LIGAND RECOGNITION BY THE MAJOR URINARY PROTEIN

Julie Roy BSc., MSc

The University of
Nottingham

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

November 2012

Declaration

I hereby declare that this thesis has been composed by myself and has not been submitted for any other degree previously. Acknowledgements of specific procedures not performed by myself are stated; otherwise the work described is my own.

Abstract

Molecular Dynamics (MD) and Quartz Crystal Microbalance (QCM) techniques can provide unique insights into what drives protein-ligand association. The major urinary protein (MUP) binds small ligands in a deeply buried hydrophobic pocket. Detailed calorimetric studies have shown that ligand binding is driven by enthalpic effects, not entropic effects [1]. Previous studies have shown that this is due to 'dewetting' of the binding site cavity even in the absence of ligands, and have also characterised the complex changes in molecular flexibility that accompany ligand binding-features that may be correlated with NMR data [2].

Recent MD revealed the hydration effects of apo-MUP and also shown where certain regions of MUP become more flexible upon ligand binding. They have also shown a water molecule remains close to the tyrosine in the binding pocket [2]. In our current MD studies and QCM experiments we have used wild type and 2 different mutants of MUP to study the binding effects of the ligand IBM. The first mutant has an OH group removed from the binding site of MUP (i.e. tyrosine to phenylalanine (Y120F)). The second mutant has an extra OH group in the binding site (i.e. alanine to serine (A103S)). For all three systems the hydration and flexibility upon ligand binding has been analysed. The hydration analysis from MD reveal (from radial distribution curves and hydration density maps) there is a small density of water that remains even without the presence of the ligand for the WT MUP whereas a larger density of water remains in the binding cavity of the A103S

hydrophilic MUP simulation. The results are based on the average structure generated from the 1 μs simulations. The Y120F MUP simulations reveal that there is no water molecules present in the binding cavity. However, as protein molecules are very dynamic in nature, water molecules are observed to hop in and out of the binding pockets for both mutant MUP (but not WT MUP) simulations over the 1 μs simulations. On the other hand the experimental QCM results reveal that on ligand binding no water loss is observed for Y120F mutant MUP whereas A103S and WT MUP have ~2 water molecules which are lost in the binding cavity.

The flexibility results from the MD simulations reveal that WT MUP have some residues which increase in flexibility whilst other residues which decrease in flexibility on ligand binding. However, the Y120F hydrophobic MUP show an overall decrease in flexibility whereas the A103S MUP shows an overall increase in flexibility on ligand binding. In contrast the experimental QCM and AFM results reveal that there is an increase in flexibility on ligand binding to all 3 different types of MUP molecules. The experimental and the simulation data have shown a variation in results but it is to be noted that the results cannot be directly compared as the analytical experiments are a surface based techniques whereas the MD simulations do not involve a surface. However, the contrast observed between computer simulation and experiments has revealed important information on the ligand binding effects on MUP.

[1] Bingham, R. J., J. B. C. Findlay, S. Y. Hsieh, A. P. Kalverda, A. Kjeliberg, C. Perazzolo, S. E. V. Phillips, K. Seshadri, C. H. Trinh, W. B. Turnbull, G. Bodenhausen, and S. W. Homans. 2004. 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.
[2] Barratt, E., R. J. Bingham, D. J. Warner, C. A. Laughton, S. E. V. Phillips, and S. W. Homans. 2005. Van der waals interactions dominate ligand-protein association in a protein binding site occluded from solvent water. J. Am. Chem. Soc. 127:11827-11834.

Acknowledgement

In the fields of observation, chance favours only the

prepared mind.

Louis Pasteur

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AC..................................alternating current

AFM................................atomic force microscope

AMBER...........................assisted model building with energy

refinement

ASP................................active site pressurization

BPTI...............................bovine pancreatic trypsin inhibitor

D.....................................dissipation

DMT................................Derjagin, Muller, Toropov

DNA................................deoxyribonucleic acid

F....... ....frequency

FFT.................................fast fourier transform

FTIR................................fourier transform infrared spectroscopy

GAFF...............................general amber force field

GPCR..............................G protein coupled receptors

GPU................................graphics processing unit

H-BONDS........................hydrogen bonds

HDM................................hydration density maps

HIV-PR............................human immunodeficiency virus-protease

IBM.................................2-isobutyl-3-methoxypyrazine

IR....................................infrared

ITC..................................iso-thermal calorimetry

K.....................................kelvin

M.....................................molar units

MD..................................molecular dynamics

MM.............................molecular mechanics

MMTEC.........................multimode tapping mode afm
electrochemical fluid cell

MUP.............................major urinary protein

NMR.............................nuclear magnetic resonance

OBP.............................odor binding protein

PBS.............................phosphate buffer solution

PCA.............................principal component analysis

PDB.............................protein data bank

PF-QNM.........................peak force-quantum nanomechanical

PME.............................particle mesh ewald

PMEMD..........................particle mesh ewald molecular dynamics

QCM-D..........................quartz crystal microbalance with dissipation

QM.............................quantum mechanics

QM/MM..........................quantum mechanics

QSAR...........................quantitative structure–activity relationship

RDF.............................radial distribution function

RMSD...........................root mean square deviation

RNA.............................ribonucleic acid

SBT.............................2-sec-butyl-4,5-dihydrothiqazole

SDS.............................sodium dodecyl sulfate

STM.............................scanning tunneling microscopy

TIP3P..........................three-site transferrable intermolecular
potential

UV.............................ultra violet

VDW...................................van der waals

VMD..................................visual molecular dynamics

WT....................................wild type

Chapter 1 – Introduction

Protein-ligand recognition has a central role in all processes in living systems. A complete understanding of protein interactions with small molecules is of great interest as it provides opportunities for faster and more effective processes in drug development. Molecular recognition is however a complex interplay of several factors such as intermolecular interactions of protein, ligand and the surrounding solvent, conformational variations of binding partners and the thermodynamics of molecular association. Over the past few decades experimental and computational techniques have been developed that shed light on the role of these factors. Our understanding of molecular recognition is still far from perfect. A brief literature review of some of the most important aspects of protein-ligand interactions is presented in this chapter followed by the aim and motivation of this research work.

1.1   Structure-based drug design

Drug discovery is an expensive and time-consuming activity. Drug discovery is an inventive process where the target is bound with potential new drugs. Retrospective analysis has shown that it can take approximately 14 years on average for a drug to come onto the market, each one costing 800 million dollars. However, only one compound out of nine that enters clinical trials eventually hits the market (Paul, S. M. et al., 2010). Drug discovery is a tough activity and wrong decisions

taken at the initial steps of the development often have a very harmful influence in the success of the drug candidate. Drug development can be classified into different stages which include drug target identification, lead component identification and clinical development. An important stage in this process is to identify and optimise a lead compound. This can be done by either high-throughput screening or by structure-based drug design. High throughput screening is used to test a large number of compounds for their ability to affect the activity of target proteins.

At present the accuracy of the predictions of structure-based drug design using molecular modelling and surface analysis techniques is limited, so there is much research into improving our understanding of the chemistry and physics of molecular recognition. On the other hand structure-based drug design is based on predicting how the structural and chemical features of a drug will determine how well it binds to a protein. Structure-based drug design originates from the lock and key concept (*figure 1*) where the key (ligand) can be specifically designed when the lock (protein) is known (Fischer, E., 1894).
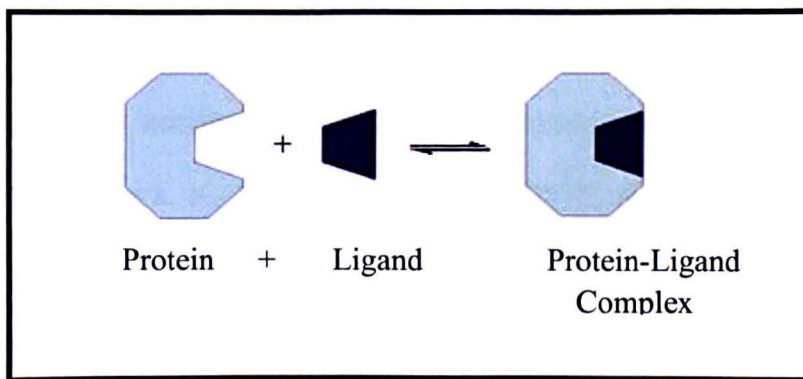
Figure 1   **A diagram representing the formation of a protein-ligand complex. The pale blue and navy blue blocks represent the lock (protein) and the key (ligand) respectively.**

Another model which describes the conformational changes of the receptor upon ligand binding is the induced fit theory. This theory postulates that first a molecule binds to a receptor and then the conformational changes in the macromolecule follow (Koshland, D. E., 1958). Such models have been applied to many protein-protein (Bui, J. M. and McCammon, J. A., 2006), protein-DNA (Levy, Y. et al., 2007) and protein-RNA (Williamson, J. R., 2000) models.

The conformational selection (or population shift) model is a popular alternative in structure-based drug design to the induced fit model. In this model, protein dynamics can lead it to constantly switch between a stable unbound conformation and a less stable bound conformation. In other words, the unbound protein spends most of its time in the lowest energy conformations, but also occupying higher-energy ones, some of which are structurally similar to the bound conformations. Due to the favourable interactions of the ligand, these conformers get partially selected during binding and the population of protein shifts in the direction of the bound conformations. This model has been a popular

model for antibody or small ligand binding systems (James, L. C et al., 2003; Arora, K. and Brooks, C. L., 2007; and Tobi, D. and Bahar, I., 2007).

Docking studies have also tried to use conformational selection by generating an ensemble of protein structures and docking small molecules against each of them in the hope of finding a bound conformation that will be stabilized by the ligand (Wong, S. and Jacobson, M. P., 2008). Another study has shown that strong and long-range protein-ligand interactions favour the induced fit model, while weak and short range interactions favour the conformational selection model (Okazaki, K. I. and Takada, S., 2008). Another recent study indicated that the determining factor in ligand binding is the timescale for transitioning between the unbound and bound states with both apo and holo forms (Zhou, H. X., 2010). They also found that switching between the two states is slow, whereas the induced fit model dominates when the switch between the two states are fast.

There have also been many studies that have shown that a blend of both models may describe more realistic systems. For example, the binding of ubiquitin to 19 different binding proteins (based on NMR structural analysis) have shown that both models play an equal role in the binding process (Wlodarski, T. and Zagrovic, B., 2009). Both the induced fit model and the conformational selection model use structural reorganisation within the protein to accommodate its binding partner. The ways in which they are differentiated are whether the structural changes occur before binding (conformational selection) or after binding

(induced fit) (*figure 2*). The main difference between the models of molecular recognition is protein flexibility and conformational heterogeneity.

In order for a protein to recognise a binding partner with the induced fit mechanism the protein must have sufficient flexibility and heterogeneity, whereas for a lock and key mechanism the protein must be rigid and homogenous. In contrast rigid proteins recognise their targets by the lock and key mechanism. However flexible proteins that recognise their targets by the induced fit or conformational selection can also recognise a whole range of other molecules. This might be detrimental if specificity of the protein to the target is required.

Induced Fit

Conformational Selection
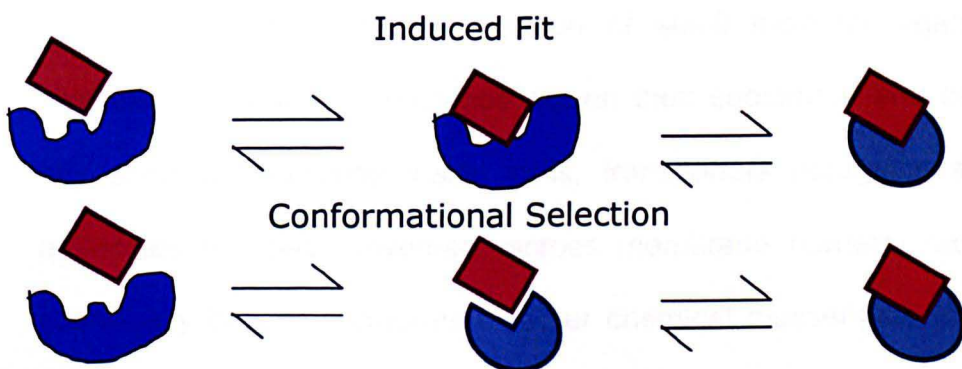
Figure 2   **Induced fit (top) and conformational selection (bottom) models of molecular recognition. The red and blue block represents the ligand and protein respectively.**

There have been many studies where protein-ligand recognition has contributed towards drug design. For example the binding of HIV protease (PR) to Nelfinavir and Amprenavir revealed that the binding of

5

Nelfinavir to HIV-PR was mainly entropy driven. It was also discovered that the entropy of binding of Amprenavir to HIV-PR is less favourable in comparison to the entropy of binding of Nelfinavir to HIV-PR (Deng, N. J. et al., 2011). This thesis aims to use computational and experimental techniques to develop knowledge of protein-ligand interactions in drug design.

1.2   Thermodynamics of Protein-ligand interaction

1.2.1 Enthalpy and Entropy

The non-covalent reversible binding of small-molecules to proteins has a dominant role in biology. Several processes which are vital to living systems involve specific recognition of small molecule ligands by proteins. For example, enzymes act on their substrates and catalyse key chemical reactions inside cells, transporters recognize specific molecules for their movement across membrane barriers, receptors specifically bind to hormones or other chemical messengers for inter- and intracellular communication and finally antibodies uniquely bind to foreign chemical agents to mount vital defence mechanisms against infection and disease. In general, the binding of a protein to a ligand in an aqueous environment is given by the reaction:

$$P_{(aq)} + L_{(aq)} \rightleftharpoons PL_{(aq)} \qquad \text{(Equation 1)}$$
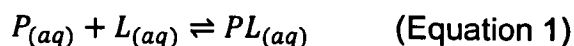
The dissociation constant K_D is described as:

$$K_D = \frac{[P][L]}{[PL]} \qquad \text{(Equation 2)}$$

The reciprocal of K_D is known as the association constant K_A and can also be used. In a simple case of a ligand binding to a protein binding site and is not affected by any other site on the protein, the value of K_D is the concentration of the ligand at which half of the binding sites are saturated (Dunn, C. D. et al., 2010). Therefore, K_D is known as the measure of affinity of a ligand towards its binding site and is measured in molar units, M. K_D can be measured experimentally (for example by isothermal calorimetry (ITC) experiments and estimated computationally. A small value of K_D would indicate that strong binding has taken place. Chemical reactions accompany a change in the free energy (ΔG) which is known as Gibbs free energy. The Gibbs free energy change can be shown in equation 3:

$$\Delta G = -RTln(K_D) \qquad \text{(Equation 3)}$$

where ΔG is the free energy of binding, T is the absolute temperature, R is the gas constant and K_D is the dissociation constant. If ΔG is large and negative then there is strong binding between the protein and the ligand and the interaction becomes favourable. It is influenced by change in two other important quantities known as the enthalpy (ΔH) which is the heat content and the entropy (ΔS) which is the degree of

disorder. The resulting relationship between these quantities can be given by:

$$\Delta G = \Delta H - T\Delta S \qquad \text{(Equation 4)}$$

The change in free energy of binding is influenced by several factors such as electrostatic and Van der Waals interactions (explained further in Chapter 2), ionization effects, conformational changes and the role of solvent. All these factors manifest themselves as favourable or unfavourable changes in enthalpy and entropy. For strong binding to occur the values obtained for ΔH would need to be large and negative whereas the values obtained for ΔS would need to be large and positive. The change in enthalpy is related to the breaking and formation of non-covalent interactions such as loss of protein-solvent and ligand-solvent (dewetting as explained further below) and the formation of protein-ligand hydrophobic contacts. The relative strengths of these interactions determine whether or not enthalpy change is favourable.

In the same way changes in entropy upon binding are related to solvent displacement and reduction in conformational degrees of freedom. The burial of lipophilic surfaces results in an increase in entropy whereas confinement of the ligand and protein-side chains has a decrease in entropy. Gain in enthalpy often also accompanies an unfavourable change in entropy as formation of precise interactions which causes structural rigidity and therefore decreases the entropy. This phenomenon is called enthalpy-entropy compensation (Perozzo, R. et al., 2004; Freire, M. G. et al., 2008). The interaction between a ligand

8

and its target (protein) maybe due to entirely non-bonded forces but in some cases covalent interaction may be involved.

1.2.2 Entropy in protein-ligand recognition

The entropy, S, refers to the measure of disorder or randomness within a system. This measurement can also be used to predict certain features of a protein structure on ligand binding and is more commonly known as the binding entropy. The two major components that contribute to the change in binding entropy are the conformational and the solvation entropies.

The conformational entropy arises from changes in the conformational degrees of freedom experienced by both the protein and ligand upon binding. The conformational entropy is usually negative (unfavourable) due to the loss of degrees of freedom resulting from the reduction in flexibility of both molecules (protein and ligand) upon binding. On the other hand the entropy of solvation is another important contribution to the entropy changes in protein-ligand recognition and can be described by the hydrophobic effect.

1.2.3 The Hydrophobic effect

The hydrophobic effect originates from the disruption of hydrogen bonds (H-bonds) between water and non-polar solutes. For example hexane does not form H-bonds with water. Therefore introducing hexane to water will disrupt the H-bond network within water molecules.

The H-bonds are partially reconstructed by building a water cage around the hexane molecule. However, these water molecules have restricted mobility. This leads to a loss in entropy of water molecules making the process unfavourable in terms of the free energy of the solute (Tanford, C., 1978). This effect is based on the tendency of (polar) water molecules to exclude non-polar molecules which leads to water and non-polar compounds to be segregated (Schravendijk, P. and van der Vegt, N. F. A., 2005).

There have been many examples of situations where the hydrophobic effect has been observed (Blokzijl, W. and Engberts, J., 1993; Pollack, G. L., 1991). Another example would be the addition of a ligand to a protein. This leads to ordered water molecules being released from the hydrophobic binding cavity, increasing the entropy of the system (*figure 3*).

Figure 3 **The hydrophobic effect in protein-ligand interactions. The blue and dark pink balls represent water and ligand molecules respectively and the light pink semi-circle represents a protein molecule.**
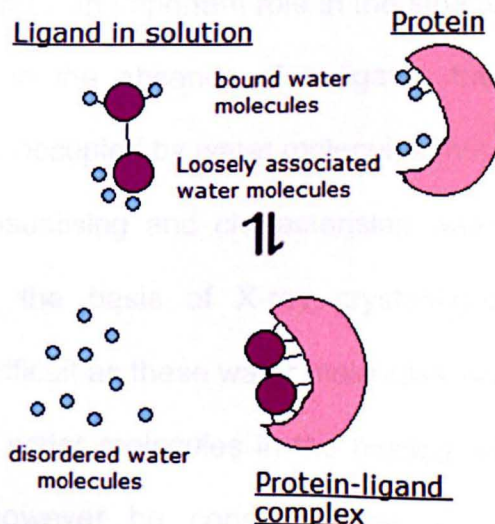
However, not all protein-ligand interactions are dominated by the hydrophobic effect, and so entropy-driven. Some examples are enthalpy driven because the target site and interacting ligand are not particularly hydrophobic; but there are also situations where, despite the protein and ligand appearing to have all the qualities required for a conventional hydrophobic-dominated recognition process, the binding is still enthalpy-driven. An example of an enthalpy-driven process is the binding of ligands to the major urinary protein (MUP) (see section 1.4).

1.3   Factors affecting protein-ligand binding affinity

In addition to shape complementarity between protein and ligands, there are other factors that affect the binding affinity of protein-ligand complexes. These are briefly discussed here.

1.3.1 Water molecules in the binding site

Water molecules play an important role in the structure and interactions of biomolecules. In the absence of a ligand, the binding site of a receptor is usually occupied by water molecules that are displaced upon ligand binding. Visualising and characterising water molecules in the binding sites on the basis of X-ray crystallographic structures is sometimes very difficult as these water molecules are highly disordered. Highly conserved water molecules in the binding sites across multiple structures can however be considered to be tightly bound. The

displacement of water molecules increases the entropy but it is offset by accompanying loss in enthalpy. The contribution of displacing a water molecule towards binding affinity therefore depends on how tightly it is bound and how efficiently the enthalpic loss by its displacement is compensated by interactions with the ligand molecule.

1.3.2 Flexibility

The conformational flexibility of proteins is a well-known phenomenon and an important consideration in molecular recognition. Proteins are flexible structures and conformational transitions of various scales play an important role in their function. Upon ligand binding, protein binding sites exhibit a variety of motions ranging from small-scale side chain rearrangements (for example amino acids such as Glycine and Alanine have small side chain as they have only one hydrogen (-H) and one methyl group (-CH_3) as their R-groups respectively) to loop movements in the active site. In some cases, protein structures with disordered regions undergo complete re-organization upon ligand binding.

1.3.2.1 Experimental approaches to flexibility

Nuclear magnetic resonance (NMR) is considered better suited to the study of structural dynamics of proteins. The main advantages are direct observation of protein in solution and the output in the form of an ensemble of low-energy conformations. An example of an NMR experiment is of *Schistosoma mansani*, a fatty acid binding protein

which indicate differences in protein flexibility between apo and holo forms. This is particularly the case within the ligand binding region (Garcia, A. E., 1992).

Systems such as G-protein coupled receptors (GPCRs) for which high resolution structural data is very challenging to generate using X-ray crystallography or NMR techniques, are widely studies using computational molecular dynamics (Dror, R. O. et al., 2011).

1.3.2.2 Computational methods to flexibility

Computational approaches such as molecular dynamics (MD) simulations are used greatly to characterize protein flexibility. Conformations generated from MD simulations can be used in other computational methods such as virtual screening, docking and scoring. Continuing developments in these areas are expected to improve our understanding of the role of protein flexibility in protein function and molecular recognition. A particular system of interest that considers these factors which influence the binding affinity of protein-ligand interactions is of MUP-ligand interactions.

1.4 MUP and its Interactions

Lipocalins are a family of globular proteins that have a primary function to bind and transport various ligands (for example volatile pheromones/odorants, fatty acids, lipids and steroids). They are

13

particularly involved in biological processes such as the protection of small hydrophobic pheromones against chemical damage in aqueous environment (Flower, D. R., 1996). Lipocalins may also play a part in cancer cell interactions (Bratt, T., 2000) as well as in the inhibition of sperm – egg binding in humans (Koistinen, H. et al., 2000) or in pregnancy (Crossett, B. et al., 1996). Additionally, lipocalins are abundant in both prokaryotic and eukaryotic cells. The structure of a lipocalin is usually based on an eight stranded beta barrel. Lipocalins do not have high similarity in amino acid sequence, but do share common sequence motifs. These motifs should be responsible for functional and structural properties among lipocalins, by playing a role in stabilizing their tertiary structure and by forming a receptor-binding site. An example of a lipocalin is the major urinary protein (MUP).

MUP is found in male mouse urine at high concentration (Finalyson, J. et al., 1968). The main function of MUP is to transport volatile pheromones in the urine (Cavaggioni, A. and Mucignat-Caretta, C., 2000). A series of ligand displacement studies were made in the late 1990's (Robertson, D. et al., 1998; Humpheries, R. E. et al., 1999) and found the MUPs played a role in the slow release of pheromones.

1.4.1 Structure of MUP

The X-ray crystal structure of MUP at 2.4 Å resolution was reported by Böcskei, Z. et al., (1992). The molecular weight of MUP is 19kDa (Flower, D. R., 1996) and it consists of 157 amino acid residues (Clissold, P. M. and Bishop, J. O., 1982). The isoelectric point of MUP is

14

between 4.2 and 4.7 (Duncan, R. et al., 1988). The major features of MUP's secondary and tertiary structure is characterised by a β-barrel structure consisting of 8 β-strands (colour coded in *Figure 4* as strand A, blue; B, yellow; C, green; D, magenta; E, cyan; F, red; G, black; and H, pink) and a single α-helix (Böcskei, Z, 1992).

Figure 4   **Structure of the major urinary protein (MUP) bound to the ligand 2-sec-butyl-4,5-dihydrothioazole (SBT) (green)**

The antiparallel beta sheets are hydrogen bonded to each other to form a closed beta barrel. The small helix (top left of *figure 4*) folds across one end of the barrel and a loop (coloured in grey) is formed which folds across the first two strands (bottom left of *figure 4*). The interior of the barrel consists of a hydrophobic cavity which forms the binding site for small hydrophobic ligand. The alpha-helix is packed against the outside of the globular beta-barrel conformation (Timm, D. E. et al., 2001).

1.4.2 Experimental Studies of MUP Interactions

There have been many experimental studies to characterize the interactions between MUP and various ligands. The experiments that are of particular interest are the ITC, NMR and X-ray crystallography data as it provides suggestions and explanations of the dewetting phenomenon in this protein-ligand recognition.

Isothermal titration calorimetry (ITC) experiments have shown that the ligand binding to MUP is enthalpy driven. The interaction of the hydrophobic ligands (in *figure 5*) with MUP can be expected to be entropy driven. The binding site of MUP is very hydrophobic and only one protein-ligand hydrogen bond is formed in each complex (between the residue tyrosine 120 in the binding cavity and the ligand) and therefore can undergo the classical hydrophobic effect (described previously in this chapter).

However this is not always the case. The interactions of MUP to various ligands have been described to be enthalpy and not entropy driven. It is necessary to look at the atomic detail to observe the important factors which are making the process to be enthalpy driven. For example, NMR studies of the interaction of MUP with various pyrazine ligands (including the ones listed in table 1) revealed entropy-entropy compensation (i.e. the residues from the binding site of the protein become more rigid whilst other residual side chains of the protein became more flexible) (Homans, S. W., 2005). A similar effect was observed for a calmodulin-peptide interaction (Lee, A. L. et al., 2000).

However, 1,8 Octandiol is an exception. The observed thermodynamical parameters, suggest that the interaction of 1,8 Octandiol with MUP encounters an enthalpy-entropy compensation which is common in bimolecular interactions in aqueous solution (Dunitz, J. D., 1995).

X-ray crystallography revealed that the binding site of the solved structure of MUP with 1,8 Octandiol was similar to other binding sites of other alcohols binding to MUP with no significant change to the protein structure as well as containing the same number of ordered water molecules in the binding site (Dunitz, J. D., 1995). It has been observed that the interaction of MUP to pentan-1-ol, octan-1-ol and nonan-1-ol to have similar low B-factors. This indicates a well ordered binding site in comparison to the main chain of the protein backbone (Day, P. J. et al., 1992; Brunger, A. T. et al., 1998). In contrast a single water was found to be in a different position for hexan-1-ol with a higher *B-factor* (40.65) indicating a more disordered binding site.

Figure 5   **The structures of ligands binding to wild-type (WT) MUP. These structures correspond to the values in table 1. The structures were produced using Chemdraw ultra 12.0 (CambridgeSoft, Cambridge, MA) (Cousins, K. R., 2011).**

pentan-1-ol

hexan-1-ol

heptan-1-ol

octan-1-ol

nonan-1-ol

1,8 octandiol

2,3 isobutylmethoxy pyrazine

6-hydroxy-6-methyl-3-heptanone

2-sec-butyl-4,5-dihydrothiazole

dehydro-exo-brevicomin

2,3 isopentylmethoxy pyrazine
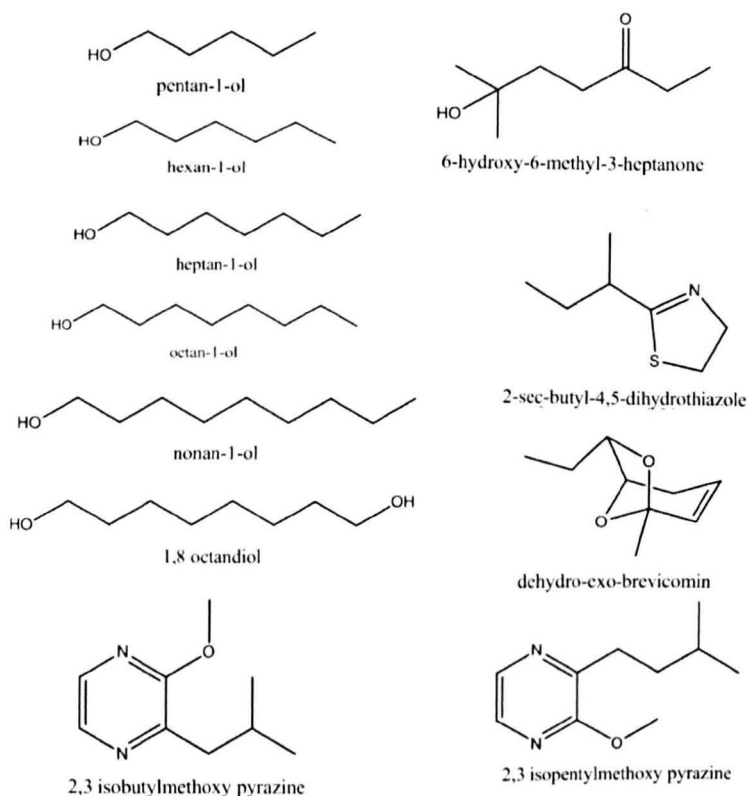
Table 1 Thermodynamic parameters of the interaction of wild-type MUP with various ligands associated to the structures in figure 5.

Molecule	K_D (μM)	ΔG (KJmol^{-1})	ΔH (KJmol^{-1})	ΔS (KJmol^{-1}K^{-1})
[c] Pentan-1-ol		-23.1	-41.0	-17.9
[c] Hexan-1-ol		-28.3	-47.6	-19.3
[c] Heptan-1-ol		-32.5	-53.4	-20.9
[c] Octan-1-ol	0.63	-35.6	-58.0	-22.4
[c] Nonan-1-ol	0.18	-38.8	-63.6	-24.8
[d] 1,8 Octandiol	1011.8	-17.2	-34.3	-17.1
[b] 2-sec-butyl 1-4,5-dihydrothazole	0.9			
[b] 6-hydroxy-6-methyl-3-heptanone	56			
[b] Dehydro-exo-brevicomin	26			
[a] 2,3 isopentylmethoxy pyrazin	1.8	-33.9	-44.5	-10.7
[a] 2,3 isobutylmethoxy pyrazine	0.3	-38.5	-47.9	-9.4

[a] Barrett, E. et al., 2006; [b] Sharrow, S. D. et al., 2002; [c] Bingham, R. J. et al., 2000, [d]Dunitz, J. D., 1995.

MUP has been studied by X-ray crystallography in both its apo- form and in complexes with a variety of ligands (Homans, S. W., 2007; Zidek, L. et al., 2002; Homans, S. W., 2004) and their crystal structures have now been published in the PDB website. Additionally, NMR data have shown that IBM (Homans, S. W., 2005) and other ligands (Zidek, L., et al., 1999; Shimokina, N. et al., 2006) binding to MUP increases the flexibility of a certain region of the protein whilst simultaneously decreasing the flexibility on other parts of the protein. This has led to Homans introducing the idea of 'entropy-entropy compensation' (Barrett, E. et al., 2005). Computational simulations have been used to investigate this phenomenon.

1.4.3 Computational Studies of MUP Interactions

Molecular modelling studies have suggested an explanation for the structural, dynamic, and thermodynamic changes associated with the binding of MUP to IBM. The ligand IBM in *figure 5* was used as it was observed by ITC experiments to be the strongest binder to MUP. The binding cavity of MUP remains largely unhydrated even in absence of a ligand (Bingham, R. et al., 2004). Therefore, on ligand binding very few water molecules are released, and so instead of being entropy driven, new non-bonded interactions are created and so the process becomes enthalpy driven which is commonly known as the dewetting phenomenon. Initially the idea of "nature abhorring a vacuum" met with some resistance in the dewetting phenomenon (Homans, S. W., 2007).

However the dewetting phenomenon has been observed in several situations and is discussed further in section 1.4.4.

1.4.4 The Dewetting Phenomenon

The dewetting of a solid by a liquid is a phenomenon which is important for a host of practical applications, (for example mineral flotation, oil recovery, detergency, lithography, and textiles to lubrication, microfluidics and self-cleaning windows). It can also play an important role in biological processes.

For example a molecular dynamics (MD) investigation on a streptavidin-biotin system found that the hydrophobic enclosures of the binding cavity of these systems would perturb the solvation of the binding site which would result in stabilising the bound complex. These hydrophobic enclosures would allow only a few energetically competitive water configurations, yielding entropic penalties of solvation (Friesner, R. A., 2006).

At a similar time dewetting was observed in the partitioning of small molecules on a lipid bilayer and they used umbrella sampling to calculate thermodynamic parameters (MacCallum, J. L. and Tieleman, D. P., 2006). Coarse-grained models were also used to show the dewetting effect in a protein-ligand complex (Michel, J. et al., 2006) and in another protein folding system (Singhal, N. et al., 2004). In an attempt to study this phenomenon more closely, an MD study of protein-ligand interaction was done using both explicit and implicit water models. The results suggested that the dewetting of the binding pocket

maybe regarded as the rate-limiting step in protein-ligand binding (Setny, P. et al, 2009).

AFM and fluorescence spectroscopy (Heier, J. et al., 2008) has shown experimental evidence of this phenomenon in polymer bilayers. Dewetting of liquid films has been developed for both simple and complex fluids (Reiter, G., 1992; Herminghaus, S. et al., 1998; Seemann, R., et al, 2005). Other examples include metastable liquid films which have been applied to partially wetted surface (Brochardwyart, F. et al., 1991; Mulji, N. et al., 2010), Solid water films on a liquid-cubic ice coexistence in droplets have observed the dewetting on a nickel surface of self-diffusion water molecules at high temperatures which was studied by IR and mass spectroscopy (Souda, R. et al., 2008), dewetting on a silica surface (Woerdeman, D. L. et al., 2002) and dewetting of perfluoropolyether was observed by FTIR and NMR (Shogrin, B., et al., 1996).

1.5   Aims and Objectives

This thesis is based on the research work undertaken under two broad themes; hydration of the protein binding site on ligand binding and protein flexibility and dynamics upon ligand binding. WT and mutant MUP-IBM interactions have proven to be a system of interest and therefore have been used as a test-bed in this thesis by computational (MD simulations) and experimental techniques (QCM and AFM experiments) which are further explained in Chapter 2.

The results section of this thesis aims to present findings of hydration (Chapter 3) of the binding cavity on MUP-IBM interactions by MD simulations and QCM experiments. Results from the changes in protein flexibility using MD simulation, QCM and AFM experiments are outlined in chapter 4. The final chapter (Chapter 5) aims to provide brief conclusions of this thesis.

1.6   References

Arora, K. and Brooks, C. L. (2007) 'Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism', Proceedings of the National Academy of Sciences of the United States of America, 104(47), 18496-18501.

Barratt, E., Bingham, R. J., Warner, D. J., Laughton, C. A., Phillips, S. E. V. and Homans, S. W. (2005) 'Van der waals interactions dominate ligand-protein association in a protein binding site occluded from solvent water', Journal of the American Chemical Society, 127(33), 11827-11834.

Barratt, E., Bronowska, A., Vondrasek, J., Cerny, J., Bingham, R., Phillips, S. and Homans, S. W. (2006) 'Thermodynamic penalty arising from burial of a ligand polar group within a hydrophobic pocket of a protein receptor', Journal of Molecular Biology, 362(5), 994-1003.

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. and Homans, S. W. (2004) 'Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein', Journal of the American Chemical Society, 126(6), 1675-1681.

Blokzijl, W. and Engberts, J. (1993) 'hydrophobic effects - opinions and facts', Angewandte Chemie-International Edition in English, 32(11), 1545-1579.

Bocskei, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E. V., Cavaggioni, A., Findlay, J. B. C. and North, A. C. T. (1992) 'pheromone binding to 2 rodent urinary proteins revealed by x-ray crystallography', Nature, 360(6400), 186-188.

Bratt, T. (2000) 'Lipocalins and cancer', Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology, 1482(1-2), 318-326.

Brochardwyart, F., Hervet, H., Redon, C. and Rondelez, F. (1991) 'spreading of heavy droplets .1. theory', Journal of Colloid and Interface Science, 142(2), 518-527.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu,

N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. (1998) 'Crystallography & NMR system: A new software suite for macromolecular structure determination', Acta Crystallographica Section D-Biological Crystallography, 54, 905-921.

Bui, J. M. and McCammon, J. A. (2006) 'Protein complex formation by acetylcholinesterase and the neurotoxin fasciculin-2 appears to involve an induced-fit mechanism', Proceedings of the National Academy of Sciences of the United States of America, 103(42), 15451-15456.

Cavaggioni, A. and Mucignat-Caretta, C. (2000) 'Major urinary proteins, alpha (2u)-globulins and aphrodisin', Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology, 1482 (1-2), 218-228.

Clissold, P. M. and Bishop, J. O. (1982) 'variation in mouse major urinary protein (mup) genes and the mup gene-products within and between inbred lines', Gene, 18(3), 211-220.

Cousins, K. R. (2011) 'Computer Review of ChemDraw Ultra 12.0', Journal of the American Chemical Society, 133(21).

Crossett, B., Allen, W. R. and Stewart, F. (1996) 'A 19 kDa protein secreted by the endometrium of the mare is a novel member of the lipocalin family', Biochemical Journal, 320, 137-143.

Day, P. J., Shaw, W. V., Gibbs, M. R. and Leslie, A. G. W. (1992) 'acetyl coenzyme-a binding by chloramphenicol acetyltransferase - long-range electrostatic determinants of coenzyme-a recognition', Biochemistry, 31(17).

Deng, N. J., Zhang, P., Cieplak, P. and Lai, L. H. (2011) 'Elucidating the Energetics of Entropically Driven Protein-Ligand Association: Calculations of Absolute Binding Free Energy and Entropy', Journal of Physical Chemistry B, 115(41), 11902-11910.

Dror, R. O., Pan, A. C., Arlow, D. H., Borhani, D. W., Maragakis, P., Shan, Y., Xu, H. and Shaw, D. E. (2011) 'Pathway and mechanism of drug binding to G-protein-coupled receptors', Proceedings of the National Academy of Sciences of the United States of America, 108(32), 13118-13123.

Duncan, R., Matthai, R., Huppi, K., Roderick, T. and Potter, M. (1988) 'GENES THAT MODIFY EXPRESSION OF MAJOR URINARY PROTEINS IN MICE', Molecular and Cellular Biology, 8(7), 2705-2712.

Dunitz, J. D. (1995) 'win some, lose some - enthalpy-entropy compensation in weak intermolecular interactions', Chemistry & Biology, 2(11), 709-712.

Dunn, C. D., Sulis, M. L., Ferrando, A. A. and Greenwald, I. (2010) 'A conserved tetraspanin subfamily promotes Notch signaling in Caenorhabditis elegans and in human cells', Proceedings of the National Academy of Sciences of the United States of America, 107(13), 5907-5912.

Finlayson, J. Mushinsk.Jf, Hudson, D. M. and Potter, M. (1968) 'components of major urinary protein complex of inbred mice - separation and peptide mapping', Biochemical Genetics, 2(2), 127-&.

Fischer, E. (1894) '- Einfluss der Configuration auf die Wirkung der Enzyme', - 27(- 3), - 2993.

Flower, D. R. (1996) 'The lipocalin protein family: Structure and function', Biochemical Journal, 318, 1-14.

Freire, M. G., Carvalho, P. J., Gardas, R. L., Marrucho, I. M., Santos, L. and Coutinho, J. A. P. (2008) 'Mutual solubilities of water and the C(n)mim Tf(2)N hydrophobic ionic liquids', Journal of Physical Chemistry B, 112(6), 1604-1610.

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C. and Mainz, D. T. (2006) 'Extra precision glide: Docking and scoring incorporating a model of

hydrophobic enclosure for protein-ligand complexes', Journal of Medicinal Chemistry, 49(21), 6177-6196.

Garcia, A. E. (1992) 'large-amplitude nonlinear motions in proteins', Physical Review Letters, 68(17).

Heier, J., Groenewold, J., Castro, F. A., Nueesch, F. and Hany, R. (2008) Enlarged bilayer interfaces from liquid-liquid dewetting for photovoltaic applications - art. no. 69991J, translated by Strasbourg, FRANCE: J9991-J9991.

Herminghaus, S., Jacobs, K., Mecke, K., Bischof, J., Fery, A., Ibn-Elhaj, M. and Schlagowski, S. (1998) 'Spinodal dewetting in liquid crystal and liquid metal films', Science, 282(5390), 916-919.

Homans, S. W. (2004) 'NMR spectroscopy tools for structure-aided drug design', Angewandte Chemie-International Edition, 43(3), 290-300.

Homans, S. W. (2005) 'Probing the binding entropy of ligand-protein interactions by NMR', Chembiochem, 6(9), 1585-+.

Homans, S. W. (2007) 'Dynamics and thermodynamics of ligand-protein interactions' in Peters, T., ed. Bioactive Conformation I, 51-82.

Humphries, R. E., Robertson, D. H. L., Beynon, R. J. and Hurst, J. L. (1999) 'Unravelling the chemical basis of competitive scent marking in house mice', Animal Behaviour, 58, 1177-1190.

James, L. C., Roversi, P. and Tawfik, D. S. (2003) 'Antibody multispecificity mediated by conformational diversity', Science, 299(5611), 1362-1367.

Koistinen, H., Koistinen, R., Hyden-Granskog, C., Magnus, O. and Seppala, M. (2000) 'Seminal plasma glycodelin and fertilization in vitro', Journal of Andrology, 21(5), 636-640.

Koshland, D. E. (1958) 'application of a theory of enzyme specificity to protein synthesis', Proceedings of the National Academy of Sciences of the United States of America, 44(2), 98-104.

Koshland, D. E., Simmons, N. S. and Watson, J. D. (1958) 'Absence of phosphotriester linkages in tobacco mosaic virus', Journal of the American Chemical Society, 80(1), 105-107.

Lee, A. L., Kinnear, S. A. and Wand, A. J. (2000) 'Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex', Nature Structural Biology, 7(1), 72-77.

Levy, Y., Onuchic, J. N. and Wolynes, P. G. (2007) 'Fly-casting in protein-DNA binding: Frustration between protein folding and electrostatics facilitates target recognition', Journal of the American Chemical Society, 129(4), 738-739.

MacCallum, J. L. and Tieleman, D. P. (2006) 'Computer simulation of the distribution of hexane in a lipid bilayer: Spatially resolved free energy, entropy, and enthalpy profiles', Journal of the American Chemical Society, 128(1), 125-130.

Michel, J., Verdonk, M. L. and Essex, J. W. (2006) 'Protein-ligand binding affinity predictions by implicit solvent simulations: A tool for lead optimization?', Journal of Medicinal Chemistry, 49(25), 7427-7439.

Mulji, N. and Chandra, S. (2010) 'Rupture and dewetting of water films on solid surfaces', Journal of Colloid and Interface Science, 352(1), 194-201.

Okazaki, K. I. and Takada, S. (2008) 'Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms', Proceedings of the National Academy of Sciences of the United States of America, 105(32), 11182-11187.

Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R. and Schacht, A. L. (2010) 'How to improve R&D productivity: the pharmaceutical industry's grand challenge', Nature Reviews Drug Discovery, 9(3), 203-214.

Perozzo, R., Folkers, G. and Scapozza, L. (2004) 'Thermodynamics of protein-ligand interactions: History, presence, and future aspects', Journal of Receptors and Signal Transduction, 24(1-2), 1-52.

Pollack, G. L. (1991) 'why gases dissolve in liquids', Science, 251(4999), 1323-1330.

Reiter, G. (1992) 'dewetting of thin polymer-films', Physical Review Letters, 68(1), 75-78.

Robertson, D., Hurst, J., Hubbard, S., Gaskell, S. J. and Beynon, R. (1998) 'Ligands of urinary lipocalins from the mouse: Uptake of environmentally derived chemicals', Journal of Chemical Ecology, 24(7), 1127-1140.

Schravendijk, P. and van der Vegt, N. F. A. (2005) 'From hydrophobic to hydrophilic solvation: An application to hydration of benzene', Journal of Chemical Theory and Computation, 1(4), 643-652.

Seemann, R., Herminghaus, S., Neto, C., Schlagowski, S., Podzimek, D., Konrad, R., Mantz, H. and Jacobs, K. (2005) 'Dynamics and structure formation in thin polymer melt films', Journal of Physics-Condensed Matter, 17(9), S267-S290.

Setny, P., Wang, Z., Cheng, L. T., Li, B., McCammon, J. A. and Dzubiella, J. (2009) 'Dewetting-Controlled Binding of Ligands to Hydrophobic Pockets', Physical Review Letters, 103(18).

Sharrow, S. D., Vaughn, J. L., Zidek, L., Novotny, M. V. and Stone, M. J. (2002) 'Pheromone binding by polymorphic mouse major urinary proteins', Protein Science, 11(9), 2247-2256.

Shimokhina, N., Bronowska, A. and Homans, S. W. (2006) 'Contribution of ligand desolvation to binding thermodynamics in a ligand-protein interaction', Angewandte Chemie-International Edition, 45(38), 6374-6376.

Shogrin, B., Jones, W. R. and HerreraFierro, P. (1996) 'Spontaneous dewetting of a perfluoropolyether(c)', Lubrication Engineering, 52(9), 712-717.

Singhal, N., Snow, C. D. and Pande, V. S. (2004) 'Using path sampling to build better Markovian state models: Predicting the folding rate and

mechanism of a tryptophan zipper beta hairpin', Journal of Chemical Physics, 121(1), 415-425.

Souda, R. (2008) 'Roles of deeply supercooled ethanol in crystallization and solvation of LiI', Journal of Physical Chemistry B, 112(9), 2649-2654.

Tanford, C. (1978) 'hydrophobic effect and organization of living matter', Science, 200(4345), 1012-1018.

Timm, D. E., Baker, L. J., Mueller, H., Zidek, L. and Novotny, M. V. (2001) 'Structural basis of pheromone binding to mouse major urinary protein (MUP-I)', Protein Science, 10(5), 997-1004.

Tobi, D. and Bahar, I. (2005) 'Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state', Proceedings of the National Academy of Sciences of the United States of America, 102(52), 18908-18913.

Williamson, J. R. (2000) 'Induced fit in RNA-protein recognition', Nature Structural Biology, 7(10), 834-837.

Wlodarski, T. and Zagrovic, B. (2009) 'Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin', Proceedings of the National Academy of Sciences of the United States of America, 106(46), 19346-19351.

Woerdeman, D. L., Parnas, R. S., Giunta, R. K. and Wilkerson, A. L. (2002) 'Dewetting of unreacted epoxy/amine mixtures on silica', Journal of Colloid and Interface Science, 249(1), 246-252.

Wong, S. and Jacobson, M. P. (2008) 'Conformational selection in silico: Loop latching motions and ligand binding in enzymes', Proteins-Structure Function and Bioinformatics, 71(1), 153-164.

Zhou, H. X. (2010) 'From Induced Fit to Conformational Selection: A Continuum of Binding Mechanism Controlled by the Timescale of Conformational Transitions', Biophysical Journal, 98(6), L15-L17.

Zidek, L., Stone, M. J., Lato, S. M., Pagel, M. D., Miao, Z. S., Ellington, A. D. and Novotny, M. V. (1999) 'NMR mapping of the recombinant mouse major urinary protein I binding site occupied by the pheromone 2-sec-butyl-4,5-dihydrothiazole', Biochemistry, 38(31), 9850-9861.

Chapter 2 - General methods

2.1 Computational Methodology - Molecular Mechanics

Molecular Dynamics (MD) simulations were first introduced in the late 1950's to study the interactions of hard spheres of an ionic crystal (Alder, B. J. et al., 1957; Alder, B. J. et al., 1959). The first protein simulation was then developed in the 1970's, where an 8.8 ps simulation was performed for a bovine pancreatic trypsin inhibitor (McCammon, J. A. et al., 1977).

2.1.1 Molecular Interactions

The interaction between atoms can be modelled using either the quantum mechanics (QM) approach or by the molecular mechanics (MM) approach. The first approach uses the modelling of electrons in a system so that the full electronic structure can be obtained. However, for a larger system (for example a protein) this approach becomes computationally very expensive. For example a quantum mechanics/molecular mechanic (QM/MM) simulation of bacteriophophytin has been performed (composing of 2 ps of QM and 17 ps of MM) (Makri, N. et al., 1995).

Another example of a QM/MM simulation of Crambin was performed to detect protein dynamics. QM was simulated for 60 ps (and 350 ps simulation was performed by pure MM) (Liu, H. Y. et al., 2001). On the other hand, the molecular mechanics (forcefield methods) method is much preferred for larger systems as it ignores the electronic motions of

a system and takes into account the force and velocity on each atom from which the energies of a system can be calculated (Leach, A. R., 2001).

2.1.2 The Forcefield

Most forcefields are composed of a series of terms that model inter and intra molecular forces in a system. These would include bonded and non-bonded terms, and as a result the potential energy can be calculated by summing up the different terms:

$$\vartheta(r^N) = \sum_{bonds} \frac{k_l}{2}(l_l - l_{l,0})^2 + \sum_{angles} \frac{k_l}{2}(\theta_l - \theta_{l,0})^2 + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) + \sum_{i=1}^{N} \sum_{j=i+1}^{N}(4\varepsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}) \quad \text{(Equation 5)}$$

where $\vartheta(r^N)$ is the potential energy which is a function of the positions (r) of N particles, l_i is the bond length which deviates from the equilibrium bond length ($l_{i,0}$), θ_i is the bond angle, $\theta_{i,0}$ is the reference bond length, k_l is the force constant, V_n is the barrier height, n is the multiplicity, ω is the torsional angle and γ is the phase factor, ε_{ij} is the well depth σ_{ij} is the collision diameter, r_{ij} separation between the two atoms, ϵ_0 is the dielectric constant, q_i and q_j are the partial charges for atoms i and j. Each of the above terms describes aspects of the nuclear motions of a system (figure 6).
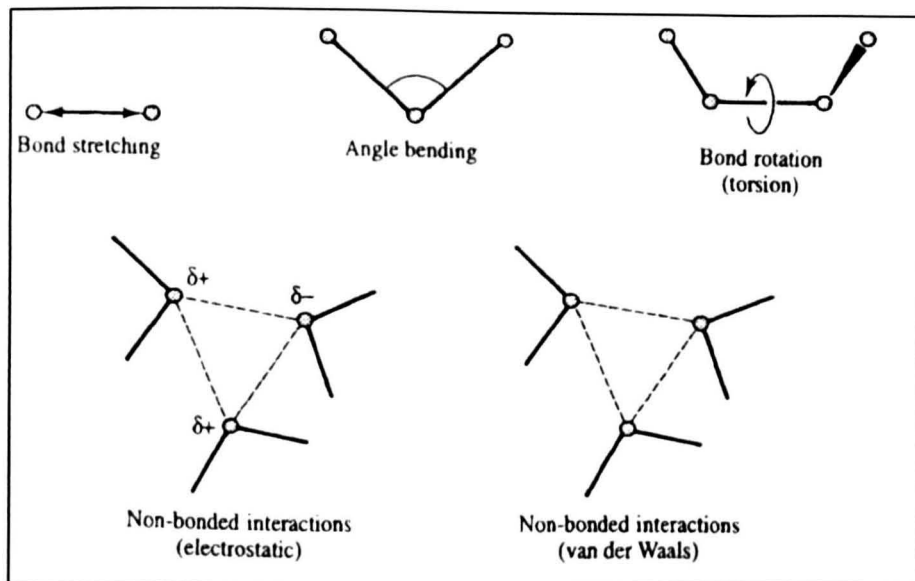
Figure 6 **The key components in a molecular mechanics forcefield. The bonded terms are on the top row whereas the non-bonded terms are on the bottom row. (Leach, A. R., 2001)**

A more detailed discussion of the bonded and non-bonded terms will now be given.

2.1.3 Bonded Terms

As figure 6 suggests the bonded terms are made of bond stretching, angle bending and of torsional bond rotation. The energetics associated with the motions described here can be represented by the first 3 terms from *equation 5*.

A good approximation of energy change associated with a pair of bonded atoms stretching can be derived from the first term in *equation 5*. However, if there is a dramatic increase in bond lengths the equation becomes less accurate. This is because according to Hooke's law, at

the bottom of the potential energy curve the harmonic potential is similar to the Morse potential (*figure 7*) at equilibrium and so at a higher bond length l_i deviates from the reference value l_{i0}.
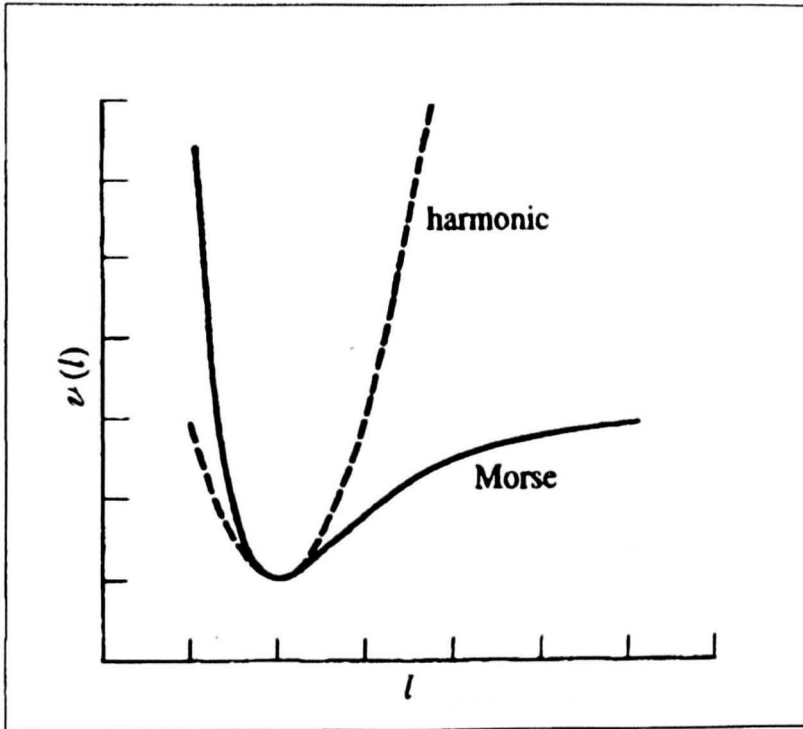
Figure 7   **Comparison between the harmonic potential and the morse curve to reveal information about the bond stretch energy (($\vartheta(l)$ on the y-axis) with the interatomic separation ((l) on the x-axis) (taken from Leach., 2001)**

In *equation 5* the harmonic potential which describes the bond energy is defined by $\vartheta(l)$, k_i is the force constant for the bond length l_i which deviates from the equilibrium bond length $l_{i,0}$.

$$\vartheta(l) = \Sigma_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 \qquad \text{(Equation 6)}$$

The energy change that accompanies angle bending can be well represented by the second term in *equation 5*. This equation is suitable

for many applications. In this equation the harmonic potential for a bond angle ($\vartheta(\theta)$) is defined by the bond angle θ_i which deviates from the equilibrium bond angle $\theta_{i,0}$. The bond angle force constant is defined as k_i in *equation 7*.

$$\vartheta(\theta) \sum_{angles} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 \qquad \text{(Equation 7)}$$

The third term in *equation 5* is the bond dihedral or torsional term. This term is more complicated than the other 2 terms and is calculated using a series of cosine functions that can describe the energy as a function of rotation angle around a given bond. The preference for either staggered or eclipsed structures is reproduced by this energy function.

$$\vartheta(\omega) = \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \qquad \text{(Equation 8)}$$

In the above equation V_n is the barrier height, n is the multiplicity, ω is the torsional angle and γ is the phase factor. An important note about the bonded terms is that the electrostatic or van der Waals effects in the bond stretching and angle bending interactions are assumed to be included. However, for the torsional term the electrostatics and the Van der Waals effects must be calculated in their own right to give full illustration of the rotation about a dihedral angle (Cheatham, T. E. and Kollman, P. A., 2000).

2.1.4 Non-bonded Terms

This is the last term in *equation 5*. It defines the interactions which occur between atoms and not through a bond. These interactions are mainly made up of Van der Waals and electrostatic interactions (as explained below).

2.1.4.1 Van der Waals Interactions

The van der Waals (VDW) interaction arises from attractive (long range) and repulsive (short range) forces. The attractive force is known as the dispersive force (London, F. and Polanyi, M., 1930) which is due to instantaneous dipoles which arise during the fluctuations in the electron clouds. These dipoles can induce a dipole in neighbouring atoms giving rise to polarizability.

When the internuclear separation is short the interaction energy is inversely proportional to r but when it is long the interaction energy decays exponentially $(exp(-2\,r/a_o))$ where a_o is the Bohr radius. Modelling the VDW interactions between atoms and molecules can be calculated by quantum mechanics calculations but becomes computationally expensive for large systems. Therefore a well-known function for VDW interactions for large systems is the Leonard-Jones 12-6 function which takes the following form between two atoms:

$$\vartheta(r) = 4\varepsilon[(\tfrac{\sigma}{r})^{12} - (\tfrac{\sigma}{r})^{6}] \qquad \text{(Equation 9)}$$

The interatomic separation for two atoms can be represented by r, the well depth by ε and a constant by σ from *equation 9*. The attractive term

can be accounted for $(\frac{\sigma}{r})^6$ whereas the repulsive term $(\frac{\sigma}{r})^{12}$ in equation 9. These parameters of the Lennard-Jones potential can be shown graphically in *figure 8*:

Figure 8 **Lennard-Jones Potential energy curve for the interaction of 2 molecules which shows both the attraction (b) and repulsion (a) forces between non-ionic particles, the well depth ε, the minimum energy interaction distance σ and the distance of separation $r(\sigma)$.**
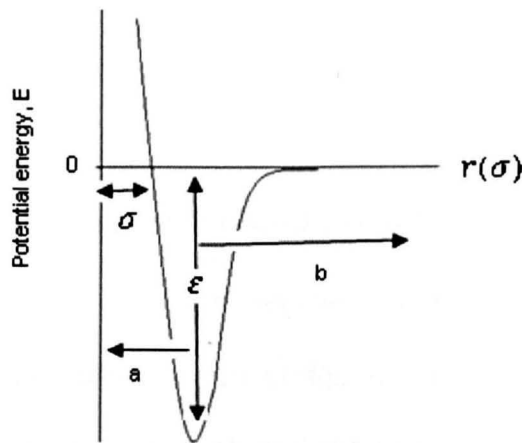
In figure 8, a deep well depth ε, indicates a stronger interaction between two molecules.

2.1.4.2   Electrostatic interactions

The remainder of *equation 5* is based on electrostatic interactions. An uneven distribution of charge within a molecule can rise if areas within a molecule have different electronegativities. This charge distribution is represented by partial atomic charges designed to reproduce the electrostatic properties of the molecule. Therefore electrostatic complimentary between the protein and the ligand at the binding

interface is vital for complex formation. The predominant types of electrostatic interactions include hydrogen bonding, salt bridges and metal interactions (Gohlke, H. and Klebe, G., 2002; Exner, O. and Bohm, S., 2003). The electrostatic interactions between a protein and a ligand are then the sum of the interaction between pairs of partial charges calculated using Coloumb's law. Electrostatic interactions originate from Coloumb's law:

$$E = \frac{q_i q_j}{4\pi\varepsilon_0 r} \frac{1}{r}$$
(Equation 10)

where q_i and q_j are the partial charges of the atoms i and j, ε_0 is the dielectric constant and r is the relative separation between the two atoms. For example when a salt bridge is established its calculated coulombic interaction energy is often three or four times higher than the total van der Waals interaction and this effect can generate aritfacts. However, if the desolvation energies are taken into account then the problem can be compensated because salt-bridges have a high desolvation penalty which counterbalances the coulombic energy (Shoichet, B. K. et al., 1999; Huang, D. and Caflisch, A., 2004).

2.1.4.3   Hydrogen bonding

Hydrogen bonding is the most important directional interaction in biological macromolecules, known for conferring stability to protein structure and selectivity to protein-ligand interactions. In general,

hydrogen bonding occurs between two electronegative atoms, one of which (donor) has a covalently bound hydrogen atom whereas the other (acceptor) has a lone pair of electrons. The strong electrostatic attraction arises from the attractive interaction between partial positive charge on the hydrogen atom and partial negative charge on the acceptor atom. The hydrogen bonding term in many interactions between the donor hydrogen atom and the heteroatom acceptor atom can be described by the 10-12 Leonard-Jones potential:

$$\vartheta(r) = \frac{A}{r^{12}} - \frac{C}{r^{10}}$$ (Equation 11)

where the non-bonded VDW energy is described as $\vartheta(r)$, A and C are constants based on the atoms involved and r^{12} and r^{10} are the interatomic separation between two molecules. However, hydrogen bonding is a form of dipole-dipole interaction, so the 10-12 potential is a correction term and is not always used in molecular dynamic simulations.

2.1.5 Parameterisation of the forcefield

To run a molecular mechanics calculation, a set of parameters is needed to model a system. These parameters are known as the forcefield parameters. These parameters are derived from a range of sources (x-ray structures, IR spectra, QM methods). To derive a forcefield from scratch, it is a very time-consuming process. An atom type is assigned to each atom in the system which is being modelled.

This atom type is dependent on the atom and what is bonded to it. For example hydrogen which is bonded to a sp^3 carbon will have a different atom type, bond length and force constants to a hydrogen bonded to a sp^2 carbon. Therefore parameters would be required to describe different bond types. These atom types are used to describe bonded (bond type, angle type and dihedral type) and non-bonded interactions (VDW and electrostatic interactions) in the parameterisation process within a system being modelled. Once the forcefield parameters and equation has been established, molecular structures and energetic data can be established.

Thermodynamic properties can be reproduced from a correctly parameterised forcefield by using computer simulation techniques (Jorgensen, W. L. and Tiradorives, J., 1988). The parameterisation of forcefields can also be completed by least squares fitting to a set of parameters that gives the optimal fit to the data (Lifson, S. and Warshel, A., 1968). For example an application of this approach is the derivation of a forcefield for peptides by fitting crystal data to a range of compounds (Hagler, A. T., 1977).

The steps that are involved in a molecular dynamics simulation will now be discussed (i.e. equilibration process and the production simulation).

2.1.6 Equilibration Process

Prior to an MD simulation, energy minimisation is required to identify a stable low energy conformation of a molecule. There can be many

different conformations of the same molecule, each with different energies.

For example, conformations with steric clashes tend to have high energies and so the conformation that has the lowest energy is likely to give the most stable state. The variation in energies is known as the potential energy surface. The energy at the bottom of the lowest point on the potential energy surface is known as the 'global minimum' which is the energy that we are interested in because at this energy the conformation found is at the most stable state. The global energy minimum is difficult to find as most minimisation algorithms can only find the energy closest to the energy minimum of the starting structure.

During the energy minimisation process there are generally two-steps employed which are known as the steepest descent and conjugate gradient method.

2.1.6.1 Steepest Descent Method

This method is based on a simple algorithm, in which it finds the steepest direction on the potential energy surface and moves the structure in a parallel direction to the force, i.e., directly downhill on the potential energy surface. At each step the gradient of each potential is calculated and a displacement is added to all the co-ordinates in a direction opposite to the gradient (i.e. in the direction of the force). The step size is increased if the lower energy results and decreased otherwise. As the algorithm uses finite displacement steps, it does not follow the gradient smoothly down to the minimum and therefore

convergence to the minima is slow. Despite this, the steepest descent method is useful when the starting conformation is considerably different from the global minimum conformation. It is also useful as the first step of a multi-step minimisation scheme.

2.1.6.2   Conjugate Gradient Method

This is a better method when the molecules are closer to the energy minimum. The initial step is equivalent to the steepest descent method. Subsequent steps involve a search along a line on the surface that is a mixture of the current negative gradient and the previous search direction (i.e. the direction of the minimisation step is determined as a weighted average of the current gradient and the direction taken in the previous iteration). A better representation of the conformation found at the global minimum of a potential energy surface can be found more quickly using this method in comparison to other methods (i.e. Newton-Raphson methods or quasi-newton methods (Leach, A. R., 2001)).

2.1.7 Multistep Equilibration

After the energy minimisation step the molecule undergoes a multistep equilibration process (Shields, G. C. et al., 1998). This involves the gradual heating of the molecule to 300 K. The equilibration is needed so that thermodynamic properties (i.e. temperature, pressure etc.) can be monitored. Equilibration is done so that the system can develop from the starting configuration to reach equilibrium.   The starting

configuration is usually chosen from structures determined by x-ray crystallography, NMR or theoretical modelling. However, it is important to add other solvent molecules so that a realistic environment for the solute is generated. For example, the biological molecules (i.e. proteins, ligands e.t.c.) can be immersed in pre-equilibrated solvent molecules and any solvent which would have steric clashes with the molecule can be discarded before the simulation commences. The periodic boundary conditions and non-periodic boundary approaches to calculating solvent boxes is further discussed in section 2.1.8.2.

2.1.8 Molecular Dynamics

MD can be used to study the motion and dynamics of a molecule over time. Newton's equations of motions can be applied to form a new set of atomic positions, which is generated as a trajectory

Newton's laws of motion state that an object (e.g. in this scenario, an atom or molecule) will move in a straight line with constant velocity unless a force has acted on it. The rate of change in a force is equal to the change in momenta, and for every action there is an equal and opposite reaction (Newton, I. et al., 1729).

$$F = ma \qquad \text{(Equation 12)}$$

By applying Newton's second law of motion (*Equation 12*) a trajectory can be derived (where F is the force, m is the mass and a is the

acceleration). The change of variables with respect to time (*dt*) in an

MD simulation can be generated by *equation 13*:

$$\frac{d^2x}{dt^2} = \frac{F_i}{m_i}$$

(Equation 13)

where m_i is the mass of a particle i and F_i is the force on the particle i,

in that direction.

2.1.8.1   Time Step

The size of the time step used to create trajectories becomes important

because if the time step is too small then only a small portion of the

phase space is covered whereas if it is too large then the system will

give errors which will lead to unrealistic structures with high-energy

overlap between atoms.

The continuous potential models used to create trajectories is assumed

to be a pairwise additive process and the integration can be broken

down into many small stages. For example, a time stepping algorithm

known as the Verlet algorithm (Verlet, L., 1967) is a common method in

a molecular dynamics simulation to integrate the equations of motion. It

uses the positions and accelerations at time *t* and the positions from the

previous step to calculate the new positions. A disadvantage to this

algorithm is the lack of a direct definition of the velocity.

There are many variations of the Verlet algorithm. For example the

leap-frog algorithm (Hockney, R. W., 1982), the velocity Verlet method

(Ferguson, D. M. et al., 1992) and the Beeman's algorithm (Thorpe, M.

F. and Beeman, D., 1976) are different variations of the Verlet algorithm. It is important to use the optimal time step as using a smaller time step would mean the simulation would take longer which means that the computational cost would become expensive. The size of the time step is dependent on the timescale of motions in a system of interest.

A typical time step is 1 *fs* is usually used because this is the smallest timescale of motions within a molecule and is correlated to bond vibrations. If the SHAKE (Rueda, M. et al., 1977) algorithm is used then the bond length between 2 atoms can be constrained so a longer time step can be employed to reduce computational costs. However, there are other algorithms that can also be used (e.g. RATTLE (Andersen, H. C., 1983))

2.1.8.2   Periodic Boundary conditions

Biological molecules are usually present in an aqueous environment and therefore to represent this system computationally a droplet of water can surround the solute. However, it may become difficult to control the edges of the water droplet as the water would have to be dependent on the shape of the molecule being represented. During the MD simulation, if a particle leaves the solvated box then an image particle replaces it by entering from the opposite side (Allen, M. P. and Imbierski, A. A., 1987) (i.e. the assumption is that the system is surrounded by an infinite number of identical copies and the particles in the simulation box move unison with their images). Particles which are

near the boundary of the simulation box interact with periodic images across the boundary. An example of a boundary is the use of the truncated octahedral water box (Adams, J. M., 1983) surrounding WT MUP (*figure 9*).

Figure 9 **MUP (PDB 2ozq (Syme, N. R. et al., 2007) in a truncated octahedral water box (Image constructed from VMD (Humprey et al., 1996))**

2.1.8.3 Long range forces

Non-bonded cutoffs are often used in non-bonded interactions and are applied to a minimum image convention. In this convention, each atom is imaged by another atom in the system (and repeated in the periodic boundary method). As the cutoff is being used, the interactions between pairs of atoms that is further than the cutoff value is set to zero. This will take into account the closest image. However the cutoff should not be so large that a particle sees its own image or even the same molecule

twice. Therefore the cutoff is limited to half the length of the simulation box length.

The interactions that do not decay faster than r^{-n}, (where n is the dimensions of the system and r is the complete set of $3N$ particle co-ordinates) can be a problem. This is because the range of interactions involved is often greater than half the simulation box length. Coulomb (r^{-1}) and dipolar (r^{-3}) interactions are examples of long range interactions. Treatment of long-range forces is important because it will help to ensure that certain properties such as the dielectric constant are correct. Different methods have been developed to handle long-range forces but in this case the particle mesh ewald (PME) will be looked at.

2.1.8.4   Particle mesh ewald molecular dynamics (PMEMD)

The study of ionic crystals led to the development of the Ewald sum method (Ewald, P. P., 1921). The method is used to calculate interaction energies of periodic systems (like crystals). This method can become computationally expensive and so to overcome the costs the fast fourier transform (FFT) algorithm has been implemented in this method. The FFT method involves replacing each particle-based point charge by a grid based charge distribution. A range of variants of PME methods exist. For example, the particle-mesh Ewald was developed in the 1990's (Darden, T., 1993) and particle-particle-particle-mesh (PPPM) approach (Bernard, C. et al., 1988; Luty, B. A. et al., 1994, 1995). Later a unification of these approaches was then developed (Deserno, M. and Holm, C., 1998).

2.1.9 Long timescale simulations

Many important biological processes occur in the micro- to millisecond time regime. Many studies have shown that in this time frame important biological processes can result from the interaction of molecules. For example protein folding, enzyme catalysis, signal transduction, ligand binding and allosteric regulation occurs in this time-scale. Studying the slow motions at this time-scale remains a challenge for both experimentalists and theoreticians. The first ever microsecond simulation of a biomolecule using classical molecular dynamics on parallel computers was performed in 1998. This simulation was done for the folding processes of the 36-residue villin headpiece in explicit solvent and represented a milestone in the field of molecular dynamics (Duan, Y. and Kollman, P. A., 1998).

However it remains the case that, even though the power of computation has increased and efficient simulation algorithms have developed over the years, MD simulations of most proteins have been restricted until recently to the 10 to 100 nanosecond time-scale. As the simulation timescale increases the validity of forcefields is a problem that still remains. For example a recent development by the Orozco group in Barcelona and collaborators found that long simulations of DNA using 'standard' forcefield (more commonly known as AMBER FF03) showed the rearrangement of a torsion angle which lead to the development of the revised "parmBSC0" forcefield. (Perez, A. et al. 2007).

Most recently there have been some major advances. D.E Shaw's group have developed a specialised computer called Anton. This machine is designed to run the entire MD simulation on specialised computer nodes which can accelerate the computation of non-bonded forces with special cores that perform the remainder of the numerical calculations (Shaw, D. E. et al., 2007). The new parallelized machine has successfully been used to simulate the dynamics in the micro-millisecond time regime of biomolecules. For example, the reversible folding and unfolding mechanism of Fi35 and the conformational behaviour of BPTI have been simulated in the long-timescale regime (Shaw, D. E. et al., 2010).

Another advance in computationally intensive simulations on biomolecules was developed by the Pande group at Stanford who launched the first distributed computing project. This project was designed to allow everyone to run short molecular dynamic simulations on their individual machines which could be sent back to the Pande group for analysis (Ensign, D. L. et al., 2007). Another recent advance is the development of the general purpose graphical processing unit (GPU) which has been harnessed to increase the computation power available for simulating biomolecules (Ganesan, N. et al., 2011).

Despite all the advances in computation speed and power, it cannot be assumed that all the problems for MD simulations on biological molecules on a long-timescale regime have been solved. Therefore, further investigation on protein-ligand recognition can validate long

timescale simulations as well as reveal anything new about the interaction.

2.2   Analysis methodology

The analysis of the long-timescale trajectories was done so that information on the hydration (radial distribution curves, hydration density maps and water molecule count in binding cavity) of the binding cavity (Chapter 3) and flexibility (RMSF fluctuations, vector analysis) of the ligand (IBM) binding to WT MUP and 2 mutants of MUP (chapter 4) could be evaluated and compared to experimental studies.

2.2.1 Reliability and Reproducibility of simulations
2.2.1.1   Secondary structure analysis

Ramachandran analysis has been a very popular way of establishing the probable "correctness" and quality of protein structures for many years (Saul, F. A. et al., 1978; Ho, B. K. and Brasseur, R., 2005). The conformation of a protein backbone is defined by three torsion angles at every residue (see *Figure 10*).

Figure 10   Torsion angles of a peptide bond (Zhou, 2011)
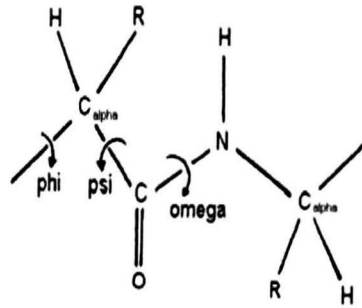
The phi and psi angles are flexible and rotation occurs. It is observed that certain combinations of phi and psi angles are more commonly observed than others. This is because they a) avoid steric clashes and b) favour particular H-bonding patterns along the chain (alpha helix or beta sheet). A plot formed with these values is known as a Ramachandran plot (after G.N. Ramachandran). Ramachandran and colleagues assessed all the different phi psi angle combinations for an alanyl dipeptide and found phi psi angles that are consistent with allowed interatomic separation. They concluded that the phi psi angles of a protein structure could be predicted without considering solvent or electrostatic interactions.

Therefore a map can be generated containing three regions that are favoured, surrounded by slightly larger areas that are 'allowed'. However, anything on the outside of these regions is called 'unfavoured' or 'disallowed' (Ramakrishnan, G. N. et al., 1963, 1965). Ramachandran plots are a very popular technique against which the stability of new crystal structures can be evaluated by the use of a program called Procheck (Laskowski, R. A. et al., 1993).

2.2.1.2   RMSD analysis

MD trajectories of a protein are commonly analysed by comparison of each snapshot with a reference structure which is known as the root mean square deviation (RMSD). It is a very commonly-used method to calculate the average difference between two (the crystal structure and the simulation frame structure is commonly used) least squares best fitted structures of superimposed proteins (Diamond, R., 1976, 1988; Kabsch, W., 1976, 1978). The RMSD can be calculated using *equation 14*:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x_{iref})^2 + (y_i - y_{iref})^2 + (z_i - z_{iref})^2} \text{ (Equation 14)}$$

where N are the number of atoms, x_i, y_i and z_i are the co-ordinates of atom i after best superposition on a structure and x_{iref}, y_{iref} and z_{iref} are the co-ordinates of the atom i in the reference structure (i.e. crystal structure).

Two identical structures would have an RMSD of 0 angstroms. The larger the RMSD value (only after fitting) the more dissimilar the structures would be. A more realistic measure of the RMSD of a protein is to calculate the difference between each snapshot of an MD trajectory rather than the starting crystal structure of the protein. This becomes very useful in long-time MD simulations where the structure may expect to incur small changes over time. An advantage of the RMSD values is that the consistency (or plateauing out on an RMSD

graph) of the RMSD values may mean that the conformation has reached stability. The molecule is known to be stable if the RMSD is lower than 3.4 Å.

2.2.2 Flexibility analysis

2.2.2.1 Root mean square fluctuations (RMSF)

The root mean square fluctuations (RMSF) have been found to be experimentally important in many proteins as they are a measure of local atomic flexibility within the protein (Cooper, A., 1976; Karplus, M. and McCammon, J. A., 1981; Allen, T. W. et al., 2004). The RMSF for an atom i can be calculated by *equation 15:*

$$RMSF_i = \sqrt{\frac{1}{N_f}\sum_{i=1}^{N_f}(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \qquad \text{(Equation 15)}$$

where N_f is the number of frames, x_j, y_j and z_j refer to the average structure calculated along the trajectory after superposition. RMSF can be derived from x-ray crystallographic B-factors (*equation 16*) and can be compared with MD simulations (Frauenfelder, H. et al., 1979).

$$B_i = \frac{8\pi^2}{3}(RMSF_i)^2 \qquad \text{(Equation 16)}$$

2.2.2.2   Principal component analysis

One possible way of studying the dynamics of a system is by principal component analysis (PCA). This method was introduced under the name of "quasi-harmonic analysis" by Karplus (Karplus, M. and Kushick, J. N. et al., 1981) to help understand the computation of entropies. It has been a valuable tool for analyzing dynamics in long time simulations of proteins (Kitao, A. et al., 1992; Amadei, A. et al., 1993 and Garcia, A. E., 1992). It is a statistical method for analysing MD trajectories and can be used to find the components that make up the greatest overall contribution to the motion within a trajectory.

A 3N x 3N covariance matrix of the Cartesian co-ordinates is generated from an MD simulation and then diagonalised to give 3N eigenvectors. The eigenvectors provide a vectorial representation of each component of structural deformation. Each eigenvector has a corresponding eigenvalue that indicates the relative contributions made by the component to the motion as a whole. The eigenvectors associated with the highest eigenvalues can be selected for further analysis.

Projections of the trajectory along the major eigenvectors give useful information on equilibrium and conformational sampling (Wlodek, S. T. et al, 1997). To ease the interpretation of the deformations associated with each principal component, short MD trajectories (animations) can be generated artificially by generating structures in which the projections vary linearly between the minimum and maximum values

observed. The resulting animations can be inspected visually using graphics packages such as VMD (Humphrey, W. et al., 1996).

PCA of a trajectory produces a description of the essential modes of deformation (eigenvectors) of the structure, as discussed above. Motion along each eigenvector relates to an orthogonal way in which the structure samples configurational space. If two separate simulations of a system are dynamically equivalent, they must sample this space the same way. This can be achieved if they have the same eigenvectors, or linear combinations of each other. Therefore the overall similarity in the dynamics of two trajectories can be determined by calculating the dot product matrices and the subspace overlap which can be represented by the Hess overlap equation (*equation 17*) (Hess, B., 2000).

$$\gamma_{AB} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}(V_i^A \cdot V_j^B)^2 \qquad \text{(Equation 17)}$$

where n is the number of snapshots, V_i^A and V_j^B is the number of eigenvectors included in the analysis.

The use of replicates is simply a way of improving computational methods. It provides a way to test out the reproducibility and replicability of the simulations. PCA allows us to see how well the dynamical behaviour of the systems converges between replicates and is conserved by ligand binding. The suitability of the dimensions for subspace must be considered for analysis as mentioned above (Rueda, M. et al, 2007). In this thesis therefore PCA is performed on each trajectory for each replicate of MUP (apo and holo) independently.

2.3   Experimental Methodology

2.3.1 Quartz crystal microbalance (QCM)

The QCM is a surface analytical technique which in this thesis has been used to detect mass and rigidity changes of a layer of molecules deposited on the surface of a quartz crystal. All measurements were performed using a technique termed quartz crystal microbalance with a dissipation (or QCM-D), (Rodahl, M. et al., 1995). An overview of the theoretical background and operation of the instrument are described in the following section.

2.3.1.1   Theoretical background

The QCM is a sensitive method for the characterisation of binding processes at surfaces (Höök, F. et al., 2002). It is a mass detection technique, in which changes in resonant frequency, Δf (Hz), of a quartz crystal are monitored as material is deposited on/lost from its surface (Saubrey, G., 1959).

In the QCM-D instrument, a quartz crystal sensor (*figure 11*) is sandwiched between two gold electrodes. An AC current applied to the electrodes produces an oscillation of the crystal, which changes as mass is deposited onto the surface; an increase in mass causes the crystal's resonant frequency, f, and it's harmonics to decrease proportionately. The quartz crystal oscillates perpendicular to the electric field.

If the absorbed layer on the crystal surface is rigid then a change in mass can be calculated from the change in frequency using the Saubrey equation (Saubrey, G., 1959):

$$\Delta m = -\frac{C}{n}\Delta f \qquad \text{(Equation 18)}$$

where Δm is the change in mass, C is the mass sensitivity constant (equal in the set up employed to 17.7 ng cm^{-2} Hz^{-1} for f=5 MHz), n is the overtone (or harmonic) number and Δf is the change in frequency (Höök, F. et al,. 2001).

In the QCM-D instrument a dissipation signal, D, is also monitored simultaneously to the frequency signal, f. D is defined by *equation 19* and has no units:

$$D = \frac{E_{lost}}{2\pi E_{stored}} \qquad \text{(Equation 19)}$$

where E_{lost} is the energy lost during each oscillation cycle and E_{stored} is the total energy stored in the oscillator.

Figure 11   Biomolecules on a quartz crystal surface.

$$\Delta f \propto \Delta m$$
$$\Delta D \propto \text{stiffness}$$

Time (µs)
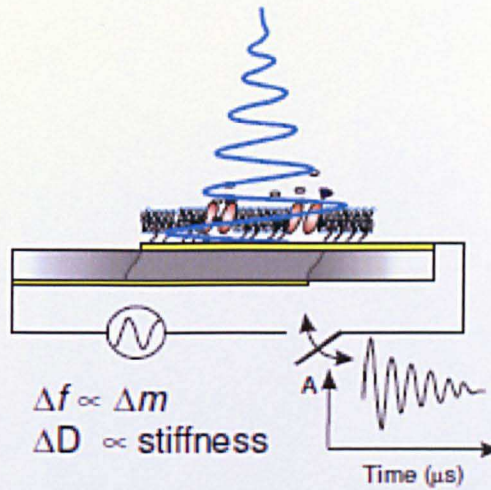
The difference in rigid and viscoeleastic sample can be demonstrated by using decay curves for the amplitude of the quartz crystal oscillations once the AC is voltage is switched off (*figure 12*).
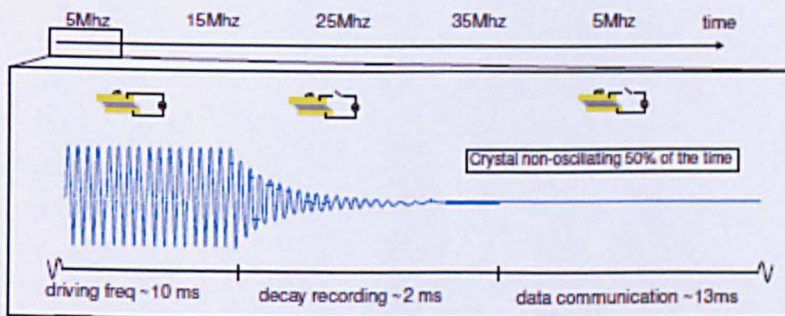
Figure 12   Decay curves for quartz crystal oscillations. An AC voltage is pulsed across a quartz crystal at the crystals resonant frequency and at several overtones, driving the crystal to oscillate at the resonant frequency. The shear wave at each frequency is allowed to dampen and the inverse of the decay time constant is recorded simultaneously to produce Δf and ΔD curves.

D can give useful information on the surface rigidity and dynamics of an immobilised molecular layer on the crystal surface. QCM-D has been utilised in a variety of experiments, for example to study lipid vesicle adsorption (Keller, C. A. and Kasemo, B., 1998, and Edvardsson, M. et al., 2005), cell adhesion (Hook, F. et al., 1998), and surface interaction studies (Wang, G. et al., 2008). Solvation effects of biological substances (Bingen, P. et al., 2008), antibody-antigen interactions (Hook, F. et al., 1998) and protein adsorption with biomolecules (von Vacano, B. et al., 2011, Sander, M. et al., 2010 and Anand, G. et al., 2011) have also been investigated.

2.3.1.2   Instrumentation

QCM-D studies were performed in Chapter 4 and 5, using the Q-Sense D300 measurement system (Q-Sense AB, Gothenburg, Sweden). The QCM-D sensor crystal consists of a 0.3 mm 5 MHz AT-cut quartz crystal, which was gold coated (*Figure 13*). The upper surface of the sensor consists of a large gold electrode with a diameter of 14 *cm* (*Figure 13*). An AC current is applied between the 2 electrodes which enables the crystal to oscillate laterally, and frequency and dissipation signals can be obtained.

Active electrode
(reaction substrate)
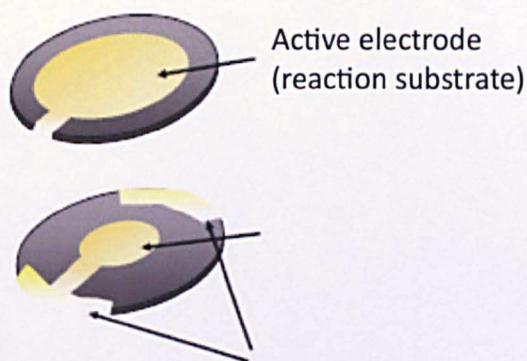
Figure 13 **The QCM-D sensor crystal. The active electrode (diameter=14mm) is the upper surface which is the sensor surface. The counter and contact electrodes are the bottom surface of the quartz crystal. When an AC current is applied between the electrodes the crystal oscillates laterally.**

In liquid environments, the limit for mass sensitivity (f) is in the order of 5 ng.cm^{-2}, and the dissipation factor (D) is approximately 3×10^{-7} for the unloaded 5 *MHz* crystal. The crystal resonant frequency (f) and the dissipation factor (D) of the oscillator were measured simultaneously at the fundamental resonant frequency (5 *MHz*) and at the 3rd overtone (15 *MHz*). The shifts in frequency and dissipation (Δf and ΔD) from the measurements performed at the 15 *MHz* 3rd overtone are presented in this thesis due to the increased sensitivity of the signal at this frequency. A Δf of 1 *Hz* corresponds to the deposition of 5.9 ng.cm^{-2} when the 3rd overtone of 15 *Hz* is used, compared to 17.7 ng.cm^{-2} when the fundamental frequency of 5 *MHz* is employed.

2.3.1.3   Quartz Crystal Cleaning and Preparation

Prior to performing the experiments described in this thesis, the quartz crystals were cleaned by approaches developed by Krozer and Rodahl (1997). The technique involved the crystals being sonicated in an ultrasonic bath with 50:50 v/v water:acetone mixture for 20 mins, followed by drying in nitrogen gas for 5 mins. This was then followed by an oxidation process, where the crystals were oxidised by exposure to ozone, induced within a UV/ozone cleaner (BioForce Nanosciences, Inc., IA, USA) for 20 mins.

The peak intensities of the UV lamp were at wavelengths of 185 and 254 *nm*. The UV light intensity was slightly below 20 mW/cm^2 at the 254 *nm* wavelength. The distance between the lamp and the quartz crystal was less than 5 *mm*. To ensure the crystals were clean the crystals were then also exposed to oxygen plasma generated by a barrel etcher (PT7100 CB10-RAD, Polaron Division) for 30 secs at 40 V, with chamber pressure of 20 Pa. The sensor crystal was stabilised in pH 7.4 phosphate buffer solution (PBS) (Sigma Aldrich, UK). A volume of ~0.5ml of the temperature-stabilised sample liquid is delivered to the chamber containing the sensor crystal (sensor loop) to ensure a complete exchange of the liquid.

Successive injections of the protein MUP was introduced to the chamber until *f* and *d* became stabilised. This was followed by PBS then IBM and then PBS again were made in the experiment. However, the addition of different solutions to the chamber (i.e. PBS/MUP/IBM) allowed the graphs to produce signal distortion (peaks of *f* < 2 *Hz* and *D*

< 0.01 10^{-4}). A simultaneous frequency and dissipation graph is formed at the 1st, 3rd, 5th and 7th overtone. All the experiments were carried out at 20°C. The quartz crystal was then inserted into the QCM chamber. The QCM liquid cell was cleaned thoroughly with SDS solution between experiments. *Figure 14* presents a photograph of the experimental setup used in this thesis.

WT and mutant MUPs were immersed in ph 7.4 PBS solution to make up a concentration of 1 µM which has been used in all QCM experiments unless otherwise stated. The IBM used in the QCM experiments in chapter 3 and 4 was also immersed in ph 7.4 PBS solutions. However, the pre-incubated WT MUP was composed by immersing 0.5g of WT MUP in 50 cm^3 of 50 µM concentration of IBM at room temperature. The solution was left for 30 mins so binding could take place between the WT MUP and IBM prior to the start of the experiments.
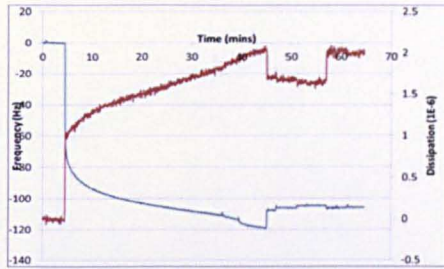
Sample frequency and dissipation graph output
of WT-MUP and IBM interaction
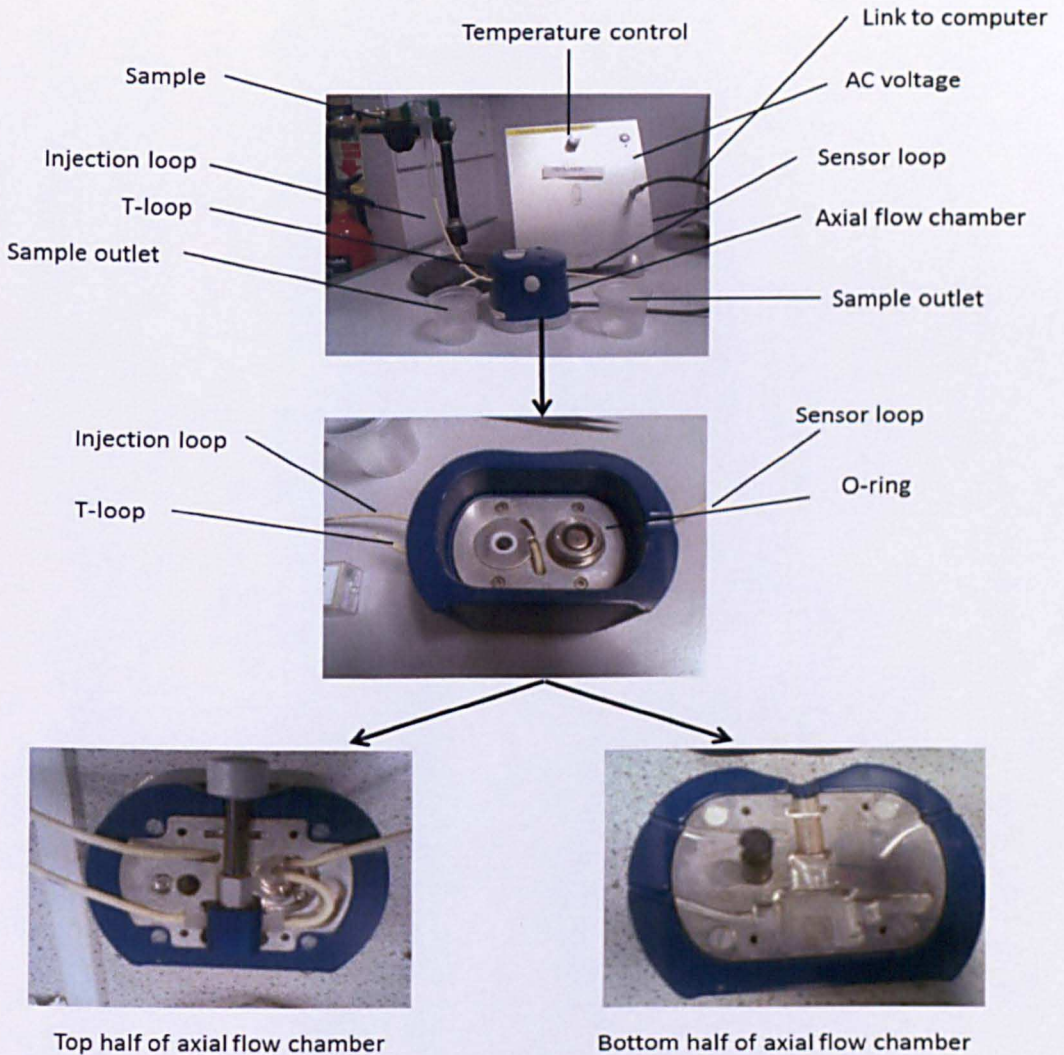
Figure 14   QCM-D experimental set-up. Simultaneous plots of frequency and dissipation versus time are acquired throughout the experiment (see example graph at top). The photographic picture directly under the graph is of the experimental setup. The second photographic picture is of the inside of the chamber where the quartz crystal would be mounted. The chamber is then split up into 2. The left is the top half of the axial flow chamber whereas the bottom half is situated on the right.

2.3.2 Atomic Force Microscope (AFM)

Binnig and colleagues in 1986 won the Nobel Prize in Physics for the design of the scanning tunnelling microscope (STM) (Binnig, G. et al., 1986). The atomic force microscope (AFM) is a descendent of this instrument combining principles from the STM and the stylus profilometer. The AFM was developed in 1986 (Binnig, G. et al., 1986) as it allowed investigations of a wider range of samples (e.g. insulating, including biological, samples) in a variety of environments (such as in liquid).

The AFM has become an increasingly popular technique for imaging surface topology to the nanometre scale, and in particular for the analysis surfaces functionalized with biological molecules (as reviewed by Alessandrini, A. and Facci, P., 2005; Jalili, N. and Laxminarayana, K., 2004). Also, force spectroscopy measurements can be made using the AFM where the forces acting between the tip and the sample can be investigated. The peak force-nanomechanical (PF-QNM) AFM will be used in this thesis and the technique used will be discussed here.

2.3.2.1   AFM mode of operation

The AFM operates through detection of the forces between the tip and sample surface. To acquire images an AFM cantilever is scanned (in the x,y plane) across the surface of the sample; during this process forces between the probe and the sample result in deflection of the cantilever spring, which results in movement of the cantilever in the z-

direction. A laser beam reflected off the back of the cantilever–tip assembly reflects off the cantilever surface to a split position sensitive photodetector, which detects the resultant cantilever deflections. In the simplest instrument configuration, the position information recorded by the detector is used as the input to a computer to form three dimensional images of surface topography. An experimental setup of the AFM is presented in *figure 15*.

Figure 15   **The AFM apparatus. A fluid tip is mounted on the end of a flexible Si_3N_4 cantilever which is attached to a piezo crystal. The laser beam light is deflected from the upper side of the end of the cantilever by a photodetector and cantilever deflection can be monitored. This is in closed-loop/constant force mode where a feedback circuit moves the piezo up and down in z to retain a constant cantiever deflection.**
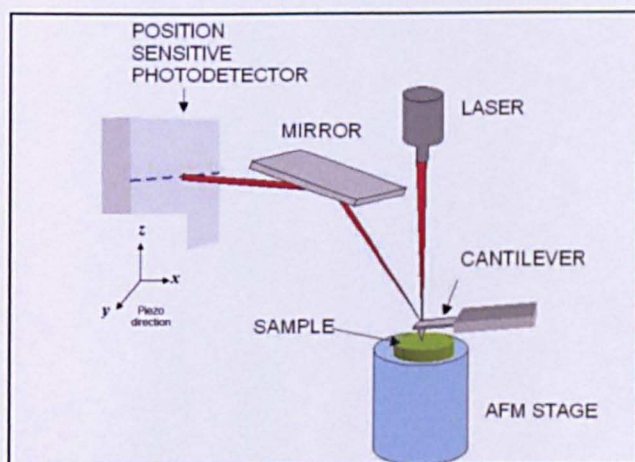
As the probe scans across the surface the cantilever undergoes deflections due to attractive and repulsive forces experienced by the tip. Therefore the deflection of the cantilever (Typically 50-200 μm long and 0.55-0.65 μm thick) is a measure of the forces experienced by the tip.

For the imaging of soft biomolecular samples the tip has a low spring constant (< 1N/m), to provide a high level of force control.

The optimum resolution of the image is greatly affected by the shape and the apex of the imaging tip (located at the end of the cantilever assembly). The shape of the tips is pyramid shaped (*figure 16*) which can be problematic, for example, it may result in the contact occurring between the sides of the tip and the sample, in addition to the apex. This would cause a tip broadening artefact whereby the features on the images appear larger than the true size. In addition, two features which are next to each other maybe perceived as one.

2.3.2.2   AFM imaging modes and force curves

There are three different types of modes for data acquisition; contact, tapping and non-contact mode. Each mode functions between different force regimes (attractive and repulsive) of the tip-sample interaction.

During the early stage of AFM development, the introduction of tapping mode (Hansma, P. K. et al., 1994 and Zhong, Q. et al., 1993), also called "intermittent contact" mode, was a significant landmark in terms of expanding the applications of AFM to delicate or weakly adsorbed samples and made AFM applicable for almost any type of solid or semi-solid sample surfaces. By doing so, the AFM overcame the strong lateral forces, associated with contact mode that damaged soft surfaces or swept the samples away from the imaging area.

In addition to providing useful information on surface topography, AFM can also be used to acquire force measurements as a measure of the tip and sample interaction.
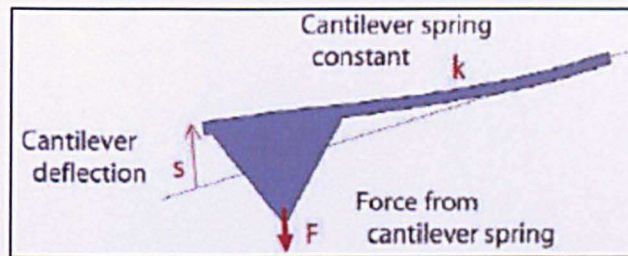
Figure 16 Force calculations from the tip (purple triangular shape) and cantilever (the bar attached to the tip).

The magnitude of the tip-sample forces during AFM imaging can be estimated using Hook's law:

$$F = -KS$$
(Equation 20)

where F is the force between the probe and sample (N), K is the cantilever spring constant (N/m) and S is the cantilever deflection (m) (*Figure 16*). The formation of force curves is further explained below.

In an AFM force measurement the probe is moved vertically at constant z-piezo velocity towards a sample surface until contact is made between the probe and sample, before being retracted away. There are two steps involved in generating a force-distance plot. The first is to measure the deflection of the cantilever as a function of vertical distance. For example, the output of the photodetector can be converted from volts to units of force for the system must be calibrated

i.e. the photodetector output is calibrated to the actual cantilever displacement (deflection sensitivity). The cantilever is driven into a hard surface assuming that the tip-sample gap is at 0. The second step is to calculate the spring constant to produce a force curve (like *Figure 17*). The slope of the force curve in the contact region on a hard surface allows for calibration of the detector sensitivity (e.g. to convert *V* to *nm*). The spring constant then allows conversion of *nm* to *F*.

The spring constant can calculated in a variety of different methods each with their own advantages and disadvantages. For example, by observing the resonant frequency changes whilst known masses are added to the free end of the cantilever (Manne, S. et al, 1993). Another possible method is by pressing the cantilever against another reference cantilever of a known spring constant (Gibson, C. T. et al., 1997). The spring constant in this work was calculated by a thermal excitation method (Hutter, J. L. and Bechhoefer, J., 1993).

Direct calculation of the spring constant can be performed by using the fundamental resonant peak of the cantilever under thermal excitation. A correction must be made to take into account the optical setup of the system i.e. how photo-detector and laser are aligned with the cantilever (these changes can produce large errors in spring constant measurement).

Figure 17 shows a typical force-distance curve in which the y-axis is the cantilever deflection force (measured in nN) versus the distance travelled by the z-piezo (measured in nm) on the x-axis). In *Figure 17*, region I shows that the tip is far from the sample as no interaction force

is detected. As the tip approaches the surface (red line) no interaction is detected until the cantilever is deflected toward the sample due to attractive VDW forces (region II).

After the contact of the tip to the surface, the cantilever deflects back (black line) due to repulsive forces (region III), as a result of overlapping electron orbitals between the atoms of the sample and the tip. After this occurrence the movement direction is inverted and the cantilever starts moving away from the sample surface. The behaviour of the cantilever during deflection should equate to the approach but due to adhesion between the tip and the sample the cantilever starts to deflect negatively (region IV) until the adhesion force is overcome by the cantilever restoring force, and the contact breaks.

AFM is one of the few techniques which can measure a force-distance curve for single molecule rupture events. This can allow mechanical properties such as individual molecule flexibility and stiffness to be quantified. Also, detailed investigation of protein and peptide folding and unfolding events may be mapped as well as protein-protein interactions and protein-ligand binding (Allison, D. P. et al., 2002 and Leckband, D. et al., 2000). However, a limitation of force-distance measurements for the application of dynamic study of samples is that due to the intrinsic complexities of the tip-sample interaction the rate of data acquisition is slow and their data is 'rich' in nature.
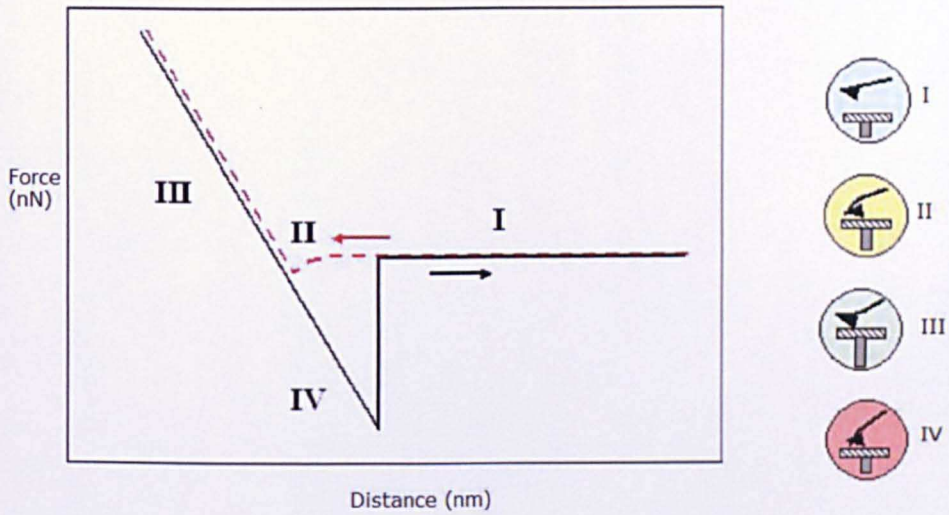
Figure 17   Force-Distance curve generated from AFM experiments. Each region I-IV is explained above in further detail. The red dotted line is the cantilever approach and the black thick line is the cantilever deflection during the retract phase of the measurement.

During tapping mode imaging, phase images that can provide qualitative information on the adhesive properties and viscoelasticity of a sample, can be simultaneously acquired. PF-QNM is an extension to the tapping mode where the oscillation is at a lower frequency which provides images at a better resolution. The main advantage of QNM is that quantitative force data can be obtained (as explained further below).

PF-QNM imaging mode AFM is a new force mapping technique (Minne, S. C. et al., 2010; Morsi, S. M. et al., 2010; Cranston, E. D. et al., 2011 and Adamick, J. et al., 2011) which can provide multiple force maps, including images of Young's modulus, deformation, adhesion and dissipation, simultaneously with the normal topographic image and with the same pixel resolution. It is advantageous in biological applications

as it works in both air and liquid environments. Another benefit to the peak force AFM apparatus is that the cantilever does not resonate (i.e. cantilever tuning is not required). Therefore, this technique becomes very useful in fluids.

SCANASYST-FLUID+ tips purchased from Bruker, UK, were used in the AFM experiments described within this thesis, which were mounted on a MMTMEC (MultiMode Tapping Mode AFM Electrochemical fluid cell) for operation in liquid buffers. The SCANASYST-FLUID+ tips are made from Si_3N_4 and are very sharp (vertical height 2.5-8 µm and apex radius 2-12 nm) so it can form high resolution images.

A multimode 8 scanning probe station with a nanoscope V controller (BruckerNano, CA) was used to obtain PF-QNM images within this thesis. NanoScope v8.1 software was used during the image acquisition, whilst NanoScope Analysis v1.20 software was used for offline data analysis in order to obtain quantitative results. These results include height and width dimensions Youngs modulous and is more commonly known as *grain analysis*.

Chapter 4 describes the application of PF-QNM AFM to examine surface topography and stiffness (Young's modulus and deformation) of biological molecules and in particular the effects of ligand binding.

2.4   References

Adamcik, J., Berquand, A. and Mezzenga, R. (2011) 'Single-step direct measurement of amyloid fibrils stiffness by peak force quantitative nanomechanical atomic force microscopy', Applied Physics Letters, 98(19), 193701-3.

Adams, J. M. (1983) 'Hydrogen-atom positions in kaolinite by neutron profile refinement', Clays and Clay Minerals, 31(5), 352-356.

Alder, B. J. and Wainwright, T. E. (1957) 'Phase transition for a hard sphere system', Journal of Chemical Physics, 27(5), 1208-1209.

Alder, B. J. and Wainwright, T. E. (1959) 'Studies in molecular dynamics .1. general method', Journal of Chemical Physics, 31(2), 459-466.

Alessandrini, A. and Facci, P. (2005) 'AFM: a versatile tool in biophysics', Measurement Science & Technology, 16(6), R65-R92.

Allen, M. P. and Imbierski, A. A. (1987) 'A molecular-dynamics study of the hard dumbbell system', Molecular Physics, 60(2).

Allen, T. W., Andersen, O. S. and Roux, B. (2004) 'On the importance of atomic fluctuations, protein flexibility, and solvent in ion permeation', Journal of General Physiology, 124(6), 679-690.

Allison, D. P., Hinterdorfer, P. and Han, W. H. (2002) 'Biomolecular force measurements and the atomic force microscope', Current Opinion in Biotechnology, 13(1), 47-51.

Amadei, A., Linssen, A. B. M. and Berendsen, H. J. C. (1993) 'ESSENTIAL DYNAMICS OF PROTEINS', Proteins-Structure Function and Genetics, 17(4).

Anand, G., Zhang, F., Linhardt, R. J. and Belfort, G. (2011) 'Protein-Associated Water and Secondary Structure Effect Removal of Blood Proteins from Metallic Substrates', Langmuir, 27(5), 1830-1836.

Andersen, H. C. (1983) 'Rattle - a velocity version of the shake algorithm for molecular-dynamics calculations', Journal of Computational Physics, 52(1), 24-34.

Bernard, C., Draper, T., Hockney, G. and Soni, A. (1988) 'LATTICE CALCULATION OF WEAK AMPLITUDES OF D-MESONS AND B-MESONS', Physical Review D, 38(11).

Bingen, P., Wang, G., Steinmetz, N. F., Rodahl, M. and Richter, R. P. (2008) 'Solvation Effects in the Quartz Crystal Microbalance with Dissipation Monitoring Response to Biomolecular Adsorption. A Phenomenological Approach', Analytical Chemistry, 80(23), 8880-8890.

Binnig, G., Fuchs, H. and Stoll, E. (1986) 'Surface-diffusion of oxygen-atoms individually observed by stm', Surface Science, 169(2-3), L295-L300.

Cheatham, T. E. and Kollman, P. A. (2000) 'Molecular dynamics simulation of nucleic acids', Annual Review of Physical Chemistry, 51, 435-471.

Cooper, A. (1976) 'Thermodynamic fluctuations in protein molecules', Proceedings of the National Academy of Sciences of the United States of America, 73(8), 2740-2741.

Cranston, E. D., Eita, M., Johansson, E., Netrval, J., Salajkoval⊐, M., Arwin, H. and WalŠgberg, L. (2011) 'Determination of Youngs Modulus for Nanofibrillated Cellulose Multilayer Thin Films Using Buckling Mechanics', Biomacromolecules, 12(4), 961-969.

Darden, T., York, D. and Pedersen, L. (1993) 'Particle mesh ewald - an n.log(n) method for ewald sums in large systems', Journal of Chemical Physics, 98(12), 10089-10092.

Deserno, M. and Holm, C. (1998) 'How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines', Journal of Chemical Physics, 109(18), 7678-7693.

Diamond, R. (1976) 'Comparison of conformations using linear and quadratic transformations', Acta Crystallographica Section A, 32(JAN1), 1-10.

Diamond, R. (1988) 'A note on the rotational superposition problem', Acta Crystallographica Section A, 44, 211-216.

Duan, Y. and Kollman, P. A. (1998) 'Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution', Science, 282(5389), 740-744.

Edvardsson, M., Rodahl, M., Kasemo, B. and Hook, F. (2005) 'A dual-frequency QCM-D setup operating at elevated oscillation amplitudes', Analytical Chemistry, 77(15), 4918-4926.

Ensign, D. L., Kasson, P. M. and Pande, V. S. (2007) 'Heterogeneity even at the speed limit of folding: Large-scale molecular dynamics study of a fast-folding variant of the villin headpiece', Journal of Molecular Biology, 374(3), 806-816.

Ewald, P. P. (1921) '- Die Berechnung optischer und elektrostatischer Gitterpotentiale', - 369(- 3), - 287.

Exner, O. and Bohm, S. (2003) 'Electrostatic calculation of the substituent effect: An efficient test on isolated molecules', Chemistry-a European Journal, 9(19), 4718-4723.

Ferguson, D. M., Pearlman, D. A., Swope, W. C. and Kollman, P. A. (1992) 'Free-energy perturbation calculations involving potential function changes', Journal of Computational Chemistry, 13(3), 362-370.

Frauenfelder, H., Petsko, G. A. and Tsernoglou, D. (1979) 'Temperature-dependent x-ray-diffraction as a probe of protein structural dynamics', Nature, 280(5723), 558-563.

Ganesan, N., Bauer, B. A., Lucas, T. R., Patel, S. and Taufer, M. (2011) 'Structural, Dynamic, and Electrostatic Properties of Fully Hydrated DMPC Bilayers From Molecular Dynamics Simulations Accelerated with Graphical Processing Units (GPUs)', Journal of Computational Chemistry, 32(14), 2958-2973.

Garcia, A. E. (1992) 'Large-amplitude nonlinear motions in proteins', Physical Review Letters, 68(17).

Gibson, C. T., Watson, G. S. and Myhra, S. (1997) 'Lateral force microscopy - a quantitative approach', Wear, 213(1-2), 72-79.

Gohlke, H. and Klebe, G. (2002) 'Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors', Angewandte Chemie-International Edition, 41(15), 2645-2676.

Hagler, A. T. (1977) 'Relation between spatial electron-density and conformational properties of molecular systems', Israel Journal of Chemistry, 16(2-3), 202-212.

Hansma, P. K., Cleveland, J. P., Radmacher, M., Walters, D. A., Hillner, P. E., Bezanilla, M., Fritz, M., Vie, D., Hansma, H. G., Prater, C. B., Massie, J., Fukunaga, L., Gurley, J. and Elings, V. (1994) 'TAPPING MODE ATOMIC-FORCE MICROSCOPY IN LIQUIDS', Applied Physics Letters, 64(13), 1738-1740.

Hess, B. (2000) 'Similarities between principal components of protein dynamics and random diffusion', Physical Review E, 62(6), 8438-8448.

Ho, B. K. and Brasseur, R. (2005) 'The Ramachandran plots of glycine and pre-proline', Bmc Structural Biology, 5.

Hockney, R. W. (1982) 'Characterization of parallel computers and algorithms', Computer Physics Communications, 26(3-4), 285-291.

Hook, F., Kasemo, B., Nylander, T., Fant, C., Sott, K. and Elwing, H. (2001) 'Variations in coupled water, viscoelastic properties, and film thickness of a Mefp-1 protein film during adsorption and cross-linking: A quartz crystal microbalance with dissipation monitoring, ellipsometry, and surface plasmon resonance study', Analytical Chemistry, 73(24), 5796-5804.

Hook, F., Rodahl, M., Brzezinski, P. and Kasemo, B. (1998) 'Energy dissipation kinetics for protein and antibody-antigen adsorption under shear oscillation on a quartz crystal microbalance', Langmuir, 14(4), 729-734.

Hook, F., Voros, J., Rodahl, M., Kurrat, R., Boni, P., Ramsden, J. J., Textor, M., Spencer, N. D., Tengvall, P., Gold, J. and Kasemo, B. (2002) 'A comparative study of protein adsorption on titanium oxide surfaces using in situ ellipsometry, optical waveguide lightmode spectroscopy, and quartz crystal microbalance/dissipation', Colloids and Surfaces B-Biointerfaces, 24(2), 155-170.

Huang, D. and Caflisch, A. (2004) 'Efficient evaluation of binding free energy using continuum electrostatics solvation', Journal of Medicinal Chemistry, 47(23).

Humphrey, W., Dalke, A. and Schulten, K. (1996) 'VMD: Visual molecular dynamics', Journal of Molecular Graphics & Modelling, 14(1), 33-38.

Hutter, J. L. and Bechhoefer, J. (1993) 'CALIBRATION OF ATOMIC-FORCE MICROSCOPE TIPS', Review of Scientific Instruments, 64(7), 1868-1873.

Jalili, N. and Laxminarayana, K. (2004) 'A review of atomic force microscopy imaging systems: application to molecular metrology and biological sciences', Mechatronics, 14(8), 907-945.

Jorgensen, W. L. and Tiradorives, J. (1988) 'The opls potential functions for proteins - energy minimizations for crystals of cyclic-peptides and crambin', Journal of the American Chemical Society, 110(6), 1657-1666.

Kabsch, W. (1976) 'Solution for best rotation to relate 2 sets of vectors', Acta Crystallographica Section A, 32(SEP1), 922-923.

Kabsch, W. (1978) 'Discussion of solution for best rotation to relate 2 sets of vectors', Acta Crystallographica Section A, 34(SEP).

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

Karplus, M. and McCammon, J. A. (1981) 'The internal dynamics of globular-proteins', Crc Critical Reviews in Biochemistry, 9(4), 293-349.

Keller, C. A. and Kasemo, B. (1998) 'Surface specific kinetics of lipid vesicle adsorption measured with a quartz crystal microbalance', Biophysical Journal, 75(3), 1397-1402.

Kitao, A., Hirata, F. and Go, N. (1991) 'The effects of solvent on the conformation and the collective motions of protein - normal mode analysis and molecular-dynamics simulations of melittin in water and in vacuum', Chemical Physics, 158(2-3).

Laskowski, R. A., Macarthur, M. W., Moss, D. S. and Thornton, J. M. (1993) 'Procheck - a program to check the stereochemical quality of protein structures', Journal of Applied Crystallography, 26, 283-291.

Leach, A. R. (2001) Molecular modeling : principles and applications, 2nd ed., Harlow: Prentice Hall.

Leckband, D. (2000) 'Measuring the forces that control protein interactions' in Annual Review of Biophysics and Biomolecular

Structure, Annual Reviews {a} , 4139 El Camino Way, Palo Alto, CA, 94303-0139, USA, 1-26.

Lifson, S. and Warshel, A. (1968) 'Consistent force field for calculations of conformations vibrational spectra and enthalpies of cycloalkane and n-alkane molecules', Journal of Chemical Physics, 49(11), 5116-&.

Liu, H. Y., Elstner, M., Kaxiras, E., Frauenheim, T., Hermans, J. and Yang, W. T. (2001) 'Quantum mechanics simulation of protein dynamics on long timescale', Proteins-Structure Function and Genetics, 44(4), 484-489.

London, F. and Polanyi, M. (1930) 'About the atomic theory analysis of the adsorption forces', Naturwissenschaften, 18, 1099-1100.

Luty, B. A., Davis, M. E., Tironi, I. G. and Vangunsteren, W. F. (1994) 'A comparison of particle-particle, particle-mesh and ewald methods for calculating electrostatic interactions in periodic molecular-systems', Molecular Simulation, 14(1), 11-20.

Luty, B. A., Tironi, I. G. and Vangunsteren, W. F. (1995) 'Lattice-sum methods for calculating electrostatic interactions in molecular simulations', Journal of Chemical Physics, 103(8), 3014-3021.

Makri, N., Sim, E. J., Makarov, D. E. and Topaler, M. (1996) 'Long-time quantum simulation of the primary charge separation in bacterial photosynthesis', Proceedings of the National Academy of Sciences of the United States of America, 93(9), 3926-3931.

Manne, S., Cleveland, J. P., Stucky, G. D. and Hansma, P. K. (1993) 'Lattice resolution and solution kinetics on surfaces of amino-acid crystals - an atomic force microscope study', Journal of Crystal Growth, 130(1-2), 333-340.

McCammon, J. A., Gelin, B. R. and Karplus, M. (1977) 'Dynamics of folded proteins', Nature, 267(5612), 585-590.

Minne, S. C., Hu, Y., Hu, S., Pittenger, B. and Su, C. (2010) 'NanoScale Quantitative Mechanical Property Mapping Using Peak Force Tapping Atomic Force Microscopy', Microscopy and Microanalysis, 16(Supplement S2), 464-465.

Morsi, S. M., Pakzad, A., Amin, A., Yassar, R. S. and Heiden, P. A. (2011) 'Chemical and nanomechanical analysis of rice husk modified by ATRP-grafted oligomer', Journal of Colloid and Interface Science, 360(2), 377-385.

Newton, I., Motte, A. and Machin, J. (1729) The Mathematical Principles of Natural Philosophy, B. Motte.

Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., III, Laughton, C. A. and Orozco, M. (2007) 'Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers', Biophysical Journal, 92(11), 3817-3829.

Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. (1963) 'Stereochemistry of polypeptide chain configurations', Journal of Molecular Biology, 7(1), 95-&.

Ramakris.C and Ramachan.Gn (1965) 'Stereochemical criteria for polypeptide and protein chain conformations .2. allowed conformations for a pair of peptide units', Biophysical Journal, 5(6), 909-&.

Rodahl, M., Hook, F., Fredriksson, C., Keller, C. A., Krozer, A., Brzezinski, P., Voinova, M. and Kasemo, B. (1997) 'Simultaneous frequency and dissipation factor QCM measurements of biomolecular adsorption and cell adhesion', Faraday Discussions, 107, 229-246.

Rodahl, M., Hook, F. and Kasemo, B. (1996) 'QCM operation in liquids: An explanation of measured variations in frequency and Q factor with liquid conductivity', Analytical Chemistry, 68(13).

Rodahl, M., Hook, F., Krozer, A., Brzezinski, P. and Kasemo, B. (1995) 'Quartz-crystal microbalance setup for frequency and q-factor

measurements in gaseous and liquid environments', Review of Scientific Instruments, 66(7), 3924-3930.

Rueda, M., Chacon, P. and Orozco, M. (2007) 'Thorough validation of protein normal mode analysis: A comparative study with essential dynamics', Structure, 15(5), 565-575.

Sander, M., Madliger, M. and Schwarzenbach, R. P. (2010) 'Adsorption of Transgenic Insecticidal Cry1Ab Protein to SiO(2). 1. Forces Driving Adsorption', Environmental Science & Technology, 44(23), 8870-8876.

Sauerbrey, G. (1959) 'Verwendung von schwingquarzen zur wagung dunner schichten und zur mikrowagung', Zeitschrift Fur Physik, 155(2), 206-222.

Saul, F. A., Amzel, L. M. and Poljak, R. J. (1978) 'Preliminary refinement and structural-analysis of fab fragment from human immunoglobulin new at 2.0 a resolution', Journal of Biological Chemistry, 253(2), 585-597.

Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., Young, C., Batson, B., Bowers, K. J., Chao, J. C., Eastwood, M. P., Gagliardo, J., Grossman, J. P., Ho, C. R., Ierardi, D. J., Kolossvary, I., Klepeis, J. L., Layman, T., McLeavey, C., Moraes, M. A., Mueller, R., Priest, E. C., Shan, Y., Spengler, J., Theobald, M.,

Towles, B., Wang, S. C. and Acm (2007) Anton, a Special-Purpose Machine for Molecular Dynamics Simulation, translated by San Diego, CA: 1-12.

Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y. and Wriggers, W. (2010) 'Atomic-Level Characterization of the Structural Dynamics of Proteins', Science, 330(6002), 341-346.

Shields, G. C., Laughton, C. A. and Orozco, M. (1998) 'Molecular dynamics simulation of a PNA center dot DNA center dot PNA triple helix in aqueous solution', Journal of the American Chemical Society, 120(24), 5895-5904.

Shoichet, B. K., Leach, A. R. and Kuntz, I. D. (1999) 'Ligand solvation in molecular docking', Proteins-Structure Function and Genetics, 34(1).

Syme, N. R., Dennis, C., Phillips, S. E. V. and Homans, S. W. (2007) 'Origin of heat capacity changes in a "Nonclassical" hydrophobic interaction', Chembiochem, 8(13), 1509-1511.

Thorpe, M. F. and Beeman, D. (1976) 'Thermodynamics of an ising-model with random exchange interactions', Physical Review B, 14(1), 188-199.

Verlet, L. (1967) 'Computer experiments on classical fluids .i. thermodynamical properties of lennard-jones molecules', Physical Review, 159(1), 98-&.

von Vacano, B., Xu, R., Hirth, S., Herzenstiel, I., Rueckel, M., Subkowski, T. and Baus, U. (2011) 'Hydrophobin can prevent secondary protein adsorption on hydrophobic substrates without exchange', Analytical and Bioanalytical Chemistry, 400(7), 2031-2040.

Wang, G., Rodahl, M., Edvardsson, M., Svedhem, S., Ohlsson, G., Hoeoek, F. and Kasemo, B. (2008) 'A combined reflectometry and quartz crystal microbalance with dissipation setup for surface interaction studies', Review of Scientific Instruments, 79(7).

Wlodek, S. T., Antosiewicz, J. and McCammon, J. A. (1997) 'Prediction of titration properties of structures of a protein derived from molecular dynamics trajectories', Protein Science, 6(2), 373-382.

Zhong, Q., Inniss, D., Kjoller, K. and Elings, V. B. (1993) 'Fractured polymer silica fiber surface studied by tapping mode atomic-force microscopy', Surface Science, 290(1-2), L688-L692.

Zhou, A. Q., O'Hern, C. S. and Regan, L. (2011) 'Revisiting the Ramachandran plot from a new angle', Protein Science, 20(7), 1166-1171.

Chapter 3 - Hydration of Binding Cavity

3.1   Introduction

Recently there has been a lot of interest in water molecules being present in the binding cavity of biological molecules. This is because water plays an important role in biomolecular association in particular the dewetting phenomenon (as explained in Chapter 1). In this chapter the aim is to detect changes in the number of water molecules in the binding pocket of the protein MUP and its mutants when interacting with the ligand IBM.

3.1.1 Hydration of proteins binding to biological molecule

Water molecules can facilitate molecular recognition in biological molecules by forming extensive hydrogen bond networks in protein complexes with not only other proteins (Reichmann, D. et al., 2008, Yokota, A. et al., 2003, Hayes, J. M. et al., 2011) but also DNA, (Bradshaw, E. M. et al., 2011; Gao, Y. G., et al., 1998; Duan, J. et al., 2002; Sonavane, S., et al., 2009) RNA, (Sonavane, S., et al., 2009; Castrignano, T., et al., 2002) and sugars (Nurisso, A. et al., 2010). Water molecules can form both intramolecular (i.e. waters between two ligand atoms or two protein atoms) and intermolecular bridges (i.e., waters between ligand and protein atoms). Water is also important in the interaction of proteins with small ligands, (de Beer, S. B. A. et al., 2010) and recognition of the hydration status of a protein binding pocket

may be used for drug design (see chapter 1) (Homans, S. A., 2007, Barillari, C. et al., 2007).

It can also be noted that not all water molecules in the protein binding cavities are equivalent. For example, some water molecules in the cavity may show rapid exchange with the bulk, while other water molecules are relatively tightly bound, (Makarov, V. et al., 2002). Some water molecules which are bound to the binding cavity of a protein prior to ligand binding are retained in complexes (i.e. with small molecules-protein (Barillari, C. et al., 2007) and protein-protein interaction (Reichmann, D. et al., 2008)). There have been many computational methods and experimental techniques of predicting solvation of a protein surface. Examples include X-ray data, (Durchschlag, H. et al., 2003, and Pitt, W. R. et al., 1991) molecular dynamics simulations, (Makarov, V. et al., 2002; Virtanen, J. J. et al., 2010; Geroult, S. et al., 2007) and grid-based simulation (Michel, J. et al., 2009).

The replacement of water molecules in the protein binding cavity by a ligand that is complementary to the protein groups lining the cavity (making hydrogen bonds and contacts where appropriate) can make an important contribution to the binding affinity. Recent studies of the periplasmic oligopeptide binding protein (OBP) (Yabuki, M. et al., 2010) have shown that water molecule displacement from the cavity with the ligand IBM decreases the binding affinity. Other OBP studies binding to di- (Ladbury, J. E. et al., 1996) tri- and tetra-peptides have also shown that the displacement of water molecules decreases the binding affinity (Tame, J. R. H. et al., 1994; Sleigh, S. H. et al., 1997). In another study

the displacement of ordered water molecules by cyclic urea inhibitors has led to a large gain in entropy (Jiao, M. et al., 2003).

The hydrophobic interaction is an important driving force in molecular recognition (as explained in Chapter 1), yet our understanding of hydrophobicity in protein binding sites, remains incomplete and may stand as an obstacle to rational drug design. Therefore, a better understanding of the hydration of the protein binding cavity in protein-ligand interactions is essential to further progress in the area.

3.2   General Methods

3.2.1   Molecular dynamics simulations

In this thesis the aim is to simulate WT and mutant forms of MUP and MUP-IBM complexes using the Amber 10 (Case, D. et al., 2008) MD package. Compared to previous work, the aim was to improve quality and reliability by a) making the simulations much longer and b) performing them in triplicate.

Previous simulations of MUP and MUP-IBM revealed very little information on the dynamics of ligand binding to MUP. For example a 30 ns simulation was performed on MUP and its complex with 2-sec-butyl-4,5-dihydrothioazole (SBT) (Macek, P. et al., 2007). Their results have shown that ligand binding can increase the mobility of many residues in the protein. On the other hand, evidence from NMR experiments has shown that on a much longer timescale (milliseconds), ligand binding appears to reduce MUP's mobility (Perazzolo, C. et al.,

2005). Therefore to perform an MD simulation on the microsecond timescale can help to further understand the ligand binding features to proteins such as MUP.

When simulations of nucleic acids became greater than 50-100 ns timescale, it became apparent that there was a problem with the AMBER forcefield, which until that time had been regarded as very accurate for nucleic acid simulations. Unusual and irreversible alterations in certain backbone torsion angles led to the simulations producing structures that were clearly erroneous. The result was a large community effort to identify and correct the problem, which was very successful (Perez, A. et al., 2007). Therefore the simulations performed in this thesis originally, was a stringent test of the Amber FF03 forcefield for proteins. At the start of the simulations it was unclear whether the simulations would result in partial denaturing of the protein reaching conformations that are clearly at odds with experimental data. Therefore to carry on with the recent development in this thesis all simulations are done in the microsecond timescale and the results are compared to the results of WT and mutant forms of MUP and MUP-IBM complexes. The hydration of the binding cavity of each complex was also analysed.

3.2.2 System setup

3.2.2.1   Formation of Replicates

The two starting structures were taken from the crystal structure of apo-MUP, PDB entry 2ozq (Syme, N. R. et al., 2007) and of MUP-IBM PDB entry 1qy1 (Bingham, R. et al., 2004). The ionisation states (Grafton, A.

K. et al., 1999) of the amino acid side chains of MUP (apo and holo) were assigned using the web-based server WHATIF tools http://swift.cmbi.ru.nl/servers/html/index.html). The web-based server detects and assigns the different protonation states of Histadine in a PDB file. The Histadines were then manually edited in the original PDB file.

The two structures were randomly turned around in space (along the different x, y and z axis) and the co-ordinates saved, to obtain a new replicate of each structure. This was done twice to obtain three replicates of each structure. Using replicates permits a better level of statistical analysis than is possible on data from a single simulation. The formation of replicates can be observed in *Figure 18*.

This was repeated for Y120F apo (PDB 1yp6 (Barrett, E. et al., 2005)) and holo (PDB 1yp7 (Barrett, E. et al., 2005)) and A103S apo and holo (crystal structure co-ordinates obtained from the Homans group at the University of Leeds).
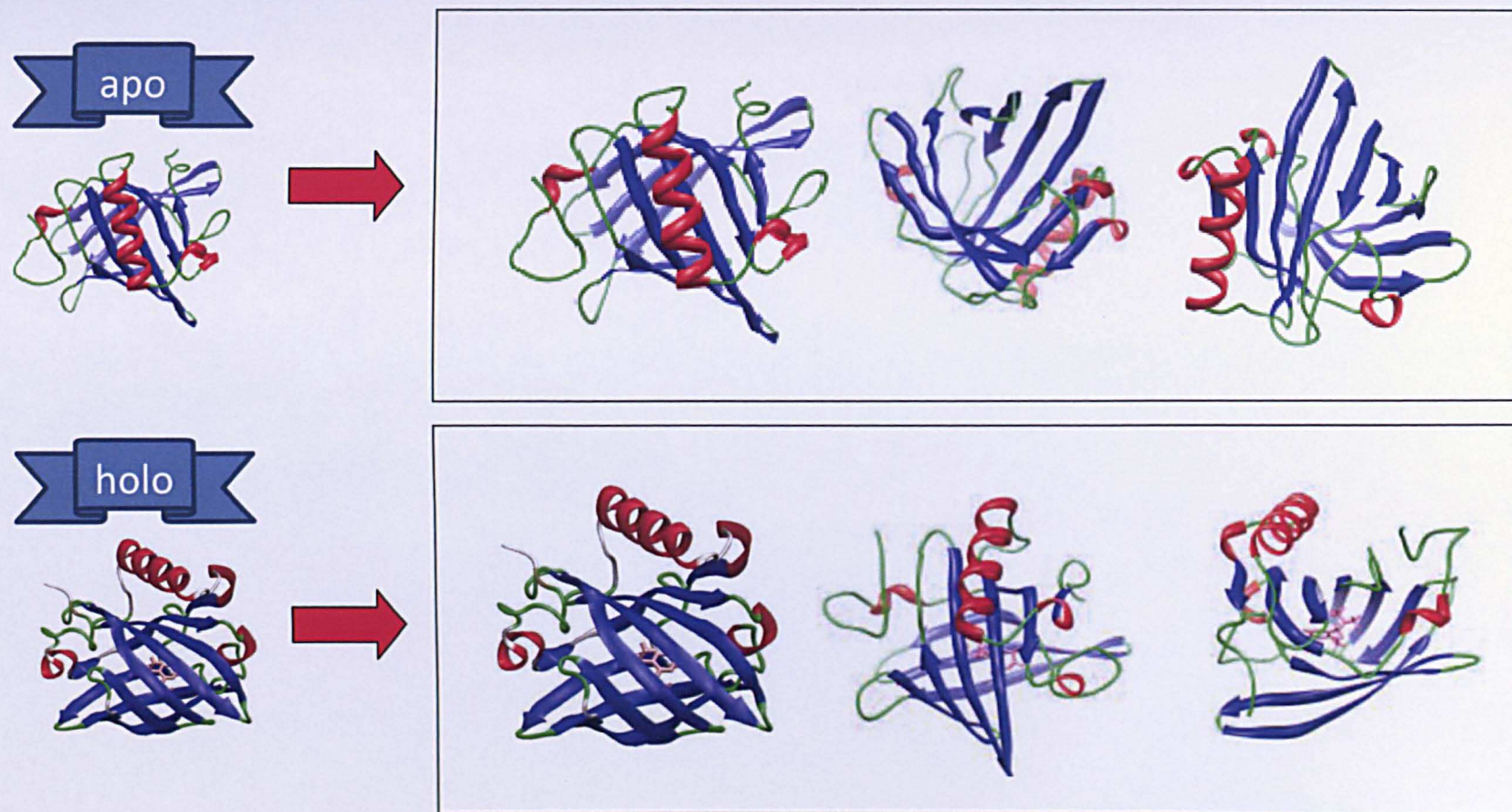
Figure 18   Formation of replicates. On the top left is the WT MUP starting configuration and on the top right are the 3 replicates formed from the starting structure. The bottom left is the MUP-IBM complex starting configuration and the box on the bottom right are the formation of the three replicates of the WT MUP-IBM complex.

3.2.2.2   Materials and Methods

The protein parameters were taken from the Amber ff03 forcefield
(Duan, J. X. et al., 2003). The ligand parameters were generated for
each replicate by using the antechamber module of amber 9 (Case, D.
et al., 2006) using the general amber forcefield (Wang, G. et al., 2008).
Each replicate was immersed in a truncated octahedral box containing
an additional ~19000-22000 TIP3P explicit water molecules and ~13
Na^+ ions to neutralize the system.

Prior to main MD runs, each replicate was subjected to energy
minimisation and a restrained-dynamics, where each replicate was
heated from 10K to 300K, over 100ps and harmonic restraints applied
(Shields, G. C. et al., 1998). Analysis was performed on data generated
through an additional 1.2 μs of unrestrained MD simulation at constant
temperature and pressure (T = 300 K; P = 1 atm) using PMEMD
module of Amber 9. SHAKE was used to constrain all bonds to
hydrogen atoms at equilibrium values, which allowed us to use a 2 fs
MD time step.

Periodic boundary conditions and particle mesh Ewald method were
used to model long range electrostatic effects. Co-ordinates and
velocities were saved every 1 ps for WT MUP and WT MUP-IBM
complexes and every 10 ps for the mutant MUPs (apo and holo forms).
The MD simulation generated an isothermal-isobaric ensemble. In this
ensemble the number of particles generated is fixed and the
temperature and pressure are maintained by a thermostat and barostat

respectively. Post trajectory analysis (RMSD and radial distribution functions) was completed using the ptraj module of amber 10.

For example the hydration density maps were formed by calculating an average structure in the form of a PDB file from the ptraj module of Amber 10 of apo WT, Y120F and A103S MUPs. The maps are than generated with a graphical program called Chimera (Pettersen, E. F. et al., 2004). The PCA was analysed using the *pcazip* tools distributed by CCPB (http://www.ccpb.ac.uk/software). A suitable and constant subspace has to be considered for PCA analysis (Rueda, M. et al., 2007). On the other hand quantitative analysis of the duration of the water molecules was performed by defining the binding cavity of MUP by using active site pressurisation (ASP) (Withers, I. M. et al., 2008). The water molecules were counted within a 1.5 Å distance that surrounds the defined binding cavity of MUP by using a FORTRAN programme developed by Dr. Charles Laughton.

The WT MUP and WT MUP-IBM simulations were run on the UK national high performance computing service HECToR. All mutant MUP simulations were done on the University of Nottingham cluster system Magnum. Due to the large number of processor's available for both HECToR and magnum, the first step was to perform some benchmarking. This would identify the most time and cost effective way of using the cluster systems.

3.2.2.3 Benchmarking

In large simulations (such as the work carried out in this thesis) it has become increasingly important to carry out simulations in the most cost effective way. Therefore benchmarking becomes important. The computer cluster system on HECToR (at the time of these studies) and magnum has two processors per node and each processor has four cores. The nodes could be linked together so they could work in parallel with each other. The speed is greater on a single core system than on a double core system for each processor as it would take longer to process information between two cores then on a single core. However, it is less cost - efficient to use a single core processor as there would be another core on the cluster which would remain inactive whilst the simulation progresses.

Figure 19 reveals that as the numbers of processors are increased for the apo WT-MUP simulation, the time taken for 100 ps would not continuously decrease with the number of processors. A 'perfect' parallel computer, doubling the number of processors used for the calculation would halve the time required, so the cost would be constant. But in 'real life' this doesn't happen – so the graph shows that although initially the computation gets faster as the number of processors is increased, there comes a point when using more processors actually slows down the rate of the simulation.
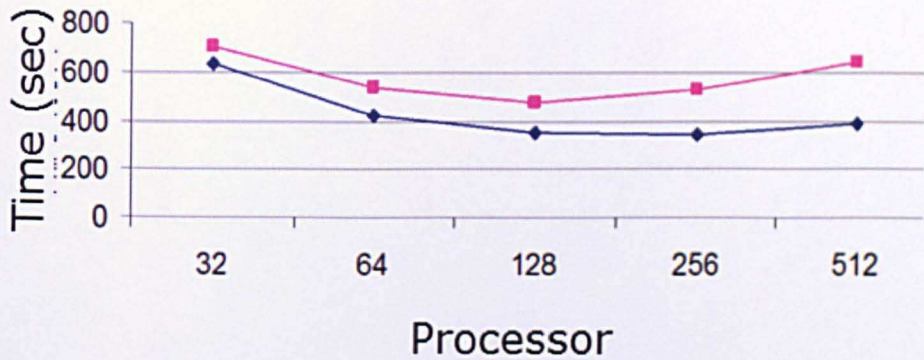
Figure 19 Time taken for each processor to run 100 ps simulation of apo WT MUP on the cluster system named HECToR. The time taken for different number of processors for one (blue line) and two (pink line) nodes are shown above.

The overall conclusion which can be made from *Figure 19* above is that 128 processor on a 1 node system (~16 ns every 24 hours) would seem the most economical and the fastest processor power to use. However, this would only be true if there were no jobs running on HECToR, so therefore no job queues. In reality there are many other people running jobs on HECToR and for maximum throughput it was found that the 32 processors 1 and 2 node would give ~8-10 ns every 24 hours. Therefore not only would this be cheaper but also, the jobs would be submitted more frequently.

On the other hand as magnum has only 240 cores available and there are 12 simulations to run, it would be most cost effective to run on 4 nodes (64 CPU's) which would give approximately 10 ns every 18 hours. For all 18 simulations approximately 80 weeks of real time would be needed and a disk space of approximately 5 Tb would be needed for the trajectories formed during the simulation.

3.2.3 General methods-QCM.

All QCM experiments were carried out on a Q-Sense D300 QCM-D using a standard gold coated quartz sensor crystal (Q-Sense AB, Sweden). The QCM-D used had a resonant fundamental frequency of 4.9 *MHz* and a mass sensitivity of the order of 5 ng cm^{-3} in a liquid environment. Changes in resonant frequency (Δf) and dissipation (ΔD) were measured simultaneously at the third, fifth and seventh overtones. For simplicity the results presented in this thesis show the data from the third overtone only (a full graph from all the replicates obtained at all different overtones are shown in Appendix 6.1). Wild type and mutant MUPs were immobilised by physisorption to the gold substrate by incubating 1 µM solutions of protein (in pH 7.4 buffer) with the sensor surface.

All samples were introduced as a volume of approximately 1.5 ml via an axial flow chamber, which included a T-loop, where the sample was thermally equilibrated at 23 °C for at least two minutes before injection of 0.5 ml into the sample chamber. Once the frequency and disspation signals were stabilized, set of (2-3) PBS injections were subsequently introduced to wash off any loosely attached material and/or multilayers of proteins formed on the surface.

Successive injections of the ligand IBM (at 50 µM) were then introduced to the sample flow chamber to allow binding to take place between the protein and the ligand, and to ensure saturation of the available binding sites. A final set of PBS injections was introduced to allow comparison to the baseline measurements prior to IBM binding

and hence determination of any changes in the properties of the immobilized layer. The entire system was thoroughly cleaned with sodium dodecyl solution (SDS) and cleaned with water (~200 ml) between each experiment. Each experiment was repeated at least 3 times to obtain replicate measurements.

3.3 Results and discussion on MD simulations of MUP binding to IBM

Previous MD simulations on MUP (Barrett, E. et al., 2005) revealed that even in the absence of a ligand, water molecules avoid entering the highly hydrophobic binding cavity, leading to the situation of a partial vacuum within the site. This dewetting process then provides an explanation for the unusual enthalpy-driven nature of the ligand-binding process. Therefore in this part of the chapter the aim is to analyse and interpret the results acquired from the long timescale simulations which have previously been described earlier.

3.3.1 Sampling and convergence in MD

On visual inspection of the simulation trajectories, the first 10 residues appeared to have unfolded from the starting crystal structure (see *Figure 20*) and therefore are not included in any analysis work in this thesis.

(a)   (b)   (c)
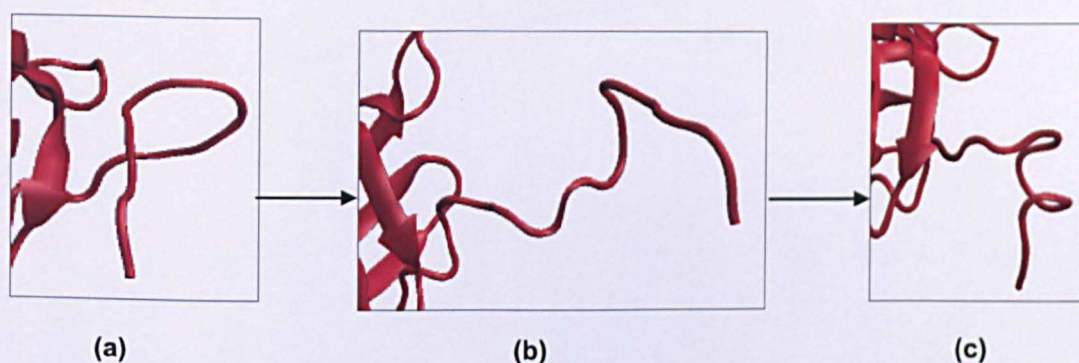
Figure 20   **The structure of the first 10 residues in the crystal structure (a), unfolded between ~530-540 ns (b) and folded to a different state from the starting structure between ~540-1000 ns (c).**

However, the rest of the protein for WT and mutant MUPs (apo and holo) no structural distortions could be observed for the 1.2 μs simulation for all 3 replicates. The RMSD graphs below (*Figure 21*) (with reference to the crystal structure) were calculated from the PTRAJ module of AMBER and can verify the visual inspection observed from the trajectories.
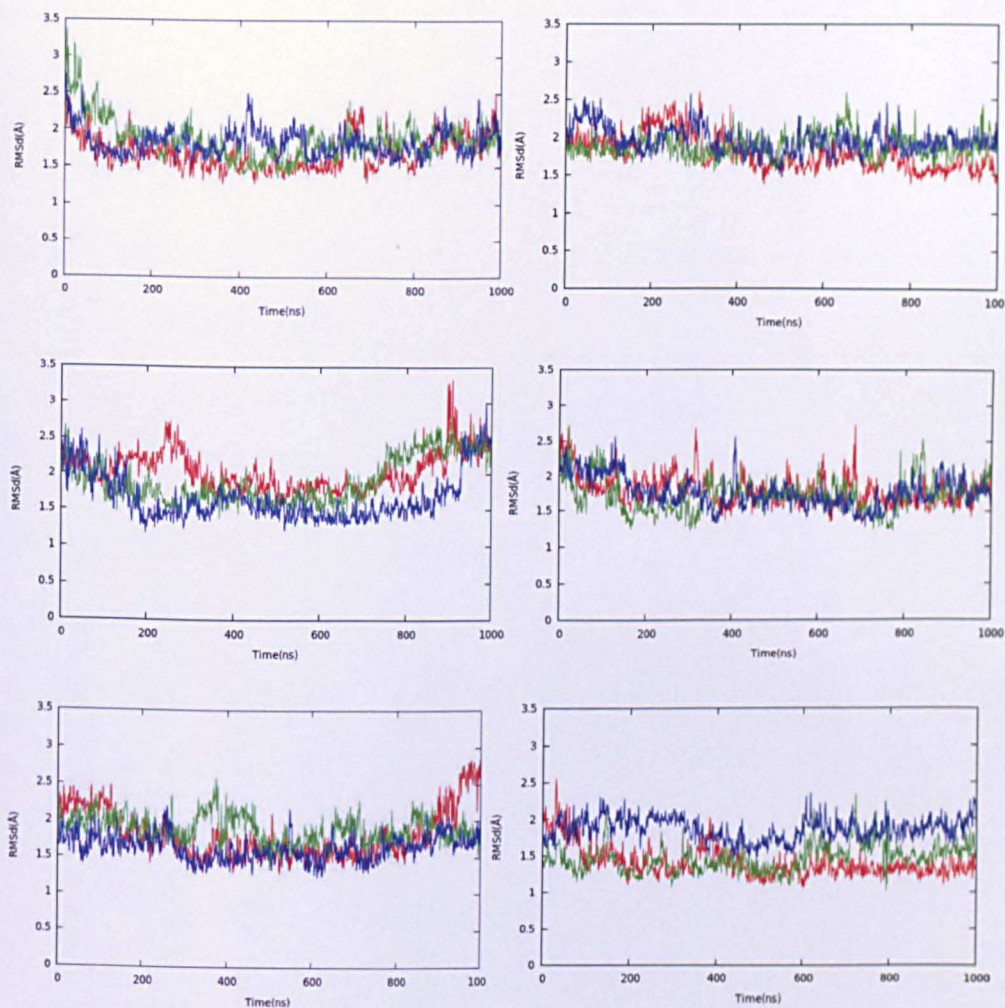
Figure 21 RMSD (with reference to the crystal structure) plots of apo (left) and holo (right) simulations of WT (top), Y120F mutant (middle) and A103S mutant (bottom) MUP with comparison to the time-average structures. The red, blue and green lines on each graph represent replicates 1, 2 and 3 respectively.

All 6 simulations eventually reached equilibrated states. This means that there is a stable and a small difference in RMSDs in each simulation over time. The heavy atom RMSDs were approximately 2.5-3.5 Å from respective crystal structure and 1.5 - 2.5 Å from respective time-average structure (as shown in *Figure 21*). The apo simulations of WT and Y120F MUP can take longer to reach equilibration in

comparison to their respective holo simulations, which appear to reach equilibration much more quickly. However, the holo A103S simulation has shown variation between the 3 replicates. Therefore, to assess this further different analysis techniques have been applied and are discussed further below.

For example a different way of analysing the equilibration and stability is to look at how well the secondary structure has been conserved. The Ramachandran plot is a good approach in analysing the stability of the secondary structure (see Appendix 6.2 for the full set of results). The results revealed that all simulations for WT and mutant MUPs remained very close to the crystal structure distributions, with 83–88% of residues in the allowed regions, and never more than 1% in the disallowed regions (the balance being in the generously allowed region).

Another way of assessing the stability and equilibration of the simulations is to look at the principal components of the simulations of WT and mutant MUPs (see *Figure 22*). As explained in chapter 1, PCA (Meyer A. S. et al., 2006) can reveal information on the mobility of the protein and their convergence in space. *Figure 22* shows the principal component 1 vs principal component 2 (PC1 vs. PC2) of all 6 replicates (apo and holo) of WT MUP, Y120F and A103S mutant MUPs. The graphs represent the differences in convergence for all the replicate simulations. For example the six different simulations (represented by the six colours of *Figure 22a*) have shown that apo and holo forms of WT MUP have a good convergence (i.e all the replicates are in the same subspace). This means that the simulations are similar to each

other. On the other hand, the PCA for Y120F mutant MUP (*Figure 22b*) simulations reveal that the apo and holo simulations are different from each other (all three apo simulations has converged to the same subspace and all three holo simulations have converged to a different subspace). However, the PCA of A103S mutant MUP has revealed that the simulations have a good convergence except for one (holo) replicate.

The differences in the convergence may have occurred for two reasons. The first reason could be that even at the one microsecond timescale, the simulations are still not long enough to reach complete equilibration. The second reason could be due to the stability of the forcefield Amber-FF03. Although, the RMSD graphs in *Figure 21* have shown that with the use of the amber forcefield ff03 the simulations have reached equilibration and the simulations are stable. Therefore, longer simulations maybe needed to provide a better understanding of the reproducibility and replicability of the systems.
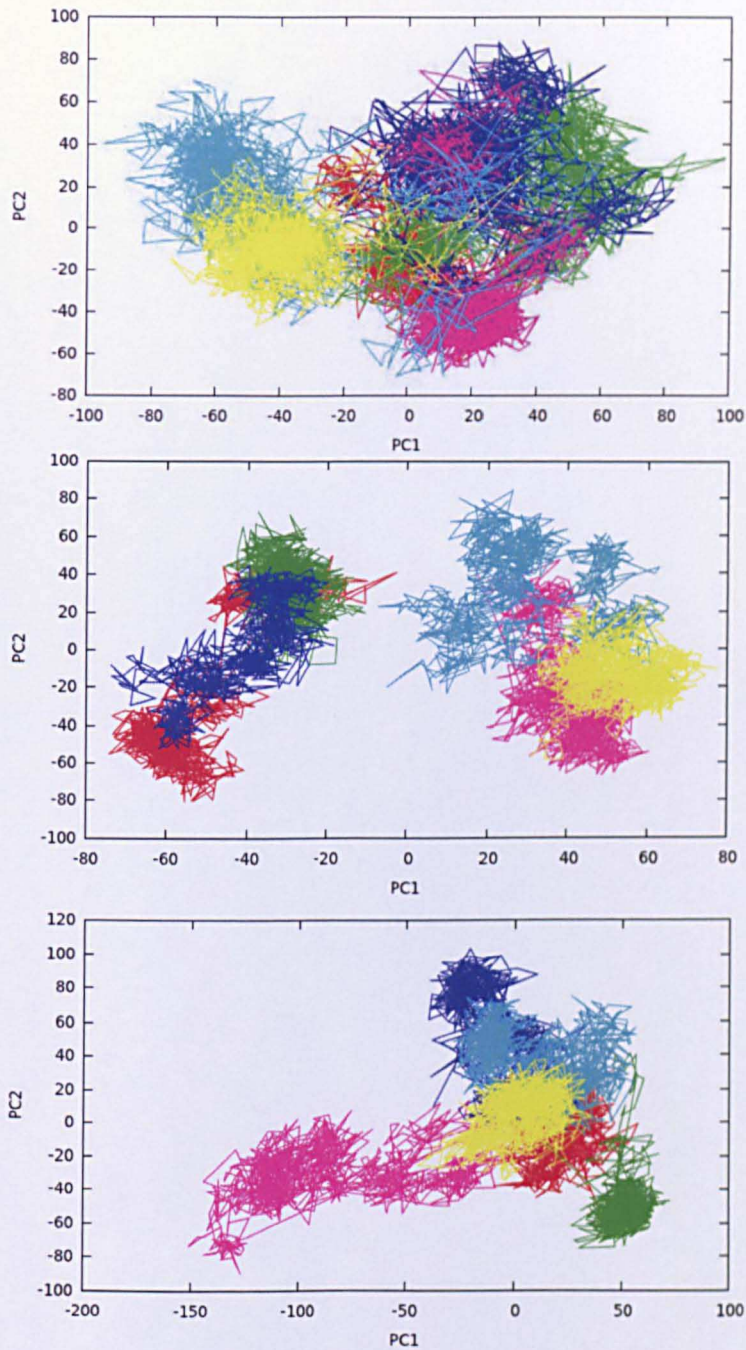
Figure 22   PC1 (nm) vs PC2 (nm) plots of apo and holo WT (top), Y120F (middle) and A103S (bottom) MUP. The six colours represent the six replicates of each simulation (red, blue and green are for apo simulations and light blue, pink and yellow are for holo simulations).

106

3.3.2 Hydration Density Maps

The application of hydration density maps is a way to visualize the binding cavity qualitatively of apo (WT, Y120F and A103S) MUPs. The hydration density maps in *Figure 23* have shown that even without the presence of a ligand there is no water in the binding pocket of Y120F (*Figure 23*b) and little water in the binding pocket of WTMUP. On the other hand the binding pocket of A103S (*Figure 23a*) shows high water density. Closer inspection of the "WT" (*Figure 23a*) hydration density map shows that there is a small water molecule which is tightly bound to the tyrosine in the binding cavity. By contouring at a low level, we can confirm that the cavity still exists and the partial vacuum has not caused the binding cavity to collapse (*Figure 23d*).

The hypothesis of dewetting is attractive in that it helps to explain the experimental data, however there is the caveat that these simulations were performed using the TIP3P water model, which is designed to reproduce the behaviour of bulk water and may possibly give some artifactual behaviour in this rather unusual environment. Nevertheless, further support for this analysis comes from independent modelling studies (Michel, J. et al., 2009), but in the future it would be useful to investigate this system using alternative water models (i.e. TIP4P and TIP5P).

The fact that the binding site is significantly dewetted is nevertheless an important feature. It means that for MUP-IBM complexes, ligand binding is associated with negligible changes in protein-water terms, but

significant changes in ligand-water and ligand-protein terms for the WTMUP and Y120F mutant MUP.

Figure 23   **(a) is the WT results (b) is the "Y120F" results (c) is the "A103S" results and (d) is contoured at a low density to reveal the full extent of the binding pocket. The residue tyrosine 120 is shown in yellow and in (c) residue Serine 103 is shown in magenta.**
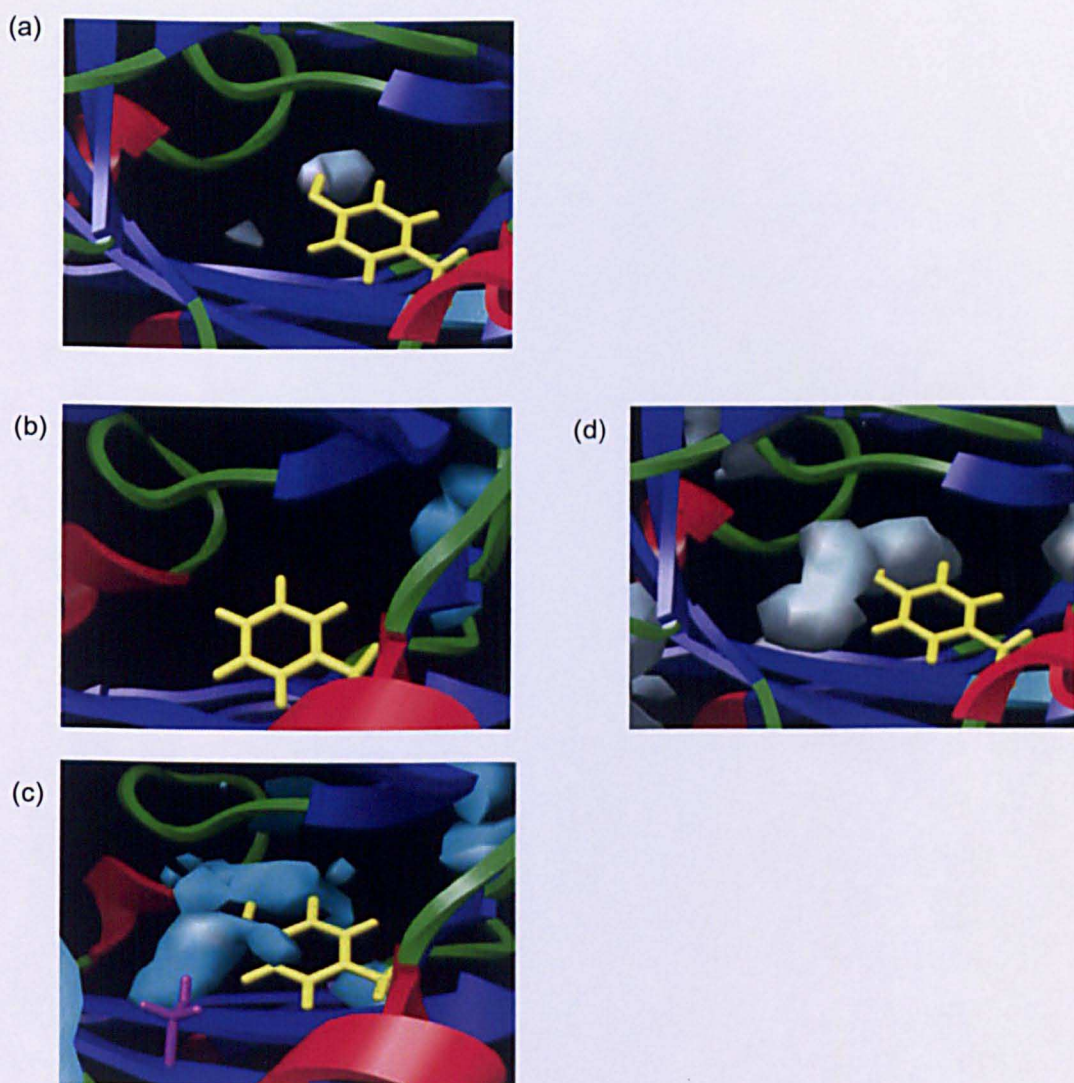
3.3.3 Radial Distribution Function

Another way of analyzing the hydration of the binding site on ligand binding is to calculate the radial distribution functions (RDF). The RDF for water oxygen atoms around the hydroxyl group of tyrosine 120 is shown in *Figure 24*. Due to the buried nature of the binding site, the density scale (y-axis) of these plots has been calibrated by measuring the RDF for the ligand in the binding site of the IBM simulations, which must integrate to one (green lines in *figures 24a, b and c*).

In comparison, the size of the binding cavity being ~9 Å from *Figure 24* it can be observed that the water occupancy of the binding site in the apo WT MUP averages ~0.4. It can also be noted that a small amount of water (occupancy ~0.3) remains close to Tyr-120 in the WT MUP (holo) simulations. This appears to be the result of water molecules transiently coming close to this residue via the slightly porous walls of the binding cavity. The results also suggest that there is less water in binding cavity of Y120F MUP simulations then the WT MUP simulations (i.e. ~0.1 for apo Y120F MUP and ~0.005 for holo Y120F MUP). On the other hand A103S mutant MUP simulations reveal that the binding cavity is more hydrated then the WT MUP and the Y120F MUP simulations (i.e. ~0.6 for apo A103S MUP and ~0.4 for holo A103S MUP). The radial distribution curves show that the apo-protein is more hydrated then the ligand bound complex for all three systems in the binding pocket. These results correspond well with the hydration density maps for all three MUP systems.

Figure 24 (a) is the WT results and (b) is the "Y120F" results and (c) is the "A103S" results. Integrated radial distribution functions for water in the MUP simulations (blue line) compared to the MUP-IBM simulations (red line). The g(r) values have been scaled so that the equivalent function for the ligand (green line) integrates to 1.

(a)
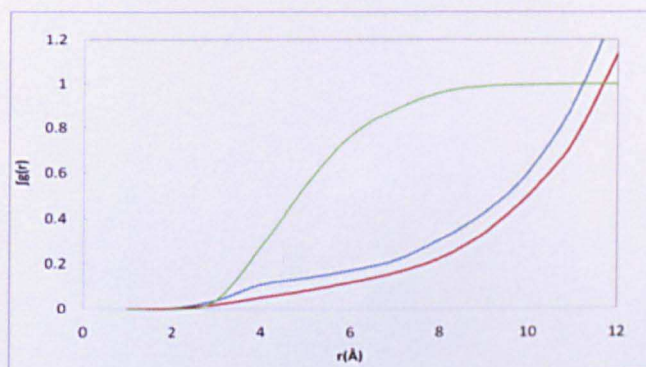
(b)
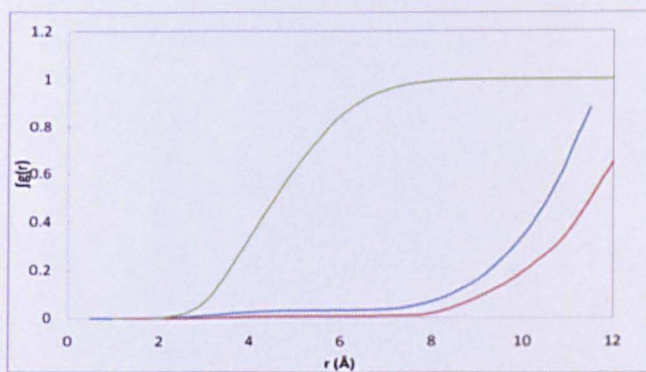
(c)
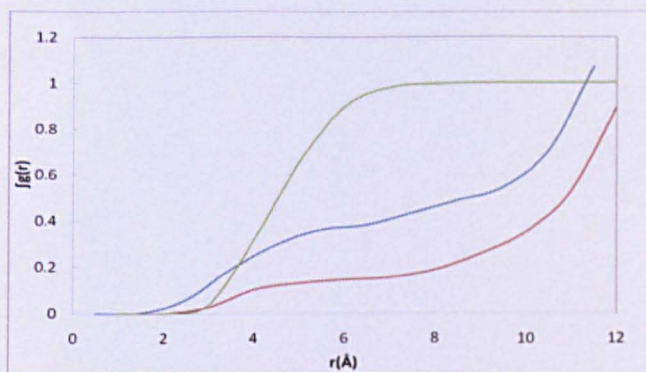
3.3.4 Duration of water molecules in binding cavity

The dewetting phenomenon suggests that even without the presence of the ligand the binding cavity should show low levels of hydration. By visual inspection the MD simulations suggest that in some cases the process is very dynamic - water molecules are constantly entering and are being ejected from the binding cavity. Quatitative analysis was carried out as explained in section 3.2.2.2.

Table 2 illustrates a range of different things (i.e. the number of different waters coming in and out of the binding cavity, the range which the total number of snapshots that water molecule was in the cavity, the number of snapshots that water molecule went from outside the cavity and the average lifetime range (total occ/hops in)) for each replicate of WT, Y120F and A103S MUP without the presence of the ligand. From *Table 2* it is evident that there are no water molecules that are found entering and removed from the cavity of WT MUP. In comparison, the hydrophobic Y120F MUP has a very small number of water molecules with a short average lifetime range of ~1 water molecule. The hydrophilic A103S MUP shows the most exchange in water molecules for replicates 2 and 3.

Table 2   The number of water molecules entering and ejecting from the binding cavity of WT, Y120F and A103S MUP without the presence of the ligand IBM

Molecule	Number of different waters in cavity	Total occ. range	Hops in range	Avg. lifetime range
WTrep1	0	0	0	0
WTrep2	0	0	0	0
WTrep3	0	0	0	0
Y120Frep1	27	1-1	1-1	1-1
Y120Frep2	8	1-1	1-1	1-1
Y120Frep3	0	0	0	0
A103Srep1	9	1-5	1-3	1-2
A103Srep2	45	1-72	1-30	1-4
A103Srep3	23	1-80	1-40	1-12

3.4   Results and Discussion of Experimental QCM-D studies of ligand IBM binding to WT and mutant MUPs

3.4.1 QCM – Control experiments

Prior to the attachment of protein to the QCM-D surface, the ligand IBM was injected into the QCM and over the gold quartz crystal surface (e.g. in the absence of protein) to determine the level of any changes in frequency and dissipation. It was important to see if the presence of IBM produced any unwanted binding of IBM to the bare gold surface

(*Figure 25*). The experiment shown in *Figure 25* shows a resonance frequency shift *(Δf)* of ~3.5 Hz (blue line) and a change in dissipation *(ΔD)* of ~0.3×10^{-6} (red line), indicating that only small changes of these values are associated with the presence of IBM in the media. The spikes in the frequency and dissipation upon IBM injection and rinsing with buffer are due to transient variations in pressure and temperature as new solutions are introduced. Insertion of buffer after exposure of the surface to the ligand causes the signal to return to the original baseline value which indicates that the ligand IBM has been taken off the surface, and therefore did not significantly bind to the gold surface.
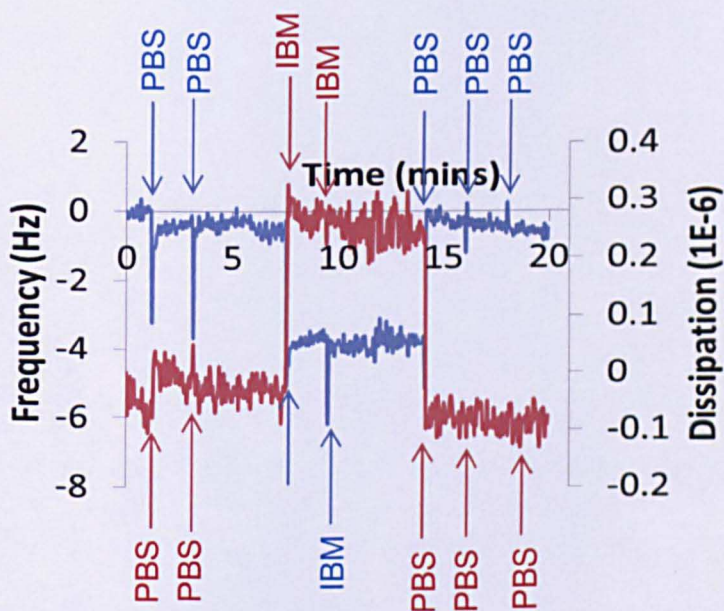
Figure 25   QCM data for the interaction of IBM on gold QCM surface. The frequency (blue line) and the dissipation (red line) have been produced for the injections of IBM (blue arrow) and PBS (red arrows).

To explore the possibly of non-specific binding of the ligand IBM to the protein another experimental set up was explored (*Figure 26*). This

consisted of pre-incubating the WT MUP in the ligand IBM prior to surface adsorption (pre-incubated WT MUP details are given in Chapter 2.3.1.3). Buffer solution was injected first into the sample chamber to acquire a stable baseline. Then the pre-incubated MUP solution was injected at t = 5 mins over the sensor and left for 10 mins, before multiple buffer washes were made to ensure a monolayer of the pre-incubated solution was formed on the surface of the gold quartz crystal. The subsequent exposure of the sensor surface functionalized with the pre-incubated MUP to ligand IBM at t = 28 mins revealed no frequency change. This confirms that once MUP ligand binding sites are occupied (e.g. by pre-incubation), further molecules of IBM are not able to non-specifically bind to/associate with the MUP coated sensor surface.

3.4.2 IBM binding to WT MUP

The Δf (blue trace) obtained upon the adsorption of WT MUP is shown in *Figure 27*. For the trace shown, prior to the adsorption of the protein, several PBS injections are used first to check for a stable baseline (e.g. prior to the introduction of MUP into the experiment), and to remove loosely attached material subsequent to protein adsorption and ligand binding. Comparison of the signals in PBS before and after MUP adsorption, allow calculation of the mass of adsorbed material.

Similarly determination of the frequency values before and after IBM exposure is used to determine specific IBM binding. Several injections of WT MUP in pH 7.4 buffer (PBS) was then added at time, t = 4 minutes to ensure maximal surface coverage. When the frequency

114

response reached a plateau at approximately -105 Hz (t = 38 mins) which suggested that protein binding to the surface had reached a maximum. The surfaces were then washed with successive 0.5 ml PBS injections (labelled as 2nd set on graph) until a plateau was reached and no more material could be washed off the surface.

The ligand IBM was then injected several times at t = 52, 54 and 56 mins until no further ligand could be bound to the protein. The third set of PBS injections at t = 60 and 62 mins indicates that the ligand IBM remains bound to the protein and is not washed off the surface. PBS injections were used to remove unbound material from the surface and analysis chamber. The comparison of the frequency values (in PBS) before and after exposure to IBM suggests that ligand binding has taken place.

A frequency change of -102 Hz has been observed for the addition of WT MUP to the crystal surface. Therefore by using the Saubrey equation, (see section 2.3.1.1 - *Equation 18*) a calculation for the adsorbed layer of WT MUP was estimated to have a mass of 928 ng. Using avogadro's number, surface coverage of the protein on the sensor was calculated to have a molecular footprint of 6.3 nm^2. From molecular modelling studies the size of MUP can be estimated to be 6.4 nm^2. The data has revealed that a full surface coverage has been gained in this experiment.

Figure 26   The frequency (blue line) and dissipation (red line) are formed simultaneously in the qcm experiment of the interaction of pre-incubated WT MUP with IBM. The injection points which are labelled on the graph (PBS, MUP and IBM) shows the injections of the solutions through the sensor loop of the axial chamber. There is no change in the IBM frequency and dissipation signal before and after the injection of PBS which indicates that no further binding can take place.

116
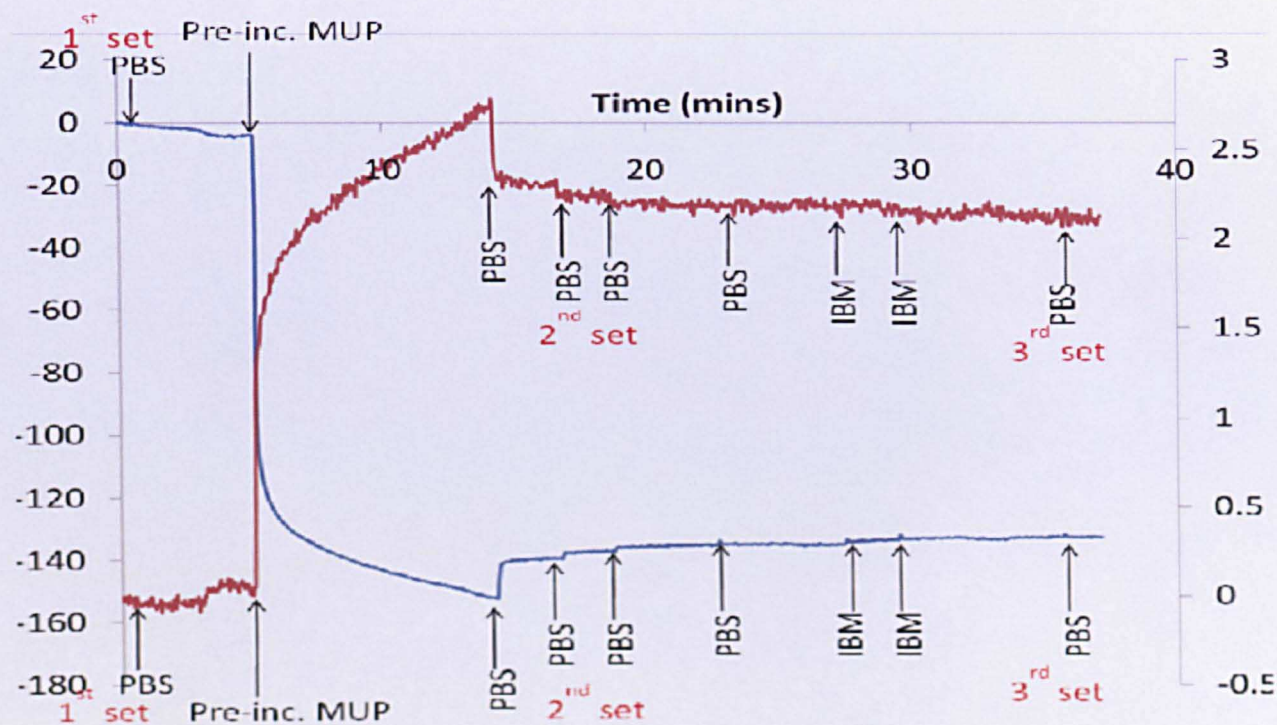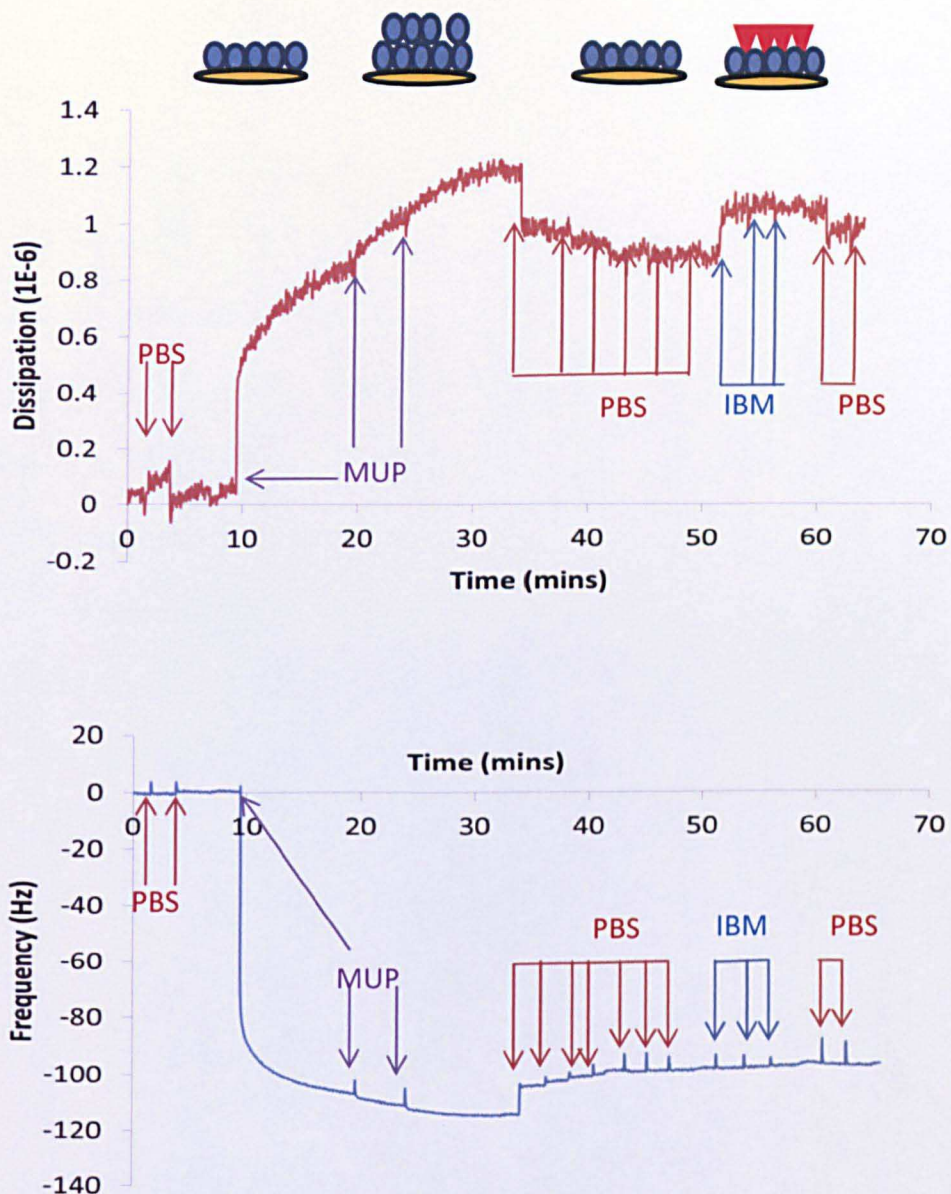
Figure 27 QCM data for the interaction of WT-MUP with IBM. The blue line indicates the frequency shift whereas the red line indicates the dissipation shift. The images at the top represent the interaction of the biological molecules with the quartz crystal with respect to the frequency and dissipation graphs below. The layers of blue circles on the yellow quartz crystal illustrates layer of MUP and the red triangle indicates the binding of the ligand IBM. The red, purple and blue arrows indicate the injection of PBS, MUP and IBM to the axial chamber of the QCM.

3.4.3 Y120F mutant MUP binding to IBM

The QCM-D experiment as described for WT-MUP was then repeated for the hydrophobic mutant Y120F MUP (*Figure 28*). For the data presented in this figure, several injections of Y120F MUP in pH 7.4 buffer (PBS) were added time, t = 5 mins. Following several PBS washes, the ligand IBM was added at t = 105 mins, before performing more PBS washes.

Figure 28 shows a gradual decrease in frequency as the MUP is added to the surface consistent with a mass increase as material is adsorbed to the surface. Using the Saubrey relationship, the adsorbed layer of Y120F MUP was estimated to have a mass of 1135 ng (from a frequency drop of -124.9 Hz). Again, using Avogadro's number, surface coverage of the protein on the sensor was calculated to have a molecular footprint of 4.7 nm^2.

Figure 28 **QCM data for the interaction of Y120F-MUP with IBM. The blue line represents the frequency signal at the 3rd overtone. The red, purple and green arrows indicate the addition of PBS, MUP and IBM respectively.**
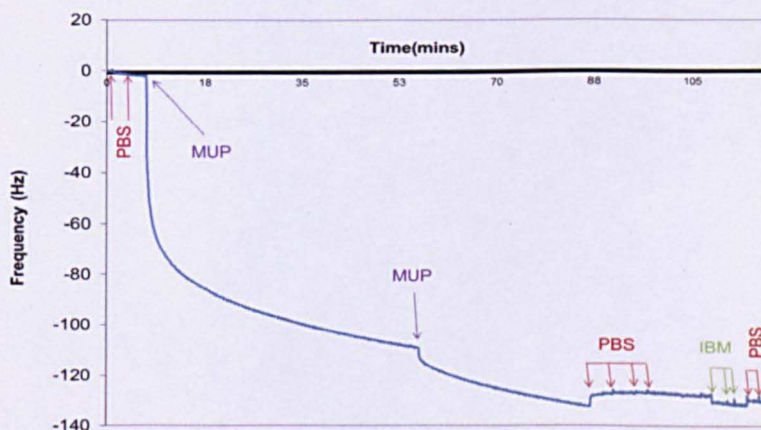
3.4.4 A103S mutant MUP binding to IBM

The same experiment was repeated for hydrophilic A103S mutant MUP. For the data shown in *Figure 29*, the mutant A103S MUP was injected at t = 8 mins and again at t = 19 mins followed by several PBS washes to ensure the protein was bound to the quartz crystal surface. Successive injections of the ligand IBM were added at t = 40 mins until no further ligand could be bound to the protein. The third set of PBS injections at t = 48 mins indicates that the ligand IBM bound to the protein and was not washed off the surface. Again using the Saubrey relationship, the adsorbed layer of A103S MUP was estimated to have a mass of 796 ng. Using Avogadro's number, the surface coverage of the protein on the sensor was calculated to have a molecular footprint of 4.76 nm^2.

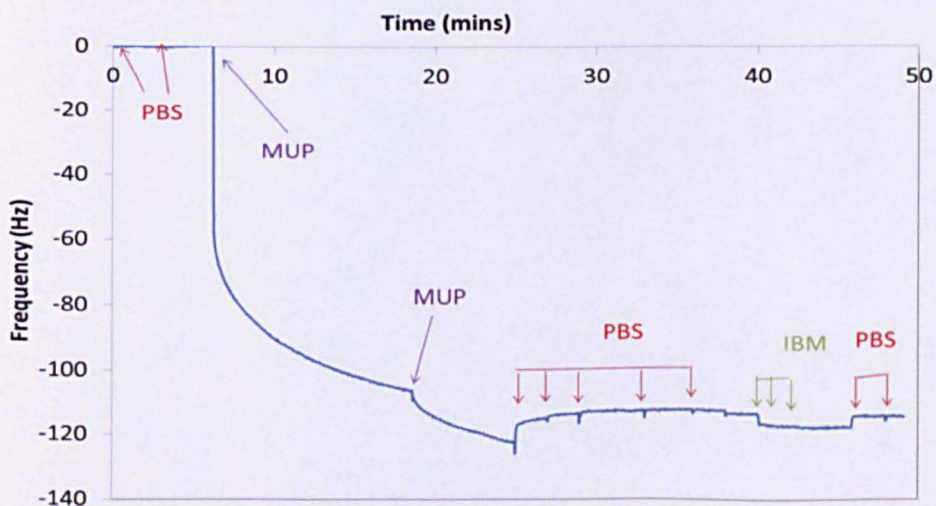**Figure 29**   **QCM data for the interaction of A103S-MUP with IBM. Similarly to *Figure 28* the blue line is the frequency signal and the red, pruple and green arrows are the addition of PBS, MUP and IBM respectively (as indicated on the graph).**

3.4.5 Water ejection in ligand binding cavity

The mutant WT and mutant MUP experiments were performed in the same way and replicated at least 3 times. From the mass changes of MUP and IBM (calculated from the frequency changes using the Saurbery equation) observed in each experiment, the number of water molecules displaced during IBM binding could be estimated. For example, an average frequency shift of -0.9 Hz has been observed for WT MUP and by using the Saubrey equation (*Equation 18*) ~8 ng is the amount of MUP on the surface. However, if it is assumed that when the MUP has been deposited on the surface and 1:1 binding takes place then it can be predicted that ~2 water molecules has been removed from the binding cavity. These results can be correlated with the observed frequency shift (with the use of the Saubrey equation from Chapter 2 –*Equation 18*) on ligand addition and it can be concluded that ~2 water molecules have been ejected from the binding cavity. The number of water molecules displaced on ligand binding during each experiment can be summarized in *Table 3*.

According to the table above IBM binding to Y120F MUP does not result in the ejection of water from the binding cavity, whereas the most water is lost from the binding cavity of hydrophilic MUP A103S and a small amount of water is lost from WT MUP. Therefore comparing and contrasting experimental results with MD results will reveal further information about the binding cavity. For example the MD simulations revealed that most water was lost from the binding cavity of A103S hydrophyillic, no water was lost from Y120F MUP (as there was no

water present in the binding cavity of MUP before ligand addition according to the hydration density maps) and a small amount of water is lost from WT MUP on ligand binding. These results are in good correlation with the experimental QCM results.

Table 3   **Observed QCM frequency shift values in PBS for the difference before and after MUP addition and before and after IBM addition to 3 different types of MUPs.**

	Wild Type	Y120F	A103S
ΔF_{MUP} (Hz)	-110±6.0	-121±6.8	-115±8.8
ΔM_{MUP} (ng)	1005±55	1105±62	1045±81
ΔF_{IBM} (Hz)	-0.9±0.1	-2.32±0.1	0±0
ΔM_{IBM} (ng)	8±1.3	21±0.9	0±0
No. of water molecules ejected	~2.1	0	~2.3

3.5   General Conclusions

In this chapter the hydration of the binding cavity of WT MUP and mutants MUPs whilst interacting with the ligand IBM has been analysed by experimental (QCM) and computational (MD) methods. The results revealed the importance of using replicate simulations as a method to validate the equilibration and stability of the simulations. The simulations also showed that equilibration can take upto 200 ns and therefore longer simulations have proven to be more valid then shorter simulations. The Ramachandran plot analysis on the other hand,

confirms that the overall geometry of the protein remains in excellent agreement with the crystal structure data, with the exception of the first ten residues which show large and variable conformational changes. The PCA reveals that all replicas for the WT MUP simulations sample in the same region of conformational space, despite having quite different trajectories. However, The Y120F simulation reveals that all the apo simulations sample in the same region of conformational space and the entire holo simulations sample in the same region of conformational space. The hydrophilic A103S on the other hand, all sample in the same region of conformational space except for one holo replicate which is a different region of conformational space. The results prove that the use of replicate simulations can provide evidence for the reliability and reproducibility of MD simulations.

The WT MUP MD simulations reveal that a very small molecule of MUP is present in the binding cavity of MUP (even without the presence of a ligand). These results support the dewetting phenomenon (i.e. even without the presence of the ligand the binding cavity remains well dehydrated). It has also been observed that no water molecules hop in and out of the binding cavity; however, the QCM data suggests that ~2.1 water molecules are lost from the binding cavity on ligand addition. The QCM results revealed that when the hydrophobic Y120F MUP was bound to the ligand IBM no water was lost from the binding cavity (i.e. the dewetting phenomenon). These results are in good correlation with the MD hydration density maps (where no water molecules are observed even without the presence of the ligand). However, the MD

simulations of Y120F MUP also show that a water molecule hops in and out of the binding pocket.

On the other hand most water was lost for A103S MUP on ligand binding which was observed from both QCM experiments. These results correlated well with the MD simulations as the hydration density maps revealed that a lot of water was present in the binding cavity even without the presence of the ligand. The duration of the water molecules in the binding cavity of A103S MUP shows evidence that water molecules are being exchanged more frequently than WT and Y120F MUP.

The experimental QCM results and computational MD simulations cannot be directly compared. This is because the QCM is a surface based approach whereas the MD simulations are simulated in explicit solvent. Also, the QCM measures in a larger timescale (minutes) whereas the MD simulations are in the microsecond time regime. However, in general the overall results from QCM and MD has shown a good correlation with each other. However further analysis (described in Chapter 4) from computational MD simulations and experimental QCM methods have been carried out to show the ligand (IBM) binding effects on WT and mutant MUPs.

3.6   References

Barillari, C., Taylor, J., Viner, R. and Essex, J. W. (2007) 'Classification of water molecules in protein binding sites', Journal of the American Chemical Society, 129(9), 2577-2587.

Barratt, E., Bingham, R. J., Warner, D. J., Laughton, C. A., Phillips, S. E. V. and Homans, S. W. (2005) 'Van der waals interactions dominate ligand-protein association in a protein binding site occluded from solvent water', Journal of the American Chemical Society, 127(33), 11827-11834.

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. and Homans, S. W. (2004) 'Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein', Journal of the American Chemical Society, 126(6), 1675-1681.

Bradshaw, E. M., Sanford, D. G., Luo, X. L., Sudmeier, J. L., Gurard-Levin, Z. A., Bullock, P. A. and Bachovchin, W. W. (2004) 'T antigen origin-binding domain of simian virus 40: Determinants of specific DNA binding', Biochemistry, 43(22), 6928-6936.

D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman (2006), AMBER 9, University of California, San Francisco.

D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C.Walker,W. Zhang, K.M. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K.F.Wong, F. Paesani, J. Vanicek, X.Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, and P.A. Kollman (2008), AMBER 10, University of California, San Francisco.

Castrignano, T., Chillemi, G., Varani, G. and Desideri, A. (2002) 'Molecular dynamics simulation of the RNA complex of a double-stranded RNA-binding domain reveals dynamic features of the intermolecular interface and its hydration', Biophysical Journal, 83(6), 3542-3552.

de Beer, S. B. A., Vermeulen, N. P. E. and Oostenbrink, C. (2010) 'The Role of Water Molecules in Computational Drug Design', Current Topics in Medicinal Chemistry, 10(1), 55-66.

Duan, J. X. and Nilsson, L. (2002) 'The role of residue 50 and hydration water molecules in homeodomain DNA recognition', European Biophysics Journal with Biophysics Letters, 31(4), 306-316.

Durchschlag, H. and Zipper, P. (2003) 'Modeling the hydration of proteins: prediction of structural and hydrodynamic parameters from X-ray diffraction and scattering data', European Biophysics Journal with Biophysics Letters, 32(5), 487-502.

Gao, Y. G., Robinson, H., Sanishvili, R., Joachimiak, A. and Wang, A. H. J. (1999) 'Structure and recognition of sheared tandem G center dot A base pairs associated with human centromere DNA sequence at atomic resolution', Biochemistry, 38(50), 16452-16460.

Geroult, S., Hooda, M., Virdee, S. and Waksman, G. (2007) 'Prediction of solvation sites at the interface of src SH2 domain complexes using molecular dynamics simulations', Chemical Biology & Drug Design, 70(2), 87-99.

Grafton, A. K. and Wheeler, R. A. (1999) 'Amino acid protonation states determine binding sites of the secondary ubiquinone and its anion in the Rhodobacter sphaeroides photosynthetic reaction center', Journal of Physical Chemistry B, 103(25), 5380-5387.

Hayes, J. M., Skamnaki, V. T., Archontis, G., Lamprakis, C., Sarrou, J., Bischler, N., Skaltsounis, A.-L., Zographos, S. E. and Oikonomakos, N. G. (2011) 'Kinetics, in silico docking, molecular dynamics, and MM-GBSA binding studies on prototype indirubins, KT5720, and staurosporine as phosphorylase kinase ATP-binding site inhibitors: The role of water molecules examined', Proteins-Structure Function and Bioinformatics, 79(3), 703-719.

Homans, S. W. (2007) 'Water, water everywhere - except where it matters?', Drug Discovery Today, 12(13-14), 534-539.

Jiao, M., Liang, Y., Li, H. T. and Wang, X. (2003) 'Studies on the unfolding of catalase induced by urea and guanidine hydrochloride', Acta Chimica Sinica, 61(9), 1362-1368.

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

Macek, P., Novak, P., Krizova, H., Zidek, L. and Sklenar, V. (2006) 'Molecular dynamics study of major urinary protein-pheromone interactions: A structural model for ligand-induced flexibility increase', Febs Letters, 580(2), 682-684.

Makarov, V., Pettitt, B. M. and Feig, M. (2002) 'Solvation and hydration of proteins and mucleic acids: A theoretical view of simulation and experiment', Accounts of Chemical Research, 35(6), 376-384.

Meyer, A. S., Blandino, M. and Spratt, T. E. (2004) 'Escherichia coli DNA polymerase I (Klenow fragment) uses a hydrogen-bonding fork from Arg(668) to the primer terminus and incoming deoxynucleotide triphosphate to catalyze DNA replication', Journal of Biological Chemistry, 279(32), 33043-33046.

Michel, J., Tirado-Rives, J. and Jorgensen, W. L. (2009) 'Prediction of the Water Content in Protein Binding Sites', Journal of Physical Chemistry B, 113(40), 13337-13346.

Nurisso, A., Blanchard, B., Audfray, A., Rydner, L., Oscarson, S., Varrot, A. and Imberty, A. (2010) 'Role of Water Molecules in Structure and Energetics of Pseudomonas aeruginosa Lectin I Interacting with Disaccharides', Journal of Biological Chemistry, 285(26), 20316-20327.

Perazzolo, C., Wist, J., Loth, K., Poggi, L., Homans, S. and Bodenhausen, G. (2005) 'Effects of protein-pheromone complexation on correlated chemical shift modulations', Journal of Biomolecular Nmr, 33(4), 233-242.

Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., III, Laughton, C. A. and Orozco, M. (2007) 'Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers', Biophysical Journal, 92(11), 3817-3829.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004) 'UCSF chimera - A visualization system for exploratory research and analysis', Journal of Computational Chemistry, 25(13).

Pitt, W. R. and Goodfellow, J. M. (1991) 'Modeling of solvent positions around polar groups in proteins', Protein Engineering, 4(5), 531-537.

Reichmann, D., Phillip, Y., Carmi, A. and Schreiber, G. (2008) 'On the contribution of water-mediated interactions to protein-complex stability', Biochemistry, 47(3), 1051-1060.

Rueda, M., Chacon, P. and Orozco, M. (2007) 'Thorough validation of protein normal mode analysis: A comparative study with essential dynamics', Structure, 15(5), 565-575.

Shields, G. C., Laughton, C. A. and Orozco, M. (1998) 'Molecular dynamics simulation of a PNA center dot DNA center dot PNA triple helix in aqueous solution', Journal of the American Chemical Society, 120(24), 5895-5904.

Sleigh, S. H., Tame, J. R. H., Dodson, E. J. and Wilkinson, A. J. (1997) 'Peptide binding in OppA, the crystal structures of the periplasmic oligopeptide binding protein in the unliganded form and in complex with lysyllysine', Biochemistry, 36(32), 9747-9758.

Sonavane, S. and Chakrabarti, P. (2009) 'Cavities in protein-DNA and protein-RNA interfaces', Nucleic Acids Research, 37(14), 4613-4620.

Syme, N. R., Dennis, C., Phillips, S. E. V. and Homans, S. W. (2007) 'Origin of heat capacity changes in a "Nonclassical" hydrophobic interaction', Chembiochem, 8(13), 1509-1511.

Tame, J. R. H., Murshudov, G. N., Dodson, E. J., Neil, T. K., Dodson, G. G., Higgins, C. F. and Wilkinson, A. J. (1994) 'The structural basis of sequence-independent peptide binding by oppa protein', Science, 264(5165), 1578-1581.

Virtanen, J. J., Makowski, L., Sosnick, T. R. and Freed, K. F. (2010) 'Modeling the Hydration Layer around Proteins: HyPred', Biophysical Journal, 99(5), 1611-1619.

Wang, G., Rodahl, M., Edvardsson, M., Svedhem, S., Ohlsson, G., Hoeoek, F. and Kasemo, B. (2008) 'A combined reflectometry and quartz crystal microbalance with dissipation setup for surface interaction studies', Review of Scientific Instruments, 79(7).

Withers, I. M., Mazanetz, M. P., Wang, H., Fischer, P. M. and Laughton, C. A. (2008) 'Active site pressurization: A new tool for structure-guided drug design and other studies of protein flexibility', Journal of Chemical Information and Modeling, 48(7), 1448-1454.

Yabuki, M., Portman, K. L., Scott, D. J., Briand, L. and Taylor, A. J. (2010) 'DyBOBS: A Dynamic Biomimetic Assay for Odorant-Binding to Odor-Binding Protein', Chemosensory Perception, 3(2), 108-117.

Yokota, A., Tsumoto, K., Shiroishi, M., Kondo, H. and Kumagai, I. (2003) 'The role of hydrogen bonding via interfacial water molecules in antigen-antibody complexation - The HyHEL-10-HEL interaction', Journal of Biological Chemistry, 278(7), 5410-5418.

Chapter 4 - Protein dynamics

In this chapter the aim is to study the role of flexibility in ligand (IBM) binding to the protein MUP by the use of computational (MD studies) and experimental methods (QCM and AFM).

4.1   Introduction

The structural flexibility of proteins is a well-known phenomenon and an important consideration in molecular recognition (Wand, A. J. 2001). Proteins are intrinsically mobile structures and their conformational change can play an important role in their function (Cozzini, P. et al, 2008). Conformational stability maybe influenced by interactions with other molecules or the thermodynamic conditions of the proteins environment. Proteins are dynamic in nature and are known to change from one conformation to another and the changes in flexibility of proteins are important especially during the unfolding process. Protein molecules are made from one or more polypeptide chains which are intrinsically flexibile structures. This flexibility arises from the ease of torsional rotation around most of the polypeptide's covalent bonds (as explained in Chapter 2).

Flexibility plays a role in how a protein interacts with other molecules. The flexibility of a protein structure can be affected by the binding of biological molecules such as substrate to the active site of an enzyme. Binding sites are often located in clefts of proteins to maximize the number of contacts it can make with other molecules to give strong and selective binding. Proteins achieve this by adapting their shape to increase the strength of the interactions between the ligand and the

protein (as explained in Chapter 1). In some cases the features that contribute to the binding site may already be wide enough to allow access to the binding site but too far apart to allow tight binding. Upon ligand binding, these features close around the ligand to produce a tightly bound protein–ligand complex.

X-ray crystallography, despite producing the highest number of experimentally solved protein structures, is limited in terms of studying conformational changes in proteins associated with the activity. Crystal structures of protein-ligand complexes can be obtained by co-crystallisation or soaking. Both of these methods can be affected by certain conditions (for example a change in temperature, protein and ligand concentration and the use of additives to improve ligand binding). However, if these methods are not successful then further optimisation of the protein expression and purification maybe necessary (Danley, D. E., 2006 and McNea, I. W., et al., 2005).

On the other hand nuclear magnetic resonance (NMR) is considered better suited to the study of structural dynamics of proteins. There are cases where the domain reorientations upon ligand binding can adopt a different conformation in the holo state and a more compact shape can cause a decrease in the rotational correlation time (Collingridge, G. L. et al., 2009; Li, Y. T. and Yang, C. M., 2005; Lee, D. et al., 2006). Another example of an NMR experiments revealed that on ligand (IBM) binding to the protein MUP there is an overall increased flexibility (Bingham, R. J., et al., 2004).

Finally, computational approaches such as molecular dynamics (MD) simulations are used extensively to characterize protein flexibility (Emperador, A. et al., 2010; Carlson, H. A. et al., 2000; Mamonova, T. et al., 2005; Hansson, T. et al., 2002). Computational approaches on their own are rarely able to provide information on flexibility in structures for typical biomolecules (although cases such as York, D. M. et al., 1993 who reported the simulation of *B-factors*) and are often applied in alliance with experimentally derived information.

Conformations generated from MD simulations can be used in other computational methods such as virtual screening, docking and scoring. For instance, virtual screening studies have shown poor ability to differentiate between binders and non-binders of a ligand to a protein (McGovern, S. L. et al., 2003; Warren, G. L. et al., 2006). Systems such as G-protein coupled receptors, (GPCRs) for which high resolution structural data is very challenging to generate using X-ray crystallography or NMR techniques, are widely studied using computational molecular dynamics (Dror, R. O. et al., 2011). Another example is observed in a study where valuable predictive models were generated to successfully interpret QSAR data (Hao, M. et al., 2011).

In this chapter we aim to study the flexibility of MUP when binding to the ligand IBM by computational and experimental techniques. Continuing developments in these areas are expected to improve our understanding of the role of protein flexibility in protein function and molecular recognition.

4.2   General Methods

The methods and techniques used for QCM experiments and MD simulations have been described in *Chapter 3.2*. All the analysis for the MD simulations has been performed in the PTRAJ module of AMBER 10. The QCM-D data has been used by monitoring the changes in flexibility through the dissipation signal which is obtained simultaneously with the frequency signal. All the QCM analysis was performed in the Qtools (Qsense, Gotenborg, Sweden). However, the experimental method for AFM experiments is described below.

4.2.1 AFM Experimental Methodology

The sample preparation for MUP at 1 µM and IBM 50 µM has been prepared from PBS (pH 7.4) buffer solution. Following adsorption of MUP onto freshly cleaved mica surfaces, they were imaged in PF-QNM mode (as described in *Chapter 2*) using an open liquid cell. Afterwards, the IBM solution was added to the liquid cell and the samples re-imaged in PF-QNM mode. The data channels which were captured to allow for visualisation/calculation of the flexibility of MUP-IBM interaction, that includes the topography, modulus and deformation. For simplicity, only one set of results is presented in the following results section, although the experiment was repeated several times and consistent data obtained.

4.3   Results and Discussion

4.3.1 Modelling evidence of protein dynamics

Despite the inherent complexity of molecular recognition, computational methods have been extensively developed over the past few years for modelling and prediction of protein-ligand interactions. These approaches can be divided into 2 main parts in this chapter of the thesis (RMS fluctuations of each residue in a protein during MUP-IBM interaction and ligand induced flexibility).

4.3.1.1   RMS Fluctuations

The RMS fluctuations of WT and mutant MUPs (apo and holo) can provide a more detailed view of protein flexibility (*Figure 30*). As expected, the most mobile regions of the structure are in the loops for MUP. We see that the fluctuation profiles are well conserved between replicates, but that some variation exists, particularly (and not unexpectedly) in the most mobile regions for MUP.

All 6 graphs also include the RMS fluctuation profiles generated from analysis of the three individual time-averaged structures (the purple line in each of the 6 graphs). This is a way to simply test the reproducibility of the observations, so the regions with low RMS fluctuation are those that are very similar in different replicates, whereas high fluctuations identify divergence in the time-averaged structures. Again, not unexpectedly, it is in the most flexible regions of the protein that it is

136

most likely that the time-averaged structures of the replicates will be somewhat divergent, since it is harder to achieve full sampling.

Overall WT (apo) MUP has not reproduced (i.e. the replicates are not very similar) as well as the ligand bound WT (holo) MUP according to the time-averaged structures on the RMS fluctuation graph (*figures 30a and 30b*-indicated by the purple line). These results are similar to the A103S MUP simulations. However, for the Y120F mutant MUP simulations the opposite is true. The time-averaged simulations reveal the apo form reproduces better than the holo form for Y120F mutant MUP. However, around L2 (residues 43–47) is a dynamic region that appears to have been very similarly sampled in the replicate simulations. This type of behaviour may indicate regions that, have flexible motions that can be well sampled over this timescale, whereas in other cases (see, e.g., around L3, residues 57–63) the motion may involve complex "jumping among minima" that make it much harder to sample the motion well. This is true for all 3 types MUP (apo and holo) simulations.

4.3.1.2   Average RMS fluctuation of each replicate

The average (apo is the green line and holo is the pink line) from each replicate from *Figure 30* above was calculated and is presented in *Figure 31* below. Again the L3 region for each simulation (*Figure 31a, b and c*) is observed to have a dynamic feature. However, only by taking the difference between the green and pink lines (apo and holo), the

flexibility of each residue of the protein can be calculated for WT, Y120F

and A103S MUPs and is shown in *Figures 32, 33 and 34* respectively.
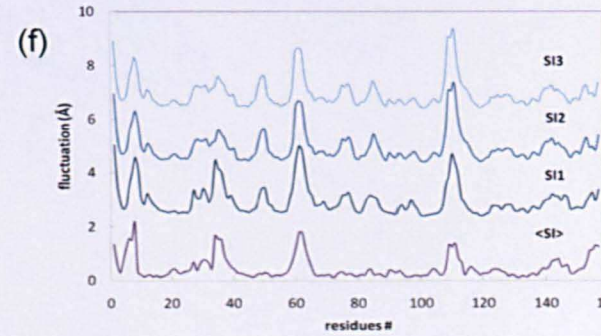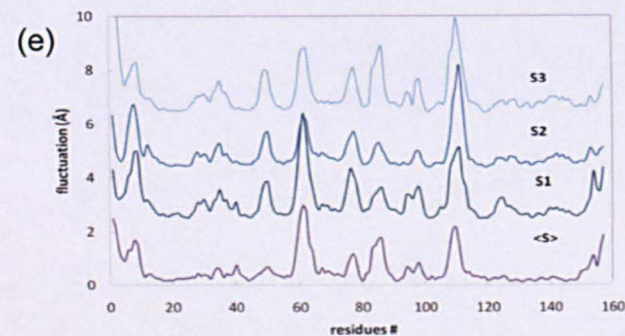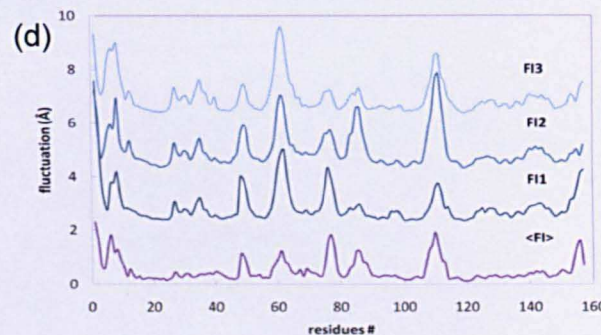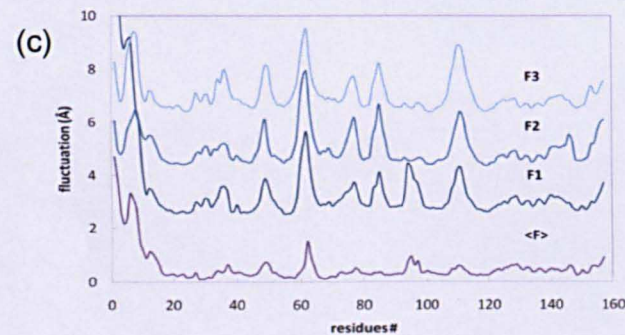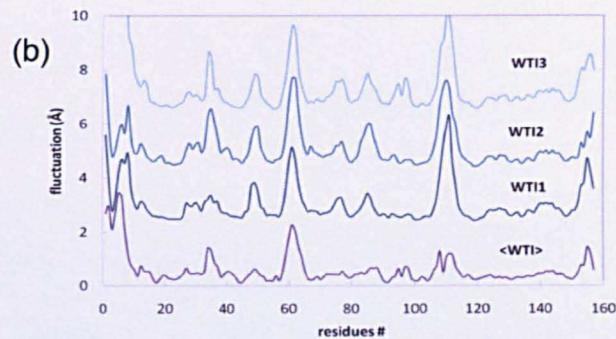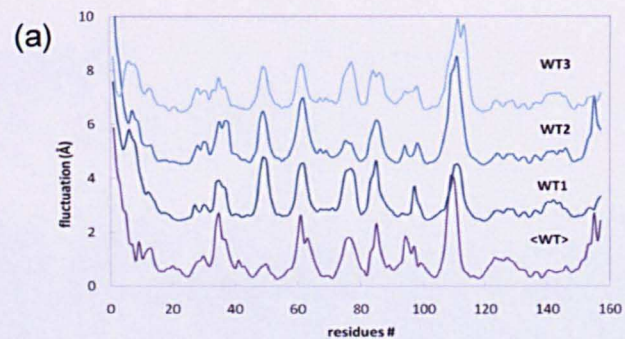
Figure 30   RMS fluctuations of individual amino acids for replicate apo (left) and ibm (right) simulations. (a) and (b) are the WT MUP simulations, (c) and (d) are the Y120F MUP simulations and (e) and (f) are the A103S MUP simulations. In each case the lowest plot is calculated from the fluctuations between the time-averaged structures from each simulation, whereas the plots stacked above are calculated from all snapshots within an individual simulation (i.e. each blue line represents a replicate simulation) and are displaced vertically by 1, 3, and 5 angstrom, respectively, for clarity.

Figure 31   Average residue fluctuations of apo (pink) and holo (green) of WT (a), Y120F (b), and A103S (c). Each holo (pink) plot has been shifted by 2 angstrom for clarity.
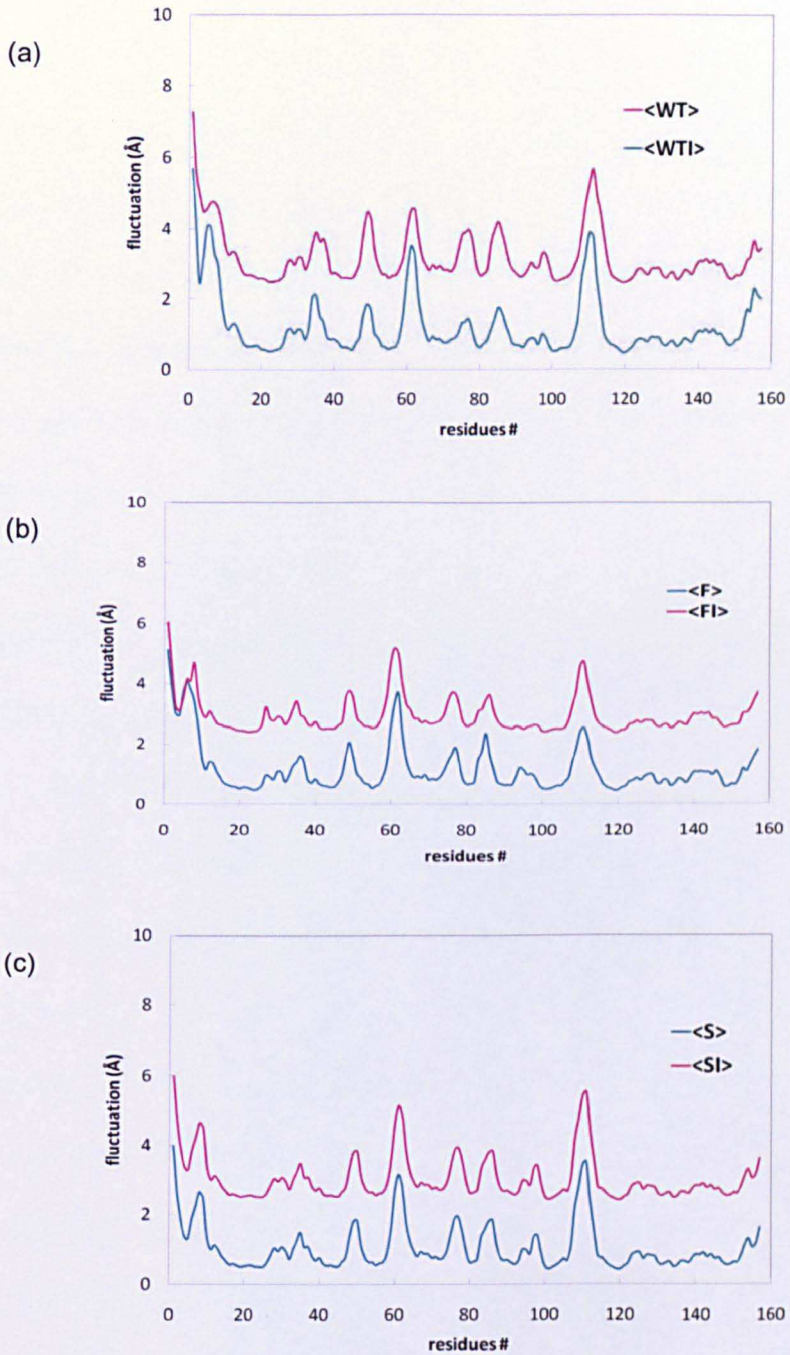
4.3.1.3   Average change in RMS fluctuation

Figures 32, 33 and 34 show the average change in RMS fluctuation (i.e. holo-apo). A useful application of having replicate simulations is to be able to carry out a statistical test. A students T-test has been carried out and the relative *p*-values are shown in *Figures 32, 33 and 34*. If the *p*-values are less than 0.2 (0.2 instead of the standard 0.05 is used as there are not many replicates available in comparison to experimental work) then the region is statistically significant. The statistically significant regions are indicated by a black bar. Some of these differences are not observed as statistically significant, and in any case we must be aware that in analyzing the NMR data, we are using a small number of structures that have been chosen for deposition by criteria other than that they are representative of an equilibrium distribution.

Figures 32, 33 and 34 shows how RMS fluctuations change on ligand binding. There is no obvious relationship between how close a residue is to the bound ligand and any changes in that residue's dynamics. Although certain regions of the protein become more rigid, others become more flexible for WT and A103S simulations. However, for the Y120F simulations only a decrease in flexibility is observed on ligand binding. This indicates that the residue tyrosine 120 in the binding cavity does play an effect to the dynamics of ligand binding to MUP.

Figure 32   RMS fluctuation of each amino acid of (apo) WT MUP on ligand binding (top), and associated *p*-values (below). RMS changes that have a *p*-value less than 0.2 are highlighted with black bars.
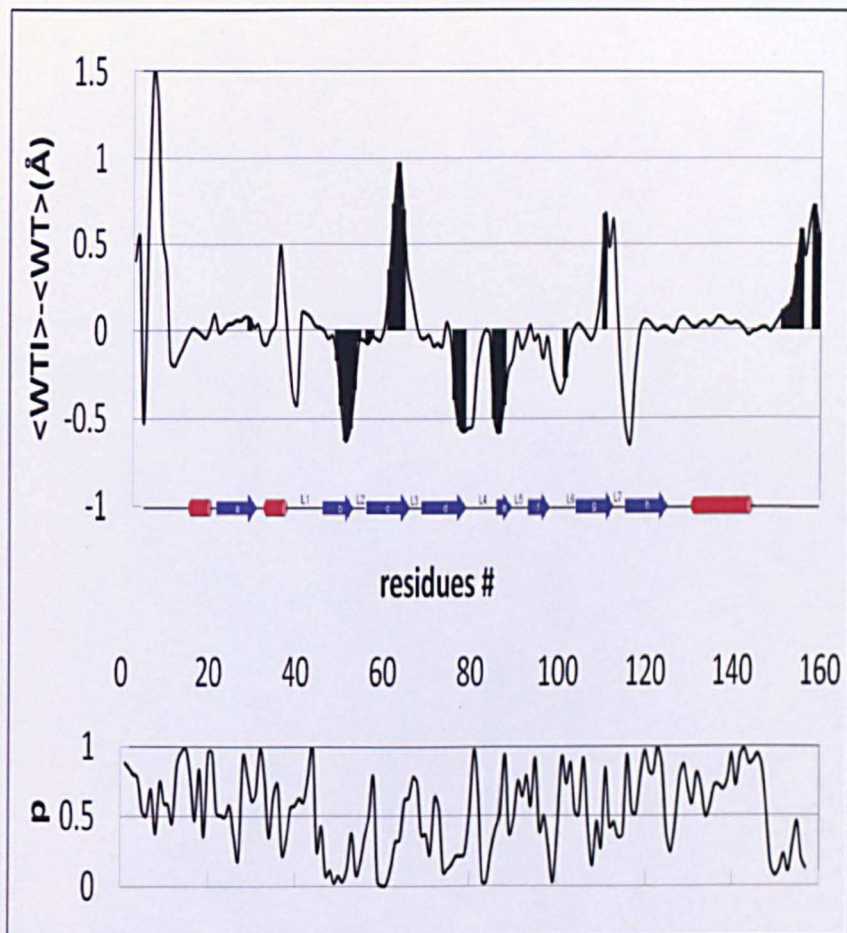
As it can be observed that for the L3 loop there is an increase in flexibility which is statistically significant whereas the loop L2 has a decrease in flexibility.

Figure 33   RMS fluctuation of each amino acid of (apo) Y120F MUP on ligand binding (top), and associated *p*-values (below). RMS changes that have a *p*-value less than 0.2 are highlighted with black bars.
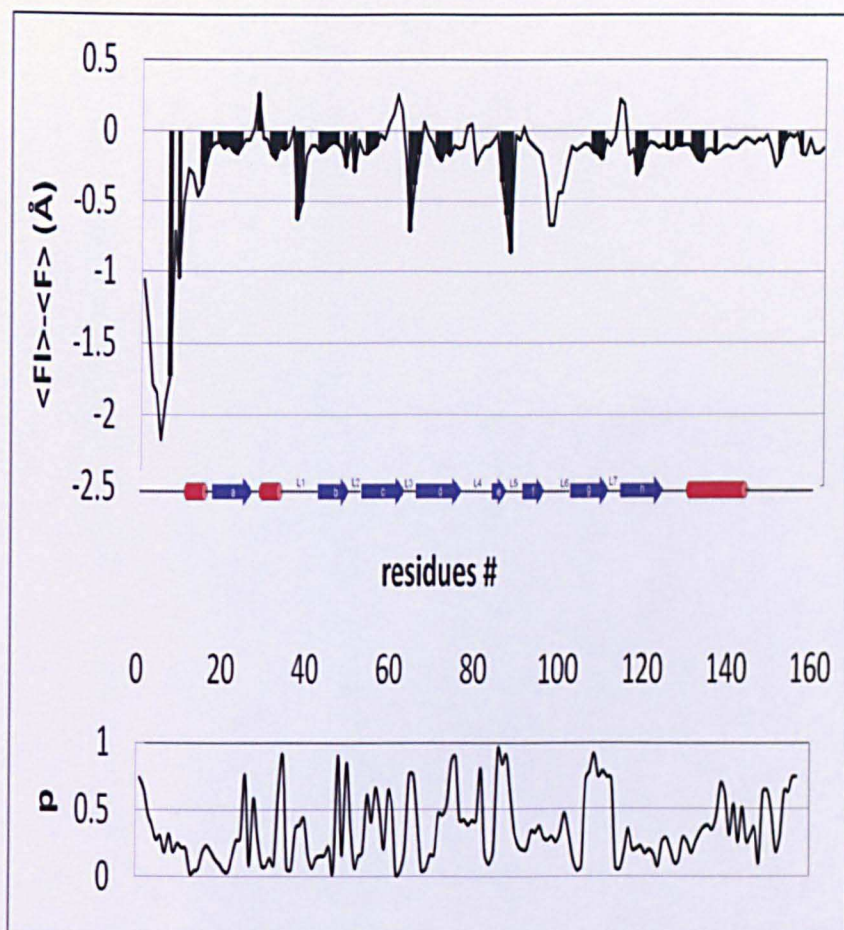
As it can be observed from *Figure 33* most of the statistically significant regions have shown a decrease in flexibility.

Figure 34   RMS fluctuation of each amino acid of (apo) A103S MUP on
ligand binding (top), and associated *p*-values (below). RMS changes that have a
p-value less than 0.2 are highlighted with black bars.
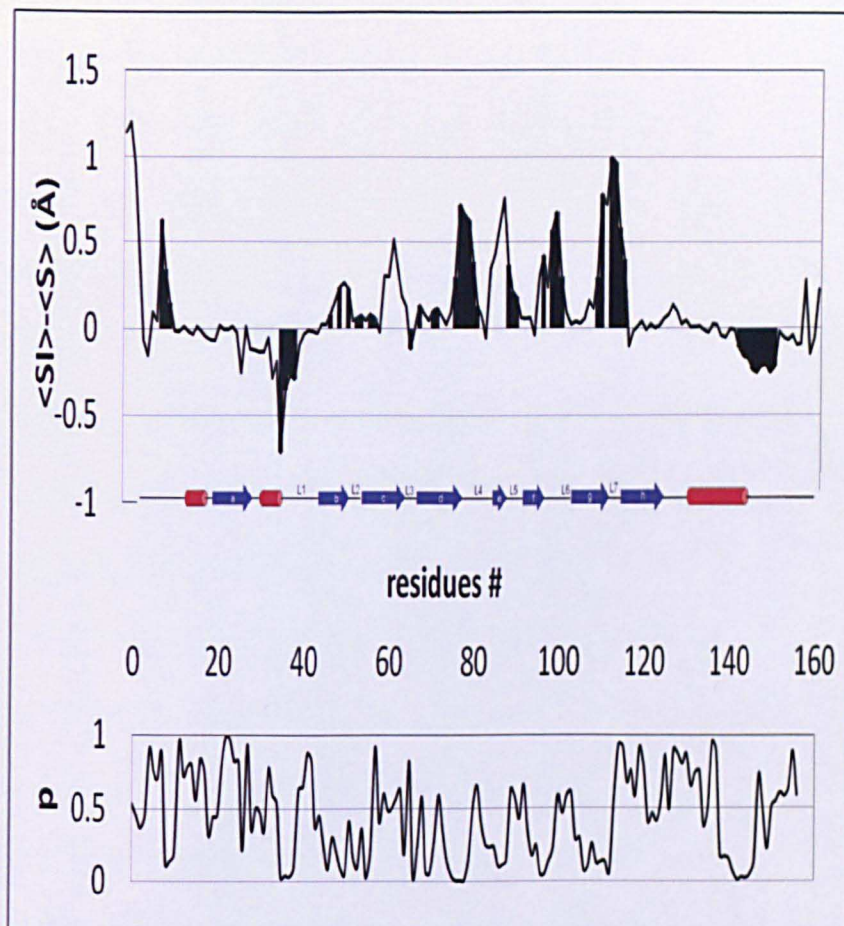
As it can be observed from *Figure 34* that some regions become flexible
whereas others show a decrease in flexibility. For example L1 shows a
significant decrease in flexibility whereas L4 shows a significant
increase in flexibility. These findings are different from the WT and
Y120F simulation results.

4.3.1.4   Ligand Dynamics

The simulations of WT and mutant MUP-IBM interactions reveal that the pyrazine is not held rigidly in the cavity, but tumbles extensively. The orientation of the pyrazine N1–N4 vector is observed as the red dots on all the graphs in *Figure 35*. H-bonding between the N1 or N4 and the Tyr-120 hydroxyl group is permitted although, other orientations without H-bond can also occur frequently. The orthogonal C2-C6 vector is observed as green dots on all the graphs on *Figure 35* and can be monitored and reveals a great deal of motion.

The ligand dynamics between each replicate for all three types of simulations is variable. However, it is difficult to say whether or not the H-bond between the N1 or N4 and the Tyr-120 plays a role in the dynamical behaviour of the ligand. This is because both the mutant simulations do not show much tumbling of the ligand on a microsecond timescale (except for Y120F-replicate 1). On the whole ligand flexibility for the mutant simulations does not show as much flexibility as the wild-type simulations (with Y120F-replicate 1 being an exception). On-going NMR experiments from the Homans group from the University of Leeds are in the process of validating these results.

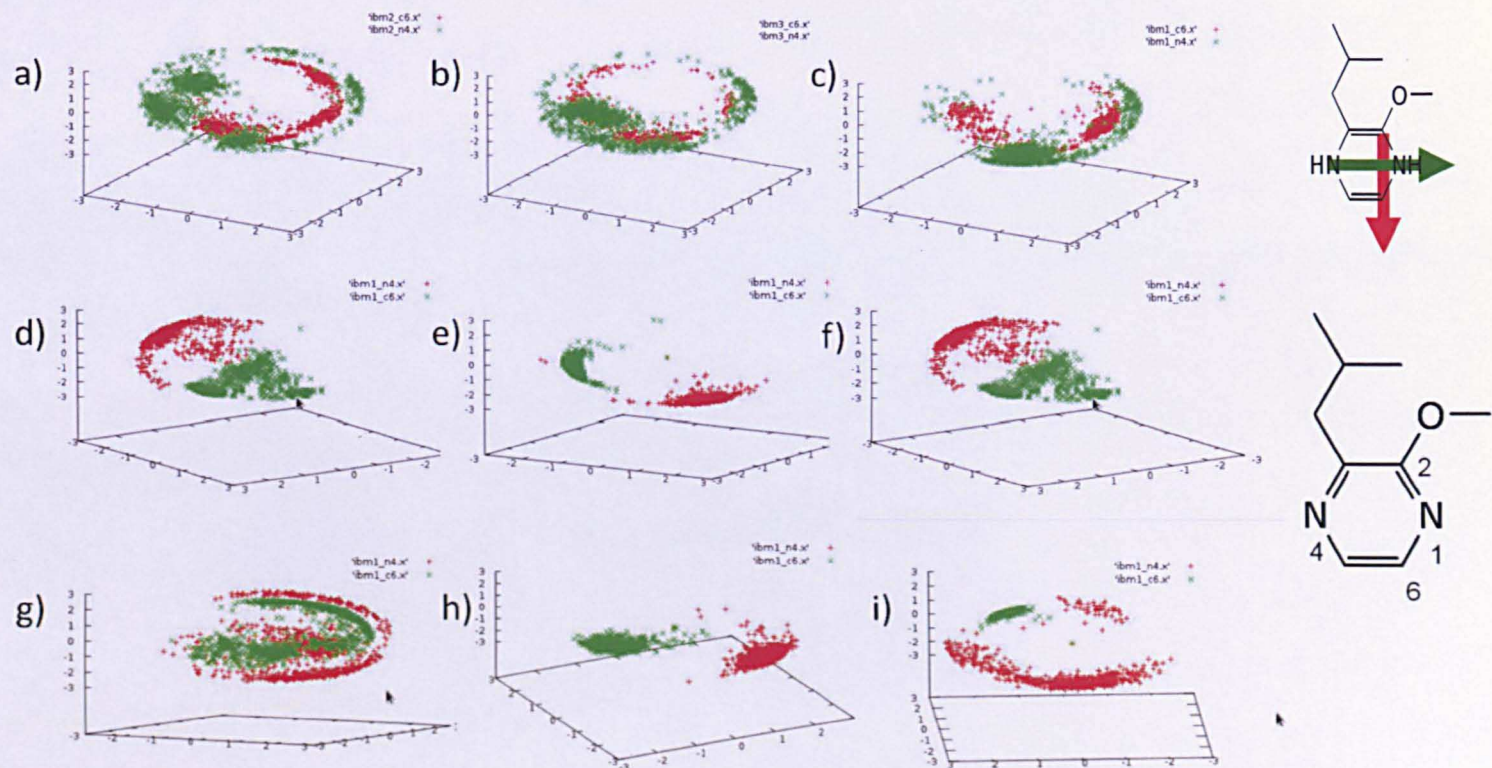**Figure 35** Polar plots to illustrate ligand tumbling. The red and green dots represent the orientation of C2-C6 vector and the N1-N4 vector respectively (illustrated on both structure of IBM on the right). The top three plots correspond to WT-MUP replicates 1, 2, and 3. The middle three plots correspond to A103S mutant MUP replicates 1, 2 and 3. The bottom three plots correspond to Y120F mutant MUP replicates 1, 2 and 3.

146

4.3.2 Experimental evidence of protein dynamics

As described earlier, NMR experiments can provide evidence on the flexibility on protein-ligand interactions. However, the values obtained are only estimated values, calculated from various parameters.

ITC experiments have the ability to measure thermodynamic parameters in protein-ligand interactions. Large amounts of data are available from ITC experiments providing thermodynamic parameters however, QCM-D and PF-AFM have not been applied to the MUP-IBM interaction before and can give direct information on the flexibility of the system. Therefore to further gain information on MUP-ligand interactions, QCM and AFM experiments and MD simulations were carried out. This can provide qualitative evidence which can support and validate the results found from NMR and ITC experiments. The results presented in this part of the thesis are from QCM-D and AFM studies of ligand IBM binding to WT and mutant MUPs.

4.3.2.1   Dissipation analysis from QCM-D on other biological molecules

QCM has been a widely used technique to study the flexibility of biological molecules on a quartz crystal surface. For example QCM has been used to bilayers formation on self- assembled monolayer surfaces

(Hook, F. et al, 1998), protein-DNA interactions (Xu, J. et al., 2011) protein-protein interactions (Praporski, S. et al., 2009).

4.3.2.2   Changes in flexibility of MUP and its mutants on ligand binding by QCM-D

The dissipation trace from *Figure 36* shows an increase in signal, first upon MUP adsorption, and subsequently on IBM addition. The dissipation signals for the PBS were observed for the before and after addition of sample (MUP/IBM). This change in the dissipation signal (ΔD) shows the overall change of the flexibility of the surface which means that an increase in the signal means that the layer formed on the crystal surface has become more flexible. Similar results (shown as a summary in *Table 4* below) were observed for the mutant MUPs.

Table 4   Dissipation changes observed upon protein adsorption (ΔD_{mup}) and ligand addition (ΔD_{ibm}) to the 3 different MUP variants studied.

	Wild Type	Y120F	A103S
ΔD_{MUP} (10^{-6})	+0.7±0.03	+2.5±0.25	+2.7±0.36
ΔD_{IBM} (10^{-6})	+0.16±0.005	+0.17±0.0042	+0.12±0.007

The results indicate that the adsorption of WT and mutant MUPs to the sensor surface increases the dissipation signal by ~+0.7 (10^{-6}). This means that the layer becomes flexible when the protein is attached to

148

the crystal surface. The size of MUP is 19 kDa and a protein of this size would be expected to be more dissipative when compared to the uncoated sensor surface. For example a dissipation shift (ΔD) of ~5 (10^{-6}) was observed for the absorption of the protein Estrogen (20 kDa protein) (Peh, W. Y. X. et al., 2007). The molecular weight is similar to this protein, but adsorption of MUP results a lower dissipation shift suggesting that MUP has a more compact rigid structure. This observation is in agreement with other experimental studies of MUP (Barrett, E. et al., 2005).

The second observation from the dissipation signal is that on ligand binding the dissipation signal increases by ca. 0.16 (10^{-6}) for WT and mutant MUPs. This means that there is an increase in flexibility of the layer formed on the crystal surface when the ligand is bound to the protein. It can also be observed qualitatively that the flexibility is the same for all 3 types of MUPs, and the mutation does not make a difference to the amount the layer has become flexible.

Also, the dissipation signal can give an indication that protein and ligand binding has actually taken place. For example, in *Figure 36*, after the addition of MUP the second set of PBS injections are made and the dissipation signal does not go back to the original starting level as beginning of the experiment. This means that a more flexible protein layer has been attached to the crystal. The same applies when the ligand is injected. The PBS injections before and after the addition of the ligand are different. Overall there is an increase in the dissipation signal. This means that the ligand has been bound to the protein and

has not been washed off the surface. However, to make sure that specific binding to the protein surface took place, complimentary control experiments were carried out as explained in Chapter 3.4.1.

Figure 36 A dissipation (red line) graph. The spikes are indicated by an injection point which is highlighted by the purple, green and red arrows which represent PBS, WT MUP and IBM respectively.
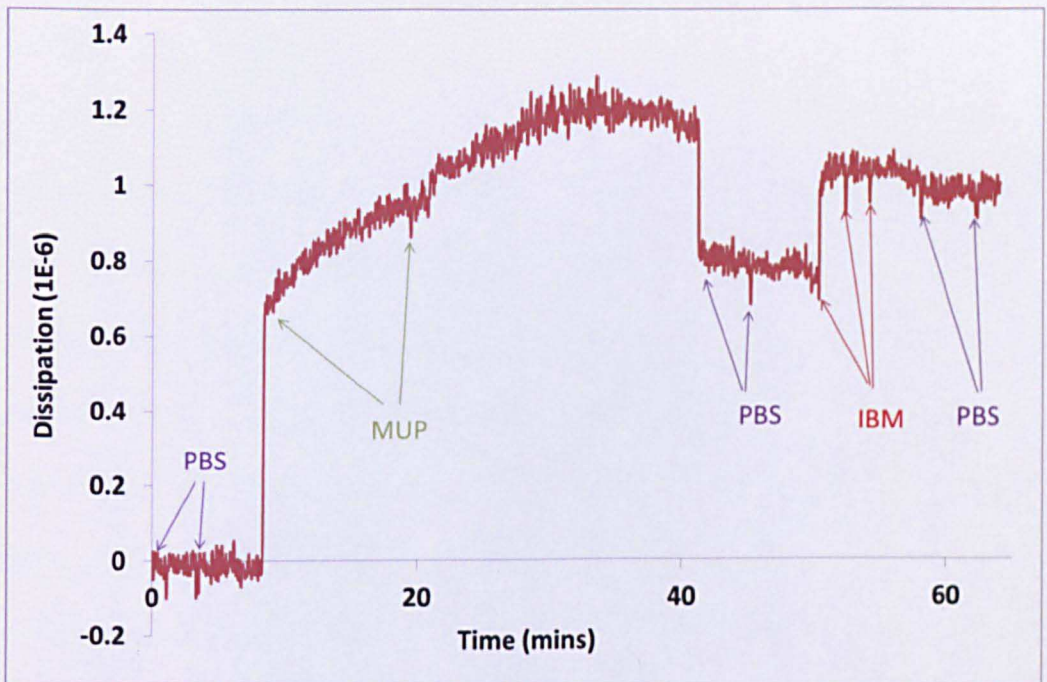
4.3.2.3 Further supporting evidence from AFM Experiments

Similar to the QCM experiments, control PF-QNM-AFM experiments were performed first to observe the effects of the IBM on the surface without the presence of protein. The results showed no change to the mica surface for images of topography, Young's modulous (as determined using the *Derjaguin, Muller, Toporov (DMT)* contact model), or deformation. This suggests that the ligand IBM does not bind to the mica surface.

In *Figure 37*, topography, DMT and deformation images of WT MUPs before and after ligand addition are shown. The experiments were repeated for mutant Y120F and A103S MUPs, with replicates in PBS buffer acquired with PF-QNM-AFM before and immediately after incubation with IBM. The concentration of MUP and IBM used were the same as the QCM experiments. Each AFM experiment was repeated at least 3 times. The typical surface characteristics are represented although the image surface locations may not be exactly the same, because it proved difficult for the tip to remain in the same sample area following the injections of IBM solution.

The images showed individual proteins or protein-IBM complexes (shown as a 'globular' feature') on a flat substrate for WT, Y120F and A103S MUPs. The lateral sizes of these biological molecules fell between 2 - 20 nm, corresponding to individual MUP molecules and aggregations of two or three molecules. By visual inspection it was difficult to tell whether there was any change in the size of the biological moleculesfollowing IBM exposure. However after performing a detailed quantitative grain analysis (see Chapter 2) on these images, we found that there is no significant change in both height and lateral dimensions.
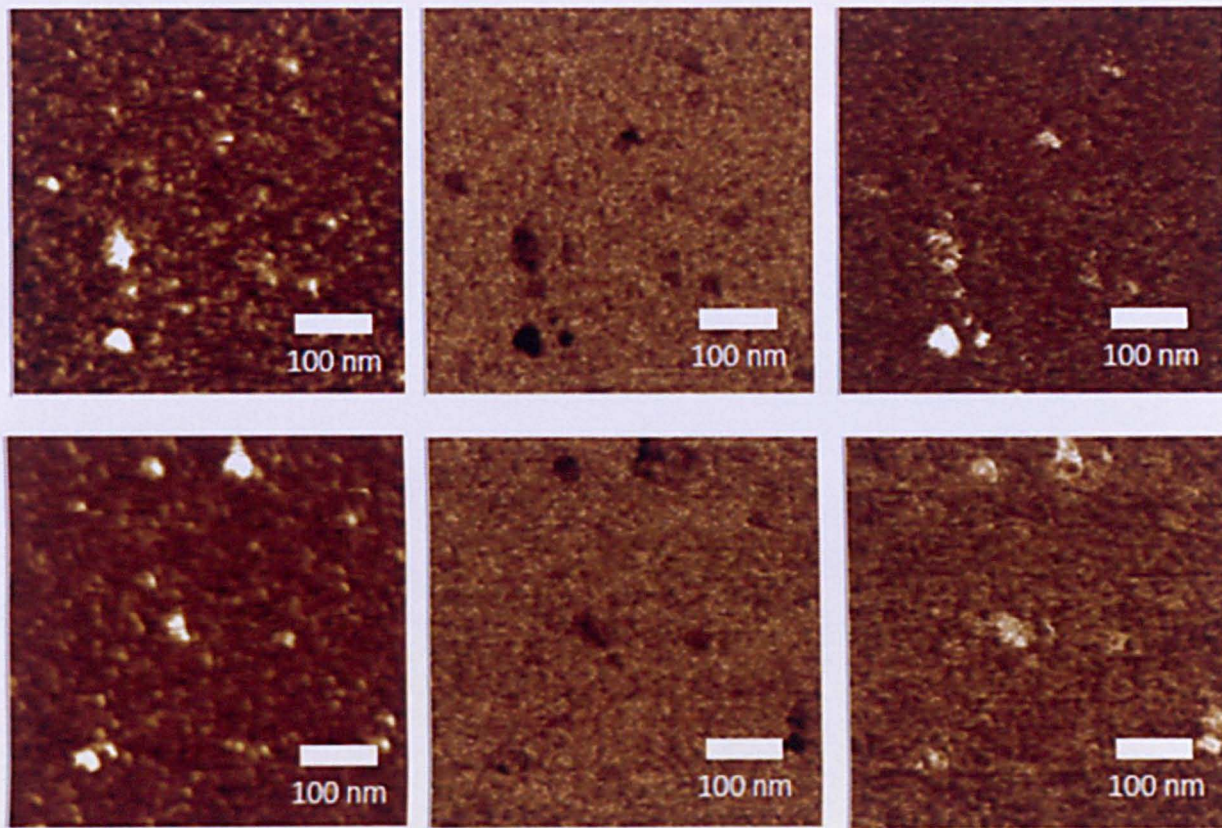
Figure 37   AFM images of Y120F mutant MUP (top) binding to IBM (bottom). The images present are topography (left), DMT (middle) and deformation (right)

The WT results shown above for before and after IBM incubation are presented in terms of topology, DMT and deformation. Similar results have been produced for Y120F and A103S mutant MUP with interaction to IBM. Statistical analysis shows that the average values for those "dark" dots corresponding to individual WT MUP particles are 180±90 MPa (before IBM incubation) and 80±50 MPa for (after IBM incubation), respectively. This suggests a "softening" effect following IBM incubation.

However, statistical analysis also shows that the background values are 270±100 MPa (before) and 125±50 MPa (after), respectively, with a similar factor of 2 as shown in the modulus values measured on the protein globular features before and after IBM incubation. One situation is that incubation of IBM resulted in modulus decrease of both MUP molecules and the mica surface, but this is unlikely considering the control experiments at the start of this section. Another possibility is a systematic shift in the measured modulus value due to tip apex condition change. In such a case, there was probably no significant change in DMT before and after IBM incubation.

However, closer inspection of the deformation images reveals that after ligand addition the surface becomes more compressible which means there is an increase in flexibility after ligand addition. The AFM experiments were repeated for mutant Y120F and A103S MUP (before and after) with IBM incubation and similar results were observed.

Another control experiment was also performed to ensure that specific binding of the IBM took place with the protein.

Figure 38   AFM images of topography (left) and deformation (right). The images on the top show the addition of pre-incubated MUP to the mica and on the bottom show pre-incubated MUP with ligand IBM addition. The images were scanned at 500nm resolution.
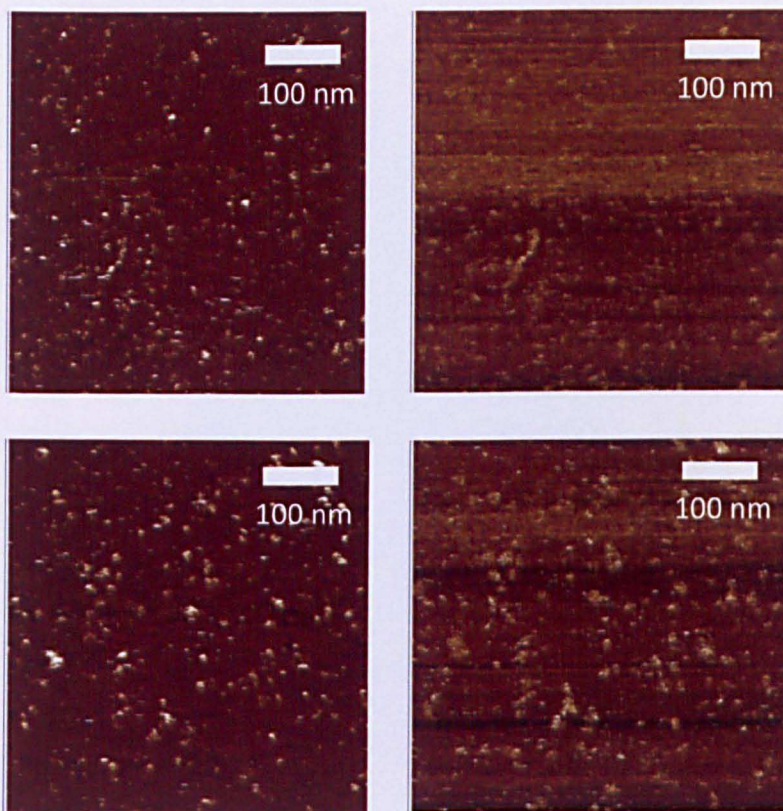
This control then shows that this effect can be 'knocked out' by pre-incubating with ligand before performing the experiment. This experiment was conducted in a similar way to the QCM experiments by exposing a pre-IBM incubated MUP preparation and then further insertions of IBM during the experiment to ensure that no more binding

can take place. The results show showed no change in topology or deformation following exposure to IBM (*Figure 38*). The data reveals similar results as the QCM experiments (i.e. further ligand binding does not take place as the binding cavity already has a ligand present).

4.4   Conclusions

From the experimental and computational analysis of results it is evident that the MUP-IBM interaction is a dynamic process.

The computational MD simulations on the other hand show that some residues become more flexible whereas other regions become more rigid for WT and A103S MUP on ligand binding. The Y120F simulation results indicate a decrease in flexibility on ligand binding.

The experimental QCM-D and AFM results indicate that there is an increase in flexibility in the WT and mutant MUP proteins upon interaction with IBM. However, the two methods cannot be directly compared (even though they give the same results). This is because in both experiments there are various estimations involved in both techniques.

However, we can also say that the experimental results and MD simulations cannot be directly correlated as the experiments are based on the protein being attached to a surface whereas the MD simulations are not based on the protein being attached to the surface.

4.5   References

Barratt, E., Bronowska, A., Vondrasek, J., Cerny, J., Bingham, R., Phillips, S. and Homans, S. W. (2006) 'Thermodynamic penalty arising from burial of a ligand polar group within a hydrophobic pocket of a protein receptor', Journal of Molecular Biology, 362(5), 994-1003.

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. and Homans, S. W. (2004) 'Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein', Journal of the American Chemical Society, 126(6), 1675-1681.

Carlson, H. A., Masukawa, K. M., Rubins, K., Bushman, F. D., Jorgensen, W. L., Lins, R. D., Briggs, J. M. and McCammon, J. A. (2000) 'Developing a dynamic pharmacophore model for HIV-1 integrase', Journal of Medicinal Chemistry, 43(11), 2100-2114.

Collingridge, G. L., Olsen, R. W., Peters, J. and Spedding, M. (2009) 'A nomenclature for ligand-gated ion channels', Neuropharmacology, 56(1), 2-5.

Cozzini, P., Kellogg, G. E., Spyrakis, F., Abraham, D. J., Costantino, G., Emerson, A., Fanelli, F., Gohlke, H., Kuhn, L. A., Morris, G. M., Orozco, M., Pertinhez, T. A., Rizzi, M. and Sotriffer, C. A. (2008) 'Target

Flexibility: An Emerging Consideration in Drug Discovery and Design', Journal of Medicinal Chemistry, 51(20), 6237-6255.

Danley, D. E. (2006) 'Crystallization to obtain protein-ligand complexes for structure-aided drug design', Acta Crystallographica Section D-Biological Crystallography, 62, 569-575.

Dror, R. O., Pan, A. C., Arlow, D. H., Borhani, D. W., Maragakis, P., Shan, Y., Xu, H. and Shaw, D. E. (2011) 'Pathway and mechanism of drug binding to G-protein-coupled receptors', Proceedings of the National Academy of Sciences of the United States of America, 108(32), 13118-13123.

Emperador, A., Meyer, T. and Orozco, M. (2010) 'Protein flexibility from discrete molecular dynamics simulations using quasi-physical potentials', Proteins-Structure Function and Bioinformatics, 78(1), 83-94.

Hansson, T., Oostenbrink, C. and van Gunsteren, W. F. (2002) 'Molecular dynamics simulations', Current Opinion in Structural Biology, 12(2), 190-196.

Hao, M., Li, Y., Wang, Y., Yan, Y., Zhang, S., Li, G. and Yang, L. (2011) 'Combined 3D-QSAR, Molecular Docking, and Molecular Dynamics Study on Piperazinyl-Glutamate-Pyridines/Pyrimidines as Potent

P2Y(12) Antagonists for Inhibition of Platelet Aggregation (vol 51, pg 2560, 2011)', Journal of Chemical Information and Modeling, 51(11), 3027-3027.

Hook, F., Rodahl, M., Kasemo, B. and Brzezinski, P. (1998) 'Structural changes in hemoglobin during adsorption to solid surfaces: Effects of pH, ionic strength, and ligand binding', Proceedings of the National Academy of Sciences of the United States of America, 95(21), 12271-12276.

Lee, D., Hilty, C., Wider, G. and Wuthrich, K. (2006) 'Effective rotational correlation times of proteins from NMR relaxation interference', Journal of Magnetic Resonance, 178(1).

Li, Y. T. and Yang, C. M. (2005) 'A tryptophan-containing open-chain framework for tuning a high selectivity for Ca_2^+ and C-13 NMR observation of a Ca_2^+-indole interaction in aqueous solution', Journal of the American Chemical Society, 127(10), 3527-3530.

Mamonova, T., Hespenheide, B., Straub, R., Thorpe, M. F. and Kurnikova, M. (2005) 'Protein flexibility using constraints from molecular dynamics simulations', Physical Biology, 2(4), S137-S147.

McGovern, S. L., Caselli, E., Grigorieff, N. and Shoichet, B. K. (2002) 'A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening', Journal of Medicinal Chemistry, 45(8), 1712-1722.

McNae, I. W., Kan, D., Kontopidis, G., Patterson, A., Taylor, P., Worrall, L. and Walkinshaw, M. D. (2005) 'Studying protein–ligand interactions using protein crystallography', Crystallography Reviews, 11(1), 61-71.

Peh, W. Y. X., Reimhult, E., Teh, H. F., Thomsen, J. S. and Su, X. (2007) 'Understanding ligand binding effects on the conformation of estrogen receptor alpha-DNA complexes: A combinational quartz crystal microbalance with dissipation and surface plasmon resonance study', Biophysical Journal, 92(12), 4415-4423.

Praporski, S., Ng, S. M., Nguyen, A. D., Corbin, C. J., Mechler, A., Zheng, J., Conley, A. J. and Martin, L. L. (2009) 'Organization of Cytochrome P450 Enzymes Involved in Sex Steroid Synthesis protein-protein interactions in lipid membranes', Journal of Biological Chemistry, 284(48), 33224-33232.

Wand, A. J. (2001) 'Pressure as a tool for protein dissection', Biophysical Journal, 80(1), 30A-30A.

Warren, G. L., Andrews, C. W., Capelli, A.-M., Clarke, B., LaLonde, J., Lambert, M. H., Lindvall, M., Nevins, N., Semus, S. F., Senger, S., Tedesco, G., Wall, I. D., Woolven, J. M., Peishoff, C. E. and Head, M. S. (2006) 'A critical assessment of docking programs and scoring functions', Journal of Medicinal Chemistry, 49(20), 5912-5931.

Xu, J., Liu, K.-W., Matthews, K. S. and Biswal, S. L. (2011) 'Monitoring DNA Binding to Escherichia coli Lactose Repressor Using Quartz Crystal Microbalance with Dissipation', Langmuir, 27(8), 4900-4905.

York, D. M., Darden, T. A., Pedersen, L. G. and Anderson, M. W. (1993) 'Molecular modeling studies suggest that zinc ions inhibit hiv-1 protease by binding at catalytic aspartates', Environmental Health Perspectives, 101(3), 246-250.

Chapter 5 - Summary and Conclusions

In this thesis the protein-ligand interactions have been studied using computational and experimental techniques. As explained in chapter 1 the dewetting phenomenon has been fascinating over the past decade. The rise of the dewetting phenomenon in protein-ligand interactions (in particular MUP-IBM interactions) has been a system of interest in the past and therefore has been studied in this thesis. The use of MD simulations, QCM and AFM experiments (as described in chapter 2) has provided evidence for the hydration of the binding cavity of MUP on ligand (IBM) binding (chapter 3). The techniques have also provided valuable new predictions of the flexibility of MUP (chapter 4).

The dewetting of the binding site is a robust observation which can be validated from the results produced in chapter 3 and 4 of this thesis. For example the hydration of the protein binding cavity has been studied using MD simulations (RDFs, HDMs and water count) as well as QCM (by analysing the frequency curve using the Saubrey equation) experiments. The QCM hydration results suggest that approximately 2 water molecules are ejected from the binding site upon ligand binding to WT and A103S MUP, whereas no change can be observed for Y120F hydrophobic MUP.

Another interesting observation from the QCM experiments is that the small increase in dissipation on MUP addition suggests that MUP is a rigid protein. On the other hand MD simulations and the analysis from HDMs and RDFs suggest that even without the presence of the ligand the binding cavity of MUP has the most water molecules in the binding

cavity of hydrophilic A103S MUP and no water molecules in the binding site of hydrophobic Y120F mutant MUP. The WT MUP was found to have a small water molecule tightly bound to the tyrosine 120 of the binding site.

Overall the flexibility results from QCM (dissipation change) and AFM (height, adhesion, deformation) experiments suggest that on ligand binding the surface increases in flexibility for WT and mutant MUPs. This is in good correlation with preliminary AFM experiments. The MD simulation (atomic fluctuations) results on the other hand suggest that some residues of the protein increase in flexibility whereas other residues decrease in flexibility for the WT MUP-IBM interactions showing entropy-entropy compensation. The hydrophobic Y120F MUP-IBM interactions show an overall decrease in flexibility whereas the hydrophilic A103S MUP-IBM interactions show a general increase in flexibility with some residues showing a decrease in flexibility (entropy-entropy compensation observed once again).

It is still difficult to directly correlate results from experimental and computational studies as there are discrepancies between the two methods. For example QCM and AFM techniques are performed on a surface whereas MD simulations are not. Therefore, a possible future study could be to perform MD simulations of MUP-IBM interactions on a surface and then comparing the results to QCM and AFM experiments. The comparison of MD simulations (in particular ligand dynamics) to NMR labelling experiments can provide valuable insight to the reliability of computational methods. Long-time MD simulations (with the use of

replicates) have proved to be important and have shown some interesting results. The results support the previously short simulations of MUP-IBM interactions. However, by studying even longer simulations of MUP-IBM interactions can be useful. However, due to the lack of computation efficiency it may be useful to look at other techniques. For example accelerated MD (a recent method studied by Shaw et al., 2010) can be used to validate long-time MD simulations.
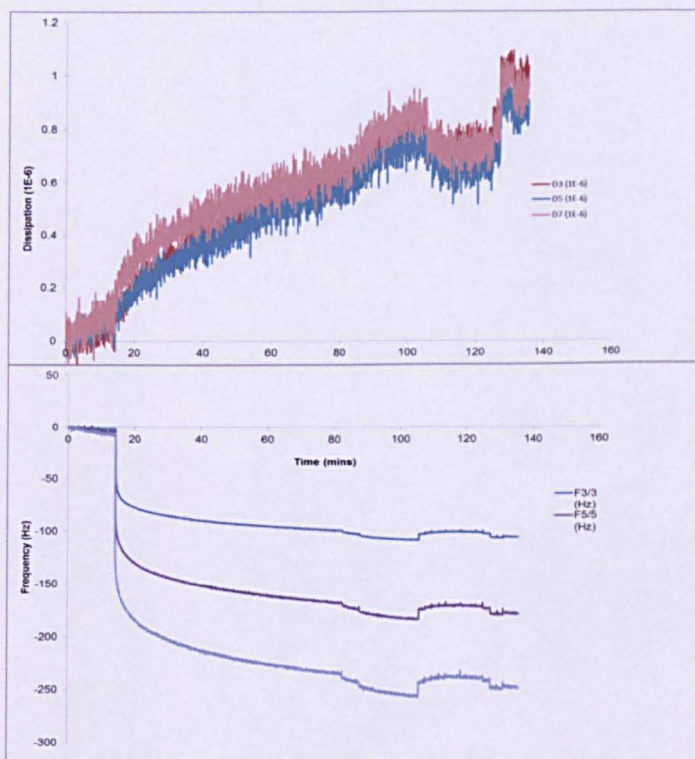
Chapter 6 - Appendixes

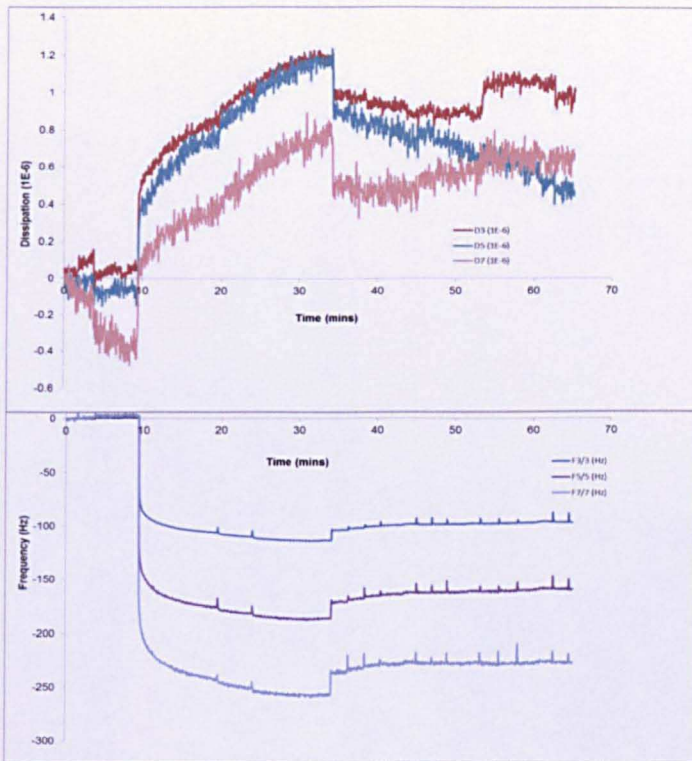## 6.1   QCM frequency-dissipation graphs

The frequency (bottom) and dissipation (top) graphs of each replicate are presented at all 3 different overtones (indicated by the 3 colours shown on each graph).
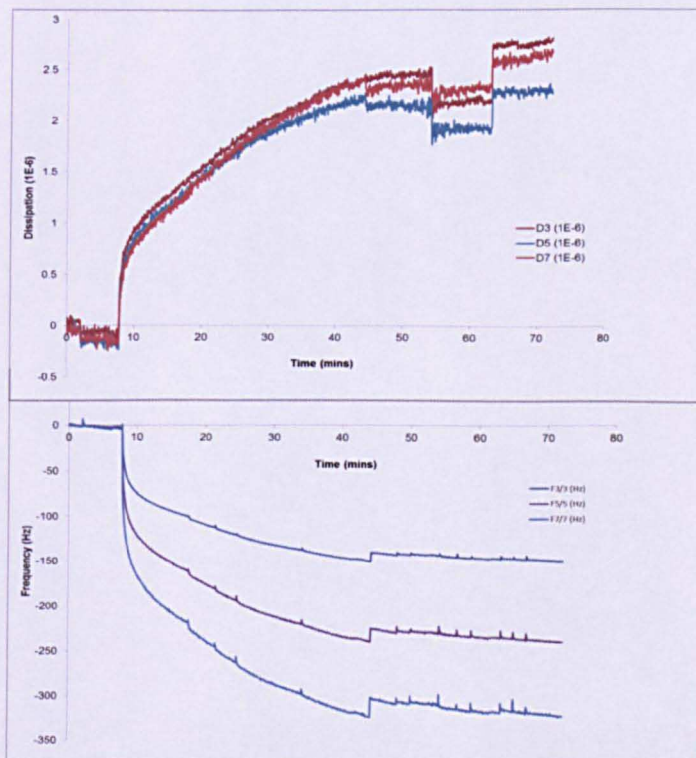
WT QCM graphs
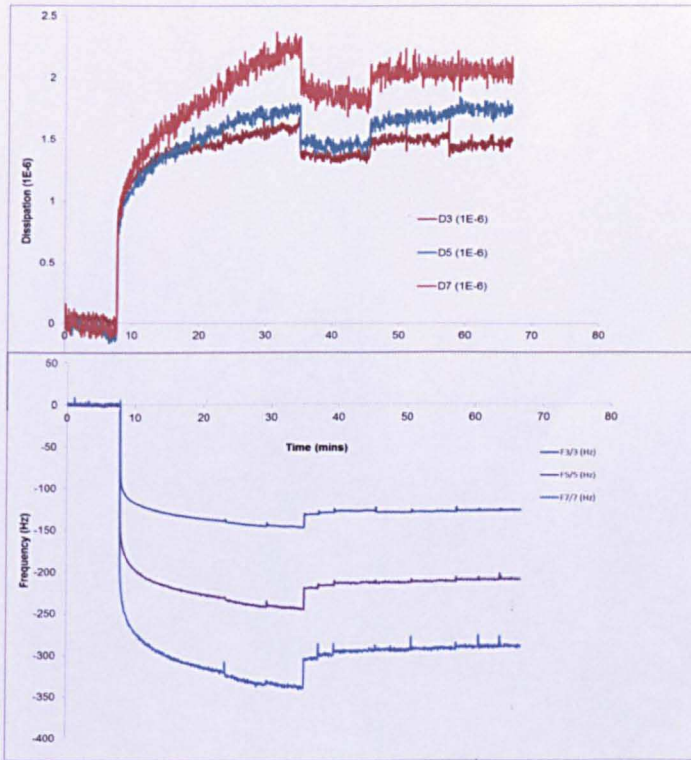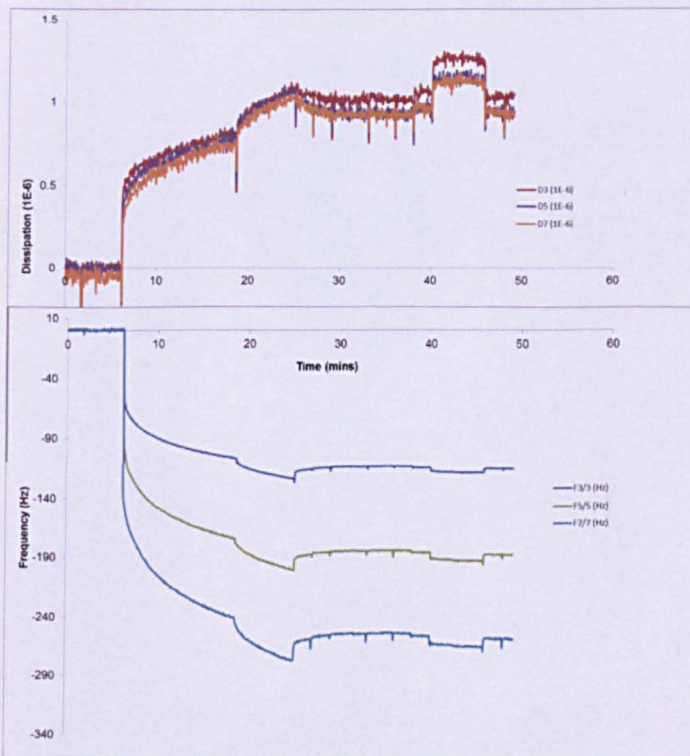
WT-Replicate 1

WT-Replicate 2
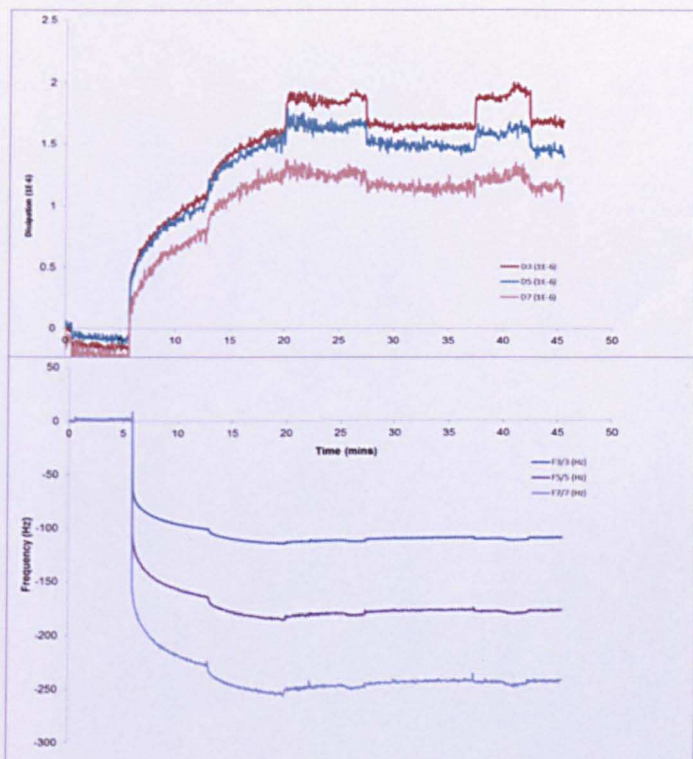

WT-Replicate 3

A103S QCM graphs

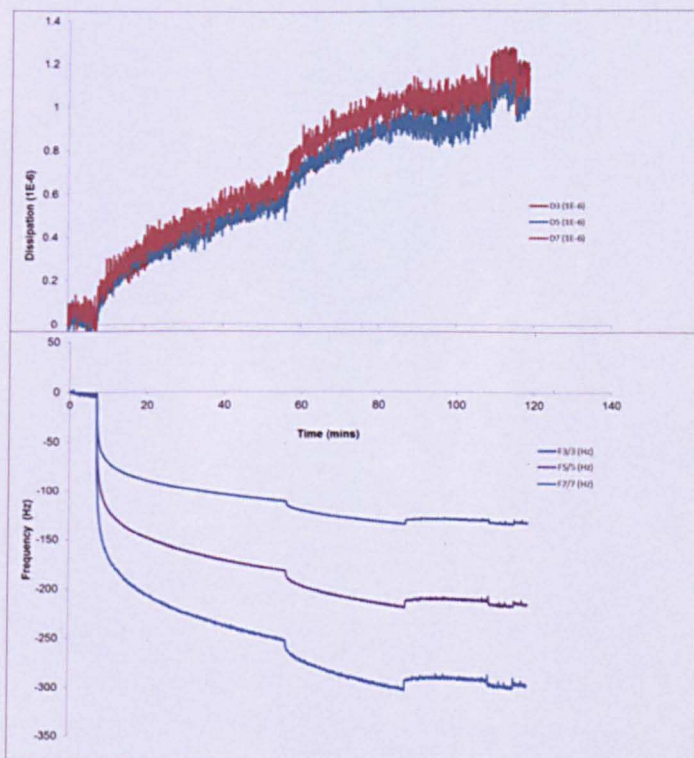## A103S-Replicate 1

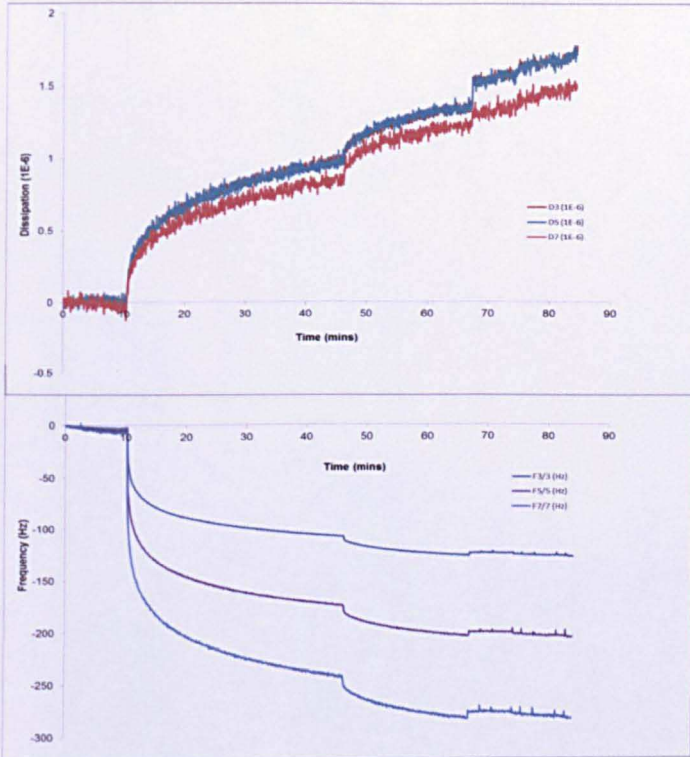## A103S-Replicate 2

A103S-Replicate 3
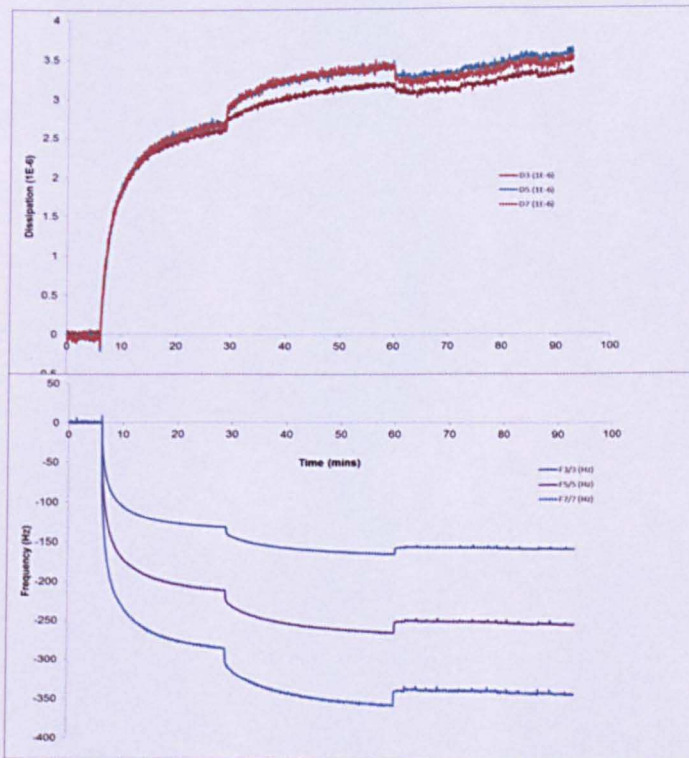

Y120F QCM graphs
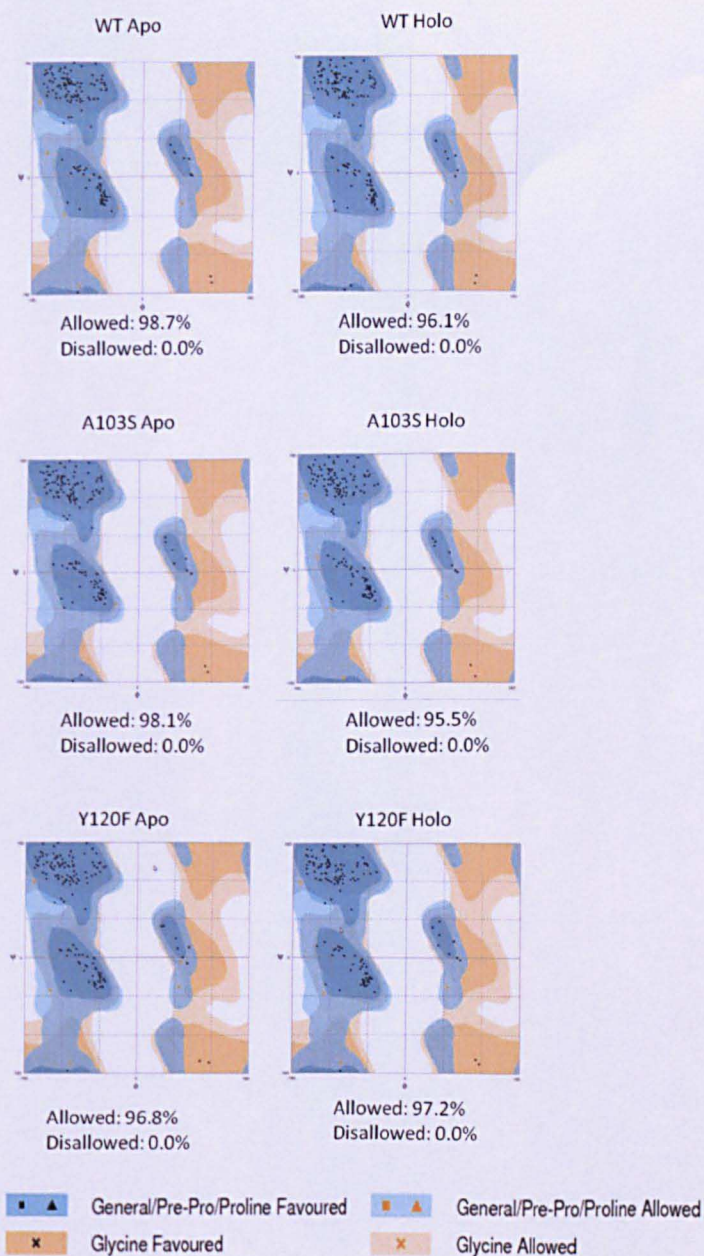
Y120F-Replicate 1

Y120F-Replicate 2

Y120F-Replicate 3

6.2   Ramachandrans plot analysis

Ramachandran's analysis was first run on the crystal structure PDB files for WT, A103S and Y120F MUP and their complexes with IBM. The phi (φ) and psi ($\emptyset$) and are indicated on the plots. The allowed and disallowed regions indicate the stability of the secondary structure (i.e. the higher the percentage for the allowed and lower the percentage for the disallowed region would indicate a more stable structure).

WT Apo   WT Holo

Allowed: 98.7%   Allowed: 96.1%
Disallowed: 0.0%   Disallowed: 0.0%

A103S Apo   A103S Holo

Allowed: 98.1%   Allowed: 95.5%
Disallowed: 0.0%   Disallowed: 0.0%

Y120F Apo   Y120F Holo

Allowed: 96.8%   Allowed: 97.2%
Disallowed: 0.0%   Disallowed: 0.0%

▪ ▲ General/Pre-Pro/Proline Favoured   ▪ ▲ General/Pre-Pro/Proline Allowed
×   Glycine Favoured   ×   Glycine Allowed

An average PDB files was saved every 10 picoseconds for each one microsecond trajectory for each replicate and a ramachandran's analysis was run on Prochecker (as indicated in chapter 2 and 3). The average allowed and disallowed regions are indicated in the table below and the results are discussed in chapter 3.

	Allowed	Disallowed
WT1	86.80	0.70
WT2	88.20	0.00
WT3	85.40	1.40
WTI1	87.50	0.70
WTI2	83.30	0.00
WTI3	85.40	0.00
	Allowed	Disallowed
A1	84.19	0.37
A2	85.04	0.80
A3	86.52	0.76
AI1	85.26	0.54
AI2	85.82	0.29
AI3	85.87	0.21
	Allowed	Disallowed
Y1	84.92	0.88
Y2	84.85	1.02
Y3	84.32	1.09
YI1	87.43	0.98
YI2	87.08	0.63
YI3	86.00	0.74

Key

A=A103s

AI=A103S-IBM

Y=Y120F

YI=Y120F-IBM

WT=WT MUP

WTI=WT MUP-IBM