Modelling Short Peptides in Solution and at the Interface

By

Daniel Andrew Cannon

Thesis presented to the Department of Pure and Applied Chemistry, University of Strathclyde, in fulfilment of the requirements for the degree of Doctor of Philosophy.

2016

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination, which has led to the award of a degree.

Signed:

Daniel Andrew Cannon

Date: 3rd May 2016

“The Lord said unto John “Come forth and receive eternal life”. But John came fifth and won a toaster”

- Deuteronomy, 25:11

Acknowledgements

Firstly, I would like to thank Dr. Tell Tuttle, I am forever grateful for his continuous support and guidance throughout my studies; indeed I could not have wished for a better mentor and supervisor.

I also extend my deepest gratitude to Prof. Rein Ulijn for his insight and encouragement in addition to his collaboration for the study presented in Chapter 1.

A special thanks goes to Prof. Nurit Ashkenasy of Ben-Gurion University of the Negev, for the deeply insightful discussions relating to both our collaboration, on which Chapters 2 to 4 are based, and the wider scope of my research.

Members of the Tuttle and Ulijn research groups past and present, many of you are my closest friends and I cannot express in words how much you have helped me.

I would be remiss if I did not explicitly thank the following people, whom have been so important to me over recent years:

Dr. Pim Frederix, thank you for helping me navigate the toughest moments of my studies and personal life, you mean more to me than I can say.

Dr. Iván Ramos Sasselli, always motivational, always patient, as wonderful a friend as I could ask for.

(Soon to be) Dr. Greg Anderson, my PhD brother. We started this journey together and I would not have reached the end without you.

Mari Paola Conte, your support and understanding gets me through the toughest times; thank you, it is a privilege to call you my friend.

Dr. Christopher Carson, whom I have known since we were seven years old, people come and go, but you have been constant, thank you.

My friends and family, without you I'm nothing.

Finally, I thank my parents; everything good in my life comes from you.

Abstract

The way in which neighbouring particles interact with each other is fundamental to all chemical phenomena, where the specific blend of intermolecular forces and geometric conformation dictate the properties of any system. This thesis discusses the importance of intermolecular interactions to phenomena witnessed for simple systems that may hark back to the origins of life, as well as complex systems where macromolecules display high affinity for solid surfaces.

In the first chapter, we explore the relationship between peptide sequence and structure. Using the example of catalytic peptides that are able to perform amide bond condensation, we examine the possibility of short peptide sequences adopting a conformation that supports an active site analogous to that found in protease enzymes such as subtilisin using molecular dynamics (MD). In this study we define a catalytic triad by convergence of an acid, histidine and hydroxyl residue below 4 Å. It was found that each of the three sequences studied could form a triad structure; however due to the flexible nature of the peptides this conformation was short-lived. We concluded that the catalytic activity of these peptides originated from their ability to form protease active site analogues and that experimentally observed K_{cat} rates, lower than that observed for enzymes, was a result of the flexible nature of the peptides.

In Chapter two, we examine the properties of single amino acids with solid surfaces. We use Au 111 as our model since there are numerous cases of peptides with a high affinity for gold. The aim of this study was to calculate the relative

binding free energies of nine amino acids that make up the gold binding peptide GBP1. Using Steered Molecular Dynamics (SMD) simulations we follow the change in free energy of each individual amino acid as it is pulled from the bound state to the bulk. We find that aromatic amino acids and those containing heteroatom side chains show greater affinity for the metal surface, by around 14 - 17 kcal mol^{-1}, compared to aliphatic amino acids. These findings provide a basis for creating and understanding the interactions between peptides and solid surfaces.

Following the study of individual amino acids with gold, we perform a comprehensive free energy study on a gold binding peptide, identified from combinatorial library experiments, in order to fully understand the influence of both electronic and conformational characteristics of peptides on the strength of interaction with the surface. By once again utilising SMD, we find that the interaction of the amino acid side chains with both the surface *and* the solvent dictate the affinity of GBPs for gold. The ability for a peptide to form strongly stabilising interactions with water will weaken its ability to bind to the surface as the adsorbed state becomes less favourable relative to the solution state. These findings demonstrate that design of peptides for binding to surfaces relies on a delicate balance between affinity for the solid and for the solution. Furthermore, we conclude that simple combination of single amino acid binding free energies do not provide sufficient insight into the affinity of the peptide for gold.

The fourth chapter focuses on conformational behaviour of GBPs under saturated conditions. Through increasing peptide concentration fourteen times, the way in

which peptides interact with both the surface and each other is investigated in order to obtain a realistic, high quality model of the system.

Table of Contents

List of Figures

List of Schemes

List of Tables

Education

2012 – 2016 **Ph.D. Thesis in Chemistry**
Title: Modelling Short Peptides in Solution and at the Interface.
Supervisors: Dr. Tell Tuttle and Prof. Rein Ulijn
University of Strathclyde, Glasgow

2011 **Honours Thesis in Theoretical Chemistry**
Title: Solving the Secrets of Synergic Synthesis
Supervisor: Dr. Tell Tuttle
University of Strathclyde, Glasgow

2006 – 2011 **Bachelor of Science**
Pure and Applied Chemistry
University of Strathclyde, Glasgow

Publications

1 Water-Assisted Tautomerisation of Hydroxylamines
D. Cannon, T. Tuttle, J. Koller, B. Plesničar
J. Phys. Chem. A, In preparation, **2016**

2* Crowding Effects on the Binding Properties of a Gold Binding Peptide
D. Cannon, N. Ashkenasy, T. Tuttle
J. Phys. Chem. B, In preparation, **2016**

3 Enzymatically Activated Emulsions Stabilized by Interfacial Nanofibre Networks
I. P. Moreira, M. Hughes, I. Ramos-Sasselli, **D. Cannon**, T. Tuttle, R. V. Ulijn
Soft Matter, **2016**, *12*, 2623

4*
Influence of Solvent in Controlling Peptide-Surface Interactions
D. Cannon, N. Ashkenasy, T. Tuttle
J. *Phys. Chem. Lett.* **2015**, *6*, 3944

5*
Discovery of Catalytic Phages by Biocatalytic Self-Assembly
Y. Maeda, N. Javid , K. Duncan , L. Birchall , K. Gibson , **D. Cannon**, Y. Kanetsuki, C. Knapp , T. Tuttle , H. Matsui.
J. Am. Chem. Soc., **2014**, *136*, 15893

6 Stabilization of Metastable Hydrogen Trioxide (HOOOH) and the Hydrotrioxyl Radical (HOOO•) by Complexation with Sulfuric Acid. A Theoretical Study
D. Cannon, T. Tuttle, J. Koller, B. Plesničar.
Comp. Theor. Chem., **2013**, *1010*, 19

7 Main Group Multiple C-H/N-H Bond Activation of a Diamine and Isolation of a Molecular Dilithium Zincate Hydride: Experimental and DFT Evidence for Alkali Metal-Zinc Synergistic Effects.
R. Campbell, **D. Cannon**, P. García-Álvarez, A.R. Kennedy, R.E. Mulvey, S.D. Robertson, J. Saßmannshausen, T. Tuttle.
J. Am. Chem. Soc., **2011**, *133*, 13706

Publications directly related with this thesis indicated with *.

Invited Talks

2014 Revealing the Conformational And Binding Properties Of Gold Binding with the Au 111 Surface
Functional Peptide Nanostructures Workshop
Ben-Gurion University of the Negev, Israel

2011 Solving the Secrets of Synergic Synthesis
ScotChem Computational Chemistry Symposium
The University of Edinburgh, UK

Posters

2015 ScotChem Computational Chemistry Symposium
University of Strathclyde, Glasgow, UK

2014 Nanopeptide 2015: International Meeting on Peptide & Protein based Materials and Technologies
University of Strathclyde, Glasgow, UK

2013 Modelling Molecules and Materials - M3
University of Strathclyde, Glasgow, UK

2012 Functional Peptide Nanostructures Workshop
Ben-Gurion University of the Negev, Israel

ScotChem Computational Chemistry Symposium
University of St. Andrews, UK

Nanopeptide 2012: Peptides as Nanomaterials & Biomaterials
The University of Manchester, Manchester Institute of Biotechnology, UK

ScotChem Computational Chemistry Symposium
University of Glasgow, UK

NAIS: State-of-the-Art Algorithms for Molecular Dynamics
International Centre for Mathematical Sciences, Edinburgh, UK

1 Catalytic Peptides

1.1 Minimal Biology

1.1.1 Introduction

From a chemist's perspective there are three fundamental components required for life: structure, molecular recognition, and reactivity; these three conditions are ubiquitous in all living systems in one way or another, but perhaps the best example of this is enzymes: biological catalysts. Enzymes are able to catalyse specific processes efficiently, where reactivity arises from recognition, arises from structure. The approach of minimal biology is to understand the fundamental aspects of each component, through the creation and evaluation of truncated analogues. These analogues may offer an insight into the basic chemical 'machinery' required for living systems and perhaps provide a glance at the origins of life.

To create analogues of proteins, minimal biology utilises amino acid building blocks, there are twenty genetically encoded amino acids that display a variety of chemical properties. All amino acids with the exception of glycine, contain a chiral 'alpha' carbon centre, however in biological systems, only the L-isomer is ever present. Figure 1.1 highlights the structural features of the amino acids and how they are typically characterised.

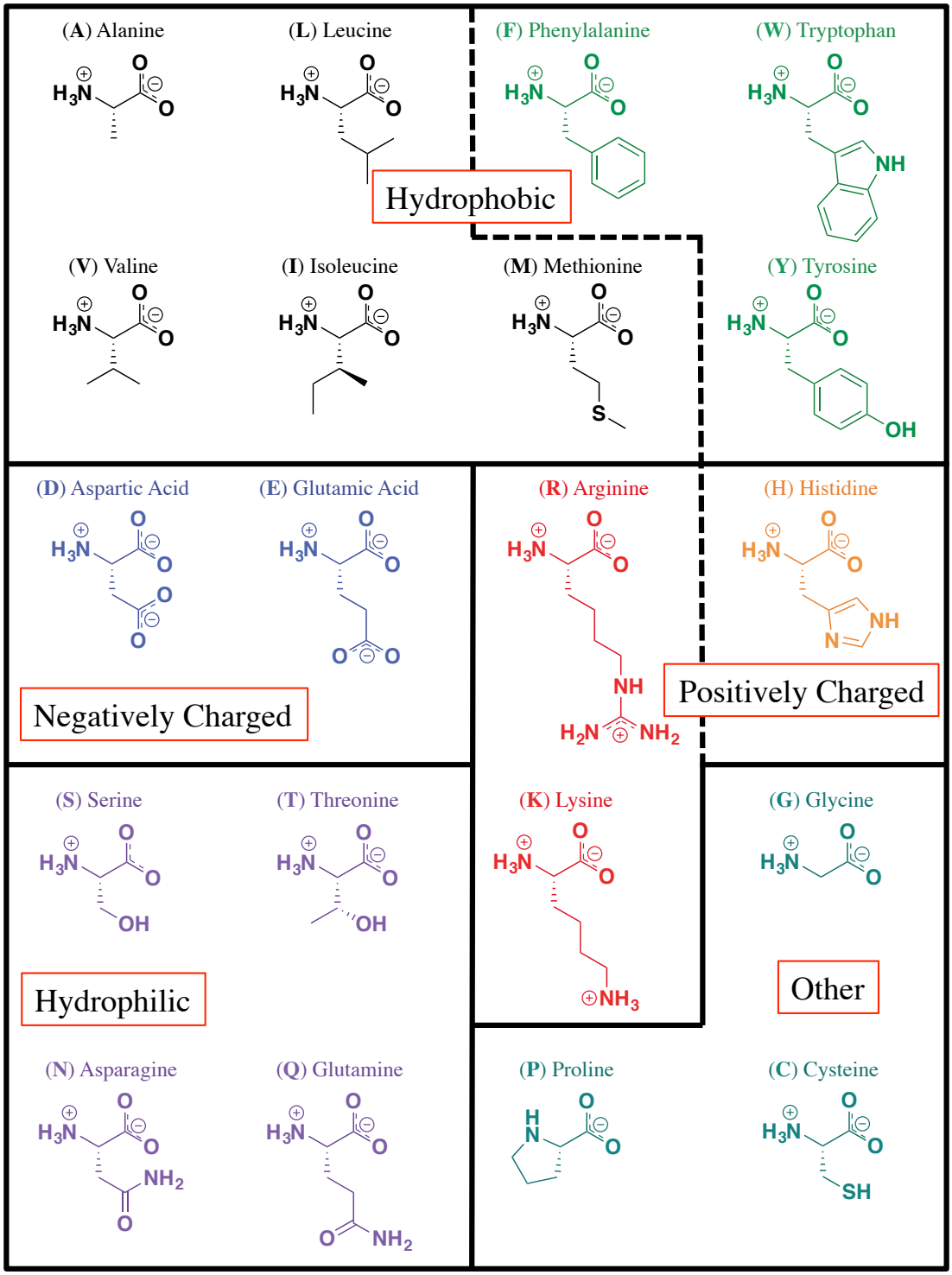
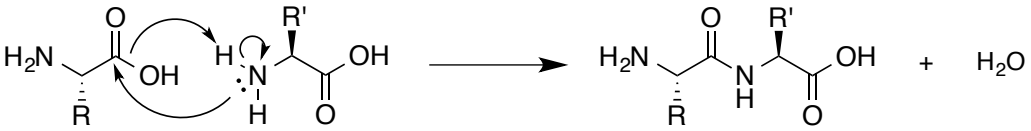


Figure 1.1. Chemical structures of 20 gene-encoded amino acids, grouped by chemical properties. (Single Letter Abbreviation) Name. Protonation state at pH 7 is indicated.

This remarkable chemical diversity leads to an effectively limitless design platform for minimal analogues. For example, in the case of a simple dodecapeptide, twelve amino acids in length, there are 4,096,000,000,000,000 different sequence possibilities, highlighting how vast the available chemical toolbox is. To build peptides, polypeptides and proteins, amino acids undergo a condensation reaction, where an amide bond is formed along with the release of water. As can be seen in Scheme 1.1, a bond is formed between the carbonyl carbon of the first amino acid and the amine nitrogen of a second amino acid. The hydroxyl moiety from the acid and a proton of the amine are released to form water.



Scheme 1.1. Mechanism of peptide bond formation via condensation (reverse hydrolysis).

This process is well understood and most peptide sequences can be produced cheaply and easily, therefore making the creation of minimal analogues of complex biological molecules a simple task.

1.1.2 Structure

The importance of structure in a biological context is best illustrated through protein folding. The folding of a chain of amino acids into a functional protein is dependent on a number of factors relating to both the chemical composition of the sequence and

the environment in which it is immersed. In living systems, correct folding of proteins is essential for efficient and selective function; misfolded structures can lead to diseased states[1-5] and thus a number of regulatory mechanisms exist to prevent their persistence.[6-8] Protein self-assembly is often viewed as extremely complex since the number of possible conformations is seemingly limitless;[9-10] however in reality this process happens quickly: in the order of microseconds.[11-15]

Folding propagates via various random conformations,[9-10] short regions of the sequence are 'programmed' to promote localised assembly[16-17] and small nucleation events that cause the protein to condense and drive the folding process towards the native structure.[17] A number of hypotheses exist as to the nature of protein folding pathways,[18-23] an example of the energy landscape for a protein folding process, illustrating the multiple routes assembly can take showing both shared and individual conformations, is shown in Figure 1.2.

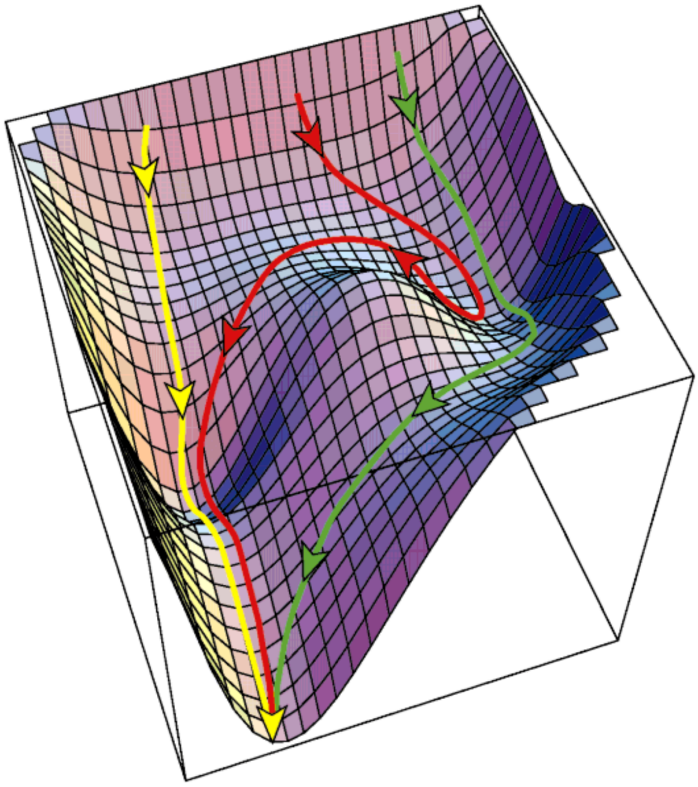


Figure 1.2. Estimated energy landscape for lysozyme folding with different folding pathways shown with arrows.[18]

Despite the apparent complexity, it is simple chemical factors that are coded within the protein sequence that give rise to its final form, first through generic interactions such as hydrophobicity then by more specific contacts like hydrogen bonds.[17, 24]

In an analogous approach to protein folding, the self-assembly of peptide amphiphiles is also sequence dependent. A review by Fleming and Ulijn[25] rationalised the relationship between chemical functionality and structure, within the context of short peptide amphiphiles, showing that minor changes in the chemistry leads to significant differences in the final self-assembled structure.

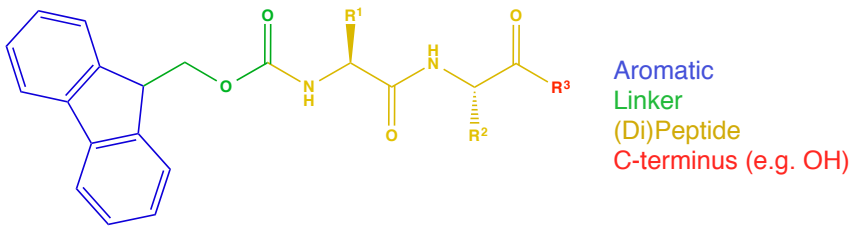


Figure 1.3. Generic form of peptide amphiphiles. Example shown is Fmoc di-peptide.

The chemical structure of peptide amphiphiles can be split into four regions: aromatic, linker, peptide and C-terminus. These species have become increasingly popular in the synthesis of supramolecular materials with tuneable and stimuli responsive properties as each of their four main moieties can be easily tailored to change the macroscopic properties of the final self-assembled material. The aromatic component serves to provide the hydrophobic region of the amphiphile whilst also promoting self-assembly through π-stacking interactions.[26-28] The linker segment can be modified to promote gelation by restricting conformational freedom of the amphiphile[29] and the C-terminus can be functionalised to modulate pH sensitivity[30] and direct structural properties[31] of the self-assembled structures. The diversity of amino acid functionality means that the peptide region is of particular importance and small changes can have a large impact on the final assembly. Some of the possible structural arrangements of peptide amphiphiles are shown in Figure 1.4.

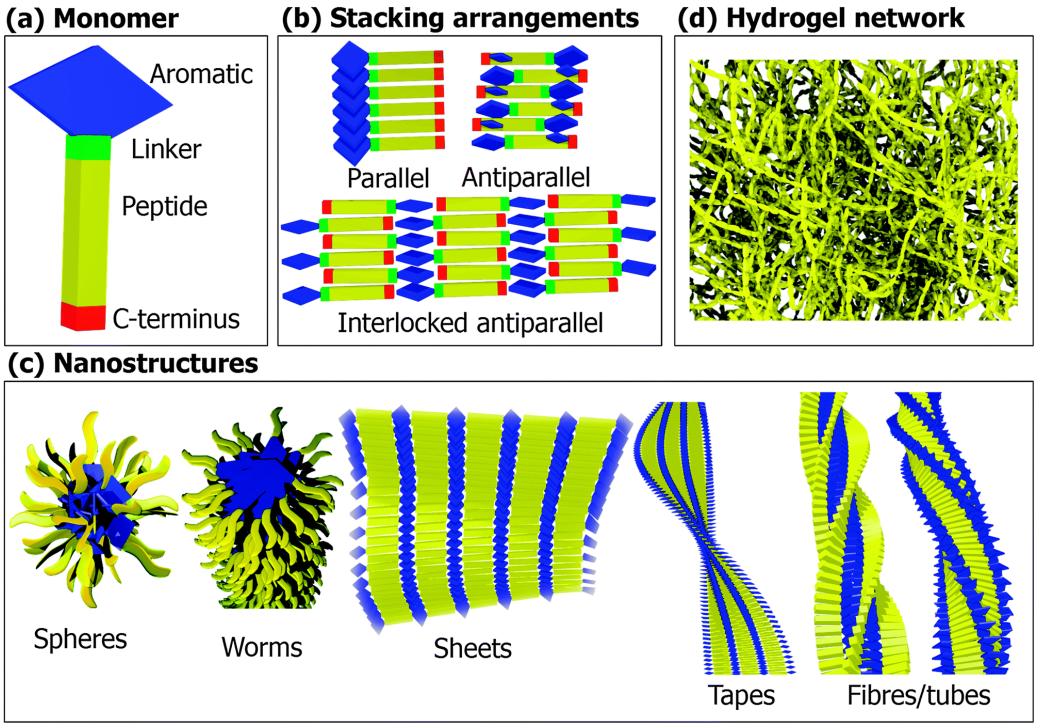


Figure 1.4. An overview of aromatic peptide amphiphile self-assembly; (a) a simplified aromatic peptide amphiphile; (b) some possible elementary stacking arrangements; (c) supramolecular nanostructures; and (d) a nanofibrous hydrogel network.[25]

One of the most common applications of peptide amphiphiles is the formation of hydrogels, which are the result of a network of self-assembled fibres becoming entangled and subsequently trapping water, Figure 4 d), and have been useful in areas such as drug delivery[32-34] and templating.[35-37] Hughes *et al.*[38] demonstrated the sequence/structure relationship of four peptide amphiphile hydrogelators: Fmoc-SF-OMe, Fmoc-SL-OMe, Fmoc-TF-OMe and Fmoc-TF-OMe, and their impact on material properties. Each of the four sequences gave rise to different morphologies, Figure 1.5; the SF derivative that gave rise to nanosheets and upon substitution of aromatic phenylalanine for leucine, resulted in the formation of planar belts. The

inclusion of threonine in the place of serine gave an extended fibre network in the case of Fmoc-TF-OMe, and short twisted fibres for Fmoc-TL-OMe.

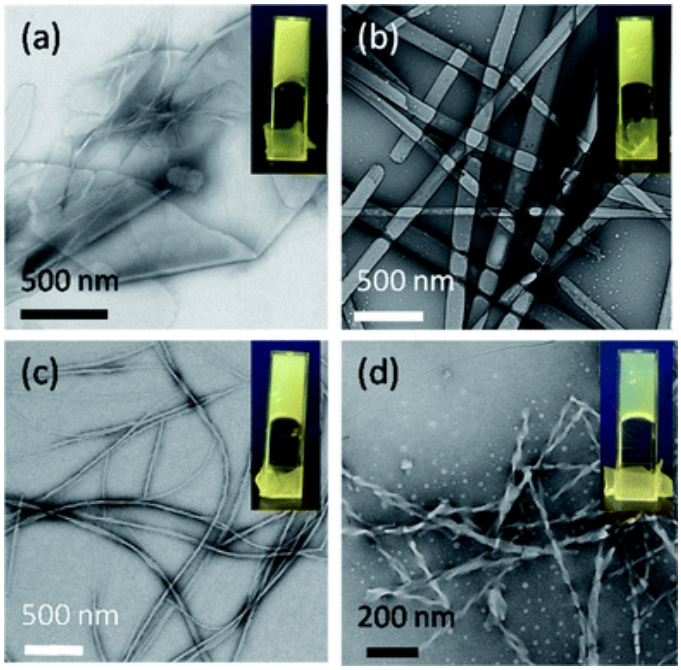


Figure 1.5. TEM micrographs of 20 mM samples of (a) Fmoc–SF–OMe, (b) Fmoc–SL–OMe, (c) Fmoc–TF–OMe, and (d) Fmoc–TL–OMe after 24 h. Inset: Images of materials at 24 h.[38]

Each hydrogel material was found to have different spectroscopic and rheological properties, highlighting the impact that small chemical modifications can have on structural properties. These subtle nuances drive structural arrangement and the high degree of organisation in proteins and peptide-based assemblies, ultimately leading to specificity and molecular recognition.

1.1.3 Molecular Recognition

Molecular recognition is the process by which specific non-covalent interactions between two molecules gives rise to a particular chemical response. The importance of this process is evident when we consider cell receptor molecules in living systems. G protein-coupled receptors (GPCRs)[39] are a series of transmembrane proteins found in eukaryotic cells and are responsible for many cell-signalling pathways.[40] GPCRs have a common bundle of seven α-helices connected by linker chains that make up their binding cavity, however subtle structural differences in the intracellular/extracellular chains alter receptor conformation[41-42] and subsequently the way in which ligands are arranged in the binding pocket.[43-46] Figure 1.6 highlights the structural similarity between bovine rhodopsin and β adrenergic receptors.

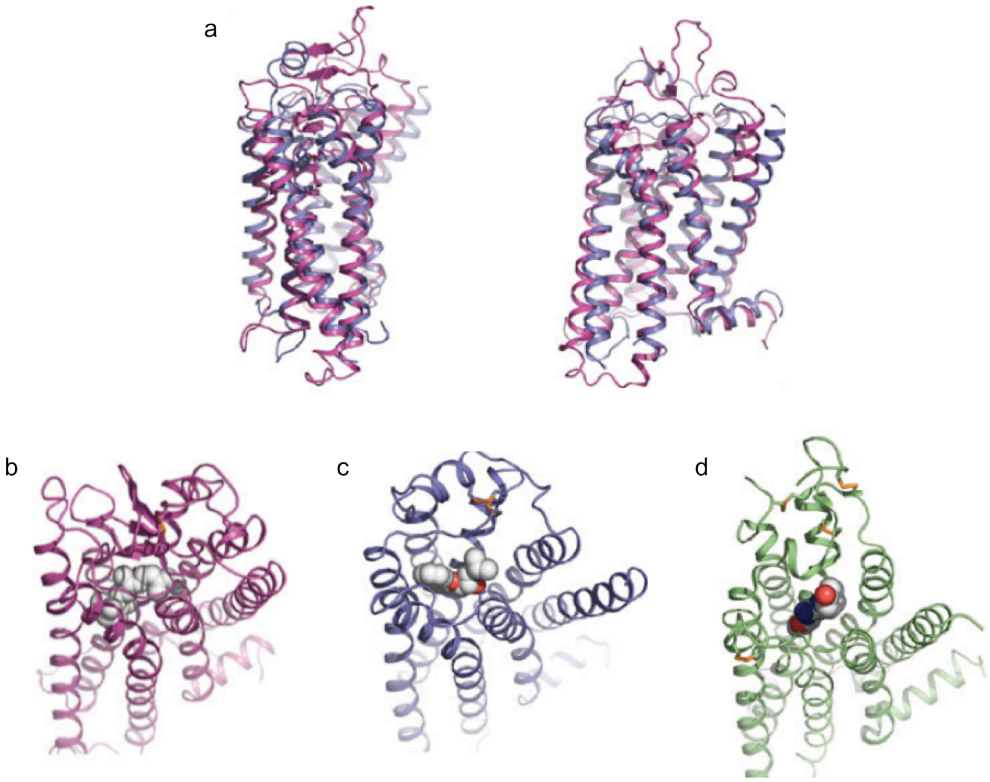


Figure 1.6. (a) Structural comparisons of Bovine rhodopsin (purple) superimposed on the human β2AR structure (blue). Extracellular views of (b) rhodopsin, (c) β2AR and (d) the A_{2A} adenosine receptor. The ligands are shown as spheres.[40]

Seemingly subtle structural variations give rise to a wide variety of signal pathways including phototransduction (Rhodopsin), muscle relaxation and contraction (β_1- and β_2-adrenergic receptors) as well as regulation of oxygen uptake and blood flow (Adenosine A_{2A}). The inherent similarity in structural arrangement of GPCRs means that *many* signalling molecules can successfully interact with an individual receptor, some may produce agonist-like response,[47-49] whilst others may elicit an antagonistic response. However, an individual GPCR's conformational setup will be biased toward precise molecular interactions with a specific agonist or antagonist, provoking the strongest response. The relationship between structure and signalling

pathway activity highlights how subtle differences in molecular interactions result in significantly different functional outcomes and hence its importance in living systems.

1.1.4 Catalysis

Both structure and recognition are closely linked and are the foundations of enzymatic catalysis. Enzymes are complex protein molecules, which act as highly specific biological catalysts to ensure reactions proceed efficiently in living systems. To highlight an enzyme's ability to enhance reactivity, a comparison of catalysed and uncatalysed reaction rates can provide an estimate of 'catalytic proficiency', defined by $(k_{cat}/K_m)/k_{non}$.[50]

In the case of peptide bond hydrolysis, the non-catalysed reaction of short polymer bound peptides in water have been shown to have a half-time in the order of ~7 years.[51] This process can also be catalysed under acidic and basic conditions[52] to the order of *days* or in the order of *hours* by metal complexes[53-54]. Enzymes however, such as Carboxypeptidase-A show a rate enhancement of 1.9×10^{11} and catalytic proficiency of 2.2×10^{15} for peptide bond hydrolysis, meaning a half-time in the order of *milliseconds*.[50] This extraordinary catalytic ability is characteristic of a number of enzymes for a number of different reactions, Table 1.1.

Table 1.1. Comparison of catalytic proficiency of a selection of proteins.[50]

Enzyme	Non-Enzymatic $t_{1/2}$		k_{non} (s^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} M^{-1})	Rate enhancement (k_{cat}/k_{non})	Catalytic proficiency [(k_{cat}/K_m)/k_{non}] (M^{-1})
OMP decarboxylase	78,000,000	y	2.8×10^{-16}	39	5.6×10^{7}	1.4×10^{17}	2.0×10^{23}
Staphylococcal nuclease	130,000	y	1.7×10^{-13}	95	1.0×10^{7}	5.6×10^{14}	5.9×10^{19}
Adenosine dearminase	120	y	1.8×10^{-10}	370	1.4×10^{7}	2.1×10^{12}	7.8×10^{16}
AMP nucleosidase	69,000	y	1.0×10^{-11}	60	5.0×10^{5}	6.0×10^{12}	5.0×10^{16}
Cytidine deaminase	69	y	3.2×10^{-10}	299	2.9×10^{6}	1.2×10^{12}	9.1×10^{15}
Phosphotriesterase	2.9	y	7.5×10^{-9}	2100	4.0×10^{7}	2.8×10^{11}	5.3×10^{15}
Carboxypeptidase A	7.3	y	3.0×10^{-9}	578	6.6×10^{6}	1.9×10^{11}	2.2×10^{15}
Ketosteroid isomerase	7.0	w	1.7×10^{-7}	66000	3.0×10^{8}	3.9×10^{11}	1.8×10^{15}
Triosephosphate isomerase	1.9	d	4.3×10^{-6}	4300	2.4×10^{8}	1.0×10^{9}	5.6×10^{13}
Chorismate mutase	7.4	h	2.6×10^{-5}	50	1.1×10^{6}	1.9×10^{6}	4.2×10^{10}
Carbonic anhydrase	5	s	1.3×10^{-1}	1000000	1.2×10^{8}	7.7×10^{6}	9.2×10^{8}
Cyclophilin, human	23	s	2.8 x 10-2	13000	1.5×10^{7}	4.6×10^{5}	5.3×10^{8}

From the examination of structural, sequential, topological and functional features of enzymes,[55] their comparison and classification has revealed certain patterns of catalytic moieties in a variety of enzyme families.[56] It was identified in the case of serine proteases, such as α-chymotrypsin,[57] trypsin[58-59] and subtilisin,[60-61] that an identical 'catalytic triad' of serine, histidine and aspartic acid was present.

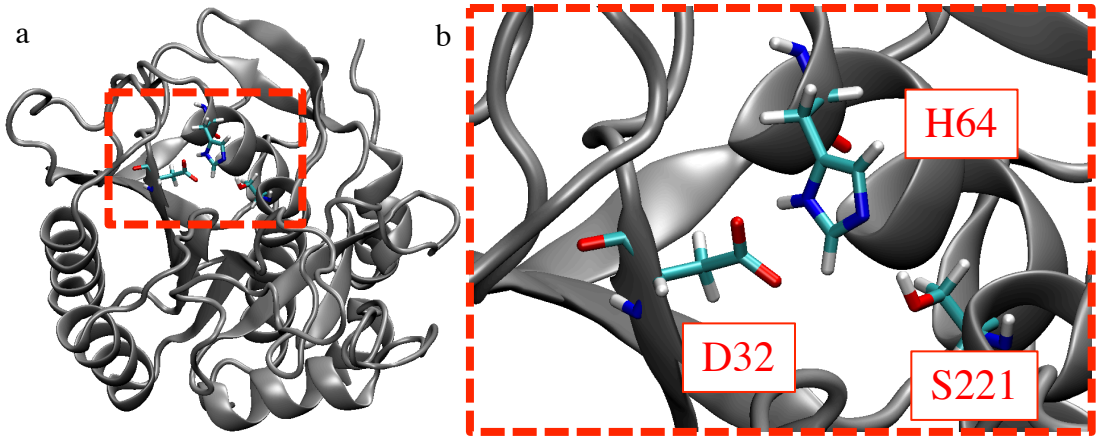
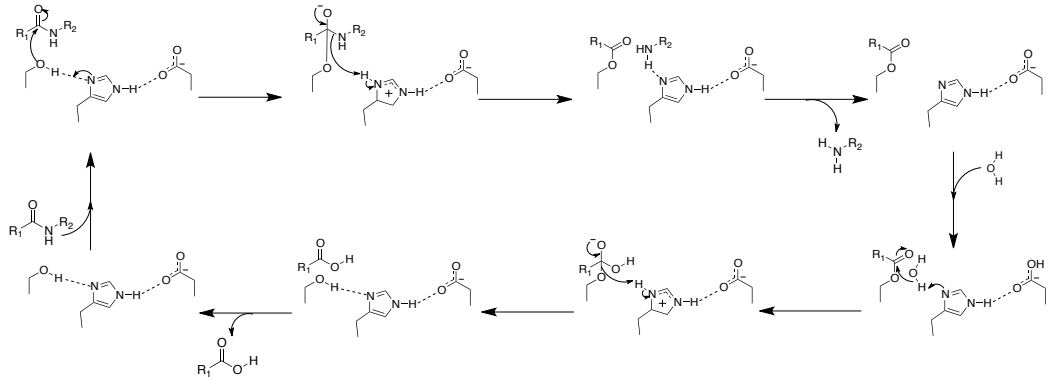


Figure 1.7. (a) Crystal structure of subtilisin (pdb: 1ST2) with catalytic triad residues highlighted (b) Active site triad residues (D32 H64 S221).

The proximity and orientation of these residues form a salt bridge that acts as a 'charge-relay system' for the activation of the serine oxygen.[62-63] In their native orientation, the interaction between aspartate and histidine creates a hydrogen bond that increases the basicity of the imidazole ring,[64] subsequently preparing the serine oxygen for electrophilic attack from the substrate upon the cleavage of the proton.



Scheme 1.2. Proposed mechanism of peptide bond hydrolysis in the context of the serine protease catalytic triad.[65]

It has been observed that while the hydrogen-bond interaction between histidine and aspartate is relatively strong in both the inhibitor-bound and unbound protease, the histidine-serine orientation promotes a weaker interaction.[66-67] Upon binding to a *substrate* however, the change in orientation favours the imidazole nitrogen and serine oxygen hydrogen bond,[62-63] consequently reducing the overall energy of the transition state, driving the reaction.

The structural continuity of active site residues throughout enzyme families indicates that electrostatic preorganization of the structure is a vital characteristic that gives rise to the impressive catalytic enhancement displayed by enzymes.[68] Enzymes are

the result of millions of years of evolution and it is possible that early precursors would have been significantly less complex and less efficient. Attempts have been made to create small enzymatically active species through computational design;[69-70] however this remains challenging. Identification of selective binders[71-74] on the other hand has become fairly easy through well-defined screening protocols based on vast libraries of proteins,[75] viruses[76-77] or antibodies.[78-79] In keeping with the theme of minimal biology, we ask the question – can the same principles used in selection of binders be applied to the selection of minimal analogues for enzymes? In other words, can we create catalytic peptides?

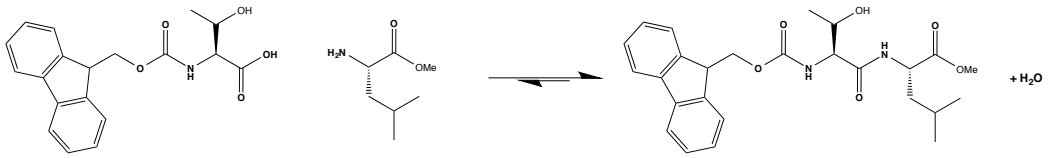
1.1.5 Minimal Catalysts

A key element of enzymatic ability relates to the large protein structure holding the active site residues in an arrangement that enhances catalysis, this structural conservation of triad arrangement is unlikely to feature in short, flexible peptides and hence catalytic ability is expected to be significantly lower. Interestingly however, the dipeptide SH is known to catalyse the hydrolysis[80-82] and reverse hydrolysis[83] of peptide bonds, suggesting that although a rigid scaffold is clearly advantageous, it is not critical. Despite the vast differences in secondary/tertiary structure observed between the serine proteases chymotrypsin and subtilisin, they are found to have remarkable similarity in function and identical active site conformations, suggesting that both divergent and convergent evolutionary processes have produced the same answer to the question of peptide hydrolysis. Therefore, the identification of catalytic peptides may provide an insight to the simple ancestors of complex enzymes.

1.1.6 Screening for Catalytic Function

One of the key aspects in identifying a minimal enzyme is the ability to isolate the catalytic candidate. In the past, catalyst-substrate reaction complexes were identified through localised colourimetric changes;[84] in this work however the first use of biocatalytic self-assembly, of which there are many examples,[84-86] as an immobilization mechanism is demonstrated.

In this study, the goal is to identify short peptide sequences that can catalyse the formation of a peptide bond, hence the reaction between soluble gel precursors Fmoc-T and L-OMe that can undergo condensation to produce high yields of the Fmoc-TL-OMe gelator was chosen.



Scheme 1.3. Reverse-hydrolysis reaction between non-gelators Fmoc-T and L-OMe to form the gelator Fmoc-TL-OMe.

In this work a recently developed technique that combines peptide functionalized phage libraries and catalytic control of molecular self-assembly was utilised. Five copies of a twelve amino acid peptide are expressed on the end of an M13 bacteriophage particle; the library of M13 phages contains approximately 10^{10}

unique peptide sequences. Phage particles are incubated in the gel precursors to give Fmoc-TL-OMe, which self-assembles to create a hydrogel network. If any of the peptide sequences presented on the phage particles can catalyse reverse hydrolysis to produce the gelator species, localised gel formation occurs and the catalytic phage particles are trapped. The heavy, trapped phage/gel aggregates are separated from non-catalytic particles via centrifugation. The aggregates are incubated with subtilisin to hydrolyse the gel and upon *E.* coli amplification of the phage particles, the catalytic peptide sequence can be identified from DNA sequencing of the viral genome.

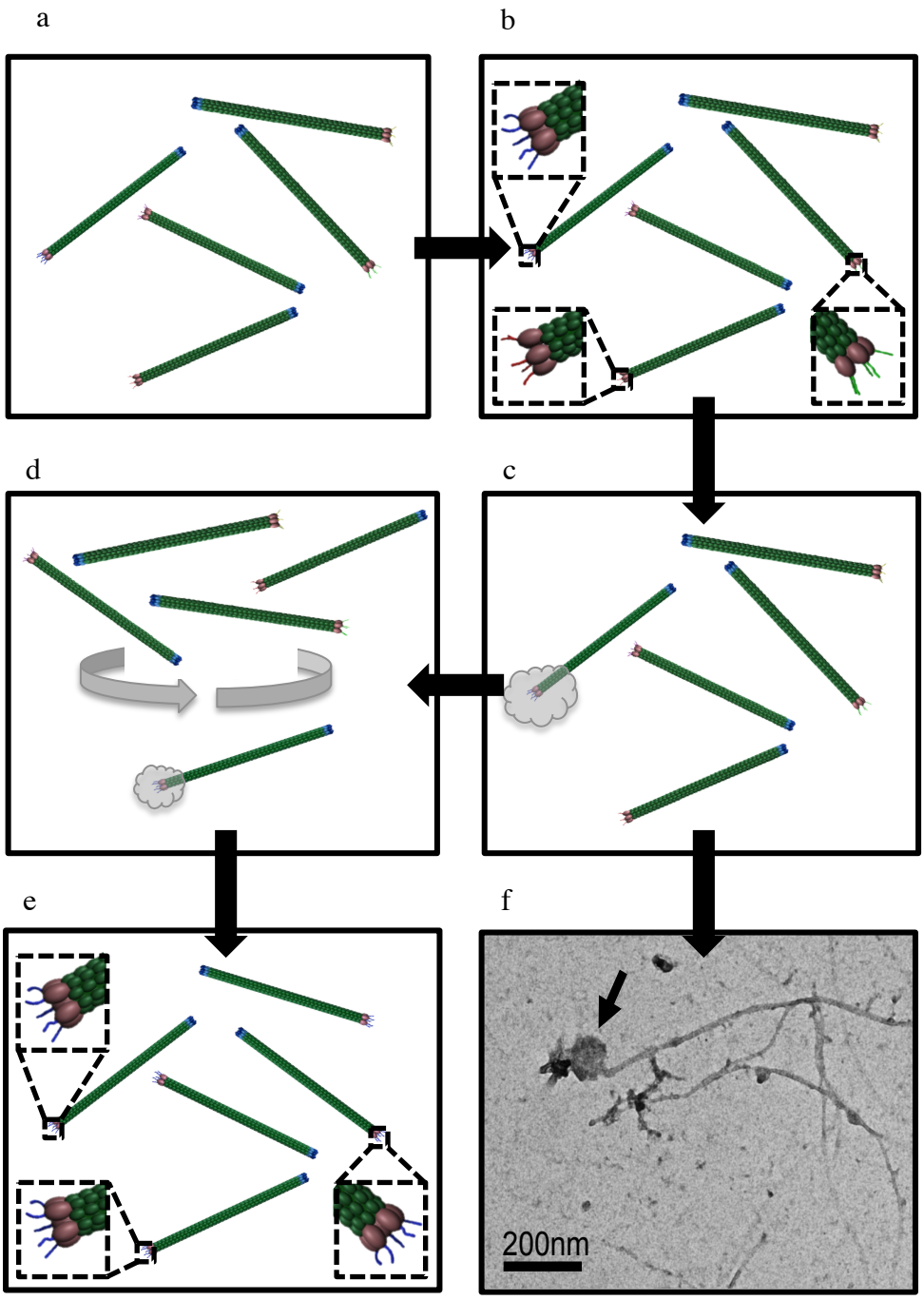


Figure 1.8. Schematic of phage panning process. a – b) From a large library of unique M13 phage particles, c) some dodecapeptide sequences can catalyse amide bond condensation to produce the gelator unit Fmoc-TL-OMe. The Fmoc-TL-OMe units can self-assemble into fibres that form hydrogels. f) Phage particles containing catalytic peptide sequences show localized gelation and d) can be isolated via

centrifugation. The gel that encapsulates the phage particle is then digested, meaning that through viral replication via E. coli (e) a library of catalytically active M13 phage particles is produced.

1.2 Experimental Results

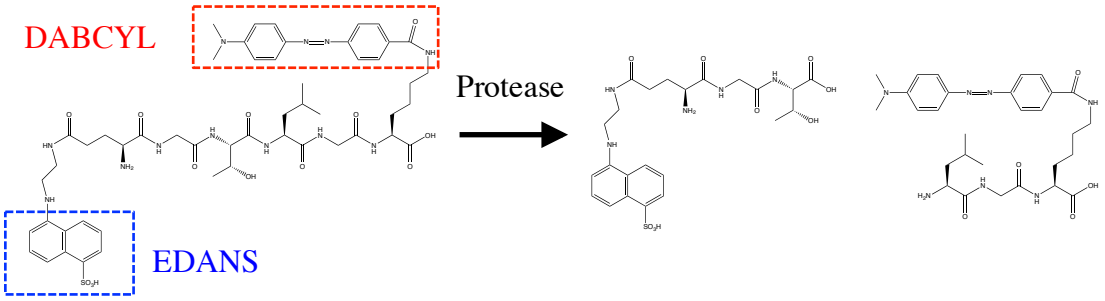
This project was carried out in collaboration with the groups of Prof. Rein V. Ulijn of the University of Strathclyde, Glasgow and Prof. Hiroshi Matsui of Hunter College, City University of New York, U.S.A. Dr. Yoshiaki Maeda carried out the phage display experiments discussed in this report. Dr. Yoshiaki Maeda, Dr. Louise Birchall, Dr. Nadeem Javid and Ms. Krystyna Duncan carried out the activity assays.

The phage panning process revealed three peptides that catalysed amide bond condensation. Although there is no apparent sequence similarity between these 'hits', it is apparent that the selected peptides contained amino acids that are typically associated with charge relay networks that enhance nucleophilicity, catalytic triads. Such triads consist of three precisely positioned and highly conserved residues: histidine (H), serine (S) and aspartic acid (D). All of the peptides identified contained at least one threonine/serine (T/S), and histidine (H), Table 1.2.

Table 1.2. Catalytic peptide sequences CP1 to CP3, identified from phage panning experiments.

Sequence	1	2	3	4	5	6	7	8	9	10	11	12
CP1	S	H	Q	A	L	Q	E	M	K	L	P	M
CP2	T	D	H	T	H	N	K	G	Y	A	N	K
CP3	S	M	E	S	L	S	K	T	H	H	Y	R

To examine whether the dodecapeptides retained catalytic activity when free in solution (i.e. not attached to the phage filaments), they were produced by solid-phase peptide synthesis; their ability to catalyse amide bond hydrolysis was examined using a Förster resonance energy transfer (FRET) assay, described in Scheme 1.4.



Scheme 1.4. Förster resonance energy transfer (FRET) assay of EGTLGK peptide with covalently bound chromophores 5-((2-aminoethyl)amino)naphthalene-1-sulfonic acid (EDANS) highlighted in blue and 4-((4-(dimethylamino)phenyl)azo)benzoic acid (DABCYL) in red. Mechanism shown indicates amide bond hydrolysis.

In a peptide based FRET assay, two chromophores – a donor and an acceptor, i.e., EDANS and DABCLY, are covalently bonded to the terminal side chain residues of

a short peptide. The adsorption spectra of the acceptor must overlap the emission of the donor so that energy transfer can occur. Upon excitation and subsequent relaxation of the EDANS molecule, a virtual photon transfers to the DABCLY acceptor through a non-radiative process, effectively quenching the emission of the donor. The efficiency of the quenching is sensitive to distance and decreases at a rate of $R_O^6 / (R_O^6 + r^6)$, where R_O is the Förster radius where half the energy is transferred and r is the separation. In the intact peptide, the close proximity of donor and acceptor facilitates efficient energy transfer; however upon cleavage of the peptide this separation increases, lowering efficiency until at distances greater than *ca.* 100 Å virtual photon energy transfer cannot occur and quenching ceases.

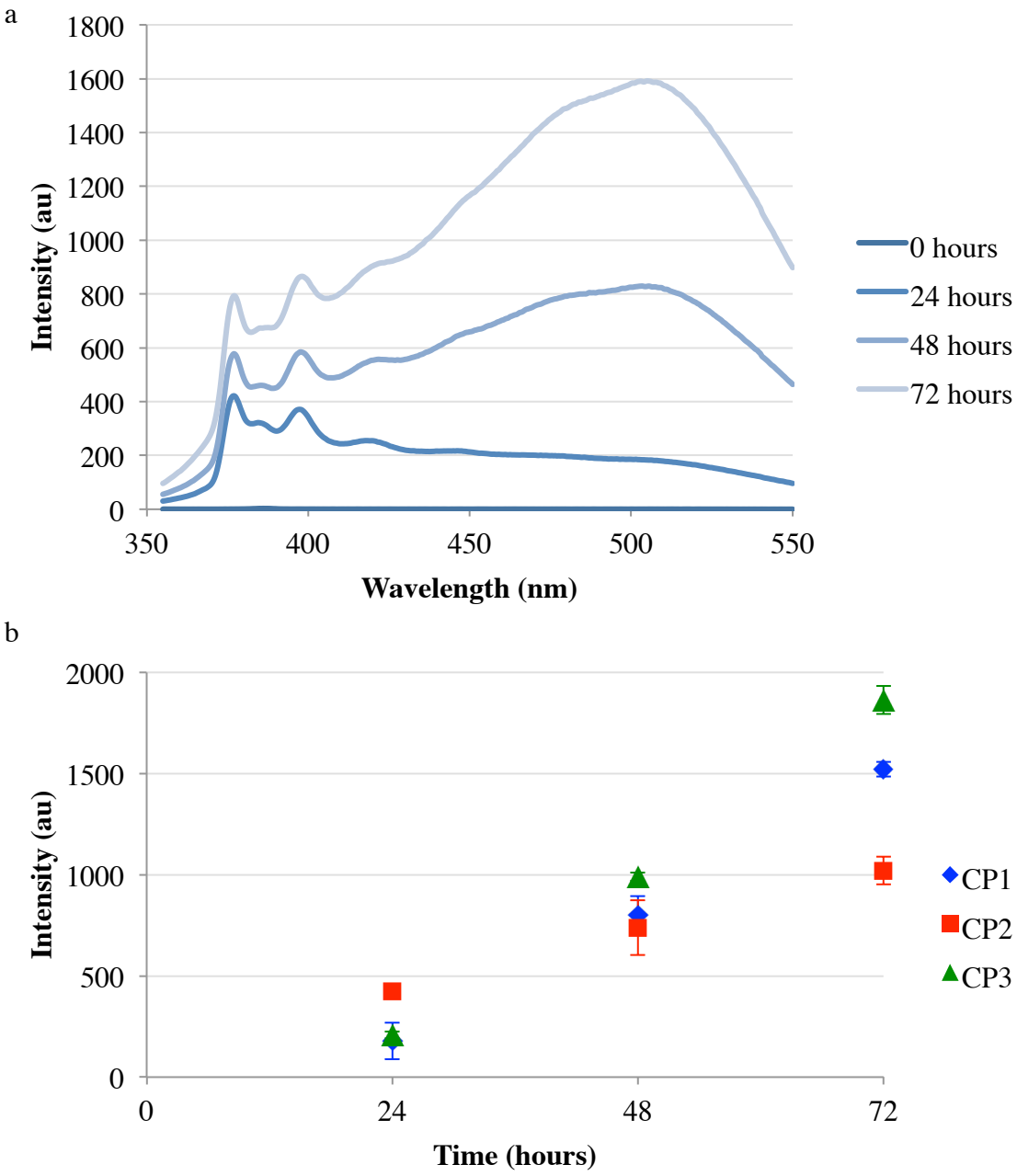


Figure 1.9. a) Fluorescence emission of FRET peptide at 24 hour time intervals upon incubation with CP1. b) Emission intensities over time for catalytic peptide FRET assays.

The main fluorescence emission peak for the EDANS is found at 493 nm, Figure 1.9 a), which can be plotted over time Figure 1.9 b); from the gradient of peak intensities over time, a catalytic rate constant, k_{obs}, can be calculated, Equation 1.1.

$$k_{obs} = \left(\frac{m_{sample}\ (hr^{-1}\ )}{(m_{calibration}\ (nM^{-1}\ ) \times [Catalytic\ Peptide]\ (nM)} \right) \quad \text{Equation 1.1}$$

Where m is the gradient of the peak intensities over time, $m_{calibration}$ is the gradient of the calibration curve and represents the molar extinction coefficient of the process.

Comparison of catalytic rate constants shows that **CP3** has the greatest amidase activity of the selected peptides, Figure 1.10, which may arise from the greater number of triad residues (6 triads = 3 Ser x 2 His x 1 Glu) that may adopt a catalytic conformation. However, the **CP1** peptide, which contains only one possible triad (S1 H2 E7), shows comparable activity to **CP3** whereas **CP2,** which may adopt 4 triads, shows lower activity; thus catalytic activity is not related directly to the propensity of catalytic residues but perhaps their ability to adopt a structure that facilitates catalysis.

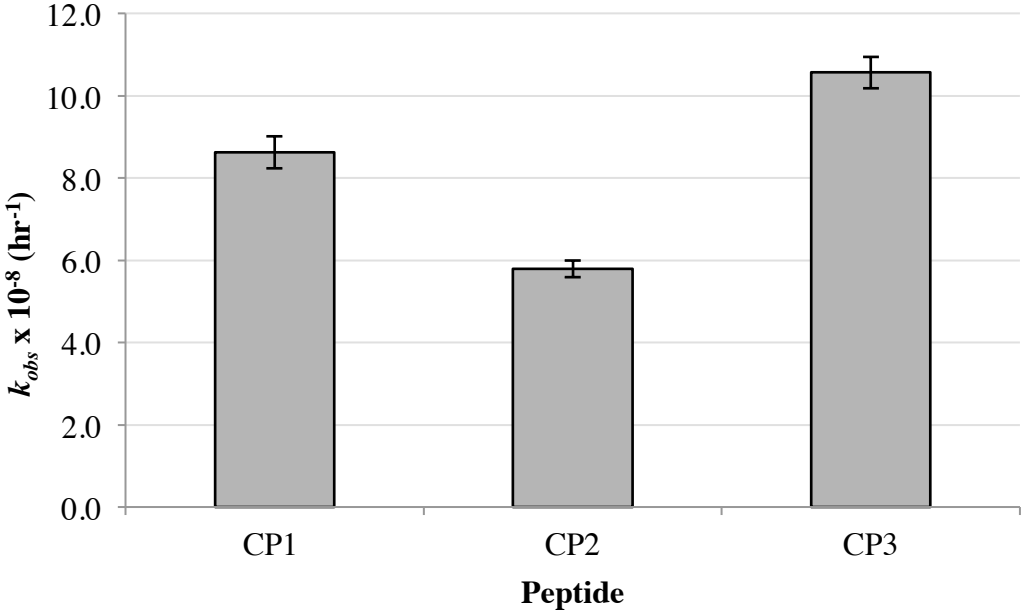


Figure 1.10. k_{obs} values for catalytic peptides based on FRET assays.

1.3 Computational Methods

1.3.1 Background

In this study the conformational properties of catalytic peptides **CP1** to **CP3** are examined to reveal the possible mechanisms by which a catalytic triad, analogous to that found in protease enzymes such as subtilisin, Figure 1.7, may form and thus catalyse amide bond hydrolysis and condensation as observed catalytic gelation and FRET experiments.

1.3.2 Theoretical Basis

One of the most prominent methods of studying biomolecules computationally is using molecular mechanics. Molecular mechanics (MM) is a method of calculating molecular structure and properties based on Newtonian mechanics, where the potential energy of a system is calculated as a function of its conformational properties. This relationship is given by a mathematical algorithm known as a force field, the form of which is shown in Equation 1.2.

$$V(r^N) = \sum_{bonds} \frac{k_l}{2}(l_i - l_0)^2 + \sum_{angles} \frac{k_\theta}{2}(\theta_i - \theta_0)^2 + \sum_{torsions} \frac{V_N}{2}(1 + \cos(n\tau - \phi))^2 + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \quad \text{Equation 1.2}$$

The force field can be broken down into five components, the first being the description of bonded interactions, where k_l is the bond stretching constant, l_i is the bond length and l_0 is the reference bond length. Analogously, a harmonic potential is also used to describe the potential energy resulting from angle bending where k_θ is the angle bending constant, θ_i is the angle and θ_0 is the reference angle. The third term describes torsional motions using a periodic function. V_N is the barrier height of the torsional potential, n is the multiplicity, τ is the torsion angle and φ is the phase factor. The remaining two components of the force field represent the non-bonded interactions of the system. Van der Waals forces are given in the form of a 12-6

Lennard-Jones potential between two neighbouring particles i and j, where σ_{ij} is the separation where the inter-particle potential is zero and r_{ij} is the distance between the centres of atoms i and j. Charge-charge interactions are represented by Coulomb's law, where q_i and q_j are the partial atomic charges of i and j and r_{ij} is the particle separation.

MM is commonly used to calculate the potential energy of a system in a particular conformation. However, as interesting biomolecular processes happen over time, it is necessary to use dynamics to fully understand systems of this nature. Molecular Dynamics (MD) employs MM in a dynamic way by propagating particle positions and momenta through time, thus the time-dependent behaviour of the system can be studied.

The potential energy is a function of all atomic positions in the system, hence the force acting on a particle is the vector sum of its interactions with neighbouring particles. The force is given by the negative gradient of the potential energy.

$$F_i = -\nabla_i \mathrm{V}(r_i) \qquad \text{Equation 1.3}$$

Where F is the force, ∇ is the gradient of the potential energy V of particle i at position r_i. By calculating the force, the acceleration (a_i) of each atom can be calculated, where m_i is the mass of particle i.

$$a_i = \frac{F_i}{m_i}$$ Equation 1.4

Equation 1.3 and Equation 1.4 can be combined to give the differential equation:

$$\frac{dV}{dr_i} = \frac{d^2 r_i}{dt^2}$$ Equation 1.5

Where t is time.

Integrating Equation 1.5 over short time intervals, δt, will give a trajectory of the particles through time by combining the forces calculated at time t with the positions and velocities at time t+δt. Employing this methodology for complex systems is effectively impossible, instead approximating the acceleration, velocities (v) and positions (r) of the particles through the Verlet algorithm is commonly used. The Velocity Verlet algorithm is based on both forward and reverse time Taylor expansions, where the positions at time t+δt are determined from those at t and t−δt:

$$r(t+\delta t) = r(t) + v(\delta t) + \frac{1}{2}a(t)\delta t^2 \qquad \text{Equation 1.6}$$

$$v(t+\delta t) = v(t) + \frac{1}{2}[a(t) + a(t+\delta t])\delta t \qquad \text{Equation 1.7}$$

The value of δt must be less than the period of the fastest degrees of freedom in the system, usually bond vibrations. Solving these equations allows the system to be studied over time.

1.3.3 Construction of System

The first consideration when constructing each system was the ionic state of charged residues of each peptide. The simulations were carried out in water at neutral pH, meaning that aspartic and glutamic acid residues as well as arginine and lysine were assembled in their charged forms. Furthermore, in this study the N-terminus is considered to be in the charged NH_3^+ state, however as each peptide is attached by the C-terminus to the phage proteins in the panning experiments, a neutral amidated C-terminus is employed, to mimic experimental conditions as closely as possible and remove the possibility of the carboxylic acid terminus participating in triad formation.

The second aspect of peptide structure that is considered is the protonation state of histidine residues. At pH *c.a.* 7 histidine can exist in either a δ or ε *singly* protonated state,[87] therefore both isomers are considered for all peptides.

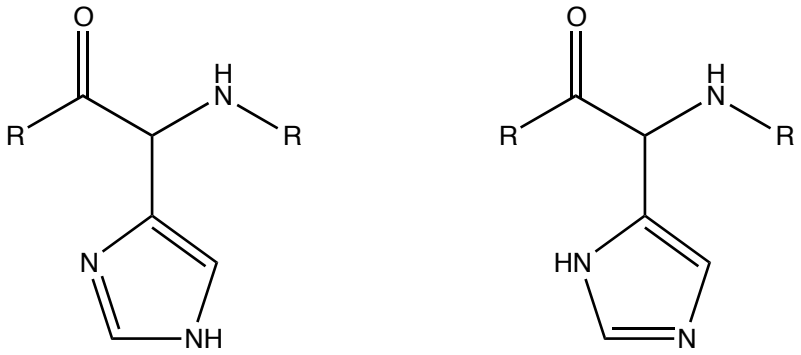


Figure 1.11. Two histidine isomers, where the ε (left) and the δ (right) nitrogen atoms of the ring are protonated.

In the case of **CP1** where there is only one histidine, a total of two systems are constructed where the H2 is built with either a protonated δ-nitrogen or ε-nitrogen on the imidazole ring. Consequently, **CP2** and **CP3** systems comprise of four simulations with δδ, δε, εδ and εε protonation state isomers considered.

Each peptide was solvated in a cubic box of TIP3P[88] water using the SOLVATE[89] module of VMD.[90] The size of the box was tailored to each system, in each case the box was aligned with the origin of the peptide and allowed a minimum spacing of 6 Å from the edge of the box. This resulted in a typical box size between 49 Å x 49 Å x 49 Å and 57 Å x 57 Å x 57 Å comprising of ca. 11000 to 18000 atoms. For peptides with a non-zero total charge a Cl^- ion was added to neutralize the system.

MD simulations were carried out for a period of 50 ns, using the NAMD[91] program with the CHARMM27[92] force field. A short minimization period was employed at the start of each simulation to remove bad contacts. The simulation was run at 310 K with the NPT ensemble and with periodic boundary conditions. The Particle-Mesh-Ewald[93] method was employed for full electrostatics, with a grid spacing of 1 Å. Covalently bonded hydrogen atoms were kept rigid using the SETTLE[94] extension of the SHAKE[95] constraint algorithm. Each simulation was run at 2 fs time steps for a total of 25,000,000 steps, giving a model simulation of 50 ns.

1.3.4 Analysis of System

In hydrolases, the catalytic mechanism involves a charge-relay network between an alcohol bearing amino acid (S or T), histidine (H), and an acidic amino acid (D or E). In the case of the serine protease subtilisin (PDB-ID: 1ST2), Ser-221 forms a hydrogen bond (3.21 Å) from the alcohol O(H) to the *non-protonated* nitrogen of His-64, which is also connected *via* a hydrogen bond from the *protonated* N(H) to the (C)OO^- of Asp-32 (2.58 Å and 3.37 Å, respectively).

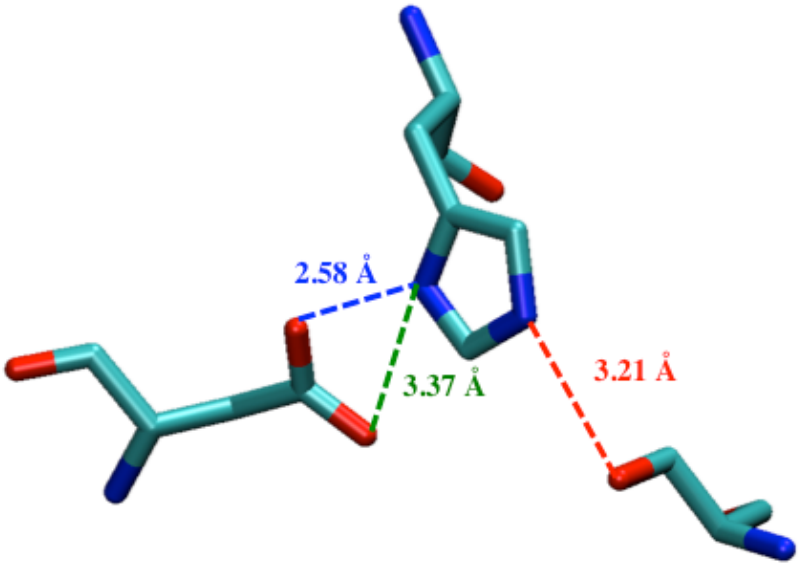


Figure 1.12. Subtilisin active site with key triad distances between heavy atoms indicated. Hydrogen atoms omitted for clarity.

To establish if an analogous triad is formed in the course of the MD simulations, the proximity of hydroxyl oxygen atoms with non-protonated imidazole nitrogen atoms and carboxylic acid oxygen atoms with protonated nitrogen atoms of the histidine was monitored. We defined the formation of a triad as when both the O(H)-N and either of the (C)OO¯N(H) distances are less than or equal to 4 Å.

Each peptide was able to temporarily fold into a conformation that allowed the catalytic triad to form. It was noted that there was no correlation between the triad formation and the energy, meaning that the peptide only adopts this conformation briefly and sporadically.

1.4 Results and Discussion

From the 50 ns simulation of the delta protonated histidine isomer of **CP1**, it was found that the peptide could adopt a triad structure analogous to that of subtilisin, Figure 1.13 a); the only possible triad – S1 H2 E7 – was found to form for only *ca.* 0.001 ns, Figure 1.13 b).

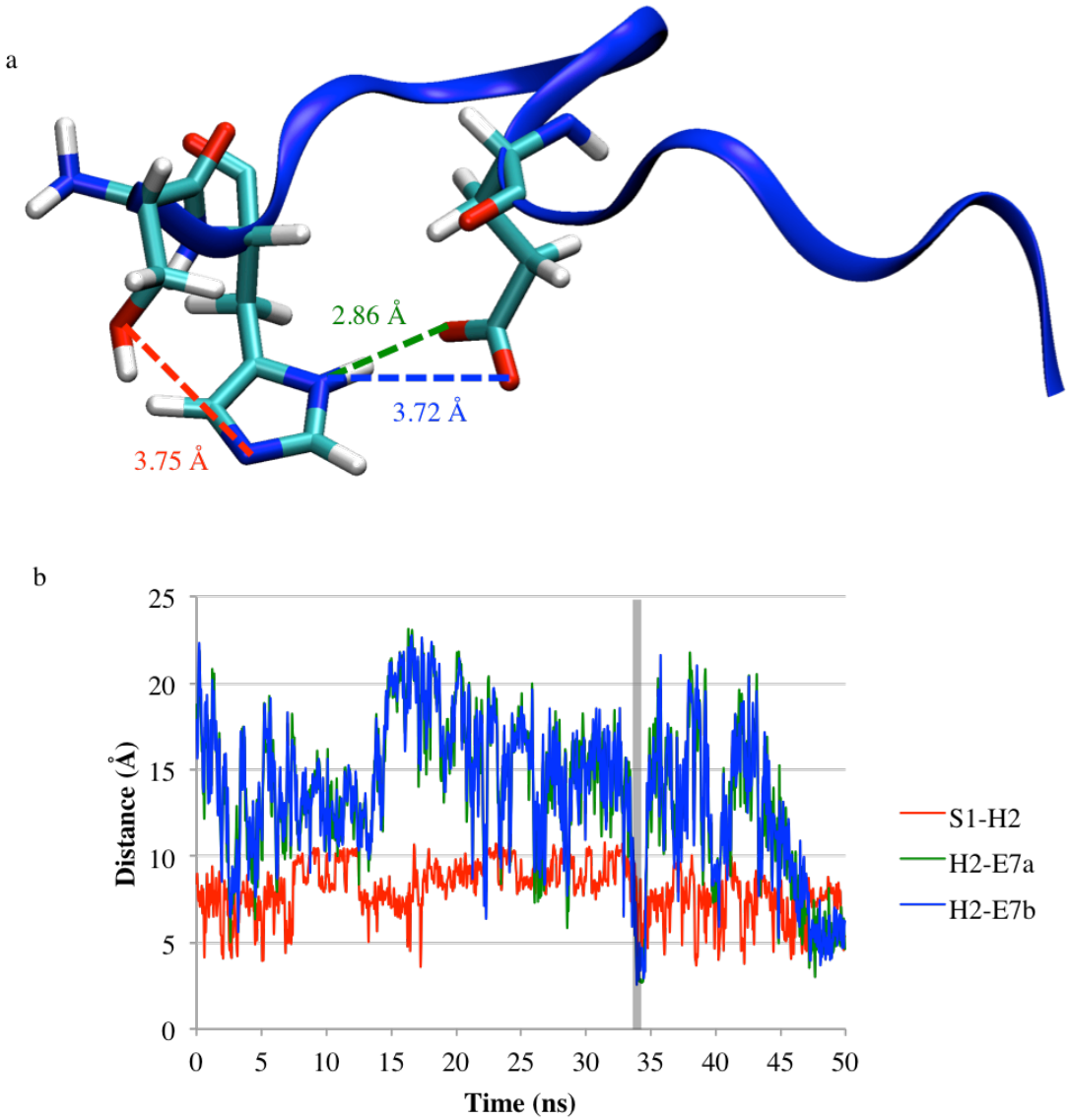


Figure 1.13. a) Exemplar S1-H2-E7 catalytic triad structure of CP1-H2δ. b) Key distance variation over time, triad formation period highlighted.

As proposed, the absence of an extended structural scaffold results in significant flexibility of the peptide, Figure 1.14; thus a triad structure is only formed sporadically and resulting amidolytic activity is much lower than a structure with a conserved catalytic motif.

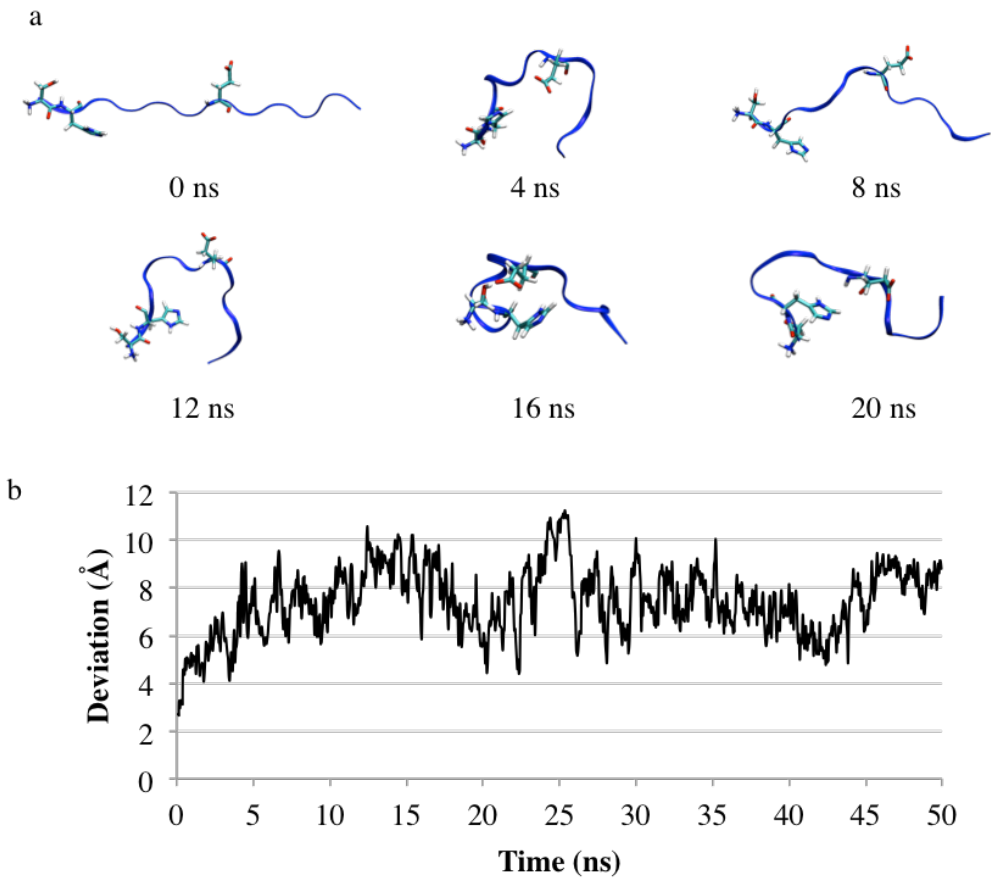


Figure 1.14. a) CP1-H2δ structure over time, catalytic residues shown explicitly. b) Deviation of CP1-H2δ relative to and aligned with the starting conformation.

Furthermore, there is no correlation between the formation of the triad and the relative energy of the peptide, i.e., there is no energy minimum when the triad forms; therefore there is no energetic penalty associated with the dissociation of the triad. This was quantified this through comparison of the average energy deviation when the triad forms with the standard deviation of the peptide during the simulation.

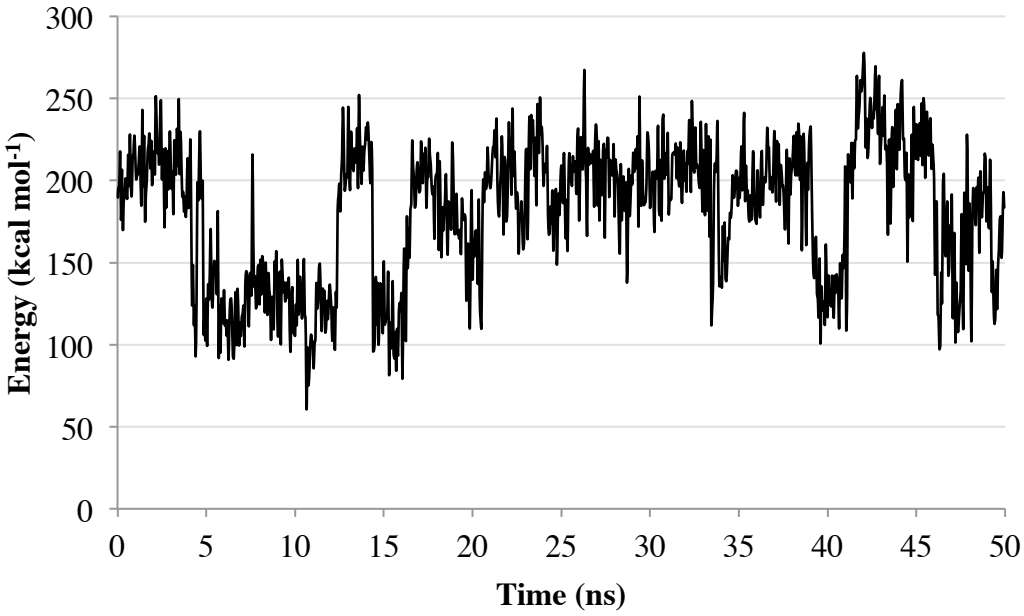


Figure 1.15. Energy of catalytic peptide CP1-H2δ over 50 ns simulation.

The average peptide energy of the simulation was found to be 180.7 kcal mol^{-1}, the standard deviation of the energy was 42.1 kcal mol^{-1} with the average deviation associated with triad formation was -38.08 kcal mol^{-1}, therefore the peptide experiences no significant stabilisation upon formation of the triad. This was found to be characteristic for all triads of all peptides formed during this study.

During the simulation of the epsilon protonated H2 isomer of **CP1**, no triad structure was identified; however it was found that the individual hydroxyl and acid – histidine distances did sporadically meet the 4 Å cut-off criteria, Figure 1.16, but did not coincide during the simulation.

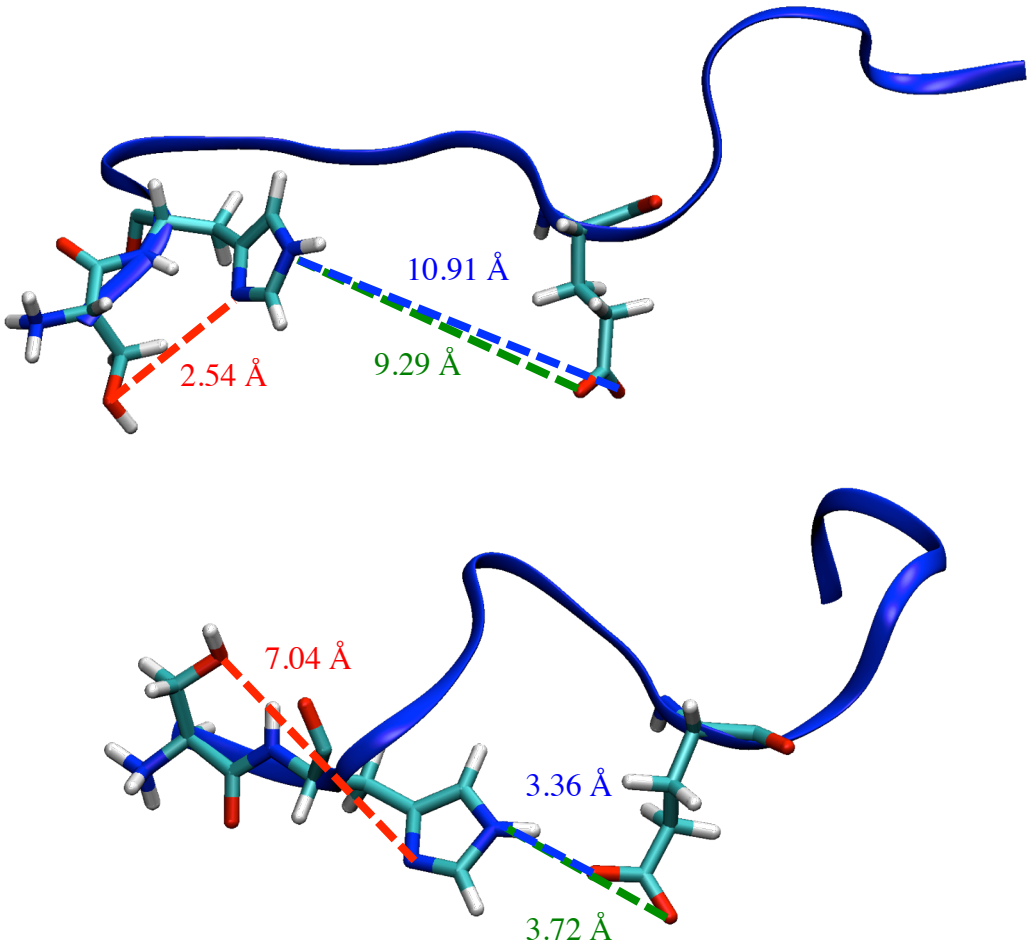


Figure 1.16. Structures of 'partial' triads formed during simulation of CP1-H2ε catalytic peptide.

The presence of such ‘partial’ triads support the notion that a simple serine–histidine ‘diad’ may catalyse amidolysis; however this study focuses *only* on the formation of the full triad, where the acidic moiety increases the basicity of the imidazole to mimic the charge-relay system observed in protease enzymes.

Unlike **CP1**, the second catalytic peptide sequence, **CP2**, has the possibility to adopt more than one triad conformation as a result of two histidine residues in the 3 and 5

positions and threonine residues in the 1 and 4 positions; indeed it is conceivable that any combination of these residues may couple with aspartic acid (D2) to form a triad. As with the examination of **CP1**, both delta and epsilon forms of histidine are considered giving rise to four isomeric combinations: H3δ-H5δ, H3δ-H5ε, H3ε-H5δ and H3ε-H5ε. From **CP2** simulations, an interesting relationship between protonation state and triad formation was observed, where a triad containing the N-terminal threonine residue (T1) prefers to form with the epsilon protonated histidine isomer, Figure 1.17.

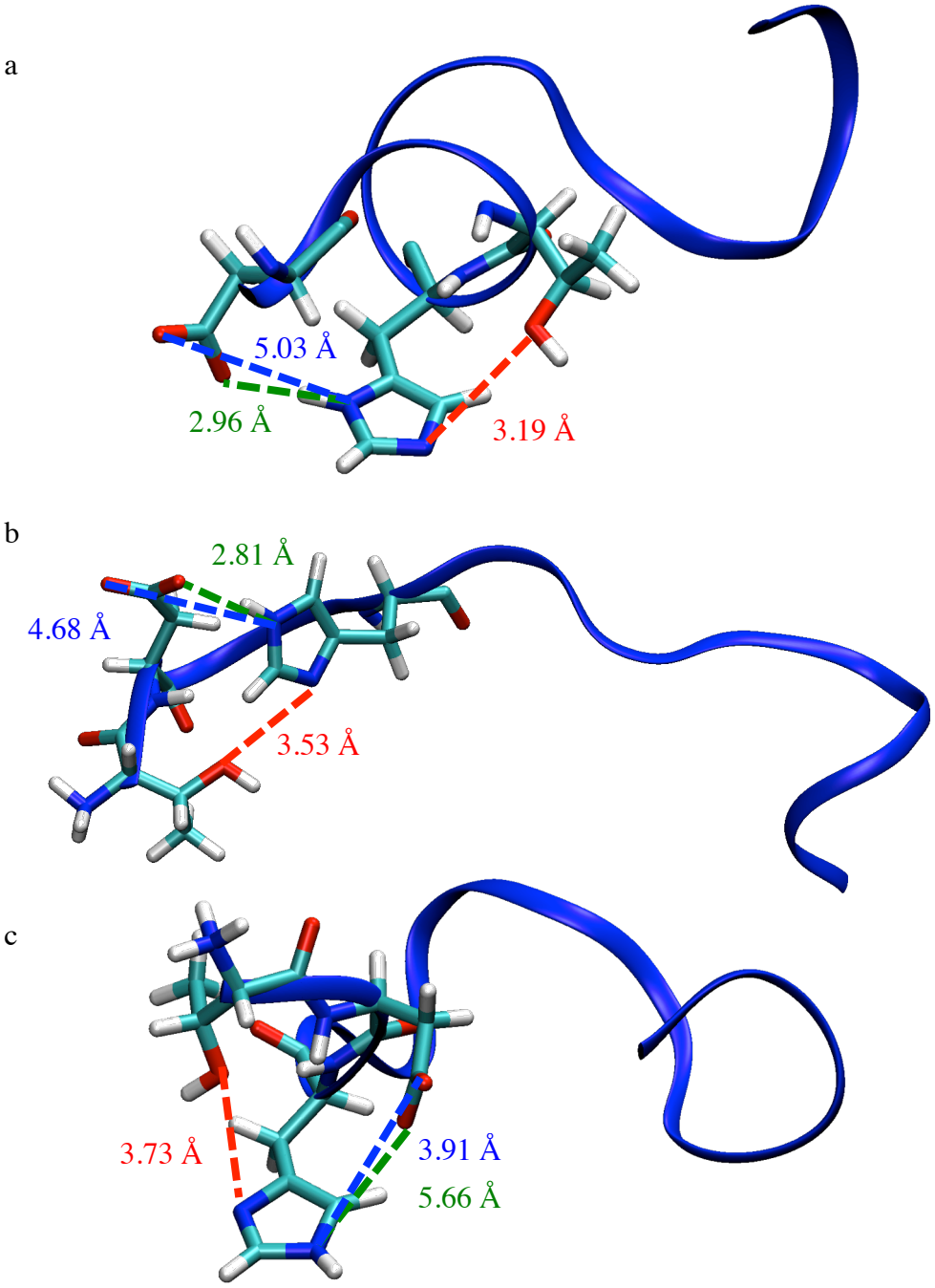


Figure 1.17. Catalytic triad structures of a) CP2-H3δ-H5δ (T4-H5-D2) b) CP2-H3δ-H5ε (T1-H5-D2) and c) CP2-H3ε-H5δ (T1-H3-D2).

Catalytic peptide **CP2**-H3δ-H5δ shows a preferential T4-H5-D2 triad formation; however the doubly epsilon protonated isomer of **CP2** shows no apparent formation of a triad structure. At neutral pH both the delta and epsilon protonated forms of histidne likely exist transiently; thus these simulations show that multiple proteolytic charge relay triads can be supported for **CP2**. These observations reinforce the simplicity and elegance of the combinational screening approach to identify catalytic peptides, in that the *order* of amino acids within the sequence is also key in facilitating catalytic function. For example, comprehensive design of a catalytic peptide, based purely on sequence rearrangement of **CP2**, would require *ca.* 30,000,000 order permutations, highlighting the efficiency of the combinatorial approach over *de novo* design by orders of magnitude.

In the case of **CP3**, as with **CP2**, there are a number of possible triads which may be adopted based on the three serine residues (S1, S4 and S6) and two histidines (H9 and H10). However, examination of all simulations indicates that only S1 E3 and H10 participate in the formation of a triad for **CP3**, Figure 1.18.

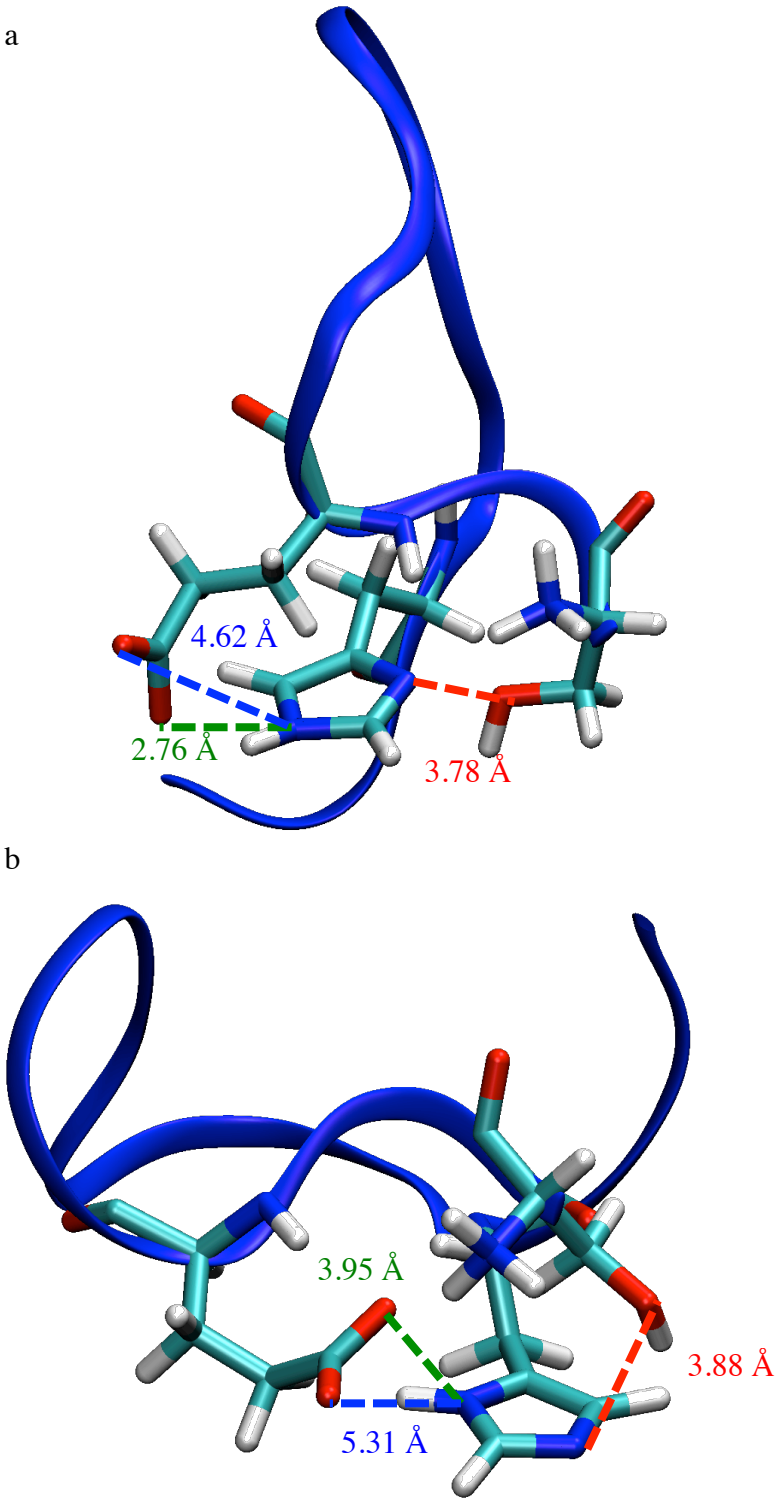


Figure 1.18. Catalytic triad structures of a) CP3-H9δ-H10ε (S1-H10-E3) and b) CP3-H9ε-H10δ (S1-H10-E3).

Again the importance of sequence order is observed in that only one of the six possible triads forms regardless of histidine protonation state; only the N-terminal serine and H10 residues arrange to give a viable triad.

This strengthens the idea that position of each residue within the peptide is a vital piece of each peptide's catalytic makeup. As **CP3** only supports one triad, it is possible that its motif would fall 'later' in the evolutionary timeline of minimal enzymes; only S1, H10 and E3 consistently compose its active site whereas the remaining hydroxyl and histidine residues may instead provide a structural support role.

1.5 Conclusions

In this work, the probable mechanisms by which three catalytic peptides perform amide bond condensation have been demonstrated through classical molecular dynamics simulations. All three peptides in this study were shown to adopt a triad analogous to those observed in protease enzyme active sites.

Through consideration of all possible tautomeric forms of the imidazole ring of histidine, it was found that not all isomeric forms of the peptides could adopt a triad structure and that for those that could, strong correlation between participating residues and protonation state were observed. The results presented also highlight the efficiency of the combinatorial screening method over *de* novo design approaches

and indicate that sequence order, and not simply triad residue propensity, is of key importance.

The formation of a triad in these simulations does not correspond to an energetic minimum, hence the structure remains flexible and peptides only form triad analogues temporarily; this correlates well with the low catalytic activity observed for these peptides in FRET based experiments since it is clear that the peptide can only support an active site transiently.

The application of molecular dynamics simulations to this investigation has provided insight into the mechanistic and conformational behaviour of minimal enzymes and has further validated the combinatorial screening method as an effective approach to identify catalytic peptides.

2 Amino Acids and the Au 111 Surface

2.1 Introduction

2.1.1 Biomolecules and Inorganic Materials

Recognition between biomolecules and solid surfaces is at the heart of many natural phenomena. Biomineralisation, the process by which living organisms produce hard tissue for structure and storage applications, relies on bio-assisted crystal growth to direct material morphology and function.[96-98] Furthermore, it has been shown that the ability for cells to adhere to inorganic materials like glass depends on the binding affinity between the cell exterior proteins and the solid surface,[99] where cell spreading is guided by the propensity of hydrophilic or hydrophobic residues. The capacity for biomolecules to bind to, and indeed recognise, metals has also been observed in immune responses toward medical implants[100-101] with protein fragments forming a coating around the artefact.

Surface-biomolecule interactions have also found widespread applicability in the field of nanotechnology and have been successfully exploited in the synthesis of metal nanoparticles[72, 102-108] with controlled size[72, 105-107] and surface morphology.[108] In the example shown in Figure 2.1, gold nanoparticles of varying size are synthesised by short peptide sequences that can reduce $AuCl^-_4$ ions and bind to Au^0.

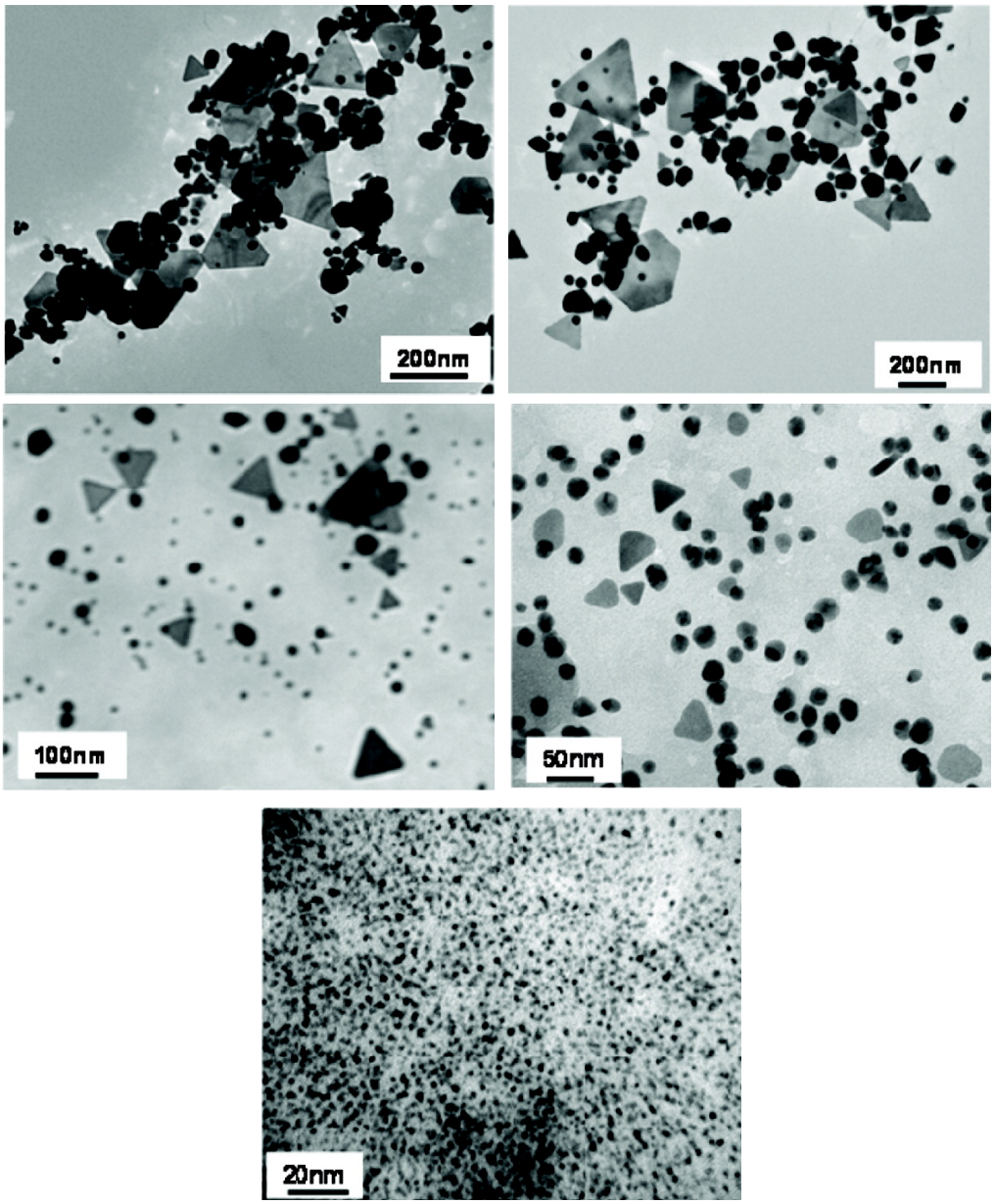


Figure 2.1. Variation of gold nanoparticle size and shape under controlled synthesis by different short peptide sequences.[102]

The importance of creating nanostructures with highly specific properties is perhaps best demonstrated via the versatility of gold nanoparticles in catalysis[109-114] and sensing applications.[115] Gold nanoparticles have strong surface plasmon resonance (SPR) and surface-enhanced Raman scattering (SERS) signals; it is these interesting

optical properties that have been exploited for use in the field of biomedicine, with functionalised gold nanoparticles widely employed for *in vivo*[116-117] and *in vitro*[118-120] diagnostic and imaging applications. For example, gold nanoparticles functionalised with DNA[121-122] and peptide[123-124] moieties that exhibit multiple affinities, can be used for the detection of various metal ions, such as Pb^{2+}, Hg^{2+} and Pd^{4+} in the nanomolar (ppb) scale. Upon binding with a specific metal ion, clear colorimetric changes can be observed and the characteristic UV/Vis spectra associated with the gold nanoparticles show well-defined peak broadening and shift patterns that are unique to each type of bound ion, Figure 2.2.

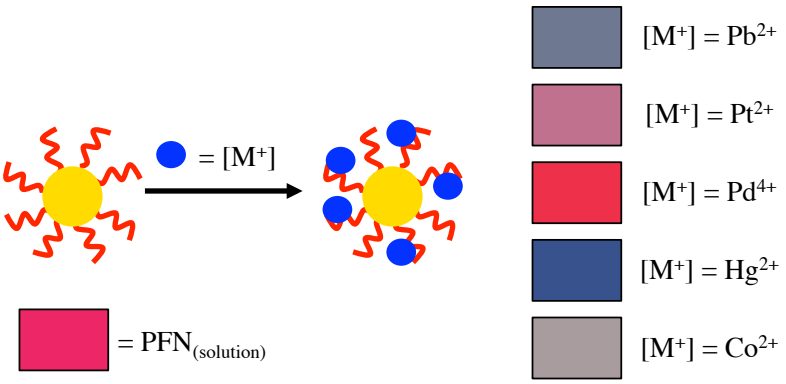


Figure 2.2. Colourimetric response of peptide functionalised nanoparticles (PFNs) to metal ions in solution.

Functionalisation of gold nanoparticles is a key element in targeting specific cells[124-125] or subcellular[126-128] components; however the mechanism of binding between biomolecules and gold is not always well understood.[129] Understanding the parameters that govern protein adsorption to surfaces is essential for design of functionalised particles for effective and specific cell targeting; using single amino

acids as a benchmark we look to explore the origins of biomolecular selectivity with gold.

2.1.2 The Gold Surface

Investigations into the interaction of biomolecules with gold focus mainly on the Au 100 and Au 111 surfaces since they are the most common facets displayed on gold nanoparticles,[130] Figure 2.3.

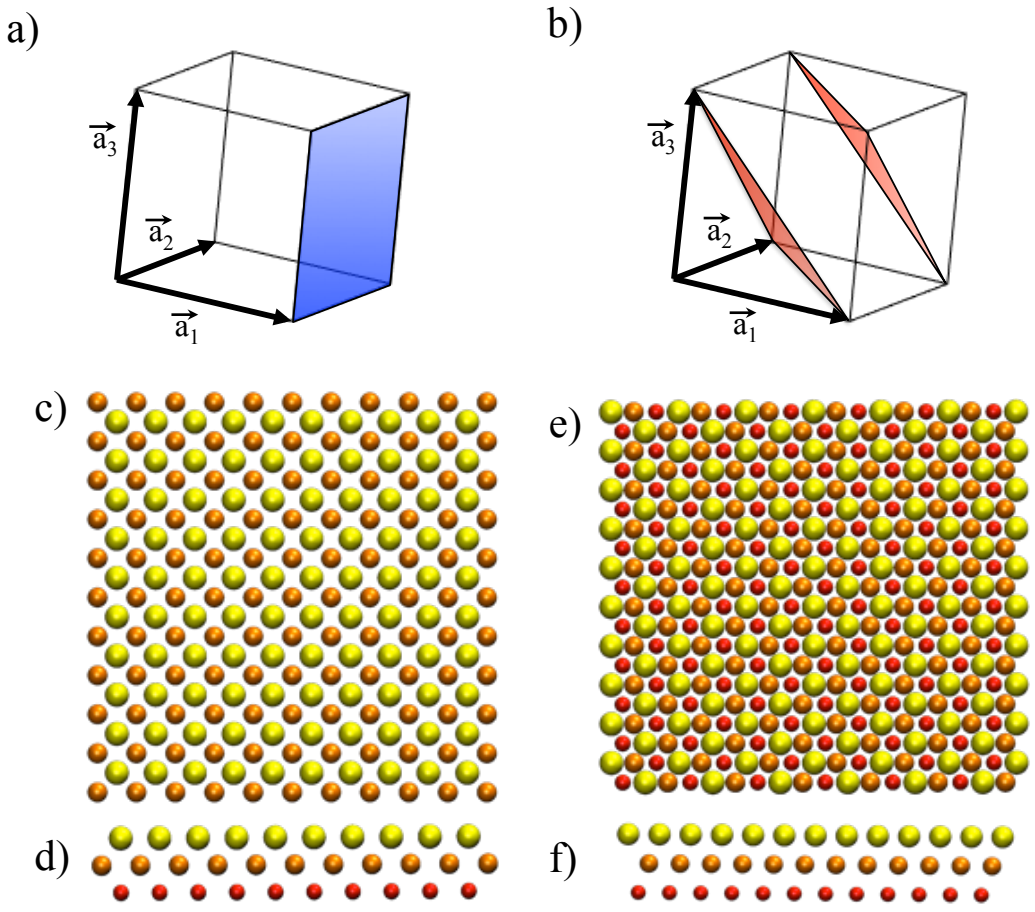


Figure 2.3. Planes of Miller indices for a) 100 and b) 111 face centred cubic (fcc) crystals. Structures of Au 100 and Au 111 surfaces shown from above c) and e) and

side-on d) and f) respectively. Yellow spheres indicate the top layer, orange the second layer and red the third layer.

Of both surface constructs, the Au 111 is the most common target for biomolecular species[131-133] and small organic molecules.[134-136] The Au 111 surface presents a number of advantages in that its surface structure is well-characterised[137-143] in the presence of air[144] and water[145] and due to its greater stability, surface reconstruction has minimal impact on reactivity.[146] Computational studies[147-148] have also demonstrated the preference for peptides to bind at the Au 111 interface over the Au 100. Furthermore, Au 111 presents a continuous flat surface, which is beneficial in microscopy studies[149-150] and makes implementation in theoretical models a relatively simple task.

2.1.3 Amino Acids on Gold

Interactions of amino acids with the Au 111 surface have been studied both quantum mechanically and with molecular dynamics. Density functional theory[151-152] (DFT) has been used to predict the conformational properties of amino acids,[153] on gold, with some studies focusing on small organic moieties that mimic amino acid side chains[153-154] whilst others examine specific residues such as cysteine,[155-158] methionine,[159] histidine[160] or proline.[161-163]

The main conclusions from these studies suggest that residues with planar side chains, such as tyrosine and tryptophan, prefer to lay flat on the surface and that

polar amino acids, e.g., serine and methionine, orientate so that the heteroatoms can interact strongly with the gold. These calculations also provide insight into the possible behaviour of peptides at the interface with gold, suggesting that unfolding would occur as the binding conformation of the amino acids is found to be a balance between backbone contact with the surface, charged termini stabilisation by interaction with the solvent and minimising contact between hydrophobic side chains and water.

The conformational properties of single amino acids with gold have also been examined by molecular dynamics.[164-167] One study in particular[165] noted that the orientation of amino acids on the surface correlated well with a soft epitaxial binding mechanism.[147] The Au 111 surface geometry and electronic properties give rise to potential wells and grooves across the surface known as epitaxial sites; the spacing between these sites corresponds well with the geometry of amino acids (to varying degrees) where polar atoms coordinate to the surface at epitaxial sites.

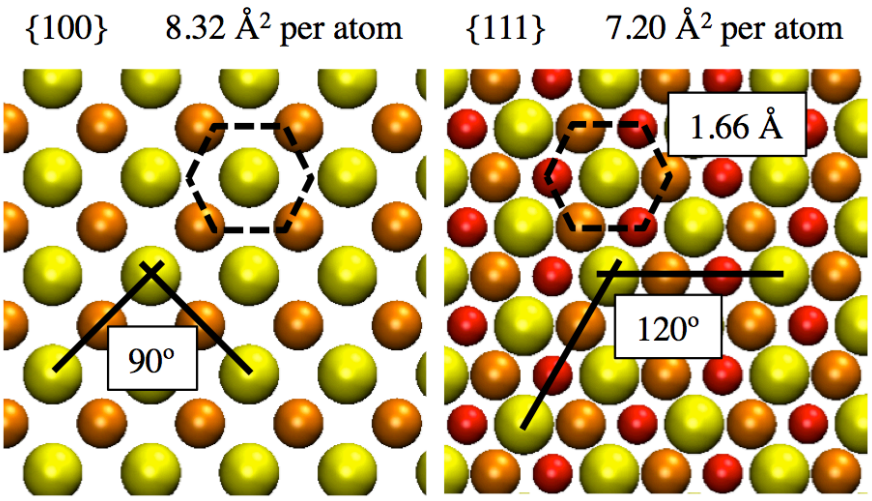


Figure 2.4. Comparison of the (a) Au 111 and (b) Au 100 geometries where dotted lines represent the orientation of amino acid side chains with aromatic rings coordinating to fcc lattice binding sites.

These investigations also showed that large planar groups prefer to lay flat on the surface, whereas polar groups prefer to orientate so that their hetetoatoms align with the surface; in particular above epitaxial binding sites. Heinz *et al.*[147-148] demonstrated that the geometric harmony between amino acids and epitaxial sites leads to binding specificity of peptides with Au 111 over Au 100.

2.1.4 Binding Energies of Amino Acids

In quantum mechanical calculations, derivation of amino acid binding energies can be calculated easily through combination of the interaction energy between the amino acid and surface and the deformation energy of the residue upon binding. Despite this, quantum mechanical studies of biomolecules and inorganic materials is

limited to small systems as QM methods are inherently computationally expensive. Therefore, the method of choice in the study of large biophysical systems is classical molecular mechanics as all components can be included at significantly lower computational effort. The main disadvantage of using such methods is the associated decrease in accuracy of the much simpler classical Hamiltonian. The forcefields used herein are parameterised for protein and protein-interface simulations and are validated using quantum mechanical calculations; however the calculation of binding energies is much more challenging.

In molecular dynamics 'compartmentalisation'[147, 165] is the most easily employed method to extract binding energies from simulation data. The interaction between a molecule and its substrate is derived from the difference in the potential energy between the complex in the bound state and in the unbound state. Using a biomolecule bound to gold as an example, Figure 2.5 illustrates the terms used to calculate the binding energy in compartmentalisation.

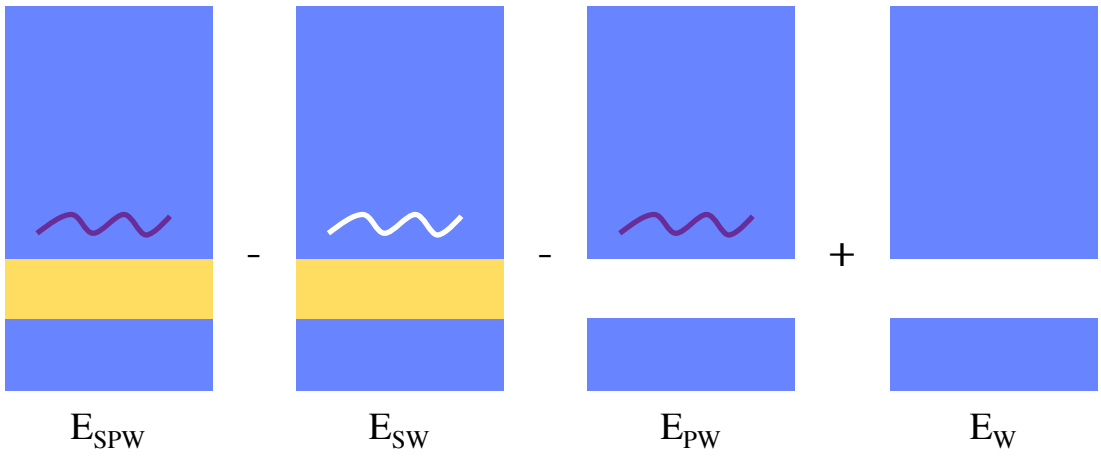


Figure 2.5. Components used in the calculation of peptide-surface binding energies. The value of ΔE is the difference between the peptide in the adsorbed state and in the

bulk state. The peptide/biomolecule (P) is shown in purple, surface (S) in yellow and water (W) in blue.

The main limitation of this approach is that it is based only on a single snapshot of the system; thus it does not account for dynamic effects and cannot be considered a true binding free energy. Any particular configuration used in this method may not be representative of a physically important state and instead may be trapped in a local minimum by large energy barriers; therefore other methods have been developed with the aim of overcoming this 'sampling' challenge.

Umbrella sampling[168-169] is a method of exploring the energy landscape of a system or process in molecular dynamics and has been used in the calculation of binding free energies of amino acids.[164, 166] The first step in performing umbrella sampling is to generate a series of starting configurations from which biased molecular dynamics simulations are performed. Steered molecular dynamics (SMD) is a method commonly used to explore states of a system or process that are normally inaccessible at equilibrium due to large energy barriers, e.g., the desorption of a biomolecule from a surface, to ensure that unfavourable states are adequately sampled. The path between the two states, i.e., adsorbed and desorbed, is split into evenly spread intervals. In each of these 'bins', the molecule of interest is held at the SMD reaction coordinate by a harmonic potential or 'spring', and a traditional molecular dynamics simulation is performed.

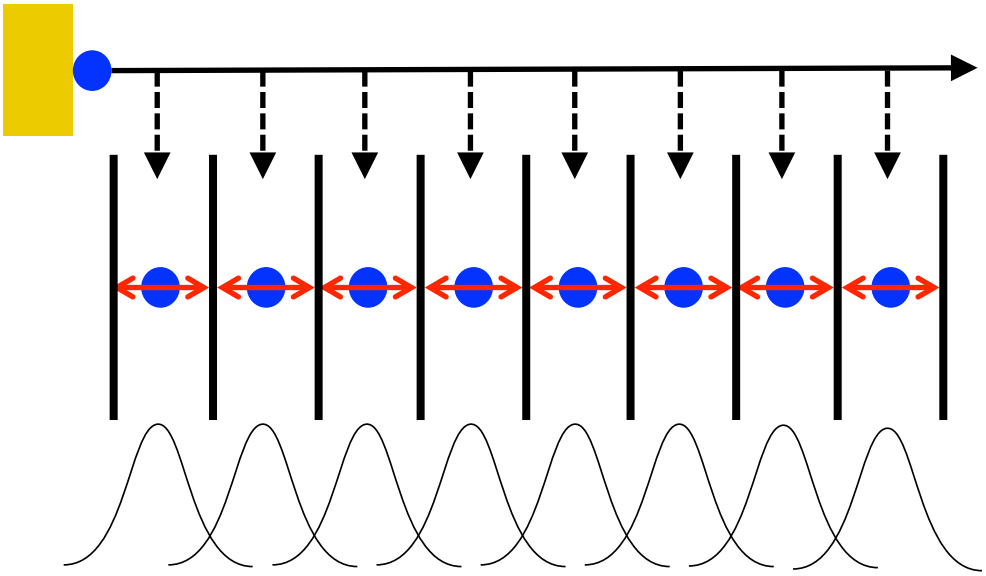


Figure 2.6. Illustration of umbrella sampling.

The force felt by the spring as the molecule interacts with its surroundings is calculated from Hooke's law:

$$F = -kx \qquad \text{Equation 2.1}$$

Where F is the force, k is the spring constant and x is the displacement.

The force acting upon the spring is collected over each simulation and the average value is calculated. The work required to pull a molecule along a particular path can be extracted by summing the forces over the bins:

$$W = \int_0^d F(x)dx \qquad \text{Equation 2.2}$$

Where d is the distance of the reaction coordinate.

This work is also known as the potential of mean force and with sufficiently narrow bin widths and long simulation lengths equates to the free energy of binding; however, choosing the most efficient parameters for umbrella sampling is often a process of trial-and-error. One consideration is the spring stiffness, if a spring is too weak the molecule will drift from its designated bin and if it is too strong it may only sample a small region of the reaction coordinate. Thus bin width, spring constant and simulation length must be tailored to ensure that all states are surveyed.

For small molecules, umbrella sampling is relatively inexpensive; however for larger molecules greater detail, i.e., more sampling bins, is required and this becomes less efficient as substrate size increases.

The adaptive biasing force[170-172] (ABF) method works on a similar principle, in that a molecule is constrained to a series of bins along a reaction coordinate with a harmonic potential; however a 'biasing force' is applied to the molecule that cancels out the force acting upon it. Thus free energy landscape is flattened and the system can overcome high energy barriers. In ABF the bin width can be varied to provide higher resolution (narrow bins) at important points on the reaction coordinate, i.e., at the interface, and lower resolution at less important points, i.e., in the bulk. This method has been employed for a variety of free energy calculations,[173-176] however it also suffers form the necessary system optimisation with respect to bin widths and simulation length.

A much easier approach is to directly calculate the change in free energy continuously from steered molecular dynamics trajectories. Another non-equilibrium method that has recently been employed for biomolecules at interfaces is the combination of non-equilibrium thermodynamic integration[177-178] (NETI) with SMD.[179]

NETI-SMD relates the work required to pull a molecule along a particular path in a non-equilibrium simulation to the equilibrium free energy through Jarzynski's equality:

$$\Delta A = -k_B T * ln\langle e^{-W/k_B T}\rangle \qquad \text{Equation 2.3}$$

Where A is the Helmholtz free energy, k_B is the Boltzmann constant, T is temperature and W is work.

In an SMD simulation a molecule, in this study an amino acid, is pulled along a trajectory at a constant velocity by a harmonic potential or 'spring'. If the molecule is pulled at an infinitely slow rate, the resulting work will be exactly equal to the equilibrium free energy of binding; however this is not practical since such a simulation would in theory take an infinite amount of time to run. Therefore, each SMD simulation is performed at a much faster velocity, i.e., not in equilibrium, but multiple times. This is advantageous in that a number of short simulations can be run simultaneously; thus reducing overall completion time.

One of the main challenges of this method however is the calculation of errors, since if there is insufficient sampling of trajectories the results will be unreliable. One method of estimating errors is via statistical bootstrapping analysis.[180]

Bootstrapping is a way of assessing the accuracy of a data set by calculating the error associated with a particular estimator, $\hat{\theta}$. From a particular data set, N samples are taken to give a vector, $\vec{x}$, consisting of N observations. The estimator is predicted by applying a function, *s*, to the vector of observations, Equation 2.4.

$$\hat{\theta} = s(\vec{x}) \qquad \text{Equation 2.4}$$

In this study, *s* is Jarzynski's equality, Equation 2.3, and the vector contains the free energies from N simulations, i.e., $\vec{x} = (x_1, x_2, x_3 \ldots x_N)$. The values within $\vec{x}$ are chosen at random from the data set and samples can be chosen more than once. The estimator is calculated for the particular vector and a distribution of estimators is obtained by resampling vectors multiple times, *ca.* 1,000,000 to eliminate random error. The overall distribution of $\hat{\theta}$ values is Gaussian due to the large resampling iterations; thus the peak represents the mean free energy. Errors are measured using t-statistics where the number of degrees of freedom is N-1; values reported herein are at the 99.9 % confidence level. Bootstrapping allows an estimation of the number of NETI-SMD simulations required to minimise errors below a certain cut-off by incrementally increasing the number of simulations used to populate $\vec{x}$.

The NETI-SMD approach has been employed with peptides on a graphene surface,[181] protein folding,[182] protein-protein[183] and protein-ligand[184] interactions, as well as amino acids bound to inorganic surfaces.[185-186] However, to date there have been no studies reported that use this method for the calculation of single amino acid binding energies on Au 111. Herein, we report a benchmarking study to examine the adsorption behaviour of a selection of single amino acids with the Au 111 surface.

2.2 Adsorption of Amino Acids

2.2.1 Background

In the following chapter, we examine the properties of a gold binding peptide with the sequence MHGKTQATSGTIQS,[132] known as GBP1; hence we have chosen to study its constituent amino acids with respect to adsorption to the gold surface and calculation of binding free energies.

The description of the CHARMM-METAL[187] force field used in this study employs a simple modification to the Lennard-Jones terms of the classical force field in Equation 1.2. The CHARMM-type force field employs a 12-6 potential to describe van der Waals interactions, Heinz et al. parameterised the σ_{ij} and ε_{ij} terms of gold interactions (where σ_{ij} is separation where the inter-particle potential is zero and ε_{ij} is the potential well depth) as an extension of CHARMM22[188] to reconstruct experimental densities[189] and surface tensions[190] of water interacting with gold under ambient conditions.

2.2.2 Construction of System

An initial equilibration of the surface and water was performed prior to the introduction of the amino acids. An Au 111 surface measuring 60 x 40 x 12 Å was generated using the PBE plugin of the GaussView 5[191] program, where the surface comprised of 6 layers of gold atoms. Two boxes of TIP3P[88] water were then added to the system, one measuring 60 x 40 x 58 Å was placed above the metal surface and another measuring 60 x 40 x 15 Å below. Each water box was generated by the genbox plugin of Gromacs[192] to a density of approximately 1.00 g cm^{-3}; the total system consisted of *ca.* 19,000 atoms, including *ca*. 5,500 water molecules.

The gold atoms of the surface were fixed to maintain surface geometry and orientation and the system was equilibrated under constant volume and temperature conditions (NVT) for 10 ns using the NAMD[91] molecular dynamics engine, employing the CHARMM-METAL[187] force field to describe the gold and CHARMM22[188] for the TIP3P water model potentials. The simulation was performed with a time step of 2fs and commenced with a 2,000-step minimization period to remove bad contacts followed by a heating phase where the temperature was increased from 10 K to 300 K at 10 K increments every 1000 steps. The Particle-Mesh-Ewald[93] method was again employed for full electrostatics, with a grid spacing of 1 Å and periodic boundary conditions were used throughout.

This step allowed the water molecules to interact with the metal surface and to adopt their equilibrium state. The water was observed to have an affinity with the metal,

where two higher density layers of water had formed at approximately 2.4 Å and 5.0 Å from the surface, Figure 2.7, Figure 2.8.

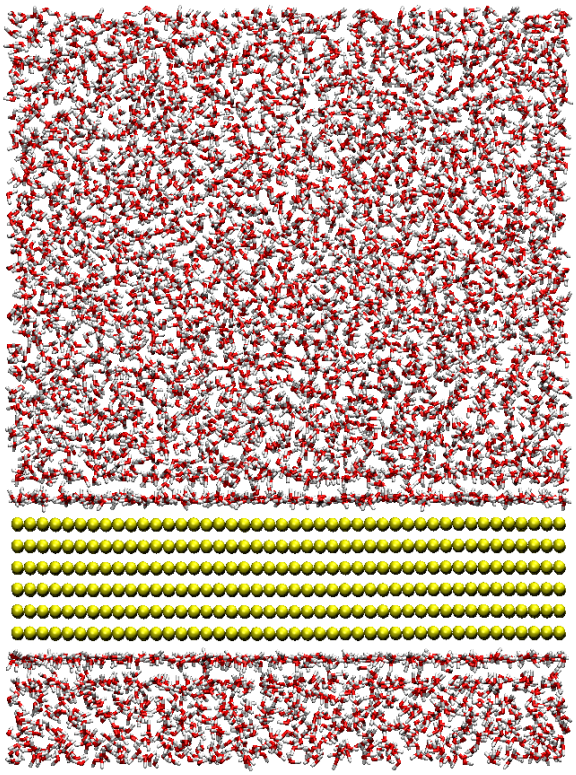


Figure 2.7. Equilibrated water – gold system with higher density water layers adjacent to the surface.

This higher density in the first wetting layer occurs as the interfacial water interacts directly with the Au 111 atoms so that at least one OH bond lies in the plane of the surface in order to minimise the number of hydrophobic interactions between the gold and the water in the second wetting layer. The CHARMM-METAL[187] force field was parameterised with reproducing experimental surface densities[189] in mind, although Figure 2.8 shows that for the system shown in Figure 2.7, this is higher, *ca.* 7.0 g cm^{-3} at 2.4 Å above the surface.

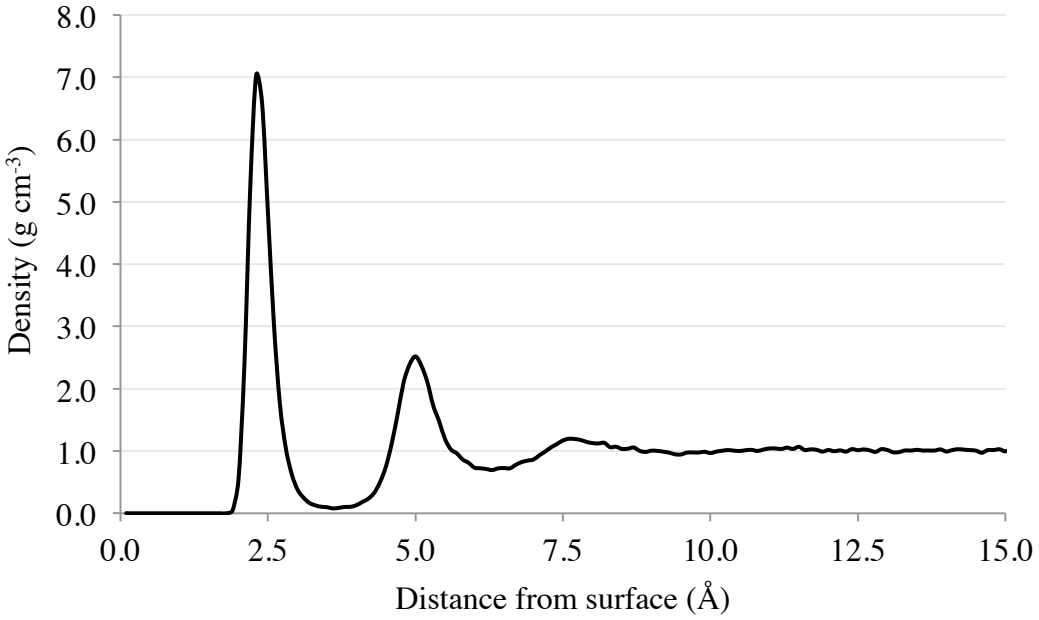


Figure 2.8. Water density as a function of the distance from the surface.

2.2.3 Adsorbed Amino Acids

Single amino acids were built in both their zwitterionic form with charged N and C termini, representative of their configuration in aqueous media at pH 7 and capped termini (acetylated N-terminus and amidated C-terminus) for comparison using the guesscoord plugin of VMD.[90] The amino acids were then placed above the metal with each residue positioned so that the side chain was pointed directly towards the surface and the closest atom was 3.5 Å above the top layer of atoms. Overlapping water molecules were removed and systems with an overall charge (i.e., Lys, +1) were neutralised with sodium or chlorine ions. Each residue was allowed to adsorb to the surface over a period of 10 ns, again under NVT conditions.

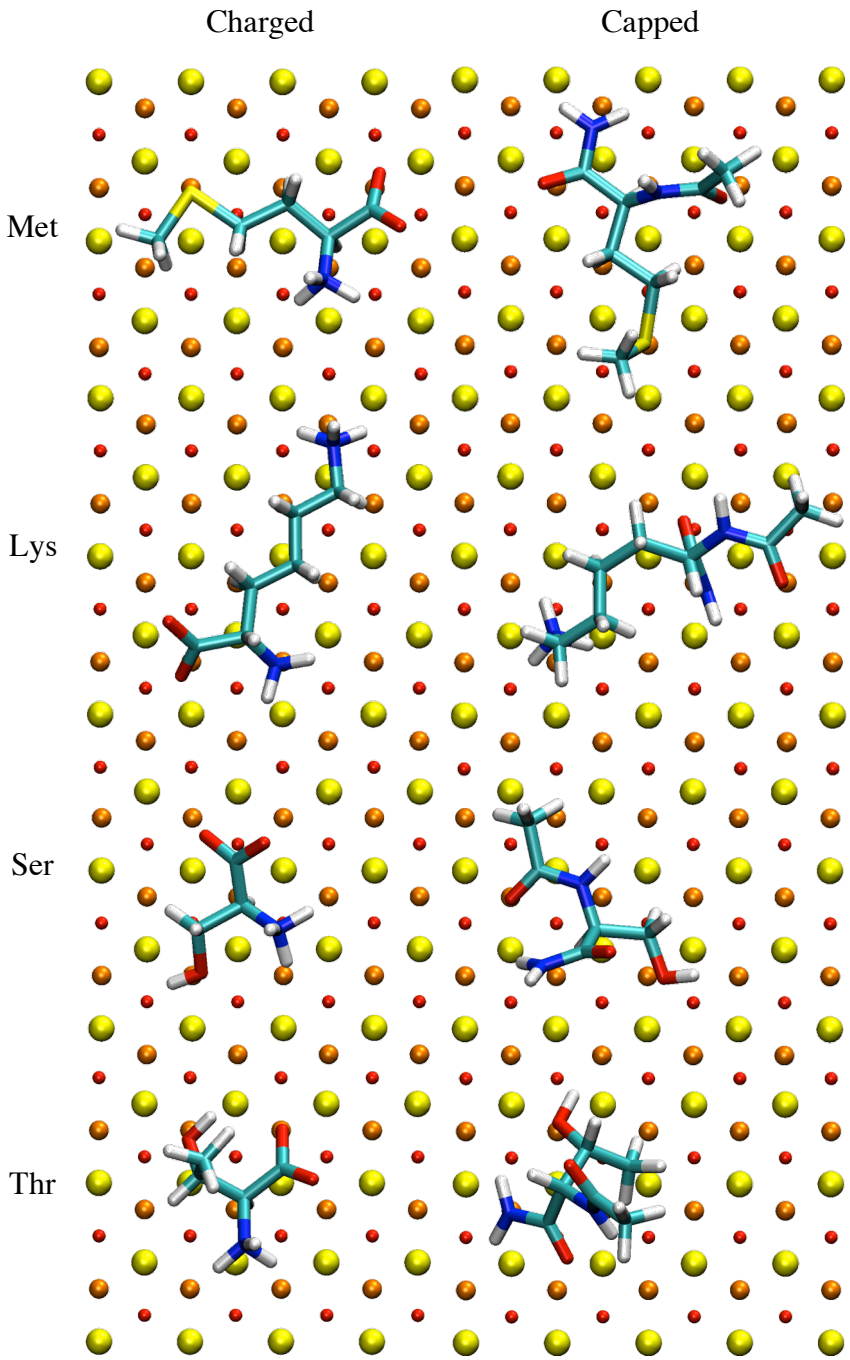


Figure 2.9. Orientation of single amino acids methionine, lysine, serine and threonine on the Au 111 surface in both charged and capped termini states.

All amino acids were found to have adsorbed to the surface, adopting a conformation consistent with soft epitaxial binding,[165] Figure 2.9, Figure 2.10 and Figure 2.11. The

polar amino acids methionine, lysine, serine and threonine were found to orientate so that the single heteroatom in the side chain lay in the epitaxial site, Figure 2.9.

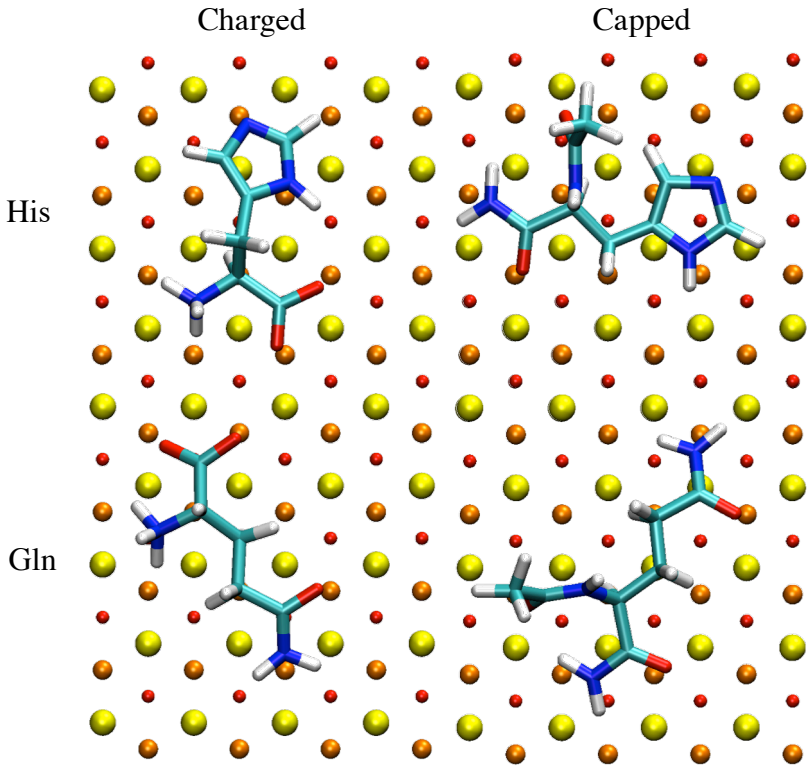


Figure 2.10. Orientation of single amino acids histidine and glutamine on the Au 111 surface in both charged and capped termini states.

Residues with two polar atoms in the side chain, i.e., histidine and glutamine, show favourable alignment with epitaxial binding sites with both side chain and terminal nitrogen and oxygen atoms aligning well with atoms in the second gold layer, Figure 2.10. Amino acids with apolar side chains alanine, glycine and isoleucine on the other hand can only form soft epitaxial binding interactions through terminal groups, Figure 2.11.

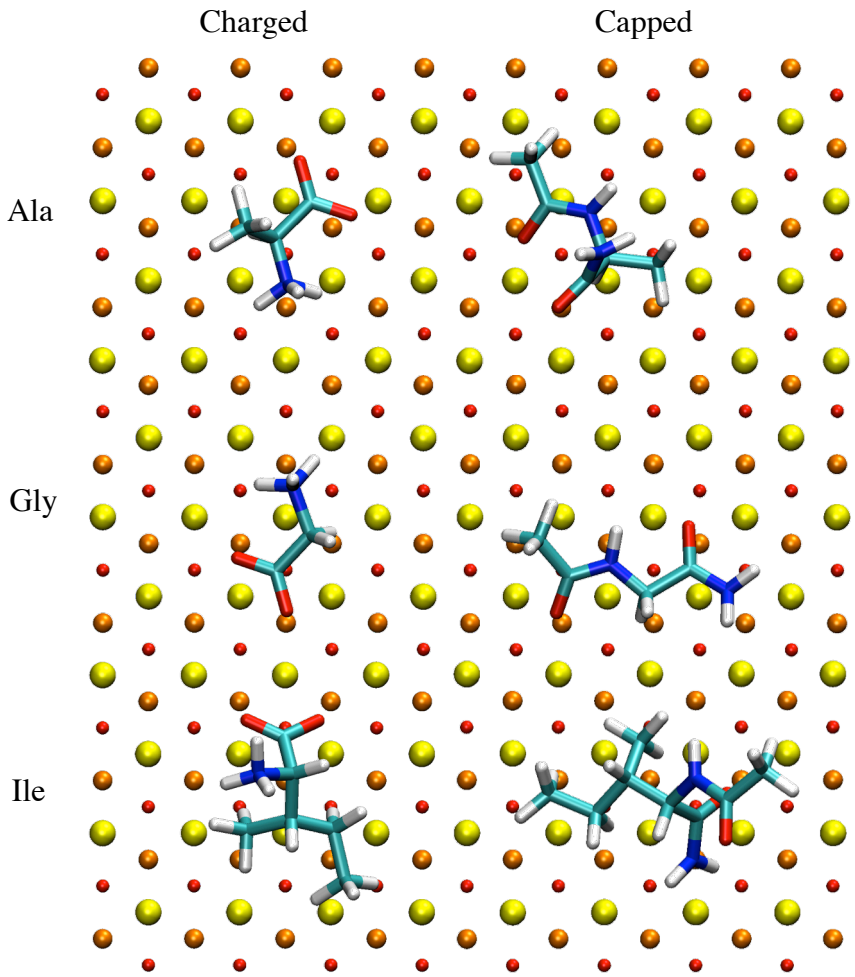


Figure 2.11. Orientation of single amino acids alanine, glycine and isoleucine on the Au 111 surface in both charged and capped termini states.

In general, backbone heterotatoms of charged terminal amino acids have a more favourable geometry for epitaxial binding than those with capped termini, which in general must sacrifice at least one ideal heteroatom-gold interaction orientation.

2.3 Desorption of Amino Acids

2.3.1 Non-Equilibrium Thermodynamic Integration

Each amino acid was pulled from the surface perpendicular to the xy plane via the C_α at a rate of 0.005 Å ps^{-1}. A spring constant of 500 kcal mol^{-1} $Å^{-2}$ was employed to ensure that the observed pulling coordinate did not lag behind the target distance; this is essential for the calculation of free energy using NETI-SMD method as the force is integrated over a regular time interval, using a weak spring constant would mean that the time-distance relationship of the constant velocity pulling would collapse. Each residue was pulled 40 Å from the surface to ensure full desorption, this process was repeated thirty times to reduce errors.

2.3.2 Binding Free Energies of Amino Acids

As each amino acid is pulled from the surface, the force on the spring is integrated using Equation 2.2 to give the corresponding work value as a function of distance, an exemplary work profile for an amino acid (alanine – capped termini) is shown in Figure 2.12.

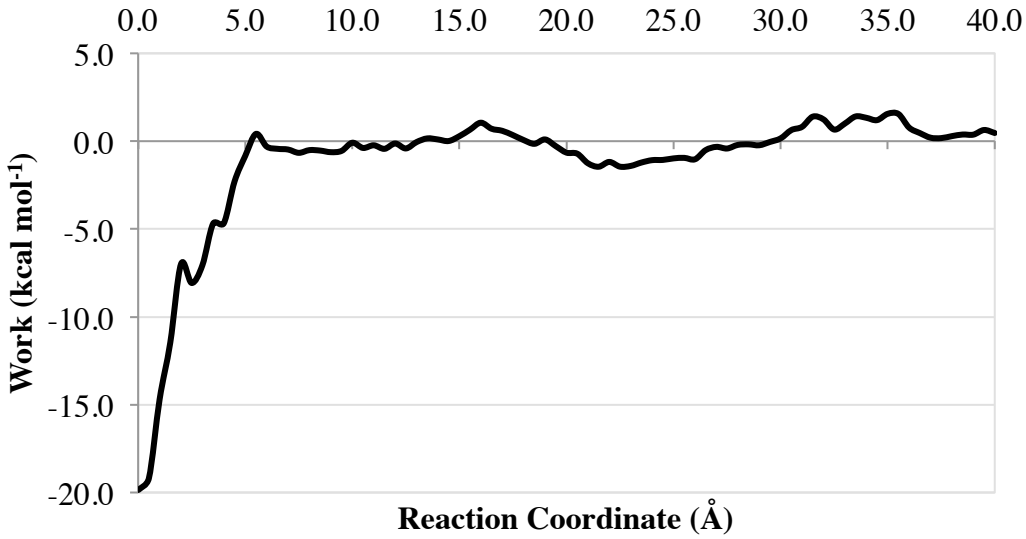


Figure 2.12. Exemplar pulling profile of a single amino acid from the Au 111 surface using the NETI-SMD method.

From the thirty desorption simulations of each amino acid, the evolution of free energies and associated errors was assessed through bootstrapping analysis, shown for alanine with capped termini in Figure 2.13. It was observed that with increasing number of simulations the errors were reduced, with the error for 30 simulations ranging from 0.42 – 1.76 kcal mol^{-1} for species with charged termini and 0.23 – 1.93 kcal mol^{-1} for those with capped termini.

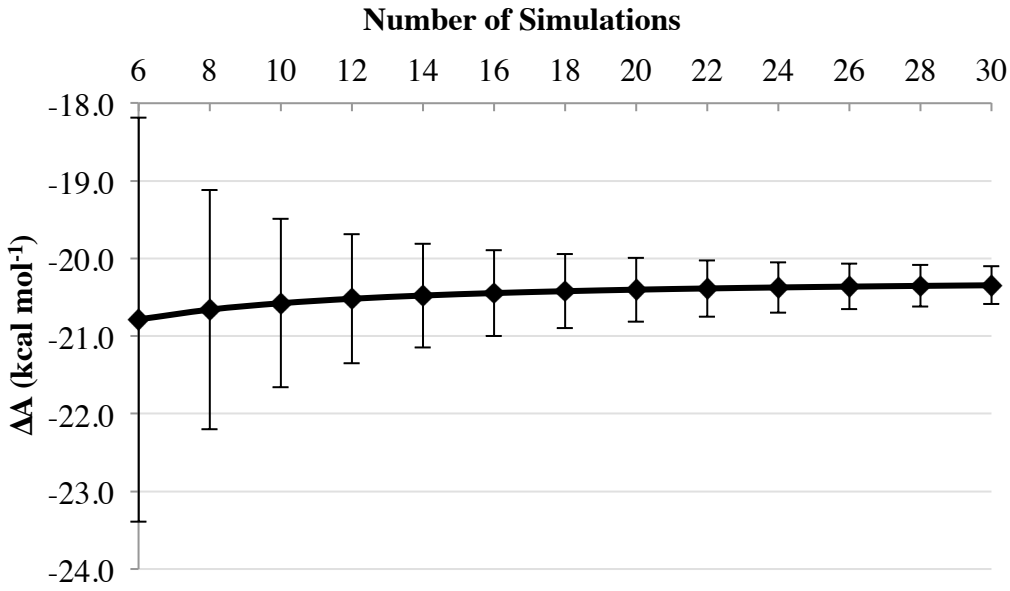


Figure 2.13. Exemplar evolution of Helmholtz free energy values and associated errors for the pulling of single amino acids (alanine with capped termini shown) with increasing number of simulations.

As can be seen in Figure 2.15 capped amino acids bind more strongly to the gold surface by between (ΔA) -6.1 and -14.3 kcal mol^{-1} over charged amino acids, likely arising from the additional six atoms of the acyl and amide terminal groups interacting with the surface.

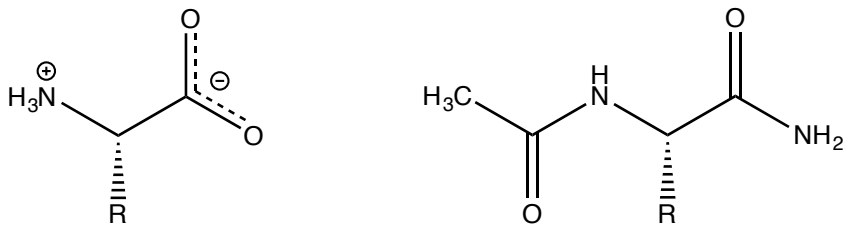


Figure 2.14. Single amino acid with both termini charged (left) and that with an acylated N-terminus and amidated C-terminus (right).

The relatively weaker binding of amino acids with charged termini may arise from the *non-polarisable* nature of the CHARMM-METAL force field. As the gold surface is entirely neutral and static, there are no additional Coulombic interactions between the surface and the adsorbate; therefore additional Van der Waals forces of the capped termini promote binding to a greater extent than charge. Furthermore, it is possible that the charged terminal amino acids are more favoured in solution, since they can form hydrogen bonds with the solvent more readily in the bulk compared to the bound state; hence weakening surface binding energy.

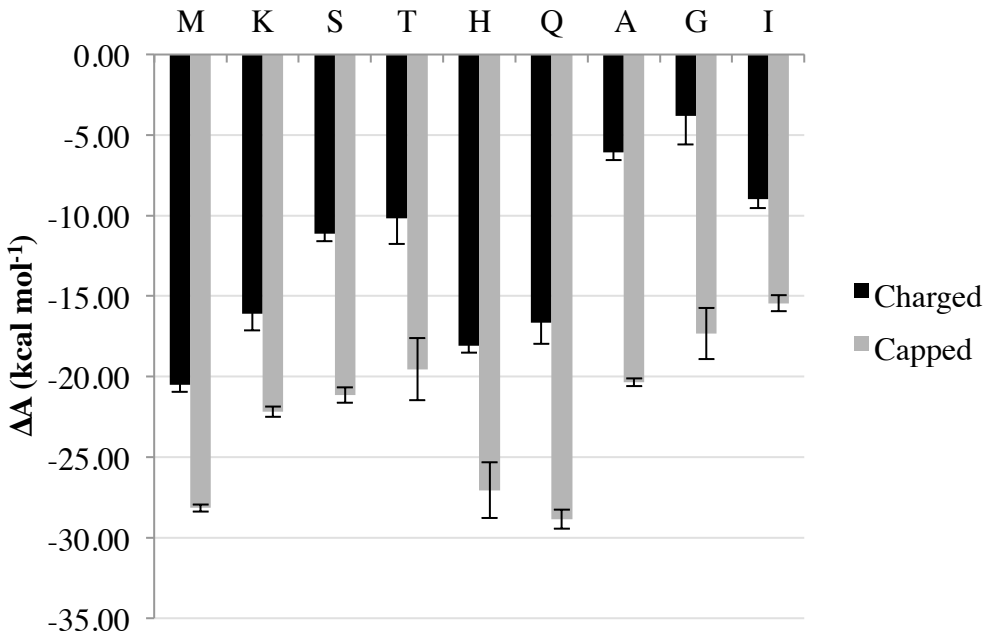


Figure 2.15. Helmholtz free energies of binding for charged and capped amino acids with the Au 111 surface.

As discussed in Section 2.2.3, the surface architecture of the Au 111 surface is well suited for the adsorption of amino acids, particularly those with polar side chains.

This is reflected in the relative binding free energies, where histidine (**H**) and glutamine (**Q**) show particularly strong binding (ΔA_H = -18.1 kcal mol^{-1}, ΔA_Q = -16.7 kcal mol^{-1}) relative to non-polar amino acids (ΔA_A = -6.1 kcal mol^{-1}, ΔA_G = -3.8 kcal mol^{-1}, ΔA_I = -9.0 kcal mol^{-1}). Interestingly, it was found that methionine (**M**) displays a high affinity for gold (ΔA_M = -20.5 kcal mol^{-1}), comparable to that of histidine and glutamine; lysine (**K**) however binds more weakly by *ca.* +4.5 kcal mol^{-1} despite also containing a moiety, although the aliphatic nature of the β to ε positions may hinder binding. Serine (**S**) and threonine (**T**) show the weakest binding of the polar residues under study (ΔA_S = -11.1 kcal mol^{-1}, ΔA_T = -10.2 kcal mol^{-1}), where the additional methyl substituent of threonine may also weaken binding, relative to serine.

The binding free energies calculated using NETI-SMD are of greater magnitude than those reported by Hoefling *et al.*[164] using the umbrella sampling method, where the greatest binding free energy was -9.4 kcal mol^{-1} (methionine). However, there is good agreement between the relative ordering of amino acid binding strength where polar amino acids methionine and histidine show the strongest affinity for gold and those with aliphatic moieties are the weakest binders; therefore, despite the discrepancies in absolute binding energy values there is consistency between the NETI-SMD approach and other, more expensive free energy methods.

2.4 Conclusions

It was found that the relative binding free energies of selected amino acids to the Au 111 surface could be calculated from a combination of non-equilibrium thermodynamic integration and steered molecular dynamics simulations.

Amino acids with polar side chains were found to orientate in accordance with a soft epitaxial binding mechanism, i.e. polar atoms sit directly above the atoms in the second gold layer, and that the introduction of bulky terminal groups increase binding strength relative to charged termini.

The relative binding free energies of the selected amino acids corroborates well with those calculated by adaptive biasing force simulations, although are observably stronger using NETI-SMD; therefore this approach can provide robust results and highlights that method continuity is of key importance when studying interfacial binding.

3 Gold Binding Peptide GBP1 and the Au 111 Surface

3.1 Introduction

3.1.1 Gold Binding Peptides

Although, the chemical diversity of peptides is advantageous as multiple functionalities can be implemented easily into a single molecule, it can also be disadvantageous in that the number of possible combinations of amino acids scales with 20^N, where N is the length of the peptide chain; hence the rational design of sequences with material binding properties is extremely difficult. This represents one of the biggest challenges in modern bionanotechnology.

Analogous to the methodology outlined in Section 1.1.6, combinatorial libraries can be used to identify biopolymers that bind to a particular surface; this is most commonly achieved with techniques such as phage display[76, 193] (PD) and cell surface display[194-195] (CSD). CSD libraries are generated by modification of DNA in a region of a bacterial genome associated with expression of protein on the cell surface, random modification and natural mutations mean that millions of unique sequences are present in each library. The library is introduced to an inorganic substrate, where sequences that show favourable binding properties are immobilised and those that do not can be washed away. The bound cells are subsequently amplified through cell culture and the protein sequence information is extracted from the genetic code.

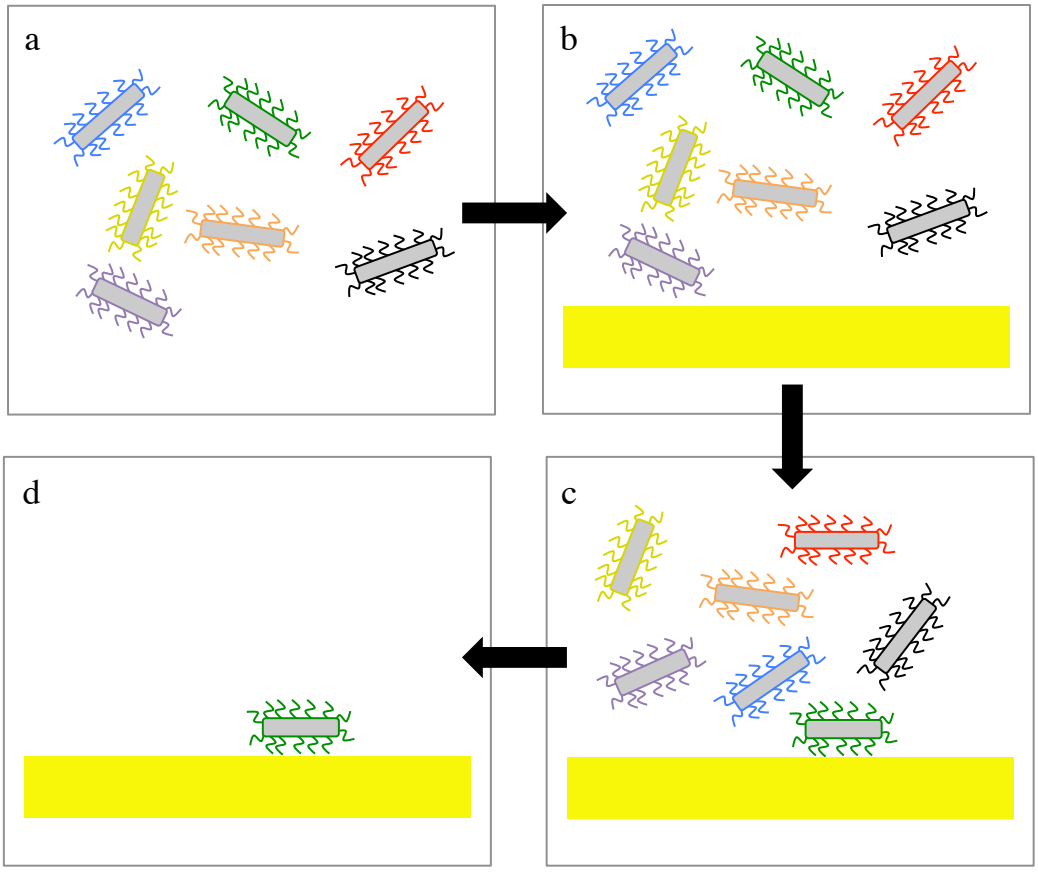


Figure 3.1. Cell surface display technique for identification of sequences that bind to inorganic surfaces: a) *E. coli* surface display library, b) introduction of library to gold substrate, c) binding of *E. coli* to gold via membrane peptide sequence with high affinity, d) non-binding cells eluted leaving binding candidates for sequence analysis.

Despite the advantages of CSD, the mutations contained within each library only explore *ca.* 0.1 % of the available sequence space of a peptide of any length; this is in part due to the variable propensity of DNA codons that express a particular amino acid as well as difficulties arising from translation of DNA into the complimentary protein. Hence, biopanning experiments can be used as a basis for identifying inorganic binders, but also leave scope for sequence design or modification.

Panning experiments have been employed successfully for the identification of a number of metal binding peptides that have subsequently been used in nanoparticle formation.[71-72, 106, 132] One of the most useful sequences is the gold-binding peptide (GBP) (MHGKTQATSGTIQS) known as GBP1, which was discovered from *Escherichia coli* CSD libraries.[132] GBP1 has been used in the controlled synthesis of gold nanoparticles,[196] boasting excellent regulation of particle size.

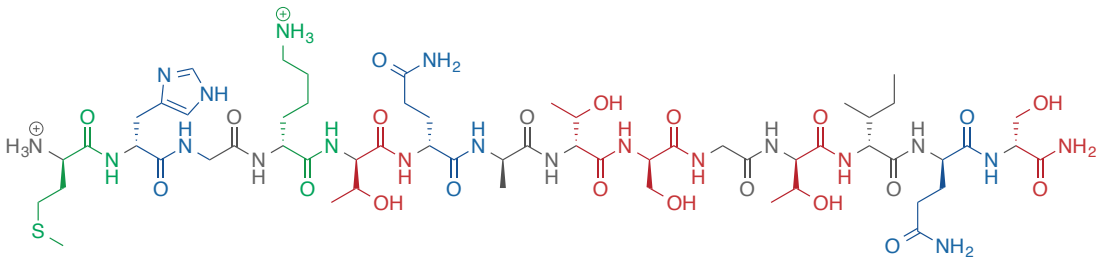


Figure 3.2. Full chemical structure of sequence GBP1, M/K residues are shown in green, S/T in red and H/Q in blue.

Studies on single amino acids have suggested that planar aromatic side chains in tryptophan and phenylalanine show the greatest affinity for gold, however they are absent in GBP1; instead there is strong representation of polar amino acids, in particular those with a hydroxyl moiety (serine and threonine). Cysteine is consciously omitted from this CSD library as it forms covalent bonds with the surface, whereas functional gold binding peptides should be able to form strong *reversible* non-bonded interactions, the lack of other residues may be a result of their limited expression in the CSD library. Despite this, GBP1 does indeed show a strong

affinity for gold[131, 196-197] although the underlying properties that drive its binding are still not fully understood.

Both experimental[107, 132, 198] and theoretical[197, 199-203] studies have focused on the important question of why certain peptides sequences display affinity and specificity for particular metals or crystal architectures. Elucidating the different parameters that govern this control remains a challenge but could ultimately facilitate *de novo* design of surface-selective binding sequences.

3.1.2 Theoretical Studies of Gold Binding Peptides

Molecular dynamics simulations have made the study of microscopic properties of peptides at interfaces accessible. Atomistic simulations have been used to examine the conformational and binding properties of a number of peptides on a range of materials such as gold,[147, 197, 199-201, 204], graphite,[205-208] and silicon-based surfaces.[209-211]

Hypotheses exist as to the origin of GBP1's ability to form strong interactions with Au 111; however these points appear to present only part of the story and leave some certain aspects inconclusive. Investigations of the conformational properties of the 42 residue, 3R-GBP1 (triple repeat of GBP1)[197, 199] propose that binding affinity is an inherent property of the high occurrence of hydroxyl moieties, via serine and threonine side chains. On the other hand, a combined experimental and theoretical study by Tang et al.[200] suggested that three anchoring residues in the N-terminal

region of the sequence (M1, H2 and Q6) allow the adsorbed peptide to experience greater conformational freedom resulting in entropically driven binding.

In a related work, the binding energy of GBP1 to gold has been derived by considering the binding free energies of the individual amino acids that constitute the peptide and assuming that these remain unchanged *in situ* within the peptide chain.[164] Alternative studies have considered the full peptide structure to calculate binding energy via the compartmentalisation method,[147, 212] which provides an indication of the binding potential energy of the peptide through single snapshots of the peptide in the adsorbed and bulk states. However, to provide binding free energies and consider multiple binding modes, dynamic methods should be employed.

3.2 Computational Methods

3.2.1 Background

To gain a comprehensive insight into the relationship between the sequence and binding strength of GBP1 we use a series of mutations, outlined in Table 3.1. In this design, the native **GBP1** sequence is compared to its analogous alanine control, **A14**. Furthermore, a substitute-out/substitute-in mutation approach was used for **GBP1** and **A14** respectively, where mutations were based on three categories of amino acid: single heteroatom side chains (M, K) in **GBP1-MK** and **A14+MK**, hydroxyl residues (S, T) in **GBP1-ST** and **A14+ST**, and two heteroatom side chains (H, Q) in **GBP1-HQ** and **A14+HQ** (see Table 3.1 for nomenclature and complete sequences).

Table 3.1. Peptide sequences under study. **GBP1** is the native sequence and A14 is the alanine control. Mutations from native and control sequences are highlighted.

Sequence	1	2	3	4	5	6	7	8	9	10	11	12	13	14
GBP1	M	H	G	K	T	Q	A	T	S	G	T	I	Q	S
GBP1-MK	A	H	G	A	T	Q	A	T	S	G	T	I	Q	S
GBP1-ST	M	H	G	K	A	Q	A	A	A	G	A	I	Q	A
GBP1-HQ	M	A	G	K	T	A	A	T	S	G	T	I	A	S
A14	A	A	A	A	A	A	A	A	A	A	A	A	A	A
A14+MK	M	A	A	K	A	A	A	A	A	A	A	A	A	A
A14+ST	A	A	A	A	T	A	A	T	S	A	T	A	A	S
A14+HQ	A	H	A	A	A	Q	A	A	A	A	A	A	Q	A

This method allows the influence of each amino acid towards peptide-surface binding affinity to be separated into its ability to form strong non-covalent interactions with the surface on the one hand, and its contribution to destabilization of the unbound peptide in solution on the other, revealing both the contribution of residue binding affinity (in the context of the peptide) and destabilization of the peptide's conformation in solution to the binding free energy of the peptide.

3.2.2 Peptide Equilibration

The GBP1 sequence was built as a straight chain using the guesscoord plugin of VMD.[90] The peptide was then placed in a box of TIP3P[88] water measuring 55 x 55 x 55 Å using the SOLVATE[89] plugin of VMD. The system comprised of *ca*. 15,500 atoms, including *ca*. 5000 water molecules giving a density of approximately 0.90

g/cm^3. Stoichiometric numbers of Na^+ or Cl^- ions were added to the system to neutralize the overall charge. The solvated peptide system was then equilibrated for 20 ns with the NAMD[91] molecular dynamics engine, employing the CHARMM22[188] force field. Each simulation was performed with a timestep of 2fs and commenced with a 2,000 step minimization period to remove bad contacts. The temperature was increased from 10 K to 300 K at 10 K increments every 1000 steps, the temperature was controlled using Langevin dynamics and the pressure was kept at 1 atm using the Langevin piston method[213] at a piston period of 100 fs and constraint of 50 fs, with piston temperature of 300 K. The Particle-Mesh-Ewald[93] method was employed for full electrostatics, with a grid spacing of 1 Å. Covalently bonded atoms were kept rigid using the SETTLE[94] extension of the SHAKE[95] constraint algorithm.

The simulation predicted an equilibrated structure with two distinct regions (Figure 3.3), with the N-terminal region (MHGKTQA) adopting a coiled structure while residues 8 to 14 of the peptide (TSGTIQS) adopt an extended conformation.

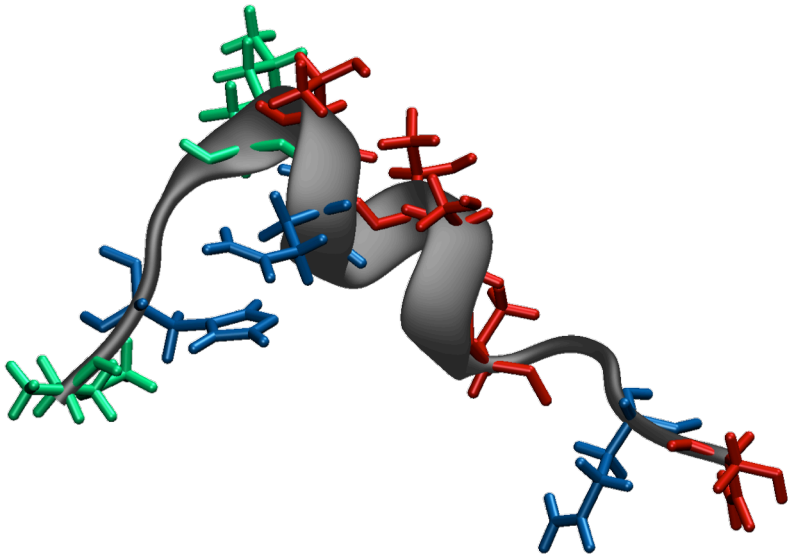


Figure 3.3. Final structure of GBP1 after a 20 ns equilibration simulation in TIP3P solvent with periodic boundary conditions. M/K residues are shown in green, S/T in red and H/Q in blue.

This structure is consistent with the available NMR data for this sequence,[214] thus validating the ability for the methods employed to predict physically stable structures of the peptide.

To ensure that a 20 ns simulation was sufficient to result in the equilibrated structure of GBP1, the simulation was extended to 250 ns. The RMSD of the structure from the initial structure, Figure 3.4 shows that there is an initial change in the overall structure of the peptide. After 20 ns, there are continued fluctuations in the structure, consistent with the relatively flexible nature of the uncoiled regions of the peptide, although the overall partially coiled and partially uncoiled aspect of the peptide

remains; 20 ns was therefore deemed sufficient for subsequent equilibrations of peptides of this length.

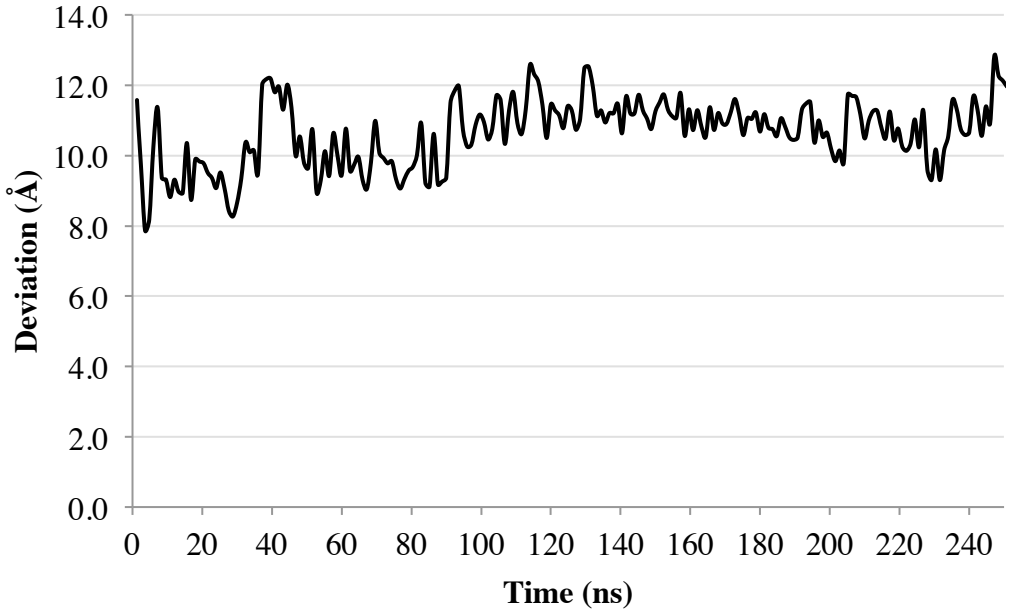


Figure 3.4 Deviation (Å) of the peptide backbone relative to the straight chain starting structure over an extended 250 ns simulation.

From the equilibrated structure of GBP1 after 20 ns, the mutated sequences were built again using the guesscoord algorithm of VMD. Each peptide was again placed in a box of TIP3P water measuring 50 x 50 x 50 Å with the genbox plugin of GROMACS[192] to a density of approximately 1.00 g cm^{-3}. This solvation method was chosen based on observations from initial tests of metal-water systems. In systems where the density of water is lower than 1.00 g/cm^3, as is the case when SOLVATE is employed, voids appear in the system upon equilibration. Therefore, re-equilibration of each sequence was performed in a box of water matching the correct

density before introduction to the metal surface. The equilibrations of the native and mutated sequences were then performed under the same temperature and pressure conditions as before with the CHARMM-Metal[187] force field being employed for comparability between the bound and unbound states.

3.2.3 Peptide Adsorption

The same pre-equilibrated water-gold system used in Section 2.2.2 for the adsorption of single amino acids was again used as the starting point for the peptide adsorption study. Each peptide in its final conformation from the equilibration simulation was added to the equilibrated water–metal system. To ensure that there was no bias between the peptide configuration and adsorption behaviour, each peptide was placed at 6 different orientations above the metal surface (Figure 3.5), so that the closest atom of the peptide was a distance of 6 Å from the metal surface. Water molecules overlapping with the peptide were removed, both the peptide and the metal were fixed and a short 5 ns equilibration of the water was carried out.

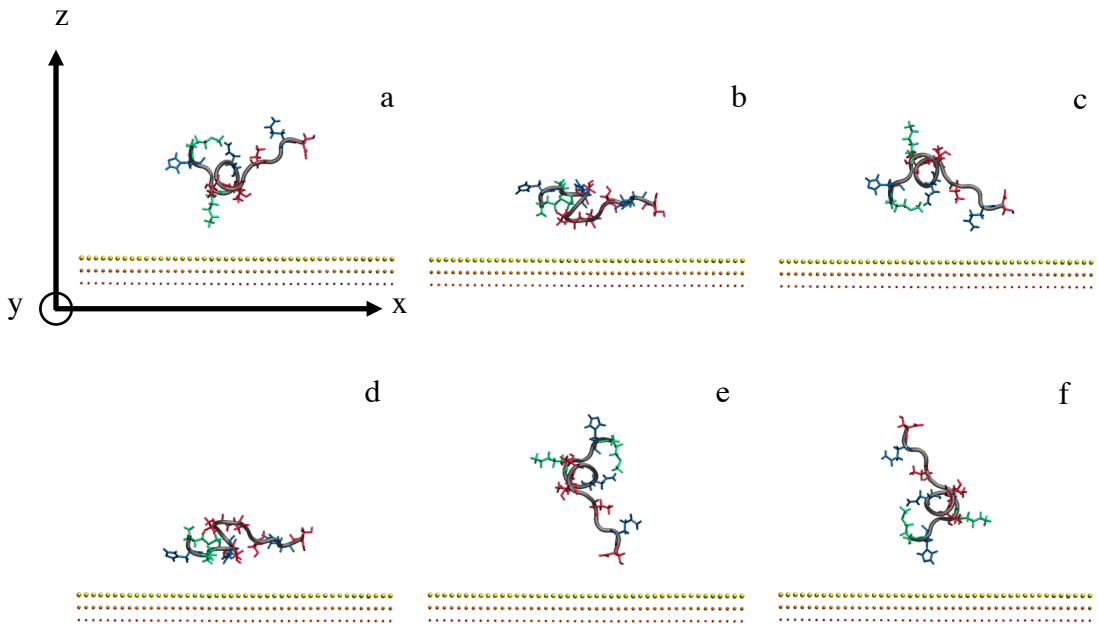


Figure 3.5 The 6 different orientations of GBP1 above the metal surface, a – d show rotations in the x-axis of 0, 90, 180 and 270 º and e – f show rotations in the y-axis of 90 and 270 º (water emitted for clarity).

After the equilibration of the water, the peptide, in each of the six orientations, was allowed to adsorb to the metal over a period of 70 ns. Such a comparatively long simulation time was employed so that the peptide could adopt its fully adsorbed configuration. All simulations were carried out at 300 K in the NAMD 2.8[91] molecular dynamics package, employing the CHARMM-METAL[187] force field to describe the gold, CHARMM22[188] for the peptides and the TIP3P[88] model for water potentials.

The peptide was considered to be adsorbed if the centre of mass was within 4.5 Å of the metal surface. Above this cut-off, binding energies were found to differ significantly and thus did not represent adsorbed states of the system. The same

adsorption protocol was employed for all sequences and using these criteria, it was found that for each mutation, at least four starting configurations adsorbed to the surface within the equilibration period. The centre of mass for the adsorbed peptides that were used in the subsequent desorption simulations varied between 3.30 Å and 4.50 Å from the surface Details of the centre of mass distances for each peptide adsorption are found in Table 3.2.

Table 3.2. Centre of mass distances (Å) of the adsorbed conformations for sequences GBP1, A14 and associated mutations. Simulations with a centre of mass > 4.50 Å are shown in red.

	Orientation					
Sequence	a	b	c	d	e	f
GBP1	4.37	4.40	4.17	4.65	4.79	4.41
GBP1-MK	3.91	3.97	3.89	4.46	3.73	4.50
GBP1-ST	4.92	4.18	3.63	4.07	4.24	4.27
GBP1-HQ	3.68	3.87	4.32	3.69	3.78	4.02
A14	3.47	3.52	3.30	3.52	3.70	3.40
A14+MK	3.52	3.71	4.08	3.34	3.58	3.91
A14+ST	3.47	3.92	4.43	4.49	3.69	3.87
A14+HQ	4.22	4.41	4.24	5.51	4.16	4.30

Examination of the adsorbed structures (Figure 3.6), indicated that the peptide prefers to orientate such that the heteroatoms are situated in the space between atoms in the top Au layer of the surface, above the atoms of the second layer

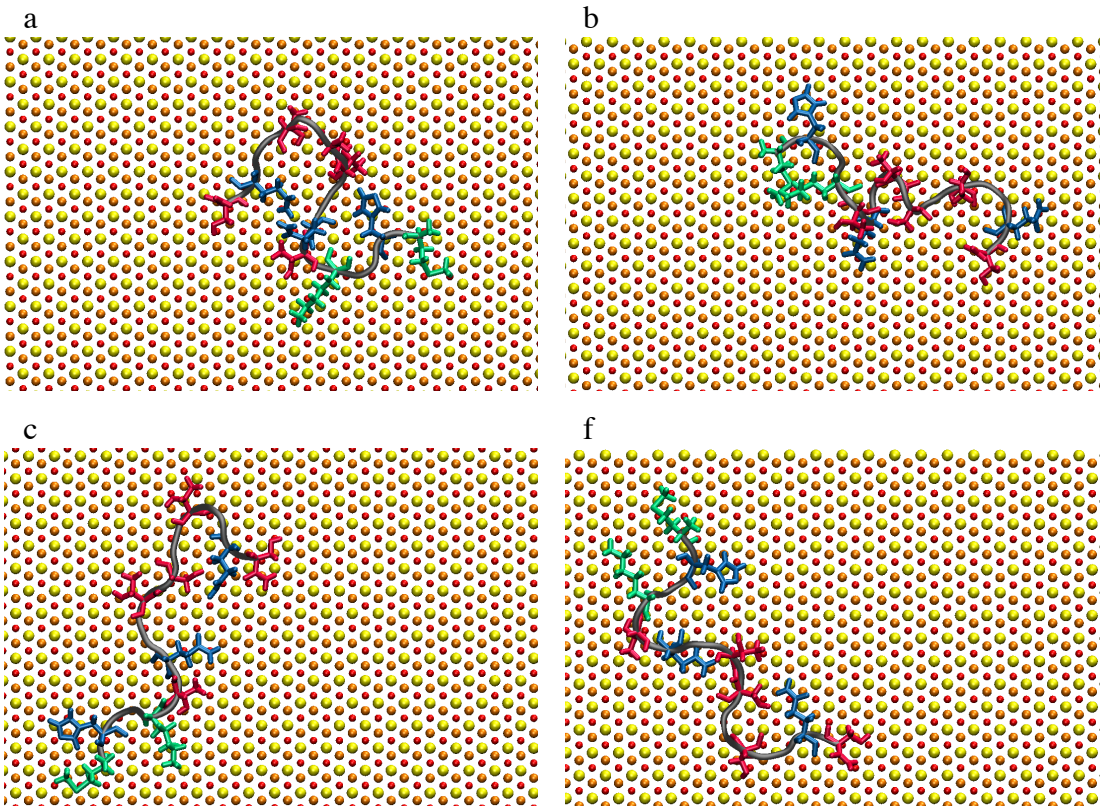


Figure 3.6. Orientations of the sequence GBP1 adsorbed to the Au 111 surface. Methionine and lysine residues are shown in green, hydroxyl residues in red and histidine and glutamine residues in blue. Yellow sphere's indicate the top layer of the gold, orange spheres the second layer and red spheres the third layer.

Furthermore, the alignment of the polar groups, especially glutamine, coincides well with the soft-epitaxial mode of adsorption, which has been observed previously in the literature[147, 165] and Chapter 2, where the Au (111) surface geometry creates binding sites for polar side chains.

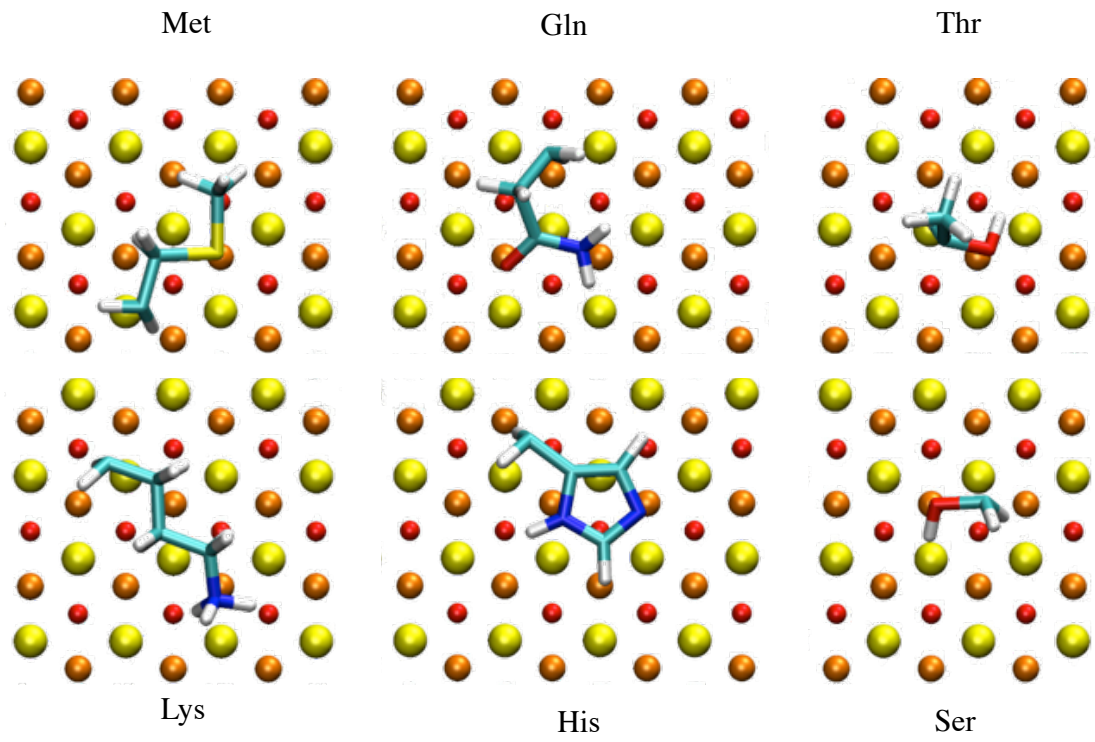


Figure 3.7 Typical adsorbed conformations of amino acids involved in **GBP1** and **A14** mutations.

3.2.4 Desorption of Peptides

Adsorbed peptides were pulled from the surface by the C_α of residue 7 in the z-axis. The peptide was pulled at a rate of 0.005 Å ps^{-1} with a spring constant of 500 kcal mol^{-1} $Å^{-2}$ so that the reaction coordinate of the simulation followed the predicted profile, the peptide was moved 40 Å from its starting orientation to ensure full desorption. This process was repeated 30 times for each adsorbed orientation, giving up to a total of 180 desorption SMD simulations for each peptide sequence.

An important feature of our methodology is that free energy calculations were performed on the peptide sequence in its entirety, as opposed to the combination of single amino acids, hence the impact of each type of residue on binding strength was measured in the context of its native peptide environment.

3.3 Binding Free Energies of GBP1 and Mutations

The resulting free energy curve for the native peptide, **GBP1**, is shown as an example in Figure 3.8.

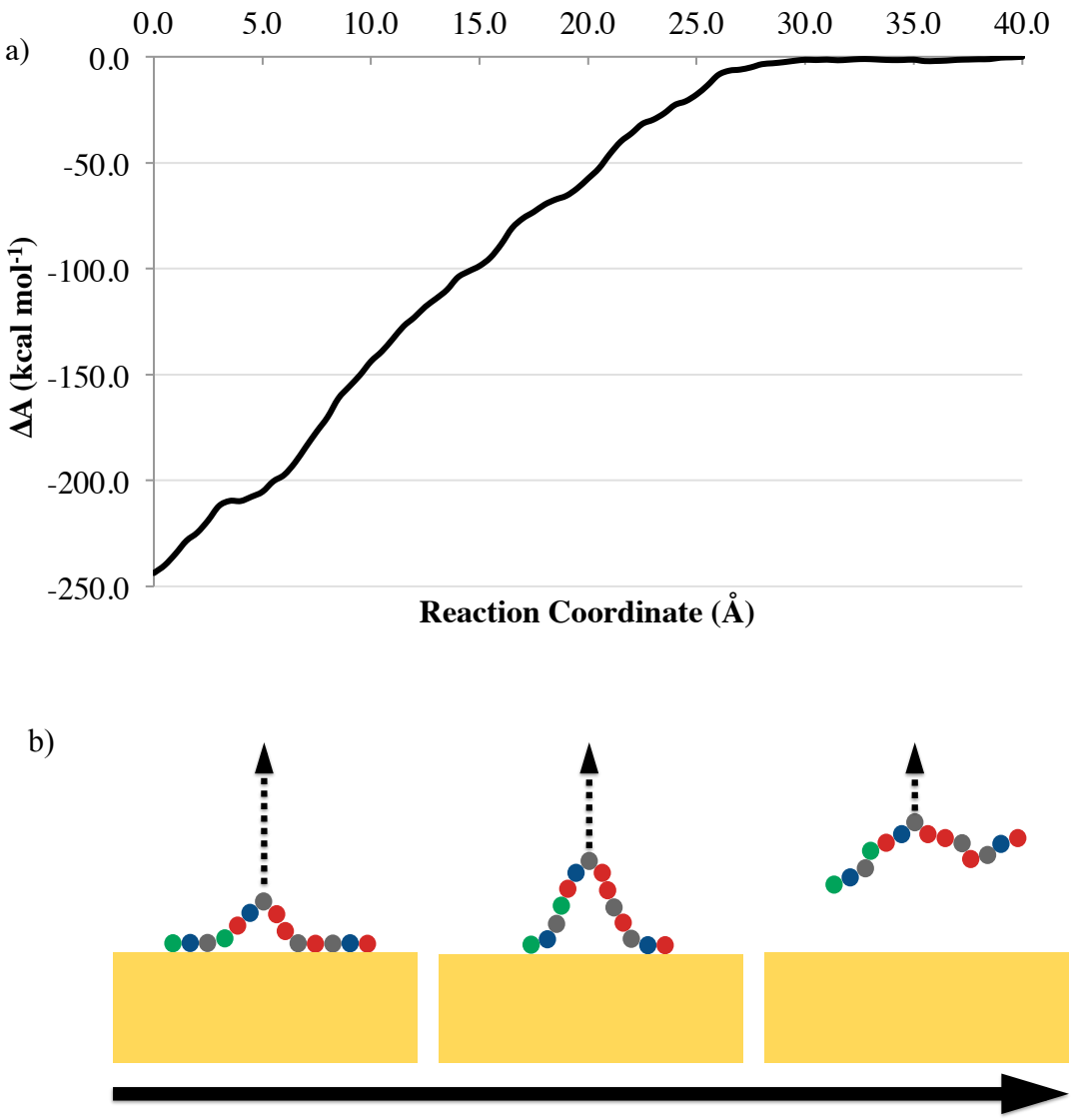


Figure 3.8. a) The change in binding free energy, ΔA, plotted as a function of desorption reaction coordinate for **GBP1**. b) Illustration of desorption process.

The convergence of binding free energy as a function of the number of simulations used in the bootstrapping analysis is shown in Figure 3.9, we defined ΔA as being converged when the associated errors were less than ± 0.5 kcal mol^{-1}.

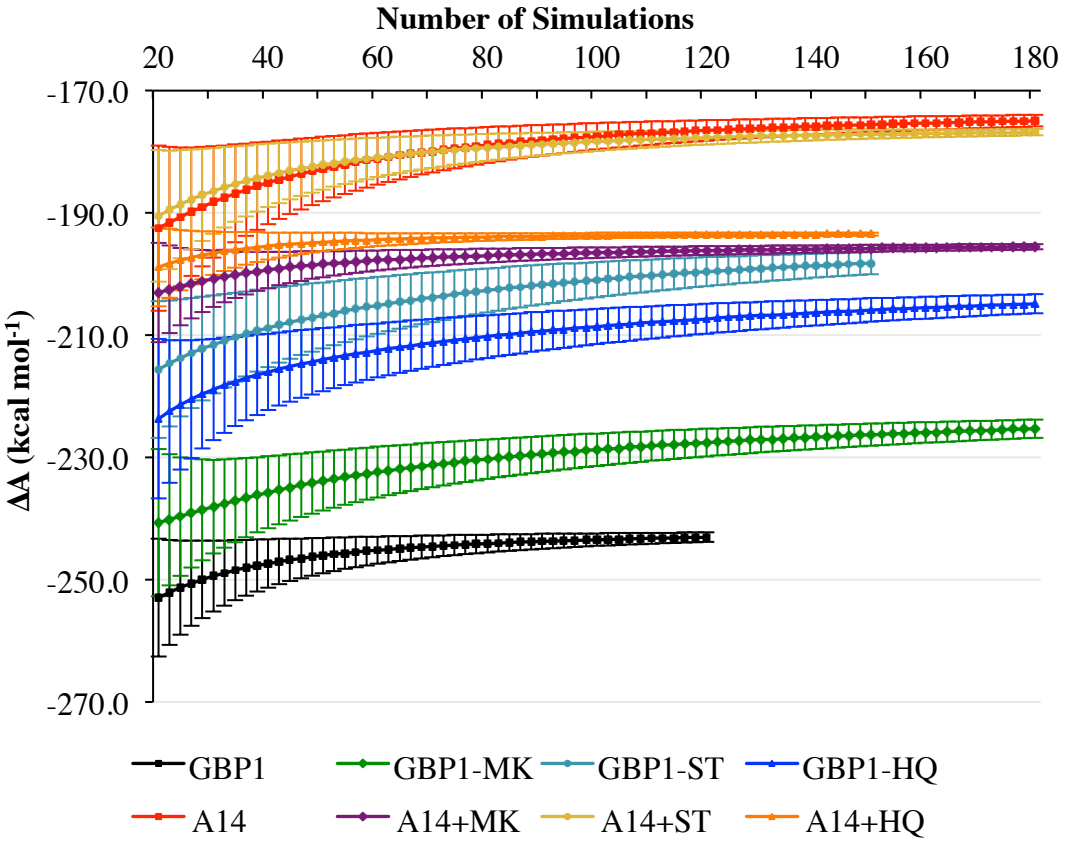


Figure 3.9. Change in ΔA as a function of the number of simulations used in the bootstrapping analysis for all adsorbed conformations of the different peptides.

ΔA values reported herein are the difference between the adsorbed state at 0 Å and the bulk state at 40 Å. The binding free energies for all eight sequences are shown in Table 3.3.

Table 3.3. Binding free energies, ΔA, for sequences **GBP1** to **A14+HQ**.

Sequence	ΔA (kcal mol^{-1})
GBP1	-243.0
GBP1-MK	-225.3
GBP1-ST	-198.2
GBP1-HQ	-204.8
A14	-174.9
A14+MK	-195.5
A14+ST	-176.8
A14+HQ	-193.5

As expected, the native **GBP1** sequence shows the greatest binding free energy (ΔA_{GBP1} = -243.0 kcal mol^{-1}) and the alanine control **A14** shows the weakest binding (ΔA_{A14} = -174.9 kcal mol^{-1}). Upon substitution of M1 and K4 for alanine (**GBP1-MK**) a loss in binding strength of +17.7 kcal mol^{-1} is observed, representing a loss of approximately +8.9 kcal mol^{-1} per residue. Mutation of the control sequence to include methionine and lysine (**A14+MK**) shows a recovery in binding strength of -20.6 kcal mol^{-1} (-10.3 kcal mol^{-1} per residue) similar to the loss of the corresponding **GBP1-MK** mutation. A loss in binding free energy of +9.0 kcal mol^{-1} per residue was found for the mutation of hydroxyl amino acids ($\Delta\Delta A_{GBP1/GBP1\text{-}ST}$ = +44.8 kcal mol^{-1}); however the introduction of hydroxyl moieties in sequence **A14+ST** does not yield any significant increase in binding strength over the control sequence **A14** (ΔA_{A14} = -174.9 kcal mol^{-1}, ΔA_{A14+ST} = -176.8 kcal mol^{-1}). A similar discrepancy is observed for the mutation of histidine and glutamine residues **GBP1-HQ**, where the loss of +38.2 kcal mol^{-1} (+12.7 kcal mol^{-1} per residue) upon removal from the native sequence is more substantial than the gain of -18.6 kcal mol^{-1} in binding free energy strength when replacing alanine in the control, (**A14+HQ**).

If the binding strength were governed solely by interaction strength of the single amino acids with the gold surface, the loss observed in substitute-out mutations would be equivalent to the gain in corresponding substitute-in sequences. Table 3.4 shows the difference in binding free energies as calculated for the full peptide and by combination of charged termini single amino acid binding free energies calculated in Chapter 2.

Table 3.4. Comparison of peptide free energies as calculated from full peptide desorption simulations and the combination of single amino acid free energies. All energies in kcal $mol^{-1.}$

Sequence	ΔA Full Peptide	ΔA Amino Acid Combination	ΔΔA
GBP1	-243.0	-163.5	-79.5
GBP1-MK	-225.3	-139.0	-86.3
GBP1-ST	-198.2	-141.1	-57.1
GBP1-HQ	-204.8	-130.3	-74.5
A14	-174.9	-85.2	-89.7
A14+MK	-195.5	-109.6	-85.9
A14+ST	-176.8	-107.5	-69.3
A14+HQ	-193.5	-118.3	-75.2

Hence, these results demonstrate that binding strength is not influenced solely, or even predominantly, by the affinity that individual amino acids have for the Au (111) surface. Furthermore, the combination of single amino acid free energies neither quantitatively nor qualitatively reproduces the relative binding free energies of the full peptide sequence. These phenomena can be explained by changes in the stability

of the peptide in the water environment upon sequence mutation in addition to the loss of anchoring points between the peptide and gold. For example, replacing polar residues for non-polar alanine in the substitute-out mutations reduces the ability of the peptide to stabilize interactions with water in the bulk solution, hence it is more favorable for the peptide to remain on the surface, enhancing binding strength. Conversely, introduction of polar residues in the substitute-in mutations stabilizes the peptide in the water bulk and consequently decreases the binding strength. The significance of this effect is clearly dependent on the sequence and resulting properties of the entire peptide.

The effects of the nature of residues on the conformation of the peptide in the bulk water are clearly observed in Figure 3.10. **GBP1-ST** retains the coiled character of the **GBP1** peptide, while **A14+ST** adopts a more open conformation; consequently serine and threonine residues can more readily form stabilizing interactions with water, meaning that binding between the peptide and the surface is almost completely diminished due to its affinity for the bulk ($\Delta\Delta A_{A14/A14+ST}$ = -1.9 kcal mol^{-1}). In general, sequences with hydroxyl side chains appear to promote the unwinding of the characteristic coiled structure of GBP1 (Figure 3.10 b, d and g), indicating stabilization of the peptide in the bulk water. Figure 3.10 a, c and h show that sequences containing histidine and glutamine retain coiled character, suggesting that in these mutations the peptide would more readily adsorb on the surface as the peptide overall is destabilized in the bulk compared to the open structures containing serine and threonine. Indeed, sequence **A14+HQ** (Figure 3.10 g), which preferentially forms a coil in solution, gives only a partial loss in binding free energy

relative to **GBP1-HQ** ($\Delta\Delta A_{GBP1/GBP1\text{-}HQ}$ = +38.2 kcal mol^{-1} $\Delta\Delta A_{A14/A14+HQ}$ = -18.6 kcal mol^{-1}) as a result of less favorable interactions with the bulk water.

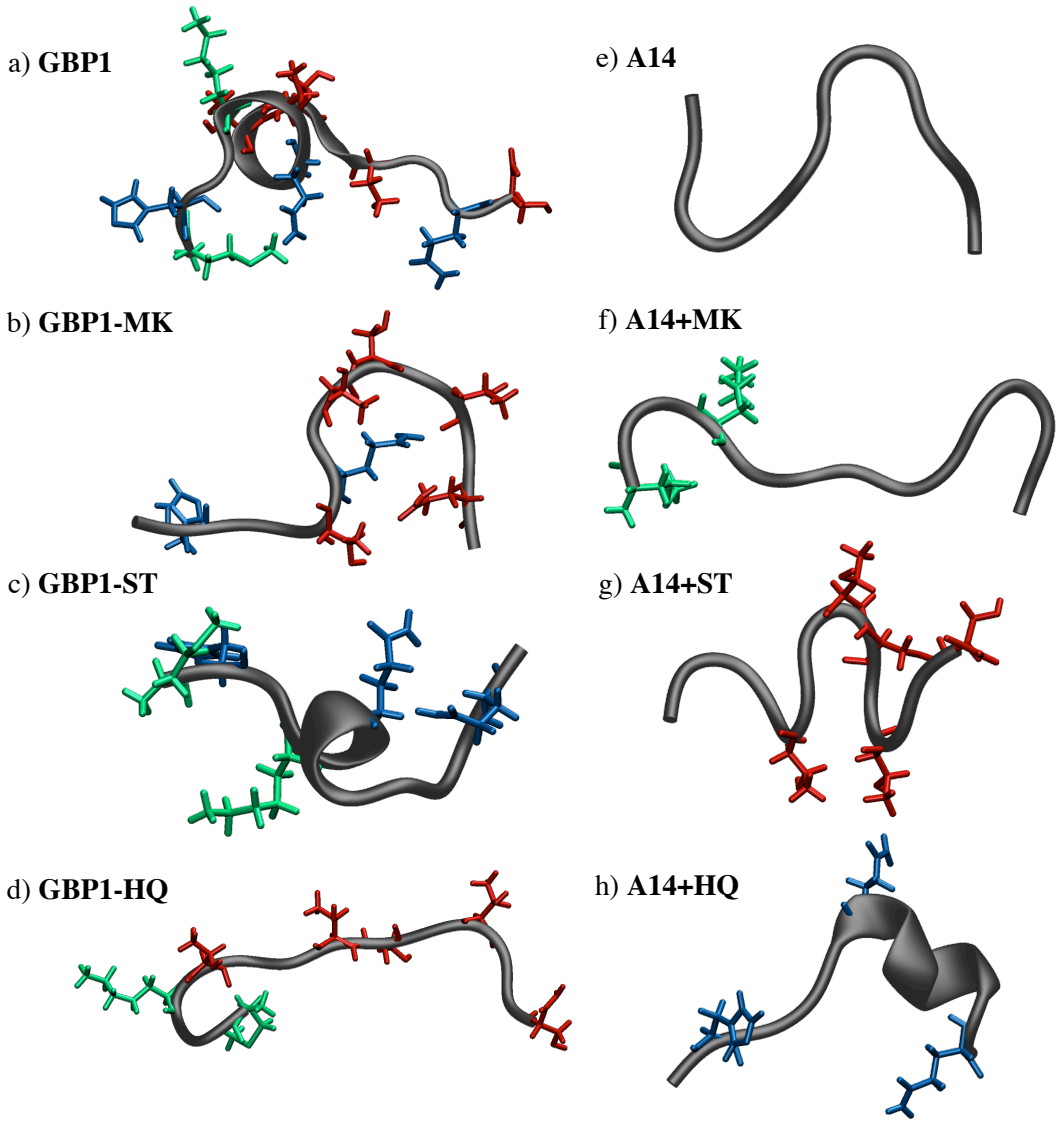


Figure 3.10. Structures of sequences **GBP1** and its three mutations (a – d) and **A14** control and its mutations (e – h) after 20 ns equilibration in water. Residues involved in mutations are shown explicitly: M/K shown in green, S/T in red and H/Q in blue.

Unlike the coiled structure of **GBP1** (Figure 3.10 a) the alanine control, **A14** (Figure 3.10 e) shows no indication of secondary structure formation. Particularly interesting

is the final structure at the completion of the 20 ns simulation of **GBP1-MK** (Figure 3.10 b), where the presence of both H, Q and hydroxyl moieties results in both a loss of coil and a closing of the peptide from positions 5 to 14. This may suggest a clash between the solvent stabilizing effects of S and T and the promotion of a closed structure by histidine and glutamine. The contribution of methionine and lysine remains almost equivalent for the substitute out (Figure 3.10 b), substitute in (Figure 3.10 f) mutations, ($\Delta\Delta A_{GBP1/GBP1\text{-}MK}$ = +17.7 kcal mol^{-1} $\Delta\Delta A_{A14/A14+MK}$ = -20.6 kcal mol^{-1}), regardless of peptide environment, indicating that for these residues binding free energy is almost entirely a result of side chain interactions with the solid surface and do not significantly alter peptide stability in water.

3.4 Conclusions

In conclusion, by employing a substitute-in/substitute-out mutation approach, it is revealed that the binding strength of peptides to surfaces is a delicate balance between the interactions of the peptide with both the surface and the aqueous solvent. In particular, the combination of strategic mutation and non-equilibrium molecular dynamics protocols reveals that both peptide gold interactions and the stability in solvent shape **GPB1** binding to gold, for which we have found that methionine and lysine contribute approximately -8.9 kcal mol^{-1}, serine and threonine -9.0 kcal mol^{-1} and histidine and glutamine -12.7 kcal mol^{-1} per residue to the binding character. Furthermore, it is demonstrated that a simple combination of amino acid binding free energies neither quantitatively nor qualitatively reproduce the relative binding free

energies of the full peptide; hence the peptide as a whole should be considered when tailoring the surface binding properties of peptides.

4 Crowding Effects on the Binding Properties of a Gold Binding Peptide GBP1

4.1 Introduction

Theoretical studies of biomolecules adsorbed to gold usually consider a single molecule and the surface.[147-148, 197, 199-200, 215] This approach allows structural and binding characteristics to be attributed purely to the peptide or protein in question as well as reducing system complexity with respect to aspects such as conformational sampling. However, biointerfacial systems rarely exist as single molecules bound to a surface in isolation, but rather as a crowded layer of adsorbates, randomly oriented or in the form of a self-assembled monolayer; therefore to compare more effectively with experiment the effects of crowding at the surface should be taken into consideration. Some studies[200] have already alluded to the potential importance of studying multiple peptides at the surface; however they concede that associated difficulties can be significant.

Computational examination of multiple adsorbates on solid surfaces has thus far been generally limited to self-assembled monolayers of alkyl thiols,[216-226] with only one instance of peptide coated structures;[227] these studies however focus only on the conformational properties of the system and do not investigate binding affinity to the gold or to neighbouring chains.

Although single molecule experiments can, and indeed have, provided significant insight into the properties of surface binding peptides, the next challenge in bridging

the gap between simulation and experiment is examining the effect of interpeptide interactions on interfacial properties. Herein, we seek to improve our model in order to strengthen the predictive power of molecular dynamics simulation for comparison with experimentally derived observations and to explore the challenges and insights associated with studying multiple peptides at the Au 111 interface.

4.1.1 Gold Surface

In the previous two chapters we have employed the CHARMM-Metal[187] force field to describe the gold surface; the relatively simple implementation of Van der Waals terms is generally considered more than adequate for studying protein interactions with gold and has been demonstrated in a number of cases;[147, 199, 212] however in such force fields polarisation effects are neglected. Polarisation in interfacial systems is normally described using the image charge effect,[228] where as a charged particle approaches a metal surface it induces an identical and opposing charge at an equal distance from the surface plane, Figure 4.1.

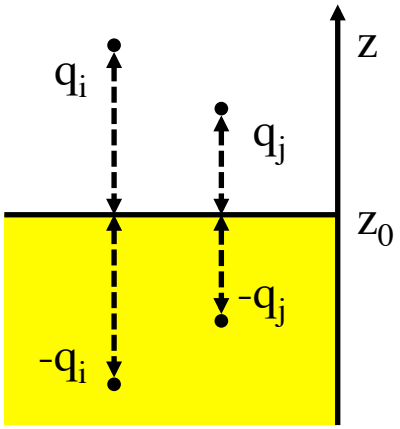


Figure 4.1. Representation of the image charge effect, where an equal and opposite charge in the surface is induced upon interaction of a charged particle.

The significance of the image charge effect is essentially system dependent, as studies have shown that its contribution is negligible in pure liquid systems.[229-232] In these examples however the solvents are both neutral and have a high dielectric constants (ethanol: $\varepsilon = 24.3$, water: $\varepsilon = 80.1$) meaning that any induced charge is very small and that the image terms are effectively negated upon summation. However, it is common for proteins/peptides to have a non-zero charge through charged termini or side chains, therefore image charge effects are more important when studying biomolecular interactions with surfaces. Image charge effects are a particularly important consideration in the case of multiple adsorbates as surface polarisation was shown to change the interaction energy between two ions in close proximity to the surface;[228] hence there may be significant impact on interpeptide attraction in addition to peptide surface binding.

Classical force fields may implicitly include image charge effects if parameterised to fit values obtained from *ab initio* calculations where polarisation of the metal is explicitly incorporated; despite this image charge effects cannot be adequately described using a classical two-body potential. One approach however is to describe the polarisability of the surface explicitly through inclusion of dipoles in the classical force field. Iori and Corni described the rigid-rod model including image charge effects in molecular dynamics simulations.[233] In this model both a partially positive and partially negatively charged particle are linked via a rigid constraint of length l_0; in the gold surface the negative particle is held static at a distance of 0.7 Å whilst the positive component can move freely within the confines of the fixed distance, Figure 4.2.

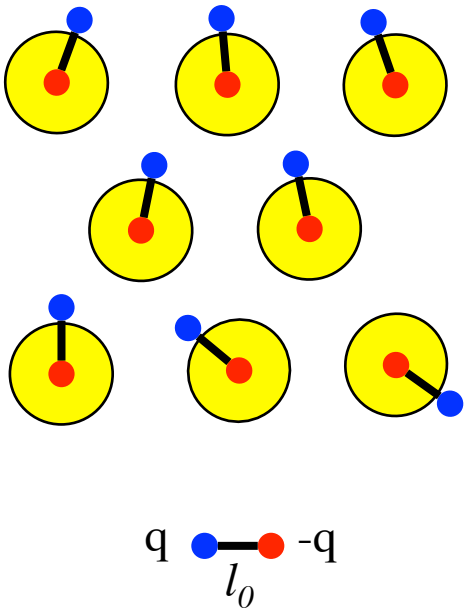


Figure 4.2. Illustration of the rigid-rod model applied to a surface where positive charges are shown in blue and negative charges in red.

The effect is similar to charge induced dipole forces and thus image charge effects are accounted for as the surface can adapt to its surroundings. This method is advantageous in that it is not computationally expensive and does not require modification of the software source code; thus can be implemented easily.

As our peptide of interest, GBP1, contains charged and polar moieties the inclusion of a polarisable surface seems an appropriate improvement to the model. The polarisable GolP force field developed by Corni *et al.* utilises the rigid rod method and has been adapted for OPLS[154] and CHARMM[234-235] type force fields, the latter being most important for consistency with our previous studies.

4.1.2 Conformational Sampling

As mentioned briefly in Section 2.1.4, sampling of large biomolecules in MD simulations is a major challenge in biophysics. Multiple minima can exist in the energy landscape of proteins and large barriers can cause a particular conformation to become kinetically trapped. This phenomenon becomes more prominent in large systems, thus advanced simulation methods have been developed to allow the potential energy surface of proteins to be surveyed more thoroughly in molecular dynamics. Replica exchange molecular dynamics (REMD), also known as parallel tempering is a method of exploring multiple states of a system by performing multiple simulations in parallel, at different temperatures. Conceptually, the replica exchange method was devised as early as 1986 by Swendsen and Wang;[236] however

it was adopted more widely upon its implementation to protein folding in MD simulations.[237]

In REMD simulations M replicas are placed at range of different temperatures, i.e., $T_0 < T_1 < T_2 < \ldots < T_M$, where T_0 is the temperature of interest. Performing simulations at higher temperature means that prohibitive energy barriers can be overcome more easily compared to at lower temperatures. Simulations are performed in the canonical ensemble and replicas at adjacent temperatures then attempt to exchange coordinates. The transition probability, P_{ij} between two temperatures *i* and *j* is determined using the Metropolis-Hastings criterion:[238]

$$P_{ij} = \left(1, e^{(E_i - E_j)\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)}\right) \qquad \text{Equation 4.1}$$

In this model if the difference in energy, ΔE_{ij}, after exchange is lower than before, the swap is automatically accepted. If not, the exchange with a probability of $(-\Delta E_{ij}/k_B T)$ is accepted; thus allowing changes that increase the energy of the system slightly, although they are less probable. Once an exchange has occurred the new momenta, p^{i}', for the particles are determined:[237]

$$p^{i\prime} = \sqrt{\frac{T_{new}}{T_{old}}}\, p^{i} \qquad \text{Equation 4.2}$$

Where p^i is the old momenta of the system. This process is repeated periodically so that low temperature replicas can move to high temperatures over the course of the sampling, for example see Figure 4.3.

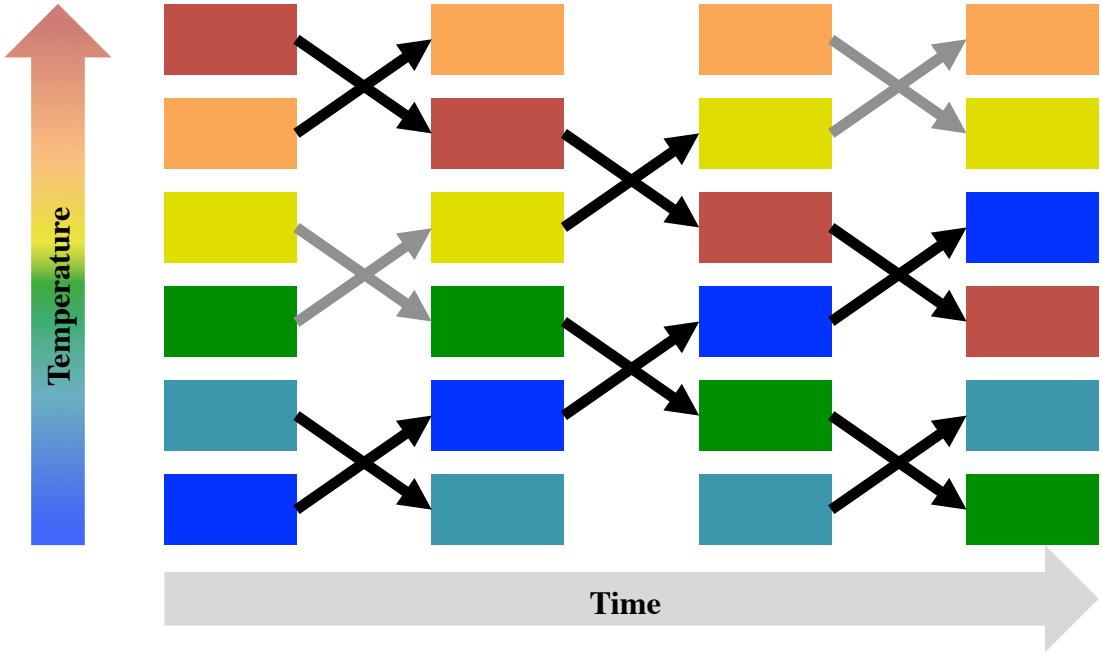


Figure 4.3. Illustration of the replica exchange process. Black arrows indicate successful exchange, grey arrow show unsuccessful exchange.

Temperature intervals in REMD should be chosen so that the potential energies of adjacent temperatures overlap; the greater the overlap, the higher the probability of exchange and greater chance that each replica will visit all temperatures. For protein simulations in explicit water generating a range of temperatures that will allow efficient exchange over all temperatures can be difficult as with such systems energy fluctuations can be large. This means that the probability of ΔE_{ij} meeting the

transition probability criteria is low unless temperature intervals are sufficiently small; thus requiring more replicas.

The large number of replica temperatures required for protein systems is one of the main difficulties in employing temperature REMD methods; hence Hamiltonian replica exchange approach has been favoured for efficient sampling of biomolecules in water. One of the most popular implementations of Hamiltonian replica exchange molecular dynamics (H-REMD) is replica exchange with solute tempering[239-240] (REST).

The source of large energetic fluctuations in biophysical simulations can generally be assigned to the dynamics of explicit water molecules, whereas the protein solute energies show much lower deviation. Since the molecule of interest in most systems is the protein component it makes sense to avoid sampling of the solvent, as when a system is in thermodynamic equilibrium, the average fluctuations of water remain fairly stable, thus the differences over time are very small.

In REST type sampling, only the solute molecule is heated, this is achieved by scaling all of the protein interactions. Firstly, the potential energy, V, is divided into solute (s) and solvent (w) components.

$$V = V_{ss} + V_{sw} + V_{ww} \qquad \text{Equation 4.3}$$

V_{ss} is the non-covalent potential energy of the multiple solute molecules, V_{sw} is the solute-solvent non-covalent potential and V_{ww} is solvent-solvent potential. The potential is then modified to reflect the desired replica temperature through modification of the potentials:

$$V_T = \frac{\beta_T}{\beta_0} V_{ss} + \sqrt{\frac{\beta_T}{\beta_0}} V_{sw} + V_{ww} \qquad \text{Equation 4.4}$$

Where T is the replica temperature, β_T is 1/T and β_0 is $1/T_0$ at the reference (lowest) temperature.

This scaling reduces the strength of the protein interactions with its surroundings as would occur in systems at higher temperature; thus potential energy barriers between states are flattened.

The REST[239-240] and REST2[241] methodologies have been applied to the sampling of proteins in solution[242-243] and at the solid interface,[200, 244] in this study we aim to expand this to multiple peptides at the gold surface.

4.2 Computational Methods

4.2.1 Background

From quartz crystal microbalance (QCM) experiments[200] the gold surface was found to be saturated by GBP1 at approximately 100 ng cm^{-2} with a 97.7 % coverage; this means that for a surface measuring 5.86 nm x 6.09 nm, approximately 14 GBP1 peptides are present on the surface, Table 4.1.

Table 4.1. Outline of terms used to calculate surface density of GBP1 on Au 111.

Mass of 1 molecule	1447.62	amu
	2.40×10^{-21}	g
	2.40×10^{-12}	ng
Molecules/ng	4.16×10^{11}	
Molecules/100 ng	4.16×10^{13}	
Molecules/cm^2	4.16×10^{13}	
Molecules/nm^2	0.42	
Surface dimension (x)	5.86	nm
Surface dimension (y)	6.09	nm
Surface area	35.69	nm^2
Molecules/Surface	14.85	
Surface Coverage	0.977	
Number of Molecules	14.50	

It was found that all peptides could not be placed flat on the surface as this would have led to significant overlap and steric clashes; therefore in order to accommodate the 14 peptides on the relatively small surface area, they were placed perpendicular to the surface with alternating termini facing towards the gold to avoid bias, Figure 4.4.

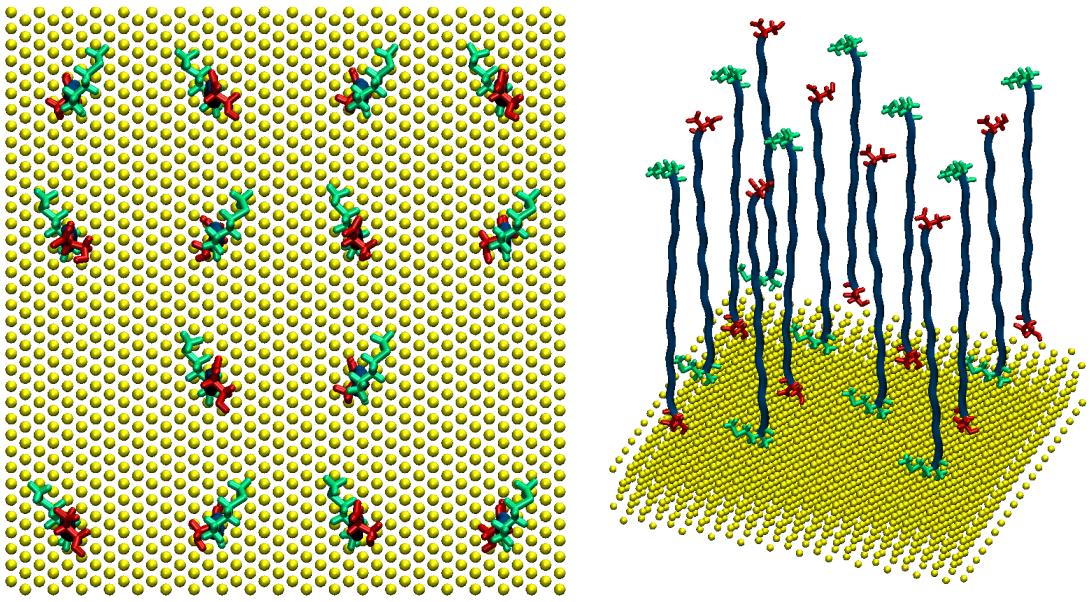


Figure 4.4. View from above (left) and from the side (right) of the surface showing the initial arrangement of 14 GBP1 peptides. Methionine (N-terminal) is shown in green, serine (C-terminal) in red and the backbone in dark blue. Water omitted for clarity.

The peptides were placed at 14.650 Å intervals in the x-axis with a 7.325 Å buffer at each side and a 15.225 Å spacing in the y-axis with a 7.613 Å buffer. Inclusion of periodic boundary conditions means that any individual peptide is adjacent to another with an opposite orientation thus providing system continuity.

By creating a system in which the concentration of peptides at the surface is the same as that found experimentally we aim to assess whether the binding free energy of GBP1 to gold will change upon interaction with adjacent peptides at the surface. Thus, to ensure the best quality model for comparison to experimental results we

improve our system by considering peptide surface concentration, surface polarisability and enhanced conformational sampling.

4.2.2 Polarisable Gold Surface

There are three key elements to the implementation of GolP-CHARMM in Gromacs 5: parameters, restraints and constraints. In the GolP-CHARMM force field there are four particle types that make up the gold surface, AUS (surface), AUB (bulk), AUC (charged) and AUI (interaction); the properties of these particles are given in Table 4.2.

Table 4.2. Parameters for four types of gold atom in GolP-CHARMM force field with respect to Au-Au interactions.

Atom Type	Mass (amu)	Charge	σ (Å)	ε (kJ mol^{-1})
AUS	196.967	-0.300	0.000	0.000
AUB	196.967	-0.300	0.380	0.480
AUC	0.500	0.300	0.000	0.000
AUI	196.967	0.000	0.380	0.480

In the CHARMM-Metal force field employed in Sections 2 and 3, it was found that polar atoms of the peptide reside upon interaction sites in the electrostatic grooves of the surface called soft epitaxial binding sites; however there is growing evidence that this phenomenon is a result of the force field rather than representative of the real adsorption mode. The strongest evidence for this comes from DFT calculations,[154, 160, 163] which show that heteroatoms prefer to sit atop of gold atoms on the top layer of gold atoms instead of between them; the GolP-CHARMM force field has been

parameterised with this in mind and implemented with the use of virtual sites. With respect to Van der Waals forces it can be seen from Table 4.2 that only the bulk (AUB) and virtual site (AUI) particles have a Lennard-Jones parameters σ and ε; there are no terms assigned to AUS or AUC.

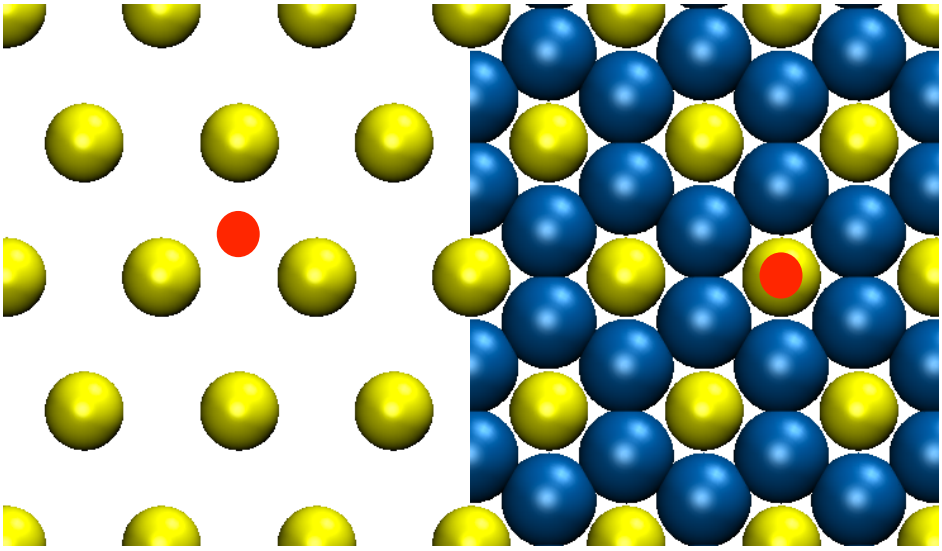


Figure 4.5. Inclusion of virtual sites (blue) drives adsorption atop the surface gold atoms (right) compared to the gaps between the atoms (left) in the standard representation that uses only 'real' gold atoms.

There are two virtual site atoms (AUI) for every surface atom (AUS) and are located in the grooves between the surface atoms; the effect of this is to direct atoms towards interaction sites atop the gold atoms, Figure 4.5, as predicted by quantum mechanical studies.[154, 160, 163] The description of gold surface polarisability has already been outlined in Section 4.1.1, here we show how it is implemented in the GolP-CHARMM force field.

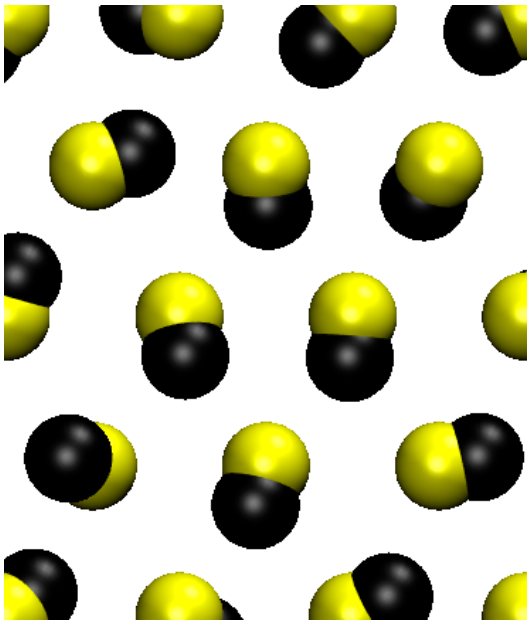


Figure 4.6. AUS/AUB (yellow) and AUC (black) atoms that are used to represent polarisable gold.

Both the surface and the bulk atoms carry a partial negative charge, each is connected by a 0.7 Å constraint to the positively charged AUC particle, Figure 4.6, this is the practical realisation of the rigid rod model shown in Figure 4.2. The surface, interaction and bulk gold atoms are held in place using a restraining force of 5 x 10^6 kJ mol^{-1} (*ca.* 1.2 x 10^6 kcal mol^{-1}) to maintain the surface arrangement of the metal slab; the counter charges (AUC) are allowed to move freely within their constraint distance creating a dipole with the gold atoms that can adapt quickly to its environment.

4.2.3 Replica Exchange with Solute Tempering

The implementation of REST2 has been described by Terakawa *et al.*[245] for the Gromacs 4[192, 246] molecular dynamics package; however its employment in Gromacs 5[247] requires slight modification. Our application of REST2 makes use of the free energy perturbation (FEP) code of Gromacs 5, which is used to scale individual bonded and non-bonded terms.

$$V_{bond} = (1-\lambda)\frac{k_l}{2}(l_i - l_0)^2 \qquad \text{Equation 4.5}$$

$$V_{angle} = (1-\lambda)\frac{k_\theta}{2}(\theta_i - \theta_0)^2 \qquad \text{Equation 4.6}$$

$$V_{torsion} = (1-\lambda)\frac{V_N}{2}(1+\cos(n\tau - (1-\lambda)\phi))^2 \qquad \text{Equation 4.7}$$

Where λ is $(1-\beta_i/\beta_0)$

The spring constants in Equation 4.5 and Equation 4.6 are scaled by 1- λ, as are the phase and barrier height components of Equation 4.7. The intramolecular non-bonded interactions of the peptide, and the intermolecular interactions between peptides, are contained in a "non-bonded neighbour list" in which defined ε_{ij} parameters are listed and again scaled by 1- λ; thus scaling the total V_{ss} potential by β_T/β_0. The solute-solvent interactions are purely non-bonded terms and are scaled in REST2 as:

$$V_{LJ} = \left[\left((1-\lambda)\frac{4\varepsilon_{ij}\sigma_{ij}}{r_{ij}}\right)^{12} - \left((1-\lambda)\frac{4\varepsilon_{ij}\sigma_{ij}}{r_{ij}}\right)^{6}\right] \qquad \text{Equation 4.8}$$

$$V_{Coul} = (1-\lambda)\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \text{Equation 4.9}$$

In Equation 4.4, the V_{sw} potential is shown to be scaled as $(\beta_T/\beta_0)^{1/2}$, this is in reference to the individual ε_{ii} and ε_{jj} terms of the Lennard-Jones potential and the individual charges, q_i and q_j, of the Coulomb potential, which are each scaled by $(1-\lambda)^{1/2}$ as the solute-solvent non-bonded terms are not held in a list. When multiplied, as in Equation 4.8 and Equation 4.9, $1-\lambda$ is returned as a common factor for each component in the potential.

The choice of lambda values is very important, as their intervals must facilitate consistent exchange between replicas. It is almost impossible to intuitively choose temperatures that will give the best exchange probabilities throughout the replica exchange system; hence a simple 'Lambda-Tuning' algorithm has been devised to adapt temperatures based on probability to improve exchange iteratively.[245] In the first instance, even temperature intervals are chosen and subsequent values are refined using Equation 4.10.

$$\lambda_i^{(new)} = \lambda_{i-1}^{(new)} + \left(\lambda_i^{(old)} - \lambda_{i-1}^{(old)}\right)(a_i/\langle a_{i-1}\rangle)^v \qquad \text{Equation 4.10}$$

Where a is the acceptance ratio between the adjacent replicas, $\langle a \rangle$ is the average acceptance ratio across all replicas and v denotes a factor that can be empirically adjusted to improve convergence, it was found that a value of 1/6 gave good convergence.

Initial tests demonstrated that approximately 200 CPU hours were required to run a single simulation for 1 ns; hence the temperature range and number of replicas were carefully chosen to maximise conformational sampling and exchange probabilities whilst managing computational resource efficiency.

Thirty-one replica temperatures ranging from 300 K to 400 K were chosen, with even increments of 3.33 K in the first instance, Figure 4.7. A 5 ns REST2 simulation was performed, requiring *ca.* 30,000 CPU hours; to measure the exchange probabilities. Temperature intervals were modified according to Equation 4.10 after each 5 ns replica exchange simulation until the difference in exchange probabilities converged to less than 0.001; it was found that to achieve convergence, 7 temperature refinements were required for this particular system, Table 4.3, requiring *ca.* 250,000 CPU hours.

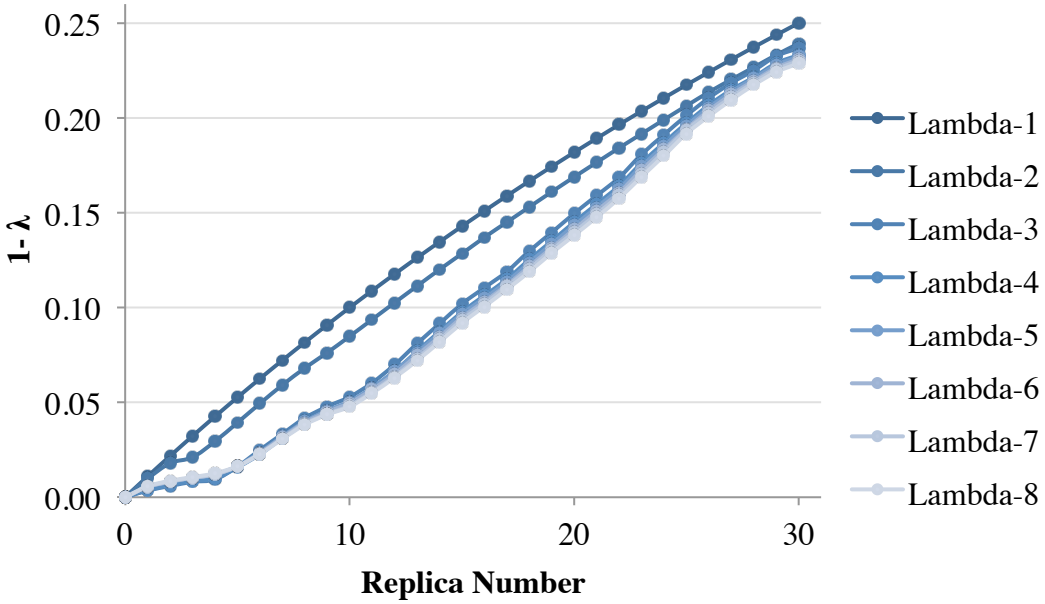


Figure 4.7. Evolution of lambda values adapted over 8 simulations.

The average exchange acceptance was measured for each replica exchange group as well as the total average, Table 4.3. Lambda tuning was used to make exchange over all replicas more homogeneous so that replicas can explore more temperatures.

Table 4.3. Acceptance ratios between adjacent replicas over 8 REST2 simulations.

Replica Exchange Groups	REST2 Simulation Number							
	1	2	3	4	5	6	7	8
0<-->1	0.48	0.13	0.55	0.42	0.39	0.41	0.48	0.32
1 <--> 2	0.53	0.11	0.52	0.38	0.46	0.45	0.53	0.53
2 <--> 3	0.34	0.25	0.35	0.44	0.29	0.22	0.34	0.60
3 <--> 4	0.60	0.05	0.69	0.69	0.71	0.62	0.60	0.45
4 <--> 5	0.23	0.25	0.14	0.11	0.19	0.31	0.23	0.34
5 <--> 6	0.36	0.31	0.09	0.44	0.33	0.25	0.36	0.17
6 <--> 7	0.30	0.31	0.40	0.36	0.32	0.35	0.30	0.31
7 <--> 8	0.32	0.34	0.25	0.35	0.37	0.38	0.32	0.35
8 <--> 9	0.49	0.27	0.34	0.33	0.27	0.23	0.49	0.42
9 <--> 10	0.23	0.21	0.36	0.28	0.35	0.26	0.23	0.46
10 <--> 11	0.41	0.30	0.37	0.29	0.34	0.40	0.41	0.21
11 <--> 12	0.34	0.39	0.22	0.42	0.22	0.41	0.34	0.36
12 <--> 13	0.31	0.46	0.27	0.22	0.43	0.25	0.31	0.33
13 <--> 14	0.33	0.43	0.34	0.34	0.25	0.49	0.33	0.32
14 <--> 15	0.31	0.45	0.32	0.36	0.38	0.43	0.31	0.34
15 <--> 16	0.41	0.36	0.41	0.30	0.40	0.41	0.41	0.39
16 <--> 17	0.36	0.38	0.38	0.39	0.36	0.52	0.36	0.32
17 <--> 18	0.38	0.49	0.28	0.33	0.30	0.27	0.38	0.36
18 <--> 19	0.37	0.44	0.35	0.37	0.37	0.44	0.37	0.33
19 <--> 20	0.37	0.46	0.35	0.35	0.34	0.29	0.37	0.36
20<-->21	0.38	0.44	0.39	0.37	0.44	0.33	0.38	0.35
21<-->22	0.42	0.45	0.39	0.42	0.32	0.44	0.42	0.34
22<-->23	0.35	0.59	0.33	0.37	0.35	0.32	0.35	0.30
23<-->24	0.32	0.50	0.42	0.38	0.42	0.53	0.32	0.34
24<-->25	0.42	0.50	0.42	0.46	0.38	0.53	0.42	0.32
25<-->26	0.45	0.46	0.43	0.36	0.42	0.24	0.45	0.41
26<-->27	0.44	0.43	0.40	0.49	0.39	0.34	0.44	0.39
27<-->28	0.43	0.37	0.39	0.51	0.40	0.64	0.43	0.35
28<-->29	0.37	0.42	0.31	0.28	0.33	0.37	0.37	0.35
29<-->30	0.41	0.25	0.40	0.55	0.41	0.32	0.41	0.39
Average	0.38	0.36	0.36	0.38	0.36	0.38	0.38	0.36

The acceptance ratios for the final REST2 simulation with ranged from 0.17 to 0.60, which is lower than that reported for single peptide systems sampled with REST/REST2 techniques (ca. 0.7);[200, 244] however the average exchange probability

of 0.36 is still deemed sufficient for the adsorption of the fourteen-peptide system to gold, given computational cost restraints.

4.2.4 Peptide Adsorption

The multiple peptide system was solvated with two boxes of TIP3P[88] water measuring 58.6 Å x 60.9 Å x 120.0 Å above the surface and 58.6 Å x 60.9 Å x 10.0 Å below consisting of 10,295 and 1,153 water molecules repectively; fourteen chlorine ions were added to the system to neutralise the overall charge of the system. The LINCS[248] algorithm was applied to constrain bonds with hydrogen atoms, the leapfrog[249] algorithm was used to integrate Newton's equations of motion with a 1 fs time step. Non-bonded interactions cut-off was set to 10 Å with the PME[93] being employed for electrostatics, the system temperature was controlled using the V-rescale[250-252] variant of the Berendsen thermostat;[253] periodic boundary conditions were employed throughout. Surface, bulk and interaction site particles were restrained by a harmonic potential of 5 x 10^6 kJ mol^{-1}, AUC particles were constrained at a length of 0.7 Å from AUS and AUB atoms again using the LINCS constraint algorithm.[248] The multiple peptide systems were run with the NVT ensemble using the REST2[241, 245] implementation described in Section 4.2.3 for a period of 20 ns giving a total of 620 ns over the 31 replica simulations.

4.2.5 Cluster Identification

Clusters of structures were generated from the 20,000 frames generated from the simulation at 300 K using the Daura[254] clustering algorithm contained within the Gromacs package. Cluster members are categorised based on a root mean squared deviation (RMSD) of 0.5 Å, the structure with the largest number of neighbours is set as the centre of the cluster. A total of 1859 clusters were identified from the analysis, 818 of which contained at least one adsorbed peptide, defined using a centre of mass cut-off of between 3.5 and 4.5 Å from the surface as in Section 3.2.3.

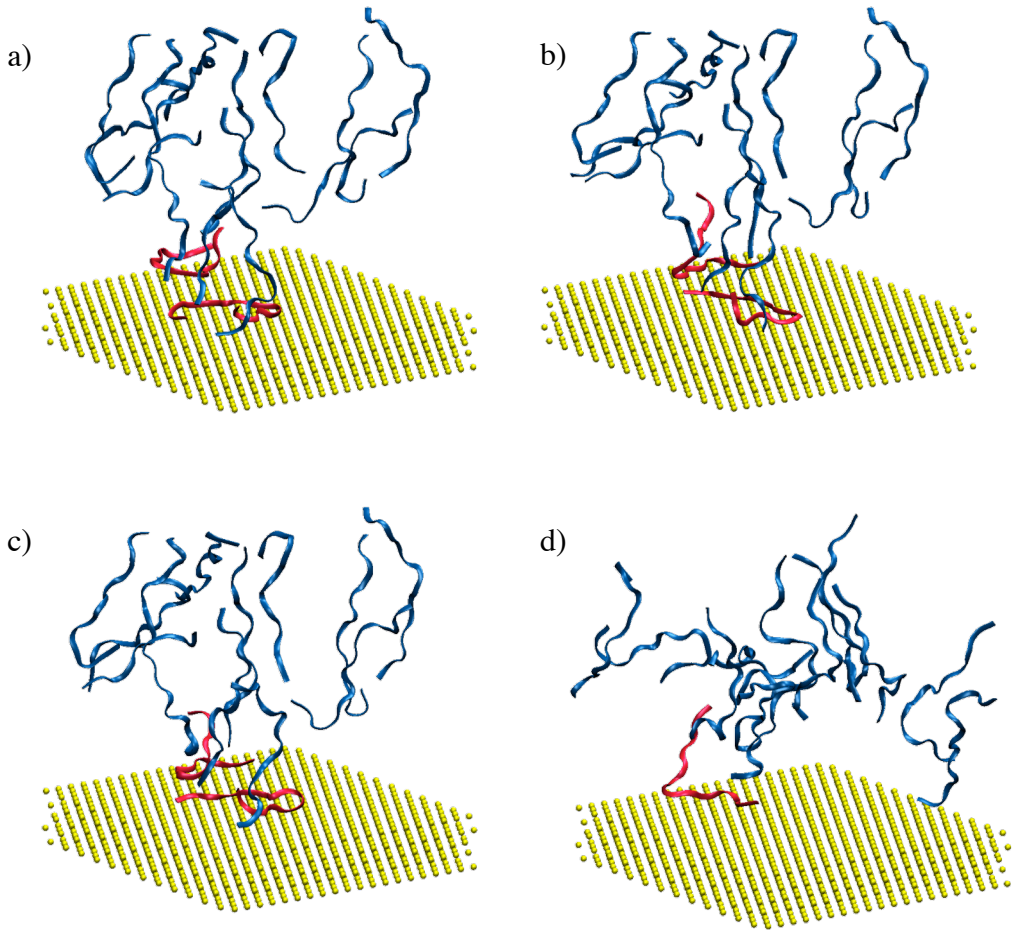


Figure 4.8. Conformations of the most populated clusters a) 360 members (3.63 % adsorbed conformations), b) 184 members (1.85 %), c) 164 members (1.65 %) and d) 162 members (1.63 %). Adsorbed peptides are shown in red and bulk peptides in blue, water omitted for clarity.

The population of clusters containing adsorbed peptides was normalised to 100 % and each cluster ranked by highest population, Figure 4.9.

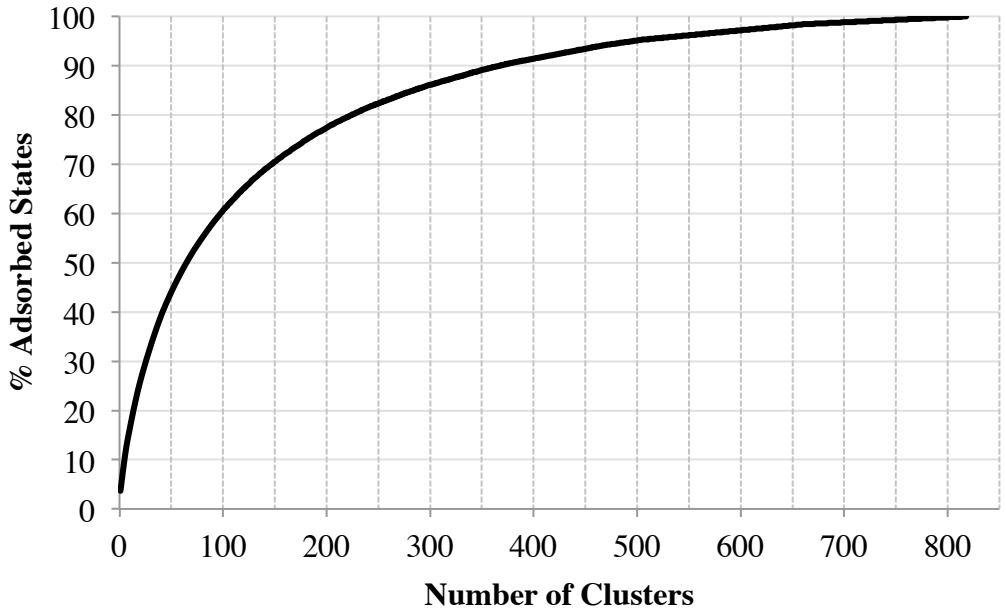


Figure 4.9. Percentage of conformations contained within clusters ranging from most to least populated.

From the data presented in Figure 4.9, it was found that a large proportion of adsorbed states can be studied using relatively few desorption simulations 10 % = 6 simulations, 25 % = 19 simulations, 50 % = 65 simulations. However, to gain an additional 10 % of adsorbed conformations a further 34 simulations would be required; hence an initial set of simulations representing 50 % of adsorbed conformations was studied using the NETI-SMD approach. This is a reasonable approach as this set of conformations contains the most relevant adsorbed states.

4.2.6 Peptide Desorption

From each of the 65 clusters that constitute half of the adsorbed states of the system, the peptide with a centre of mass closest to that of the gold surface was chosen for desorption. Each peptide was pulled from the surface at a rate of 0.001 Å ps^{-1} over a distance of 40 Å with a harmonic potential of 500 kcal mol^{-1} attached to the alpha carbon of residue 7 – alanine; one pulling simulation was performed for each configuration (65 x 2600 CPU hours = *ca.* 170,000 CPU hours), Figure 4.12 a). Compared to Chapter 3, a slower pulling rate (0.001 Å ps^{-1} vs. 0.005 Å ps^{-1}) was employed to avoid artefacts in the free energy calculation arising from steric clashes between the desorbing peptide and neighbouring peptides in the bulk.

4.3 Binding Free Energies in Crowded System

Statistical bootstrapping analysis in conjunction with Jarzynski's equality was used to determine the binding free energy and estimate errors. Interestingly, it was found that the binding free energies converged to less than 0.5 kcal mol^{-1} after utilising only 20 of the 65 simulations, Figure 4.10; this convergence is significantly quicker than for the single peptide systems discussed in Chapter 3.

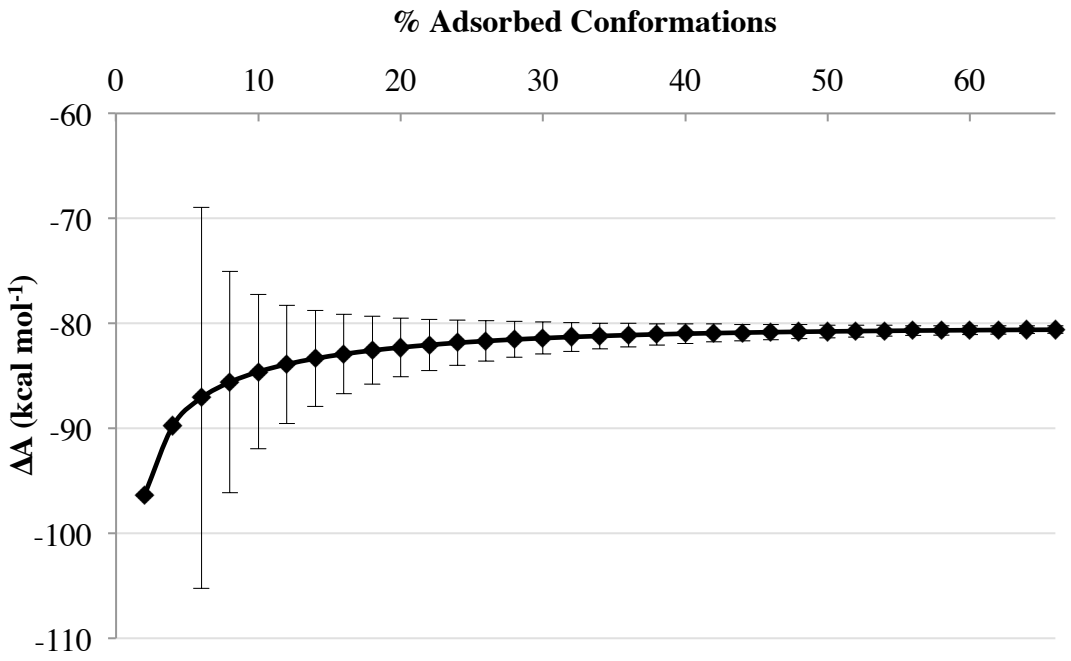


Figure 4.10. Evolution of Helmholtz free energy of binding as a function of number of simulations employed in statistical bootstrapping analysis. Errors for 2 (± 5769.5 kcal mol^{-1}) and 4 (± 53.5 kcal mol^{-1}) simulations omitted for clarity.

Analysis of the binding free energy taken from the top 65 cluster simulations reveals that, in addition to rapidly converging, the bind free energy from these crowded systems (ΔA = -80.6 kcal mol^{-1}), is significantly lower compared to that calculated for the single peptide system in Chapter 3 (ΔA = -243.0 kcal mol^{-1}). The weaker binding energy, (ΔΔA = +162.4 kcal mol^{-1}) may arise from the a number of factors including desorption rate, polarisable surface and interaction between the peptide being pulled and those in solution. To examine this, the non-bonded interaction energy (ΔE) between the peptide that is being pulled and those in solution was extracted, Figure 4.11.

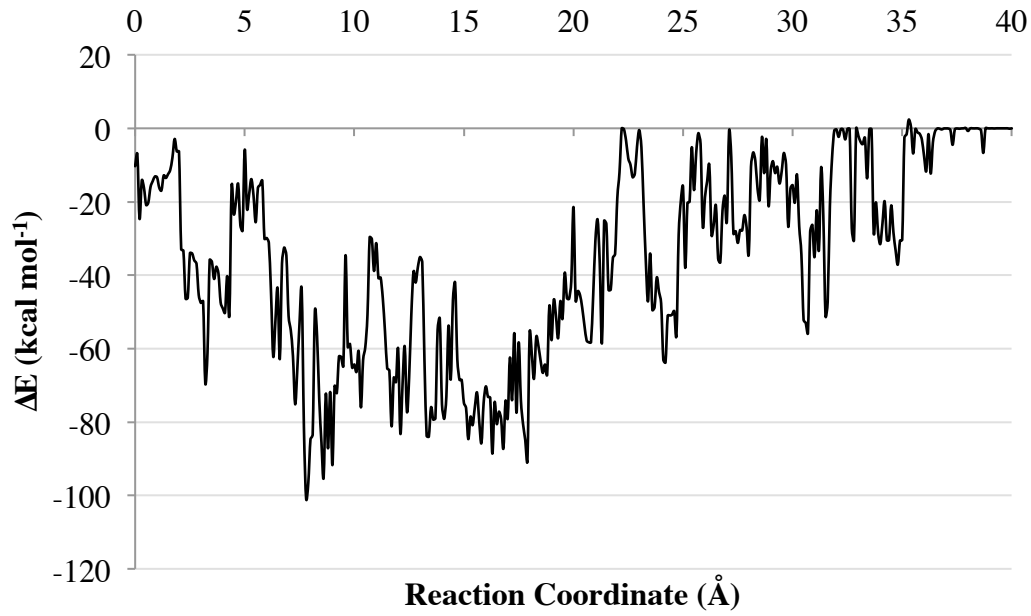


Figure 4.11. Interaction energy between desorbing peptide and bulk peptides as a function of pulling reaction coordinate.

The interaction energy (ΔE) is found to be as high as -100 kcal mol^{-1} in the first stages of the pulling simulation (*ca*. 2.5 Å to 17.5 Å); thus the bulk peptides clearly interact with the peptide that is being pulled from the surface, effectively weakening its binding free energy with the surface. In the latter stages of the desorption process however, the interaction between the desorbing peptide and those in the bulk tends towards 0 kcal mol^{-1}, which coincides with the increase in interaction energy between the thirteen bulk peptides and the gold surface, Figure 4.12 c).

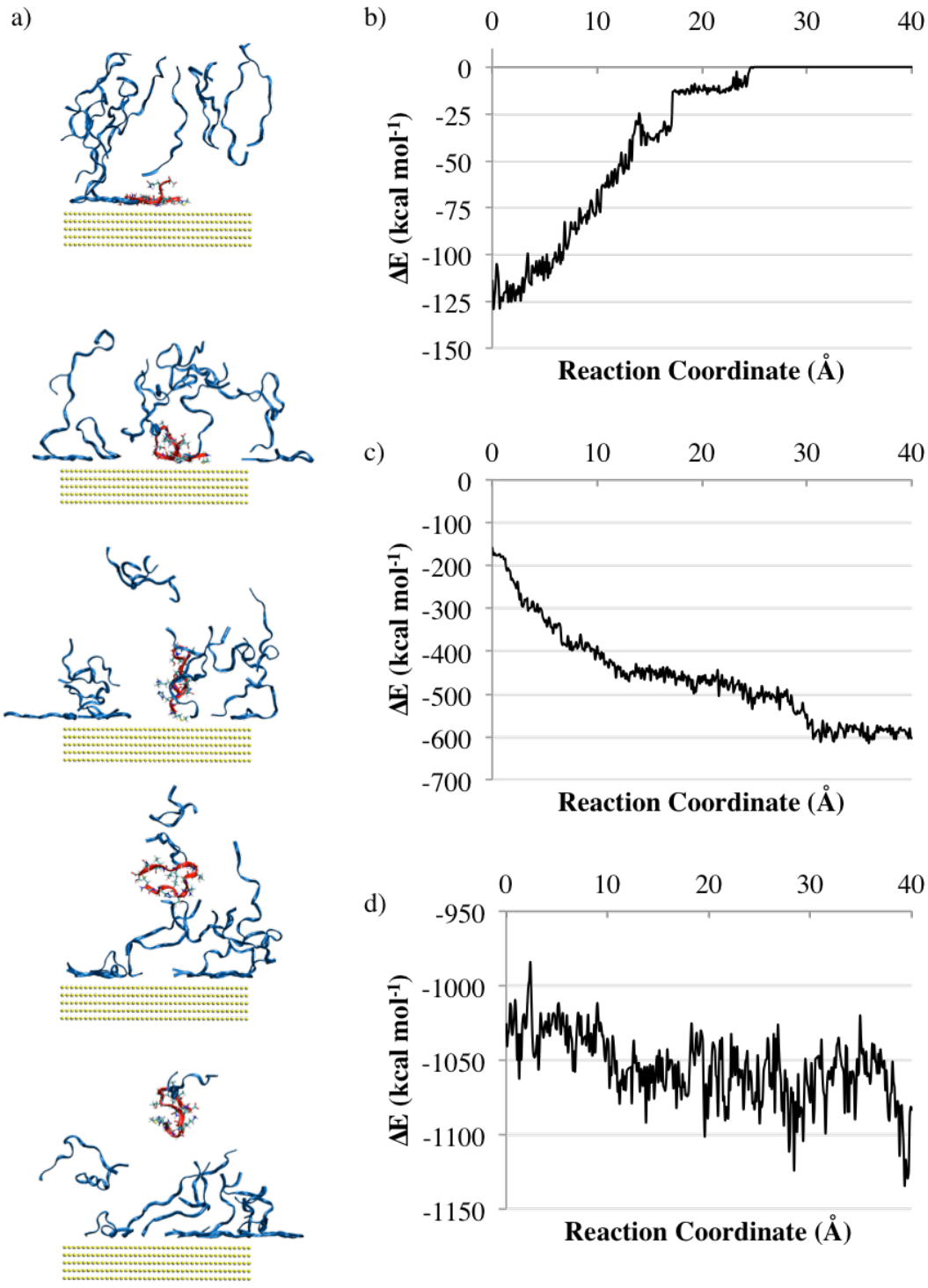


Figure 4.12. a) Snapshots at 10 ns intervals of the desorption process of the multiple peptide system (pulling peptide shown in red and bulk peptides in blue). b) Intermolecular interaction energy change (ΔE) between desorbing peptide and the gold surface as a function of reaction coordinate. c) Change in ΔE between the gold

surface and bulk peptides and d) change in intermolecular ΔE of bulk peptides as a function of pulling coordinate.

As the bulk peptides begin to form strong interactions with the surface, Figure 4.12 c) the intermolecular interactions between the peptides also become stronger as their proximity increases, Figure 4.12 d) and the pulling peptide no longer interacts with the surface, Figure 4.12 b). Therefore, the importance of the bulk peptides cannot be understated when considering both binding energy and system stability.

4.4 Conclusions

The inclusion of multiple peptides within the simulation corresponding to experimental surface concentrations has revealed interesting binding phenomena in the behaviour of peptides interacting with solid surfaces. It is observed that peptides in the bulk weaken the affinity of surface bound peptides, the adsorption of peptides to the gold surface is stabilised through strong peptide-peptide interactions as well as those between the peptide and the gold.

Throughout this work, progressive improvement of model quality has revealed several important phenomena in the binding of peptides to inorganic surfaces.

From the work carried out in Chapters 2 and 3, and through strategic mutation of GBP1, it was shown that the binding free energy of a peptide with a solid surface cannot simply be inferred from the summation of its single amino acid component

free energies; indeed the stability of the peptide both on the surface and in the bulk solvent is key to its binding affinity.

In Chapter 2, it was observed that the 'static' CHARMM-Metal forcefield was unable to fully capture all of the molecular mechanisms involved in the binding of charged adsorbates; hence in Chapter 4 the polarisable GolP-CHARMM force field was used for the gold surface to give a more complete representation of biomolecular binding, especially for peptides containing charged side chains.

In Chapter 4, employment of the Hamiltonian replica exchange technique, REST, allowed the most relevant conformations of the system to be identified. For example, from the six-orientation approach adopted in Chapter 3, four adsorbed conformations of GBP1 on the gold surface were obtained; the enhanced sampling approach on the other hand, gave rise to more than 800 unique adsorbed conformations weighted by population, representing the most physically relevant states of the system.

In the pursuit of understanding the binding behaviour of peptides at solid interfaces, the calculation of binding free energies using NETI-SMD in conjunction with statistical bootstrapping proved to be a consistently reliable approach. In the final chapter, a slower pulling speed was employed to avoid free energy artefacts from a system where steric clashes may conceal true binding phenomena. Additionally, this modification resulted in faster convergence of errors to below a stringent threshold of ± 0.5 kcal mol^{-1}. Indeed in Chapter 3, GBP1 required 134 simulations to meet

convergence criteria when pulled at a rate of 0.005 Å ps^{-1} compared to the 54 simulations when pulled at 0.001 Å ps^{-1}.

The significance of surface crowding has previously been speculated;[200] however, the magnitude and nature of this phenomenon has not been explicitly determined until now. In Chapter 4, the presence of experimental concentrations of peptide are shown to be extremely important in determining the binding free energy, adsorption behaviour and system stabilisation; hence for comparison with experimental systems and improved reliability of binding free energies, surface crowding should also be explicitly considered when examining surface binding.

5 References

1. Horwich, A. *The Journal of Clinical Investigation,* **2002,** *110*, 1221.

2. Dobson, C. M. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* **2001,** *356*, 133.

3. Thomas, P. J.; Qu, B.-H.; Pedersen, P. L. *Trends Biochem. Sci.,* **1995,** *20*, 456.

4. Knowles, T. P. J.; Vendruscolo, M.; Dobson, C. M. *Nat Rev Mol Cell Biol,* **2014,** *15*, 384.

5. Luheshi, L. M.; Crowther, D. C.; Dobson, C. M. *Curr. Opin. Chem. Biol.,* **2008,** *12*, 25.

6. Ellis, R. J. *Curr. Opin. Struct. Biol.,* **2001,** *11*, 114.

7. Hartl, F. U.; Hayer-Hartl, M. *Science,* **2002,** *295*, 1852.

8. Bukau, B.; Horwich, A. L. *Cell,* **1998,** *92*, 351.

9. Levinthal, C. *Extrait du Journal de Chimie Physique,* **1968,** *65*.

10. Levinthal, C. In *How to Fold Graciously*, Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois, Debrunnder, J. T. P.; Munck, E., Eds. University of Illinois Press: 1969; pp 22.

11. Zwanzig, R.; Szabo, A.; Bagchi, B. *Proc. Natl. Acad. Sci. U. S. A.,* **1992,** *89*, 20.

12. Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. *Nature,* **2003,** *421*, 863.

13. Yang, W. Y.; Gruebele, M. *Nature,* **2003,** *423*, 193.

14. Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature,* **2002,** *420*, 102.

15. Eaton, W. A.; Muñoz, V.; Thompson, P. A.; Henry, E. R.; Hofrichter, J. *Acc. Chem. Res.,* **1998,** *31*, 745.

16. Fersht, A. R. *Proc. Natl. Acad. Sci. U. S. A.,* **2000,** *97*, 1525.

17. Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. *Nature,* **2001,** *409*, 641.

18. Dobson, C. M.; Šali, A.; Karplus, M. *Angew. Chem. Int. Ed.,* **1998,** *37*, 868.

19. Plotkin, S. S.; Onuchic, J.; eacute; N *Q. Rev. Biophys.,* **2002,** *35*, 111.

20. Dill, K. A.; Chan, H. S. *Nat Struct Mol Biol,* **1997,** *4*, 10.

21. Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct., Funct., Bioinf.,* **1995,** *21*, 167.

22. Sali, A.; Shakhnovich, E.; Karplus, M. *Nature,* **1994,** *369*, 248.

23. Englander, S. W.; Mayne, L. *Proc. Natl. Acad. Sci. U. S. A.,* **2014,** *111*, 15873.

24. Dobson, C. M. *Nature,* **2003,** *426*, 884.

25. Fleming, S.; Ulijn, R. V. *Chem. Soc. Rev.,* **2014,** *43*, 8150.

26. Debnath, S.; Shome, A.; Das, D.; Das, P. K. *J. Phys. Chem. B,* **2010,** *114*, 4407.

27. Zou, Y.; Razmkhah, K.; Chmel, N. P.; Hamley, I. W.; Rodger, A. *RSC Advances,* **2013,** *3*, 10854.

28. Zhang, H.; Wang, H.; Xu, G.; Yuan, S. *Colloids and Surfaces A: Physicochemical and Engineering Aspects,* **2013,** *417*, 217.

29. Fleming, S.; Debnath, S.; Frederix, P. W. J. M.; Tuttle, T.; Ulijn, R. V. *Chem. Commun. (Cambridge, U. K.),* **2013,** *49*, 10587.

30. Tang, C.; Smith, A. M.; Collins, R. F.; Ulijn, R. V.; Saiani, A. *Langmuir,* **2009,** *25*, 9447.

31. Ryan, D. M.; Doran, T. M.; Anderson, S. B.; Nilsson, B. L. *Langmuir,* **2011,** *27*, 4029.

32. Gao, Y.; Kuang, Y.; Guo, Z.-F.; Guo, Z.; Krauss, I. J.; Xu, B. *J. Am. Chem. Soc.,* **2009,** *131*, 13576.

33. Liang, G.; Yang, Z.; Zhang, R.; Li, L.; Fan, Y.; Kuang, Y.; Gao, Y.; Wang, T.; Lu, W. W.; Xu, B. *Langmuir,* **2009,** *25*, 8419.

34. Naskar, J.; Palui, G.; Banerjee, A. *J. Phys. Chem. B,* **2009,** *113*, 11787.

35. Reches, M.; Gazit, E. *Science,* **2003,** *300*, 625.

36. Lamm, M. S.; Sharma, N.; Rajagopal, K.; Beyer, F. L.; Schneider, J. P.; Pochan, D. J. *Adv. Mater. (Weinheim, Ger.),* **2008,** *20*, 447.

37. Sharma, N.; Top, A.; Kiick, K. L.; Pochan, D. J. *Angew. Chem. Int. Ed.,* **2009,** *48*, 7078.

38. Hughes, M.; Frederix, P. W. J. M.; Raeburn, J.; Birchall, L. S.; Sadownik, J.; Coomer, F. C.; Lin, I. H.; Cussen, E. J.; Hunt, N. T.; Tuttle, T.; Webb, S. J.; Adams, D. J.; Ulijn, R. V. *Soft Matter,* **2012,** *8*, 5595.

39. Pierce, K. L.; Premont, R. T.; Lefkowitz, R. J. *Nat. Rev. Mol. Cell Biol.,* **2002,** *3*, 639.

40. Rosenbaum, D. M.; Rasmussen, S. G. F.; Kobilka, B. K. *Nature,* **2009,** *459*, 356.

41. Azzi, M.; Charest, P. G.; Angers, S.; Rousseau, G.; Kohout, T.; Bouvier, M.; Piñeyro, G. *Proc. Natl. Acad. Sci. U. S. A.,* **2003,** *100*, 11406.

42. Wisler, J. W.; DeWire, S. M.; Whalen, E. J.; Violin, J. D.; Drake, M. T.; Ahn, S.; Shenoy, S. K.; Lefkowitz, R. J. *Proc. Natl. Acad. Sci. U. S. A.,* **2007,** *104*, 16657.

43. Cohen, G. B.; Oprian, D. D.; Robinson, P. R. *Biochemistry,* **1992,** *31*, 12592.

44. Jaakola, V.-P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. *Science,* **2008,** *322*, 1211.

45. Shenoy, S. K.; Drake, M. T.; Nelson, C. D.; Houtz, D. A.; Xiao, K.; Madabushi, S.; Reiter, E.; Premont, R. T.; Lichtarge, O.; Lefkowitz, R. J. *J. Biol. Chem.,* **2006,** *281*, 1261.

46. Shi, L.; Liapakis, G.; Xu, R.; Guarnieri, F.; Ballesteros, J. A.; Javitch, J. A. *J. Biol. Chem.,* **2002,** *277*, 40989.

47. Baker, J. G. *Br. J. Pharmacol.,* **2010,** *160*, 1048.

48. Galandrin, S.; Oligny-LongprÈ, G.; Bouvier, M. *Trends Pharmacol. Sci.,* **2007,** *28*, 423.

49. Sugimoto, Y.; Fujisawa, R.; Tanimura, R.; Lattion, A. L.; Cotecchia, S.; Tsujimoto, G.; Nagao, T.; Kurose, H. *J. Pharmacol. Exp. Ther.,* **2002,** *301*, 51.

50. Radzicka, A.; Wolfenden, R. *Science,* **1995,** *267*, 90.

51. Kahne, D.; Still, W. C. *J. Am. Chem. Soc.,* **1988,** *110*, 7529.

52. Lawrence, L.; Moore, W. J. *J. Am. Chem. Soc.,* **1951,** *73*, 3973.

53. Hegg, E. L.; Burstyn, J. N. *J. Am. Chem. Soc.,* **1995,** *117*, 7015.

54. Zhu, L.; Qin, L.; Parac, T. N.; Kostic, N. M. *J. Am. Chem. Soc.,* **1994,** *116*, 5218.

55. Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M. *Structure,* **1997,** *5*, 1093.

56. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. *Protein Sci.,* **1996,** *5*, 1001.

57. Matthews, B. W.; Sigler, P. B.; Henderson, R.; Blow, D. M. *Nature,* **1967,** *214*, 652.

58. Huber, R.; Kukla, D.; Bode, W.; Schwager, P.; Bartels, K.; Deisenhofer, J.; Steigemann, W. *J. Mol. Biol.,* **1974,** *89*, 73.

59. Stroud, R. M.; Kay, L. M.; Dickerson, R. E. *Cold Spring Harbor Symp. Quant. Biol.,* **1972,** *36*, 125.

60. Kraut, J. *Annu. Rev. Biochem.,* **1977,** *46*, 331.

61. Kraut, J.; Robertus, J. D.; Birktoft, J. J.; Alden, R. A.; Wilcox, P. E.; Powers, J. C. *Cold Spring Harbor Symp. Quant. Biol.,* **1972,** *36*, 117.

62. Paetzel, M.; Dalbey, R. E. *Trends Biochem. Sci.,* **1997,** *22*, 28.

63. Paetzel, M.; J. Strynadka, N. C. *Protein Sci.,* **1999,** *8*, 2533.

64. Moult, J.; Sussman, F.; James, M. N. G. *J. Mol. Biol.,* **1985,** *182*, 555.

65. Berg, J. M.; Tymoczko, J. L.; Stryer, L., *Biochemistry: international edition*. WH Freeman & Company Limited: 2006.

66. Dauter, Z.; Betzel, C.; Genov, N.; Pipon, N.; Wilson, K. S. *Acta Crystallographica Section B,* **1991,** *47*, 707.

67. Rypniewski, W. R.; Dambmann, C.; von der Osten, C.; Dauter, M.; Wilson, K. S. *Acta Crystallographica Section D,* **1995,** *51*, 73.

68. Warshel, A.; Florián, J. *Proc. Natl. Acad. Sci. U. S. A.,* **1998,** *95*, 5950.

69. Nanda, V.; Koder, R. L. *Nature Chem.,* **2010,** *2*, 15.

70. Golynskiy, M. V.; Seelig, B. *Trends Biotechnol.,* **2010,** *28*, 340.

71. Whaley, S. R.; English, D. S.; Hu, E. L.; Barbara, P. F.; Belcher, A. M. *Nature,* **2000,** *405*, 665.

72. Naik, R. R.; Stringer, S. J.; Agarwal, G.; Jones, S. E.; Stone, M. O. *Nat. Mater.,* **2002,** *1*, 169.

73. Ellington, A. D.; Szostak, J. W. *Nature,* **1990,** *346*, 818.

74. Tuerk, C.; Gold, L. *Science,* **1990,** *249*, 505.

75. Arnold, F. H. *Acc. Chem. Res.,* **1998,** *31*, 125.

76. Kehoe, J. W.; Kay, B. K. *Chem. Rev. (Washington, DC, U. S.),* **2005,** *105*, 4056.

77. Forrer, P.; Jung, S.; Plückthun, A. *Curr. Opin. Struct. Biol.,* **1999,** *9*, 514.

78. Pollack, S. J.; Jacobs, J. W.; Schultz, P. G. *Science,* **1986,** *234*, 1570.

79. Tramontano, A.; Janda, K. D.; Lerner, R. A. *Science,* **1986,** *234*, 1566.

80. Chen, J.; Wan, R.; Liu, H.; Cheng, C.-m.; Zhao, Y.-f. *Lett. Pept. Sci.,* **2000,** *7*, 325.

81. Li, Y.; Zhao, Y.; Hatfield, S.; Wan, R.; Zhu, Q.; Li, X.; McMills, M.; Ma, Y.; Li, J.; Brown, K. L.; He, C.; Liu, F.; Chen, X. *Bioorganic & Medicinal Chemistry,* **2000,** *8*, 2675.

82. Ma, Y.; Chen, X.; Sun, M.; Wan, R.; Zhu, C.; Li, Y.; Zhao, Y. *Amino Acids,* **2008,** *35*, 251.

83. Gorlero, M.; Wieczorek, R.; Adamala, K.; Giorgi, A.; Schininà, M. E.; Stano, P.; Luisi, P. L. *FEBS Lett.,* **2009,** *583*, 153.

84. Krattiger, P.; McCarthy, C.; Pfaltz, A.; Wennemers, H. *Angew. Chem. Int. Ed.,* **2003,** *42*, 1722.

85. Winkler, S.; Wilson, D.; Kaplan, D. L. *Biochemistry,* **2000,** *39*, 12739.

86. Yang, Z.; Gu, H.; Fu, D.; Gao, P.; Lam, J. K.; Xu, B. *Adv. Mater. (Weinheim, Ger.),* **2004,** *16*, 1440.

87. Li, S.; Hong, M. *J. Am. Chem. Soc.,* **2011,** *133*, 1534.

88. Price, D. J.; Brooks Iii, C. L. *J. Chem. Phys.,* **2004,** *121*, 10096.

89. Grubmuller, H. *SOLVATE*, v 1.2; Theoretical Biophysics Group, Institute for Medical Optics, Ludwig-Maximilian University: Munich, 1996.

90. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.,* **1996,** *14*, 33.

91. Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.,* **1999,** *151*, 283.

92. MacKerell, A. D.; Banavali, N.; Foloppe, N. *Biopolymers,* **2000,** *56*, 257.

93. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.,* **1993,** *98*, 10089.

94. Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.,* **1992,** *13*, 952.

95. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.,* **1977,** *23*, 327.

96. Belcher, A. M.; Wu, X. H.; Christensen, R. J.; Hansma, P. K.; Stucky, G. D.; Morse, D. E. *Nature,* **1996,** *381*, 56.

97. Thompson, J. B.; Paloczi, G. T.; Kindt, J. H.; Michenfelder, M.; Smith, B. L.; Stucky, G.; Morse, D. E.; Hansma, P. K. *Biophys. J.,* **2000,** *79*, 3307.

98. Addadi, L.; Weiner, S. *Angewandte Chemie International Edition in English,* **1992,** *31*, 153.

99. Iuliano, D. J.; Saavedra, S. S.; Truskey, G. A. *Journal of biomedical materials research,* **1993,** *27*, 1103.

100. Kasemo, B. *Current Opinion in Solid State and Materials Science,* **1998,** *3*, 451.

101. Rabe, M.; Verdes, D.; Seeger, S. *Adv. Colloid Interface Sci.,* **2011,** *162*, 87.

102. Tan, Y. N.; Lee, J. Y.; Wang, D. I. C. *J. Am. Chem. Soc.,* **2010,** *132*, 5677.

103. Murphy, C. J.; Sau, T. K.; Gole, A. M.; Orendorff, C. J.; Gao, J.; Gou, L.; Hunyadi, S. E.; Li, T. *J. Phys. Chem. B,* **2005,** *109*, 13857.

104. Daniel, M.-C.; Astruc, D. *Chem. Rev. (Washington, DC, U. S.),* **2004,** *104*, 293.

105. Wang, Z.; Chen, J.; Yang, P.; Yang, W. *Appl. Organomet. Chem.,* **2007,** *21*, 645.

106. Sarikaya, M.; Tamerler, C.; Jen, A. K. Y.; Schulten, K.; Baneyx, F. *Nat Mater,* **2003,** *2*, 577.

107. Tamerler, C.; Duman, M.; Oren, E. E.; Gungormus, M.; Xiong, X.; Kacar, T.; Parviz, B. A.; Sarikaya, M. *Small,* **2006,** *2*, 1372.

108. Matmor, M.; Ashkenasy, N. *J. Mater. Chem.,* **2011,** *21*, 968.

109. Haruta, M. *CATTECH,* **2002,** *6*, 102.

110. Lopez, N.; Janssens, T. V. W.; Clausen, B. S.; Xu, Y.; Mavrikakis, M.; Bligaard, T.; Nørskov, J. K. *J. Catal.,* **2004,** *223*, 232.

111. Fujitani, T.; Nakamura, I.; Akita, T.; Okumura, M.; Haruta, M. *Angew. Chem. Int. Ed.,* **2009,** *48*, 9515.

112. Zanchet, A.; Dorta-Urra, A.; Roncero, O.; Flores, F.; Tablero, C.; Paniagua, M.; Aguado, A. *Phys. Chem. Chem. Phys.,* **2009,** *11*, 10122.

113. Barrio, L.; Liu, P.; Rodríguez, J. A.; Campos-Martín, J. M.; Fierro, J. L. G. *J. Chem. Phys.,* **2006,** *125*, 164715.

114. Haruta, M.; Daté, M. *Applied Catalysis A: General,* **2001,** *222*, 427.

115. Saha, K.; Agasti, S. S.; Kim, C.; Li, X.; Rotello, V. M. *Chem. Rev. (Washington, DC, U. S.),* **2012,** *112*, 2739.

116. Qian, X.; Peng, X.-H.; Ansari, D. O.; Yin-Goen, Q.; Chen, G. Z.; Shin, D. M.; Yang, L.; Young, A. N.; Wang, M. D.; Nie, S. *Nat Biotech,* **2008,** *26*, 83.

117. Maiti, K. K.; Dinish, U. S.; Fu, C. Y.; Lee, J.-J.; Soh, K.-S.; Yun, S.-W.; Bhuvaneswari, R.; Olivo, M.; Chang, Y.-T. *Biosensors and Bioelectronics,* **2010,** *26*, 398.

118. Zhou, W.; Gao, X.; Liu, D.; Chen, X. *Chem. Rev. (Washington, DC, U. S.),* **2015**.

119. Cognet, L.; Tardin, C.; Boyer, D.; Choquet, D.; Tamarat, P.; Lounis, B. *Proc. Natl. Acad. Sci. U. S. A.,* **2003,** *100*, 11350.

120. Loo, C.; Lowery, A.; Halas, N.; West, J.; Drezek, R. *Nano Lett.,* **2005,** *5*, 709.

121. Liu, J.; Lu, Y. *Chem. Mater.,* **2004,** *16*, 3231.

122. Lee, J.-S.; Han, M. S.; Mirkin, C. A. *Angew. Chem. Int. Ed.,* **2007,** *46*, 4093.

123. Slocik, J. M.; Zabinski, J. S.; Phillips, D. M.; Naik, R. R. *Small,* **2008,** *4*, 548.

124. Kumar, A.; Ma, H.; Zhang, X.; Huang, K.; Jin, S.; Liu, J.; Wei, T.; Cao, W.; Zou, G.; Liang, X.-J. *Biomaterials,* **2012,** *33*, 1180.

125. Heo, D. N.; Yang, D. H.; Moon, H.-J.; Lee, J. B.; Bae, M. S.; Lee, S. C.; Lee, W. J.; Sun, I.-C.; Kwon, I. K. *Biomaterials,* **2012,** *33*, 856.

126. Kodiha, M.; Wang, Y. M.; Hutter, E.; Maysinger, D.; Stochaj, U. *Theranostics,* **2015,** *5*, 357.

127. Feldherr, C. M.; Akin, D. *J. Cell Sci.,* **1997,** *110*, 3065.

128. Feldherr, C. M.; Akin, D.; Cohen, R. J. *J. Cell Sci.,* **2001,** *114*, 4621.

129. Nakanishi, K.; Sakiyama, T.; Imamura, K. *J. Biosci. Bioeng.,* **2001,** *91*, 233.

130. Marks, L. D. *Reports on Progress in Physics,* **1994,** *57*, 603.

131. Brown, S.; Sarikaya, M.; Johnson, E. *J. Mol. Biol.,* **2000,** *299*, 725.

132. Brown, S. *Nat Biotech,* **1997,** *15*, 269.

133. Sarikaya, M.; Tamerler, C.; Schwartz, D. T.; Baneyx, F. *Annu. Rev. Mater. Res.,* **2004,** *34*, 373.

134. Wetterer, S. M.; Lavrich, D. J.; Cummings, T.; Bernasek, S. L.; Scoles, G. *J. Phys. Chem. B,* **1998,** *102*, 9266.

135. Syomin, D.; Koel, B. E. *Surf. Sci.,* **2002,** *498*, 61.

136. Syomin, D.; Koel, B. E. *Surf. Sci.,* **2002,** *498*, 53.

137. Salmeron, M.; Kaufman, D. S.; Marchon, B.; Ferrer, S. *Appl. Surf. Sci.,* **1987,** *28*, 279.

138. Dovek, M. M.; Lang, C. A.; Nogami, J.; Quate, C. F. *Phys. Rev. B.,* **1989,** *40*, 11973.

139. Barth, J. V.; Brune, H.; Ertl, G.; Behm, R. J. *Phys. Rev. B.,* **1990,** *42*, 9307.

140. Hallmark, V. M.; Chiang, S.; Rabolt, J. F.; Swalen, J. D.; Wilson, R. J. *Phys. Rev. Lett.,* **1987,** *59*, 2879.

141. Jaklevic, R. C.; Elie, L. *Phys. Rev. Lett.,* **1988,** *60*, 120.

142. Hasegawa, Y.; Avouris, P. *Science,* **1992,** *258*, 1763.

143. Nie, H. Y.; Mizutani, W.; Tokumoto, H. *Surf. Sci.,* **1994,** *311*, L649.

144. Haiss, W.; Lackey, D.; Sass, J. K.; Besocke, K. H. *J. Chem. Phys.,* **1991,** *95*, 2193.

145. Tao, N. J.; Lindsay, S. M. *J. Appl. Phys.,* **1991,** *70*, 5141.

146. Hanke, F.; Björk, J. *Phys. Rev. B.,* **2013,** *87*, 235422.

147. Heinz, H.; Farmer, B. L.; Pandey, R. B.; Slocik, J. M.; Patnaik, S. S.; Pachter, R.; Naik, R. R. *J. Am. Chem. Soc.,* **2009,** *131*, 9704.

148. Feng, J.; Slocik, J. M.; Sarikaya, M.; Naik, R. R.; Farmer, B. L.; Heinz, H. *Small,* **2012,** *8*, 1049.

149. Domke, K. F.; Zhang, D.; Pettinger, B. *J. Am. Chem. Soc.,* **2006,** *128*, 14721.

150. Dahanayaka, D. H.; Wang, J. X.; Hossain, S.; Bumm, L. A. *J. Am. Chem. Soc.,* **2006,** *128*, 6052.

151. Hohenberg, P.; Kohn, W. *Phys. Rev.,* **1964,** *136*, B864.

152. Kohn, W.; Sham, L. J. *Phys. Rev.,* **1965,** *140*, A1133.

153. Hoefling, M.; Iori, F.; Corni, S.; Gottschalk, K.-E. *ChemPhysChem,* **2010,** *11*, 1763.

154. Iori, F.; Di Felice, R.; Molinari, E.; Corni, S. *J. Comput. Chem.,* **2009,** *30*, 1465.

155. Pakiari, A. H.; Jamshidi, Z. *J. Phys. Chem. A,* **2007,** *111*, 4391.

156. Xie, H.-J.; Lei, Q.-F.; Fang, W.-J. *J Mol Model,* **2012,** *18*, 645.

157. Di Felice, R.; Selloni, A.; Molinari, E. *J. Phys. Chem. B,* **2003,** *107*, 1151.

158. Jing, C.; Fang, Y. *Chem. Phys.,* **2007,** *332*, 27.

159. Crespilho, F. N.; Lima, F. C. A.; da Silva, A. B. F.; Oliveira Jr, O. N.; Zucolotto, V. *Chem. Phys. Lett.,* **2009,** *469*, 186.

160. Iori, F.; Corni, S.; Di Felice, R. *J. Phys. Chem. C,* **2008,** *112*, 13540.

161. Rai, S.; Singh, H. *J Mol Model,* **2013,** *19*, 4099.

162. Rai, S.; Suresh Kumar, N. V.; Singh, H. *Bull. Mater. Sci.,* **2012,** *35*, 291.

163. Hong, G.; Heinz, H.; Naik, R. R.; Farmer, B. L.; Pachter, R. *ACS Appl. Mater. Interfaces,* **2009,** *1*, 388.

164. Hoefling, M.; Iori, F.; Corni, S.; Gottschalk, K.-E. *Langmuir,* **2010,** *26*, 8347.

165. Feng, J.; Pandey, R. B.; Berry, R. J.; Farmer, B. L.; Naik, R. R.; Heinz, H. *Soft Matter,* **2011,** *7*, 2113.

166. Nawrocki, G.; Cieplak, M. *J. Phys. Chem. C,* **2014,** *118*, 12929.

167. Hughes, Z. E.; Wright, L. B.; Walsh, T. R. *Langmuir,* **2013,** *29*, 13217.

168. Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.,* **1977,** *23*, 187.

169. Souaille, M.; Roux, B. t. *Comput. Phys. Commun.,* **2001,** *135*, 40.

170. Darve, E.; Pohorille, A. *J. Chem. Phys.,* **2001,** *115*, 9169.

171. Comer, J.; Gumbart, J. C.; Hénin, J.; Lelièvre, T.; Pohorille, A.; Chipot, C. *J. Phys. Chem. B,* **2015,** *119*, 1129.

172. Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. *J. Chem. Phys.,* **2008,** *128*, 144120.

173. Wei, C.; Pohorille, A. *J. Am. Chem. Soc.,* **2009,** *131*, 10237.

174. Dehez, F.; Tarek, M.; Chipot, C. *J. Phys. Chem. B,* **2007,** *111*, 10633.

175. Chipot, C.; Hénin, J. *J. Chem. Phys.,* **2005,** *123*, 244906.

176. Bhatnagar, N.; Kamath, G.; Chelst, I.; Potoff, J. J. *J. Chem. Phys.,* **2012,** *137*, 014502.

177. Jarzynski, C. *Phys. Rev. Lett.,* **1997,** *78*, 2690.

178. Jarzynski, C. *Physical Review E,* **1997,** *56*, 5018.

179. Israelewitz, B.; Gao, M.; Schulten, K. *Curr. Opin. Struct. Biol.,* **2001,** *11*, 224.

180. Efron, B.; Tibshirani, R. *Statistical science,* **1986**, 54.

181. Mijajlovic, M.; Penna, M. J.; Biggs, M. J. *Langmuir,* **2013,** *29*, 2919.

182. Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. *J. Chem. Phys.,* **2003,** *119*, 3559.

183. Cuendet, M. A.; Michielin, O. *Biophys. J.,* **2008,** *95*, 3575.

184. Baştuğ, T.; Chen, P.-C.; Patra, S. M.; Kuyucak, S. *J. Chem. Phys.,* **2008,** *128*, 155104.

185. Chu, X.; Jiang, W.; Zhang, Z.; Yan, Y.; Pan, H.; Xu, X.; Tang, R. *J. Phys. Chem. B,* **2011,** *115*, 1151.

186. Pan, H.; Tao, J.; Xu, X.; Tang, R. *Langmuir,* **2007,** *23*, 8972.

187. Heinz, H.; Vaia, R. A.; Farmer, B. L.; Naik, R. R. *J. Phys. Chem. C,* **2008,** *112*, 17281.

188. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B,* **1998,** *102*, 3586.

189. Lide, D. R., *CRC handbook of chemistry and physics*. CRC press: 2004.

190. Tyson, W. R.; Miller, W. A. *Surf. Sci.,* **1977,** *62*, 267.

191. Dennington, R.; Keith, T.; Millam, J. *GaussView, Version 5*, Semichem Inc.: Shawnee Mission, KS, 2009.

192. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.,* **2008,** *4*, 435.

193. Smith, G. P. *Science,* **1985,** *228*, 1315.

194. Daugherty, P. S. *Curr. Opin. Struct. Biol.,* **2007,** *17*, 474.

195. Georgiou, G.; Stathopoulos, C.; Daugherty, P. S.; Nayak, A. R.; Iverson, B. L.; Iii, R. C. *Nat Biotech,* **1997,** *15*, 29.

196. Zin, M. T.; Ma, H.; Sarikaya, M.; Jen, A. K. Y. *Small,* **2005,** *1*, 698.

197. Braun, R.; Sarikaya, M.; Schulten, K. *J. Biomater. Sci., Polym. Ed.,* **2002,** *13*, 747.

198. Wei, Y.; Latour, R. A. *Langmuir,* **2008,** *24*, 6721.

199. Verde, A. V.; Acres, J. M.; Maranas, J. K. *Biomacromolecules,* **2009,** *10*, 2118.

200. Tang, Z.; Palafox-Hernandez, J. P.; Law, W.-C.; E. Hughes, Z.; Swihart, M. T.; Prasad, P. N.; Knecht, M. R.; Walsh, T. R. *ACS Nano,* **2013,** *7*, 9632.

201. Palafox-Hernandez, J. P.; Tang, Z.; Hughes, Z. E.; Li, Y.; Swihart, M. T.; Prasad, P. N.; Walsh, T. R.; Knecht, M. R. *Chem. Mater.,* **2014,** *26*, 4960.

202. Raut, V. P.; Agashe, M. A.; Stuart, S. J.; Latour, R. A. *Langmuir,* **2005,** *21*, 1629.

203. Latour, R. A. *Biointerfaces,* **2008,** *3*, FC2.

204. Bellucci, L.; Corni, S. *J. Phys. Chem. C,* **2014,** *118*, 11357.

205. Kuang, Z.; Kim, S. N.; Crookes-Goodson, W. J.; Farmer, B. L.; Naik, R. R. *ACS Nano,* **2010,** *4*, 452.

206. Kim, S. N.; Kuang, Z.; Slocik, J. M.; Jones, S. E.; Cui, Y.; Farmer, B. L.; McAlpine, M. C.; Naik, R. R. *J. Am. Chem. Soc.,* **2011,** *133*, 14480.

207. Katoch, J.; Kim, S. N.; Kuang, Z.; Farmer, B. L.; Naik, R. R.; Tatulian, S. A.; Ishigami, M. *Nano Lett.,* **2012,** *12*, 2342.

208. Tomásio, S. M.; Walsh, T. R. *J. Phys. Chem. C,* **2009,** *113*, 8778.

209. Patwardhan, S. V.; Emami, F. S.; Berry, R. J.; Jones, S. E.; Naik, R. R.; Deschaume, O.; Heinz, H.; Perry, C. C. *J. Am. Chem. Soc.,* **2012,** *134*, 6244.

210. Emami, F. S.; Puddu, V.; Berry, R. J.; Varshney, V.; Patwardhan, S. V.; Perry, C. C.; Heinz, H. *Chem. Mater.,* **2014,** *26*, 5725.

211. Ramakrishnan, S. K.; Martin, M.; Cloitre, T.; Firlej, L.; Gergely, C. *J. Chem. Inf. Model.,* **2014,** *54*, 2117.

212. Heinz, H. *J. Comput. Chem.,* **2010,** *31*, 1564.

213. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.,* **1995,** *103*, 4613.

214. Kulp Iii, J. L.; Sarikaya, M.; Spencer Evans, J. *J. Mater. Chem.,* **2004,** *14*, 2325.

215. Pandey, R. B.; Heinz, H.; Feng, J.; Farmer, B. L.; Slocik, J. M.; Drummy, L. F.; Naik, R. R. *Phys. Chem. Chem. Phys.,* **2009,** *11*, 1989.

216. Hautman, J.; Klein, M. L. *J. Chem. Phys.,* **1989,** *91*, 4994.

217. Ghorai, P. K.; Glotzer, S. C. *J. Phys. Chem. C,* **2007,** *111*, 15857.

218. Djebaili, T.; Richardi, J.; Abel, S.; Marchi, M. *J. Phys. Chem. C,* **2015,** *119*, 21146.

219. Piserchia, A.; Zerbetto, M.; Salvia, M.-V.; Salassa, G.; Gabrielli, L.; Mancin, F.; Rastrelli, F.; Frezzato, D. *J. Phys. Chem. C,* **2015,** *119*, 20100.

220. Kim, H.; Saha, J. K.; Zhang, Z.; Jang, J.; Matin, M. A.; Jang, J. *J. Phys. Chem. C,* **2014,** *118*, 11149.

221. Djebaili, T.; Richardi, J.; Abel, S.; Marchi, M. *J. Phys. Chem. C,* **2013,** *117*, 17791.

222. Saha, J. K.; Ahn, Y.; Kim, H.; Schatz, G. C.; Jang, J. *J. Phys. Chem. C,* **2011,** *115*, 13193.

223. Ahn, Y.; Saha, J. K.; Schatz, G. C.; Jang, J. *J. Phys. Chem. C,* **2011,** *115*, 10668.

224. Jiménez, A.; Sarsa, A.; Blázquez, M.; Pineda, T. *J. Phys. Chem. C,* **2010,** *114*, 21309.

225. Yang, A.-C.; Weng, C.-I. *J. Phys. Chem. C,* **2010,** *114*, 8697.

226. Meena Devi, J. *Progress in Natural Science: Materials International,* **2014,** *24*, 405.

227. An, D.; Su, J.; Weber, J. K.; Gao, X.; Zhou, R.; Li, J. *J. Am. Chem. Soc.,* **2015,** *137*, 8412.

228. Heimel, G.; Brédas, J.-L. *Nat. Nanotechnol.,* **2013,** *8*, 230.

229. Barabino, G.; Marchesi, M. *Chem. Phys. Lett.,* **1984,** *104*, 478.

230. Shelley, J. C.; Patey, G. N.; Bérard, D. R.; Torrie, G. M. *J. Chem. Phys.,* **1997,** *107*, 2122.

231. Spohr, E. *Acta Chem. Scand.,* **1995,** *49*, 189.

232. Qian, J.; Hentschke, R.; Knoll, W. *Langmuir,* **1997,** *13*, 7092.

233. Iori, F.; Corni, S. *J. Comput. Chem.,* **2008,** *29*, 1656.

234. Wright, L. B.; Rodger, P. M.; Corni, S.; Walsh, T. R. *J. Chem. Theory Comput.,* **2013,** *9*, 1616.

235. Wright, L. B.; Rodger, P. M.; Walsh, T. R.; Corni, S. *J. Phys. Chem. C,* **2013,** *117*, 24292.

236. Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.,* **1986,** *57*, 2607.

237. Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.,* **1999,** *314*, 141.

238. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.,* **1953,** *21*, 1087.

239. Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U. S. A.,* **2005,** *102*, 13749.

240. Huang, X.; Hagen, M.; Kim, B.; Friesner, R. A.; Zhou, R.; Berne, B. J. *J. Phys. Chem. B,* **2007,** *111*, 5405.

241. Wang, L.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem. B,* **2011,** *115*, 9431.

242. Shao, Q.; Shi, J.; Zhu, W. *J. Chem. Phys.,* **2012,** *137*, 125103.

243. Ostermeir, K.; Zacharias, M. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics,* **2013,** *1834*, 847.

244. Wright, L. B.; Walsh, T. R. *Phys. Chem. Chem. Phys.,* **2013,** *15*, 4715.

245. Terakawa, T.; Kameda, T.; Takada, S. *J. Comput. Chem.,* **2011,** *32*, 1228.

246. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics,* **2013**.

247. Abraham, M. J.; van der Spoel, D.; Lindahl, E.; Hess, B.; Team, G. D. *GROMACS,* **2014**.

248. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.,* **1997,** *18*, 1463.

249. Quinn, T.; Katz, N.; Stadel, J.; Lake, G. *arXiv preprint astro-ph/9710043,* **1997**.

250. Bussi, G.; Parrinello, M. *Comput. Phys. Commun.,* **2008,** *179*, 26.

251. Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.,* **2007,** *126*, 014101.

252. Raiteri, P.; Gale, J. D.; Bussi, G. *J. Phys.: Condens. Matter,* **2011,** *23*, 334213.

253. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.,* **1984,** *81*, 3684.

254. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem. Int. Ed.,* **1999,** *38*, 236.