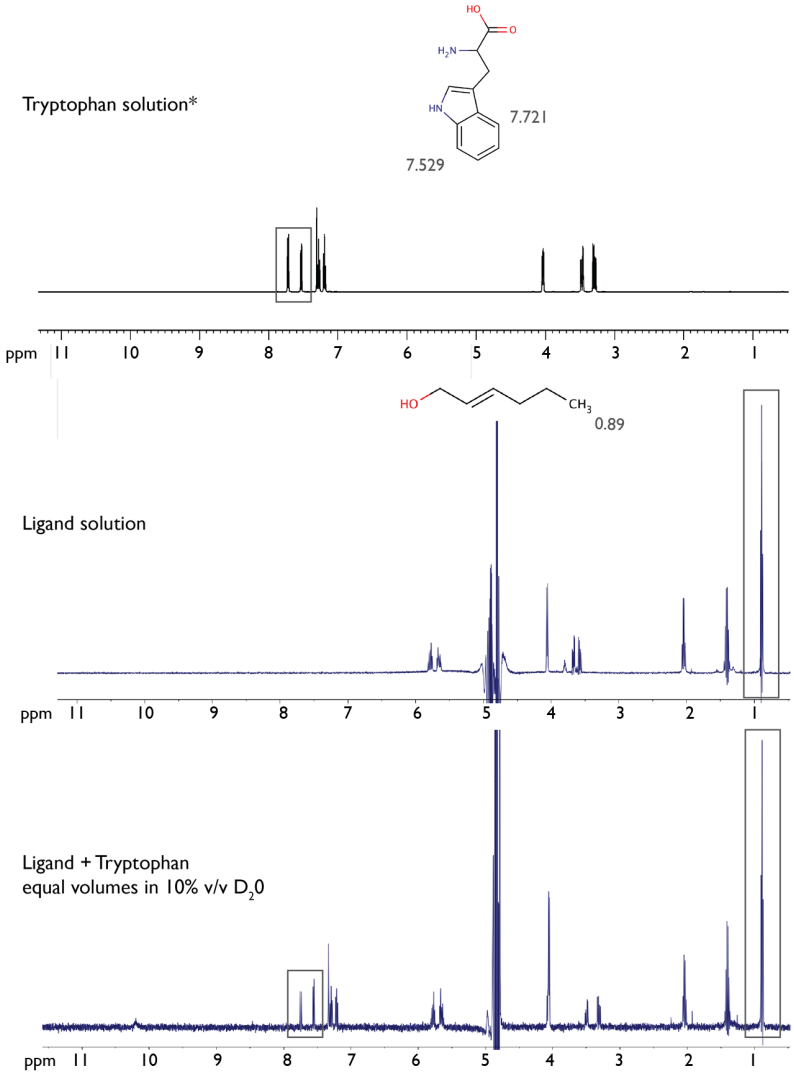


Figure 3.2 ^{1}H NMR measurements of alkenol ligand concentration using a tryptophan solution. Molecular structures of tryptophan (top) and t26 (middle) are shown. ^{1}H resonance frequencies are indicated next to the relevant proton positions in the structures, and correspond to the boxed signals in the spectra. *^{1}H assignment for L-Tryptophan was obtained from the Biological Magnetic Resonance Databank

(BMRB).[76] The ratio of the indicated peak integrals from each species was measured to determine alkenol concentration for ITC experiments. A 60 s delay was used between scans to ensure the system had returned fully to equilibrium and therefore avoid relaxation artefacts in the peak intensities. See §3.2.1.1 for further details.

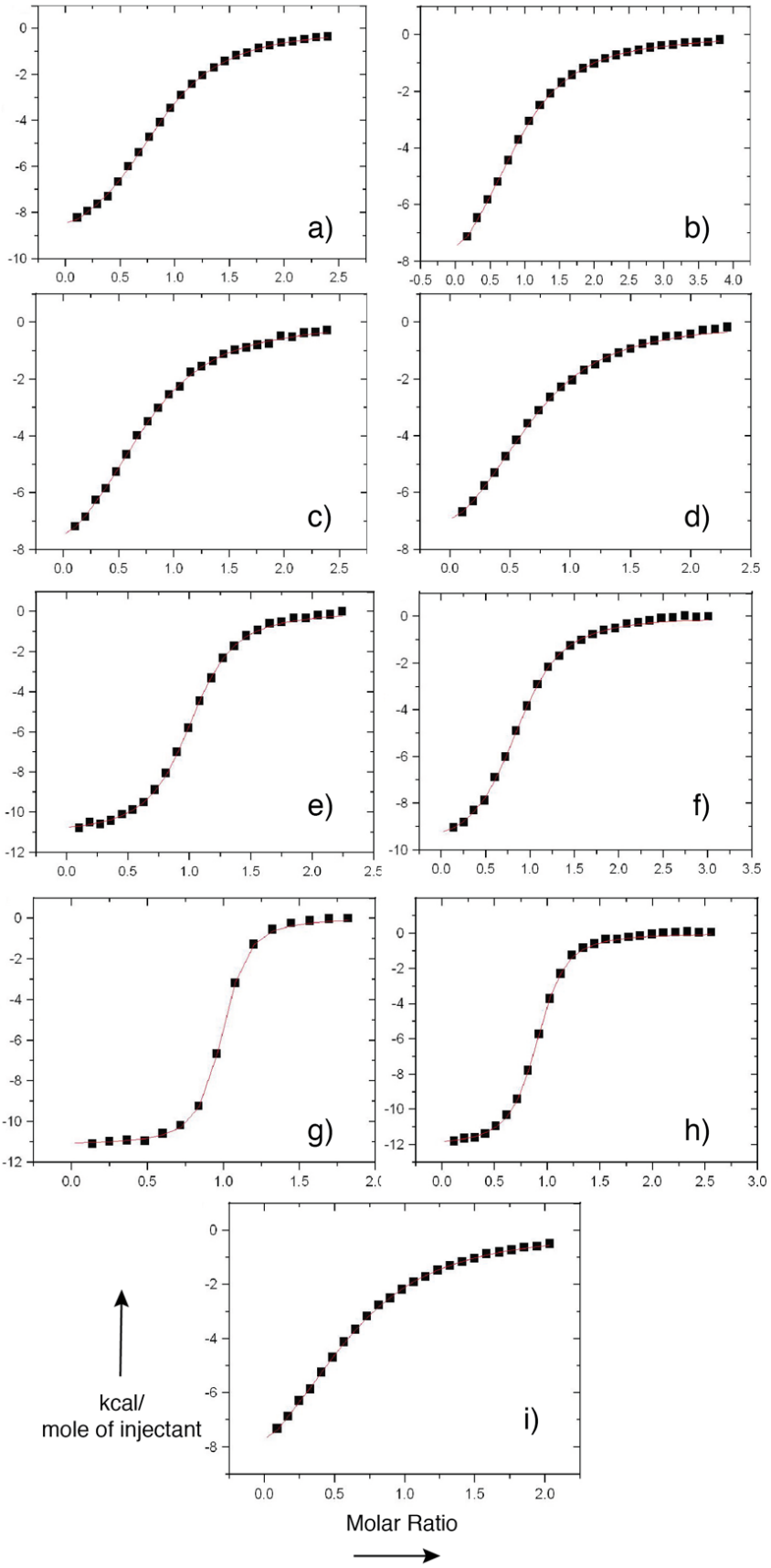


Figure 3.3 Representative ITC isotherms for each ligand binding to MUP. a) t26, b) c26, c) t36, d) c36, e) t27, f) c37, g) t28, h) c38, i) 56. Isotherms for ligands 6, 7 and 8 are displayed in reference [28].

3.2.2 X-Ray Crystallography

Crystallisation was performed by R Malham. Dr Chi Trinh performed all ligand soaking, data collection and data processing. Accordingly, methods for §3.2.2.2 and §3.2.2.3 were provided by Dr Chi Trinh. Ligand soaking protocols were provided by Dr Caitriona Dennis.

3.2.2.1 Crystallisation

Optimal crystallisation conditions were previously identified; therefore crystals were obtained using reservoir solutions containing 45 to 70 mM $CdCl_2$ and 0.1 M malate at pH 4.6 to 5.2. Using the hanging drop method, drops containing 1 µL of MUP solution (10 mg/mL) and 1 to 2 µL of reservoir solution were equilibrated against 1 mL reservoir solution. Crystals with space group $P4_32_12$ grew within 2 days at room temperature. Ligand soaks were performed by immersing crystals in a 2 µL drop of reservoir solution before adding 1 µL of 30 mM aqueous ligand solution. Crystals were then equilibrated with the ligand for ~30 seconds. Prior to flash freezing in liquid nitrogen the crystals were transferred to another 2 µL drop, comprising reservoir solution with 1 µL of 30 mM aqueous ligand and 25% (v/v) PEG 400 as a cryoprotectant.

3.2.2.2 Data collection and processing

All MUP-ligand complex data were recorded at 100 K and gave resolutions between 1.3 and 1.6 Å from a single crystal on the macromolecular crystallography beamline stations I02 and I03 at Diamond Light Source. For all data recorded, the diffraction images were integrated using MOSFLM[89], then scaled and reduced using SCALA[90] and CTRUNCATE[91] from the CCP4 program suite[91]. Five percent of the reflections were selected and excluded from the refinement using the program FREERFLAG[92] and constituted the R_{free} set. The R_{free} set from the previously recorded data for the MUP structure 1ZNH was used.[1] The data processing statistics are shown in Table 3.1.

3.2.2.3 Structure determination

All the crystal structures of MUP ligand complexes were determined using the Direct Fourier Transform method with the structure of a previously deposited MUP structure as a starting model (PDB accession code 1ZNH[1]). After initial rounds of rigid body and restrained refinement using REFMAC5[93], iterative manual model building and refinement were carried out using COOT[94] and REFMAC5. The initial coordinates and restraint library file for the ligands were obtained from the PRODRG server[95]. All the ligands were manually fitted into both $2F_o$-F_c and F_o-F_c maps in the program COOT. Water molecules were manually added in COOT for peaks over 3.5σ in the F_o-F_c map, and where appropriate hydrogen bonds could be made to surrounding residues or other water molecules. Refinement was judged complete when the R factor had converged and no significant interpretable features remained in the F_o-F_c map. Structure validations were carried out with MOLPROBITY[96]. For all the MUP-ligand complexes, residue 1 of the N terminus is missing from the final structure due to poorly connected electron density for this region. All refinement statistics are shown in Table 3.1.

3.2.2.4 Crystal Structure Analysis

RMSD values between structures of MUP-ligand complexes were calculated using the Pymol 'align' command. All atoms in residues 1 to 155 were specified for alignment, and the default iterative outlier removal was disabled to avoid artificially low RMSD values being returned.

B-factors were compared between complexes after adjusting each dataset as described by Ringe and Petsko.[97] Briefly, of those structures to be compared, that with the highest overall B-factor is identified, and its lowest B-factor noted. B-factors of all other structures are scaled such that their lowest B-factor matches this value.

The number of water molecules in the binding pocket was measured by visual inspection of the final structure.

Ligand Molecular Surface Area (MSA) calculations were performed by Dr J Clements at the University of Texas at Austin. Connolly or molecular surface areas[98] were calculated for the ligands in their bound conformations using Macromodel v9.1 (Schrödinger, LLC, New York, NY, 2007).

R^2 values were calculated using the Microsoft Excel RSQ function.

	MUP-T26	MUP-T27	MUP-T28	MUP-C36	MUP-C37	MUP-C38
Diamond beamline station	I02 23/10/10	I03 05/12/10	I02 23/10/10	I03 05/12/10	I03 05/12/10	I03 05/12/10
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
a (Å)	53.3	53.5	53.3	53.7	53.4	53.5
b (Å)	53.3	53.5	53.3	53.7	53.4	53.5
c (Å)	137.4	137.6	137.6	137.6	137.6	137.6
Resolution (Å)	49.69-1.80	42.23-1.30	37.67-1.60	21.27-1.30	11.56-1.30	13.66-1.30
R_{merge} (%)§*	8.8 (33.8)	7.1 (37.1)	7.0 (31.2)	6.4 (34.5)	6.8 (41.1)	6.3 (35.5)
R_{pim} (all I^+ & I^-) (%)+*	3.6 (13.4)	2.8 (14.4)	2.6 (10.7)	2.2 (11.3)	2.3 (13.4)	3.3 (19.0)
Observed reflections	137,325	368,624	222,624	498,181	472,585	211,581
Unique reflections	19,174	49,611	27,077	49,948	48,914	49,558
Completeness (%)*	99.8 (100.0)	99.0 (100.0)	99.9 (100.0)	99.0 (100.0)	98.3 (97.7)	99.2 (99.7)
Multiplicity *	7.2 (7.3)	7.4 (7.6)	8.2 (9.4)	10.0 (10.2)	9.7 (10.0)	4.3 (4.4)
$\langle I \rangle / \sigma \langle I \rangle$ (I/σ)*	13.0 (4.7)	14.9 (5.0)	16.3 (5.4)	26.0 (6.1)	18.6 (5.8)	12.0 (3.7)
Refinement						
R_{factor} (%)	18.8	15.8	17.6	16.1	15.8	16.1
R_{free} (%) †	20.9	17.6	20.7	17.2	17.1	17.6
No. of protein atoms	1304	1304	1304	1304	1304	1304
No. of solvent molecules	169	169	168	169	169	169
No. of ligand atoms	7	8	9	7	8	9
Average overall *B*-factor (Å^2)	23.0	16.8	18.1	16.7	15.8	16.7
RMS bond lengths (Å) ξ	0.011	0.011	0.013	0.010	0.010	0.011
RMS bond angles (°) ξ	1.2	1.4	1.4	1.4	1.4	1.4
Ramachandran analysis, the percentage of residues in the regions of plot (%) ‡						
Most favoured	97.4	96.8	97.4	97.4	96.8	96.8
Outliers	0	0	0	0	0	0

* Values given in parentheses correspond to those in the outermost shell of the resolution range.

§ $R_{merge} = \sum_{hkl} \sum_i \left| I_i(hkl) - \langle I(hkl) \rangle \right| / \sum_{hkl} \sum I_i(hkl)$

+ *Rpim* - precision-indicating (multiplicity-weighted) *Rmerge*.

† R_{free} was calculated with 5% of the reflections set aside randomly.

ξ Based on the ideal geometry values of Engh & Huber (1991).

‡ Ramachandran analysis using the program MolProbity (Lovell *et al.,* 2003).

Table 3.1 Crystallographic data collection and refinement statistics, provided by Dr Chi Trinh.

3.3 Results and Discussion

3.3.1 Changes in observed thermodynamic parameters

Previous ITC (300 K, pH 7.4) experiments from 2005 investigated MUP binding to primary aliphatic alcohols from pentanol (C5) to nonanol (C9). Affinity was found to decrease incrementally by 3.9 kJ mol^{-1} with each methylene removed, spanning a range of -38.8 kJ mol^{-1} (C9) to -23.1 kJ mol^{-1} (C5). This was due to a less favorable $\Delta\Delta H°_{obs}$ of 5.6 kJ mol^{-1} per methylene that dominated more favourable $T\Delta\Delta S°_{obs}$ of 1.7 kJ mol^{-1}.[1]

For this work the observed parameters of binding, $\Delta\Delta G°_{obs}$, $\Delta\Delta H°_{obs}$ and $T\Delta\Delta S°_{obs}$, were measured using ITC at 298 K and pH 7.4 for all ligands in Figure 3.1. Data for 6, 7 and 8, had previously been measured.[28] Tables 3.2 and 3.3 show the observed parameters and the change in observed parameters calculated using Equation 3.1. Table 3.3 also shows average changes of each parameter grouped by type of modification and restrictions grouped by isomer type.

Changes in thermodynamics parameters resulting from methylene removal, averaged over all series, are the same within error as those previously reported. A small entropic benefit of 0.9 kJ mol^{-1} is overwhelmed by an enthalpic penalty of 5 kJ mol^{-1} resulting in an unfavourable change in affinity of 4.1 kJ mol^{-1}.

The introduction of an internal double bond displays a similar trend of enthalpy-entropy compensation resulting in an unfavourable change to affinity. However there is a less unfavourable affinity change, 1.9 kJ mol^{-1}, compared to methylene removal. This is because of a lessened entropy-enthalpy offset, which is slightly more pronounced for trans isomers due to slightly lower enthalpic penalties. It is noteworthy that compared to methylene removal, introduction of an internal double bond results in a much larger entropic benefit of 6.9 kJ mol^{-1}, and a slightly larger enthalpic penalty of about 8.8 kJ mol^{-1}. Introducing a terminal double bond has a distinct entropic effect, closer to that of methylene removal, also seen in the intrinsic values.

Ligand	Number of experiments	ΔG°_{obs} (kJ/mol)	ΔH°_{obs} (kJ/mol)	$T\Delta S^\circ_{obs}$ (kJ/mol)	ΔG°_{h} (kJ/mol)	ΔH°_{h} (kJ/mol)	$T\Delta S^\circ_{h}$ (kJ/mol)
6*		-28.5 ± 0.3	-49.3 ± 1.1	-20.8 ± 1.1	-18.2 ± 0.5	-66.1 ± 1.5	-48.0 ± 1.6
t26	4	-27.3 ± 0.0	-44.3 ± 1.8	-17.0 ± 1.8	-22.0 ± 0.5	-65.4 ± 1.5	-43.4 ± 1.6
c26	2	-25.8 ± 0.0	-40.4 ± 1.4	-14.7 ± 0.1	-22.0 ± 0.5	-65.4 ± 1.5	-43.4 ± 1.6
t36	2	-26.4 ± 0.1	-39.4 ± 1.1	-13.0 ± 0.2	-22.0 ± 0.5	-65.4 ± 1.5	-43.4 ± 1.6
c36	2	-26.0 ± 0.1	-39.0 ± 1.7	-13.1 ± 0.4	-22.0 ± 0.5	-65.4 ± 1.5	-43.4 ± 1.6
56	2	-25.6 ± 0.1	-45.5 ± 1.5	-19.9 ± 0.3	-21.0 ± 0.5	-65.3 ± 1.5	-44.3 ± 1.6
7*		-33.1 ± 0.2	-56.2 ± 0.4	-23.1 ± 0.4	-17.5 ± 0.5	-69.9 ± 1.5	-52.4 ± 1.6
t27	15	-34.3 ± 2.1	-53.0 ± 3.9	-18.7 ± 1.9	-21.3 ± 0.5	-69.2 ± 1.5	-47.9 ± 1.6
c37	5	-30.1 ± 0.1	-41.8 ± 0.2	-11.8 ± 0.3	-21.3 ± 0.5	-69.2 ± 1.5	-47.9 ± 1.6
8*		-36.9 ± 0.6	-60.5 ± 0.5	-23.6 ± 0.8	-16.7 ± 0.5	-73.7 ± 1.5	-56.9 ± 1.6
t28	16	-35.5 ± 0.2	-52.6 ± 1.3	-17.1 ± 1.3	-20.6 ± 0.5	-72.9 ± 1.5	-52.4 ± 1.6
c38	4	-33.7 ± 0.1	-49.4 ± 0.6	-15.7 ± 0.6	-20.6 ± 0.5	-72.9 ± 1.5	-52.4 ± 1.6

Table 3.2 Thermodynamics parameters: observed by ITC upon ligand binding (obs) and calculated desolvation parameters (h). Ligand names are the abbreviations from Figure 3.1. All values are in kJ mol^{-1} and to 1.d.p. Values denoted with a * are reproduced from reference [28]. Observed parameters are from repeat experiments. Desolvation parameters were calculated as described above. See §3.2.1.2 for calculation of errors.

3.3.2 Changes in intrinsic thermodynamic parameters

The difference between observed and intrinsic values is simply the change in ligand desolvation thermodynamics upon modification. This is represented by the central horizontal arrows and accompanying values displayed in Figure 3.4. The type of modification results in significantly different changes to ligand desolvation. Though both modifications result in equally unfavourable changes to desolvation entropy, removal is preferable to restriction, due to a preferable desolvation enthalpy. Ultimately, introduction of a double bond incurs a greater overall desolvation penalty than removal of a methylene.

Calculations were performed and error analysis carried out as detailed in §3.2.1.2. Desolvation and intrinsic parameters are displayed in Tables 3.2 and 3.3 respectively, and graphed in Figure 3.4. The values discussed in this section may contain contributions from differential complex solvation or differential protein dynamics, which are considered in §3.3.3.

Both types of modification follow a similar pattern to that described in §3.3.1, namely that entropic benefits are offset by enthalpic penalties resulting in modest changes to affinity. Nonetheless, intrinsic affinity is improved by 2.0 kJ mol^{-1} with the introduction of an internal double bond. However, this is not translated into observed affinity gains, due to the critical impact of desolvation. As displayed in Figure 3.4, introducing a double bond results in a 3 to 4 kJ mol^{-1} penalty to the free energy change of ligand desolvation.

Modification	$\Delta\Delta G°_{obs}$	$\Delta\Delta H°_{obs}$	$T\Delta\Delta S°_{obs}$	$\Delta\Delta G°_{i}$	$\Delta\Delta H°_{i}$	$T\Delta\Delta S°_{i}$
8→7	3.8 ± 0.6	4.3 ± 0.6	0.6 ± 0.9	3.0 ± 0.9	8.1 ± 2.2	5.0 ± 2.4
7→6	4.5 ± 0.4	6.9 ± 1.1	2.3 ± 1.2	3.8 ± 0.8	10.6 ± 2.4	6.8 ± 2.5
t28→t27	1.2 ± 2.1	-0.4 ± 4.1	-1.6 ± 2.3	0.5 ± 2.2	3.4 ± 4.6	2.9 ± 3.2
t27→t26	7.0 ± 2.1	8.7 ± 4.3	1.6 ± 2.7	6.3 ± 2.2	12.4 ± 4.8	6.1 ± 3.5
c38→c37	3.7 ± 0.2	7.6 ± 0.7	3.9 ± 0.7	3.0 ± 0.7	11.4 ± 2.2	8.4 ± 2.3
c37→c36	4.1 ± 0.2	2.8 ± 1.7	-1.3 ± 0.5	3.4 ± 0.7	6.5 ± 2.7	3.2 ± 2.3
6→t26	1.3 ± 0.3	5.0 ± 2.1	3.7 ± 2.1	-2.6 ± 0.7	5.7 ± 3.0	8.3 ± 3.1
6→c26	2.8 ± 0.3	8.9 ± 1.8	6.1 ± 1.1	-1.1 ± 0.7	9.6 ± 2.8	10.7 ± 2.5
6→t36	2.1 ± 0.3	9.9 ± 1.5	7.8 ± 1.1	-1.7 ± 0.7	10.6 ± 2.6	12.4 ± 2.5
6→c36	2.6 ± 0.3	10.3 ± 2.0	7.7 ± 1.2	-1.3 ± 0.7	11.0 ± 2.9	12.3 ± 2.5
7→t27	-1.2 ± 2.1	3.2 ± 3.9	4.4 ± 2.0	-5.1 ± 2.2	3.9 ± 4.4	9.0 ± 3.0
7→c37	3.0 ± 0.2	14.3 ± 0.4	11.3 ± 0.5	-0.8 ± 0.7	15.1 ± 2.2	15.9 ± 2.3
8→t28	1.3 ± 0.6	7.9 ± 1.4	6.6 ± 1.5	-2.6 ± 0.9	8.6 ± 2.5	11.1 ± 2.7
8→c38	3.1 ± 0.6	11.0 ± 0.8	7.9 ± 1.0	-0.7 ± 0.9	11.8 ± 2.3	12.5 ± 2.5
6→56	3.0 ± 0.3	3.8 ± 1.8	0.9 ± 1.2	0.1 ± 0.7	4.6 ± 2.8	4.5 ± 2.5
Grouped by type of modification						
<Remove>	4.1 ± 0.5	5.0 ± 1.1	0.9 ± 0.7	3.3 ± 0.6	8.7 ± 1.4	5.4 ± 1.1
<Restrict>	1.9 ± 0.3	8.8 ± 0.7	6.9 ± 0.5	-2.0 ± 0.4	9.5 ± 1.0	11.5 ± 0.9
Restrict modifications grouped by isomer						
<Trans>	0.9 ± 0.6	6.5 ± 1.2	5.6 ± 0.9	-3.0 ± 0.7	7.2 ± 1.6	10.2 ± 1.4
<Cis>	2.9 ± 0.2	11.1 ± 0.7	8.3 ± 0.5	-1.0 ± 0.4	11.9 ± 1.3	12.8 ± 1.2

Table 3.3 Changes in observed and intrinsic thermodynamic parameters, and means grouped according to modification type. The 'modification' column uses the abbreviations from Figure 3.1, repeated in Figure 3.4. 6, 7 and 8 values taken from reference [28]. Observed and intrinsic values were calculated, and errors propagated, using Equation 3.1 and data in Table 3.2. Group means do not include values for ligand '56'. All values are in kJ mol^{-1} and to 1 d.p. The same data are displayed graphically in Figure 3.4.

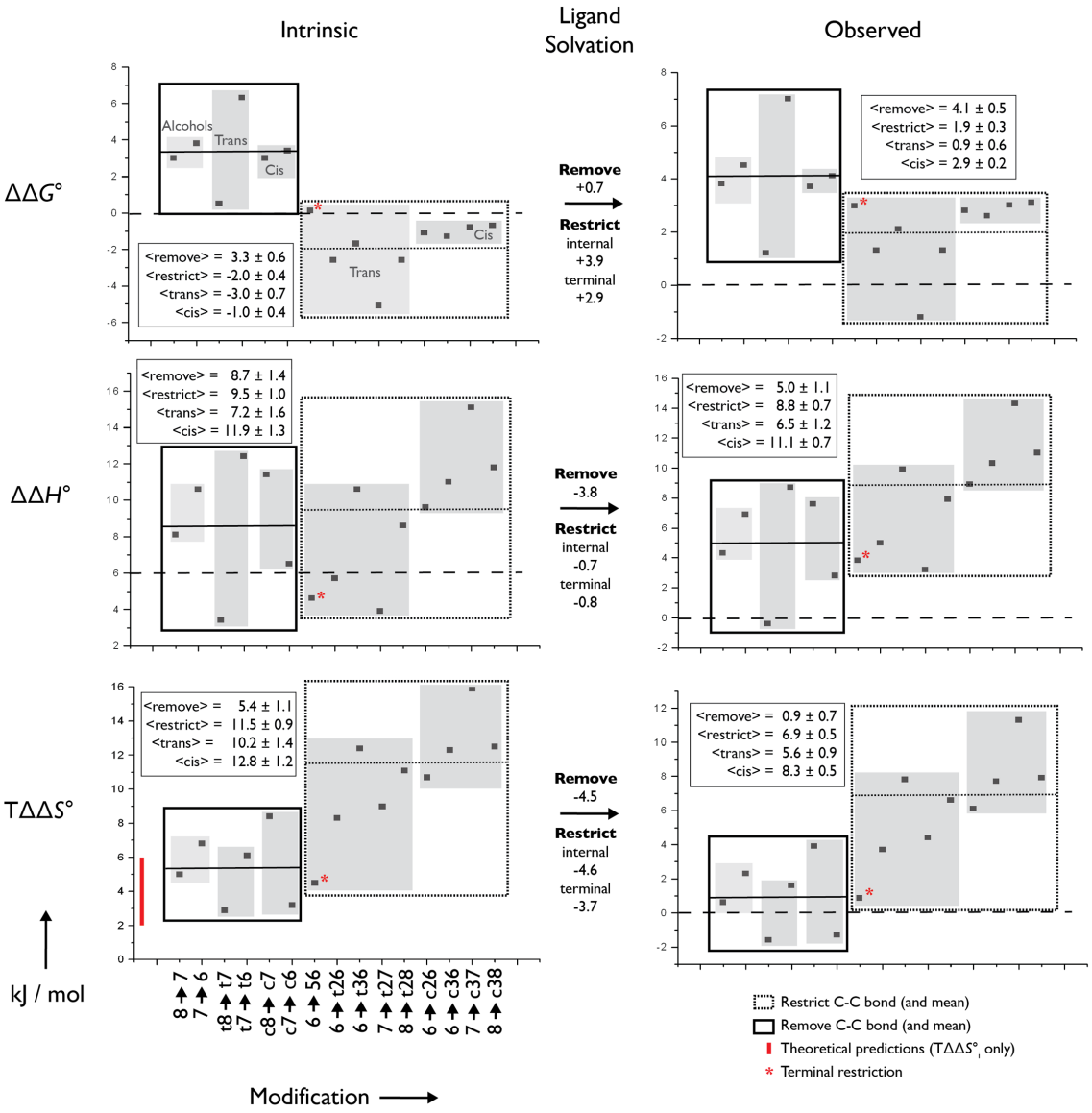


Figure 3.4 Intrinsic and observed values of $\Delta\Delta G°$, $\Delta\Delta H°$ and $T\Delta\Delta S°$, and group means. Values, from Table 3.3, are grouped by modification (solid and dashed boxes), and different isomers are indicated (grey squares). The horizontal line in each box is the group mean. Group means do not include values for 56, which is indicated with an asterisk. Zero is indicated by a dashed line. Central arrows and values indicate the calculated desolvation parameters. All values are in kJ mol^{-1} and to 1.d.p. Abbreviations from Figure 3.1 are used: hexanol (6), heptanol (7), octanol (8), trans-2-hexenol (t26), trans-3-hexenol (t36), trans-2-heptenol (t27), trans-2-octenol (t28), cis-2-hexenol (c26), cis-3-hexenol (c36), cis-3-heptenol (c37), cis-3-octenol (c38), 5-hexenol (56).

Both types of modification incur a significant ~9 kJ mol^{-1} $\Delta\Delta H°_{i}$ penalty on average, as observed previously for methylene removal.[1] Unlike previously, individual methylene

removal values span a wide range of 3.2 to 12.4 kJ mol^{-1}, variation that is unexpected in a system dominated by weak, non-directional, non-polar interactions.

A $T\Delta\Delta S°_i$ of 5.5 kJ mol^{-1} per methylene removed was previously suggested for saturated alcohols in this system at 300 K.[1] The average $T\Delta\Delta S°_i$ value for methylene removal is the same within error, 5.4 kJ mol^{-1}, and lies within the range of theoretical predictions. Introducing an internal double bond, i.e. any bond with two adjacent C-C bonds, results in significantly greater $T\Delta\Delta S°_i$ than methylene removal, on average 11.5 kJ mol^{-1}. Contrastingly, introducing a terminal double bond results in a $T\Delta\Delta S°_i$ of 4.5 kJ mol^{-1}, in the range of the methylene removal values. This suggests that the $T\Delta\Delta S°_i$ values resulting from methylene removal are indeed due to the loss of a rotatable bond, and that the higher values for internal double bonds include additional entropic effects.

It has been calculated that the effect of restricting or removing rotors on $T\Delta\Delta S°_i$ is more due to changes in the vibrational (torsional fluctuations within a conformation) than rotameric component of the entropy: in fact it has been proposed that up to 90% of the change arises from the vibrational component.[99] However, these calculations were performed for inhibitor binding to HIV protease, which is not dominated by weak, non-polar, non-directional interactions like MUP. Adopting a vibrationally restrictive single bound conformation is less necessary in MUP. Analysis of crystallographic electron density (discussed in §3.3.3) supports this, indicating distinct ligand flexibility in the pocket for 'restrict' ligands.

3.3.3 Crystallographic analysis to assess potential for contributions from differential complex solvation and protein dynamics

An appreciation of contributions from differential complex solvation and protein dynamics to the intrinsic values may indicate the requirement for further work. This is achieved by observing changes in crystallographically-derived parameters and calculating their correlation to changes in intrinsic thermodynamic parameters, as detailed in §3.2.2.4. As $\Delta\Delta G°_i$ is simply a function of $\Delta\Delta H°_i$ and $T\Delta\Delta S°_i$, correlations were only calculated for the latter two parameters. Modification type was also correlated to these parameters. A comparison of how much variance is explained by

changes in crystallographic observables as opposed to the type of modification indicates the degree to which trends in the reported intrinsic parameters can be simply related to the type of modification. Crystal structures were obtained for the complexes marked in Figure 3.1, to a resolution of < 1.5 Å, as detailed in §3.2.2.3. Crystal structures for 6, 7 and 8 were previously published.[1]

Modification	Δ Bound water	Δ <B-factor>	Δ Global RMSD	Δ Ligand MSA	Modification type
8→7	2	-1.70	0.32	-18.30	1
7→6	0	0.95	0.68	-14.80	1
t28→t27	1	0.76	0.11	-17.75	1
t27→t26	0	2.29	0.14	-18.25	1
c38→c37	0	-0.93	0.03	-16.50	1
c37→c36	0	-0.10	0.05	-15.35	1
6→t26	-1	-0.53	0.56	-5.60	2
7→t27	-1	-5.93	0.61	-2.15	2
8→t28	0	-2.05	0.63	-2.70	2
6→c36	-1	-3.41	0.61	-6.90	2
7→c37	-1	-7.07	0.61	-6.35	2
8→c38	1	-3.54	0.63	-8.15	2
R^2					
$\Delta\Delta H^\circ_i$	0.01	0.04	0.03	0.00	0.01
$T\Delta\Delta S^\circ_i$	0.23	0.58	0.47	0.51	0.65

Table 3.4 Changes in crystallographically-derived observations with each ligand modification, and correlation to corresponding values in Table 3.3 (which are not reproduced in this Table). All crystal structures have resolution < 1.5 Å. Abbreviations from Figure 3.1, listed again in Figure 3.4, are used in the 'modification' column. MSA is ligand molecular surface area. See §3.2.2.4 for calculation of values and R^2.

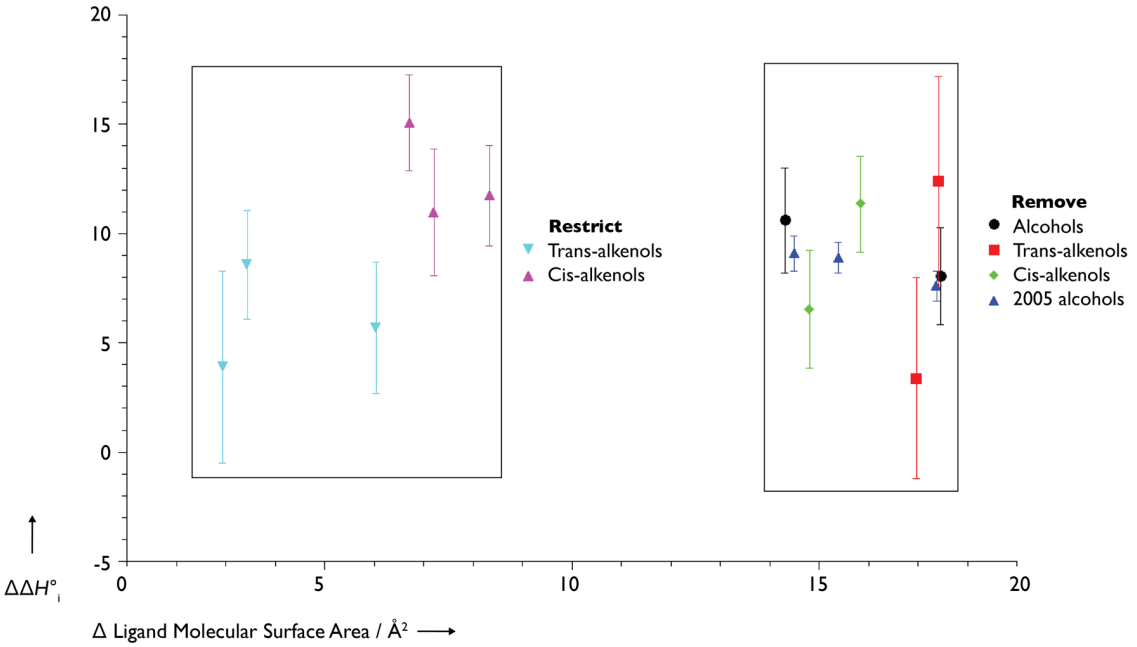


Figure 3.5 Absolute change in bound ligand molecular surface area (ΔMSA) versus change in intrinsic enthalpy of binding ($\Delta\Delta H°_i$) for all crystallographically-analysed ligand modifications. Data and errors are tabulated in Tables 3.3 and 3.4. Alcohols data are from reference [28]. '2005 alcohols' data, acquired at 300 K, are taken from reference [1].

Differential complex solvation was evaluated by comparing the number of high-residency water molecules in the MUP binding pocket for each complex, as discussed in §2. The resolution of these structures is high enough to resolve high-residency static water molecules (below 2 Å). However, highly dynamic water molecules that could have important energetic contributions are currently experimentally unobservable, and therefore their potential contribution is an important caveat. An estimation of differential protein dynamics was performed by calculating both RMSD values and differences in protein all-atom B-factor averages between complexes. Though crystallographic B-factors can correlate to solution dynamics[100], comparing these values between different data sets is non-trivial. The magnitudes of these values are influenced by multiple factors, including crystal defects and differential processing. Consequently there is no definitive method for performing such comparisons, and any results need to be treated with caution. The values reported here arise from datasets which were adjusted according to the method of Ringe *et al.*[97]

Differences were seen in these crystallographic observables, Table 3.4. None of these changes correlated to the intrinsic enthalpy changes ($R^2 < 0.05$), and therefore do not account for the wide range of values observed. The MUP-alcohol experiments

from 2005 show a narrow range of $\Delta\Delta H°_i$ across four modifications, 7.6 to 9.7 kJ mol^{-1}, attributed to a change in ligand molecular surface area, ΔMSA, interacting with the MUP pocket.[1] This work shows no correlation between change in ΔMSA, calculated from a single conformation in the binding pocket, and $\Delta\Delta H°_i$ using either the whole dataset or just the 'remove' values, as explored below.

Figure 3.5 shows absolute ΔMSA of the bound ligand versus $\Delta\Delta H°_i$, Table 3.4, for those modifications analysed crystallographically, Figure 3.1, including the data from reference [1]. For the latter data, the hexanol to pentanol modification is discounted because two pentanol molecules bind in the pocket. The absolute ΔMSA for the nonanol to octanol transition was calculated as 16 Å^2 using UCSF Chimera.[101] Focusing on the 'remove' values initially, $\Delta\Delta H°_i$ values for methylene removal in saturated alcohols (black circles) sit close to the narrow spread of for the 2005 data (blue triangles). However, values for the two unsaturated series (green diamonds and red squares) diverge from these values and greatly widen the data range: this is greatest for trans-alkenols, which display the smallest and largest $\Delta\Delta H°_i$ values for methylene removal. The small range of ΔMSA probed through methylene removal may preclude observation of any correlation. Nonetheless, all values are consistent with the 2005 range within error, perhaps suggesting a standard $\Delta\Delta H°_i$ for removing a methylene. Restrict values also incur a similar $\Delta\Delta H°_i$ penalty, revealing that removing the entire methylene surface is enthalpically equivalent to removing two H atoms and introducing a polarisable bond. That the penalty is higher for cis than trans alkenols may reflect the relatively increased strain of a cis isomer. The equivalence of $\Delta\Delta H°_i$ with modification shows that a simple correlation to ΔMSA is misleading because it ignores the equally important chemical nature of the ligand surface.

Correlations of $T\Delta\Delta S°_i$ to the same crystallographic observables are much higher. The strongest correlation is with modification type (R^2 = 0.65). There is only weak correlation (R^2 = 0.23) to differential complex solvation, meaning that explicitly accounting for differential solvation could adjust the exact magnitudes of individual data points but would not affect trends in the data. However, differential protein dynamics upon ligand modification appear to explain almost as much of the variance (R^2 = 0.51, 0.58) in $T\Delta\Delta S°_i$ as the type of modification. Though the B-factor values must be treated with caution, this result clearly indicates that the reported intrinsic entropy changes contain potentially significant contributions from differential protein dynamics in addition to changes in ligand conformational entropy.

Furthermore, ligand electron density for alkenol complexes is less defined than for their alcohol counterparts, indicating that introduction of an internal double bond increases the number of ligand binding configurations. Though this observation does not affect the methylene removal data, an increase in ligand conformational entropy in the bound state could explain the additional $T\Delta\Delta S°_i$ benefit of restriction above and beyond removal.

3.4 Summary and Conclusions

This is the first systematic experimental evaluation, in a protein-ligand system, of the suggested entropic and affinity benefits of decreasing rotor numbers in a ligand molecule. The work also evaluates two possible modifications for decreasing rotor numbers, both removal and restriction of C-C bonds. The necessary tolerance to ligand modification and a thermodynamic decomposition approach identified MUP as suitable system. Observed and intrinsic thermodynamics of binding were obtained using measured ITC and calculated desolvation parameters. Crystal structures were obtained and analysed to appreciate potential contributions from differential complex solvation and protein factors to the intrinsic parameters.

Analysis of crystal structures indicates a contribution from differential protein dynamics to intrinsic entropy changes. This would usually prompt further investigation using NMR relaxation experiments, however in this system errors are too high with this technique to be beneficial. A weak correlation between intrinsic parameters and differential complex solvation was found.

Previously predicted penalties for rotamerically 'freezing' a single ligand C-C bond upon binding were observed as an average 5.4 kJ mol^{-1} $T\Delta\Delta S°_i$ benefit from methylene removal across multiple ligand series. The introduction of a terminal double bond gives a similar result of 4.5 kJ mol^{-1}, vindicating methylene removal as effectively the deletion of a rotor in terms of intrinsic entropy. However, a significantly higher $T\Delta\Delta S°_i$ benefit of 11.5 kJ mol^{-1} is obtained by introducing an internal double bond. This may arise from an increase in the number of bound configurations of the ligand upon introducing an internal double bond, an effect that relies on the specific

architecture and size of the pocket. These gains are considerable in terms of ligand design, where 5.7 kJ mol^{-1} represents an order of magnitude in affinity.

In the case of this system these considerable entropic gains are compensated by unfavourable enthalpic and solvation contributions resulting in an unfavourable effect on affinity. Desolvation offsets to these gains are pronounced for restrictions, due to a large enthalpic penalty of desolvation.

Despite being more comprehensive than other work in the field, there is still too little data to identify conclusive and reliable trends. Overall, these data support predictions of a $T\Delta\Delta S°_i$ benefit of ~5.4 kJ mol^{-1} from removing a rotor. However, in hydrophobic alkyl chains that undertake nonpolar interactions with the protein, this modification also incurs desolvation and intrinsic enthalpy penalties that result in an unfavourable effect on affinity. Introduction of an internal double bond has a pronounced $T\Delta\Delta S°_i$ benefit, possibly due to a combination of both rotor restriction and an increase in bound configurations of ligand. This pronounced benefit is mostly offset by an increased intrinsic enthalpy penalty. Restriction modifications also incur a greater desolvation penalty than methylene removal, resulting here in an unfavourable affinity, albeit less unfavourable than for methylene removal.

Reproducing the enhanced benefits of restriction over removal observed in this system will be unlikely if this effect does arise from an increase in the flexibility of bound ligand, because most other systems will not have similarly spacious and non-specific binding pockets. However, introducing double bonds may bear fruit in certain disease-relevant systems: previous work showed a 10-fold improvement in binding affinity of a peptide ligand to a protein receptor due to a single internal bond restriction like those reported here[102].

Chapter 4

Ligand rotational and translational entropy

4.1 Introduction

4.1.1 The biggest penalty to binding entropy

It is regularly stated that the most prominent entropic penalty in protein-ligand interactions is that of decreased translational and rotational entropy of the ligand. Though these losses are partially replaced by vibrations in the bound state, they have been previously calculated as ~56 kJ mol^{-1} in solution at 25 °C, with a range of -40 to -60 kJ mol^{-1} for many small organic molecules.[5] These predictions arise from a statistical-mechanical approach that extrapolates from gas phase, resulting in very high entropy for the ligand in its free state. However, recent analysis of the effect in the literature, one paper of which utilised a decomposed MUP interaction for experimental comparison[103], demonstrates that models which consider free ligand to be constrained by solvent result in predictions closer to experimental estimates.[104]

Either type of model requires accurately accounting for ligand translation and rotation when bound. Enzyme-substrate binding requires that ligands be held in a single bound conformation (binding mode) to enable a specific chemical reaction. Furthermore, computational, crystallographic and spectroscopic data indicate that most interactions still have either a single or a small number of defined binding modes.[56] Therefore in many cases it is assumed that a (small) ligand loses all, or almost all, of its rotational and translational degrees of freedom upon binding to a (much larger) protein. However, because thermal energy can be on the same order as the energy of non-covalent interactions, the ligand bound state may have a degree of residual freedom.[105] Consequently residual motion, including the volume the bound ligand can access, is sometimes estimated using computational methods to assess relative motion of ligand and protein in the complex.

It is very difficult to experimentally measure and confirm predictions of ligand rotational and translational entropy loss upon binding, which can vary from 56 down to

~5 kJ mol^{-1}.[106] Until now, direct experimental assessment of predicted residual motion has never been attempted, due to a combination of practical difficulty and an expectation that any such motion would be below experimental resolution.

4.1.2 Loss of ligand rotational and translational entropy upon binding to MUP

Two MUP ligands have been sufficiently studied experimentally that decomposition has estimated a value for the loss of ligand rotational and translational entropy. These are 2-methoxy-3-isopropyl-pyrazine (IPMP) and 2-methoxy-3-isobutyl-pyrazine (IBMP), displayed below.

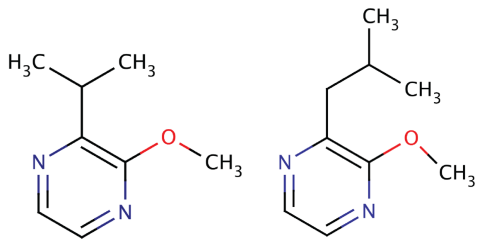


Figure 4.1 Chemical structures of 2-methoxy-3-isopropyl-pyrazine (IPMP) and 2-methoxy-3-isobutyl-pyrazine (IBMP).

Ligand	$\Delta G°_b$ (kJ/mol)	$\Delta H°_b$ (kJ/mol)	$T\Delta S°_b$ (kJ/mol)	Loss of ligand rot. and trans. entropy (kJ/mol)
IPMP	-33.9 ± 0.28	-44.5 ± 0.4	-10.7 ± 0.5	-25
IBMP	-38.5 ± 0.86	-47.9 ± 0.9	-9.4 ± 0.9	-27 to -78

Ligand	<Protein B-factor>	<Ligand B-factor>
IPMP	28.91	28.60
IBMP	22.04	23.89

Table 4.1 Thermodynamic parameters from ITC at 308 K, and B-factors from crystal structures, for MUP interactions with IPMP and IBMP. Data reproduced from references [23] and [18]. Average B factors include all atoms in relevant segments, taken from pdb files 1QY1 (IBMP) and 1QY2 (IPMP).

IBMP binds more favourably to MUP, mainly due to a more favourable enthalpy of binding that may arise from burial of an increased non-polar surface area upon interaction. The entropies of binding are the same within error.

Crystallographic B factors can be used as indicators of dynamics.[100] Intra-structure comparison of the average all-atom ligand B-factor to that of the protein indicates the relative motion of the ligand in the pocket. Table 4.1 shows that for IPMP the ligand value is lower than that for the protein, whereas the opposite is true for IBMP. However, the difference in the IBMP B-factor values is small. Additionally, the occupancy of the ligand is 1, indicating it is in the same position in all units of the crystal. Consequently, both ligands were assumed essentially rigid when bound, despite the considerable size of the pocket as demonstrated in Figure 4.2.[18,23]

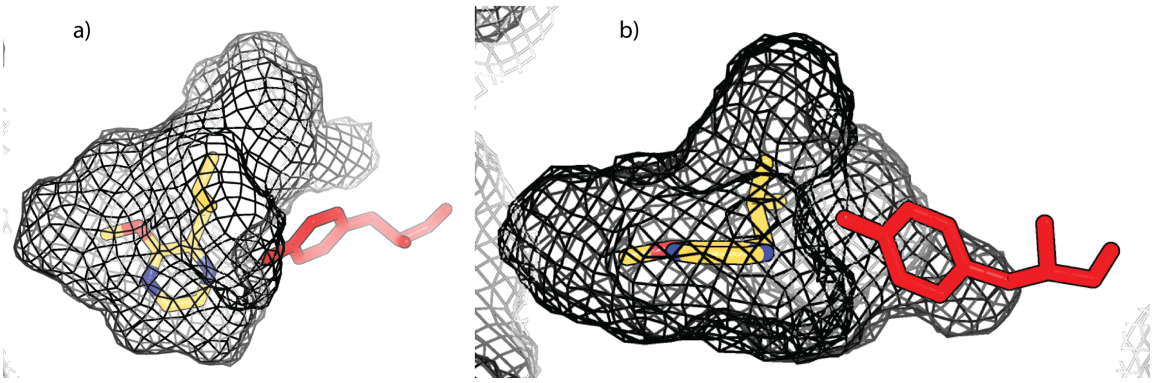


Figure 4.2 Ligand in binding pocket from IBMP-MUP crystal structure, 1qy1. a) top view b) side view. The pocket is represented by black mesh indicating the amount of space for movement. Tyr120 is shown in red. It is the only polar pocket residue, to which a hydrogen bond is assumed.

The predicted values of ligand rotational and translational entropy loss detailed in Table 4.1 incorporate this assumption. It was also assumed that all rotation of all bonds other than in the methyl groups were completely constrained upon binding. The final predictions, displayed in Table 4.1, were then generated using values from the literature that are based on the original statistical-mechanical approach for evaluating free ligand entropy.[107,108]

A recent theoretical paper, mentioned in §4.1.1, used the MUP-IPMP interaction as an experimental benchmark to test different approaches for calculating entropy of free ligand.[103] A rigid binding pose was again assumed, where, upon binding, all translational and rotational motions were reduced to vibrations and librations respectively. Using an approach with a solvent-constrained free ligand entropy, a value of -25.7 kJ mol^{-1} was calculated, almost replicating the -25 kJ mol^{-1} generated by experimental decomposition.

4.1.3 Significant residual translation and rotation predicted from μs molecular dynamics

Because of the wealth of biophysical data regarding the MUP-IBMP interaction, e.g. NMR and ITC, it was recently investigated using long-timescale 1 μs molecular dynamics simulations.[26] MD forcefields are not calibrated for or assessed at such long timescales, requiring that these simulations were carefully checked for forcefield artefacts. This was achieved for the protein by corroborating experimentally observed NMR chemical shifts with trajectory averaged calculated chemical shifts from the whole 1 μs trajectory.[23,26]

Unexpectedly, substantial ligand rotation and translation was observed in the binding pocket. The polar plot in Figure 4.3 (reproduced from the article), describes ligand movement through the projection of two orthogonal inter-ligand atom vectors over the 1 μs simulation. Though the movement is clustered, thus not completely isotropic, there is clearly substantial rotation of the ligand, and extended occupation of multiple distinct ligand poses. Clustering was performed by Dr Charlie Laughton (personal communication), and generated 10 populations. Representative poses from these clusters, above the population of each cluster as a timecourse for each replicate trajectory, Figure 4.4, reveal considerable movement between clusters even on a relatively short timescale. Given the MUP pocket's non-polar character, it is expected that ligands which can interact with the single hydrogen bond donor Tyr120 would be energetically required to satisfy this interaction. Surprisingly, there are times during each of three replicate trajectories when none of the three IBMP H-bond acceptor atoms (the two ring nitrogens and the methoxy oxygen) are within the necessary distance of the Tyr120 hydroxyl proton to indicate a H-bond.

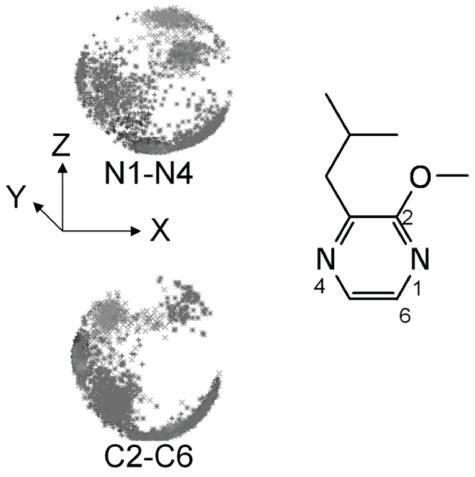


Figure 4.3 Polar plot of orthogonal ligand vectors over 1.2 μs simulation timecourse, reproduced from reference [26]. Much rotation is observed, contrary to the assumption of a single ligand conformation indicated in the crystal structure.

The loss of rotational entropy upon binding was reported from the 1 μs trajectories, using the Schlitter (analysis of covariance) method to compare the bound trajectory and a 1 μs trajectory of free ligand in water.[109] Translational changes were estimated using an approximate 10 kJ mol^{-1} penalty generated from the ideal gas approximation. The final reported value was ~22 kJ mol^{-1}, below the lower end of the IBMP prediction shown in Table 4.1, and similar to the recent calculation for IPMP assuming rigid binding. This unexpected motion, particularly the prolonged occupation of alternative poses, is herein investigated by direct experimental observation. This work provides the first dataset for experimental corroboration of ligand behaviour in long timescale MD simulations.

Additionally, to assess whether this unexpected ligand motion is a simulation artefact arising from the MD forcefield (AMBER) used[26], an explicitly solvated 100 ns trajectory of the MUP-IBMP complex was generated in CHARMM using parameters generated by the CHARMM general forcefield.[47] Motion over 100 ns is compared between the two trajectories to observe whether significantly less motion is observed in CHARMM than that evident in AMBER, Figure 4.4.

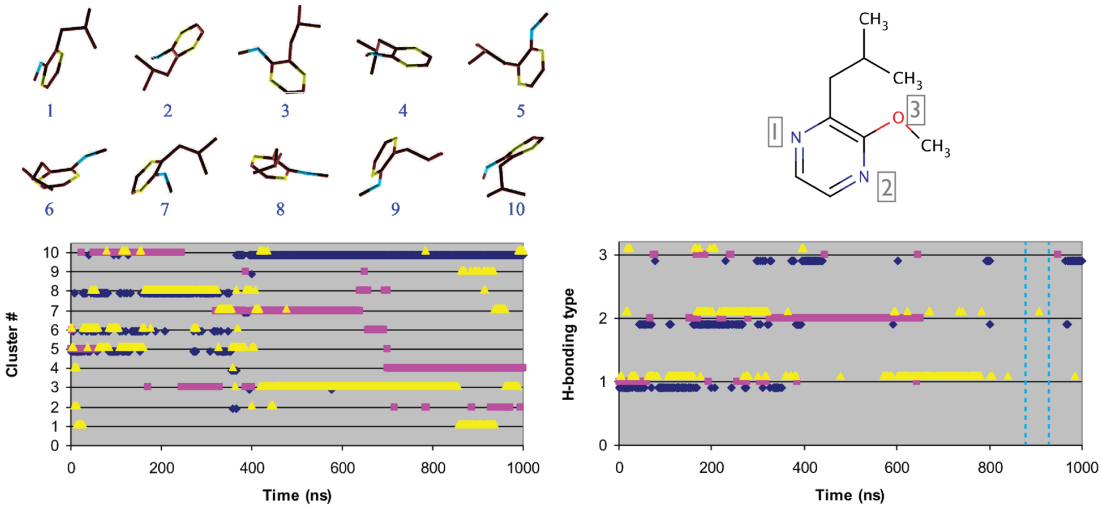


Figure 4.4 Analysis of residual IBMP rotation and translation throughout the 1 μs simulations. Figure provided via personal communication from Dr Charlie Laughton, University of Nottingham. (top left) 10 ligand clusters were generated, and representative structures are shown (protein in same orientation in each but not displayed). (bottom left) The three different coloured symbols represent the three replicate trajectories of the MUP-IBMP complex. Population of each cluster over time is shown. (top right) The three h-bond acceptor atoms on IBMP. (bottom right) Trajectory frames where the ligand position allows hydrogen bonding to the Tyr120 hydroxyl are shown from the three atoms indicated above. There are points in each trajectory where none of these three atoms are within H-bonding distance of Tyr120, indicated by dashed blue vertical lines.

4.1.4 Directly assessing predicted residual ligand motion for the first time using NMR

NMR residual dipolar couplings (rdcs) report the relative orientation of interatomic vectors within a molecular system for NMR-active nuclei, as described in §1.3.2.3. Using a sample containing NMR-active nuclei in both ligand and protein therefore allows the direct experimental observation of their relative orientation in complex, providing the closest experimental counterpart, albeit time-averaged, to the MD-derived polar plot in Figure 4.3. This is obtained by observing the deviation of the measured ligand rdcs from those predicted using the protein alignment tensor, indicative of a different orientation with respect to the external magnetic field, §1.3.2.3.

Measurement of rdcs requires production of protein and ligand samples which are labelled with NMR-active stable isotope. To achieve protein labelling, recombinant protein expression can be induced in 'minimal' medium, where the sole carbon and nitrogen sources are ^{13}C-glucose and ^{15}N-ammonium chloride respectively. The past two decades of protein NMR demonstrate that this is tractable for a large range of systems. Conversely, ligands must be synthesisable from labelled reagants. Fortunately IBMP can be synthesised organically from leucine, ammonium chloride, glyoxal and iodomethane, using published methodology.[110-113]

Rdcs were successfully measured previously for ^{15}N-labelled MUP protein in the PF1 phage alignment medium[114], dispensing the need to undertake lengthy alignment medium optimisation. When measuring rdcs, it is necessary to ensure that couplings are measured from the molecule in the required state, i.e. in complex and not in solution. Overlapping signals can frustrate this if both bound and unbound states are well populated during the experiment. Ideally, ligand rdcs need to be measured in the absence of free ligand, and protein rdcs need to be measured in the absence of free protein. Given the μM binding affinity of IBMP, it is expected that the exchange between free and bound ligand will be slow on the NMR timescale. Therefore ligand couplings were measured on a sample where ligand concentration is sub-stoichiometric, both before and after alignment. Careful titration of ligand into the NMR sample containing protein is therefore necessary. Then after full saturation with ligand post alignment, protein couplings will be measured, with isotropic protein couplings obtained by removing the phage via centrifugation.

Observed rdcs are averaged over all orientations sampled during the experiment, meaning they do not give atomic detail of each pose sampled, but rather indicate rigidity or deviation therefrom. NOE measurements also provide structural information regarding protein and ligand, but in the form of inter-nuclear distances, as detailed in §1.3.2.5. NOE spectra were measured to complement the rdc measurements in the assessment of residual rotation and translation of IBMP in the MUP pocket.

4.1.5 Work undertaken

Whilst our understanding of the MUP-IBMP interaction so far presumes rigid binding, long MD simulations show pronounced occupation of heterogeneous orientations by IBMP, in the MUP binding pocket on the μs timescale. For the first time, predicted residual ligand motion has been directly experimentally observed/assessed by measuring NMR rdcs and NOEs for a sample of ^{15}N MUP protein bound to ^{13}C-^{15}N IBMP. Complementary NOEs are obtained using a sample containing ^{13}C-^{15}N MUP and unlabelled IBMP, which allow assignment of crosspeaks to binding pocket protons. This work provides the first experimental dataset for corroboration of ligand behaviour in long timescale MD simulations. Ligand is synthesised from ^{13}C-^{15}N L-Leucine and ^{15}N ammonium chloride. ^{15}N-MUP protein was provided very generously by Girish Tampi. For comparison to observed NOE data, simulated NOEs were generated for the ligand pose from the crystal structure and representative poses from the 1 μs trajectory, using Prof Steve Homans 'relaxmd' software. Finally, ligand motion as a forcefield artefact is assessed by comparing a 100 ns explicitly solvated trajectory performed using CHARMM forcefield and parameters to the first 100 ns of the published AMBER results.

4.2 Material and Methods

4.2.1 Organic synthesis of IBMP

The entire protocol is graphically summarised along with weights at each step, Figure 4.5. ^{1}H NMR spectra (Bruker, 500 MHz) for each step are shown with assignments for molecule and solvent peaks in Appendix 1.

4.2.1.1 L-Leucine → BOC-L-Leucine [110]

1M sodium hydroxide (NaOH, 7.6 mL) and dioxane (3 mL) solution containing Leucine (500 mg, 3.61 mmol) was cooled to 0 °C, before the addition of further dioxane (2.3 mL) containing BOC anhydride (915 mg, 4.19 mmol). After overnight stirring at room

temperature, the mixture was diluted with water and extracted twice with hexanes. Solid citric acid was used to acidify the aqueous component before extraction with ethyl acetate. The organic component was then washed with water and brine, dried over magnesium sulphate ($MgSO_4$), filtered and finally concentrated under reduced pressure to yield BOC-L-Leucine (550 mg, 2.31 mmol).

4.2.1.2 BOC-L-Leucine → BOC-L-Leucine amide [111]

BOC-L-Leucine (550 mg, 2.31 mmol) was dissolved in methanol (MeOH, 20 mL) with ammonium chloride (NH_4Cl, 190mg, 3.47 mmol) and triethylamine (Et_3N, 484 μL) before addition of the amide coupling agent 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM, 766 mg, 2.77 mmol). After stirring at room temperature overnight, the solvent was removed using a rotary evaporator. The result was dissolved in dichloromethane (DCM, CH_2Cl_2) and washed sequentially with saturated sodium carbonate, water and brine. After drying with $MgSO_4$ and filtration, the sample was concentrated by rotary evaporator. Finally, a silica column in 1:1 hexane : ethyl acetate was used to purify the BOC-L-Leucine amide. The sample was contaminated with trimethoxy triazine (visible in ^{1}H NMR as 9H at 4.0 ppm[115]) due to a side-reaction of DMT-MM and MeOH. Because of lack of reactivity in subsequent reactions, this contaminant was left in the sample until the final IBMP purification step. However, mmol values cannot be calculated from weights due to this contamination, from this step until the final step.

4.2.1.3 BOC-L-Leucine amide → L-Leucine amide [112]

Trifluoroacetic acid (TFA, 15 mL) was added to the BOC-L-Leucine amide (341 mg, 1.43 mmol) and stirred at room temperature for an hour before rotary evaporation of solvent to yield L-Leucine amide (323 mg, 2.3 mmol). Half of this was taken forward to the next step.

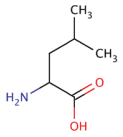
4.2.1.4 L-Leucine amide → 2-hydroxy-3-isobutyl-pyrazine (IBHP) [113]

Glyoxal (40% aq., 173 μL) was added with rapid stirring to a solution of L-Leucine amide (161 mg, 1.15 mmol) in MeOH (2.7 mL) at -35 °C, before dropwise addition of 12M NaOH (150 μL) over 5 minutes, maintaining a temperature of -35 °C. This mixture was left stirring at -35 °C for 30 minutes, then warmed to room temperature and stirred for 2 hours. The reaction vessel was cooled to 0 °C before neutralisation with 12M HCl (150 μL), addition of sodium bicarbonate (160 mg) and filtration. Water was added to filtrate before removal of MeOH via rotary evaporation. The solution was extracted with DCM, dried over $MgSO_4$ and filtered before removal of solvent via rotary evaporation to yield IBHP (70 mg, 0.44 mmol).

4.2.1.5 2-hydroxy-3-isobutyl-pyrazine (IBHP) → 2-methoxy-3-isobutyl-pyrazine (IBMP) [113]

IBHP (70 mg, 0.44 mmol) was suspended in dry THF (1 mL) and cooled to 0 °C before addition of sodium hydride (60% in mineral oil, NaH, 30 mg, 0.75 mmol). After stirring for 30 minutes at 0 °C, iodomethane (MeI, 50 μL, 0.75 mmol) was added and the reaction was warmed to room temperature and stirred for 48 hours. The reaction mixture was then diluted with water, and THF removed using a rotary evaporator. The aqueous solution was extracted with DCM. The organic layers were then combined, washed with a 5% aqueous solution of sodium thiosulfate, dried over $MgSO_4$, filtered and solvent removed via rotary evaporation. This crude product was then partially purified on a gel column in MeOH. After concentration through rotary evaporation, the partially pure product was purified on a silica gel column using 6% MeOH in DCM. Removing solvent through rotary evaporation yielded pure IBMP. Product analyses by ^1H NMR, and for the final product, ESI-HRMS (Electrospray ionisation high resolution mass spectrometry), are detailed in Appendix 1.

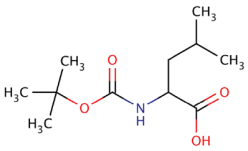
500 mg
M_w = 138.17
3.6 mmol



Leucine

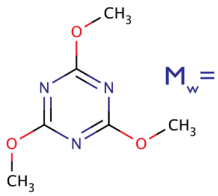
1M NaOH
Dioxane
Boc anhydride

550 mg
M_w = 238.17
2.3 mmol
64% yield



BOC-Leucine

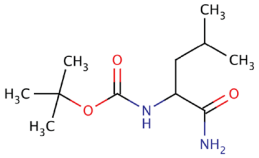
MeOH
Et_3N
DMT-MM
NH_4Cl



M_w = 171.15

Trimethoxy triazine
unexpected DMT-MM + MeOH
side-reaction

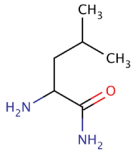
341 mg
M_w = 238.17



BOC-Leucine amide

TFA

323 mg
M_w = 139.17

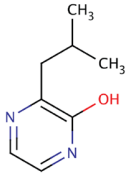


Leucine amide

Only half taken forward

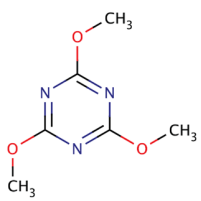
MeOH
Glyoxal
12M NaOH
12M HCl

70 mg
M_w = 160.17

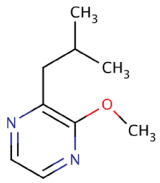


2-hydroxy-3-isobutyl pyrazine (IBHP)

THF
NaH (60% min. oil)
MeI



23 mg
M_w = 174.20
0.1 mmol



2-methoxy-3-isobutyl pyrazine (IBMP)

Figure 4.5 (previous page) Organic synthesis protocol for IBMP. Reagants are listed next to arrows, and those in bold are sources of heteroatoms present in the final product. Mass weight, molecular weight, total mmol (millimoles) and yield are shown next to each product. Due to trimethoxy triazine contamination, weights in red cannot be accurately converted into mmol product and yield.

4.2.2 NMR

4.2.2.1 Samples

Protein buffer is 50 mM potassium phosphate pH 7.4, 1mM sodium azide and 0.2 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). IBMP stock solution concentration was measured as 13.4 mM spectrophotometrically using A_{220} (e_{220} = 4980 M^{-1} cm^{-1}). During titration, aqueous ^{15}N-^{13}C IBMP stock solution was added to the NMR sample using a 10 μL pipette, before the NMR tube was centrifuged and then the solution mixed using a long glass pipette. MUP concentration was estimated at ~0.7 mM.

4.2.2.2 Experiments

All NMR spectra were easured using Varian Inova (Varian Inc., CA, USA) NMR spectrometers with Z axis gradients and triple resonance probes at 298 K. All experiments use sequences from the Biopack pulse sequence library supplied by Varian Inc. ^{1}H 1D spectra were measured using DPFGSE[116] water suppression and the water experiment from the Biopack library. 1D spectra were phase and baseline corrected. Peak integration was carried out using iNMR software.[63] All other spectra were processed using NMRPipe.[64] In these latter cases, phase correction was performed manually for ^{1}H using NMRDraw (part of NMRPipe), or automatically by Biopack (Varian) for heteronuclear spectra. NOE, and ongoing rdc, experiments were performed by Mr Phil Morrison and Dr Arnout Kalverda. NOE crosspeak intensities were calculated from the relevant spectral slices using NMRView.[117]

4.2.3 CHARMM MD simulation of MUP-IBMP complex

The crystal structure of MUP-IBMP (1qy1), with Cd ions removed, was used as the starting structure for the simulation. The 'Quick MD Simulator' functionality of CHARMM-GUI was used to generate ligand parameters, neutralise and solvate the system, and set up periodic boundary conditions.[118] Protein parameters were taken from the CHARMM22 forcefield and the ligand parameters were generated by CHARMM-GUI using the CHARMM Generalised Forcefield (CGENFF).[47] A disulphide

bridge was specified between residues 64 and 157 as observed in the crystal structure. Sufficient K^+ and Cl^- were placed using a Monte-Carlo method to achieve neutrality. The molecule was solvated in an octahedral box of ~7000 TIP3 water molecules with periodic boundary conditions. Long range electrostatic interactions were treated using the particle mesh Ewald method. SHAKE was applied to constrain all hydrogen bonds, allowing a 2 fs timestep.

Minimisation involved 1000 steps using the Steepest Descent algorithm followed by 100 steps using the Adopted Basis Newton-Raphson algorithm. The system was subsequently heated from 50 K to 298 K over 25000 steps, before performing 100 ns of equilibrium trajectory at constant temperature and pressure (298 K, 1 atm). Coordinates were saved every 1 ps. Wordom was used in generating a 100 ns desolvated and aligned trajectory of 1 frame per 100 ps for visualisation.[48] Julie Roy provided copies of the AMBER 1 μs trajectories[26], edited to 1 frame per 100 ps. Wordom was also used on these trajectories to generate a desolvated and aligned trajectory of the first 100 ns of 1 frame per 100 ps. All trajectories were subsequently visualised and compared using VMD.[119] RMSD and RMSF analysis were also undertaken using Wordom.

4.2.4 NOE prediction from AMBER trajectory structures

The trajectories provided by Julie Roy were also converted to 1 μs desolvated and aligned trajectories with the aid of Wordom, composed of 1 frame per ns. These trajectories were subsequently visualised and compared using VMD.[119] Ligand poses populated for long periods were identified via visual inspection and representative frames were extracted, Figure 4.14. These frames were edited using Pymol to produce structures comprising only the IBMP ligand and any protein atoms within 7 Å of ligand atoms. Simulated NOES for each of these representative poses from the AMBER trajectory were generated using 'relaxmd' software, using a spectrometer frequency of 600 MHz, a correlation time of 9 ns[23] and a mixing time of 0.12 s (Prof. Steve Homans, personal communication). Relaxmd is a package for the calculation of homonuclear and heteronuclear relaxation and NOE parameters via a full relaxation matrix approach.[120] Only simulated NOEs > 0.5% intensity were included in the analysis. For each pose NOEs were calculated from three groups probed experimentally: the methoxy and both methyls.

4.3 Results and Discussion

4.3.1 Ligand synthesis and titration into protein sample

Synthesised IBMP was characterised and purity was confirmed by a number of techniques. The ^{1}H NMR spectrum of the ^{15}N-^{13}C IBMP stock solution is as published[113], except for minimal (~1%) residual contamination by trimethoxy triazine, Figure 4.6. ^{1}H -^{13}C HSQC and ^{13}C TOCSY HSQC experiments performed on the same sample probe the labelled butyl chain of IBMP, showing the correct resonances and connectivity thereof, Figure 4.7. High resolution electrospray ionisation mass spectrometry of the final sample measured a mass equivalent to the expected value (within 0.01 Da).

As described in §4.1.4, to minimise the chance of overlapping signals from free ligand frustrating the accurate measurement of couplings arising from bound ligand, the sample was titrated to sub-stoichiometric saturation. Titration of IBMP into MUP was monitored by three NMR experiments; ^{1}H spectra to observe increasing MUP saturation; the first transient of ^{1}H-^{13}C HSQC specta to observe free ligand; and ^{1}H-^{15}N HSQC were used to observe protein backbone changes accompanying IBMP binding.

Changes in the methyl region of a 1D ^{1}H NMR spectrum of MUP upon titration of IBMP were observed, Figure 4.8a. The relative peak intensities of the methyl peaks at approximately -0.5 and -0.8 ppm were used to measure the fraction of protein in the free and bound states respectively. ^{1}H 1D spectra were obtained after each titration step and the relevant peaks integrated to obtain the approximate protein saturation, Figure 4.8b. The titration was stopped at ~65% protein saturation, which was obtained after three titration steps.

A previous publication in 2004 has assignments for the ^{1}H-^{15}N HSQC spectrum for both unbound and IBMP-bound ^{15}N MUP.[23] In this study, ^{1}H-^{15}N HSQC spectra were obtained after each titration step. Overlaid spectra of the pre-titration and final HSQCs with the bound and unbound assignments reveals that the majority of peaks overlay exactly, Figure 4.9. Some peaks move or split, due to a residual unbound protein population, in accordance with the MUP-IBMP assignment.

The first transient of a ^{1}H-^{13}C HSQC was obtained pre-titration and after each titration step to monitor for the presence of unwanted emergence of free ligand, Figure 4.10. Resonances for the butyl chain protons are clearly observed in the

aqueous stock solution of ^{15}N-^{13}C IBMP, and weak natural abundance breakthrough of the protein is observed in the ^{15}N MUP sample pre-titration. Upon titration, peaks at -0.1 and 0.15 ppm appear and increase in intensity, corresponding to the bound state methyl protons. A small amount of free ligand may be appearing at the final step due to a sharper peak at ~0.75 ppm.

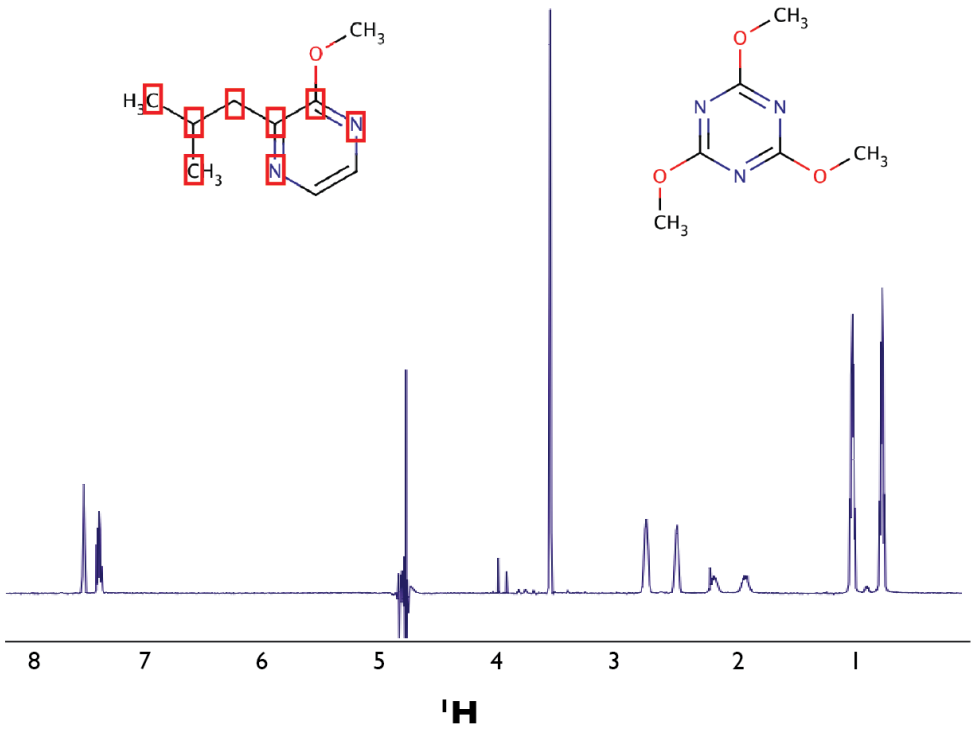


Figure 4.6 ^{1}H NMR spectrum of ^{13}C-^{15}N IBMP in H_2O stock solution. Red boxes on the IBMP structure (top left) indicate ^{13}C and ^{15}N heteroatoms. A small amount of trimethoxy triazine (shown top right) is visible. δH (500 MHz, H_2O/D_2O); 7.5 (m, 1H, ArH), 7.37 (m, 1H, ArH), 4.02 (s, 9H, trimethoxy triazine), 3.57 (s, 3H, OCH_3), 2.64 (d of m, 2H, J_{CH} 127.4 Hz, CH_2), 2.06 (d of m, 1H, J_{CH} 132 Hz, CH), 0.91 (d of m, 6H, J_{CH} 124.4 Hz, CH_3). See Appendix I for ^{1}H NMR details of all synthesis intermediates. Appendix I also contains a $CHCL_3$ ^{1}H NMR spectra for commercially available unlabelled IBMP, for comparison.

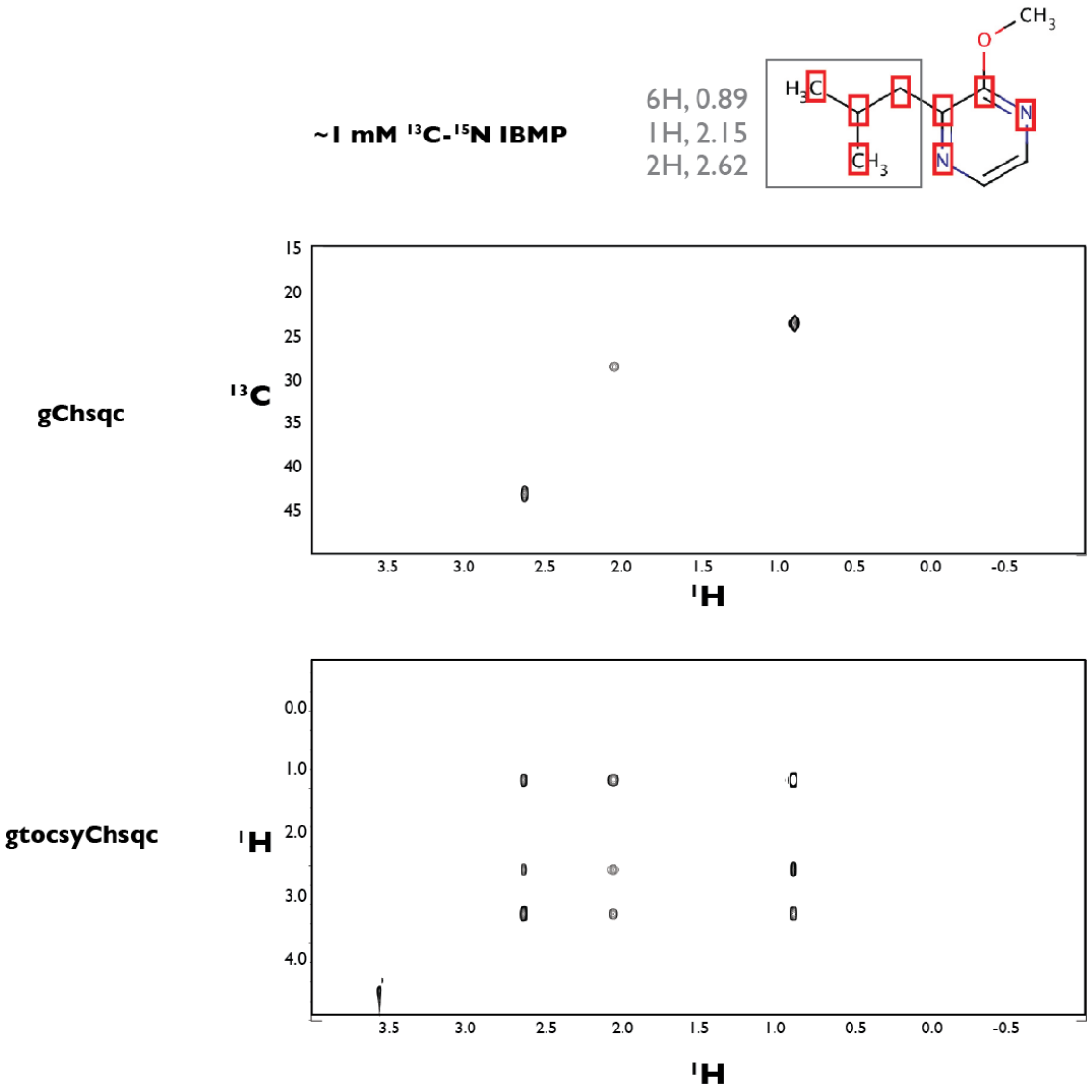


Figure 4.7 ^{1}H-^{13}C HSQC and ^{13}C TOCSY HSQC NMR spectra of synthesised ^{13}C-^{15}N IBMP in H_2O stock solution. Only the butyl chain contains protons bound to ^{13}C, in accordance with the resonances listed. The TOCSY shows the expected crosspeaks between all three peaks on the butyl chain.

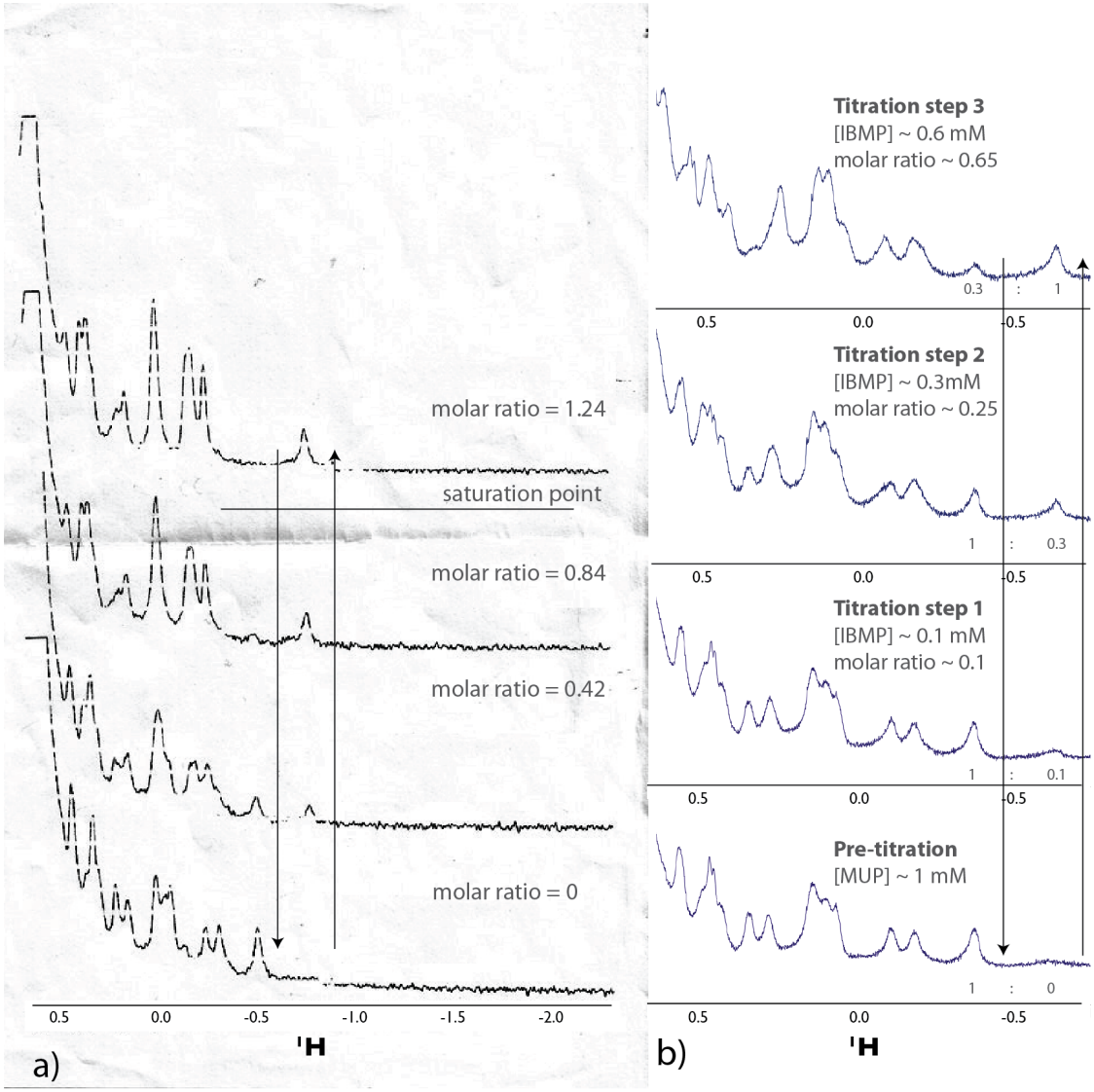


Figure 4.8 NMR experiments tracking MUP saturation during titration of ^{15}N-^{13}C IBMP into ^{15}N MUP. a) ^{1}H 1D NMR spectra reveal a change in peak profile upon MUP titration. Relative intensities of the two rightmost peaks indicate saturation (indicated by arrows). b) Relative peak intensities over three titrations estimate the final sample as ~65% saturated.

Figure 4.9 NMR experiments tracking protein backbone changes during titration of ^{15}N-^{13}C IBMP into ^{15}N MUP. Amide protons are observed using ^{1}H-^{15}N HSQC NMR spectra, which are colour coded as per the figure legend. The titration does not significantly change the HSQC. Some peaks move or split, due to a residual unbound protein population.

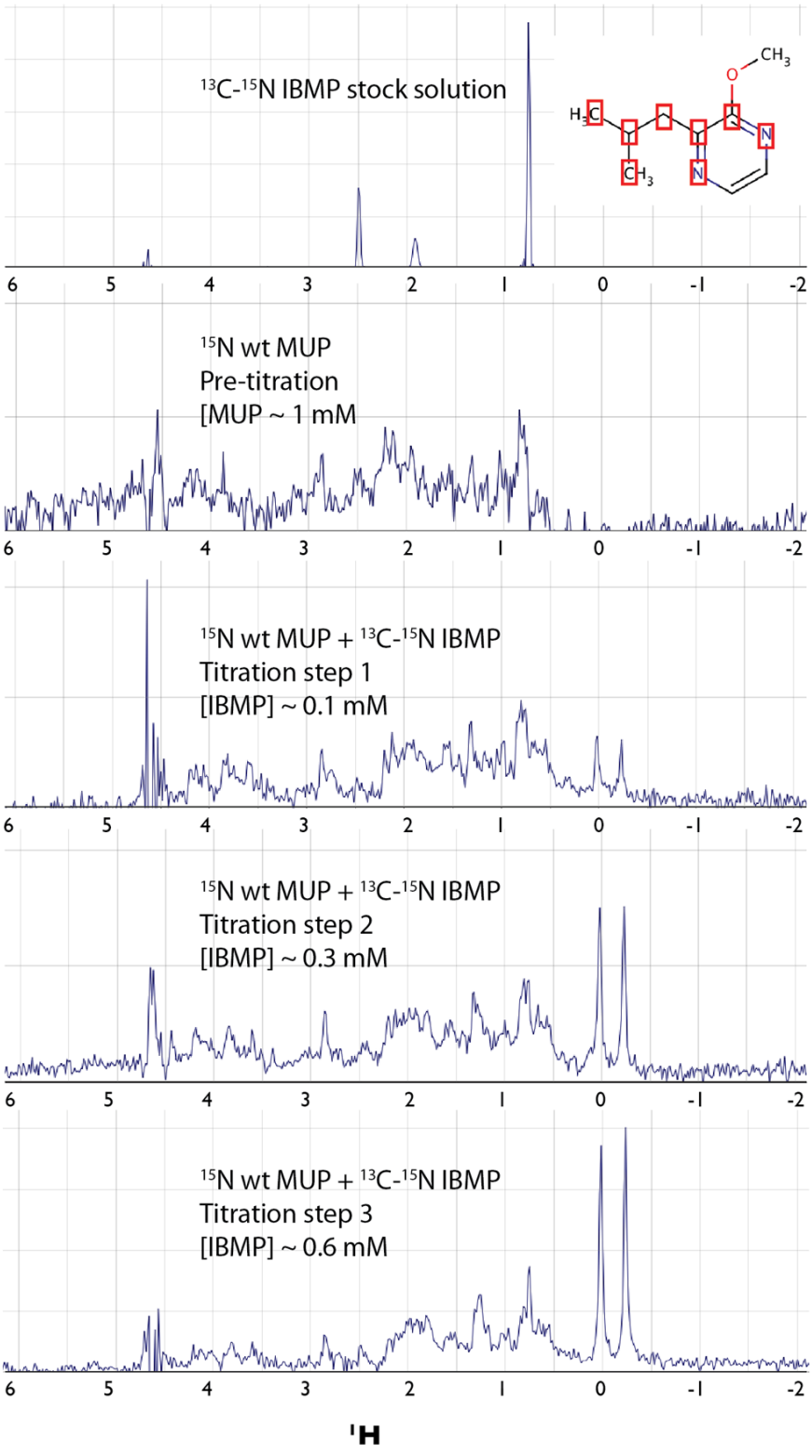


Figure 4.10 NMR experiments monitoring ligand resonances during titration of ^{15}N-^{13}C IBMP into ^{15}N MUP. Ensuring no free ligand accumulates over the titration is achieved by obtaining the first transient of a ^{1}H-^{13}C HSQC NMR spectrum. Upon titration, two peaks at -0.1 and 0.15 ppm appear and increase in intensity, corresponding to the bound state methyl protons. A small amount of free ligand may be appearing at the final step as indicated by a sharper peak at ~0.75 ppm.

4.3.2 Assignment of ^{13}C-^{15}N IBMP in complex

NMR experiments were performed to obtain ^{1}H-^{13}C HSQC and HCCH TOCSY spectra for the post-titration sample, Figure 4.11. Though some of the peaks in the sample are not of ligand origin and remain unassigned, two sets of peaks corresponding to free (green) and bound (blue) ligand are clear. The TOCSY indicates the free CH peak at 2.15 ppm, which is not shown on the HSQC, presumably due to low intensity, conclusively revealing the presence of a small amount of free ligand. Dispersion of the CH_3 and CH_2 peaks in the bound state indicates that the ligand is moving less freely than in solution.

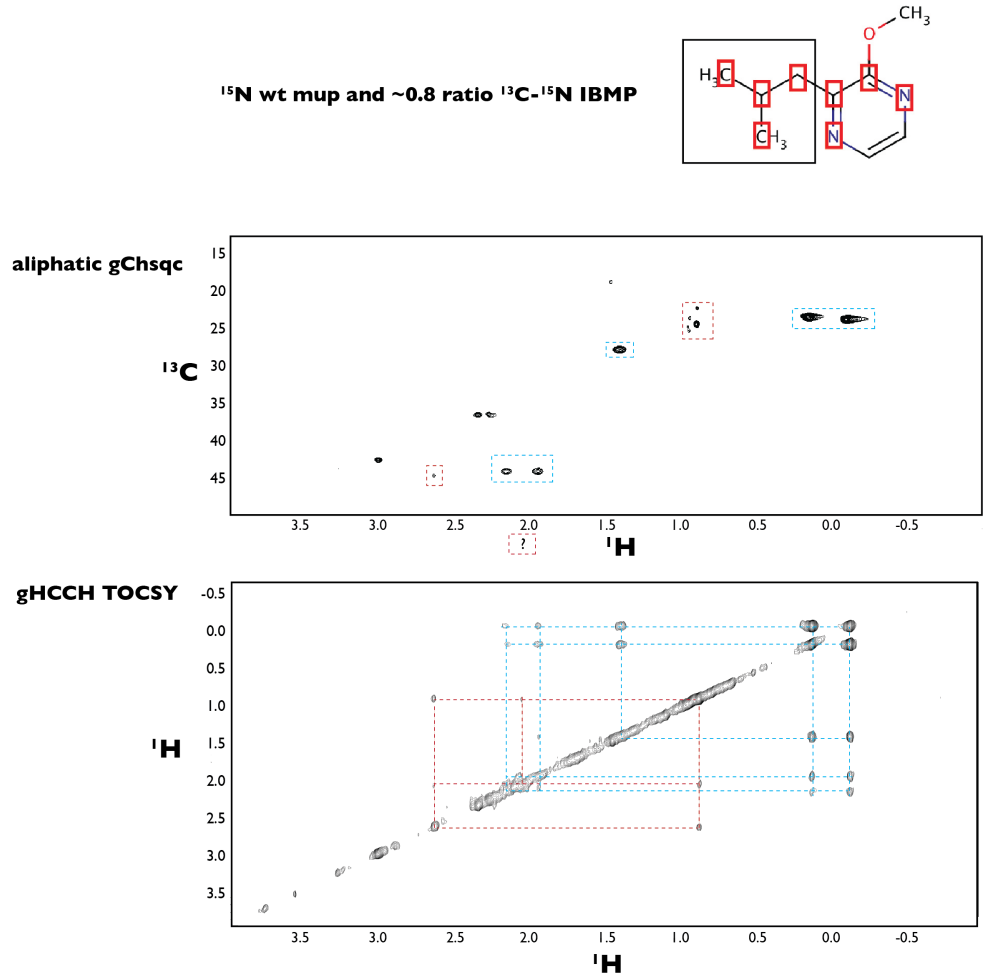


Figure 4.11 Assigning ligand resonances in the MUP-IBMP complex using HSQC and TOCSY NMR experiments. Two sets of peaks corresponding to free (green) and bound (blue) ligand are clear, revealing the presence of a small amount of free ligand. The dispersion of the CH_3 and CH_2 peaks in the bound state indicates that the ligand is moving less freely than in solution. This may indicate a well-defined pose in the pocket associated with different resonances for each methyl group and methylene proton.

4.3.3 2D ^{13}C HSQC NOESY of ^{13}C-^{15}N IBMP bound to ^{15}N MUP

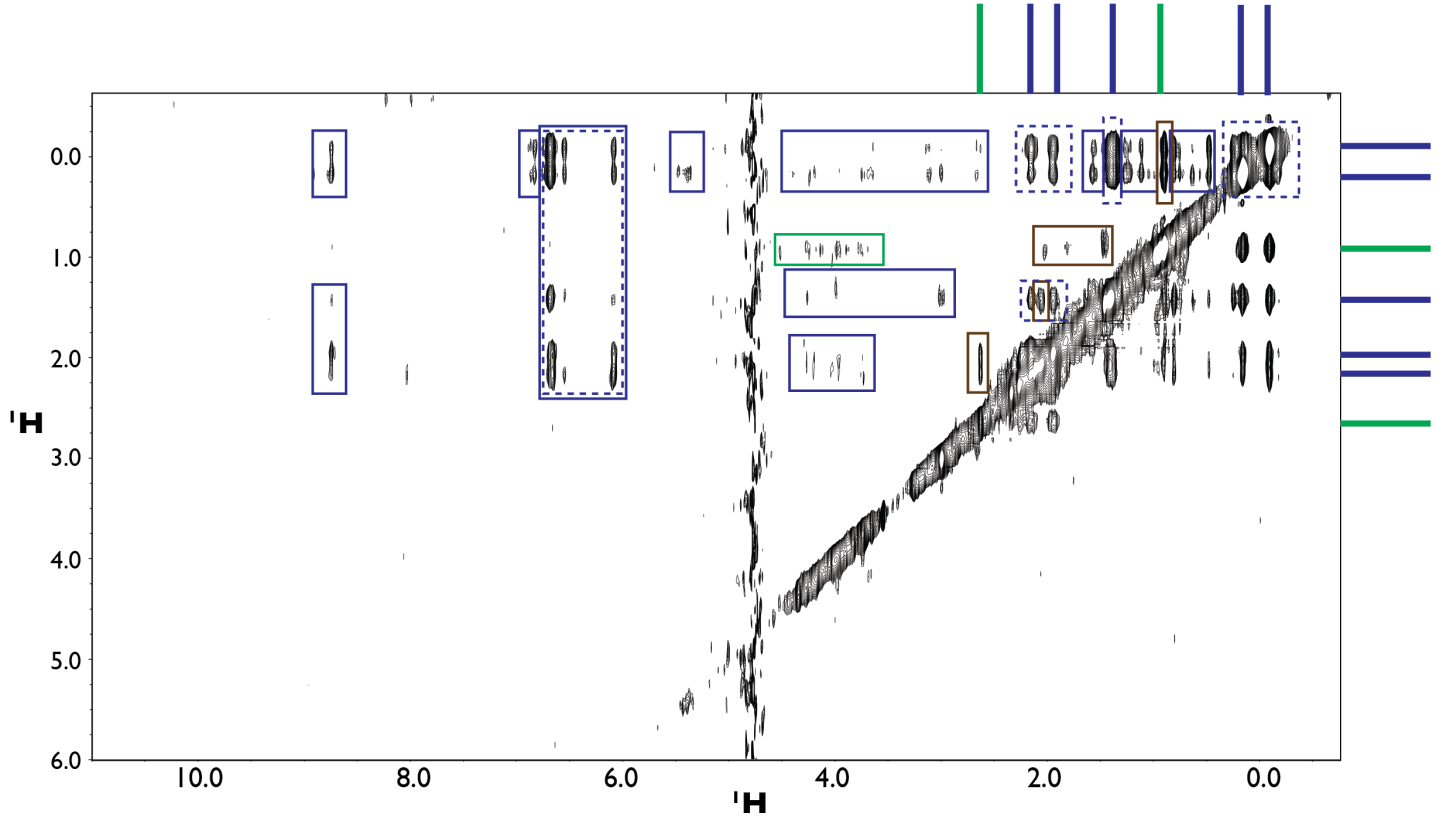
As described in §1.3.2.5, NOE crosspeaks indicate nuclei that are, upon average, within ~6 Å of each other. NOE measurements were obtained before alignment of the sample to assess agreement of experimental and simulated NOEs, the latter derived from MD-observed poses, Figure 4.14 and §4.3.5.

A ^{13}C HSQC NOESY spectrum was obtained at 600 MHz with a 120 ms mixing time, Figure 4.12. The HSQC component results in only those crosspeaks seen to protons attached to ^{13}C appearing 'below' the diagonal. Crosspeaks between bound ligand resonances are observed as expected. No crosspeaks between residual free ligand resonances are observed. This may be due to the small quantity of free ligand, which will have smaller intensity NOEs that take longer than 120 ms to build up. It is noteworthy that there are exchange crosspeaks between bound ligand and free ligand, revealing that the ligand exchanges between the free and bound state within the experimental timescale.

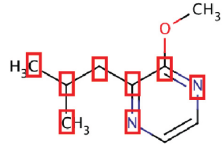
There are some crosspeaks between the bound ligand and the protein, 'above' the diagonal. Those seen between 6 and 7 ppm may be from IBMP methyl protons to either the unassigned IBMP aromatic protons or a phenylalanine sidechain in the protein pocket. The crosspeaks at ~9 ppm are to an unknown pocket residue: from the crystal structure the most likely candidate would be the Tyr120 hydroxyl proton.

Directly assigning the observed NOEs was prevented by the protein having no ^{13}C label. Consequently, acquiring a NOESY spectrum for a second sample containing ^{13}C-^{15}N MUP and an excess of unlabelled ligand complemented these NOE measurements.[23]

Figure 4.12 (next page) 2D ^{13}C HSQC NOESY NMR spectrum of ^{13}C-^{15}N IBMP bound to ^{15}N MUP, provided by Phil Morrison. Resonances of the ligand protons in the free and bound state are listed, and are indicated next to the spectrum using blue and green bars for bound and free ligand respectively. Crosspeaks between 6 and 7 ppm are boxed with solid and dashed lines because it is unknown whether they correspond to intra-ligand or ligand-protein NOEs.



Proton Resonances



Free	Bound
6H, 0.89	6H, -0.1 + 0.15
1H, 2.15	1H, 1.4
2H, 2.62	2H, 1.9 + 2.1
3H, 3.45	3H, ?
2H, 6.96 + 7.15	2H, ?

NOEs

- Free ligand - Protein
- Bound ligand - Protein
- Bound ligand - Bound ligand
- Bound ligand - Free ligand

4.3.4 3D ^{13}C NOESY HSQC of unlabelled IBMP bound to ^{13}C-^{15}N MUP

The ^{1}H-^{15}N HSQC spectrum of the sample containing ^{13}C-^{15}N MUP bound to unlabelled IBMP differs from that of the sample containing ^{15}N MUP bound to ^{13}C-^{15}N IBMP, Figure 4.13a. The differences between the two arise from the differential saturation of the samples: the former contains saturated protein and therefore the chemical shifts correspond completely to the bound state, whereas the latter contains a mixture of bound and unbound protein, with accordant changes in chemical shift.

To observe NOE crosspeaks from protein sidechain protons to ligand protons, a ^{13}C NOESY HSQC spectrum was obtained at 600 MHz with a 120 ms mixing time, Figure 4.13b. Protein-protein NOEs are on the diagonal, whereas protein-ligand NOEs appear as crosspeaks. As expected, no intra-ligand NOE peaks appear due to the ligand being unlabelled.

Crosspeaks are observed to the resonances at ~0 ppm and 3.5 ppm. The peaks around 0 ppm correspond to the IBMP methyl protons, as seen in Figure 4.11, whereas 3.5 ppm corresponds to a methoxy shift. NOEs to the ligand CH, CH_2 and aromatic protons are presumably sufficiently weak to be below experimental detection. Spectra in the third dimension at the two above mentioned resonances of ~0 and 3.5 ppm reveal NOEs to previously assigned pocket residues.[23] Protons attached to the following sidechain carbon atoms show NOE crosspeaks to the ligand methyl protons: L42δ1, L42δ2, L54δ1, L54δ2, I92δ1, A103β, and L116δ2, Figure 4.13c. L40δ2, M69, V82γ2, L105δ1 show crosspeaks to the ligand methoxy protons, Figure 4.13d. Upon simple inspection, the spatial distribution of these residues with respect to the relevant ligand protons suggests an average ligand pose close to that observed by crystallography, Figure 4.13e. However, the next section more comprehensively assesses the agreement between these observed NOEs and simulated NOEs arising from the many MD-observed ligand poses.

The interpretation of relative NOE intensities as directly representing relative internuclear distances can be complicated by spin diffusion. A 120 ms mixing time was used as a compromise, to minimise spin diffusion whilst maximising useable NOE intensity. In the next section, experimental NOEs are compared to simulated NOEs that are calculated using the relaxmd software package that takes into account spin diffusion.

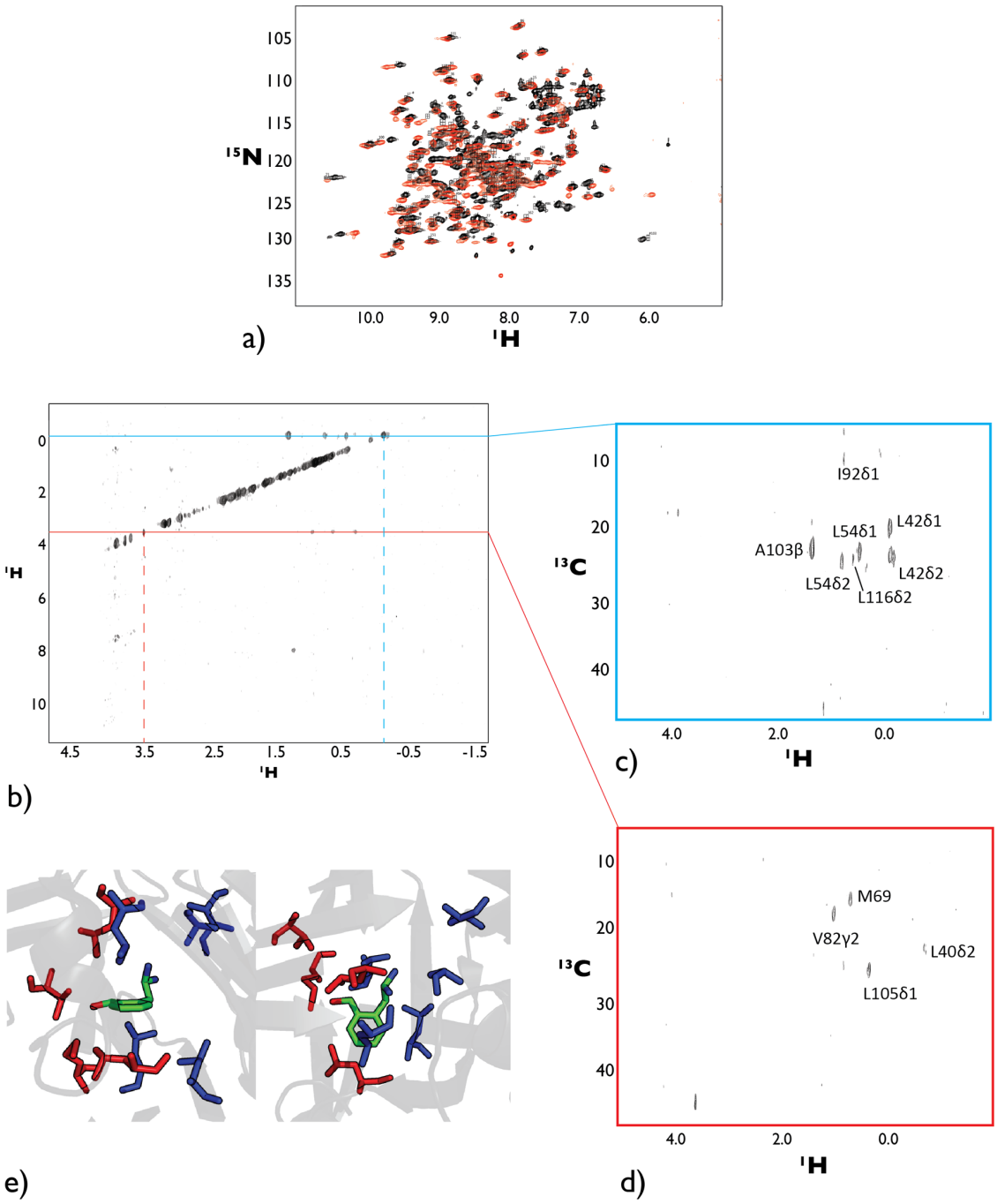


Figure 4.13 NMR spectra of unlabelled IBMP bound to ^{13}C-^{15}N MUP, provided by Phil Morrison. a) ^{1}H-^{15}N HSQC reveals much overlay of this sample to the ^{13}C-^{15}N IBMP bound to ^{15}N MUP sample. b) 3D ^{13}C NOESY HSQC spectrum with the carbon dimension collapsed. Blue and red lines represent crosspeaks to the bound IBMP methyl and methoxy peaks respectively. c) + d) 2D spectra extracted from the points indicated in b), showing crosspeaks to the relevant ligand protons. e) 1qy1 MUP-IBMP crystal structure indicating crosspeak residues from c) and d) in blue and red respectively. Methyl and methoxy protons are coloured blue and red accordingly. NOE data may support a ligand pose similar to that observed crystallographically.

4.3.5 Comparison of simulated and observed NOEs

Representative structures of well-populated ligand poses from the 1 μs trajectories analysed in Figures 4.3, 4.4, and reference[26] were analysed to generate simulated NOEs, as described in §4.2.4. Nine ligand poses populated for long periods were identified via visual inspection of MD trajectories, Figure 4.14, the first of which represents the crystallographically-observed pose. Three groups of ligand protons were chosen from which to simulate the NOEs: the methoxy and two methyl groups.

Poses 1 and 2 show the best agreement between simulated NOEs and observed NOEs at a residue-level comparison, Table 4.2. They have the highest 'observed/simulated' values, demonstrating that more of the simulated NOEs are observed for these poses. Furthermore, they have the lowest 'strong unobserved' values. This number represents how many of the strongest 6 simulated NOEs are not observed in the experimental data. If the pose is correct, then the strongest simulated NOEs should definitely be present in the experimental data. Consequently a low value, wherein the strong simulated peaks are mostly observed, indicates good agreement. Pose 1 is almost identical to the crystal structure, and pose 2 is the structurally closest to Pose 1: the methyl remains in almost the same position but the ring and methoxy are moved, Figure 4.15. Accordingly, whereas pose 1 has the best overall agreement between simulated and observed NOEs, pose 2 still has good agreement for methyl NOEs due to its methyl group remaining in almost the same position.

The above observation is strengthened by the comparison of the intensities of simulated and observed NOEs, as shown in Table 4.3. These data also confirm that the residue-level pattern in Table 4.2 arises from simulated NOEs to the exact protons assigned in the observed NOEs. The relative intensities of the simulated and observed NOEs were compared for those poses with multiple (>2) comparable simulated NOEs, i.e. poses 1 and 2. The only simulated NOEs missing for pose 1 are the two experimentally weakest NOEs: 92δ1 and 116δ2. Though NOEs to 42δ1+2 are weaker in simulation than experiments, the ranking is almost exactly the same as experiment. Pose 2 has a stronger 42δ1+2, but an overestimated 92δ1, leading to close but not perfect agreement with experiment. The correlation of methyl NOE intensities is equivalent for both, R^2 ~0.6. Figure 4.15 displays poses 1 and 2, and demonstrates that despite having similar methyl NOEs, they still contain various conformations of the butyl chain. This scope for conformational flexibility of the butyl chain is important for interpretation of rdc data: the butyl chain may sample multiple bond orientations with

respect to the protein whilst the ligand remains in effectively the same pose. A comparison of methoxy NOE peak intensities is only possible for pose 1, wherein the ranking is reproduced except for a simulated over-estimation of NOE intensity to residue 69. It is worth noting that poses 3, 8 and 9, all of which have two comparable simulated NOEs, perform badly by possessing relative intensities inverse to those seen in experiment.

The nine poses sample multiple different methyl and methoxy positions, Figure 4.14, yet only poses 1 and 2 give the above mentioned agreement. Therefore these data support a situation wherein the ligand is oriented very close to the crystallographically-observed pose for the majority of the time.

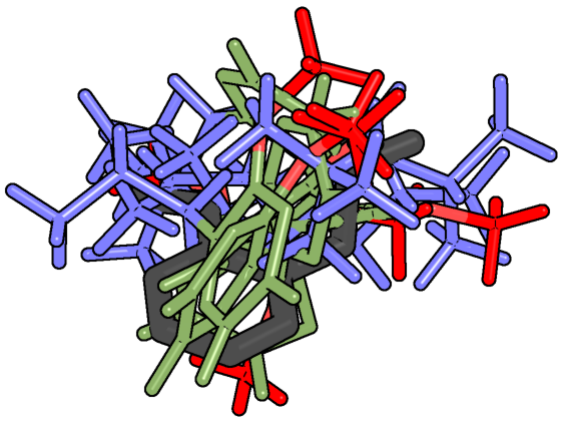


Figure 4.14 Crystal (1qy1) pose overlaid with all nine AMBER MD-observed poses, demonstrating the translational and rotational heterogeneity of the latter. The crystal pose is presented in grey. All MD-observed poses are coloured as follows: green for ring atoms, red for methoxy atoms and blue for butyl chain atoms.

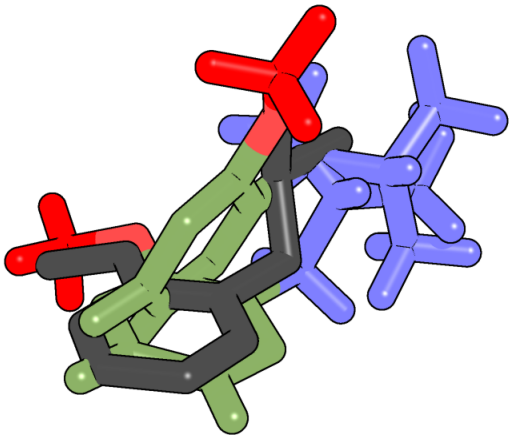


Figure 4.15 Crystal (1qy1) pose with AMBER MD-observed poses 1 and 2 overlaid, coloured as per Figure 4.14. Pose 1 almost exactly reproduces the crystal pose. Pose 2 has a similar butyl chain position, but the ring is rotated to significantly reposition the methoxy, as seen in the greater agreement of methyl but not methoxy NOEs in Tables 4.2 and 4.3.

Pose	Methoxy	Methyl A	Methyl B	Observed / simulated	Strong unobserved
1‡	40*, 54, 56, **69***, **82***, 83, 84, 88, 105*	42*, 45, 54*, **56**, **103***, 120	54, 56, 90, **103**, **105**	7 / 17	2
2	103, 104, **105***, **116**, 117, 118	42*, 45, **54***, **56**, 90, 103*, 120	54*, **90**, 92*, 101, **103***, 104, 120	5 / 16	3
3	**42**, **90**, 92, 101, 103, 120	84, 90, 103*, **105**, **116***, 117, 120	24, 90, 103*, 104, 105, **116***, 117, 118, **120**	2 / 15	4
4	90, 91, 103, **104**, **105***, 116, 117, 118	54*, 56, 82, **84**, 88, 90, **105**	**54***, **56**, 82, 84, 90, 105	2 / 15	4
5	24, 39, **40***, 41, **56**, 116, 120	56, 69, **82**, 84, **105**	**40**, 56, **69**, 105, 116*	2 / 14	5
6	90, **103**, 104, 105*, 116, **117**, 118, 120	**38**, **40**, 56, 69, 84, 116*	38, 69, 82, **84**, 105, **116***	2 / 16	5
7	**90**, 91, 103, **104**, 105*, 116, 117, 118, 120	40, 56, **69**, 82, 88, **105**, 116*	38, **40**, 69, 105, **116***	2 / 17	5
8	42, 54, **56**, **90**, 103, 120	38, **40**, 56, 84, **116***	54*, 56, **69**, 82, 83, **84**	2 / 15	5
9	24, **38**, 39, **40***, 41, 56, 116	38, **90**, 103*, 104, **105**, 116*, 117, 118, 120	24, 56, 90, 103*, 104, 105, **116***, 117, 118, 119, **120**	3 / 19	4

Table 4.2 Simulated ligand-protein NOEs by protein residue number for nine well-populated ligand poses taken from the 1 μs AMBER trajectories.[26] Only intensities above 0.5% were included. Residue numbers are grey if the intensity is calculated as between 0.5 and 1%, black if above 1%. Bold residues are the two with the strongest NOEs from the relevant proton. Values denoted with a * indicate residues with observed NOEs, Figure 4.13. ‡ Pose 1 represents the crystallographically-observed pose. Observed/simulated is the number of simulated residues with observed NOEs / the total number of residues in the simulated NOEs. 'Strong unobserved' is the number of strong (bold) simulated NOEs that are not observed experimentally.

Pose	1	2	3	4	5	6	7	8	9	Observed
Atoms	Methoxy									
40δ2	1.87	-	-	-	10.68	-	-	-	2.53	1.51
69	3.66	-	-	-	-	-	-	-	-	2.39
82γ2	2.42	-	-	-	-	-	-	-	-	2.52
105δ1	3.25	2.83	-	1.63	-	-	1.62	-	-	2.71
	Methyl									
42δ1	0.61	2.06	-	-	-	-	-	-	-	5.62
42δ2	0.56	1.90	-	-	-	-	-	-	-	6.17
54δ1	5.53	4.91	-	8.72	-	-	-	3.54	-	5.21
54δ2	2.77	3.44	-	1.22	-	-	-	-	-	5.33
92δ1	-	2.69	-	-	-	-	-	-	-	1.89
103β	8.42	10.33	4.96	-	-	-	-	-	5.5	9.54
116δ2	-	-	13.63	-	1.82	8.98	7.62	7.62	10.63	2.95

Observed	Pose 1	Pose 2
	Methoxy	
40δ2	40δ2	
69	82γ2	
82γ2	105δ1	
105δ1	69	
	Methyl	
92δ1	92δ1	116δ2
116δ2	116δ2	42δ2
54δ1	42δ2	42δ1
54δ2	42δ1	92δ1
42δ1	54δ2	54δ2
42δ2	54δ1	54δ1
103β	103β	103β
R^2	0.59	0.61

Table 4.3 (above) Intensities of simulated and observed NOEs (for simulated, between ligand groups used for calculation and protons attached to observed assigned carbons). Values are generated by addition of either percentage intensity values for simulated NOEs, or peak intensities for observed NOEs. For simulated NOEs, only values with intensity > 0.5% were included. (left) Ranking of NOE crosspeaks according to intensity (weaker to stronger) for poses with multiple comparable simulated NOES, i.e. poses 1 and 2. R^2 are calculated using the intensities above. Experimental data supports poses close to, or population weighting close to 1, which has a similar methyl position to pose 2.

4.3.6 Rdc experiments: inconclusive data so far

Experiments by Phil Morrison and Dr Arnout Kalverda to generate a sufficient rdc dataset to describe the relative rigidity of the ligand with respect to the protein are ongoing. Rdc prediction and analysis were performed by Dr Gary Thompson using the PALES program.[121] The PF1 phage introduced partial alignment to the protein sample without affecting the structure, observed via comparison of HSQC spectra. A sufficient number of isotropic backbone amide protein couplings attributable to secondary structure residues in the bound state were resolved and measured before ligand saturation. Accordingly, unlike the plan described in §4.1.4, the phage did not need to be removed after ligand saturation to obtain bound state protein couplings. Protein backbone amide couplings were measured using J-modulated ^{15}N HSQC experiments, and in conjunction with the 1QY1 crystal structure of the IBMP-MUP complex were used to calculate predicted rdc values for the ligand in the crystallographically-observed pose, Table 4.4.

The measurement of ligand rdcs has been problematic. Only rdcs for the butyl chain have been acquired so far, using a combination of coupled HSQC and J-modulated Constant Time Period experiments. The two unlabelled ring carbons and their associated protons have remained unassigned and therefore inaccessible, so no data exists for the orientation of the pyrazine ring. This is problematic, because the accuracy of the alignment tensor calculation scales with the number of non-parallel vectors measured. Moreover, so far the observed rdcs are weak and some have quite large errors. For illustrative purposes, a comparison of the predicted and current experimentally observed ligand rdcs is provided in Table 4.4, courtesy of Phil Morrison. The important observation so far in these preliminary data is the partial disagreement of expected and predicted rdcs (some predicted rdcs are within the error bounds of the observed values), perhaps indicating a non-crystal like orientation of the butyl chain bond vectors. Nonetheless, as mentioned in §4.3.5, the butyl chain may sample different conformations, and therefore multiple bond vector orientations relative to protein, whilst the ligand effectively remains in a crystal-like pose consistent with the NOE data. Limitations in the ligand rdc acquisition may be surmountable, for example increasing rdc strength through addition of more alignment medium, or accessing ring protons through use of natural abundance ^{13}C experiments. The option also exists to use the residual labelled leucine amide, §4.2.1.3, with labelled glyoxal to

synthesise IBMP with all the aromatic heteronuclei labelled, but because the reagant is expensive other options will be considered first.

Coupled Atoms	Coupling Pair	Experimental rdc (Hz)	Predicted rdc (Hz)
C-C	Me(0) and CH	0.3	1.7/-0.9
C-C	Me(1) and CH	0.5	1.7/-0.9
C-C	CH2 (0) and CH	2.7 ± 1.3	-0.2
C-C	CH2 (1) and CH	-0.1 ± 1.6	-0.2
C-C	CH2 (0) and C(ar)	-1.2 ± 0.7	1.7
C-C	CH2 (0) and C(ar)	2.4 ± 1.1	1.7
C-C	Me(0) and CH	-2.1	1.7/-0.9
C-C	Me(1) and CH	0	1.7/-0.9
C-H	CH3 (0)	0	n.a.
C-H	CH3 (1)	-2.0	n.a.
C-H	CH	0.005	-6.8

Table 4.4 Comparison of experimentally observed and predicted rdcs for IBMP butyl chain bond vectors. The predicted values were generated with the PALES program[121] using the 1qy1 crystal structure and backbone amide rdcs from residues in secondary structure elements, i.e. rigid parts of the structure. The experimentally observed rdc values are low. Where no error is quoted for experimental rdcs, HSQC peak optimisation quoted error as 0.00.

4.3.7 CHARMM–AMBER MD comparison

A limited assessment of whether the observed ligand motion is an artefact of the AMBER forcefield was performed. 100 ns of explicitly solvated trajectory was generated using the CHARMM forcefield and parameters and compared to the first 100 ns of AMBER trajectory, as detailed in §4.2.3. The CHARMM trajectory was checked before analysis. Temperature and total energy are stable across the 100 ns. The RMSD, calculated with respect to the first frame, increases from ~1 Å up to ~2 Å as the trajectory explores conformational space further from the native state. There are some brief excursions to higher RMSD. Nonetheless, an increase of ~1 Å in the RMSD value over the trajectory is not indicative of protein instability. Therefore these three data reveal the trajectory to contain no obvious errors, Figure 4.16.

Ligand atom average RMSF values calculated from aligned trajectories are 3.0 $Å^2$ for CHARMM, and 15.3 $Å^2$ for AMBER. This indicates that even within the first 100 ns the ligand is much more mobile using AMBER than using CHARMM, and is confirmed by visual inspection of the trajectories, which reveals that using CHARMM the ligand mostly occupies a single pose, Figure 4.17a.

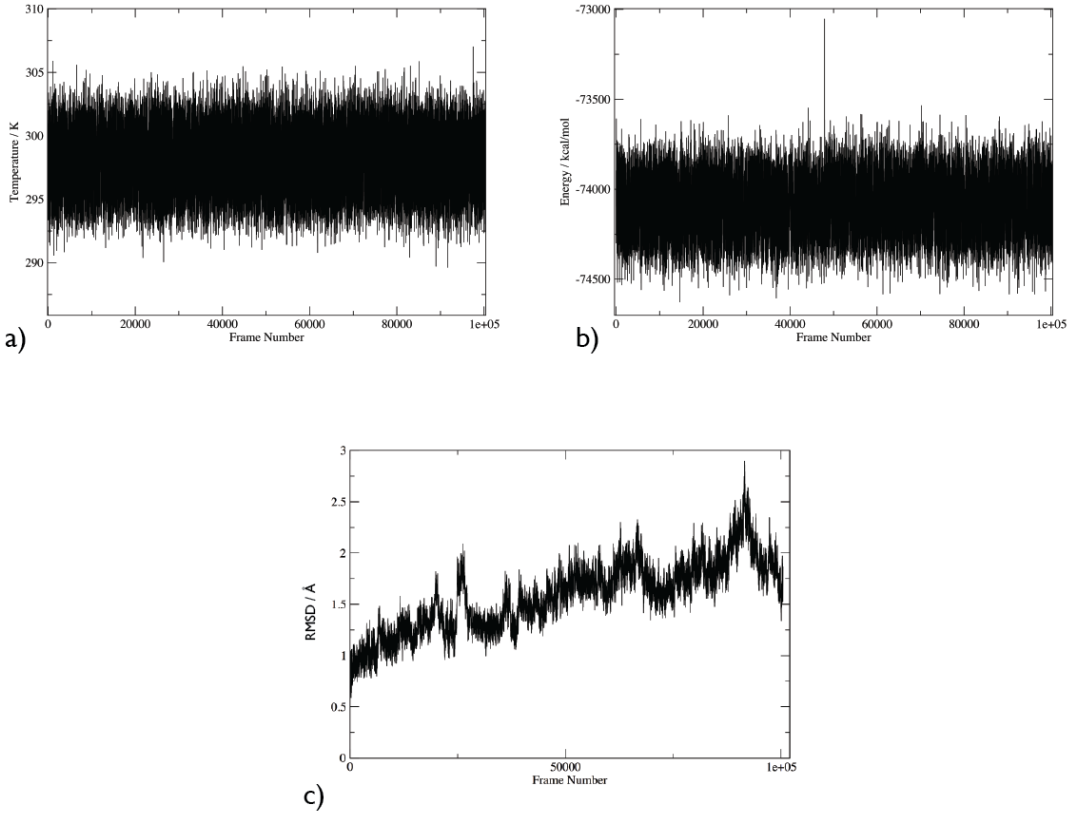


Figure 4.16 Quality control of 100 ns MUP-IBMP CHARMM MD trajectory. a) temperature (K) vs time. b) total energy (kcal/mol) vs time. c) RMSD (Å) vs time, wherein each frame is aligned with the first frame. Energy and temperature are stable. The system explores further from the native state with time, with the RMSD value increasing by 1 Å over the course of the trajectory. These three data reveal the trajectory to contain no obvious errors.

Though the ligand moves more using AMBER, there are only two heavily populated poses for both 100 ns AMBER trajectories analysed, Figure 4.17b+c. Figure 4.17 shows all poses from the first 100 ns compared to poses 1 and 2 from the AMBER trajectory, §4.3.5, whose simulated NOEs agreed best with experiment. It is

clear that the CHARMM pose is quite different from pose 1 (the crystal pose) or 2, whereas the AMBER poses are close to either. Therefore the first 100 ns of AMBER could generate average NOEs close to experiment, whereas CHARMM could not, despite less ligand movement.

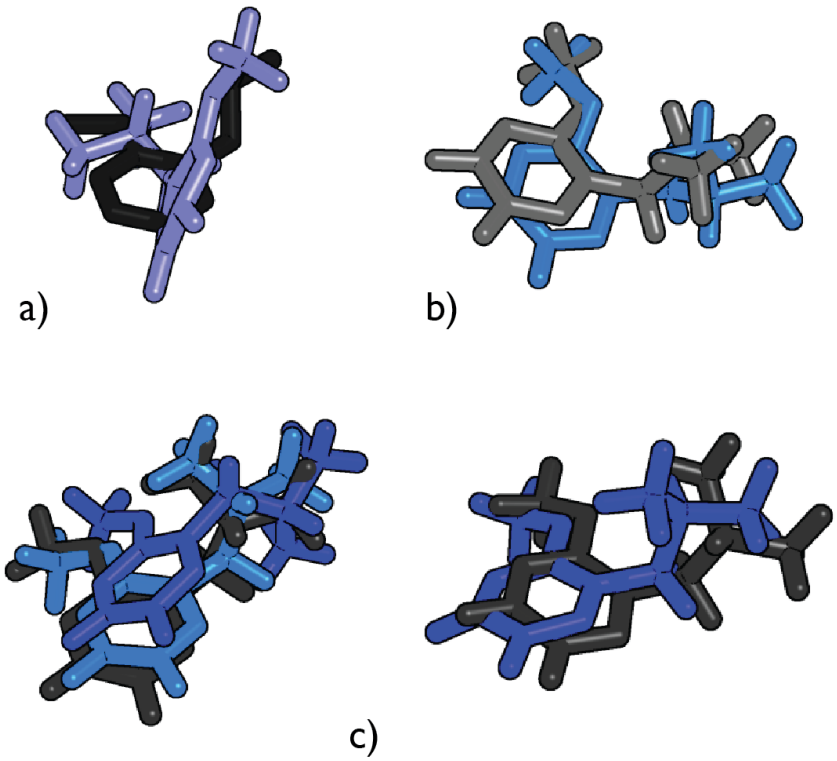


Figure 4.17 Heavily populated poses from first 100 ns of MUP-IBMP MD simulations. a) crystal pose (black) with CHARMM pose (blue). b) pose 2 (grey) with AMBER pose (blue). c) pose 1 (grey) with AMBER poses (blue). Though the CHARMM pose is very different, the AMBER poses are close to pose 1 and 2, which gave the best agreement to observed NOEs, Tables 4.2 and 4.3.

If the AMBER cluster residency times over the whole 1 μs are heavily weighted toward crystal-like poses or positions of methyl and methoxy groups, the AMBER simulations would not strongly disagree with the observed NOEs, and therefore not be obviously artefactual. However, if non-crystal-like poses are well populated, the trajectory would diverge from experimental agreement. Considering the good agreement between simulation and NMR for the protein reported by Roy and

Laughton[26], such divergence would indicate that the observed ligand movement is a ligand-specific artefact for μs timescales.

Clusters generated in analysis by Dr Charlie Laughton are shown with accompanying residence times over the 1 μs trajectory, Figure 4.4. The most highly populated clusters are 3, 4, 5, 6, 7, 8 and 10. By comparing average poses for these clusters with those for which NOEs were simulated, it is possible to assess whether, on average, crystal-like poses or methyl and methoxy positions are observed. Clusters 5 and 10 are close to the crystal structure, Figure 4.18a. However the other 5 highly populated clusters are not, Figure 4.18b. Cluster 3 is similar to poses 4, 6 and 7. Cluster 4 is similar to pose 3. Cluster 7 is similar to pose 8. Cluster 6 is not particularly similar to any of the poses for which NOEs were simulated. Cluster 8 has a methyl position similar to pose 2, therefore resulting in NOEs close to those observed, however a methoxy position similar to poses 5 and 9. Clusters 5, 6, 8 and 10 are those most sampled at short timescales in the AMBER simulation, Figure 4.4, reinforcing the potential sub 100 ns simulation-experimental agreement mentioned above.

In conclusion, the cluster populations observed by MD include significant population of non-crystal-like poses, indicating that the range of ligand orientations observed in the whole 1 μs simulation is inconsistent with experimentally observed NOEs. Rdc data is currently inconclusive, but the partial disagreement of current values may represent butyl chain conformational flexibility, rather than sampling of multiple poses.

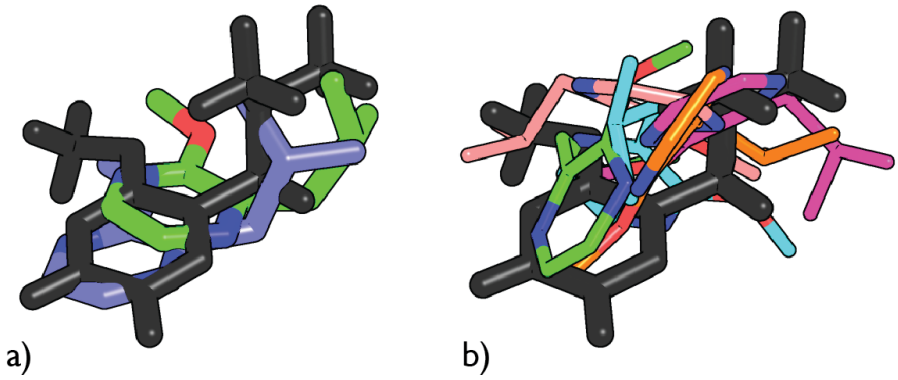


Figure 4.18 Representative structures from clusters in Figure 4.4 compared to the crystallographically-observed ligand pose. a) crystal (grey), cluster 5 (blue) and 10 (green). b) crystal (grey) with clusters 3, 4, 6, 7 and 8.

4.4 Summary and Conclusions

Loss of rotational and translational entropy is a major contributor to protein-ligand binding thermodynamics. 1 μs MD simulations observed significant IBMP movement and population of multiple distinct poses in the MUP pocket, and challenged the previous assumption, used in thermodynamic decomposition, that IBMP lost almost all rotational and translational degrees of freedom upon binding to MUP. For the first time, an attempt was made to directly experimentally observe/assess predicted residual ligand rotation and translation by measuring NMR rdcs and NOEs on a sample of ^{15}N-MUP protein and ^{13}C-^{15}N IBMP. These experiments were complemented by NOE spectra performed on a sample containing ^{13}C-^{15}N MUP bound to unlabelled IBMP. The ligand was synthesised successfully from reagents labelled with stable isotopes, and titrated into labelled MUP to a sub-stoichiometric concentration with minimal excess ligand. Strong butyl peaks were observed for the bound ligand, with dispersion of methyl resonances indicating a restriction of rotation and translation compared to bulk.

The NOESY spectrum acquired for the sample containing ^{15}N ^{13}C IBMP bound to ^{15}N MUP demonstrated a bound-unbound ligand exchange occurring within a 120 ms timescale. Assigning observed NOEs was obstructed by the protein having no ^{13}C label, an intentional labelling strategy to aid resolution of ligand only rdcs. Complementary NOE spectra acquired for a sample containing ^{13}C-^{15}N MUP and an excess of unlabelled IBMP allowed assignment of NOEs from the ligand to pocket residue protons. Ligand orientations in the 1 μs trajectory were assessed for experimental agreement by simulating NOEs for heavily populated poses. The simulated NOEs were calculated taking spin diffusion into account, and indicated that poses close to crystal structure give good experimental agreement. Dr Charlie Laughton had clustered poses from the trajectory, with cluster vs residency time shown in Figure 4.4. A comparison of cluster average structures to simulated NOEs revealed that most of the extensively populated clusters are close to poses with poor experimental agreement, indicating that the orientational sampling represented in Figure 4.4, and thus the 1 μs trajectory, would not result in NOEs with good experimental agreement. However, the first 100 ns contains sampling closer to the crystal structure, suggesting that the forcefield is only insufficient at generating trustworthy ligand behaviour for longer timescale simulations. To control for AMBER specific issues in ligand sampling, a 100 ns trajectory was also produced in CHARMM.

This revealed less ligand movement than AMBER over 100 ns, but a comparison at longer timescales is needed.

Residual dipolar coupling measurements are being undertaken by Dr Arnout Kalverda, Dr Gary Thompson and Mr Phil Morrison. Butyl chain RDCs partially disagree with predicted values for the crystal pose. However, this does not necessarily mean that different ligand orientations are sampled. Experiments are ongoing to acquire a conclusive rdc dataset.

Whilst protein behaviour seems to be well reproduced[26], these experiments provide unique data that raises questions about the validity of ligand behaviour observed in long timescale simulations, which are becoming more commonly used.[49] These experiments also support the understanding, derived from the crystal structure, used in previous thermodynamic decomposition and challenged by the 1 μs simulation, that IBMP has minimal rotational and translation entropy when bound to MUP. However, residual ligand motion may not be completely absent. More extensive measurement of ligand rdcs in the complex should provide a better picture of the amount of residual motion present.

Chapter 5

Protein Dynamics

5.1 Introduction

5.1.1 The importance of developing novel probes of site-specific protein dynamics

Many protein functions depend critically upon structural fluctuations (dynamics) of a picosecond or longer. Examples are the hinge motion critical to the enzymatic cycle of lysozyme[122], or the transport cycles of ion-channels.[123] These motions can be affected by ligand binding. This is most dramatically demonstrated by dynamic allostery, such as that observed in the binding of cyclic AMP to a mutant catabolite activator protein, wherein the protein's affinity for DNA is altered through modification of ensemble-averaged dynamics rather than structure.[124] Site-specific changes in protein dynamics induced by ligand binding are well documented, such as in MUP.[23]

Interesting both mechanistically and thermodynamically, changes in protein dynamics, comprising a number of different types of motion, can potentially be addressed using a range of theoretical and experimental techniques, Figure 5.1. These techniques, which can characterise structural or dynamic states of a protein to different degrees, include inelastic neutron scattering[125], X-ray scattering[126], ion-mobility mass-spectrometry[127], fluorescence polarisation[128], Förster resonance energy transfer[128] and analytical ultracentrifugation. Though some of these techniques report high time resolution, e.g. 100ps for x-ray scattering[126], they generate only global protein parameters for size and shape such as radius of gyration, effective force constant, hydrodynamic volume or cross-sectional area. These can be useful in providing constraints for molecular dynamics (MD). However, higher resolution is required to investigate site-specific changes, and in this case NMR, working at atomic resolution and down to picosecond timescales, is the only sufficiently detailed experimental technique to corroborate MD. The benefits and limitations of NMR are considered below, highlighting the need for development of novel techniques with the potential for probing site-specific protein dynamics.

NMR gives information about dynamics in solution state both at atomic resolution and timescales easily accessible by modern MD. Analyses of NMR relaxation derived S^2 values as a function of ligand binding constitute the experimental 'gold standard' for investigating site-specific changes to protein dynamics upon ligand binding, §1.3.2.6. A range of studies have used this approach, as detailed in Table 1 of reference [129]. NMR residual dipolar couplings provide dynamics information on the nanosecond-millisecond timescale, §1.3.2.3. Furthermore, using NMR and MD in combination has started to address previously unanswerable questions regarding protein structure and dynamics; refining MD data using NMR data to generate experimentally-consistent conformational ensembles for intrinsically disordered proteins[130]; constraining MD simulations with NMR distance and orientation parameters to determine the structure and dynamics of a protein's native state[8]; and identifying long-range correlated motions involved in functional allosteric mechanisms[36].

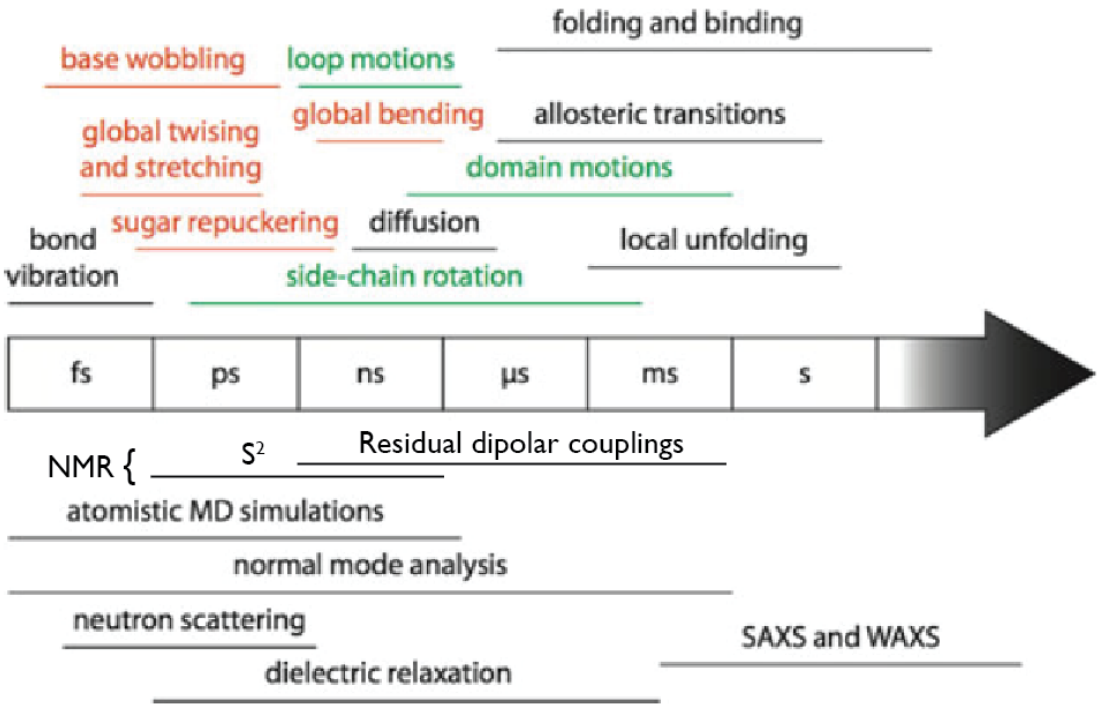


Figure 5.1 Timescales of molecular motions and techniques for their investigation. Modified from reference [131].

However, performing NMR relaxation experiments is time-consuming and involves multiple obstacles. The protein must be successfully produced in a form either fully or selectively labelled with stable isotope. There must exist NMR-compatible solution conditions in which the relaxation parameters can be measured. Furthermore, a sufficient number of the protein resonances must be assigned. Studies do not always conform to this ideal and can provide limited data sets. This is not a problem if sufficient information is acquired to answer the question at hand, but demonstrates the importance of developing quicker novel probes of site-specific protein dynamics when the opportunity arises.

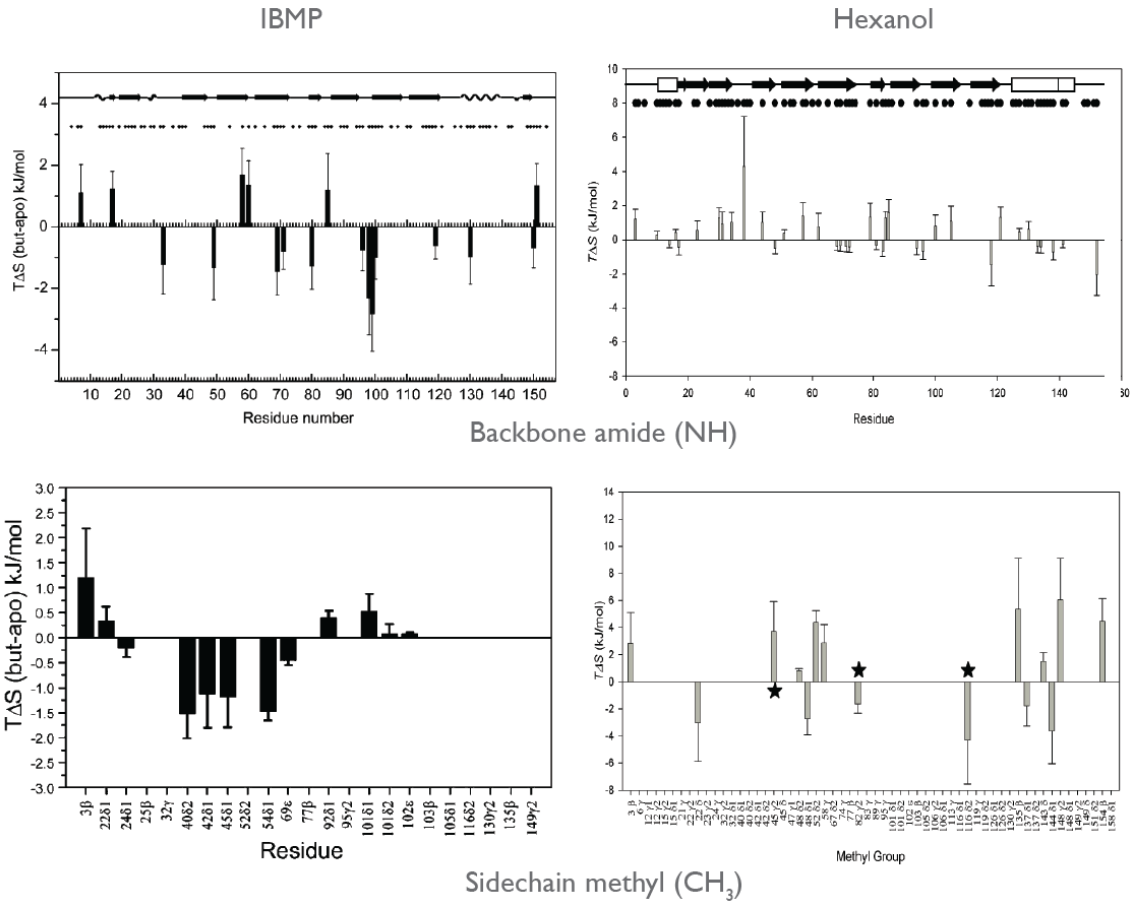


Figure 5.2 Entropy changes upon ligand binding calculated from NMR S^2 values, calculated as described in §1.3.2.6. IBMP data are from reference [23] and hexanol data are from reference [28]. Stars in the bottom right pane are binding site methyl groups. Dots under the structural schematic on the NH represent the residues for which measurements were performed, indicating the proportion of the protein covered by these measurements.

5.1.2 Site-specific changes to MUP protein dynamics upon ligand binding

Site-specific changes in protein dynamics upon ligand binding to MUP have been observed using both NMR and MD. S^2 for backbone amide and sidechain methyl bond vectors were acquired at 298 K for MUP-IBMP[23] and MUP-hexanol[28], both before and after ligand binding. Those backbone and sidechain ΔS^2 above error were converted to $T\Delta S$ values using the procedure outlined in §1.3.2.6, and the resulting figures from the respective studies are reproduced in Figure 5.2. The incomplete coverage of the measurements is clear, wherein out of a total 157 residues, 113 backbone amide protons were measured for hexanol and 82 for IBMP. Nonetheless, despite the small RMSD of the protein atoms before and after binding, < 0.7 Å, clear site-specific changes in flexibility were observed. This led to the suggestion of entropy-entropy compensation in MUP-ligand binding, whereby distal residues increase in flexibility to offset decreasing flexibility in the binding pocket.[23] The 1 μs trajectories produced by Roy and Laughton similarly predict an entropy change of zero within error upon IBMP binding, through an offsetting of site-specific flexibility changes, with the most significant decrease in flexibility occurring around residue 49.[26] Importantly, though the alignment of the protein crystal structures before and after ligand binding shows a small RMSD and therefore little conformational change, there are clearly site-specific changes in protein dynamics upon ligand binding.

5.1.3 THz time domain spectroscopy as a novel probe of protein dynamics: a new system for protein crystals

THz time domain (henceforth referred to as THz) spectroscopy of proteins has been proposed as an experimental technique for the investigation of collective protein vibrations (vibrational modes) relating to conformational or dynamic states.[132] Consequently, changes in such vibrations, for example as a result of ligand binding, are observable in THz spectra.[133] This presents THz as a potential probe of site-specific dynamics, if the spectra can be interpreted in terms of residue level fluctuations. The capacity to observe changes in THz spectra as a function of conformation is proven, for example THz investigation of a photo-controlled reversible conformational change

in bacteriorhodopsin revealed distinct and reproducible THz spectra for each conformation.[132] THz spectra also reflect changes in protein flexibility upon ligand binding, comprising a novel ligand binding assay[133,134]. Evidence of picosecond timescale collective protein motions in THz spectra has been observed very recently.[135]

Though mainly used for small molecules, THz spectroscopy of protein samples has been performed using hydrated thin films, pellets and aqueous solutions.[132,136,137] These sample types suffer in that the structure and hydration of the sample being interrogated is variable. Pellets and films can contain denatured protein.[138] Water absorbs greatly in the THz range, and as such accounting for overwhelming water artefacts is an obstacle for solution studies.[139] Additionally, hydration dependence of THz spectra is well established, due to the participation of hydration water molecules in vibrational motions[140], making hydration control a key element of acquiring spectra.[138] In all cases, a lack of narrow band features seen for small molecules is consistently observed with protein samples, due to vibrational mode overlap arising from the larger protein system.[136,138]

Dr Kasia Tych at the University of Leeds recently developed a THz spectroscopy system that obtains THz spectra from protein crystals at 110 K. This results in exact knowledge of the protein structure because x-ray crystal structures are routinely obtained at 110 K. Likewise, the hydration of the sample is known from the crystal structure, and is kept constant throughout the experiment due to low temperature. Furthermore, whereas neutron scattering measurements may take many hours and possibly hundreds of milligrams of protein, these spectra are acquired within minutes using only micrograms of protein. Briefly describing the THz system, a protein crystal is mounted directly over a metal pinhole aperture and excess solution removed, before being flash-frozen using a nitrogen cryostream. The aperture, which resulted in no diffraction effects, is subsequently placed at the focal point of the THz beam, allowing the entire beam to interact with the sample. A time domain signal is then measured, Fourier transformed and converted into an absorption coefficient using Equation 5.1.

$$\alpha(\nu) = -\frac{2}{d}\left\{\frac{A(\nu)}{A_0(\nu)(1-R(\nu))}\right\}$$

Equation 5.1

This describes, at a given THz frequency, ν, the absorption coefficient, α, in terms of the crystal thickness, d, the time domain signal through crystal, $A(\nu)$, the time domain signal through empty sample cell, $A_0(\nu)$, and the reflection coefficient at the air-sample interface $R(\nu)$, which is calculated using the refractive index.[141] For further details see references [141] and [142].

Subtracting spectra for two states of the crystal results in a difference spectrum, which can therefore be acquired for any crystallographically tolerable intervention, e.g. ligand binding or pH modification. The frequencies present in difference spectra can be interpreted as representing changes in collective vibrational motions of equivalent frequency as a function of ligand binding. Reference to IR spectroscopy, because THz is far-IR spectroscopy, justifies this approach by demonstrating that the spectrum is defined by the absorbance of the vibrational modes at a given frequency. In the commonly used double harmonic approximation, the intensity of absorption arising from mode i, A_i, is proportional to the square of the change in dipole moment as a function of the vibration, Equation 5.2.[143]

$$A_i = \frac{N\pi\, d_i}{3\, c^2} \left|\frac{\partial \mu}{\partial Q_i}\right|^2$$

Equation 5.2

c is the speed of light, d_i is the degeneracy of the mode, N is Avogadro's number and the term in brackets is the differential of the dipole moment, μ, with respect to the mode coordinate Q_i. The time-dependent dynamics of multiple vibrational coordinates generate the spectrum, and thus the absorption coefficient. To briefly demonstrate this, Equation 5.3 is reproduced from reference [135], showing a definition of the absorption coefficient in terms of harmonic oscillations and the refractive index, the latter of which is accounted for in $R(\nu)$ in Equation 5.1. It defines the product of the absorption coefficient, $\alpha(\omega)$, and refractive index, $n(\omega)$, as a time-correlation function of the total dipole operator, where ω is frequency, c is the speed of light, $\beta = (k_B T)^{-1}$, V is the volume, M is the total dipole of the system and M(t) is the time dependence of the total dipole moment.

$$\alpha(\omega)n(\omega) = \frac{2\pi\omega^2\beta}{3cV} \int_{-\infty}^{\infty} dt e^{-i\omega t} \langle \mathbf{M}(0) \cdot \mathbf{M}(t) \rangle$$

Equation 5.3

Recent work with moist film samples has complicated the relationship of frequency to vibrational mode by observing that relaxational terms could also contribute to THz difference spectra, if there are changes in the solvent exposure and rotational motions of surface side chains as a function of binding.[138] However, obtaining THz difference spectra at 110 K where motion is more limited, and between crystals wherein binding is known to effect no change to the crystalline lattice and negligible change to the hydration or structure of the protein, RMSD < 0.7 Å, means that these potential contributions are assumed as negligible for this work.

Though the crystal environment at 110 K undoubtedly limits protein dynamics, it has been demonstrated that solution MD and crystallographic B-factors can both represent the same dynamic trends across a protein backbone.[100] Accordingly, NMR S^2 and crystallographic B-factors can correlate.[144,145] Therefore, this work is a preliminary analysis of THz difference spectra frequencies in terms of changes in collective vibrational modes, asking whether such analysis can indicate site-specific changes to dynamics similarly to the NMR observations detailed in §5.1.2.

5.1.4 Normal mode analysis (NMA): a theoretical counterpart to THz spectroscopy

The calculation of the frequency and collective motion of vibrational modes for a system is achieved through performing a normal mode analysis (NMA). A non-linear system has 3N-6 normal modes, where N is the number of atoms. The first 6 modes are removed because they detail translational and rotational motions that do not report on the internal dynamics of the molecule. Modes do not interact with each other: each mode is independent of all others. The lowest frequency modes involve global motions with many atoms undergoing larger displacements. The highest frequency modes involve localised motions with displacement of smaller numbers of atoms. For proteins, a well-defined native state necessitates that the protein is in an energy basin, allowing the energy potential to be assumed as harmonic when the native structure is thoroughly minimised; the harmonic oscillations around the minimum are the normal modes. Neutron scattering and THz spectroscopy measurements reveal a 'dynamical transition' whereby protein fluctuations start to deviate from harmonicity above 180 K.[146,147] This limit has been recently shown to extend down to 110 K in some cases[148], but for MUP was measured as ~130 K.[141] Accordingly, at higher

temperatures the harmonic approximation of NMA is unsuitable. However, due to the 110 K operating temperature of the new THz system, NMA is suitable to model the harmonic oscillations of MUP.

In NMA, a Taylor expansion of the MD potential energy function around the minimum is performed. The second-order partial derivatives of the potential energy function, the force constants for the harmonic oscillations, are placed in a mass-weighted matrix. Diagonalisation of this matrix results in the eigenvalues and eigenvectors corresponding to the collective atomic displacements and frequencies that define the modes.[149] The procedure is quick, taking only a few hours on modern computers. Collective motions can be analysed for a single mode or superposition of multiple modes. The relative amplitudes of the collective motions are independent of temperature: the absolute amplitudes of the motions, sometimes useful for experimental comparisons, can be scaled using a temperature factor. In this work the amplitude of the modes is not considered, as only the frequencies in difference spectra are necessary to identify affected modes.

The structure and the hydration level used both affect the NMA, and therefore these two factors need to be considered carefully. The native well of an energy landscape contains multiple sub-minima, each of which can be approximated harmonically, Figure 5.3. Consequently, traditional single-structure NMA (SS NMA) results in some modes idiosyncratic to a single minimum in the native well. However, averaging NMA from multiple sub-minima in the native well has been shown to diminish idiosyncratic modes whilst enhancing the modes common to sub-minima, thus generating a more representative picture of native vibrational fluctuations.[122,150] The speed of NMA is not greatly undermined because short trajectories are sufficient to generate the non-identical poses required.[122] This refinement, native ensemble NMA (NE NMA), is used herein for the first time with THz spectroscopy. All-atom explicitly solvated MD trajectories were generated to sample the native basin, at both the THz and NMR experimental temperatures of 110 K and 298 K. Additionally, SS NMA is performed using the crystal structures. NE NMA trajectories simulating the crystal asymmetric unit are currently being developed, and trajectories of the crystal unit cell are being considered.

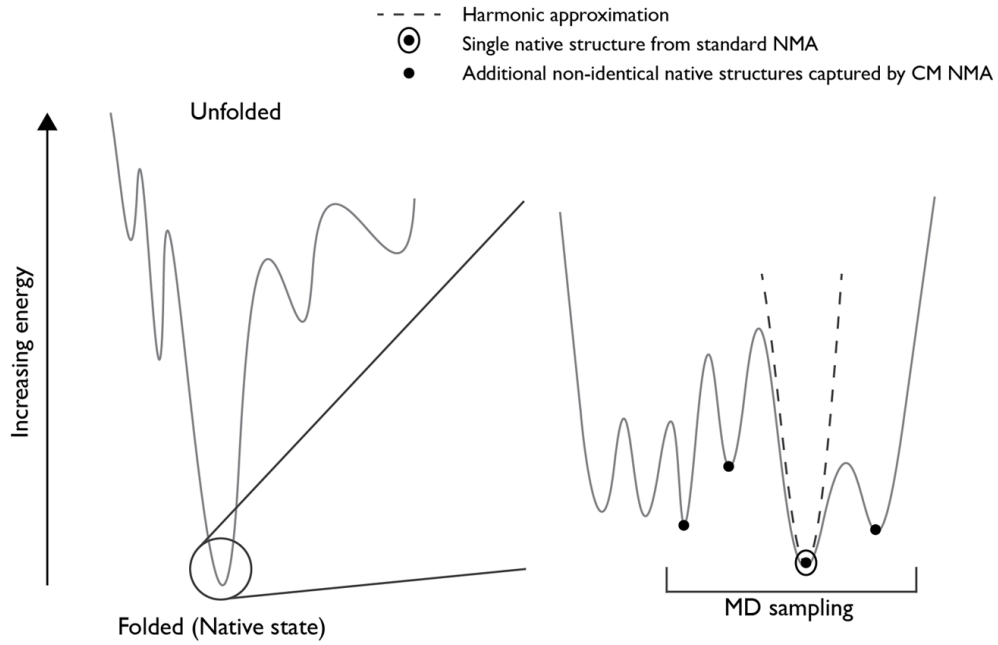


Figure 5.3 A protein energy landscape with detail of the native state energy well. Traditional single-structure NMA (SS NMA) results in some modes idiosyncratic to a single minimum in the native well. However, averaging NMA from multiple minima in the native well diminishes idiosyncratic modes whilst enhancing common modes. Multiple minima are sampled by generating a short MD trajectory.

Water molecules are included firstly to avoid the collapse of protein surface elements during the thorough minimisation procedure. Secondly, the hydration dependence of THz spectra, whereby increasing water increases absorbance, demonstrates that hydration water contribute to vibrational modes.[142] Likewise, increasing hydration increases normal mode densities at lower frequencies. Therefore a realistic hydration level must be applied in NMA. This reveals another benefit of the new THz system, wherein accurate (i.e. closer to experimental) NMA hydration is simpler to achieve than for previous systems. This is because the unit cell hydration is known from the crystal structure, and remains consistent across the experiment due to low temperature. Accordingly, for SS NMA, the protein was hydrated to replicate the asymmetric unit water density observed in the crystal. This resulted in asymmetric hydration due to the structure of the asymmetric unit, with a depth of water molecules extending between 2 and 10 Å from the protein surface. Consequently, for this preliminary investigation, structures from the explicitly solvated MD trajectories

used for NE NMA were edited to only contain water molecules within 5 Å of the protein surface, approximately the middle of this range.

Importantly for this work, the qualitative correspondence of fluctuations described by NMA and NMR S^2 values has been described[151,152]. Therefore using NMR S^2 as benchmark data for this preliminary investigation is justified. Herein NMA is used as a method of interpreting the frequencies in THz difference spectra in terms of changes in collective vibrational modes, and thus site-specific dynamics, upon ligand binding. Additionally, the capacity of NMA to predict ligand-induced site-specific RMSF changes without the use of THz difference spectra is considered, alongside the effect on mode densities of the different NMA approaches.

5.1.5 Work undertaken

This work is a preliminary investigation into the combination of two complementary and developing techniques to comprise a novel probe of site-specific dynamic changes upon ligand binding, namely NMA-interpreted THz difference spectra. 'Gold standard' NMR S^2 data are used as a benchmark for site-specific changes in protein dynamics upon ligand binding to MUP. Protein, crystals and ligand solutions were produced for THz spectroscopy. Dr Kasia Tych acquired THz difference spectra for the binding of both IBMP and hexanol to MUP. The THz system acquires spectra at 110 K, thus NMA's harmonic approximation of protein fluctuations is justified. MD trajectories are generated to provide conformational samples for NE NMA at temperatures corresponding to the THz and NMR experiments, 110 K and 298 K, and are checked for unexpected behaviour before being analysed. NE NMA is compared to traditional SS NMA with regards to the agreement of THz difference spectra-derived predictions with NMR S^2 data. The capacity of NMA to predict ligand-induced changes without the use of THz difference spectra, and the effect on mode densities of the different NMA techniques are also considered.

5.2 Materials and Methods

5.2.1 THz spectroscopy

5.2.1.1 Production of crystals for THz spectroscopy

Proteins and crystals were produced as described in §2.2.1.1 and §3.2.2.1. IBMP was introduced to the crystals through addition to reservoir solution and overnight vapour-diffusion as described in reference [21]. Hexanol was introduced by soaking crystals in ligand-doped reservoir solution as described in §3.2.2.1. Soaking was performed by Dr Kasia Tych.

5.2.1.2 Generation of THz difference spectra

These experiments were performed by Dr Kasia Tych, with details summarised from reference [141]. Crystals were transferred onto a pinhole aperture and excess solution removed using a paper wick before being flash frozen using a nitrogen cryostream. The aperture is positioned at the focal point of the THz beam, such that the entire beam interacts with the sample. No diffraction effects were observed due to the use of the aperture. A time domain signal is measured from a broad-bandwidth THz frequency pulse applied to the sample, Fourier transformed and the absorption coefficient calculated using Equation 5.1 and parameters measured as detailed in reference [141]. The frequency components measured were from 0.3 up to 7.5 THz. All measurements were obtained at ~110 K. Reported absorption coefficients are from eight measurements of four crystals of each complex, and ten measurements of five crystals of the unbound protein, wherein each measurement itself is an average of five THz scans. Uncertainties in the THz absorption coefficients are propagated from uncertainties in the variables of Equation 5.1. Subtracting absorption coefficients and propagating their errors at each frequency resulted in difference spectra, Figure 5.6b. Frequencies at which the difference is zero within error were then discarded. Changes above error are displayed as absolute values minus the error, Figure 5.6c. A list of the accompanying frequencies was used with NMA as described in §5.2.2.3.

5.2.2 Normal mode analyses (NMA)

5.2.2.1 Single structure normal mode analysis (SS NMA)

The crystal structure of MUP (2ozq), MUP-IBMP (1qy1) and MUP-hexanol (Dr Caitriona Dennis, unpublished), were processed using CHARMM GUI 'Quick MD Simulator'.[118] A disulphide bridge was specified between residues 64 and 157 as observed in the crystal structure. The structure was solvated to the density observed experimentally in the asymmetric unit and neutralised using Na^+ placed using a distance method. Protein parameters were taken from the CHARMM22 forcefield. The cadmium ion (from crystallisation solution) and ligand parameters for IBMP and hexanol were generated by CHARMM-GUI using the CHARMM Generalised Forcefield (CGENFF).[47] These structures were then thoroughly minimised, to an energy gradient lower than 10^{-12} kcal mol^{-1} $Å^{-1}$, and subsequently analysed using the VIBRAN module of CHARMM. RMSF values are obtained per mode or for all modes using scripts obtained from Dr Roland Stote (personal communication).

5.2.2.2 Equilibrium simulations and native ensemble normal mode analysis (NE NMA)

5.2.2.2.1 Generating trajectories

The crystal structure of MUP (2ozq), MUP-IBMP (1qy1) and MUP-hexanol (unpublished), were used as the starting structure for the simulation. The 'Quick MD Simulator' functionality of CHARMM-GUI was used to generate ligand parameters, neutralise and solvate the system, and set up periodic boundary conditions.[118] Protein parameters were taken from the CHARMM22 forcefield. The ligand parameters for IBMP and hexanol were generated by CHARMM-GUI using the CHARMM Generalised Forcefield (CGENFF).[47] A disulphide bridge was specified between residues 64 and 157 as observed in the crystal structure. Sufficient K^+ and Cl^- were placed using a Monte-Carlo method to achieve neutrality. The molecule was solvated in an octahedral box of ~7000 TIP3 water molecules with periodic boundary conditions. Long range electrostatic interactions were treated using the particle mesh Ewald method. SHAKE

was applied to constrain all hydrogen bonds, allowing a 2 fs timestep. Minimisation involved 1000 steps using the Steepest Descent algorithm followed by 100 steps using the Adopted Basis Newton-Raphson algorithm. The system was subsequently heated from 50 K to 298 K over 25000 steps, before performing 18 ns of equilibrium trajectory at constant temperature and pressure (298 K, 1 atm). Coordinates were saved every 1 ps.

5.2.2.2.2 Checking trajectories

Total energy and temperature as a function of time were obtained directly from CHARMM, Figure 5.4. RMSD and RMSF calculations, Figure 5.5 and Table 5.1 respectively, were performed using Wordom.[48] The latter were calculated for three subsets of atoms: alpha carbon (C_α), backbone nitrogen (N) or sidechain carbon ($C_{\beta,\delta,\gamma}$), using trajectories wherein each frame was aligned using the same subset of atoms.

5.2.2.3 Processing of normal mode analyses

Using Wordom[48], 100 frames at equal spacing (1 per 180 ps, starting at frame 1) were extracted from each trajectory, and reoriented and aligned with respect to the first frame of the trajectory. All TIP3 water not within 5 Å of protein were removed using CHARMM. These structures were then minimised and processed as detailed in §5.2.2.1.

The following procedures were performed on three datasets for both ligands: the single structure (§5.2.2.1) and the average over all 100 structures extracted from both 110 K and 298 K trajectories. RMSF values are obtained per mode or for all modes, averaged over all 100 structures, using scripts mentioned in §5.2.2.1. Heat maps in Figure 5.7 were generated using JColorGrid software.[153]

5.2.2.3.1 Mode density calculations

CHARMM reports the normal mode frequencies as wavenumbers. Equation 5.5 was used to convert wavenumbers to THz frequencies for comparison to difference spectra. Equation 5.5 is derived from Equation 4.4, which shows that frequency, ν equals the speed of light, c, divided by the wavelength, λ. Because the wavelength is the reciprocal of the wavenumber, Equation 5.4 becomes Equation 5.5.

$$\nu = c / \lambda$$

Equation 5.4

$$\nu = c * wavenumber$$

Equation 5.5

For each structure, a histogram was generated of THz frequency versus number of modes, using a bin width of 0.05 THz. For the NE NMA datasets, the bin values were averaged over all 100 structures. The data are displayed in Figure 5.8.

5.2.2.3.2 Treatment of NMR S^2 values

Raw NMR S^2 values for hexanol-MUP were obtained from reference [28], and for IBMP-MUP[23] were obtained from Prof Steve Homans. Only those bond vectors with S^2 for both bound and unbound samples were kept. Subtracting unbound from bound values yielded the change upon binding, ΔS^2. The errors in the S^2 were propagated and only those values with changes above error were kept. To report an increase in flexibility as a positive change, the signs of the ΔS^2 values were inverted.

Heat maps shown in Figure 5.7 use the following colour scheme. Those residues with no ΔS^2 values are shown in red. ΔS^2 values that are zero within error are shown as white, whereas those above error are represented in greyscale according to magnitude (dark = greater change).

5.2.2.3.3 RMSF calculations

Per-atom RMSF values averaged over all modes (and over all 100 structures for the NE NMA calculations) were calculated, and the average per-residue value obtained. The per-residue variance was used as an estimate of error. Averages were subtracted (ligand bound versus unbound) and error propagated. Those residues for which an above-error RMSF change is observed as a function of ligand binding are displayed as black boxes in Figure 5.7c+d.

5.2.2.3.4 Interpretation of THz difference spectra using normal mode analyses

For each structure, modes were selected whose frequency corresponded to the THz difference spectrum above-error changes as defined in §5.2.1.2. For each selected mode, the per-residue RMSF was calculated and the ten residues with the highest values selected. These 'top ten' from every selected mode were then pooled, and the ten most often mentioned residues from these lists taken as the 'top ten' list for that structure. For NE NMA, these lists were then pooled for all 100 structures and the ten most often mentioned residues again selected. For each difference spectrum, this procedure was performed on both bound and unbound structures, as the mode-RMSF relationship is different in each case, and both will contribute to the observed difference spectrum. These 'top ten' datasets are presented in Figure 5.7a+b.

5.3 Results and Discussion

5.3.1 Analysis of trajectories used for native ensemble normal mode analysis

To provide conformational sampling for NE NMA corresponding to the THz and NMR experimental temperatures, MD trajectories of MUP, IBMP-MUP and hexanol-MUP

were generated at 110 K and 298 K, as detailed in §5.2.2.2.1. It is critical that MD trajectories are checked for any unexpected behaviour before being analysed, as per §4.3.7, involving monitoring energy, temperature, RMSD and RMSF over the course of the simulation.

Energy and temperature values, Figure 5.4, are consistent across the entire trajectory. The energy reveals a very short (< 500 ps) equilibration period at 110 K due to slower sampling at the lower temperature. Energies at higher temperatures are expectedly observed due to increased thermal energy in the system. Energy differences are observed between the simulations, albeit smaller than those induced by temperature variation. These likely arise from bound ligand energies because no relation to greater protein instability is revealed by RMSD and RMSF analysis (below).

To monitor drift from native state and sampling of partially unfolded states, RMSD versus time were analysed for all six MD simulations, and averaged over three subsets of atoms: C_{α}, N and $C_{\beta,\delta,\gamma}$, Figure 5.5. These values, calculated with respect to the initial frame of the trajectory, show no difference between the three systems despite the energy difference seen above. A temperature dependence is observed, whereby the greater temperature at 298 K allows increased conformational sampling than at 110 K, i.e. 0.5-2 Å, compared to 0.3-0.6 Å. Sidechain atoms are more variable than backbone atoms, though all increase gradually on average over the course of the trajectory as conformational sampling moves further from the native state with simulation length. However, these values are low enough to not indicate unfolding.

The adherence to native like structure and lack of partial unfolding shown by the RMSD values is supported by protein RMSF analysis, performed for all atoms and different structural segments, Table 5.1. Low RMSF values are observed for all analyses whilst repeating the pattern of higher RMSF at higher temperature and higher values for sidechain atoms. Loop regions are expectedly far more mobile than residues in sheets and the binding pocket. There is no RMSF difference between the systems, rendering the energy differences observed in Figure 5.4 as most likely arising from the bound ligand itself.

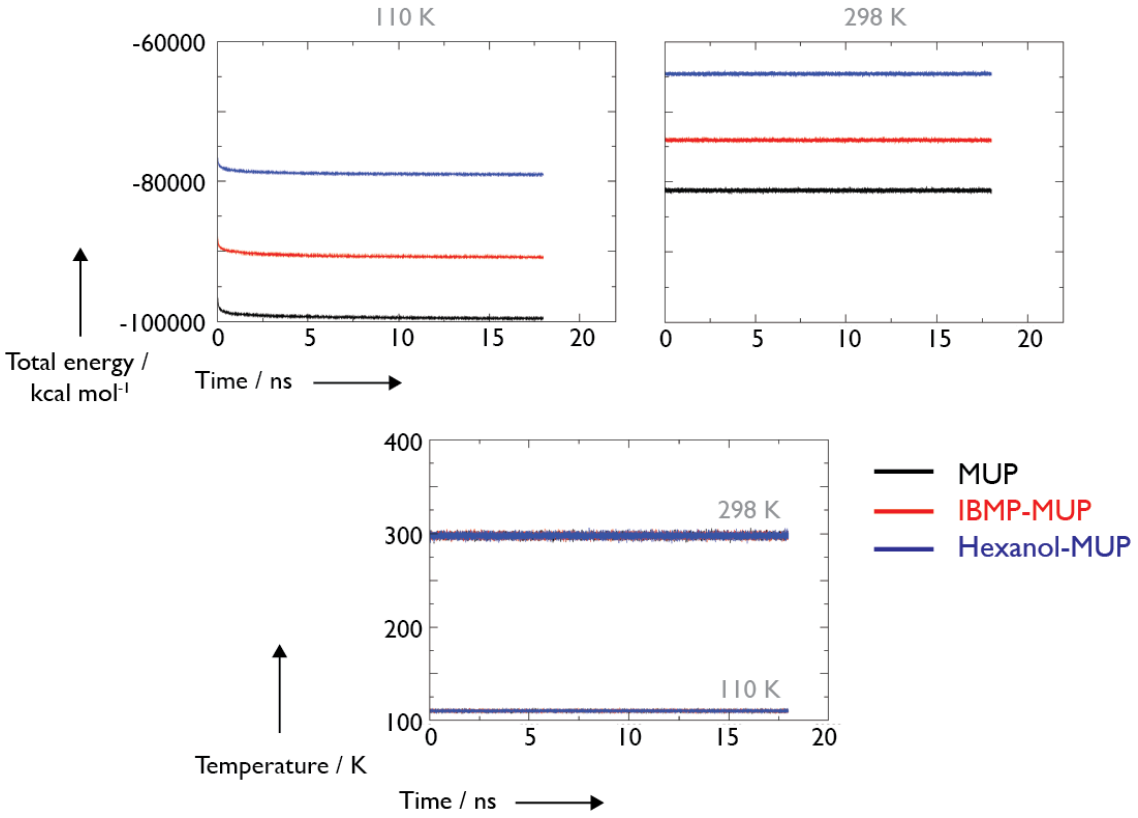


Figure 5.4 NE NMA CHARMM MD simulations: total energy and temperature. Black is MUP, red is IBMP-MUP and blue is hexanol-MUP. (top) total energy versus time for all six simulations, with 110 K on the left and 298 K on the right. Values are in kcal/mol. (bottom) Temperature versus time. Values are in K. All values are consistent across the trajectory, and only a very short (< 500 ps) equilibration is observed at 110 K, due to the slower sampling at lower temperature. The energies are higher at 298 K than 110 K, as a result of increased thermal energy in the system. Differences are observed in the total energy of different complexes, due to the energy of the bound ligands.

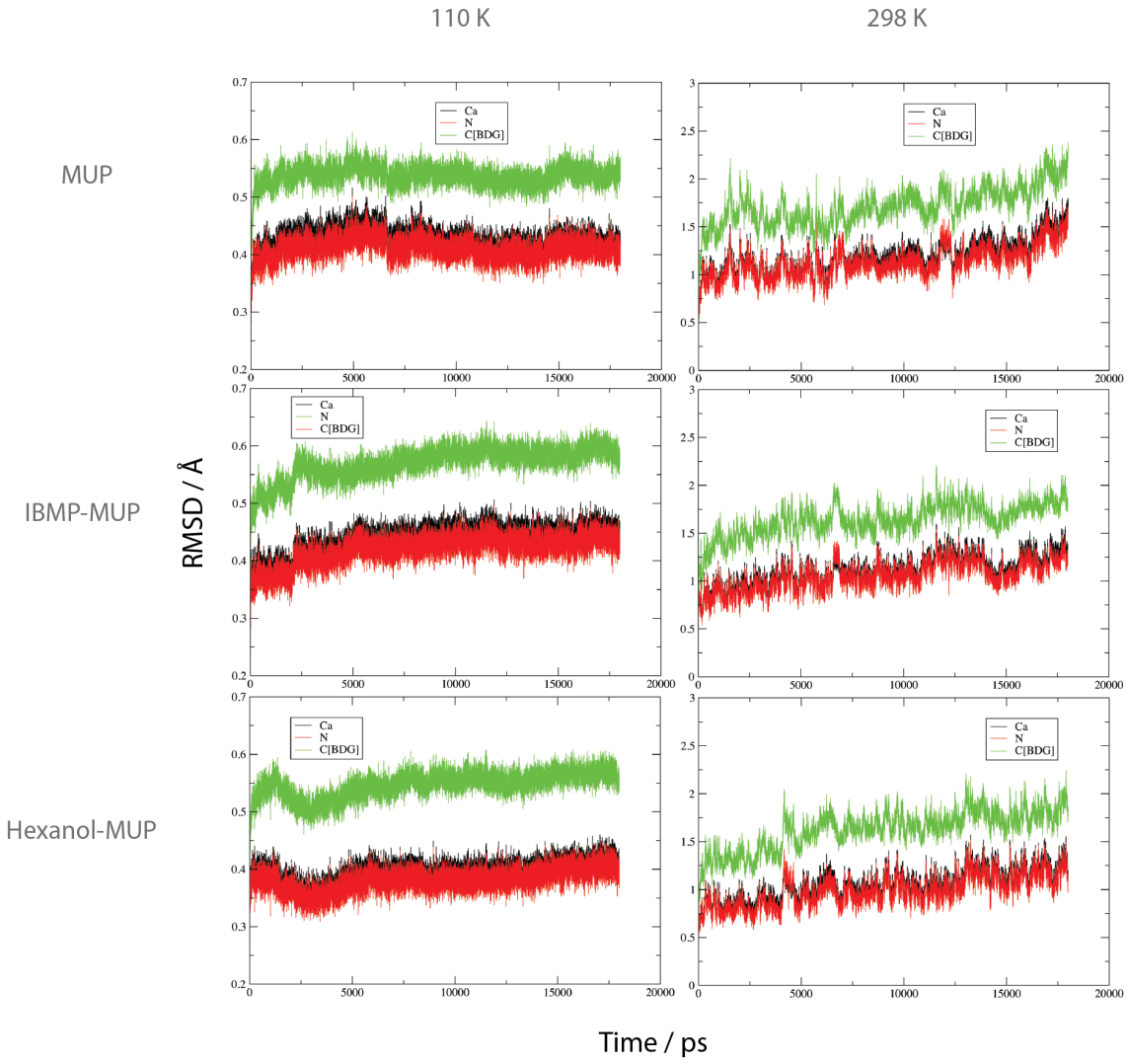


Figure 5.5 NE NMA CHARMM MD simulations: RMSD versus time for all six MD simulations averaged over three subsets of atoms: C_α (black), N (red) and $C_{\beta,\delta,\gamma}$ (green). All values are in Å. RMSD are calculated for each frame with respect to the initial frame of the trajectory. Sidechain atoms are more variable than backbone atoms, though all increase gradually on average over the course of the trajectory as conformational sampling moves further from the native state. The increased temperature at 298 K results in greater conformational sampling than 110 K: 0.5-2 Å, compared to 0.3-0.6 Å. However, these values are too low to indicate unfolding.

	MUP		IBMP-MUP		Hexanol-MUP	
	110 K	298 K	110 K	298 K	110 K	298 K
			OVERALL			
C_{α}	0.39	1.03	0.40	0.95	0.36	0.91
N	0.38	0.99	0.39	0.92	0.35	0.87
$C_{\beta,\delta,\gamma}$	0.49	1.50	0.51	1.39	0.49	1.37
			LOOP			
C_{α}	0.43	1.23	0.45	1.07	0.37	1.11
N	0.41	1.19	0.43	1.04	0.35	1.04
$C_{\beta,\delta,\gamma}$	0.47	1.45	0.49	1.27	0.44	1.38
			HELIX			
C_{α}	0.33	0.82	0.32	0.82	0.37	0.86
N	0.31	0.80	0.32	0.82	0.38	0.83
$C_{\beta,\delta,\gamma}$	0.41	1.20	0.40	1.21	0.49	1.17
			POCKET			
C_{α}	0.30	0.76	0.28	0.66	0.31	0.56
N	0.30	0.72	0.28	0.60	0.30	0.57
$C_{\beta,\delta,\gamma}$	0.37	1.00	0.34	0.86	0.39	0.70
			SHEET			
C_{α}	0.32	0.76	0.32	0.70	0.31	0.67
N	0.32	0.73	0.32	0.67	0.31	0.66
$C_{\beta,\delta,\gamma}$	0.40	1.07	0.40	1.02	0.41	0.92

Table 5.1 NE NMA CHARMM simulations: RMSF values by atom type and structural element for each simulation. All values are in $Å^2$. Values were generated as described in §5.2.2.2.2. All values are low, indicating no large fluctuations that would suggest partial or complete unfolding. Values are expectedly relatively greater at higher temperature and for sidechain atoms. Loop regions are expectedly far more mobile than sheet residues and those in the binding pocket.

5.3.2 THz difference spectra

Protein and crystals were produced for use in THz spectroscopy, as detailed in §5.2.1.2. Dr K Tych acquired THz spectra for MUP, IBMP-MUP and hexanol-MUP in crystalline form at 110 K. Uncertainties in the THz absorption coefficient are propagated from all parameters in Equation 5.1, yet mostly arise from uncertainty in the crystal thickness, d. Each spectrum results from transmission of a broad-bandwidth THz frequency pulse through the crystal, averaged over multiple experiments, Figure 5.6a. The frequency components measured were from 0.3 up to 7.5 THz, at intervals of 0.07 THz. Difference spectra were generated by subtracting absorption coefficients and propagating errors accordingly, Figure 5.6b. Changes above error can be seen as those values where the error bar does not cross the zero line. Figure 5.6c displays those changes above error as their absolute values. Despite the small RMSD values of the complexes to the apo protein, i.e. no conformational change, the crystalline environment and the 110 K temperature, the vibrational changes upon ligand binding are sufficient to generate ligand-dependent above-error difference spectra. Lists of the frequencies shown in Figure 5.6c are used with NMA in §5.3.3 to attempt their interpretation in terms of collective vibrational motion changes upon ligand binding, which can be related back to changes in site-specific fluctuations.

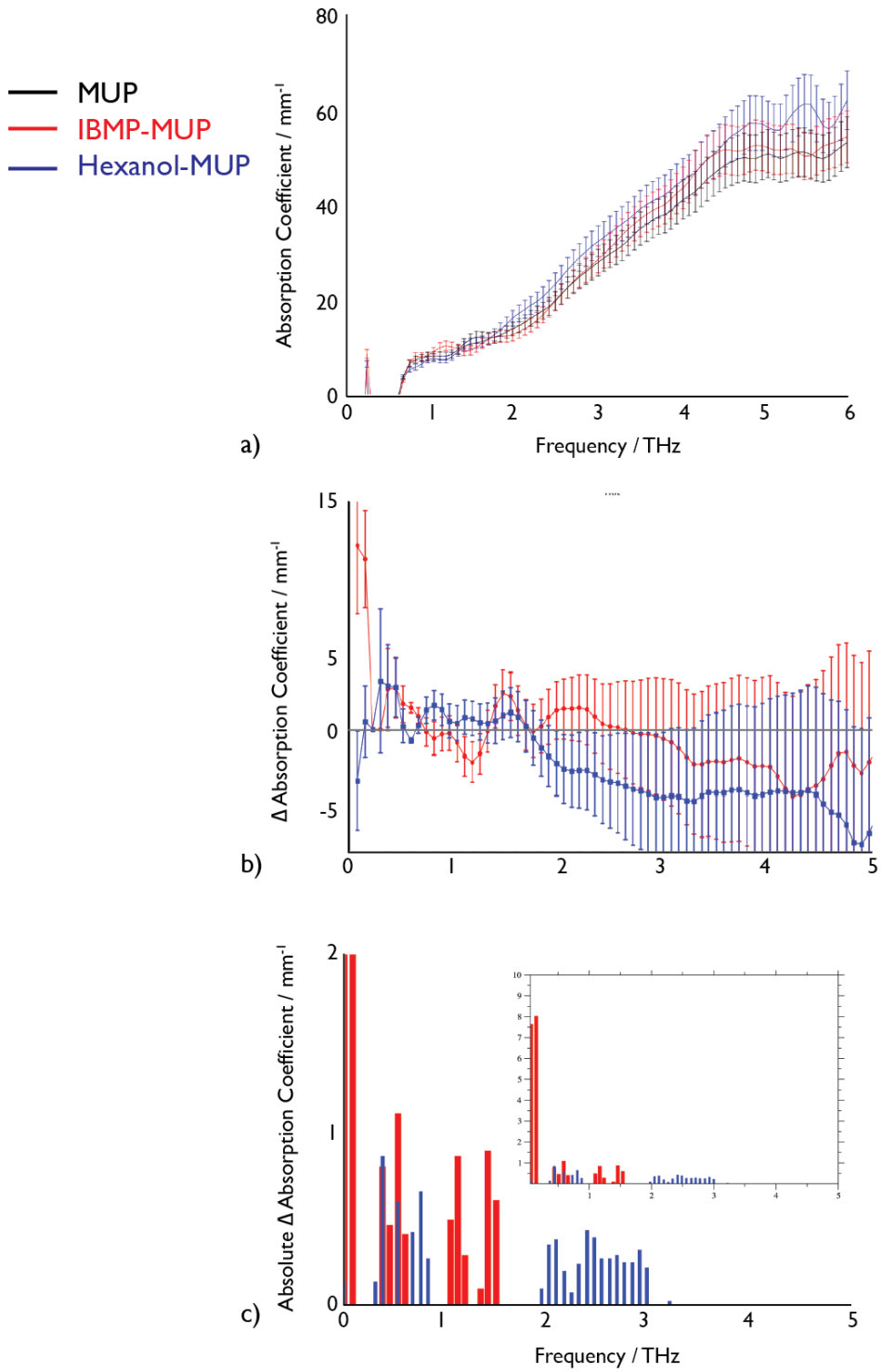


Figure 5.6 THz spectra and difference spectra for MUP (black), MUP-IBMP (red) and MUP-hexanol (blue). a) THz spectra for crystalline MUP with IBMP and hexanol. Only positive values are shown. Error bars arise as described in §5.2.1.2. b) THz difference spectra generated from a). Propagated error bars reveal which frequencies have changes above error, i.e. the error bars do not cross the zero line. c) Only those differences above error from b) are shown as absolute values. The inset shows the full scale of the leftmost bars.

5.3.3 Normal mode analysis-interpreted THz vs NMR data: site-specific changes in protein dynamics upon ligand binding

The frequencies at which above-error changes were observed in the THz difference spectra were used to provide an NMA interpretation at residue resolution, as detailed in §5.2.2.3.4. This was achieved by picking the modes with corresponding frequencies, identifying the residues whose fluctuation is most affected by those modes, and then averaging these results over modes and structures, the latter for NE NMA only. Due to the incompleteness of the comparison NMR data, only the ten most affected residues were considered for comparison. This provides a higher chance of successfully assessing whether NMA-interpreted THz difference spectra generally agrees with NMR, despite the incomplete data. Here the common interpretation of trends in ΔS^2 across the backbone is used, i.e. representing areas of differential flexibility, in this case upon ligand binding.[34]

Results are displayed in Figure 5.7a+b, using the following colour scheme. Those residues with no ΔS^2 values are shown in red. ΔS^2 values that are zero within error are shown as white, whereas those above error are represented in greyscale according to magnitude (dark = greater change). For all three NMA, i.e. SS and NE at both temperatures, the interpretation of THz difference spectra was performed on both bound and unbound structures, because the mode-RMSF relationship will differ in each whilst both will contribute to the observed difference spectrum. The ten most affected residues are displayed as black boxes in Figure 5.7a+b, by ligand and type of NMA. There are 20 predictions per NMA type per ligand, 10 for apo, 10 for holo. It is interesting to note that irrespective of the MD temperature, the NE NMA averaging results in a greater spread of predictions across the protein backbone compared to SS NMA or even the first frame of the NE NMA, for which all predictions are in the C terminus.

The alignment of these THz-NMA predictions with the NMR data is summarised in Table 5.2 according to two metrics: 'incorrect' and 'hit sum'. 'Incorrect' represents the proportion of NMA predictions (black squares) that NMR data reveals to be incorrect, i.e. for residues with ΔS^2 that are zero within error (white squares). This is measured out of 20, because there are 20 predictions per NMA type per ligand: 10 for apo and 10 for holo. Due to there being two ΔS^2 columns, each white square next to a black square scores half a point, i.e. if a residue is predicted for which both

NH and CH_3 ΔS^2 are zero within error, the score is 0.5+0.5 = 1. Therefore this value should be minimal for this technique to demonstrate utility, because the NMA predictions represent only the top ten affected residues, which should not have ΔS^2 that are zero within error. Table 5.2 demonstrates that in fact these values range from 1 to 6, wherein the value decreases, i.e. improves, as SS NMA → NE NMA 110 K → NE NMA 298 K. To aid interpretation, the expected value from random placement of predictions was calculated. Only NE NMA 298 K performs better than random. 'Hit sum' represents the degree to which predictions are for residues with higher ΔS^2 values. The 'hit sum' value reports the sum of all above-error ΔS^2 values aligning with predictions. Therefore the higher the value, the better the prediction of residues whose site-specific dynamics are known (using NMR S^2 data) to change upon ligand binding. NE NMA 110 K has the highest total hit sum for both ligands, and NE NMA 298 K the worst.

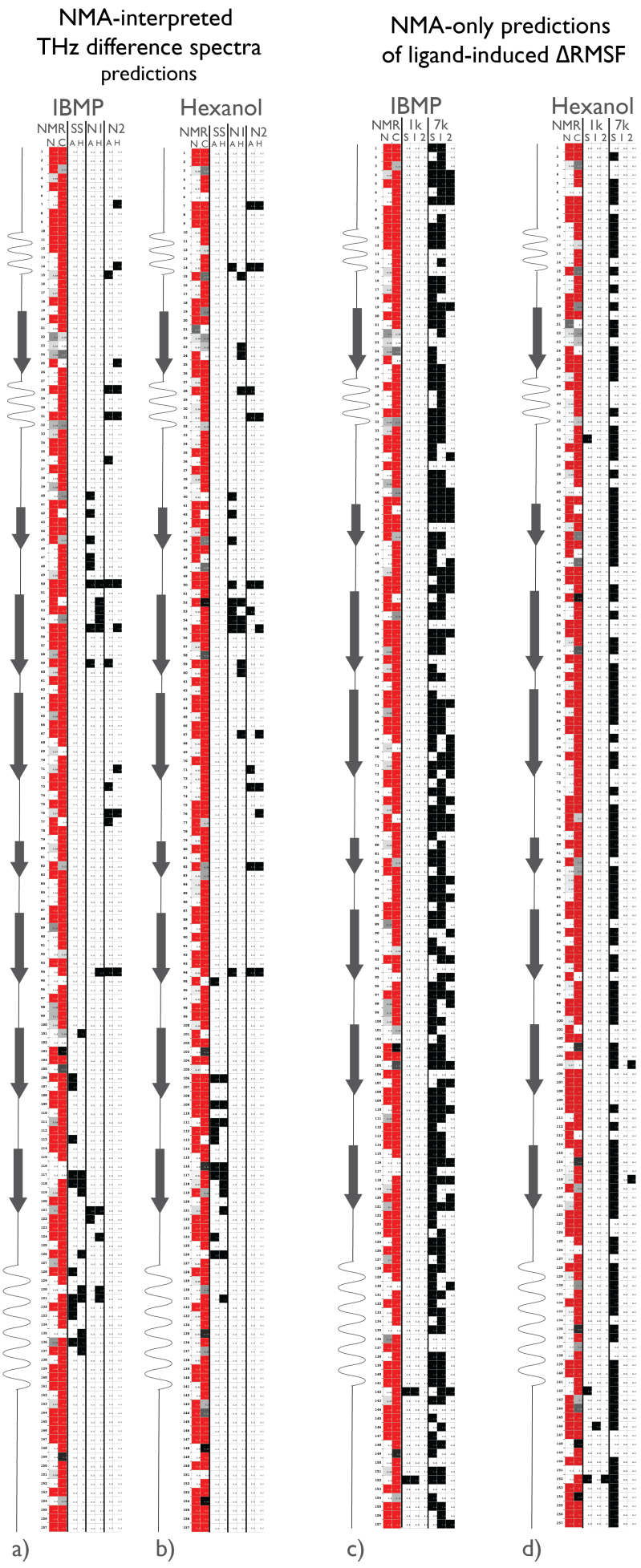
Together, these metrics reveal two observations with regard to generating predictions closer to the most dynamically affected residues as observed by NMR; both that NE NMA performs better than SS NMA, and that sampling at the THz temperature of 110 K may be superior. The power of both metrics would scale with increasingly comprehensive NMR data. The mediocre performance in this preliminary assessment highlights the requirement for experimental systems with more complete NMR S^2 data for conclusive assessment of this approach as a reliable probe of site-specific changes in protein dynamics. Additionally, improving the NMA modelling of the system, by performing NE NMA using trajectories performed on the crystal asymmetric unit, and potentially optimising hydration, could lead to improvement of this agreement.

	IBMP				Hexanol			
	r	SS	NE NMA 110 K	NE NMA 298 K	r	SS	NE NMA 110 K	NE NMA 298 K
Incorrect	3.4	6.0	4.0	3.0	2.7	5.5	3.5	1.0
Hit sum		0.46	0.59	0.28		1.46	2.32	0.74

Table 5.2 NMA-interpreted THz vs NMR: summary of agreement from Figure5.7a+b. 'Incorrect' reports the amount of known incorrect predictions, i.e. alignment with zero within error ΔS^2. Given the incompleteness of the NMR dataset, the 'r' column gives the expected value if the predictions were random. 'Hit sum' reports the sum of all above-error ΔS^2 values aligning with predictions, the relative magnitude indicating the type of NMA whose predictions best align with above-error ΔS^2. Out of all NMA types, NE NMA 298 K has the lowest incorrect, and NE NMA 110 K has the highest total hit sum.

Figure 5.7a+b (next page) NMA-interpreted THz difference spectra: top 10 ΔRMSF residues upon ligand binding. Data are sorted by increasing residue number with secondary structure elements displayed. ΔS^2 above error are displayed in the NMR columns: NH S^2 and CH_3 S^2 are labelled as N and C respectively. Those residues with no ΔS^2 values are shown in red. ΔS^2 values that are zero within error are shown as white, whereas those above error are represented in greyscale according to magnitude (darker = greater change). The other columns display the 10 most affected residues derived from the difference spectrum as black boxes, details in §5.2.2.3.4. The other strips of columns are single structure (SS), NE NMA 110 K (N1) and NE NMA 298 K (N2), with apo (unbound) and holo (bound) data represented as A and H respectively. a) is for IBMP binding and b) is for hexanol binding.

Figure 5.7c+d (next page) NMA-only predictions of ΔRMSF upon ligand binding. The figure follow the same scheme as 5.7a+b Residues whose average NMA-derived RMSF changes upon ligand binding are above error are shown as black boxes in the relevant columns. These run in the order of single structure, NE NMA 110 K and NE NMA 298 K (labelled S, 1 and 2 respectively), firstly averaged over 1000 modes (1k), then averaged over 7000 modes (7k). a) is for IBMP binding and b) is for hexanol binding.


NMA-interpreted
THz difference spectra
predictions
NMA-only predictions
of ligand-induced ΔRMSF
IBMP
Hexanol
IBMP
Hexanol
NMR
SS
NI
N2
1k
7k
a)
b)
c)
d)

5.3.4 Changes in positional fluctuations upon ligand binding using only normal mode analysis

The capacity of and difference between SS and NE NMA in blindly predicting changes in RMSF upon ligand binding, in this system where little conformational change occurs, was considered against the NMR data. The RMSF of a given atom as a function of a given mode or range of modes can be calculated. MUP has a maximum number of modes, i.e. 3N-6, corresponding to ~ 7500. Ligand-induced changes in RMSF were calculated using two different mode ranges: the first 1000 modes (1k), representing more global modes, and the (almost full) 7000 modes (7k). The per-residue variance (calculated over all 100 structures in NE NMA) was used as an error, and propagated when the differences between bound and unbound were calculated. Above-error changes are displayed as black boxes in Figure 5.7c+d. As in §5.3.3, NMA predictions aligning with red or greyscale NMR values indicate either no data or known agreement respectively, whereas alignment with white boxes represent known disagreement, i.e. NMR ΔS^2 is zero within error.

Generally more changes are observed when considering all 7000 (7k) modes as compared to just the first 1000 (1k), wherein almost no changes are observed. This is unsurprising when considering that the higher modes are more localised, and therefore can differ as a result of the smaller structural changes, as in this case, where RMSD > 0.7 Å. Alternatively, the more global first 1000 modes require a more dramatic conformational change to differ. Due to this sensitivity, the results for SS NMA are less trustworthy than for NE NMA, due to the prominence of idiosyncratic modes. This is evidenced by 7k SS NMA's prediction of many residues that are observed by NMR not to change. Considering only the 7k NE NMA results, hexanol binding results in no changes at 110 K and only at two residues at 298 K, despite clear ΔS^2. Conversely, IBMP shows many changes, especially at 110 K. A large number of these predictions align with residues where ΔS^2 was measured as zero within error, i.e. white boxes in the NMR columns. These results demonstrate that NMA alone, SS or NE, is incapable of predicting important changes upon ligand binding.

5.3.5 Differences in normal mode densities between single structure and native ensemble normal mode analyses

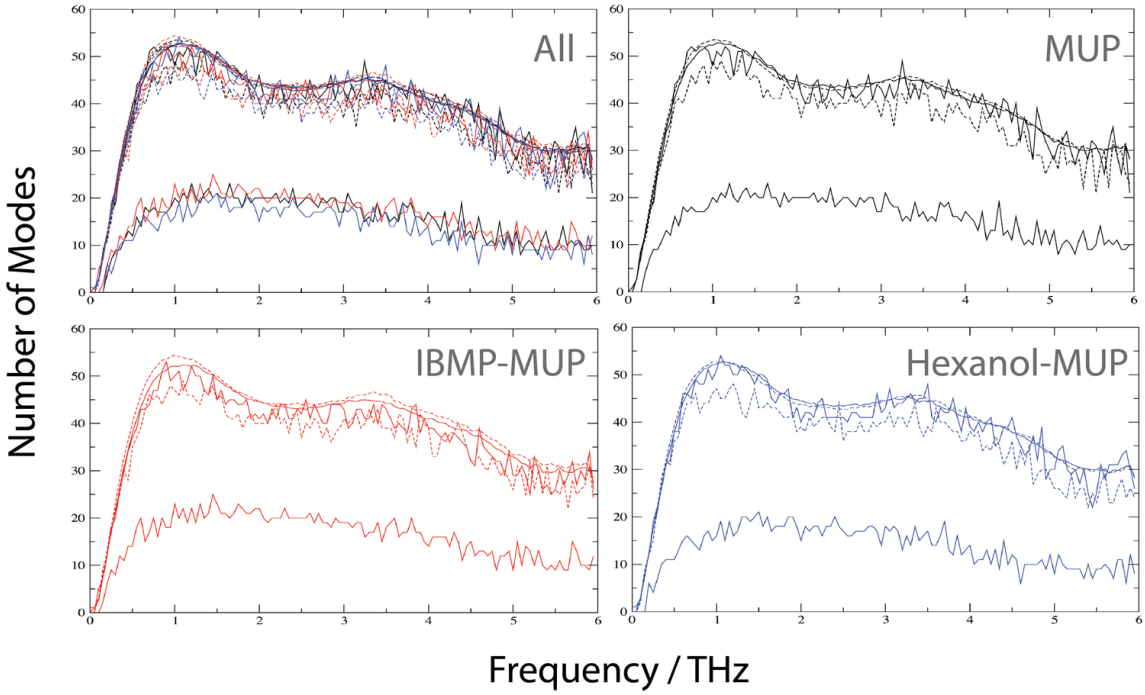


Figure 5.8 Normal mode densities for NE and SS NMA for MUP, MUP-IBMP and MUP-hexanol over the 0 to 6 THz range. Normal mode densities are presented as histograms of bin width = 0.05 THz. The apo and holo normal mode densities all overlay, revealing no distinguishing features. The lower solid lines are from SS NMA. The higher non-smooth lines represent the first frames of the trajectories, dashed is 110 K, solid is 298 K. The higher smooth lines represent the NE NMA, where again 110 K is dashed and 298 K solid. The temperature refers to the temperature of the trajectory from which the NE NMA structures were sampled. Difference between the first frame of the NE NMA and the SS NMA is due to the difference in hydration.

Normal mode densities are shown for SS NMA, NE NMA, and the first frames of the NE NMA trajectories, Figure 5.8. Comparing SS NMA to the first frame reveals a marked difference between asymmetric unit hydration and the 5 Å solvent shell used for each NE NMA structure. Increased hydration in the NE NMA structures than in the asymmetric unit increases the number of low frequency modes and is associated with clearer features at 1 and 3.5 THz. Therefore this initial estimate of an appropriate

symmetrical hydration shell, based on the average amount of asymmetric unit hydration, will need further optimisation. This could result in a large impact on the quality of agreement between NMA-interpreted THz difference spectra and NMR seen in 5.3.3. It is also observed that NE NMA produces much smoother curves with clearer features, though these are generally the same as for the first frame, due to the effect of the averaging over multiple minima. The difference in predictions arising from NE NMA at the two temperatures and for both ligands, §5.3.3, revealed that modes at equivalent frequencies contained non-equivalent collective motions. The equivalence of the mode densities for all structures and temperatures, and their difference to SS NMA, therefore indicates that mode density comparisons are useful for optimising hydration in NMA.

5.4 Summary and Conclusions

Changes in protein positional fluctuations upon ligand binding can constitute an important entropic contribution to binding thermodynamics. Currently, NMR S^2 values are the best experimental method for observation of these changes. However, obtaining S^2 values is a lengthy and sometimes difficult process. Frequencies observed in THz difference spectra reveal which collective protein vibrational modes are most affected by ligand binding. Normal mode analysis (NMA) is an MD-based technique that analyses the collective vibrational modes in atomic detail, providing a tool for interpreting the global THz-derived information as changes in positional fluctuations. A new THz spectroscopy system allows the acquisition of spectra using protein crystals at 110 K. Unlike NMR, this system requires no protein labelling, spectra are quick to obtain, and dynamics of the entire protein are captured. For the first time in protein THz spectroscopy, the protein structure and hydration in the experimental sample is accurately known, and the spectra are acquired below the anharmonic limit. Because NMA makes a harmonic approximation regarding the collective motions, and is dependent on the hydration and structure of the protein analysed, this is the best THz system for interpretation using NMA.

This work constitutes a preliminary investigation of the capacity of NMA-interpreted THz difference spectra to predict changes in positional fluctuations upon ligand binding, using MUP binding to IBMP and hexanol, and NMR S^2 data (acquired at

298 K) as a benchmark. Despite negligible structural differences between the apo and holo crystal structures of MUP bound to IBMP and hexanol, the low (110 K) temperature and the crystal packing constraints, the dynamic changes upon ligand binding are sufficient to generate above-error THz difference spectra that can be analysed by NMA. The difference spectra are also different dependent upon the ligand bound, despite the negligible structural differences between the crystal structures of the two complexes.

Traditional normal mode analysis uses a single structure (SS NMA), whereas the protein native state is defined by an ensemble of structures. Averaging normal modes for an ensemble of native structures enhances those modes representative of the native ensemble, and diminishes modes idiosyncratic to one particular structure. Native ensemble normal mode analysis (NE NMA) was herein performed using structural ensembles extracted from MD trajectories produced at both 110 and 298 K.

NMA was used to predict the ten most likely changes in positional fluctuations upon ligand binding (MUP to IBMP and hexanol), by analysing the apo and holo collective vibrational motions corresponding to the THz difference spectrum frequencies. NE NMA predictions are spread across the protein, whereas SS NMA predictions cluster at the C terminus. This pattern is also observed for the first NE NMA frames, thus demonstrating the benefit of the ensemble averaging of NE NMA, which diminishes the unrepresentative idiosyncratic modes of single structures. NE NMA 298 K was the only type of NMA better than random at avoiding incorrect predictions, i.e. for those residues where ΔS^2 is zero within error. However, the structural ensemble generated at the THz experimental temperature of 110 K (NE NMA 110 K) generated predictions that align best with the largest changes in positional fluctuations described by the NMR S^2 data. This may indicate that better agreement could arise through ensuring that NE NMA sampling better represents the experimental sample, such as through generating ensembles from trajectories performed for the crystalline asymmetric unit or full unit cell, both of which are being considered. The quality of the comparison between NMR S^2 and the NMA-interpreted THz difference spectra is restricted by the incompleteness of the NMR data, suggesting that further assessments may be best performed using systems with more comprehensive NMR S^2 data. Without using THz difference spectra as a guide, NMA alone is incapable of predicting important changes upon ligand binding.

Normal mode densities differed between NE and SS NMA due to hydration. SS NMA used the crystal asymmetric unit hydration, i.e. that of the experimental sample.

This mode density discrepancy therefore reveals that the NE NMA hydration is not equivalent to the experimental sample, and that optimisation of the hydration shell for NE NMA structures extracted from solvated MD trajectories may result in better performance of NMA-interpreted THz.

This work reveals that proteins in a crystal lattice at 110 K exhibit changes in harmonic fluctuations as a result of ligand binding, even with a negligible structural difference between apo and holo crystal structures. These changes are also ligand dependent, despite the similarly negligible difference between the holo crystal structures. Nonetheless, decreasing the THz absorption coefficient error through improved measurements of crystal dimensions may generate better resolved difference spectra and is therefore worthy of attention. This system is limited to studying crystallisable systems which undergo crystallographically tolerable changes, e..g ligand binding. Proteins that are difficult to crystallise may not benefit from the relative ease and speed of the proposed system compared to NMR. However, on chip THz systems showing increased resolution and sensitivity could constitute an avenue for low-temperature THz of non-crystalline proteins.[154]

Additionally, this work reveals the benefit of averaging over the native ensemble in producing representative normal mode analyses. Though this preliminary study resulted in only moderate performance, improvements can be made to the sampling and hydration of the NMA structures, allowing conclusive validation of the potential for this approach as a novel probe of changes in protein dynamics upon ligand binding.

Chapter 6

Summary and Future Work

The work presented in this thesis addressed four issues regarding the dynamics and thermodynamics of protein-ligand binding, with the goal of improving our understanding of biomolecular associations. MUP is used for this work both because it enables a tractable perturbation and therefore thermodynamic decomposition approach, and also due to its wealth of pre-existent interaction thermodynamics data. A range of established techniques were brought to bear, namely ITC, NMR, x-ray crystallography and MD, whilst the preliminary assessment of a new probe of protein dynamic changes upon ligand binding was addressed, NMA-interpreted THz difference spectra.

The fundamental, yet insufficiently thermodynamically described, role of water in biomolecular interactions was the focus of §2.[16] A better entropic estimate of water ejection from the MUP binding site was sought to both aid future decomposition in MUP and to address the suggestion of entropic desolvation in this system, as observed in other proteins with non-polar pockets.[12,15,21,28] Previous estimates suggested conflicting values of 0.1 or -5.8 kJ mol^{-1} desolvation entropy per water molecule in MUP.[18,21] The difference in binding thermodynamics between two interactions that differ by the ejection of a single water molecule and a potential change in sidechain entropy was obtained. Accounting for the sidechain contribution is possible using NMR experiments, however these were aborted due to a defective sample. Therefore the production of new sample and repeat experiments are required to determine the better thermodynamic estimate desired. More than 6.9 kJ mol^{-1} of sidechain entropy would have to be lost upon ligand binding to render desolvation entropic, due to a difference of 3.9 ± 3.0 kJ mol^{-1} being observed between the interactions by ITC.

Lipocalins, such as hydrophobic binder MUP and hydrophilic binder HBP, are used as structural scaffolds for biotechnological applications due to their structural stability and capacity to tolerate functional mutations in loop regions.[67,68] Engineering MUP for a novel capacity, to enable histamine binding in a mode similar to related hydrophilic HBP, was previously attempted. This was performed by introducing either or both the polar sidechain Ser103 and the ionisable sidechain Asp40 into the MUP

binding pocket. Crystal structures of the series of MUP mutants, all obtained to a resolution of below 2.1 Å, revealed an increased apo pocket water content, accompanied by capacity to bind histamine, charged at the crystal pH of 5.5, and a partial incapacity to bind IBMP.

Crystallography is an attractive screening strategy for monitoring mutation-derived changes in MUP binding profile because large quantities of crystals can be produced with a small amount of protein. An attempt was made to validate this approach by performing ITC on the MUP variants binding IBMP and histamine, to assess whether the same trends in binding across the panel of MUP mutants are observed. The ITC solution conditions were changed to partially mimic the crystalline conditions, to attempt the closest possible comparison between the techniques: namely the pH was lowered to 5.5, but the salinity was not increased due to the potential risk of aggregation. Computational methods predicted a rare high pK_a for the D40 residue in MUP, presenting a risk of neutralisation at pH 5.5 that would likely ablate the binding of charged histamine.

Contrary to the crystallographic observations, no change in IBMP or histamine affinity were observed across the series, with histamine affinity being consistently very weak. These data appear to support the computational predictions that D40 is neutralised at pH 5.5. Possibly the salinity needs to be adjusted to replicate the binding trends observed crystallographically, as it affects pK_a values and differs between ITC and crystallography. Alternatively, the trends may be reproducible at higher pH. An attractive avenue for ongoing work is to directly measure the *in situ* D40 pK_a using NMR, by performing a pH titration on a ^{15}N-^{13}C labeled protein sample and monitoring the chemical shift of the D40 sidechain atoms. In summary, this study has demonstrated that validating the use of crystals to screen interactions that are heavily pH dependent is non-trivial. The validity of the crystallographic observations in this case remains unresolved.

The first systematic experimental evaluation, in a protein-ligand system, of the often voiced yet rarely and inconsistently observed entropic and affinity benefits of decreasing rotor numbers in a ligand molecule was performed in §3.[6,7] Removing a single ligand rotor was investigated by comparing MUP binding within and between panels of saturated and unsaturated alcohols, which allowed two methods for decreasing rotor numbers to be assessed: rotor removal through shortening the molecule by one methylene group ('remove'), and rotor rigidification through introduction of a double bond ('restrict'). Restrict modifications were also assessed in

both cis and trans isomers, and with the double bond placed at different positions. Entropically, previously predicted penalties for rotamerically 'freezing' a single ligand C-C bond upon binding, used for important decisions in drug development and 5 to 6 kJ mol^{-1} in magnitude, were vindicated as an average 5.4 kJ mol^{-1} $T\Delta\Delta S°_i$ benefit from methylene removal across multiple ligand series.[5,82-85] The introduction of a terminal double bond gives a similar result of 4.5 kJ mol^{-1}, suggesting that methylene removal is effectively the deletion of a rotor in terms of intrinsic entropy. However, a significantly higher $T\Delta\Delta S°_i$ benefit of 11.5 kJ mol^{-1} is obtained by introducing an internal double bond. This is likely due to an increase in bound ligand conformations with these modifications. All observed gains are considerable in terms of ligand design, where 5.7 kJ mol^{-1} represents an order of magnitude in affinity. However, reproducing the enhanced intrinsic entropic benefits of restriction over removal observed in this system will be unlikely if this effect does arise from an increase in the flexibility of bound ligand, because most other systems will not have similarly spacious and non-specific binding pockets. Differential desolvation and intrinsic enthalpy conspire to generate unfavourable changes in affinity for both of these modification types. Analysis of crystal structures indicates a contribution from differential protein dynamics to intrinsic entropy changes, which would likely be insufficiently resolved by NMR relaxation experiments. Despite being the first systematic investigation of its kind, and more comprehensive than other work in the field, there is still too little data to identify conclusive and reliable trends. Encouragingly, in a peptide-protein interaction, the introduction of an internal double bond into the ligand was reported to result in a ten-fold improvement in affinity.[102]

Loss of ligand rotational and translational entropy, a major contributor to binding thermodynamics, estimated at -40 to -60 kJ mol^{-1}, was addressed in §4.[5] The previous assumption used in thermodynamic decomposition was that IBMP lost almost all rotational and translational degrees of freedom upon binding to MUP.[18,23] This was recently challenged by 1 μs MD simulations that observed significant IBMP movement in the MUP pocket when bound.[26] Consequently, the first attempt to directly observe and assess residual ligand motion was undertaken using a combination of NMR NOE and rdc measurements obtained for a sample of ^{15}N-MUP protein and ^{13}C-^{15}N IBMP. Labelled IBMP was synthesised successfully. Strong butyl peaks were observed for the bound ligand, with dispersion of methyl resonances indicating a restriction of rotation and translation compared to bulk.

NOE measurements for a sample containing ^{13}C-^{15}N MUP and an excess of

unlabelled ligand allowed assignment of pocket residue protons with NOEs from the ligand. This enabled comparison of simulated NOES from the 1 μs trajectory with those observed experimentally. The simulated NOEs were calculated taking spin diffusion into account. The analysis indicated that orientations close to that observed crystallographically resulted in simulated NOEs in close qualitative and quantitative (relative intensity) agreement with experimental NOEs. Clustering of ligand poses from the 1 μs MD simulations (Dr Charlie Laughton, personal communication) reveals that most of the extensively populated clusters represent poses with poor experimental agreement. Thus, the NOE analysis concludes that the trajectories inaccurately sample motion of bound IBMP. Analysis of the first 100 ns of the simulations shows closer sampling to the crystal pose, suggesting that only at long timescales is ligand behaviour inaccurate. This is likely due to previously unrevealed limitations of the forcefield. Reproducing the first 100 ns with a second forcefield, CHARMM, showed less movement, but a comparison at timescales beyond 100 ns is needed.

Residual dipolar coupling measurements for butyl chain bond vectors were acquired by Dr Arnout Kalverda and Phil Morrison. Whilst the rdc values are low, they disagree with predicted rdc values calculated by Dr Gary Thompson using the protein backbone couplings and the crystal structure (PDB 1QY1). This reveals these bond vectors to be sampling non crystallographically-observed orientations. However, this does not necessarily indicate that the ligand samples multiple orientations. These observations remain consistent with the NOE simulations, which show that the butyl chain can sample multiple conformations whilst the ligand remains in effectively the same pose. However, residual ligand motion may not be completely absent. More extensive measurement of ligand rdcs in the complex are being sought, and should provide a better picture of the amount of residual motion present.

Therefore, whilst protein behaviour seems to remain well reproduced[26] by increasingly commonplace long timescale simulations, these experiments provide unique data that questions the validity of ligand behaviour observed at longer MD timescales. This work therefore highlights the necessity to continuously experimentally corroborate evolutions in MD simulations, despite the practical difficulty of obtaining such data. Otherwise, the attraction of MD as a technique can be undermined by its inaccuracy.

Changes in protein positional fluctuations upon ligand binding can constitute an important entropic contribution to binding thermodynamics. A preliminary

investigation of a potential novel protein dynamics probe was the focus of §5. The capacity of normal mode analysis (NMA) interpretation of THz spectroscopy difference spectra to identify residues whose fluctuations change upon ligand binding was investigated. NMR S^2 data, which reveal site-specific changes in dynamics upon ligand binding, were used as a benchmark for the study of IBMP and hexanol binding to MUP. Currently, NMR S^2 values, which are lengthy and sometimes difficult to obtain, are the best experimental method for observation of these changes. Unlike NMR, THz spectroscopy requires no protein labelling, spectra are quick to obtain, and dynamics of the entire protein are captured.

A new THz spectroscopy system, developed in Leeds by Dr Kasia Tych, allows the acquisition of spectra using protein crystals at 110 K. It was firstly demonstrated that ligand-dependent THz difference spectra are obtainable between unbound and bound protein crystals, despite the crystalline environment, the 110 K temperature and the close overlay of the apo and holo structures for these interactions (RMSD > 0.7 Å). Likewise, difference spectra are also ligand-dependent, despite the negligible structural differences between the crystal structures of the two complexes. Successful acquisition of difference spectra allowed changes in harmonic vibrations to be analysed using NMA for similar patterns of differential residue flexibility to those observed by NMR S^2. Harmonic vibrations were analysed using both traditional single-structure (SS) NMA, and native ensemble averaged (NE) NMA. Averaging normal modes for an ensemble of native structures enhances those modes representative of the native ensemble, and diminishes modes idiosyncratic to one particular structure. NE NMA used structures generated by MD trajectories performed at NMR and THz experimental temperatures of 110 K and 298 K.

Those residues most affected by changes in collective motions indicated by the THz difference spectra were compared to NMR ΔS^2, resulting in moderate agreement. NE NMA performs better overall than SS NMA, demonstrating the benefit of the ensemble averaging of NE NMA, which diminishes the unrepresentative idiosyncratic modes of single structures. NE NMA 298 K is the only NMA that performs better than random at avoiding incorrect predictions. In generating predictions that align best with the largest changes in positional fluctuations described by the NMR S^2 data, for both ligands NE NMA 110 K performs best. This improvement in predictions when using a native ensemble generated at the THz experimental temperature is encouraging, and may indicate that better agreement could arise by generating NE NMA sampling more representative of the experimental sample. This may be achieved by generating

ensembles from trajectories performed for the crystalline asymmetric unit or full unit cell, both of which are being considered. Without using THz difference spectra as a guide, NMA alone was demonstrated to be incapable of predicting important changes upon ligand binding.

Optimisation of the hydration level for NE NMA structures derived from solvated trajectories could also improve agreement. Scope for hydration optimisation was revealed by the difference in normal mode densities between SS NMA and a single NE NMA structure, which differ in little other than hydration. The SS NMA is hydrated according to the crystal asymmetric unit and is therefore more representative of the experimental sample. Therefore optimisation of symmetric with respect to asymmetric hydration may be achieved through comparison of mode densities.

Though this preliminary study reveals moderate performance, work continues on improving the NMA models to conclusively validate the potential for this approach as a novel probe of changes in protein dynamics upon ligand binding. Further work needs to be done improving the experimental comparability of the NMA model by varying NMA structural sampling and hydration level. Uncertainty in the THz absorption coefficient error could also be improved through increasing the accuracy of crystal dimension measurements. To aid comparison with NMR S^2 data, systems whose NMR S^2 data are more complete are being considered for further investigation.

Hopefully the work contained in this thesis, which has addressed previously unaddressed or unanswered questions, opens up further avenues of questioning and contributes to a better understanding of the different contributions to the dynamics and thermodynamics of protein-ligand interactions.

Appendix I

Analysis of intermediates in IBMP organic synthesis (§4.2.1)

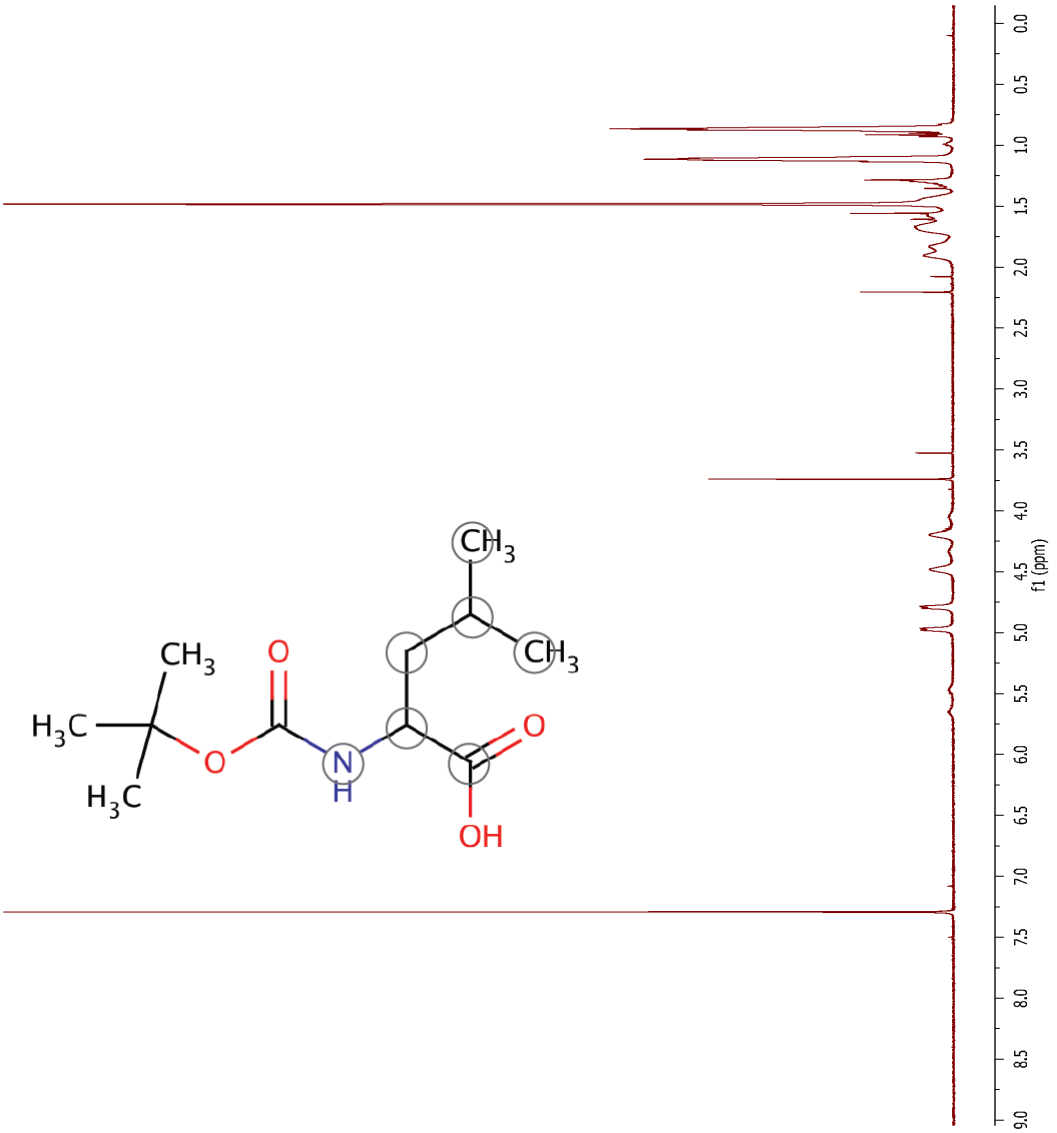


Figure A1.1 ^{1}H NMR of ^{13}C-^{15}N Boc-L-Leucine. ^{13}C and ^{15}N atoms are circled on the molecule. δH (500 MHz, $CDCl_3$); 7.28 (s, 1H, $CHCl_3$), 4.88 (d of m, 1H, J_{CH} 91 Hz, $C_{\alpha}H$), 4.34 (d, 1H, J_{CH} 141 Hz, CH), 1.95-1.5 (br m, 2H, CH_2), 1.48 (s, 9H, BOC), 0.99 (d of m, 6H, J_{CH} 124.88 Hz, CH_3).

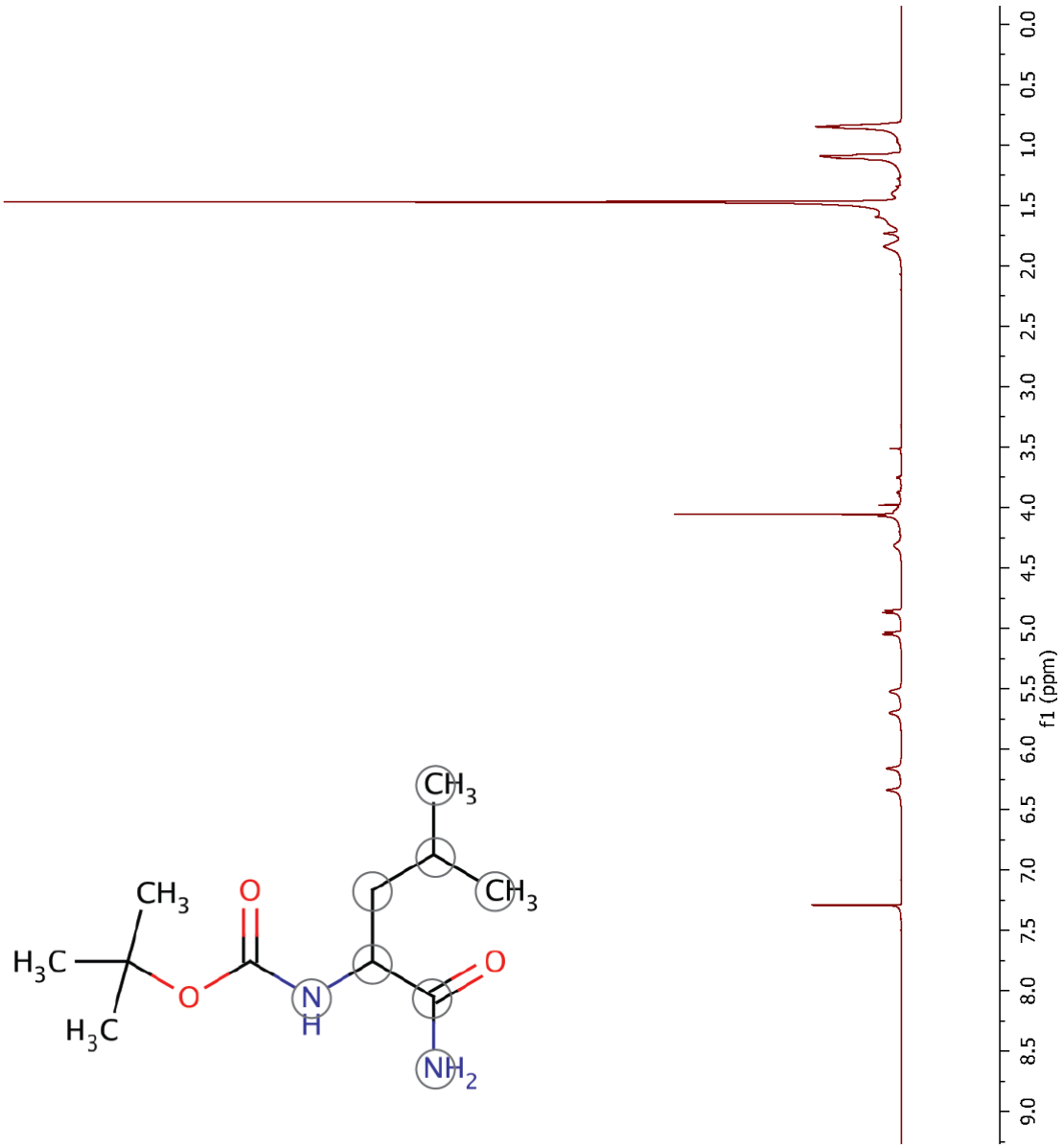


Figure A1.2 ^{1}H NMR of ^{13}C-^{15}N Boc-L-Leucine amide. ^{13}C and ^{15}N atoms are circled on the molecule. δH (500 MHz, $CDCl_3$); 7.3 (s, 1H, $CHCl_3$), 6.25 (d, 1H, J_{NH} 89.8 Hz, CONH), 5.61 (d, 1H, J_{NH} 89.2 Hz, $C_{\alpha}NH$), 4.95 (dd, 1H, J_{HH} 8.02 Hz, J_{CH} 90.2 Hz, $C_{\alpha}H$), 4.19 (d, 1H, J_{CH} 129.6 Hz, CH), 4.06 (s, 9H, trimethoxy triazine), 1.95-1.5 (br m, 2H, CH_2), 1.47 (s, 9H, BOC), 1.28 (s, 2H, grease), 0.97 (d of m, 6H, J_{CH} 124.4 Hz, CH_3).

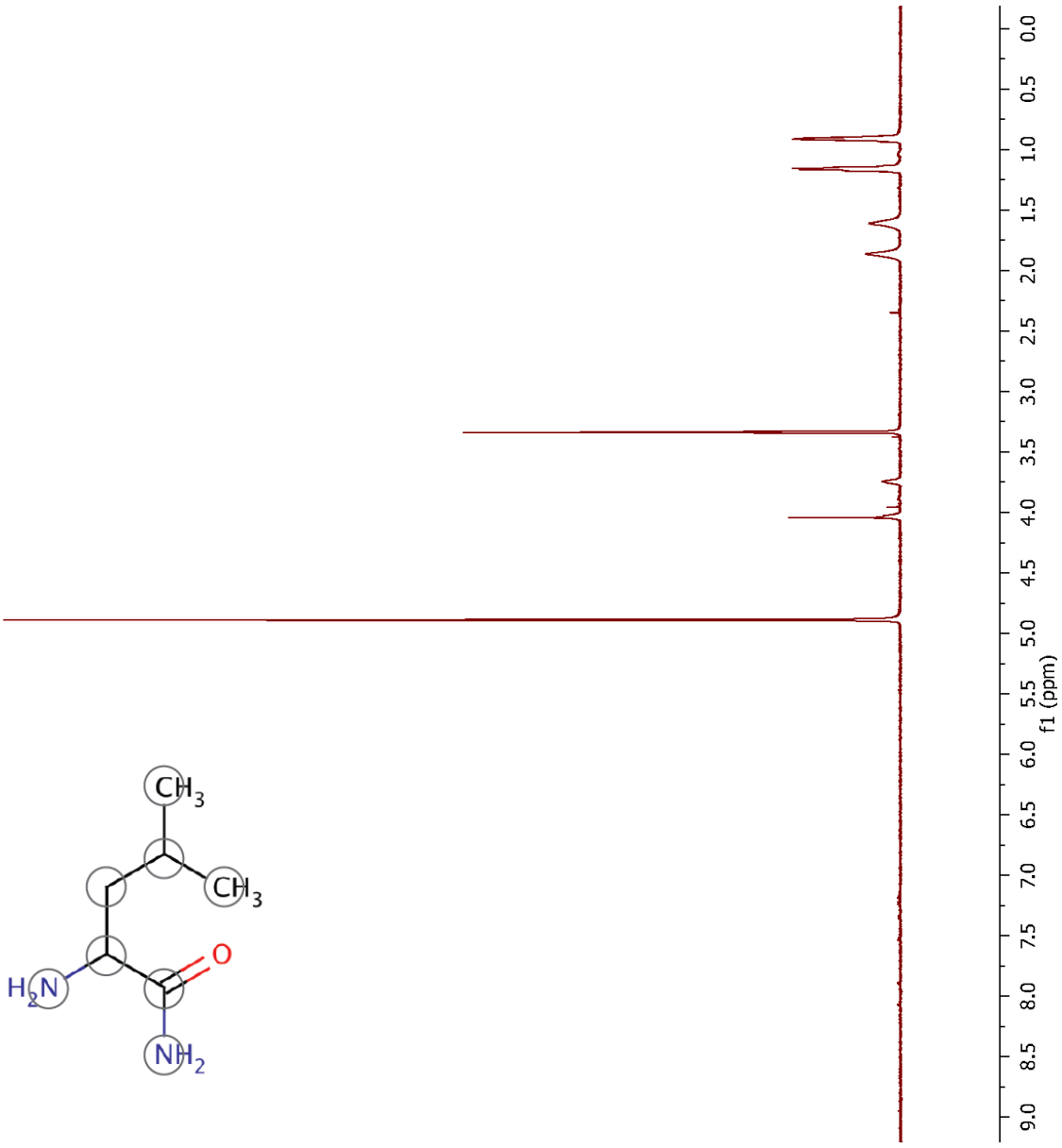


Figure A1.3 ^{1}H NMR of ^{13}C-^{15}N L-Leucine amide. ^{13}C and ^{15}N atoms are circled on the molecule. δH (500 MHz, CD_3OD); 4.88 (s, 1H, CD_3OH), 4.04 (s, 9H, trimethoxy triazine), 3.89 (d of m, 1H, J_{CH} 145 Hz, $C_\alpha H$), 3.34 (s, 3H, CD_2HOD), 1.74 (d of m, 3H, J_{CH} 125.9 Hz, CH and CH_2), 1.04 (d of m, 6H, J_{CH} 125 Hz, CH_3).

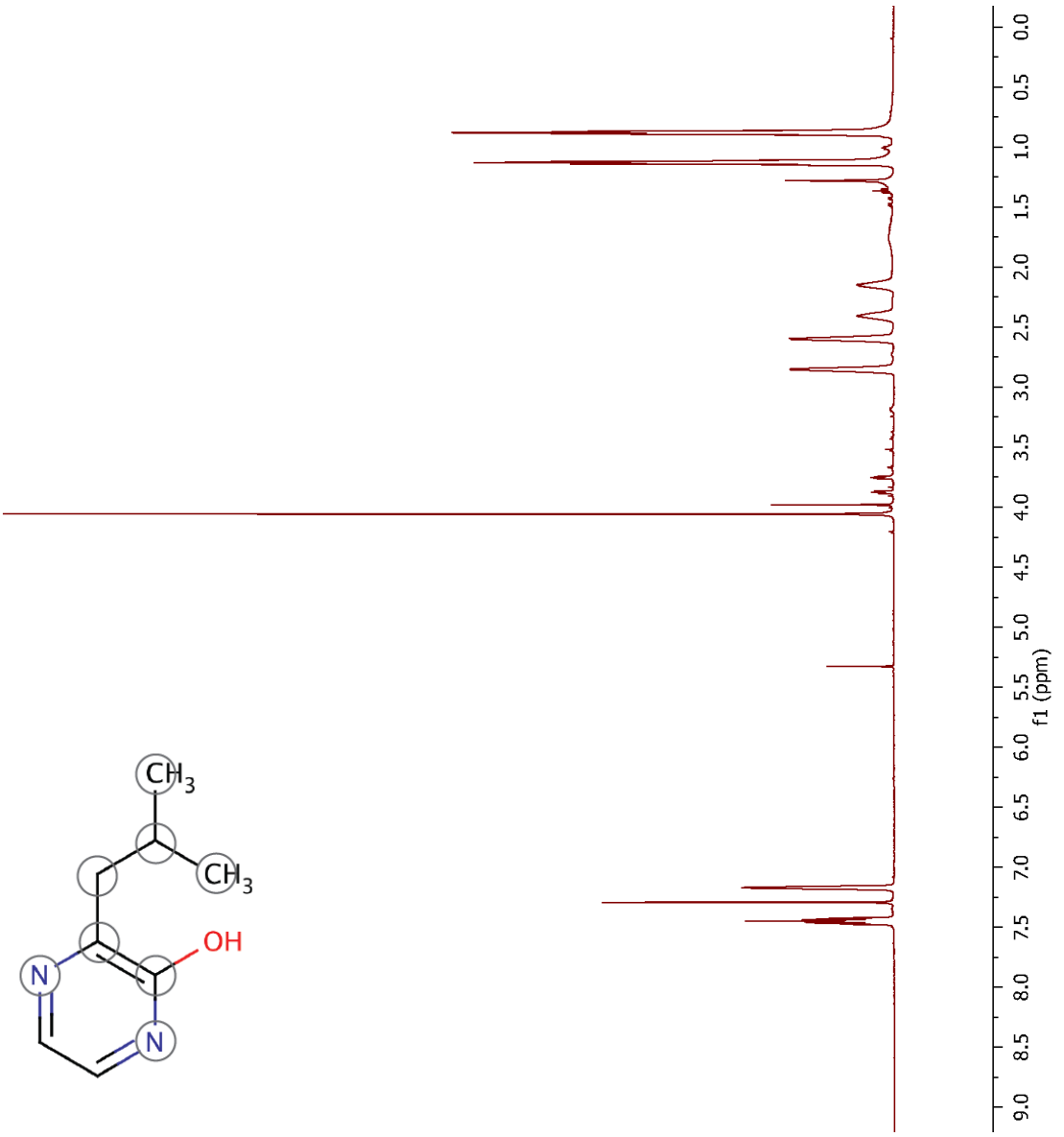


Figure A1.4 ^{1}H NMR of ^{13}C-^{15}N 2-hydroxy-3-isobutyl-pyrazine (IBHP). ^{13}C and ^{15}N atoms are circled on the molecule. δH (500 MHz, $CDCl_3$); 7.45 (m, 1H, ArH), 7.29 (m, 1H, $CHCl_3$), 7.17 (s, 1H, ArH), 5.33 (s, 2H, DCM), 4.06 (s, 9H, trimethoxy triazine), 2.73 (d of m, 2H, J_{CH} 127.50 Hz, CH_2), 2.28 (d of m, 1H, J_{CH} 128.68 Hz, CH), 1.01 (d of m, 6H, J_{CH} 124.75 Hz, CH_3).

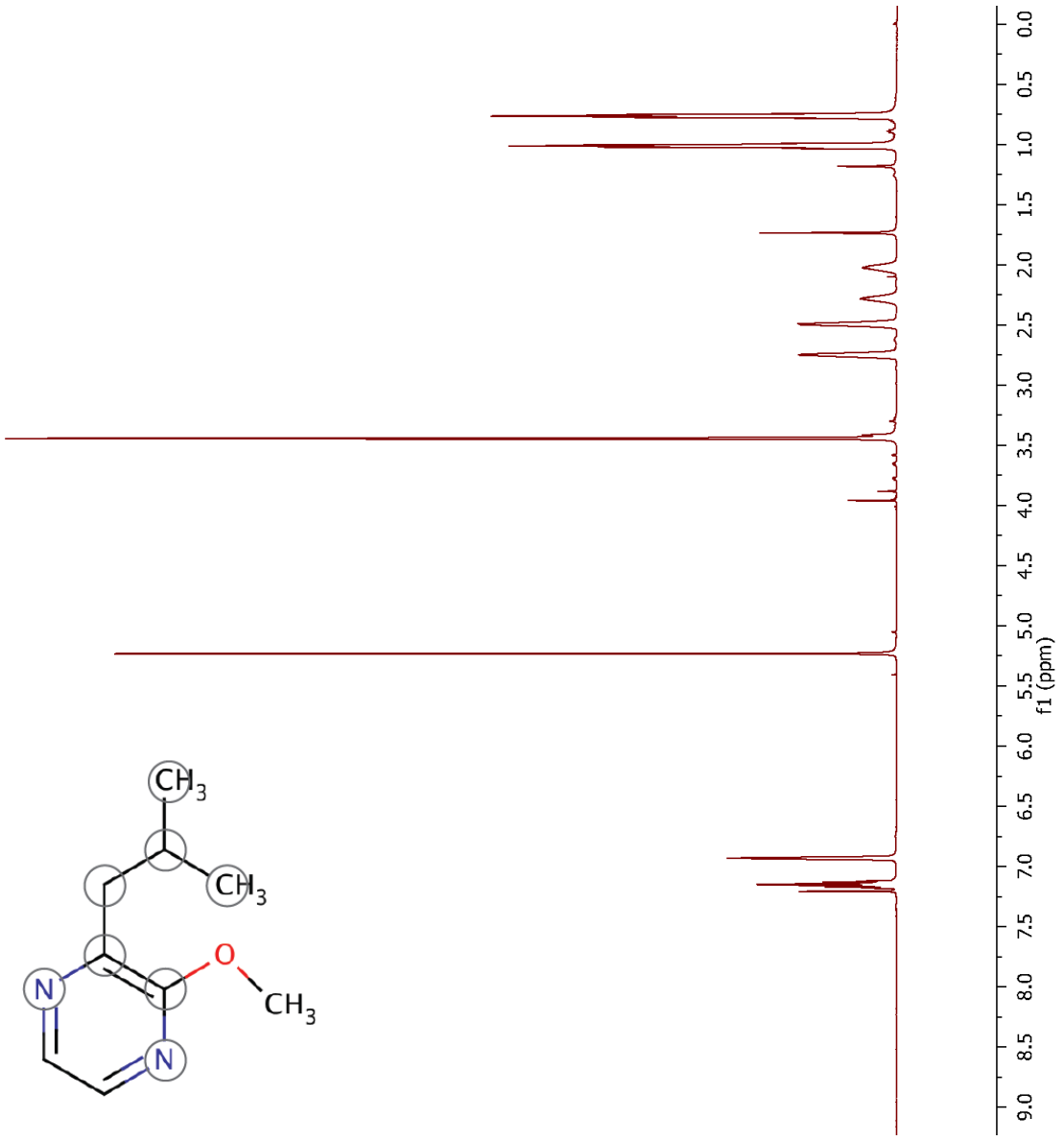


Figure A1.5 ^{1}H NMR of ^{13}C-^{15}N 2-methoxy-3-isobutyl-pyrazine (IBMP). ^{13}C and ^{15}N atoms are circled on the molecule. δH (500 MHz, $CDCl_3$); 7.21 (s, 1H, $CHCl_3$), 7.15 (m, 1H, ArH), 6.93 (m, 1H, ArH), 5.23 (s, 2H, DCM), 3.96 (s, 9H, trimethoxy triazine), 3.44 (s, 3H, OCH_3), 2.62 (d of m, 2H, J_{CH} 127.6 Hz, CH_2), 2.15 (d of m, 1H, J_{CH} 128.7 Hz, CH), 1.74 (s, 8H, THF), 0.89 (d of m, 6H, J_{CH} 124.7 Hz, CH_3).

ESI-HRMS (Electrospray ionisation high resolution mass spectrometry) of ^{13}C-^{15}N 2-methoxy-3-isobutyl-pyrazine (IBMP)

$C_3H_{14}NaO^{15}N_2{}^{13}C_6$, calculated 197.1141, observed 197.1140.

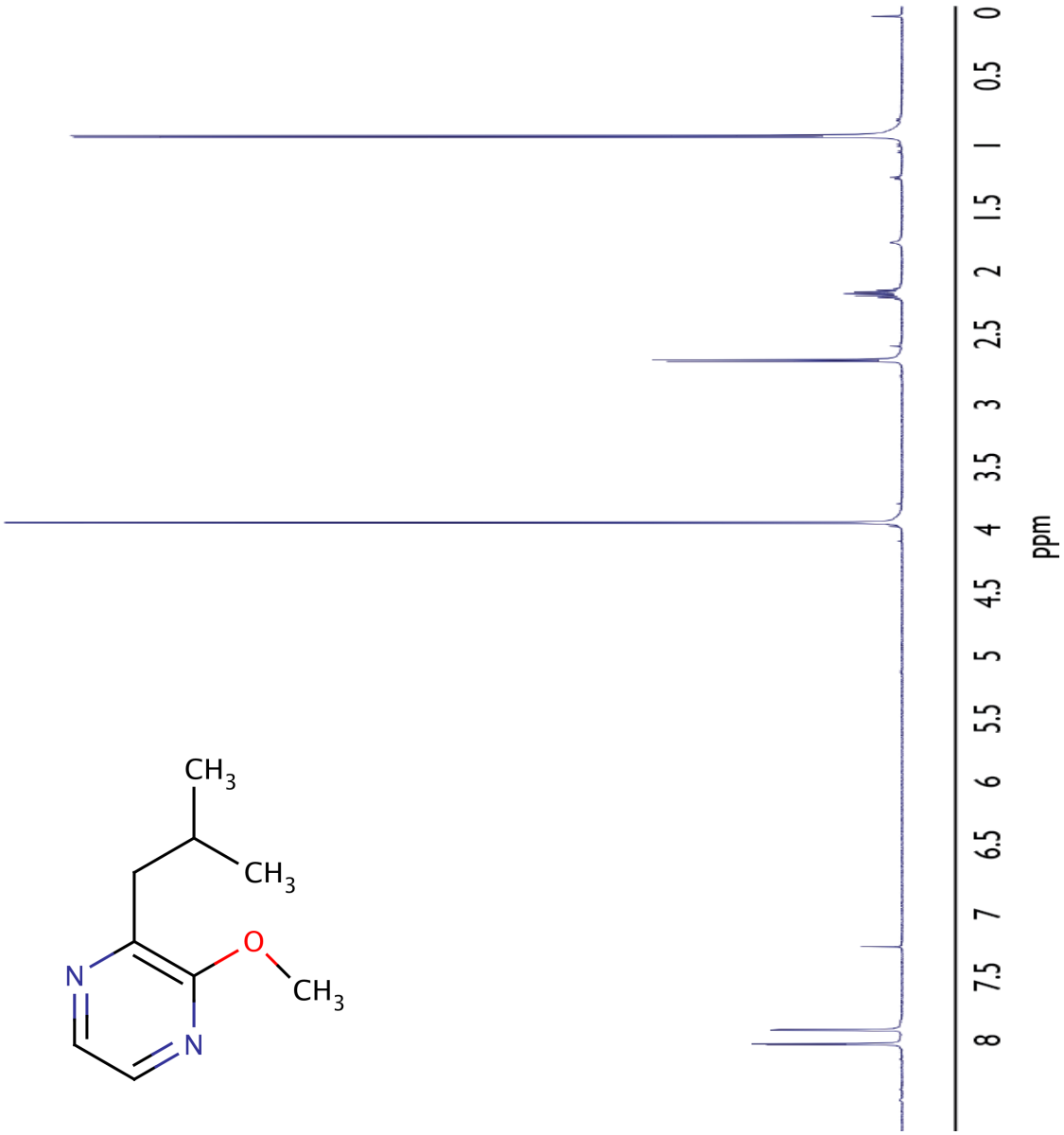


Figure A1.6 ^{1}H NMR of commercially available 2-methoxy-3-isobutyl-pyrazine (IBMP, Sigma-Aldrich). δH (500 MHz, $CDCl_3$); 8.04 (m, 1H, ArH), 7.92 (m, 1H, ArH), 7.27 (s, 1H, $CHCl_3$), 3.96 (s, 3H, OCH_3), 2.69 (d, 2H, J_{HH} 7.2 Hz, CH_2), 2.17 (m, 1H, CH), 0.94 (d, 6H, JHH 6.7 Hz, CH_3).

References

1 Malham, R., Johnstone, S., Bingham, R. J., Barratt, E., Phillips, S. E., Laughton, C. A. & Homans, S. W. Strong solute-solute dispersive interactions in a protein-ligand complex. *J. Am. Chem. Soc.* **127**, 17061-17067, (2005).

2 Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M. & Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545-1614, (2009).

3 Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham Iii, T. E., DeBolt, S., Ferguson, D., Seibel, G. & Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **91**, 1-41, (1995).

4 Sousa, S. F., Fernandes, P. A. & Ramos, M. J. Protein-ligand docking: current status and future challenges. *Proteins* **65**, 15-26, (2006).

5 Page, M. I. & Jencks, W. P. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc. Natl. Acad. Sci. U. S. A.* **68**, 1678-1683, (1971).

6 Martin, S. F. Preorganization in biological systems: Are conformational constraints worth the energy? *Pure Appl. Chem.* **79**, 193-200, (2007).

7 Udugamasooriya, D. G. & Spaller, M. R. Conformational constraint in protein ligand design and the inconsistency of binding entropy. *Biopolymers* **89**, 653-667, (2008).

8 Lindorff-Larsen, K., Best, R. B., Depristo, M. A., Dobson, C. M. & Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128-132, (2005).

9 Tanford, C. The hydrophobic effect and the organization of living matter. *Science* **200**, 1012-1018, (1978).

10 Ashworth, R. A., Howe, G. B., Mullins, M. E. & Rogers, T. N. Air-water partitioning coefficients of organics in dilute aqueous solutions. *J. Hazard. Mater.* **18**, 25-36, (1988).

11 Plyasunov, A. V. & Shock, E. L. Thermodynamic functions of hydration of hydrocarbons at 298.15 K and 0.1 MPa. *Geochim. Cosmochim. Acta* **64**, 439-468, (2000).

12 Denisov, V. P., Venu, K., Peters, J., Horlein, H. D. & Halle, B. Orientational Disorder and Entropy of Water in Protein Cavities. *J. Phys. Chem. B* **101**, 9380-9389, (1997).

13 Ernst, J. A., Clubb, R. T., Zhou, H. X., Gronenborn, A. M. & Clore, G. M. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science* **267**, 1813-1817, (1995).

14 Olano, L. R. & Rick, S. W. Hydration free energies and entropies for water in protein interiors. *J. Am. Chem. Soc.* **126**, 7991-8000, (2004).

15 Setny, P., Baron, R. & McCammon, J. A. How Can Hydrophobic Association Be Enthalpy Driven? *J. Chem. Theory Comput.* **6**, 2866-2871, (2010).

16 Li, Z. & Lazaridis, T. Water at biomolecular binding interfaces. *Physical chemistry chemical physics : PCCP* **9**, 573-581, (2007).

17 Baron, R., Setny, P. & McCammon, J. A. Water in cavity-ligand recognition. *J. Am. Chem. Soc.* **132**, 12091-12097, (2010).

18 Shimokhina, N., Bronowska, A. & Homans, P. S. W. Contribution to Ligand Desolvation to Binding Thermodynamics in a Ligand-Protein Interaction. *Angew. Chem. Int. Ed.* **45**, 6374-6376, (2006).

19 Zidek, L., Stone, M. J., Lato, S. M., Pagel, M. D., Miao, Z., Ellington, A. D. & Novotny, M. V. NMR mapping of the recombinant mouse major urinary protein I binding site occupied by the pheromone 2-sec-butyl-4,5-dihydrothiazole. *Biochem. (Mosc.)* **38**, 9850-9861, (1999).

20 Pertinhez, T. A., Ferrari, E., Casali, E., Patel, J. A., Spisni, A. & Smith, L. J. The binding cavity of mouse major urinary protein is optimised for a variety of ligand binding modes. *Biochem. Biophys. Res. Commun.* **390**, 1266-1271, (2009).

21 Barratt, E., Bingham, R. J., Warner, D. J., Laughton, C. A., Phillips, S. E. & Homans, S. W. Van der Waals interactions dominate ligand-protein association in a protein binding site occluded from solvent water. *J. Am. Chem. Soc.* **127**, 11827-11834, (2005).

22 Barratt, E., Bronowska, A., Vondrasek, J., Cerny, J., Bingham, R., Phillips, S. & Homans, S. W. Thermodynamic penalty arising from burial of a ligand polar group within a hydrophobic pocket of a protein receptor. *J. Mol. Biol.* **362**, 994-1003, (2006).

23 Bingham, R. J., Findlay, J. B. C., Hsieh, S. Y., Kalverda, A. P., Kjeliberg, A., Perazzolo, C., Phillips, S. E. V., Seshadri, K., Trinh, C. H., Turnbull, W. B., Bodenhausen, G. & Homans, S. W. Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein. *J. Am. Chem. Soc.* **126**, 1675-1681, (2004).

24 Perazzolo, C., Verde, M., Homans, S. W. & Bodenhausen, G. Evidence of chemical exchange in recombinant Major Urinary Protein and quenching thereof upon pheromone binding. *J. Biomol. NMR* **38**, 3-9, (2007).

25 Bocskei, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E., Cavaggioni, A., Findlay, J. B. & North, A. C. Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature* **360**, 186-188, (1992).

26 Roy, J. & Laughton, C. A. Long-timescale molecular-dynamics simulations of the major urinary protein provide atomistic interpretations of the unusual thermodynamics of ligand binding. *Biophys. J.* **99**, 218-226, (2010).

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

28 Syme, N. R. *A comparison of the thermodynamics of hydrophobic and hydrophilic ligand-protein interactions* PhD thesis, University of Leeds, (2009).

29 Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131-137, (1989).

30 Turnbull, W. B. & Daranas, A. H. On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc.* **125**, 14859-14866, (2003).

31 Zhang, Y. L. & Zhang, Z. Y. Low-affinity binding determined by titration calorimetry using a high-affinity coupling ligand: a thermodynamic study of ligand binding to protein tyrosine phosphatase 1B. *Anal. Biochem.* **261**, 139-148, (1998).

32 Andujar-Sanchez, M., Jara-Perez, V. & Camara-Artigas, A. Thermodynamic determination of the binding constants of angiotensin-converting enzyme inhibitors by a displacement method. *FEBS Lett.* **581**, 3449-3454, (2007).

33 Sigurskjold, B. W. Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Anal. Biochem.* **277**, 260-266, (2000).

34 Ishima, R. & Torchia, D. A. Protein dynamics from NMR. *Nat. Struct. Biol.* **7**, 740-743, (2000).

35 Prestegard, J. H., Bougault, C. M. & Kishore, A. I. Residual dipolar couplings in structure determination of biomolecules. *Chem. Rev.* **104**, 3519-3540, (2004).

36 Fenwick, R. B., Esteban-Martin, S., Richter, B., Lee, D., Walter, K. F., Milovanovic, D., Becker, S., Lakomek, N. A., Griesinger, C. & Salvatella, X. Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. *J. Am. Chem. Soc.* **133**, 10336-10339, (2011).

37 Karplus, M. Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* **85**, 2870-2871, (1963).

38 Chou, J. J., Case, D. A. & Bax, A. Insights into the mobility of methyl-bearing side chains in proteins from (3)J(CC) and (3)J(CN) couplings. *J. Am. Chem. Soc.* **125**, 8959-8966, (2003).

39 Perez, C., Lohr, F., Ruterjans, H. & Schmidt, J. M. Self-consistent Karplus parametrization of 3J couplings depending on the polypeptide side-chain torsion chi1. *J. Am. Chem. Soc.* **123**, 7081-7093, (2001).

40 Saupe, A. & Englert, G. High-Resolution Nuclear Magnetic Resonance Spectra of Orientated Molecules. *Phys. Rev. Lett.* **11**, 462-464, (1963).

41 Saupe, A. Recent Results in the Field of Liquid Crystals. *Angew. Chem. Int. Ed.* **7**, 97-112, (1968).

42 Blackledge, M., Hus, J.-C. & Dosset, P. *Module 1.0: A Novel Interactive Tool for Rigid-Body Modeling of Multi-Domain Macromolecules using Residual Dipolar Couplings*, <http://protchem.lic.gorlaeus.net/ubbink/Module/theory.html> (2001).

43 Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* **104**, 4559-4570, (1982).

44 Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546-4559, (1982).

45 Yang, D. & Kay, L. E. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. Mol. Biol.* **263**, 369-382, (1996).

46 Homans, S. W. Probing the binding entropy of ligand-protein interactions by NMR. *ChemBioChem* **6**, 1585-1591, (2005).

47 Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I. & Mackerell, A. D., Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671-690, (2010).

48 Seeber, M., Cecchini, M., Rao, F., Settanni, G. & Caflisch, A. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* **23**, 2625-2627, (2007).

49 Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O. & Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **19**, 120-127, (2009).

50 Buck, M., Bouguet-Bonnet, S., Pastor, R. W. & MacKerell, A. D., Jr. Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. *Biophys. J.* **90**, L36-38, (2006).

51 Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712-725, (2006).

52 Mackerell, A. D., Jr., Feig, M. & Brooks, C. L., 3rd. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400-1415, (2004).

53 Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **100**, L47-L49, (2011).

54 Ferrara, P., Apostolakis, J. & Caflisch, A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins* **46**, 24-33, (2002).

55 Tozzini, V. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* **15**, 144-150, (2005).

56 Mobley, D. L. & Dill, K. A. Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". *Structure* **17**, 489-498, (2009).

57 Dunitz, J. D. The entropic cost of bound water in crystals and biomolecules. *Science* **264**, 670, (1994).

58 Shimokhina, N. *Dissecting contributions to binding thermodynamics in ligand-protein interactions* PhD thesis, University of Leeds, (2007).

59 ALTABiosciences. *Amino Acid Analysis*, <http://www.altabioscience.bham.ac.uk/services/amacid/index.shtml> (

60 Gill, S. C. & von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319-326, (1989).

61 Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411-2423, (1995).

62 Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131-137, (1989).

63 *iNMR*, <http://www.inmr.net> (

64 Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293, (1995).

65 Doig, A. J. Thermodynamics of amino acid side-chain internal rotations. *Biophys. Chem.* **61**, 131-141, (1996).

66 Paesen, G. C., Adams, P. L., Harlos, K., Nuttall, P. A. & Stuart, D. I. Tick histamine-binding proteins: isolation, cloning, and three-dimensional structure. *Mol. Cell* **3**, 661-671, (1999).

67 Skerra, A. Anticalins as alternative binding proteins for therapeutic use. *Curr. Opin. Mol. Ther.* **9**, 336-344, (2007).

68 Skerra, A. Alternative binding proteins: anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. *FEBS J.* **275**, 2677-2683, (2008).

69 Schlehuber, S. & Skerra, A. Lipocalins in drug discovery: from natural ligand-binding proteins to "anticalins". *Drug Discov. Today* **10**, 23-33, (2005).

70 Croke, R. L., Patil, S. M., Quevreaux, J., Kendall, D. A. & Alexandrescu, A. T. NMR determination of pKa values in alpha-synuclein. *Protein Sci.* **20**, 256-269, (2011).

71 Grimsley, G. R., Scholtz, J. M. & Pace, C. N. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci.* **18**, 247-251, (2009).

72 Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S. & Onufriev, A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33**, W368-371, (2005).

73 Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **61**, 704-721, (2005).

74 Onufriev, e. a. *H++ accuracy table*, <http://biophysics.cs.vt.edu/H++_accuracy_table.pdf> (

75 <http://omlc.ogi.edu/spectra/PhotochemCAD/abs_html/tryptophan.html> (

76 BMRB. *L-Tryptophan*, <http://www.bmrb.wisc.edu/metabolomics/mol_summary/show_data.php?molName=L_tryptophan&id=bmse000050&whichTab=1> (

77 BMRB. *Histamine*, <http://www.bmrb.wisc.edu/metabolomics/mol_summary/show_data.php?molName=histamine&id=bmse000744&whichTab=1> (

78 Williamson, M. P. Secondary-structure dependent chemical shifts in proteins. *Biopolymers* **29**, 1423-1431, (1990).

79 Moreau, V. H., Valente, A. P. & Almeida, F. b. C. L. Prediction of the amount of secondary structure of proteins using unassigned NMR spectra: a tool for target selection in structural proteomics. *Genet. Mol. Biol.* **29**, 762-770, (2006).

80 Wishart, D. S. & Nip, A. M. Protein chemical shift analysis: a practical guide. *Biochem. Cell Biol.* **76**, 153-163, (1998).

81 Putnam, C. *Scripps Institute Protein Calculator*, <http://www.scripps.edu/~cdputnam/protcalc.html> (

82 Gao, C., Park, M. S. & Stern, H. A. Accounting for ligand conformational restriction in calculations of protein-ligand binding affinities. *Biophys. J.* **98**, 901-910, (2010).

83 Gillet, V., Johnson, A. P., Mata, P., Sike, S. & Williams, P. SPROUT: a program for structure generation. *J. Comput. Aided Mol. Des.* **7**, 127-153, (1993).

84 Gilson, M. K. & Zhou, H. X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 21-42, (2007).

85 Raha, K. & Merz, K. M. Large-scale validation of a quantum mechanics based scoring function: Predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. *J. Med. Chem.* **48**, 4558-4575, (2005).

86 Benfield, A. P., Teresk, M. G., Plake, H. R., DeLorbe, J. E., Millspaugh, L. E. & Martin, S. F. Ligand preorganization may be accompanied by entropic penalties in protein-ligand interactions. *Angew. Chem. Int. Ed.* **45**, 6830-6835, (2006).

87 Yang, D., Mok, Y. K., Forman-Kay, J. D., Farrow, N. A. & Kay, L. E. Contributions to protein entropy and heat capacity from bond vector motions measured by NMR spin relaxation. *J. Mol. Biol.* **272**, 790-804, (1997).

88 Krishnamurthy, V. M., Bohall, B. R., Semetey, V. & Whitesides, G. M. The paradoxical thermodynamic basis for the interaction of ethylene glycol, glycine, and sarcosine chains with bovine carbonic anhydrase II: an unexpected manifestation of enthalpy/entropy compensation. *J. Am. Chem. Soc.* **128**, 5802-5812, (2006).

89 Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **62**, 48-57, (2006).

90 Evans, P. Scaling and assessment of data quality. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **62**, 72-82, (2006).

91 The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **50**, 760-763, (1994).

92 Brunger, A. T. Free R value: cross-validation in crystallography. *Methods Enzymol.* **277**, 366-396, (1997).

93 Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **53**, 240-255, (1997).

94 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **60**, 2126-2132, (2004).

95 Schuttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **60**, 1355-1363, (2004).

96 Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **66**, 12-21, (2010).

97 Ringe, D. & Petsko, G. A. Study of protein dynamics by X-ray diffraction. *Methods Enzymol.* **131**, 389-433, (1986).

98 Tunon, I., Silla, E. & Pascual-Ahuir, J. L. Molecular surface area and hydrophobic effect. *Protein Eng.* **5**, 715-716, (1992).

99 Chang, C. E. A., Chen, W. & Gilson, M. K. Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1534-1539, (2007).

100 Rueda, M., Ferrer-Costa, C., Meyer, T., Perez, A., Camps, J., Hospital, A., Gelpi, J. L. & Orozco, M. A consensus view of protein dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 796-801, (2007).

101 Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612, (2004).

102 Harrison, B. A., Gierasch, T. M., Neilan, C., Pasternak, G. W. & Verdine, G. L. High-affinity mu opioid receptor ligands discovered by the screening of an exhaustively stereodiversified library of 1,5-enediols. *J. Am. Chem. Soc.* **124**, 13352-13353, (2002).

103 Irudayam, S. J. & Henchman, R. H. Entropic cost of protein-ligand binding and its dependence on the entropy in solution. *J. Phys. Chem.. B* **113**, 5871-5884, (2009).

104 Murphy, K. P., Xie, D., Thompson, K. S., Amzel, L. M. & Freire, E. Entropy in biological binding processes: estimation of translational entropy loss. *Proteins* **18**, 63-67, (1994).

105 Homans, S. W. Dynamics and Thermodynamics of Ligand–Protein Interactions. *Top. Curr. Chem.* **272**, 51-82, (2007).

106 Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* **8**, 243-256, (1994).

107 Turnbull, W. B., Precious, B. L. & Homans, S. W. Dissecting the cholera toxin-ganglioside GM1 interaction by isothermal titration calorimetry. *J. Am. Chem. Soc.* **126**, 1047-1054, (2004).

108 Lundquist, J. J. & Toone, E. J. The cluster glycoside effect. *Chem. Rev.* **102**, 555-578, (2002).

109 Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **215**, 617-621, (1993).

110 Lassen, K. M., Lee, J. & Joullie, M. M. Synthetic Studies of Tamandarin B Side Chain Analogues. *J. Org. Chem.* **75**, 3027-3036, (2010).

111 Mizuhara, T., Hioki, K., Yamada, M., Sasaki, H., Morisaki, D. & Kunishima, M. Direct Preparation of Primary Amides by Reaction of Carboxylic Acids and Ammonia in Alcohols Using DMT-MM. *Chem. Lett.* **37**, 1190-1191, (2008).

112 Leite, A. C. L., Vieira, R. F., de Faria, A. R., Wanderley, A. G., Afiatpour, P., Ximenes, E. C. P. A., Srivastava, R. M., de Oliveira, C. F., Medeiros, M. V., Antunes, E. & Brondani, D. J. Synthesis, anti-inflammatory and antimicrobial activities of new 1,2,4-oxadiazoles peptidomimetics. *Farmaco* **55**, 719-724, (2000).

113 Gerritsma, D. A., Brindle, I. D., Jones, T. R. B. & Capretta, A. Preparation of labelled 2-methoxy-3-alkylpyrazines: synthesis and characterization of deuterated 2-methoxy-3-isopropylyrazine and 2-methoxy-3-isobutylpyrazine. *J. Labelled Comp. Rad.* **46**, 243-253, (2003).

114 Hansen, M. R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.* **5**, 1065-1074, (1998).

115 Fridman, N., Kapon, M., Sheynin, Y. & Kaftory, M. Different packing in three polymorphs of 2,4,6-trimethoxy-1,3,5-triazine. *Acta Crystallogr., Sect. B: Struct. Sci* **60**, 97-102, (2004).

116 Hwang, T. L. & Shaka, A. J. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J. Magn. Reson., Ser A* **112**, 275-279, (1995).

117 Johnson, B. A. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313-352, (2004).

118 Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859-1865, (2008).

119 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33-38, 27-38, (1996).

120 Keepers, J. W. & James, T. L. A theoretical study of distance determinations from NMR. Two-dimensional nuclear overhauser effect spectra. *J. Magn. Reson. (1969)* **57**, 404-426, (1984).

121 Zweckstetter, M. & Bax, A. Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase:,Äâ Aid to Protein Structure Determination by NMR. *J. Am. Chem. Soc.* **122**, 3791-3792, (2000).

122 van Vlijmen, H. W. T. & Karplus, M. Analysis of Calculated Normal Modes of a Set of Native and Partially Unfolded Proteins. *J. Phys. Chem. B* **103**, 3009-3021, (1999).

123 Treptow, W., Marrink, S. J. & Tarek, M. Gating motions in voltage-gated potassium channels revealed by coarse-grained molecular dynamics simulations. *J. Phys. Chem. B* **112**, 3277-3282, (2008).

124 Tzeng, S. R. & Kalodimos, C. G. Dynamic activation of an allosteric regulatory protein. *Nature* **462**, 368-372, (2009).

125 Zaccai, G. Neutron scattering perspectives for protein dynamics. *J. Non-Cryst. Solids* **357**, 615-621, (2011).

126 Cho, H. S., Dashdorj, N., Schotte, F., Graber, T., Henning, R. & Anfinrud, P. Protein structural dynamics in solution unveiled via 100-ps time-resolved x-ray scattering. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7281-7286, (2010).

127 Smith, D. P., Giles, K., Bateman, R. H., Radford, S. E. & Ashcroft, A. E. Monitoring copopulated conformational states during protein folding events using electrospray ionization-ion mobility spectrometry-mass spectrometry. *J. Am. Soc. Mass Spectrom.* **18**, 2180-2190, (2007).

128 Michalet, X., Weiss, S. & Jager, M. Single-molecule fluorescence studies of protein folding and conformational dynamics. *Chem. Rev.* **106**, 1785-1813, (2006).

129 Stone, M. J. NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding. *Acc. Chem. Res.* **34**, 379-388, (2001).

130 Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M. R. & Blackledge, M. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J. Am. Chem. Soc.* **131**, 17908-17918, (2009).

131 Fenwick, R. B., Esteban-Martin, S. & Salvatella, X. Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur. Biophys. J.* **40**, 1339-1355, (2011).

132 Markelz, A., Whitmire, S., Hillebrecht, J. & Birge, R. THz time domain spectroscopy of biomolecular conformational modes. *Phys. Med. Biol.* **47**, 3797-3805, (2002).

133 Knab, J. R., Chen, J. Y., Mader, M. & Markelz, A. G. in *American Physical Society, March Meeting 2004* (Montreal, Quebec, Canada, 2004).

134 Chen, J.-Y., Knab, J. R., Ye, S., He, Y. & Markelz, A. G. Terahertz dielectric assay of solution phase protein binding. *Appl. Phys. Lett.* **90**, 243901-243901-243903, (2007).

135 He, Y., Chen, J. Y., Knab, J. R., Zheng, W. & Markelz, A. G. Evidence of protein collective motions on the picosecond timescale. *Biophys. J.* **100**, 1058-1065, (2011).

136 Xu, J., Plaxco, K. W. & Allen, S. J. Collective dynamics of lysozyme in water: terahertz absorption spectroscopy and comparison with theory. *J. Phys. Chem. B* **110**, 24255-24259, (2006).

137 Knab, J. R., Jing-Yin, C., Yunfen, H. & Markelz, A. G. Terahertz Measurements of Protein Relaxational Dynamics. *P. IEEE* **95**, 1605-1610, (2007).

138 Markelz, A. G. Terahertz Dielectric Sensitivity to Biomolecular Structure and Function. *IEEE J. Sel. Topics Quantum Electron.* **14**, 180-190, (2008).

139 Kindt, J. T. & Schmuttenmaer, C. A. Far-Infrared Dielectric Properties of Polar Liquids Probed by Femtosecond Terahertz Pulse Spectroscopy. *J. Phys. Chem.* **100**, 10373-10379, (1996).

140 Leitner, D. M., Gruebele, M. & Havenith, M. Solvation dynamics of biomolecules: modeling and terahertz experiments. *HFSP journal* **2**, 314-323, (2008).

141 Tych, K. M. *Terahertz Time-Domain Spectroscopy of Biological Macromolecules* PhD thesis, University of Leeds, (2011).

142 Tych, K. M., Burnett, A. D., Wood, C. D., Cunningham, J. E., Pearson, A. R., Davies, A. G. & Linfield, E. H. Applying broadband terahertz time-domain spectroscopy to the analysis of crystalline proteins: a dehydration study. *J. Appl. Crystallogr.* **44**, 129-133, (2011).

143 Darling, C. L. & Schlegel, H. B. Dipole Moments, Polarizabilities, and Infrared Intensities Calculated with Electric Field Dependent Functions. *J. Phys. Chem.* **98**, 5855-5861, (1994).

144 Powers, R., Clore, G. M., Garrett, D. S. & Gronenborn, A. M. Relationships Between the Precision of High-Resolution Protein NMR Structures, Solution-Order Parameters, and Crystallographic B Factors. *J. Magn. Reson., Series B* **101**, 325-327, (1993).

145 Clore, G. M. & Schwieters, C. D. Concordance of Residual Dipolar Couplings, Backbone Order Parameters and Crystallographic B-factors for a Small α / β

Protein: A Unified Picture of High Probability, Fast Atomic Motions in Proteins. *J. Mol. Biol.* **355**, 879-886, (2006).

146 Doster, W., Cusack, S. & Petry, W. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature* **337**, 754-756, (1989).

147 Khodadadi, S., Pawlus, S., Roh, J. H., Garcia Sakai, V., Mamontov, E. & Sokolov, A. P. The origin of the dynamic transition in proteins. *J. Chem. Phys.* **128**, 195106, (2008).

148 Kim, C. U., Tate, M. W. & Gruner, S. M. Protein dynamical transition at 110 K. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 20897-20901, (2011).

149 Case, D. A. Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* **4**, 285-290, (1994).

150 Batista, P. R., Robert, C. H., Marechal, J. D., Hamida-Rebai, M. B., Pascutti, P. G., Bisch, P. M. & Perahia, D. Consensus modes, a robust description of protein collective motions from multiple-minima normal mode analysis--application to the HIV-1 protease. *PCCP* **12**, 2850-2859, (2010).

151 Brueschweiler, R. Normal modes and NMR order parameters in proteins. *J. Am. Chem. Soc.* **114**, 5341-5344, (1992).

152 Palmer, A. G. & Case, D. A. Molecular dynamics analysis of NMR relaxation in a zinc-finger peptide. *J. Am. Chem. Soc.* **114**, 9059-9067, (1992).

153 Joachimiak, M. P., Weisman, J. L. & May, B. JColorGrid: software for the visualization of biological measurements. *BMC Bioinformatics* **7**, 225, (2006).

154 Tiang, C. K., Cunningham, J., Wood, C., Hunter, I. C. & Davies, A. G. Electromagnetic simulation of terahertz frequency range filters for genetic sensing. *J. Appl. Phys.* **100**, 066105-066103, (2006).