University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

A Thesis Submitted for the Degree of PhD at the University of Warwick

http://go.warwick.ac.uk/wrap/61709

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

THE UNIVERSITY OF WARWICK

AUTHOR: **Louise B Wright** DEGREE: **Ph.D.**

TITLE: **Modelling the Adsorption of Peptides at aqueous Quartz and Gold Interfaces**

DATE OF DEPOSIT: ....

I agree that this thesis shall be available in accordance with the regulations governing the University of Warwick theses.

I agree that the summary of this thesis may be submitted for publication.

I **agree** that the thesis may be photocopied (single copies for study purposes only).

Theses with no restriction on photocopying will also be made available to the British Library for microfilming. The British Library may supply copies to individuals or libraries. subject to a statement from them that the copy is supplied for non-publishing purposes. All copies supplied by the British Library will carry the following statement:

> "Attention is drawn to the fact that the copyright of this thesis rests with its author. This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's written consent."

AUTHOR'S SIGNATURE: ....

USER'S DECLARATION

1. I undertake not to quote or make use of any information from this thesis without making acknowledgement to the author.
2. I further undertake to allow no-one else to use this thesis while it is in my care.

DATE SIGNATURE ADDRESS

....

....

....

....

....

Modelling the Adsorption of Peptides at aqueous Quartz and Gold Interfaces

by

Louise B Wright

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Chemistry

March 2014

THE UNIVERSITY OF
WARWICK

Contents

List of Tables vi

List of Figures xiv

Acknowledgments xxiii

Declarations xxiv

Abstract xxv

Abbreviations xxvii

Chapter 1 Introduction 1

1.1 The Potential of material and facet selective Peptides 1
1.2 Selection of Material Binding Peptides 5
1.2.1 Biocombinatorial Peptide Selection 5
1.2.2 Towards the Rational Design of Peptide Sequences 6
1.2.3 Biointerfacial Experimental Techniques 7
1.2.4 Biointerfacial Simulation 10
1.3 Aqueous Quartz Interface 13
1.3.1 Experiment 13
1.3.2 Simulation 17
1.3.3 Crystallographic Surfaces 19
1.4 Aqueous Gold Interface 21
1.4.1 Experiment 21
1.4.2 Simulation 23
1.4.3 Crystallographic Surfaces 25

Chapter 2 Methods 28

2.1 Describing the Potential Energy Landscape 28

2.1.1 First Principles Methods 28
2.1.2 Atomistic Force-Fields 33
2.2 Molecular Dynamics 36
2.2.1 Time Evolution 37
2.2.2 Thermostats 37
2.2.3 Barostats 38
2.2.4 First-Principles MD 39
2.3 Replica Exchange Methods 40
2.3.1 T-REMD 41
2.3.2 H-REMD 42
2.3.3 Replica Exchange with Solute Tempering 43
2.4 Free Energy Methods 44
2.4.1 Potential of Mean Force 45
2.4.2 Metadynamics 46
2.5 Analysis 48
2.5.1 Hydrogen Bonding 48
2.5.2 Peptide Structure Clustering 49
2.6 Peptide Secondary Structure 51

Chapter 3 Free Energy Calculations of Amino-Acid Analogue Adsorption at Aqueous α-Quartz Interfaces. 53
3.1 Introduction 53
3.2 Methods 54
3.2.1 Analysis 55
3.3 Results 55
3.3.1 Methane 56
3.3.2 Ammonium 57
3.3.3 Butyl ammonium 59
3.3.4 Benzene 61
3.3.5 Ethanoate 63
3.4 Discussion 66
3.5 Conclusions 67

Chapter 4 First-Principles Molecular Dynamics Simulations of Ammonium and Ethanoate Adsorption at the Aqueous (100) α-Quartz Interface. 68
4.1 Introduction 68
4.2 Methods 69

4.2.1 Car-Parrinello Molecular Dynamics Simulations 70
4.2.2 Force-field MD Simulations 71
4.2.3 Analysis 72
4.3 Results 73
4.3.1 Ethanoate 73
4.3.2 Ammonium 74
4.4 Discussion 77
4.5 Conclusions 80

Chapter 5 Benchmarking Replica Exchange with Solute Tempering for Bio-interfacial Systems: QBP-1 Adsorption onto the (100) α-Quartz Surface. 81
5.1 Introduction 81
5.2 Methods 82
5.2.1 System Details 82
5.2.2 Temperature-REMD 83
5.2.3 Replica Exchange with Solute Tempering 84
5.2.4 Analysis 86
5.3 Results 87
5.3.1 Replica Mobility 87
5.3.2 Clustering Analysis 88
5.3.3 Peptide Structures and Ramachandran Analysis 91
5.3.4 Hydrogen Bonding 96
5.3.5 Binding Residues and Motifs 96
5.3.6 Energetics 98
5.4 Discussion 99
5.5 Conclusions 101

Chapter 6 GolP-CHARMM: A First-Principles Based Force-field for modelling protein adsorption to Au(111), Au(100)(1×1) and Au(100)(5×1).103
6.1 Introduction 103
6.2 Methods 104
6.2.1 *in Vacuo* Force-Field Calculations 113
6.2.2 Force-field MD Simulations of Aqueous Gold Interfaces . . . 115
6.2.3 Analysis 115
6.3 Results 116
6.3.1 Evaluation of vdW-DFT 116
6.3.2 Force-field Parameterisation 118

6.3.3 Aqueous Au(111) and Au(100) Interfaces 129
6.4 Conclusions 133

Chapter 7 Material Selectivity of QBP-1 Adsorption. 135
7.1 Disclaimer 135
7.2 Introduction 135
7.3 Methods 136
7.3.1 System Details 136
7.3.2 Replica Exchange with Solute Tempering 137
7.3.3 Analysis 138
7.4 Results 139
7.4.1 Peptide Conformation 139
7.4.2 Surface Binding 144
7.5 Discussion 146
7.6 Conclusions 149

Chapter 8 Facet Selectivity of AuBP-1 adsorption onto Gold: A Replica Exchange with Solute Tempering Metadynamics Study. 150
8.1 Disclaimer 150
8.2 Introduction 150
8.3 Methods 151
8.3.1 System Details 151
8.3.2 Replica Exchange with Solute Tempering 153
8.3.3 Analysis 153
8.4 Schemes for Reweighting Structural Data from a Metadynamics Simulation 156
8.4.1 Scheme 1: Total bias 'V' 157
8.4.2 Scheme 2: Relative Bias 'ΔV' 157
8.4.3 Scheme 3: Relative Bias 'ΔV equilib' 157
8.4.4 Scheme 4: 'Time period' 158
8.4.5 Scheme 5: Tiana method 160
8.4.6 Scheme 6: Bonomi Method 161
8.4.7 Evaluation 162
8.5 Results 166
8.5.1 Binding Strength 166
8.5.2 Cluster Analysis 170
8.5.3 Secondary Structure 171
8.5.4 Binding Mechanism 171

8.6 Conclusions 174

Chapter 9 Conclusions and Future Prospects 176

Appendix A Additional Information: Free Energy Calculations of Amino-Acid Analogue Adsorption at Aqueous α-Quartz Interfaces. 181

Appendix B Additional Information: First-Principles Molecular Dynamics Simulations of Ammonium and Ethanoate Adsorption at the Aqueous (100) α-Quartz Interface. 188

Appendix C Additional Information:Benchmarking Replica Exchange with Solute Tempering for Bio-interfacial Systems: QBP-1 Adsorption onto the (100) α-Quartz Surface. 194

Appendix D Additional Information:GolP-CHARMM: A First-Principles Based Force-field for modelling protein adsorption to Au(111), Au(100)(1×1) and Au(100)(5×1). 203

Appendix E Additional Information:Material Selectivity of QBP-1 Adsorption. 225

Appendix F Additional Information:Facet Selectivity of AuBP-1 adsorption onto Gold: A Replica Exchange with Solute Tempering Metadynamics Study. 236

List of Tables

1.1 Silica-binding peptide sequences along with either the substrate against- or method by which they were selected ('PD' denotes phage display and 'NP' denotes nanoparticle). Positively charged residues have been highlighted in red, negatively charged ones in blue and polar in green. Aromatic residues have been underlined. [1] Reference Kroger et al. [1999], [2] Reference Naik et al. [2002a], [3] Reference Eteshola et al. [2005], [4] Reference Sano et al. [2005], [5] Reference Chen et al. [2006], [6] Reference Patwardhan et al. [2012], [7] Reference Puddu and Perry [2012], [8] Reference Tamerler et al. [2007], [9] Reference Oren et al. [2007]. 15

1.2 Gold-binding peptide sequences along with either the substrate against- or method by which they were selected ('PD' denotes phage display). Positively charged residues have been highlighted in red, negatively charged ones in blue and polar in green. Aromatic residues have been underlined. [1] Reference Brown [1997], [2] Reference Naik et al. [2002b], [3] Reference Peelle et al. [2005], [4] Reference Huang et al. [2005], [5] Reference Nam et al. [2006], [6] Reference Hnilova et al. [2008], [7] Reference Kim et al. [2010]. 22

3.1 Optimal PMF free energy of adsorption, and vertical position of the PMF minimum for amino-acid analogues on (100), (001) and (011) α-Quartz surfaces, under aqueous conditions. Errors were obtained via block averaging as outlined in Section 2.4.1. 56

3.2 Hydrogen bond correlation life-time (ps) formed by the adsorbates ammonium and ethanoate at the PMF minimum distance on (100), (001) and (011) α-Quartz. 58

5.1 Total number and number of ‘significant’ clusters identified for each REMD-based simulation. (Cluster analysis carried out over the last 4×10^6 MD steps only.) The % of total structure population belonging to each of the three most populated clusters identified over the same time period is also recorded. 89

6.1 Optimal PW-DFT calculated interaction energies of small molecules adsorbed onto the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces *in vacuo* ($\mathrm{kJ\,mol^{-1}}$) and equilibrium separations in direction normal to the surface (Å). ‘n/a’ refers to a calculation which was not performed. 109

7.1 Total number of, and the number of ‘significant’, backbone clusters calculated over the final 5 ns. Percentage population of top 3 clusters identified by clustering analysis conducted over the same portion of the trajectory, is also given. 139

8.1 Percentage of frames in which residues R10 and R11 make direct contact with the Au(100)(1×1) surface under aqueous conditions, calculated using the different reweighting schemes outlined above. ‘raw’ corresponds to bias data extracted from the metaD simulation without reweighting;[a] reference Bonomi et al. [2009b]. No data is available for Scheme 5 (Tiana [Tiana [2008]]) due to the numerical instability of the algorithm. 162

8.2 Free energy of adsorption of AuBP-1 at the aqueous Au(111), Au(100)(1×1) and Au(100)(5×1) interfaces estimated using both the ΔG_{f1} and ΔG_{f5} binding profiles (Section 8.4.4 and Figure 8.2). 169

A.1 Average number of hydrogen bonds formed by the adsorbates, methanol, ammonium and ethanoate, at the PMF minimum distance above the (100), (001) and (011) α-quartz surfaces. Hydrogen bonds to methanol were defined as the simultaneous satisfaction of the following criteria: O-O$_{MeOH}$ distance of less than 3.46 Å; either H-O$_{MeOH}$ distance of less than 2.50 Å or O-H$_{MeOH}$ distance of less than 2.32 Å; and H-O-O angle of less than 30°; hydrogen bonds to ammonium as the simultaneous satisfaction of the following criteria: N-O distance of less than 3.72 Å; H-N distance of less than 2.24 Å; and H-N-O angle of less than 30°; and hydrogen bonds to ethanoate as the simultaneous satisfaction of the following criteria: O-O distance of less than 3.28 Å; H-O distance of less than 2.46 Å; and H-O-O angle of less than 30°. 183

B.1 Cut-off distances for hydrogen bond definitions, taken from radial distribution functions for each approach. 190
B.2 Trajectory average number (per timeframe) and average length (Å) of hydrogen bonds formed by CH_3COO^- to water or the quartz surface from both CPMD and FMD simulations. Hydrogen bonds were defined as per the criteria detailed in Section 2.5.1 and defined in Table B.1. 192
B.3 Trajectory average number (per timeframe) and average length of hydrogen bonds formed by NH_4^+ to water or the quartz surface from both CPMD and FMD simulations. Hydrogen bonds were defined as per the criteria detailed in Section 2.5.1 and defined in Table B.1. . . 193

C.1 Average numbers of hydrogen bonds per MD frame. 198
C.2 Average RMSD of backbone atom positions of all structures in the three most populated clusters identified by each REMD-based method from the centriod structure of the four most populated solution clusters of QBP-1 [Notman et al. [2010]]. 'Identical' clusters are highlighted in red, whilst 'similar' ones are highlighted in green (see Section 2.5.2). 199

C.3 Residues underlined are those which are on average bound to the quartz surface. Results are shown for the three most populated backbone clusters identified by clustering analysis carried out over the last 4x10^6 MD steps of each simulation. (Index: 'n'= normal T-REMD, 'l_w'= light T-REMD, 'l_0.6'= light sc 0.6 T-REMD, 'l_0.8'= light sc 0.8 T-REMD, 're_1'= REST1, 're nh'= REST no h, 're 2 off'= REST 2 off, 're_l'= REST light, 're_l nh'= REST light no h,'re_l 2 off'= REST light 2 off.) 200

C.4 Top 4 binding dyad motifs in the REMD-based simulations, calculated over the last 2x10^6MD steps of each trajectory. Percentage of trajectory for which the dyad makes binding contact is given in parenthesises. Dyads present in S1 which were found to be enriched dyads in quartz binding peptides [Oren et al. [2010]] are highlighted as follows: PW, WL, PP, LP, (given in order of reducing Motif Enrichment Factor). 201

C.5 Top 4 binding triad motifs in the REMD-based simulations, calculated over the last 2x10^6MD steps of each trajectory. Percentage of trajectory for which the dyad makes binding contact is given in parenthesises. Dyads present in S1 which were found to be enriched dyads in quartz binding peptides [Oren et al. [2010]] are highlighted as follows: PPW, PWL and WLP(given in order of reducing Motif Enrichment Factor). 202

D.1 Buckling amplitudes b_1 and b_2 relative to bulk d_{bulk}. Values given in parentheses from Havu *et al.* [Havu et al. [2010]] are those obtained using the PBE functional, those not in parentheses were obtained using LDA only. Data to calculate b_2 from Jacob [Jacob [2007]] is not available. ^{a} reference Havu et al. [2010]; ^{b} reference Jacob [2007]; ^{c} reference Ocko et al. [1991]; † are values calculated in this work using a k-point mesh of 2×2×1 and * are values calculated in this work using a k-point mesh of 4×4×1. 209

D.2 Average interlayer distances d_{12} and d_{23} relative to bulk d_{bulk}. Values given in parentheses from Havu *et al.* [Havu et al. [2010]] are those obtained using the PBE functional, those not in parentheses were obtained using LDA only. [a] reference Havu et al. [2010]; [b] reference Shi and Stampfl [2008]; [c] reference Jacob [2007]; [d] reference Ocko et al. [1991]; † are values calculated in this work using a k-point mesh of 2×2×1 and * are values calculated in this work using a k-point mesh of 4×4×1. 210

D.3 Interaction energies and equilibrium separations of all the small molecules used to fit and test FF parameters for the Au(100)(5×1) surface found from PW-DFT calculations using the vdW-DF functional. 'Tx' and 'Hx' refer to the initial location of the molecule on the surface; 'T23' denotes a geometry in which the two heteroatoms in methanoic acid and methyl amide are adsorbed atop different sites (see Section 6.2). 212

D.4 GolP-CHARMM LJ parameters for the Au(111) surface. Interaction strengths ϵ (kJ mol^{-1}) and distances σ (Å) are given for both surface virtual sites (AUI) and bulk gold atoms (AUB). Note the Au–H interaction is for hydroxyl and amide H only. 213

D.5 GolP-CHARMM LJ parameters for the Au(100)(1×1) surface. Interaction strengths ϵ (kJ mol^{-1}) and distances σ (Å) are given for both surface virtual sites (AUI) and bulk gold atoms (AUB). Note the Au–H interaction is for hydroxyl, amide and thiol H only. 214

D.6 GolP-CHARMM LJ parameters for the Au(100)(5×1) surface. Interaction strengths ϵ (kJ mol^{-1}) and distances σ (Å) are given for both hollow (AUIU) and bridge (S0, S1, S2, S3) surface interaction sites. 'NR2' corresponds to an unprotonated N in imidazole, 'NHT2' to N in an unprotonated N-terminus or lysine side-chain, 'S' to S atoms in dialkyl sulfide moieties, 'ST' to S atoms in thiols. Note the Au–H interaction is for hydroxyl, thiol and amide H only. All cross terms involving bridge interaction sites that have not been explicitly given above are zero. 215

D.7 Interaction energies and equilibrium separations of set of molecules used to fit GolP-CHARMM parameters for the Au(111) surface. Energies in kJ mol^{-1} and distances in Å. ^{a} PW-DFT energy calculated from this work; ^{b} Experimental interaction energy taken from reference [Wetterer et al. [1998]]; ^{c} Experimental interaction energy taken from reference [Syomin et al. [2001]]; ^{d} Experimental interaction energy taken from reference [Lavrich et al. [1998]]; ^{e} Experimental interaction energy taken from reference [Nuzzo et al. [1987]]. 217

D.8 Comparison between experimental interaction energies (kJ mol^{-1}) and those obtained using GolP-CHARMM parameters for the Au(111) surface. ^{a} Experimental interaction energy taken from reference [Syomin and Koel [2002a]] ; ^{b} Experimental interaction energy taken from reference [Wetterer et al. [1998]]; ^{c} Experimental interaction energy taken from reference [Syomin and Koel [2002b]]. RMSD between GolP-CHARMM and experiment was 4.15 kJ mol^{-1} using the above 'test set' whereas the RMSD between GolP and experiment was 4.97 kJ mol^{-1} [Iori et al. [2009]] using a 'test set' of molecules that comprised of cyclohexane, cyclohexene, cycloctane, toluene, non-1-ene, acetone, trans-but-2-ene, diethyl sulfide and undec-1-ene. 218

D.9 Interaction energies for test set of molecules adsorbed onto the Au(111) surface. Energies in kJ mol^{-1}. 219

D.10 Interaction energies and equilibrium separations of the set of molecules used to fit GolP-CHARMM parameters for the Au(100)(1×1) surface. Energies in kJ mol^{-1} and distances in Å. 220

D.11 Interaction energies for the test set of molecules adsorbed onto the Au(100)(1×1) surface. Energies in kJ mol^{-1}. 221

D.12 Optimal interaction energies (kJ mol^{-1}) and equilibrium closest surface separations (Å) for all small molecules investigated adsorbed onto Au(100)(5×1), calculated using vdW-DF and optimal FF parameters given above. 222

D.13 Interaction energies (kJ mol^{-1}) and equilibrium closest surface separations (Å) for molecules containing an unprotonated N atom adsorbed onto Au(100)(5×1) *in vacuo*, calculated using vdW-DF and optimal FF parameters given above. Labels 'Tx' refer to the top site location 0-3 on the Au(100)(5×1) surface. 223

D.14 Interaction energies (kJ mol^{-1}) and equilibrium closest surface separations (Å) for sulfur containing molecules adsorbed onto Au(100)(5×1) *in vacuo*, calculated using vdW-DF and optimal FF parameters given above. Labels 'Tx' refer to the top site location 0-3 on the Au(100)(5×1) surface. 224

E.1 Atoms or groups of atoms used to calculate residue–surface separations and the cut-off distances defining the adsorbed state of the residue at each interface. (In the case of more than one atom being used to define a functional group, the binding distance cut-off refers to the surface-separation of the centre of mass of the atoms involved.) 229

E.2 Percentage of frames in the last 5 ns of the trajectory in which each residue is adsorbed to the quartz/gold surface. The four residues with the greatest degree of interfacial contact have been highlighted in red. 231

E.3 Percentage of frames in the last 5 ns of the trajectory in which each diad is adsorbed to the quartz/gold surface. Diads which are statistically enriched in quartz binding peptides, compared to the Ph.D.-12 peptide library as a whole are PW, WL, PP, LP [Oren et al. [2010]]. Motifs found to have significant interfacial contact during the simulations carried out in this work have been highlighted according to their sequence using the same colour scheme or underlined. 232

E.4 Percentage of frames in the last 5 ns of the trajectory in which each triad is adsorbed to the quartz/gold surface. Triads which are statistically enriched in quartz binding peptides, compared to the Ph.D.-12 peptide library as a whole are PPW, PWL and WLP [Oren et al. [2010]]. Motifs found to have significant interfacial contact during the simulations carried out in this work have been highlighted according to their sequence using the same colour scheme or underlined. 233

E.5 Percentage of frames in the last 5 ns of the trajectory in which each the backbone of each residue is adsorbed to the quartz/gold surface. Surface-separation cut-off distances of 5.1 Å and 4.4 Å were used to define the adsorbed state of a residues' backbone atoms (defined as the centre of mass of the amide group) at the (100) α-Quartz and Au(111) interfaces respectively. 234

F.1 Amino-acid functional group gold distances used as cut-off to define direct gold adsorption at the aqueous Au(111), Au(100)(1×1) and Au(100)(5×1) interfaces. In the case of Arg and Trp, the *c.o.m.* of the heavy atoms in the functional group described above was used in determining residue side-chain surface separation, whereas for Glu, Gly, Leu and Val it was the *closest* carboxylate oxygen atom, hydrogen or methyl carbon, respectively. 238
F.2 Percentage population of each reference cluster of AuBP-1. The most populated clusters have been highlighted as follows: C1, C2, C3. . . 241
F.3 Percentage likelihood for each potential binding diad to be in direct contact with the gold surface. The binding characteristics of a dyad have been highlighted following the scheme: strong 76-100%; significant 51-75%; and moderate 26-50%. 241
F.4 Percentage likelihood for each potential binding triad to be in direct contact with the gold surface. The binding characteristics of a triad have been highlighted following the scheme: strong 76-100%; significant 51-75%; and moderate 26-50%. 241

List of Figures

1.1 Scheme to depict how material- and facet-selective peptide sequences can be used to direct the fabrication of a) a regular array of NPs above a uniform surface, b) an array of NPs of mixed composition on a patterned surface and c) a matrix of NP orientationally aligned. . 3

1.2 Plan and side-on views of the (100), (001) and (011) α-Quartz surfaces. Silicon atoms shown in yellow, oxygen atoms in red and hydrogen atoms in white. 20

1.3 Plan and side-on views of the a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1) surfaces. 26

2.1 Schematic diagram to depict the movement of a second partial charge (blue) out of a sphere of radius r_c surrounding the first charge (red) from time t to time $t + \Delta t$. 35

2.2 Boundaries in ϕ/ψ space demarking the principal regions in a Ramachandran plot for analysis of QBP-1 structure. 51

2.3 Boundaries in ϕ/ψ space demarking the principal regions in a Ramachandran plot (1 β, 2 PPII, 3 α, 4 α_L and 5 γ_L) for analysis of AuBP-1 structure. 52

3.1 a) PMF free energy of adsorption butyl ammonium on the (100), (001) and (011)) α-quartz surfaces as a function of surface separation. Typical configurations of butyl ammonium adsorbed onto the b) (100) (‘N-up’), c) (001) (‘flat’) and d) (011) (‘flat’) surfaces. Inset: PMF free energy of adsorption of ammonium on the (100), (001) and (011) α-quartz surfaces as a function of vertical surface separation. Data in red, purple and yellow correspond with (100), (001) and (011) respectively. 60

3.2 a) PMF free energy of adsorption benzene on (100), (001) and (011) α-Quartz surfaces as a function of surface separation. Typical configurations of benzene adsorbed onto the b) (100) (‘upright’), c) (001) (‘upright’) and d) (011) (‘upright’) surfaces. 62
3.3 a) PMF free energy of adsorption ethanoate on (100), (001) and (011) α-Quartz surfaces as a function of surface separation. b)-d) Typical configurations of ethanoate on (100) at b) a separation of 3.0 Å (‘carboxylate-up’) and c) PMF minimum (‘methyl-up’), d) on (001) (‘flat’), and e) on (011) (‘flat’). 64

4.1 Vertical separation between the quartz surface and centre of mass of a) ethanoate, and b) ammonium, taken from the CPMD and FMD simulations. 74
4.2 ‘Hydrogen-bond like’ distances between the quartz surface and a) ethanoate ((CO)$O^{-}\cdots$H(OSi) distance), and b) ammonium (NH$\cdots$O(H)Si distance), taken from the CPMD and LFF simulations. In both plots the black-dashed line denotes the RDF-determined cut-off used in the criteria to define each type of hydrogen-bond. 75
4.3 Continuous hydrogen-bond time-autocorrelation functions for: a) ethanoate–quartz (eq) and ethanoate–water (ew) hydrogen-bonds, and, b) ammonium–quartz (aq) and ammonium–water (aw) hydrogen-bonds, taken from the CPMD and LFF simulations. 76
4.4 Snapshots of typical configurations of a) ethanoate and b) ammonium at the aqueous (100) α-quartz interface (taken from ethanoate CPMD-run1 and ammonium CPMD-run2 respectively). a) Shows ethanoate bound in a bidentate orientation whilst in b) ammonium is bound monodentate to the surface. Silicon atoms are shown as yellow, oxygen as red, hydrogen as white, nitrogen as blue and carbon as green. Water is omitted for clarity. 79

5.1 Number of clusters of QBP-1 backbone structure found from a) T-REMD and b) REST simulations as a function of MD steps. 90

5.2 Commonly observed structures of QBP-1 in the most populated clusters after REMD simulations of the peptide adsorbed onto the (100) α-quartz surface: a) extended, b) extended with kinks (akin to solution clusters 2 and 4 Reference [Notman et al. [2010]]), c) 'C' shaped (akin to solution cluster 1), and d) helix (found as a top-3 populated cluster using the 'REST 2 off', 'REST light' and 'REST light 2 off' methods). Water has been omitted for clarity. 93

5.3 Ramachandran plots for QBP-1 adsorbed on the quartz surface for each RE-based variant. 94

5.4 Potential Energy difference, $\Delta\mathrm{PE}_{sys-ext}$, Equation 5.1. 98

6.1 Scheme depicting the general steps taken to derive GolP-CHARMM FF parameters for the Au(111) and Au(100) surfaces. 105

6.2 Arrangement of virtual sites on a) the Au(111), b) the Au(100)(1×1) and c) Au(100)(5×1) surfaces. Real gold surface atoms are shown in gold and hollow virtual interaction sites in blue. The same colour scheme has been used for bridge virtual interaction sites on the Au(100)(5×1) surface as that in Figure 6.4. 108

6.3 Scheme depicting the steps taken to identify optimal adsorbate-gold geometries on the Au(100)(5×1) surface. See Figure 6.4 for description of adsorption site labels. 110

6.4 The Au(100)(5×1) surface after relaxation using vdW-DF [Dion et al. [2004]; Thonhauser et al. [2007]] viewed a) from above and b) from the side. For the plan view (left) surface layer Au atoms are depicted as largest and in medium gold, 2nd layer Au atoms dark gold and medium size, while the 3rd layer of Au atoms are the smallest size and light gold in colour. 111

6.5 Correlation between experimental interaction energies for small molecules adsorbed onto the Au(111) surface and those calculated with vdW-DF [Dion et al. [2004]; Thonhauser et al. [2007]; Roman-Perez and Soler [2009]]. 117

6.6 The most favourable orientation of a single water molecule adsorbed onto Au(111) a) using the FF without extra Au–O LJ parameters, b) using vdW-DF and c) using the new FF with Au–O LJ parameters. 122

6.7 Comparison between the interaction energies of small molecules with Au(111) *in vacuo* calculated using GolP-CHARMM FF and either experimental data (green) or those calculated using the vdW-DF exchange functional (black). Both molecules used to derive FF parameters (crosses) and those in the test set (open circles) are shown. The solid blue line denotes 1:1 correlation. 123
6.8 Comparison between the interaction energies of small molecules with Au(100) *in vacuo* calculated using GolP-CHARMM to those calculated using the vdW-DF exchange functional. Both molecules used to derive FF parameters (crosses) and those in the test set (open circles) are shown. The solid blue line denotes 1:1 correlation. 125
6.9 A comparison of the structure of water above the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces using three metrics: a) normalised mass density profiles, b) total number and number of donor/acceptor hydrogen bonds as a function of surface separation and c) water dipole moment. 130
6.10 Spatial arrangement of Au atoms in the a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1) surfaces, together with the density of water in the plane of the interface, for the first adsorbed layer only. 133

7.1 Snapshots of the centroid structure belonging to the three most populated clusters of QBP-1 in solution and adsorbed at either the aqueous α-Quartz or Au(111) interfaces. Water has been omitted for clarity. 140
7.2 Potential energy difference between centroid structure of cluster CX and that of the cluster representing QBP-1 helical conformations. . . 142
7.3 Left: Ramachadran plots of QBP-1 a) in solution and adsorbed at the aqueous, b) α-Quartz, or c) Au(111) interfaces. Right: Pie charts depicting the frequency with which each region of ϕ/ψ space was sampled. 143
7.4 Schematic depicting the degree to which each residue contributes to surface binding at the aqueous a) α-Quartz and b) Au(111) interfaces. The key corresponds to the percentage of frames in which the residue is bound to the surface (Table E.2). 145

7.5 The orientation of aromatic residues with respect to a) the α-Quartz and b) the Au(111) surface. θ was defined as the angle between the normal to the aromatic plane of each residue and the z direction. Data from the tryptophan residues W4 (dark) and W11 (light) were plotted in green, whereas that from Y7 was plotted in purple. Right: Snapshots depicting a trp residue adsorbed onto the α-Quartz surface in an upright/flat conformation (top/middle) and onto the Au(111) surface in a flat conformation (bottom). 147

8.1 Reference structures a-i used to cluster conformations of AuBP-1 at the aqueous Au(111), Au(100)(1×1) and Au(100)(5×1) interfaces, during the REST metaD simulations. 155
8.2 Free Energy Profiles G_{f1} and G_{f5} of AuBP-1 adsorbed at the aqueous a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1) interfaces. 165
8.3 The free energy of adsorption profile for AuBP-1 binding at the aqueous a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1) interfaces. Profiles shown after 60, 70, 80, 90 and 100 ns of metaD simulation. . 168
8.4 Secondary structural characteristics of AuBP-1 in solution and upon adsorption to the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces under aqueous conditions. Top left depicts the regions of ϕ/ψ phase-space on a Ramachandran plot used to define each characteristic- β, PPII, α, α_L, γ_L and random coil (RC). 172
8.5 The binding propensity of individual residues within AuBP-1 at the aqueous a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1) interfaces rated according to the percentage likelihood of being in direct contact with the gold surface. Residues labelled as ‘strong’ binders are highlighted in magenta, ‘significant’ binders in ‘cyan’ and ‘moderate’ in yellow. 173

A.1 a) PMF free energy of adsorption methane on (100), (001) and (011) quartz surfaces as a function of surface separation. b) Typical configurations of methane adsorbed onto the (100), c) (001) and d) (011) surfaces. 182
A.2 PMF free energy of adsorption of ammonium on the (100) α-quartz surface as a function of surface separation calculated with CHARMM and AMBER parameters for NH_4^+. 184

A.3 Orientational analysis of butyl ammonium as a function of vertical surface separation from the a) (100), b) (001) and c) (011) α-quartz surfaces. The angle the CH_3–NH_3^+ vector of butyl ammonium made to the surface normal was calculated as a function of surface separation. Angles less than 60° were classed as 'N-up', between 60° and 120° as 'flat', and greater than 120° as 'C-up'. The surface-separation corresponding to the minimum in the PMF profile of butyl ammonium above each surface is indicated by the blue line. The green lines on the (100) orientation plot a) define the region of the plateau in the PMF profile above this surface. 185

A.4 Orientational analysis of benzene as a function of vertical surface separation from the a) (100), b) (001) and **c)** (011) α-quartz surfaces. The angle between the normals to the planes made by the benzene aromatic ring and the quartz surface was used to classify adsorbed structures as follows: benzene plane made to the surface normal was classified as follows: structures with angles of less than 30° or greater than 150° correspond to 'upright', those greater than 60° but less than 120° to 'flat', and all others to 'tilted'. The surface-separation corresponding to the minimum in the PMF profile of benzene above each surface is indicated by the blue line. The green lines on the (100) orientation plot a) define the region of the plateau in the PMF profile above this surface. 186

A.5 Orientational analysis of ethanoate as a function of vertical surface separation from the a) (100), b) (001) and **c)** (011) α-quartz surfaces. The angle the CH_3–CO_2^- vector of ethanoate made to the surface normal was calculated as a function of surface separation. Angles less than 60° were categorised as 'carboxylate-up', between 60° and 120° as 'flat', and greater than 120° as 'methl-up'. The surface-separation corresponding to the minimum in the PMF profile of ethanoate above each surface is indicated by the blue line. The green line on the (100) orientation plot a) corresponds to the local minimum in the PMF profile above this surface. 187

B.1 Displacement of a) NH_4^+ and b) CH_3COO^- from their initial positions, during CPMD simulations, along the $[\bar{1}20]$ and [001] vectors in the (100) quartz surface. These vectors point along the major grooves in the surface. 189

B.2 Radial Distribution Function plots for water surrouding ethanoate and ammonium from which the distance criteria for hydrogen bonding were derived, for both CPMD and FMD approaches. 191

C.1 Temperature mobilities for 4 out of the 36 replicas in each T-REMD simulation. Replicas shown are (clockwise from top left) 1 (black), 12 (green), 24 (red) and 36 (gold). 195

C.2 Effective temperature mobilities for 4 out of the 11 replicas in each REST simulation. Replicas shown are (clockwise from top left) 1 (black), 4 (green), 8 (red) and 11 (gold). 196

C.3 Percentage of the total structure population belonging to each of the three most populated clusters as a function of MD steps shown for 'REST 2 off', 'REST light' and 'REST light 2 off' simulations. Data shown for cluster analysis performed over the entire trajectory and over the last 4x10^6 MD steps only (inset). Clusters comprised of helical structures are shown in red. 197

D.1 Hollow sites surrounding top sites T1 and T2 tested in methane–Au(100)(5×1) PW-DFT calculations. Surface layer Au atoms are depicted as largest and in medium gold, 2nd layer Au atoms dark gold and medium size, while the 3rd layer of Au atoms are the smallest size and light gold in color. 204

D.2 Optimal geometries of a number of small molecules adsorbed onto Au(111) *in vacuo* obtained from PW-DFT calculations using the vdW-DF exchange functional. Images rendered using XCrySDen. Gold atoms gold, hydrogen cyan, oxygen red, nitrogen blue, carbon yellow and sulphur cream. 205

D.3 Optimal geometries of a number of small molecules adsorbed onto Au(100)(1×1) *in vacuo* obtained from PW-DFT calculations using the vdW-DF exchange functional. Images rendered using XCrySDen. Gold atoms gold, hydrogen cyan, oxygen red, nitrogen blue, carbon yellow and sulphur cream. 206

D.4 Most favorable geometry of each small molecule adsorbed onto the Au(100)(5×1) surface found from PW-DFT calculations using the vdW-DF functional. Left depicts a view from above the surface whilst right the system is shown from the side. Gold atoms are shown in gold (for the plan view only: surface layer are largest in medium gold, 2nd layer are dark gold and 3rd layer are smallest and light gold), carbon in green, hydrogen in white, oxygen in red, nitrogen in blue and sulfur in yellow. 207
D.5 Most favorable geometry of each small molecule adsorbed onto the Au(100)(5×1) surface found from PW-DFT calculations using the vdW-DF functional. Left depicts a view from above the surface whilst right the system is shown from the side. Gold atoms are shown in gold (for the plan view only: surface layer are largest in medium gold, 2nd layer are dark gold and 3rd layer are smallest and light gold), carbon in green, hydrogen in white, oxygen in red, nitrogen in blue and sulfur in yellow. 208
D.6 Side view of the gold slab highlighting the Au-Au-Au angle about the T3 top site. 211
D.7 Correlation between experimentally-determined interaction energies of alkanes of varying chain length (C_nH_{2n+2}) [Wetterer et al. [1998]] with Au(111), and those calculated using GolP-CHARMM parameters for gold. 216

E.1 Trajectories of four representative replicas (0, 5, 10, 15) through effecive temperature space in simulations a) 'sol', b) 'quartz' and c) 'gold'. 225
E.2 Number of clusters of QBP-1 structure as a function of simulation time. 226
E.3 Distribution of functional group—gold distances used to define residue specific adsorption cut-offs at the aqueous (100) α-Quartz interface. 227
E.4 Distribution of functional group—gold distances used to define residue specific adsorption cut-offs at the aqueous Au(111) interface. 228
E.5 Pie charts depicting the frequency with which each region of ϕ/ψ space was sampled in REST simulations of QBP-1 adsorbed at the aqueous (100) α-Quartz interface modelled using a) CHARMM22* and b) CHARMM27. 230

E.6 Normalized frequency distribution of residue backbone atoms above the (100) α-Quartz and Au(111) surfaces. Cut-offs for backbone group surface binding are demarked by dotted lines (dark green- quartz and red- gold). 235

F.1 Trajectories of representative replicas (0, 5, 10, 15) through effective temperature space during a) REST MD simulations (of AuBP-1 in solution and at the aqueous Au(111) interface) and during b) REST metaD simulations (at the aqueous Au(111), Au(100)(1×1) and Au(100)(5×1) interfaces). Data shown for first 10 ns of all simulations only. 237

F.2 Evolution of ln(W) (Equation 2.31) as a function of simulation time at the aqueous a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1) interfaces. 239

F.3 Time evolution of the free energy of adsorption of AuBP-1 to a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1). Plotted in black is the minimum energy based on the peptide binding to either the top or bottom surface of the gold slab; error bars represent the energetic differences between the two interfaces. The average energy of adsorption to the two identical gold surfaces present in each simulation cell is demarked by a green line, while the free energy of adsorption, with its associated error, calculated self consistently using the 'time period' reweighting method after five iterations is plotted in brown. . 240

Acknowledgments

First and foremost I must gratefully thank both Dr Tiff Walsh and Professor Mark Rodger for their support, advise and guidance during my PhD. I owe a huge deal to Tiff–professionally and personally–for the many opportunities she has given me: be it to attend the CECAM conference in Lausanne, May 2011; to initiate the collaboration with Stefano Corni in 2011; to publish endless(?!) papers; visit Deakin for 5 months; and (not least) encourage me to learn to drive(!). Her unwavering enthusiasm and continuous wealth of ideas has motivated me tremendously throughout. Likewise, I couldn't have got to this point without Mark, not least him being there to 'step into Tiff's shoes' as my primary supervisor at Warwick as of January 2012. The depth of his knowledge has broadened my horizons significantly and with his guidance I have developed as an individual. One of the most enjoyable memories of my PhD is the 2 hours spend in a room in the VLSCI centre in Melbourne discussing citrate monolayers and REST metaD with both Tiff and Mark.

All of the work at the aqueous gold interface reported in this thesis would not have been possible without collaboration with Professor Stefano Corni, Centro S3CNR Istituto Nanoscienze, Modena, Italy. For the chance to visit Modena twice to learn the skills needed for deriving GolP-CHARMM and, on the second occasion, to use the newly commissioned FERMI supercomputer, I am very grateful.

Whilst not directly funded by the flagship EPSRC grant I have been lucky enough to work closely with 'Materials Interfaced with Biology' consortium over the last 3 years. The chance to talk and meet with others in a similar field regularly–both experimentalists and modellers–was motivational. I particularly acknowledge Colin Freeman, for the opportunity to work on a collaborative paper with him during this time.

At Warwick I thank group members (Walsh and Rodger) past and present and those I have shared an office with. Particular thanks goes to Jasmine Desmond, David Bray and Aaron Finney, for useful discussions, 'being there' for each other and company during all those long hours. Equally, I am very grateful to those in Deakin who all contributed to making my 5 months in Australia both very enjoyable and academically successful. Special thanks to Aaron Brown for his friendship and help.

Last and not least, thanks to my parents–for their support throughout–and friends Nicola Cumley and Beth Hanson–for, periodic 'sanity breaks' involving many cups of tea and coffee!

Declarations

This thesis is all my own work except where explicitly stated in the text (Chapters 7 and 8). No part of this report has previously been submitted for examination for any other higher degree.
To date Chapters 3, 4, 5 and 6 have been published as the following journal articles:

Facet Selectivity of Binding on Quartz Surfaces: Free Energy Calculations of Amino-Acid Analogue Adsorption, L. B. Wright and T. R. Walsh, J. Phys. Chem. C, 116, p2933.

First-principles Molecular Dynamics simulations of NH_4^+ and CH_3COO^- adsorption at the aqueous Quartz Interface, L. B. Wright and T. R. Walsh, J. Chem. Phys., 2012, 137, p224702.

Efficient Conformational Sampling of Peptides Adsorbed onto Inorganic Surfaces: Insights from a Quartz Binding Peptide, L. B. Wright and T. R. Walsh, Phys. Chem. Chem. Phys., 2013, 15, p4715.

GolP-CHARMM: First-principles based Force-fields for the Interaction of Proteins with Au(111) and Au(100), L. B. Wright, P. M. Rodger, S. Corni and T. R. Walsh, J. Chem. Theory Comput., 2013, 9, p1616.

Abstract

The aqueous interface between biomolecules and inorganic substrates is of interest to many cross-disciplinary areas of science, ranging from fundamental biological research into biomineralisation processes, to the more application-driven fields of nanotechnology and biomimetic materials synthesis. In particular, by harnessing the selectivity observed in Nature, proteins and peptides make ideal candidates for directing the assembly of nanoparticles into nanostructured, multifunctional materials with pre-defined physical properties. The rational design of peptide sequences with tunable affinities either for a substrate of a specific composition, or for a specific crystallographic plane of a given material, would mark an important step towards realising these goals. Before this is possible, however, the fundamental mechanisms involved in peptide-substrate binding under aqueous conditions must be understood.

Molecular simulation, used throughout this thesis, is well suited for studying biointerfacial systems at the level of detail needed to advance research. The work presented herein is primarily focused on answering the question of whether facet-selective peptide adsorption is indeed possible at the aqueous quartz and gold interfaces. A range of different simulation techniques, all based on atomistic molecular dynamics, are employed. As part of the study, two of the ‘grand challenges’ currently facing biointerfacial simulation–the deviation of force-fields suitable for the interfacial environment, and enhancing conformational sampling of an adsorbed biomolecule–are addressed.

Potential of Mean constraint Force free energy of adsorption calculations show that, out of the amino-acid analogues tested, all display energetic and/or spatial selectivity between the (100), (001) and (011) surfaces, on adsorption to Quartz. Facet specificity in binding, for the building blocks from which peptides are comprised, is highly suggestive of the biomolecules themselves displaying similar characteristics. The general trend in small molecule adsorption strength–non-polar aromatic > negatively-charged > non-polar aliphatic > positively-charged–is common to all three aqueous Quartz interfaces. The propensity for negatively-charged ethanoate to bind more strongly to the aqueous, fully hydroxylated (100) α-Quartz surface than positively-charged ammonium is confirmed by first-principles simulations.

Selectivity between the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces is also observed when the gold-binding peptide AuBP-1 [Hnilova et al. [2008]] adsorbs to gold under aqueous conditions. Metadynamics, in combination with the advanced sampling technique Replica Exchange with Solute Tempering, is used to study this system. The impact of Au(100) reconstruction on peptide binding is considered here for the first time. This aspect of gold binding was made possible by a suite of force-fields derived within this work, GolP-CHARMM, for modelling the interactions between proteins and peptides and gold.

Abbreviations

AFM	Atomic Force Microscopy
AuNP	Gold Nanoparticle
BOMD	Born Oppenheimer Molecular Dynamics
COM	Centre of Mass
CPMD	Car Parrinello Molecular Dynamics
DFT	Density Functional Theory
FF	Force-field
GGA	Generalised Gradient Approximation
H-REMD	Hamiltonian Replica Exchange Molecular Dynamics
LDA	Localised Density Approximation
LJ	Lennard-Jones
MD	Molecular Dynamics
metaD	Metadynamics
NMR	Nuclear Magnetic Resonance
NP	Nanoparticle
NPT	Isothermal-Isobaric ensemble (constant number, pressure and temperature)

NVE	Microcanconical Ensemble (constant number, volume and energy)
NVT	Canconical Ensemble (constant number, volume and temperature)
PEL	Potential Energy Landscape
PME	Particle Mesh Ewald
PMF	Potential of Mean Force
PW-DFT	Plane Wave Density Functional Theory
QCM	Quartz Crystal Microbalance
RDF	Radial Distribution Function
REMD	Replica Exchange Molecular Dynamics
REST	Replica Exchange with Solute Tempering
RMSD	Root Mean Squared Deviation
SFG	Sum Frequency Generation Spectroscopy
SPR	Surface Plasmon Resonance
T-REMD	Temperature Replica Exchange Molecular Dynamics

Chapter 1

Introduction

Fundamental understanding of the aqueous interface between biological molecules and inorganic substrates is of interest to many cross-disciplinary fields of science, ranging from biology and medicine to biomimetic material synthesis and nanotechnology. In this work I have used molecular simulation to investigate the mechanisms governing peptide adsorption at an oxide–quartz–and a metallic–gold–interface. Specifically this report focuses on the potential of a peptide preferentially to adsorb to one material, or one crystallographic surface of a given material, over another.

The future applications which motivate this research are outlined in Section 1.1 before a brief review of the current methods used to select or design material binding peptides, Section 1.2. The current challenges to gaining greater insight into peptide-surface adsorption at the aqueous quartz and gold interfaces are introduced in Sections 1.3 and 1.4, respectively.

1.1 The Potential of material and facet selective Peptides

In Nature, biomolecules perform a key role in the formation, structure and function of biological hard minerals (*e.g.* hydroxyapatite [Capriotti et al. [2007]], nacre [Fu et al. [2005]; Collino and Evans [2008]; Metzler et al. [2010]; Amos et al. [2011]] and silica [Kroger et al. [1999, 2001, 2002]; Sumper and Brunner [2008]]), while interfaces between biomolecules and non-biological inorganic substrates are increasingly prevalent in the modern world (*e.g.* joint replacements). In recent years, protein-surface interactions, in particular, have been the subject of much research [Latour [2008]; Tamerler et al. [2010]; Seker and Demir [2011]; di Felice and Corni [2011]]. Since the structure of a protein is integral to its function, understanding the mechanisms driving surface adsorption–including binding affinity, structure and aggregation–is valuable for many reasons. For example, in the realm of medicine, this information could be used to improve the compatibility of medical implants [Walczyk et al. [2010]], predict the toxicity of nanoparticles (NPs) [Shemetov et al. [2012]; Vecchio et al. [2012b,a]] and aid the design of NP-based therapeutics [Colombo et al. [2012]; Kumar et al. [2012]; Liao et al. [2012]; Black et al. [2013]] or diagnostics [Chakraborty et al. [2011]; Mahmood et al. [2012]]. Alternatively, if harnessed, the selectivity of biomolecule-substrate binding observed in Nature makes proteins and peptides (short chains of up to ∼50 residues) ideal candidates for controlling the synthesis and subsequent organization of NPs into pre-defined nanostructures. Not only does this biomimetic approach to material synthesis [Tamerler and Sarikaya [2009]; Tamerler et al. [2010]; Hnilova et al. [2012c,a,b]; Schoen et al. [2011]; Coppage et al. [2011]; Briggs and Knecht [2012]] offer the potential for designing new smart materials, but it can also be carried out under ambient temperature and pressure; significantly less energy-intensive conditions than many of the procedures currently used in manufacturing.

Fusion proteins comprising of peptide sequences with selective affinity to different materials are one way in which to direct the assembly of NPs into nanomaterials of mixed composition. For example, by exploiting substrate-selectivity, the surface of one material could be patterned in a controlled manner by NPs of an alternative composition using a protein that features both material-binding domains [Hnilova et al. [2012b]] (Figure 1.1a). Alternatively, heterogeneous surfaces could be decorated by NPs using two or more sets of fusion proteins, each presenting a different material-binding sequence (Figure 1.1b). In recent years the quartz–gold system has been used several times as an exemplar test case of these biomimetic synthetic strategies [Hnilova et al. [2012c,a,b]; Tamerler et al. [2010]]. Interest in these interfaces in particular was not just a scientific exercise: the potential to exploit the unique physical properties of noble metal NPs to their full potential in a new material, whilst retaining economic feasibility, is key to the practical realization of many nanotechnological applications. Silica–an abundant raw resource–and gold–an expensive commodity–are therefore an optimal combination.

While thus far identification of material-selective peptides has been the primary focus of research, interest in finding sequences able to discriminate between different crystallographic planes of the same substance is growing [Heinz et al. [2009]; Yu et al. [2010, 2012]; Feng et al. [2012]; Chiu et al. [2011]; Coppage et al. [2012]; Ruan et al. [2013]]. Facet-selective peptides offer the potential to regulate NP synthesis such that only NPs of a given shape are produced. NP growth is widely

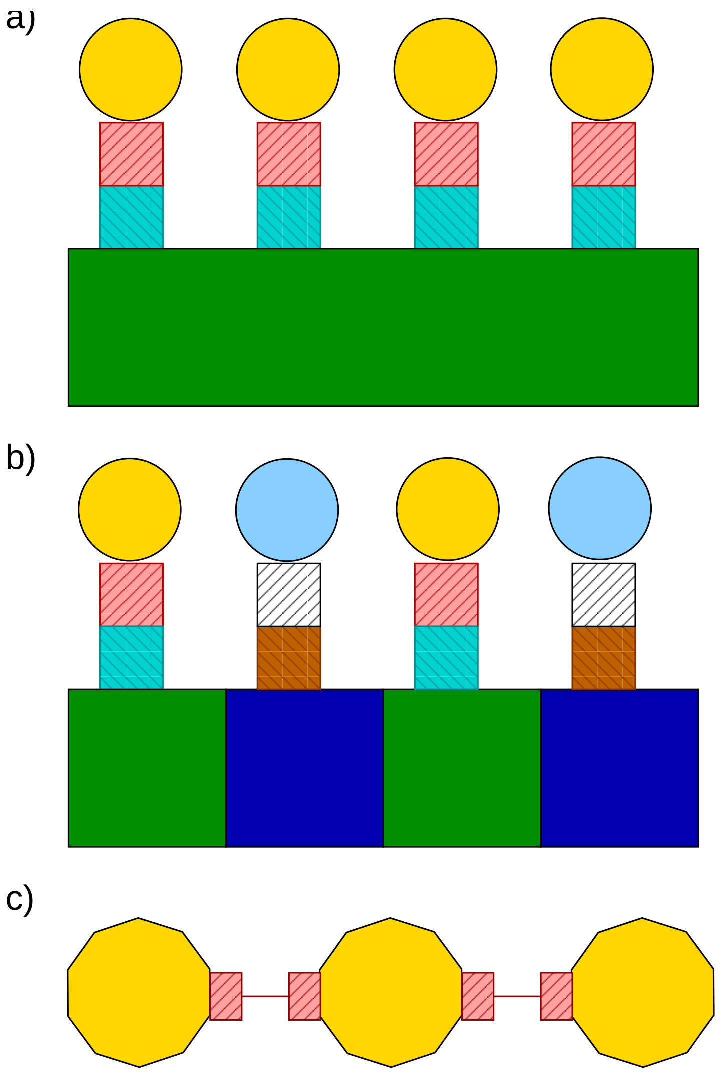


Figure 1.1: Scheme to depict how material- and facet-selective peptide sequences can be used to direct the fabrication of a) a regular array of NPs above a uniform surface, b) an array of NPs of mixed composition on a patterned surface and c) a matrix of NP orientationally aligned.

thought to be a kinetically-driven process, with low energy surfaces growing faster and thus featuring more prominently on the final particle, than high energy ones. On adsorption of a capping ligand, the surface energy of the facet in question is modified. Therefore, by selectively controlling which crystallographic planes are passivated by peptide binding potentially allows us to control the overall morphology of a NP. Furthermore, facet-selective peptide sequences could help regulate NP assembly in an orientational manner so that nanostructures of anisotropically ordered NP are produced (Figure 1.1c). While advantageous for NPs of any composition, the greatest promise of facet-selective peptides comes in the case of noble metals where the unique optical and chemical properties of NPs depends critically on size, shape and assembly. For instance, the plasmonic adsorption frequency of a nanosensor could be finely tuned if composed only of NPs of a specific size and shape, systematically assembled. Similarly, it has recently been shown that the selectivity and efficiency of NP-based catalysts can be improved by the rational design of the NPs incorporated [Shuai et al. [2013]].

To date, synthetic strategies for size and shape selective production of NPs reported in the literature can be broadly grouped into two classes: seed mediated growth (*e.g.* the synthesis of five-fold stellar AuNPs featuring 110 facets [Jiang et al. [2013]]); and those aided by the addition of additives–ions [Peng et al. [2013]], surfactants [De and Mandal [2013]; Danger et al. [2012]] and peptides [Kim et al. [2010]; Coppage et al. [2012]; Gkikas et al. [2013]; Chiu et al. [2011]; Ruan et al. [2013]; Chiu et al. [2012]]. While 'additive-mediated' synthetic strategies to anisotropic AuNPs with rod-like dimensions exist–for example Au nanorods and nanopods produced in the presence of 4-dimethylaminopyridine (DMAP) [Danger et al. [2012]]–, unlike platinum [Chiu et al. [2011]; Ruan et al. [2013]] and palladium [Peng et al. [2013]], those to exert fine control over the morphology and facet composition of spherical AuNPs are lacking. Recent progress in this regard was made by Kim *et al.* however, who reported that the size and shape of AuNPs precipitated in the presence of the gold binding peptide 'Midas-11' (sequence 9 Table 1.2) and its variants could be regulated in a controlled manner by systematically varying the experimental conditions [Kim et al. [2010]]. For example, the mutant G11Y produced mainly large (1-2 μm) trigonal, truncated trigonal or hexagonal platelets while the mutant G11H gave rise to platelets of mixed size and shape (both at pH3 and 37°C).

1.2 Selection of Material Binding Peptides

Since the pioneering work of Brown, who reported the first genetically engineered material-binding peptides [Brown [1997]], numerous sequences which selectively adsorb to a range of other non-biological substrates (*e.g.* metals, polymers, semiconductors and oxides) have been identified (see a recent review by Seker and Demir for references [Seker and Demir [2011]]; silica and gold binding peptides are discussed in detail in Sections 1.3 and 1.4,respectively). At present, the most commonly used selection processes involve screening large peptide libraries by biocombinatorial techniques. In what follows, the principles behind, and the problems associated with, these processes are outlined, before the challenges which must be met before peptide sequences with tunable affinity for a given substrate can be rationally designed are discussed.

1.2.1 Biocombinatorial Peptide Selection

To date two biocombinatorial methods, originally designed to probe biological protein-ligand interactions, have been used to identify material-binding peptides: phage [Brown [1997]] and cell surface [Wittrup [2001]] display. The principles behind both are the same. First, random sequences of DNA are inserted into the virus or cell genome. The insertion point is chosen to be a site within the genetic information corresponding to a protein expressed on the surface of the phage or cell. The phages/cells are then exposed to the inorganic substrate of interest. Those with coat proteins featuring peptides with high surface-affinity adsorb to the interface, whilst those with little or no affinity are removed by washing. Bound phages/cells are eluted and amplified. In the case of phage display, amplification is achieved by bacterial infection. The biopanning process described is repeated–each round using only peptide sequences expressed by those phages/cells which were eluted from the interface at the end of the previous iteration–until a consensus is reached. Finally, by sequencing the genome of the phage/cell which adsorbs the most strongly to the substrate, high affinity material-binding peptides can be identified.

There are a number of problems associated with only using biocombinatorial techniques to identify peptide sequences. Common to both procedures, the display peptide libraries that can be screened only contain $\sim$0.1% of the residue combinations possible for a peptide of a given length. The shortfall arises from either rejection of the inserted DNA or the inability of the phage/cell to express the modified coat protein. In addition, many amino-acids can be represented by more than one DNA codon (a triplet of DNA bases). However, the number of DNA codons encoding an amino-acid varies–for example, arginine, leucine and serine are each represented by 6 different codons, while methionine and tryptophan are associated with only one codon each. Therefore, even if all the possible combinations of DNA bases for a given length of peptide were expressed by a phage/cell display library, some peptide sequences would be represented more often than others. In the case of cell surface display, a third layer of bias is introduced at the translation stage–tRNA complimentary to each amino-acid are not uniformly available. On the other hand, the peptide sequence genetically inserted into the coat protein of a virus may influence the rate at which it can infect host cells used in the amplification stage. Peptides which mediate strong surface adsorption but hinder bacterial infection are hence less likely to be identified at the end of the procedure. Each of these issues can be used to account for the disparity between the primary sequences of material-binding peptides identified using different peptide libraries (see Tables 1.1 and 1.2 silica and gold binding sequences, respectively). Moreover, the most common method used to elute bound phages/cells favours selection of charged or polar peptides over hydrophobic ones, independent of the peptide library screened [Puddu and Perry [2012]]. Differences in conformation, between a peptide expressed as part of a phage/cell surface coat protein and fully solvated, are also not accounted for by these identification processes [Hnilova et al. [2008]].

Although a third biocombinational technique, ribosome display [Hanes and Pluückthun [1997]; He and Taussig [1997]]–yet to be applied to interfacial systems–can support larger peptide libraries and is neither biased by tRNA availability nor the rate of bacterial infection, it is more difficult to implement and does not overcome all of the problems associated with phage and cell surface display mentioned above.

1.2.2 Towards the Rational Design of Peptide Sequences

The ability rationally to design material-binding peptide sequences would mark a significant breakthrough in the realm of biomimetic research, by-passing the lengthy, costly, and sometimes inconsistent, experimental peptide selection process (Section 1.2.1). Two approaches are possible: one in which sequences are constructed from knowledge of the fundamental mechanisms governing peptide adsorption to an interface; and the other in which algorithms are used to predict strong-binding sequences based on the statistical occurrence of residues and motifs to feature in the high-affinity peptides identified by biocombinatorial experiments. The latter, bioinformatics, has already been successfully implemented at the aqueous silica interface [Oren et al. [2007]]. Using data from phage display selection experiments at the (100) α quartz interface [Tamerler et al. [2007]], the peptide, QBP-1 (PPP-

WLPYMPPWS) [Oren et al. [2007]], was designed. In this case the bioinformatics approach was validated by subsequent surface plasmon resonance (SPR) measurements which confirmed the high affinity of QBP-1 for silica. However there is a potential for biases in the experimental data-set employed in a bioinformatics study [Puddu and Perry [2012]] to be extrapolated by this method. The first approach, rationally constructing sequences using our knowledge of the physics and chemistry of the system, is, on the other hand, more robust.

Both the affinity of individual residues for a substrate and the conformation of a peptide as a whole are critical to surface adsorption. In an attempt to probe the first of these effects–the direct contribution of each residue to binding–in isolation from the second–peptide conformation–two approaches have been adopted: amino-acid [Alaeddine and Nygren [1996]; Zimmerman et al. [2004]; Churchill et al. [2004]; Vlasova and Golovkova [2004]; Gao et al. [2008]; Kitadai et al. [2009]; Lopes et al. [2009]; Holinga et al. [2011]] and host-guest peptide [Cohavi et al. [2011]; Wei and Latour [2010]; Thyparambil et al. [2012]; Willet et al. [2005]; Peelle et al. [2005]] adsorption experiments. The results from both methods must be interpreted with care. For instance, the adsorption of an amino-acid may be predominantly mediated by its amino- or carbonyl-termini. These moieties, however, are sterically hindered from interacting with a substrate when part of a peptide backbone. Whilst this is not an issue for host-guest systems–protein or peptide sequences where one or more residues are systematically replaced–disentangling the effect of amino-acid substitution on the conformation of the host-guest system itself from the affinity of the residue in question for the interface is difficult [Skelton et al. [2009]]. The influence of the 'host' must also be considered–a fact highlighted when interpreting the work of Willet *et al.* [Willet et al. [2005]]. In this study, surface adsorption of homopeptides was quantified by the fluorescence of an attached marker [Willet et al. [2005]]. Unwanted interactions between the fluorophore and either the surface or the homopeptide gave rise to anomalies in their results (*e.g.* the negligible affinity of cysteine for gold).

1.2.3 Biointerfacial Experimental Techniques

Experimental characterization of bio-interfacial systems, such as determining the free energy of adsorption or adsorbate-surface geometry, presents a significant challenge in itself. At present, surface plasmon resonance (SPR) and quartz crystal microbalance (QCM) are the methods most commonly used to monitor the kinetics of, and infer thermodynamic data about, biomolecule adsorption. In the first, SPR, photons of light are passed through a glass prism, exciting surface electrons in the conduction band of a thin sheet of metal attached. The resonance frequency for surface plasmon adsorption is dependent on the refractive index of the media in contact with the metal film. Therefore the rate of biomolecule-surface adsorption and desorption can be determined by monitoring SPR signal intensity. Fundamental to SPR is the use of a metal which supports surface plasmons (*e.g.* Ag or Au) although techniques to extend this method to alternative inorganic substrates have been pioneered. For example, Sarikaya and *co-workers* have quantified peptide-silica adsorption by applying a thin layer of $\mathrm{Si}_x\mathrm{O}$ to the gold surface of an SPR chip [Oren et al. [2007, 2010]], while Cohavi *et al.* immobilized AuNPs onto a NeutrAvidin-coated surface to monitor protein-NP binding. However, the range of non-metallic interfaces to which SPR can be applied is limited to those which are topologically flat and thinner than the penetration depth of the incident light source.

A QCM device is composed of a piezoelectric quartz wafer placed between two gold plates. The resonance frequency of the quartz, in the presence of an AC electric field, is dependent on the mass of adsorbate adhered to the surface. Like SPR, with QCM, the range of substrates, alternative to quartz, is limited to those which can form a uniform coating of nanoscale dimensions.

Wei and Latour recently devised a protocol for using Atomic Force Microscopy (AFM) to measure the binding affinity of peptides to substrates which cannot be studied by SPR or QCM [Wei and Latour [2010]]. Previously there were two fundamental obstacles to relating the desorption force measured in an AFM experiment to the free energy of adsorption: determining the number of peptides attached to an AFM probe tip that actually interact with a surface; and removing the contribution from the tip alone to surface adhesion. The latter can be overcome by using long flexible tethers between AFM tip and peptide of interest. To tackle the former hurdle, host-guest TGTG-X-GTCT peptides were used to calibrate AFM data against SPR measurements [Wei and Latour [2010]]. The aqueous alkanethiol self-assembled monolayers (SAMs) interface, suitable for SPR experiments as well as AFM, was used in this calibration stage. This enabled the free energy of adsorption of the same host-guest peptides to a range of other surfaces, including (100) quartz and fused silica glass, to be estimated from AFM experiment alone [Thyparambil et al. [2012]].

A full description of both the internal conformation of a biomolecule and its geometry in relation to a substrate, at the level of detail required for designing material-binding peptide sequences, remains elusive to obtain experimentally in all but a small number of cases. Many of the techniques commonly used to determine peptide conformation in solution can not be extrapolated to the bound state.

Circular Dichroism (CD) spectroscopy is one such example. Using this technique, adsorption in the far UV region of the electromagnetic spectrum is monitored. The ensemble- and sequence-averaged likelihood that a peptide has a particular secondary structural characteristic is found using reference databases containing the CD spectra of proteins of known conformation. However, the method relies on light passing all the way through a sample, making it intrinsically unsuitable for interfacial systems. However, promising recent advances have been made in two alternative fields–Sum Frequency Generation (SFG) Spectroscopy and Nuclear Magnetic Resonance (NMR).

Two photons of light are projected at an interface–one in the infra red (IR) and the other in the visible region–to obtain a SFG spectrum. Vibrational transitions arise from resonance in the second order susceptibility, χ, of the system. Interfacial media can be selectively studied over bulk since χ=0 for all entities which include a point of inversion. By assigning adsorption bands to different X-H bonds (X=C, O, N), the average orientation of a biomolecule with respect to an interface can be determined. For example, Somorjai and *co-workers* used SFG to contrast the adsorption of amino acids [Holinga et al. [2011]] and homopeptides [York et al. [2009]] to a hydrophobic (polystyrene) and a hydrophilic (fused silica) surfaces under aqueous conditions. The absence of C-H stretching bands in a SFG spectrum can mean one of two things: no adsorbate is present; or that the adsorbate is randomly oriented. One of the major drawbacks of SFG therefore, is that it cannot be used to elucidate information on a single molecule level.

Out of all the experimental methods discussed, NMR perhaps holds the most promise for future experimental characterization of biointerfacial systems, with the possibility of obtaining both spatial and orientational data. Unlike SPR, however, it can not intrinsically discriminate between interfacial and bulk phases. This at present therefore limits its use to substrates with high surface areas (*e.g.* porous materials) and consequently necessitates with use of solid-state magic angle spinning (MAG) NMR, for which the resolution is much lower. Overcoming these difficulties, Ben Shir *et al.* studied the adsorption of alanine [Ben Shir et al. [2010]] and glycine [Ben Shir et al. [2012]] to amorphous silica in dry and microsolvated conditions using ^{13}C and ^{15}N labelled amino-acids. Adsorbate-silica geometry was determined in each case from inter- and intra-molecular distances.

Nuclear Overhauser Effect NMR spectroscopy (NOSEY), a technique commonly used to investigate the native state of proteins in solution, was in 2011 successfully used by Mirau *et al.* to elucidate the adsorbed conformation of a peptide for the first time [Mirau et al. [2011]]; saturation-transfer difference (STD) NMR was used in the same study to determine peptide-substrate orientation. The TBP-12 [Sano et al. [2005]] peptide adsorbed to either silica or titania NPs under conditions of rapid exchange was investigated. In this cutting-edge study, the sign and magnitude of the NOE signal–dependent on the rate of molecular reorientation in solution–was used to discriminate between bound and free states of the adsorbate. Cross peaks in a NOSEY spectrum identify residues in close spatial proximity to each other, while protons bound to a surface can be recognized from the difference between the standard NMR spectrum and one in which the signal from the hydroxyl groups and water is saturated. In a similar vein to STD, Calzolai *et al.* had previously employed chemical shift perturbation (CSP) NMR to identify patches on the surface of an isotopically labelled protein (ubiquitin) through which AuNP-binding was mediated [Calzolai et al. [2010]].

1.2.4 Biointerfacial Simulation

While at present many of the difficulties discussed above in Section 1.2.3 still hinder experimental determination of peptide conformation and/or aggregation upon adsorption, such information is, in principle, readily attainable from molecular simulation. Numerous studies modeling a range of different substrates have been carried out in recent years–see Sections 1.3 and 1.4 for a review of recent quartz and gold biointerfacial simulations. However, the 'success' of each–determined by the accuracy with which the essential physics and chemistry of the system are reproduced–depends critically on the model and simulation protocol adopted [Latour [2008]].

First-principles, quantum mechanical methods can be used to study a system at the most fundamental level with no (or in the case of Density Functional Theory , DFT, minimal) empirically fitted parameters (Section 2.1.1). These calculations can be highly accurate but are as a consequence, intrinsically very computationally demanding. This therefore limits their use to static calculations or simulations on the order of ps (Section 2.2.4), for systems of small dimension ($\ll$1000 atoms) only. Both the small size and short duration of a first-principles simulation impose implementation and interpretation difficulties–namely, the potential to introduce unphysical artifacts due to cell periodicity; and the ability adequately to equilibrate and sample different conformational states of a system, respectively. In addition many biointerfacial electronic structure theory (EST) calculations in the past have been performed *in vacuo* or implicit solvent. However, it is widely thought that the structure of water at an interface, neglected by both approximations, can play a critical role in peptide binding [Jena and Hore [2010]; Skelton et al. [2009]; Schneider and Ciacchi [2012]]. In the case of DFT, the performance of the exchange-correlation functional used must also be considered: without corrective terms, dispersion interactions (such as those between an amino-acid with a hydrophobic side-chain and a surface) are poorly described. Despite these caveats, first-principles simulation is the only method in which electrons are explicitly accounted. As a consequence of all the possible levels of theory, they give the most faithful representation of atomic polarization and charge transfer, both of which are especially important when modeling metallic interfaces or chemisorption events.

Atomistic models employing empirical force-fields (FFs) to describe the interactions between 'atoms' (entities of point mass and charge) are more computationally accessible, enabling scientists to study larger systems (on the order of millions of atoms) over longer time-scales (up to 1 μs). Although over recent years this type of simulation has become the method of choice for many in the field, a number of challenges remain.

Many well-established FFs exist for solid inorganic materials (for example, see Ref [Herzbach et al. [2005]] for a review of silica potentials), while those parameterised to be compatible with existing biomolecular FFs are fewer in number. In the latter, the description of the bulk phase of the solid substrate is often compromised in favour of compatibility and accurate representation of the surface itself [Lopes et al. [2006]]. To overcome this issue, some either employ methods such as those outlined by Schroder *et al.* [Schroder et al. [1992]] to fit cross-terms between organic and inorganic phases [Freeman et al. [2007]] or derive parameters for a biointerfacial FF in which the co-ordinates of the substrate are held fixed in space [Iori et al. [2009]; Iori and Corni [2008]; Wright et al. [2013b,c]]. The latter approach may introduce artifacts into the model [Phan et al. [2012]; Wright et al. [2013a]], however.

Like in solution, the choice of biomolecular FF employed to model the organic component of the system and with which the interfacial FF is designed to be compatible, is also critical, especially for peptide conformational studies. Although refinement of bio-organic FFs is a rapidly advancing area of research, a number of reports have shown that the most recent version of CHARMM (CHARMM22* [Piana et al. [2011]]) performs well for predicting the structure of proteins and peptides in solution [Piana et al. [2011]; Collier et al. [2012]; Lindorff-Larson et al. [2012]; Cino et al. [2012]]. Either CHARMM27 [MacKerell et al. [1998, 2004]] or CHARMM22* [Piana et al. [2011]] were used in all the simulations reported here. Questions remain, however, over the transferability of standard biomolecular FF parameters (derived for solution-based studies) to bio-interfacial simulations [Vellore et al. [2010]; Collier et al. [2012]]. To tackle this issue, Latour and *co-workers* very recently proposed a dual parameter protocol in which a refined set of intra-peptide interaction terms are used when the peptide is within the interfacial region, while standard CHARMM parameters are adopted in solution [Snyder et al. [2012]].

The ability adequately to sample all the relevant conformational states of a system during a molecular simulation poses one of the biggest challenges for the future of both bio-interfacial and solution-based studies. The need explicitly to account for water in the former case intrinsically makes the systems both large and complex. Thus, in many cases, unless advanced sampling techniques are employed, it is not possible to explore all of the important low energy states of the system within the time-scale of a simulation. As a worst case scenario, the system may remain trapped in a single deep well in its potential energy landscape throughout an entire simulation. Unless these difficulties are overcome the use of computational modeling to probe both the hypothesis of peptide-substrate binding induced peptide-folding [Evans et al. [2008]; Capriotti et al. [2007]; Collino and Evans [2008]; Rimola et al. [2012]]–thoroughly, and calculate peptide binding affinities accurately, is hindered.

1.3 Aqueous Quartz Interface

Due to the ubiquitous prevalence of silica in the natural environment, the aqueous biomolecule-silica interface is of particular interest. Taking inspiration from marine organisms such as diatoms, which can form intricate silica nanostructures, silica is an environmentally friendly and economically advantageous resource from which new biomimetic materials could potentially be made. In this work, the affinity of amino-acid analogs for Quartz, the lowest energy crystalline polymorph of silica under ambient conditions, (Chapters 3 and 4) and the conformation of a quartz-binding peptide (QBP-1 [Oren et al. [2007]]) upon adsorption (Chapters 5 and 7) have been studied using three different simulation techniques: potential of mean force (PMF) MD, Car Parrinello MD (CPMD) and replica exchange MD (REMD), respectively.

First, in this section reviews of both experimental and computational work carried out to date at the peptide-silica interface (Section 1.3.1 and 1.3.2, respectively) are presented, before the specific crystallographic planes of quartz modeled herein are described (Section 1.3.3).

1.3.1 Experiment

Biomineralisation of silica in diatoms occurs by a class of homologically-conserved proteins, silaffins, that have been extensively studied [Kroger et al. [1999, 2001, 2002]; Sumper and Brunner [2008]]. Silaffin mimics, the most studied of which is the R5 peptide [Kroger et al. [1999]], have been widely used by several groups to control the synthesis of silica nanostructures [Brott et al. [2001]; Patwardhan and Clarson [2002]; Lutz et al. [2005]; Soto-Cantu et al. [2010]; Xia and Li [2011]; Patwardhan [2011]]. In addition, peptides capable of binding to silica nanoparticles (either those synthesized biomimetically [Naik et al. [2002a]] or using standard procedures [Patwardhan et al. [2012]; Puddu and Perry [2012]]), the (100) α-quartz surface [Tamerler et al. [2007]] and thermally grown silica films [Eteshola et al. [2005]] have been identified using biocombinatorial techniques (Table 1.1).

Positively charged residues are prevalent in many of the silica binding peptide sequences reported [Brott et al. [2001]; Patwardhan and Clarson [2002]; Naik et al. [2002a]; Eteshola et al. [2005]; Sano and Shiba [2003]; Xia and Li [2011]; Patwardhan et al. [2012]; Puddu and Perry [2012]]. For example, silaffins are enriched in lysine residues [Kroger et al. [1999, 2001]]. However, in silaffins lysines are post-translationally alkylated, and so their interactions with silica could be hydrophobic in character rather than mediated by the terminal ammonium group.

In addition, it has been postulated that the zwitterionic nature of silaffins, with positively-charged ammonium groups and negatively-charged phosphorylated serines, aids silaffin aggregation, rather than the charged groups necessarily having a direct influence on silica precipitation [Kroger et al. [2002]; Sumper and Brunner [2008]]. Early experimental evidence to support this hypothesis was the observation that silaffins formed aggregates of approximately 700 molecules at low ionic strength [Kroger et al. [2002]]. Later studies have shown that phosphate or sulfate ions can induce aggregation of synthetic polyamines in a concentration and pH dependent manner [Brunner et al. [2004]; Lutz et al. [2005]]. The amount and size of the silica nanoparticles precipitated from silicic acid solution by these polyamines correlated with polyamine aggregate size. Hence the prevalence of positively charged residues in silica precipitating peptides reported in the literature [Brott et al. [2001]; Patwardhan and Clarson [2002]; Sano et al. [2005]; Hayashi et al. [2009]; Xia and Li [2011]] is not necessarily an indication that they interact directly with silica surfaces because in each case a phosphate-containing buffer was used.

Moreover, in the case of biocombinatorally selected silica-binding peptides, the nature of the positively charged residue enriched differs depending on the selection conditions. Hayashi *et al.* used AFM to measure adsorption forces between ferritin-fused mutated versions of the N-terminal hexapeptide of TBP-1 (RKLPDA) and silica surfaces [Hayashi et al. [2009]]. In otherwise identical conditions, the R1K mutant (the TBP-1 hexapeptide in which the first residue, arginine, was replaced by lysine) adsorbed less strongly to silica than the wild-type. In the pioneering work by Mirau *et al.* however, peaks in the STD NMR spectrum of TBP-1, the full 12 residue peptide, indicate that hydrogens at the charged ends of both the R1 and K2 side-chains, as well as those belonging to the methyl groups of L3 and A6, were in close proximity to the surface of a silica nanoparticle when adsorbed [Mirau et al. [2011]]. Histidine, rather than lysine or arginine, was enriched in peptides identified by phage display experiments carried out by Eteshola *et al.* [Eteshola et al. [2005]]. Under the selection conditions employed in the latter study, pH 7.4, histidine is predominantly present in solution in its neutral, rather than protonated, form. These findings, taken together with the prevalence of hydrophobic residues in some sequences, especially those that bind strongly to quartz (16-20 Table 1.1), suggest that at least two different peptide-silica binding mechanisms could exist–one mediated by electrostatics and the other by hydrophobic shielding; a hypothesis supported by the recent work of Puddu and Perry [Puddu and Perry [2012]].

To probe the intrinsic affinity of individual residues within a peptide for silica, independently of peptide conformational effects, experiments to study amino-

		sequence	substrate
1	R5	SSKKSGSYSGSKGSKRRIL[1]	PD: silafin derivative
2		MSPHPHPRHHHT[2]	PD: silica ppt. by R5
3		RGRRRRLSCRLL[2]	PD: silica ppt. by R5
4		HPPMNASHPHMH[3]	PD: thermally grown SiO_2
5		HTKHSHTSPPPL[3]	PD: thermally grown SiO_2
7	TBP-1	RKLPDAPGMHTW[4]	PD: NP (surface area $0.4\,m^2kg^{-3}$)
8		HKKPSKS[5]	PD: SiO_2 and TiO_2 NPs (∼50 nm)
9		TKRNNKR[5]	PD: SiO_2 and TiO_2 NPs (∼50 nm)
10		YITPYAHLRGGN[6]	PD: NP (15.0 ± 0.4 nm)
11		KSLSRHDHIHHH[6]	PD: NP (82.1 ± 3.6 nm)
12		LDHSLHS[6,7]	PD: NP (82.1 ± 3.6 nm)
13		MHRSDLMSAAVR[6]	PD: NP (450 ± 22 nm)
14		KLPGWSG[7]	PD: NP (82 ± 4 nm)
15		AFILPTG[7]	PD: NP (82 ± 4 nm)
16		RLNPPSQMDPPF[8]	PD: (100) quartz
17		QTWPPPLWFSTS[8]	PD: (100) quartz
18	QBP-1	PPPWLPYMPPWS[9]	Designed: Quartz
19		LPDWWPPPQLYH[9]	Designed: Quartz
20		SPPRLLPWLRMP[9]	Designed: Quartz

Table 1.1: Silica-binding peptide sequences along with either the substrate against- or method by which they were selected ('PD' denotes phage display and 'NP' denotes nanoparticle). Positively charged residues have been highlighted in red, negatively charged ones in blue and polar in green. Aromatic residues have been underlined. [1] Reference Kroger et al. [1999], [2] Reference Naik et al. [2002a], [3] Reference Eteshola et al. [2005], [4] Reference Sano et al. [2005], [5] Reference Chen et al. [2006], [6] Reference Patwardhan et al. [2012], [7] Reference Puddu and Perry [2012], [8] Reference Tamerler et al. [2007], [9] Reference Oren et al. [2007].

acid adsorption onto various silica surfaces have been conducted by several groups [Zimmerman et al. [2004]; Churchill et al. [2004]; Vlasova and Golovkova [2004]; Alaeddine and Nygren [1996]; Kitadai et al. [2009]; Lopes et al. [2009]]. Churchill *et al.* found that lysine bound to quartz more strongly than the other amino-acids they studied (tyrosine, alanine, glycine, glutamate and aspartate, within the pH range 2–11) [Churchill et al. [2004]], whilst Vlasova and Golovkova found that arginine bound most strongly to highly dispersed silica surfaces out of the positively charged residues, lysine, ornithine, arginine and histidine (within the pH range 2–8) [Vlasova and Golovkova [2004]]. Contrary to Churchill *et al.* and Vlasova and Golovkova, however, Zimmerman *et al.* reported that neither lysine nor dilysine adsorbed onto mesoporous and nonporous silica, unlike the hydrophobic amino-acid tryptophan; this was found to interact strongly with both surfaces (at pH 5.7) [Zimmerman et al. [2004]]. The former result corroborates the work of Alaeddine *et al.* who measured the free energy of adsorption of the amino-acids leucine and serine onto both hydrophilic and hydrophobic quartz surfaces. They found that the binding affinity was greater for the former, non-polar amino-acid at both interfaces (at pH 7.2) [Alaeddine and Nygren [1996]]. In a more recent attenuated total reflectance IR spectroscopy study, $81 \pm 5\,\%$ of the L-lysine molecules adsorbed to an amorphous silica surface were in the cationic state, with the remaining $19 \pm 5\,\%$ in the zwitterionic state (pH 7.1 - 9.8) [Kitadai et al. [2009]]. Although Kitadai *et al.* concluded that primarily the electrostatic attraction between lysine and the silica drives adsorption, the fact that at pH below 8.5 a higher proportion of adsorbed lysine is in the zwitterionic state than lysine free in solution suggests that hydrophobic interactions may also be important. Overall, the above results highlight that the exact nature of the silica surface investigated–*e.g.* amorphous *v.s.* crystalline, silanol density, silanol type (vicinal, geminal or isolated) and charge state–can have a profound effect on adsorbate affinity. In addition, it must be noted that comparison between studies at a specific silica interface (*e.g.* quartz) is hindered due to the impact of sample preparation on surface properties [Kosmulski [2002]].

As well as surface binding affinities, the orientation and geometry of selected amino-acids adsorbed onto silica has also been experimentally investigated. Employing SFG, the adsorption of 8 amino-acids (phenylalanine, leucine, glycine, lysine, arginine, cysteine and alanine) onto a model hydrophilic SiO_2 surface was probed [Holinga et al. [2011]]. For all adsorbates, C–H vibrational modes were absent from the SFG spectra. Since QCM measurements confirmed the presence of the amino-acids at the interface, this observation suggests a lack in the net orientation and ordering of the adsorbed molecules. In dry conditions, on the other hand, alanine and glycine were found to bind to amorphous silica in a specific geometry: the ammonium terminus was co-ordinated to 3 or 4 surface Si sites in each case, while the carboxylate terminus pointed away from the interface and exhibited slightly more orientational freedom [Ben Shir et al. [2010, 2012]]. However, under conditions more comparable to those used in the SFG experiments performed by Holinga *et al.* (*i.e.* ambient temperature and in the presence of water) the results of both SFG and NMR experiments were consistent; the NMR spectra of alanine and glycine suggest that surface desorption is likely to occur and that the adsorbates are orientationally disordered at the interface.

Experimental [Hayashi et al. [2009]] and bioinformatics [Oren et al. [2007]] studies have shown that peptides with identical amino-acid content but different sequences have different surface binding characteristics, underlining the role of peptide conformation on adsorption. This is exemplified by recent MD simulations in which the calculated gold-binding affinity of several peptides for gold were not simple sums of their constituent amino-acids binding energies [Verde et al. [2011]; Feng et al. [2012]]. Proline, known to restrict the conformational freedom of a peptide backbone due to its ring structure, features widely in both predominantly positively charged and charge neutral silica-binding peptides (Table 1.1). Its presence could indicate peptide stiffness is a pre-requisite to strong silica binding–the opposite of that hypothesized to be the case at the gold interface [Hnilova et al. [2008]; So et al. [2009]; Verde et al. [2011]].

1.3.2 Simulation

Like their experimental counterparts, the results of computational studies of the biomolecule-silica interface are highly dependent on the nature of the surface modeled and the conditions used–*in vacuo* or aqueous. Several groups have used DFT to calculate the binding affinities of amino-acids and their analogs to various silica surfaces *in vacuo* [Han et al. [2008]; Han and Sholl [2009, 2010]; Rimola et al. [2008, 2009]]. Although these studies are interesting, their findings cannot easily be extrapolated to shed light on the mechanisms governing peptide adsorption onto silica under aqueous conditions. As mentioned in Sections 1.2.3 and 1.2.4, the reasons for this are two-fold: first the absence of liquid water–thought to play an important role in biomolecule adsorption [Skelton et al. [2009]; Jena and Hore [2010]; Schneider and Ciacchi [2012]]; and second the potential for an amino-acid to adopt a binding geometry not possible for a residue part of a longer peptide chain. Only the first of these issues has previously been addressed at the first-principles level of theory [Costa et al. [2008]; Nonella and Seeger [2008]; Zhao et al. [2011]]. In Chapter 4 the results of first-principles simulations of ammonium and ethanoate adsorbed at the aqueous (100) α-quartz interface are presented. This work was carried out in an attempt to address *both* of the problems commonly associated with interfacial DFT simulations and applied to the aqueous quartz (the specific form of silica studied herein) interface for the first time. By simulating the adsorption of small molecules featuring the same functional groups present in peptides, rather than the whole amino-acid, it was possible to model residue side-chain adsorption specifically.

Force-field (FF) based (rather than first-principles) simulations have been used in the majority of the work carried out in this thesis. Several interfacial FFs exist for silica [Brodka and Zerda [1996]; Lopes et al. [2006]; Cygan et al. [2004]; Lorenz et al. [2008]; Hassanali and Singer [2007]; Hassanali et al. [2010]; Patwardhan et al. [2012]], although only a small proportion of these have been specifically parameterized to be biomolecule compatible [Lopes et al. [2006]; Lorenz et al. [2008]; Butenuth et al. [2012]; Patwardhan et al. [2012]]. Out of the second subset, the silica FF derived by Lopes *et al.* was specifically parameterized for charge-neutral hydroxylated quartz interfaces [Lopes et al. [2006]], while those developed by Butenuth *et al.* and Patwardhan *et al.* were designed for modeling amorphous silica [Butenuth et al. [2012]] and β-cristoballite [Patwardhan et al. [2012]] surfaces, respectively, both of which feature deprotonated silanol groups. Since the simulations reported in this thesis were performed at the aqueous quartz interface, it was appropriate to use the first FF [Lopes et al. [2006]], herein denoted 'LFF'. LFF was comprehensively parameterised following the CHARMM methodology [MacKerell et al. [1998]] and contains terms which reasonably reproduce the structure of the bulk substrate, as well as the surface, thus permitting fully flexible-solid simulations.

Due to its important role in biomolecule adsorption, interfacial water structuring can be used to benchmark a FF. Recently, the performance of three interfacial silica FFs–LFF [Lopes et al. [2006]], CLAYFF [Cygan et al. [2004]] and CWCA [Lorenz et al. [2008]]–relative to both first-principles simulation (AIMD) and experiment was assessed at the aqueous (101) [Skelton et al. [2011a]] and (100) [Skelton et al. [2011b]] α-quartz surfaces. Both studies found that CLAYFF [Cygan et al. [2004]] was better able than LFF to reproduce the degree to which water was structured at the silica interface observed in AIMD simulation. On the other hand, LFF was more apt for accurately modeling the structure of the quartz interface itself. In light of these findings, the ability of LFF and CLAYFF to model the characteristics of the two functional groups present in all peptides–NH_3^+ and -COO^-– on binding was assessed in Reference Wright and Walsh [2012b]. The FF results were benchmarked against CPMD simulations at the aqueous (100) α-quartz interface.

(It must be noted however, that the specific combination of LJ mixing rules and water model used in the parameterisation of CLAYFF is not also employed by any existing biomolecular FF calling into question its use in biointerfacial simulations.) Taken together with the results from previous MD simulations of the aqueous quartz interface using LFF [Lopes et al. [2006]; Notman and Walsh [2009]], all these findings suggest that LFF is at present the best FF with which to model peptide-quartz adsorption. Work by Latour and *co-workers* to extend their interfacial, dual parameter set FF [Snyder et al. [2012]] to the aqueous quartz interface may lead to a more accurate description of the system in the future.

Unlike other aqueous bio-interfaces, relatively few reports of silica adsorption FF-based simulations exist in the literature. The most recent modeled the adsorption of the N-terminus hexapeptide of TBP-1 [Schneider and Ciacchi [2012]], and sequences 11 (wild-type and mutated) and 13 [Patwardhan et al. [2012]] in Table 1.1 to an oxidized silicon and cristoballite surfaces, respectively. In both cases the interface featured unprotonated silanols, and hence carried an overall negative charge. As a consequence, the predominantly positively-charged peptides studied were observed to bind to the surface by electrostatics. In contrast, Oren *et al.* found that the hydrophobic residues of QBP-1 and sequence 19 in Table 1.1 bound strongly to the charge-neutral hydroxylated (100) α-quartz surface when they performed simulations at this interface [Oren et al. [2010]]. The latter results are consistent with the calculated free energies of adsorption of both amino-acid analogs [Notman and Walsh [2009]; Wright and Walsh [2012a]] (see Chapter 3).

In this thesis, two aspects of the aqueous peptide-quartz interface have been addressed by MD simulation for the first time: the potential for individual residues to energetically or spatially discriminate between different crystallographic planes of quartz on adsorption (Chapter 3); and the selectivity of QBP-1 to bind to quartz over gold (Chapter 7).

1.3.3 Crystallographic Surfaces

Since experimental evidence suggests that the degree of surface deprotonation is minimal at pH 7 [Eggleston and Jordan [1998]], fully hydroxylated, charge-neutral quartz surfaces have been modeled throughout this work. Snapshots of the three crystallographic planes studied–(100), (011) and (001)–are shown in Figure 1.2. The (001) hydroxylated surface has the highest silanol density (0.765×10^{-3} OH nm^{-2}) of the three surfaces investigated and can be considered to be essentially topologically flat. In contrast, the (100) surface has larger interstitial sites into which water penetration has been observed in previous MD simulations [Lopes et al. [2006];

Notman and Walsh [2009]]. Both the (001) and (100) (α termination) surfaces feature geminal silanols. The third surface, (011), is topologically stepped with surface silicon atoms being present at two distinct heights. It possesses vicinal silanols and features the lowest silanol density of the three (0.471×10^{-3} OH nm^{-2}).

Adsorption at all three aqueous α-quartz interfaces–(100), (011) and (001)–has been investigated in Chapter 3, while the (100) surface only was modeled in Chapters 4, 5, 7.

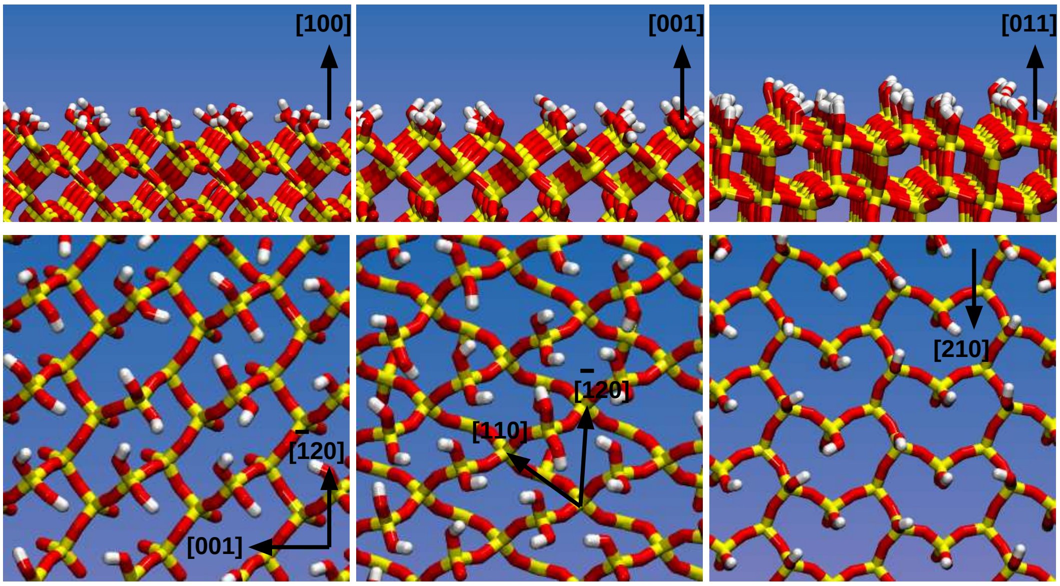


Figure 1.2: Plan and side-on views of the (100), (001) and (011) α-Quartz surfaces. Silicon atoms shown in yellow, oxygen atoms in red and hydrogen atoms in white.

1.4 Aqueous Gold Interface

The full potential of NP in future nanotechnological applications can only be reached if their synthesis and subsequent organization is carefully controlled. The promise of AuNPs in particular, due to their unique chemical and physical properties, was sufficiently great that some of the first biocombinatorial peptides to be identified were selected against gold surfaces [Brown [1997]]. Since then a number of other gold-binding sequences have been found (Table 1.2) and the aqueous biomolecule-gold interface is one of the most studied by computational modeling [Heinz et al. [2009, 2010]; Feng et al. [2011]; Verde et al. [2009, 2011]; Yu et al. [2010, 2012]; Feng et al. [2012]; Hoefling et al. [2010a,b, 2011]; di Felice and Corni [2011]; Rosa et al. [2012]; Brancolini et al. [2012]; Causa et al. [2013]] out of all bio-interfacial systems. In this thesis, a new FF, specifically parameterized for peptide adsorption at the aqueous Au(111) and Au(100) (native and reconstructed) interfaces [Wright et al. [2013b,c]] is described in Chapter 6, while the material selectivity of QBP-1- and the facet selectivity of AuBP-1- [Hnilova et al. [2008]] adsorption onto gold are investigated in Chapters 7 and 8, respectively.

Within this section a review of significant previous experimental (Section 1.4.1) and computational (Section 1.4.2) work is given, before the crystallographic planes of gold investigated herein are discussed (Section 1.4.3).

1.4.1 Experiment

Unlike at the aqueous quartz interface, both physi- and chemi-sorption of peptides on gold is possible; the latter mediated by the thiol functional group present in the side-chain of the amino-acid cysteine. Thus, gold-binding peptides fall into two groups: those which contain cysteine and those which do not. In this thesis it is the mechanisms governing the adsorption of peptides in the second group–those which only phyisorb onto gold–which has been studied. This subset of peptides offer a number of advantages in the realms of biomimetic material synthesis and nanomedicine over those which chemisorb: chiefly, the potential for selectivity (either between different crystallographic planes of gold or between noble metal substrates) and reversible binding.

Not only do all of the sequences listed in Table 1.2 have high affinity for gold, but many of them are also able to reduce $HAuCl_4$ to form biomimetically synthesized AuNPs [Brown [1997]; Slocik et al. [2005b]; Huang et al. [2005]; Kim et al. [2010]]. However, like quartz-binders there is little overall consensus in the primary sequences of gold binding peptides. Some general trends can be observed, though. Specifically, hydroxyl containing residues, serine, threonine and tyrosine, feature in most sequences, while a smaller subset of peptides contain motifs of two or more consecutive positively charged residues. The latter observation, together with the relative prevalence of glycine and proline, disorder promoting residues [Dyson and Wright [2005]; Uversky and Dunker [2010]], is consistent with the hypothesis that, in general, material-binding peptides belong to the intrinsically disordered class of proteins (IDPs) [Evans et al. [2008]]. In fact, AuBP-1, AuBP-2 and GBP-1 have all been shown experimentally to adopt labile structures in solution [Hnilova et al. [2008]; Kulp et al. [2004]], supporting this conjecture by Evans *et al.*.

		sequence	substrate
1	GBP-1	MHGKTQATSGTIQS[1]	PD: Au not specified
2	A3	AYSSGAPPMPPF[2]	PD: acid-etched Ag particles
3		KHKHWHW[3]	designed
4		RMRMKMK[3]	designed
5		VSGSSPDS[4]	PD: thin gold films
6		LKAHLPPSRLPS[5]	PD:5 nm colloidal gold suspension
7	AuBP-1	WAGAKRLVLRRE[6]	PD: polycrystalline Au films
8	AuBP-2	WALRRSIRRQSY[6]	PD: polycrystalline Au films
9		TGTSVLIATPYV[7]	PD: metallic Au powder

Table 1.2: Gold-binding peptide sequences along with either the substrate against- or method by which they were selected ('PD' denotes phage display). Positively charged residues have been highlighted in red, negatively charged ones in blue and polar in green. Aromatic residues have been underlined. [1] Reference Brown [1997], [2] Reference Naik et al. [2002b], [3] Reference Peelle et al. [2005], [4] Reference Huang et al. [2005], [5] Reference Nam et al. [2006], [6] Reference Hnilova et al. [2008], [7] Reference Kim et al. [2010].

In contrast to silica, to date, only one experimental report of the binding propensity of individual amino-acids at the aqueous gold interface exists [Li et al. [2001]]. In this study Lipkowski and *co-workers* investigated the characteristics of phenylalanine adsorbed to a Au(111) electrode using electrochemical measurements and Fourier Transform IR spectroscopy. Host-guest peptide systems have, however, been used twice to predict the intrinsic contribution of each residue within a gold-binding peptide to its overall surface affinity. In the first study, cell surface-coverage assays of yeast cells displaying the interdigitated XHXHXHX heptapeptide suggested that the positively charged and aromatic residues His, Trp, Lys and Arg all promoted strong binding, while The, Pro and Glu have the weakest affinities [Peelle et al. [2005]]. These results are qualitatively in agreement with Cohavi *et al.* who used SPR to determine the amount of BLIP-fused (TEM1-β-lactamase inhibitor protein) and free XXX homotripeptides adsorbed at the aqueous interface of both AuNP and gold films [Cohavi et al. [2011]]. Their results again indicate that positively-charged and aromatic amino-acids–His, Trp, Lys and Tyr–are strong binding residues, while small or negatively-charged amino-acids–Glu, Asp, Ser and Ala–directly contribute negligibly to gold adsorption. (The homotripeptides PPP was not investigated.) In both studies, the amino-acid methionine–which features a methyl sulfide functional group and thus has the potential to chemisorb to gold–was found to promote strong gold adsorption. As mentioned previously however, the results of any host-guest peptide binding assay must be interpreted with care since each guest residue may influence the conformation of the entire host-guest peptide by a different amount.

Bioinformatics provides an alternative way to predict which residues are important–either in terms of their intrinsic surface affinity or conformational influence–for a peptide to adsorb strongly to a surface. Hnilova *et al.* quantified residues enriched in the strong and weak gold-binding sequences identified in their phage display experiments [Hnilova et al. [2008]]. Broadly consistent with host-guest studies, Arg, Trp and Tyr were all over-expressed in peptides with high gold affinity (relative to the peptide library from which they were selected) while Lys, Pro and The were under-expressed. (Arg was also over-expressed in weak binding peptides however.) Taken together with the fact that tyrosine has been shown to mediate peptide-Au(111) reduction [Slocik et al. [2005a], the experimental evidence suggests that: 1) the positively charged and aromatic residues in the gold binding peptides in Table 1.2 directly interact with the gold surface; while 2) serine and threonine perhaps play an important conformational role in these sequences.

1.4.2 Simulation

Great experimental interest in biomolecule-gold interfacial systems is mirrored by a large number of computational studies. In this thesis the *aqueous peptide*-gold interface, specifically, has been studied using FF-based simulations. At present two FFs are widely used in the literature for such simulations: CHARMM-METAL [Feng et al. [2011]] (also compatible with CVFF (CVFF-METAL [Pandey et al. [2009]]) and a number of other biomolecular FFs) [Heinz et al. [2008]], parameterized to reproduce bulk metal density and surface tension; and GolP [Iori et al. [2009]; Iori and Corni [2008]] (compatible with OPLS-AA [Jorgensen et al. [1996]]), specifically targeted to protein-surface interactions. The latter also includes terms to describe the dynamic polarization of gold atoms induced by biomolecule adsorption. Covalent interactions with gold, such as the chemisorption of thiols, are not within the scope of most atomistic MD approaches (see Chapter 2) and are not modeled by either FF.

The adsorption of all 20 amino-acids onto the Au(111) surface under aqueous conditions has been studied using both CHARMM-METAL and GolP [Hoefling et al. [2010a]; Feng et al. [2011]], providing a metric against which the two potentials can be assessed. However, the predicted binding trends are in qualitative agreement: the free energy of adsorption for capped amino-acids were found to be aromatic < sulphur < positively-charged < negatively-charged < aliphatic [Hoefling et al. [2010a]] when modeled by GolP, while amino-acids (uncapped, with zwitterionic termini) with large planar side-chains (Arg, Trp, Gln, Met, Tyr) adsorbed more strongly than aliphatic ones (Leu, Asp, Thr, Gly, Ala), when modeled by CHARMM-METAL [Feng et al. [2011]]. Although the relative tendencies of each amino-acid to adsorb are also in qualitative agreement with experiment in both instances, GolP possibly provides a better prediction of absolute binding strength, than CHARMM-METAL. This is at least the case for Phe where the free energy of adsorption calculated using GolP ($\Delta\mathrm{G}= -43.6\,\mathrm{kJ\,mol^{-1}}$) lies just outside of the experimental range ($-18 < \Delta\mathrm{G} < -37\,\mathrm{kJ\,mol^{-1}}$) while CHARMM-METAL overestimates surface attraction by approximately 35%.

As with the aqueous quartz interface, the properties of water adsorbed onto gold can be used as a second criterion for evaluating the performance of the two FFs. Recent first-principles simulations carried out at the aqueous Au(111) interface [Cicero et al. [2011]; Nadler and Sanz [2012]] provide good benchmarks for this purpose. However, the same weakness in their description of this interface exists for both FFs, namely that neither can reproduce the increased tendency of water molecules very close to the gold surface to act as hydrogen bond donors, relative to those in bulk solution [Cicero et al. [2011]]. In the case of GolP, a modification to the FF proposed by Cicero *et al.* [Cicero et al. [2011]] can be used to correct this deficiency; no such correction is available for CVFF/CHARMM-METAL. Moreover, for the latter FF, CVFF/CHARMM-METAL, water density is preferentially located above surface hollow sites, in sharp contrast to first-principles simulation, where atop gold atom adsorption is favoured [Cicero et al. [2011]; Nadler and Sanz [2012]]. In addition, CVFF/CHARMM-METAL has two further shortcomings in its description of the interactions between biomolecules and gold: first, it was not specifically parameterized to capture individual molecule-gold interactions and thus does not include terms to account for the weak chemisorption of the sulfur- and nitrogen-containing functional groups present in peptides; and second, the dynamic polarization of metal atoms is neglected. Recent calculations carried out *a posteri-*

ori on MD simulations employing CVFF-METAL [Heinz et al. [2010]] revealed that on surfaces such as Au(100), where the epitaxial match between a biomolecule and underlying gold atoms is poor, the energy associated with the image-charge effect could be significant, thus suggesting that it cannot be disregarded. For all of the reasons mentioned, it was therefore decided that the CVFF/CHARMM-METAL FF would not be used in any of the simulations presented herein.

Whilst several groups have contrasted the adsorption of peptides onto gold surfaces of different geometries–Au(111) and Au(100) planes [Heinz et al. [2009, 2010]; Yu et al. [2010, 2012]], AuNP of different dimensions (composed of Au(111) and Au(100) planes) [Feng et al. [2012]; Yu et al. [2010, 2012]] and stepped [Feng et al. [2012]] interfaces–by means of MD simulation, the CVFF/CHARMM-METAL FF was used in each case. Adsorption onto Au(100) was found to be weaker than that to Au(111), with discrimination between the two surfaces thought to be mediated by molecular epitaxy. This argument was based on the adsorption of heteroatoms within residue side-chains to hollow sites on the Au(111), but not the Au(100), surface in a complementary fashion due to the underlying geometry of the gold atoms. However, as mentioned previously in the case of water, the hollow-site binding predicted using CVFF/CHARMM-METAL is in contrast to first-principles calculations [Hong et al. [2009]; Wright et al. [2013b,c]; Rosa et al. [2012]], weakening this argument. Taken together with the fact that reconstructed status of the Au(100) surface has not been considered and the other fore-mentioned shortcomings of the FF, the potential for peptides to discriminate between different aqueous gold interfaces warrants further study. A biocompatible interfacial gold FF suitable for such work was, however, lacking since GolP parameters are restricted to the Au(111) surface only. As a consequence, a new FF, GolP-CHARMM FF [Wright et al. [2013b,c]] was derived (see Chapter 6 for more details).

As well as using the newly-derived FF to model the adsorption of the AuBP-1 peptide at the aqueous Au(111) and Au(100) (both native and reconstructed forms) interfaces (Chapter 8), the hypothesis that peptide flexibility is important for strong gold-binding [Hnilova et al. [2008]; So et al. [2009]; Verde et al. [2011]] has been tested by the simulation of a conformationally stiff peptide–QBP-1 [Oren et al. [2007]]–at the aqueous Au(111) interface (Chapter 7).

1.4.3 Crystallographic Surfaces

The Au(111) surface is the lowest energy crystallographic plane of gold, while spherical AuNPs are mainly composed of {111} and {100} faces [Marks [1994]]. It is well established that the planar Au(100) surface is present in its reconstructed form

(Au(100)-*hex* [Gibbs et al. [1990]; Mochrie et al. [1990]; Ocko et al. [1991]]) under the conditions relevant to biomolecule adsorption [Kolb [1996]] (aqueous environment, 300 K, pH 7). Unlike the Au(111) surface, also present in its reconstructed form, Au(111)$(22\times\sqrt{3})$ [Harten et al. [1985]], under these conditions, there are significant structural differences between the Au(100)(1×1) and Au(100)-*hex* surfaces. Namely, the reconstructed interface features an hexagonal array of atoms, while atoms in the plane of the native surface have only four-fold in-plane symmetry. As a consequence, the density of atoms in the surface layer increases by 25% in the transition from Au(100)(1×1) to Au(100)-*hex*. In addition, the reconstructed surface features atomic-scale undulations [Ocko et al. [1991]; Jacob [2007]; Shi and Stampfl [2008]; Havu et al. [2010]], whilst the native interface is atomically flat.

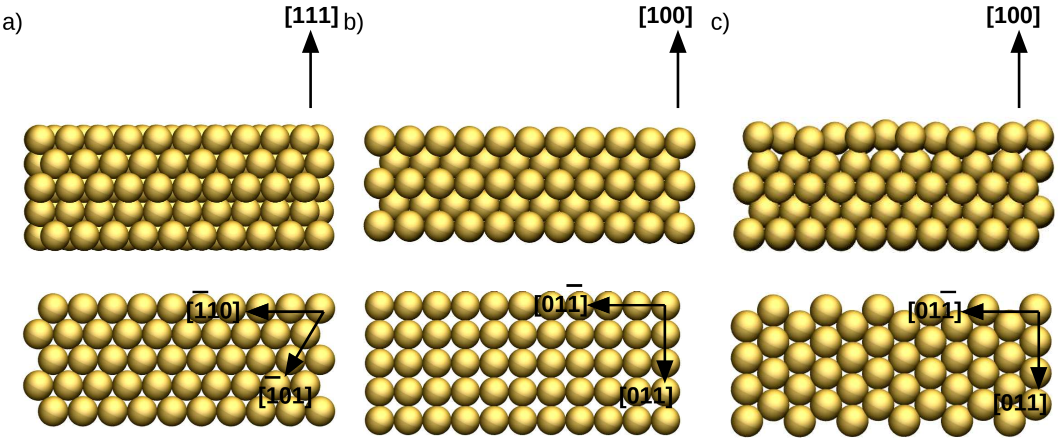


Figure 1.3: Plan and side-on views of the a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1) surfaces.

Experimentally, it has been shown that the adsorption of some small molecules, containing functional groups that are also present in biomolecules, onto Au(100) can promote [Skoluda [2004]] the Au(100)-*hex* reconstruction of planar surfaces [Gibbs et al. [1990]; Mochrie et al. [1990]; Ocko et al. [1991]] under aqueous conditions, whilst others stabilise the ideal surface [Skoluda [2011]]. The reconstructed status of AuNP facets, either in the presence or absence of adsorbates, is at present unknown [Zehner et al. [1975]; Shi and Stampfl [2008]; Li et al. [2008]; Wang and Palmer [2012]; Surrey et al. [2012]]. Thus, in the shortage of conclusive evidence, it was considered appropriate to consider both the native and reconstructed forms of the Au(100) surface when investigating the potential for peptide adsorption to be used as a means of discriminating between Au(111) and Au(100). However, the large size of the unit cell of the reconstructed surface (Au(100)(28×5) [Wang [1991]]) makes it inpractically large for computational simulation in its precise form. While some have approximated the Au(100)-*hex* surface by Au(111) [Lin and Gross [2012]]–both feature Au atoms hexagonally arranged–this approach neglects the influence of surface rumpling–a feature of the former interface only–which could play an important role in liquid water adsorption [Phan et al. [2012]]. Many groups, have used the Au(100)(5×1) approximation when studying the reconstructed Au(100) interface at the first-principles level of theory [Venkatachalam et al. [2008]; Jacob [2007]; Feng et al. [2005]; Shi and Stampfl [2008]; Sun et al. [2011]; Baber et al. [2012]]. Since their results were in good agreement with experiment, this approximation has also been used here. Therefore, the three aqueous gold interfaces modelled in this thesis are: Au(111), Au(100)(1×1) and Au(100)(5×1) (Figure 1.3).

Chapter 2

Methods

2.1 Describing the Potential Energy Landscape

The Potential Energy Landscape (PEL) is a description of the potential energy of a system as a function of the co-ordinates of its constituent particles. Stable configurations are located at minima on the PEL, whilst the saddle points between minima correspond to transition states. The minimum on the PEL with the most negative potential energy, the global minimum, is the thermodynamically most favourable state of the system; higher energy, metastable states may be observed under experimental conditions, however. Since the ultimate goal of most computational modelling studies is to either reproduce experimental and/or generate complimentary novel data, it is important that as much of the PEL is considered as possible in these works, not just the global minimum state.

The results of any modelling investigation depend critically on how the PEL is defined. Two approaches have been adopted in this thesis: first-principles and atomistic force-field, described in Sections 2.1.1 and 2.1.2, respectively.

2.1.1 First Principles Methods

Using first-principles methods, the PEL of a system is defined quantum mechanically. The basic entities–nuclei and their surrounding electrons–are modelled as wave-functions, $\phi(\mathbf{r}, s)$ (functions of position, $\mathbf{r}$, and spin state s). Unlike classical particles which are described in terms of position and momentum, the exact location of an entity described using quantum mechanics (QM) is unknown. Instead, only the *probability* of finding it in a given *region* of space at any one time can be determined. Moreover, from the Heisenberg Uncertainty Principle, the accuracy with which a particle's position and momentum can be known are inversely proportional.

The wave-function describing a system in state i, Φ_i, is, by definition, a solution of Schrödinger's equation:

$$\hat{H}\Phi_i = E_i\Phi_i \tag{2.1}$$

where $\hat{H}$ is the Hamiltonian operator (comprised of kinetic, $\hat{T}$, and potential, $\hat{U}$, energy operators) and E_i is the energy eigenvalue of state i. This eigenvalue equation must be solved for both the nuclei and electrons in a system. However, since the mass of an electron is significantly less than that of an atomic nucleus, the motions of the two can be separated. Using what is commonly referred to as the Born-Oppenheimer approximation, Schrödinger's equation for electronic states of a system as a function of fixed nuclei co-ordinates, R, simplifies to:

$$\hat{H}\Phi_i = -\frac{\hbar^2}{2m_e}\sum_j \nabla_j^2\Phi_i + \frac{1}{2}\sum_{j,kj\neq k}\frac{e^2}{4\pi\epsilon_0|r_j - r_k|}\Phi_i + v(r)\Phi_i = E(R)_i\Phi_i \tag{2.2}$$

where $\hbar$, m_e and ϵ_0 are the fundamental quantities the reduced Planck's constant (or Dirac constant), electron mass and the permittivity of a vacuum, respectively; $|r_j - r_k|$ is the distance between the jth and k^{th} electron; and $v(R)$ is the external field experienced by the electrons due to the presence of the nuclei. For all but the simplest cases–one or two electron systems–exact analytical solutions to Equation 2.2 do not exist. Instead, iterative numerical strategies based on the variational principle are adopted. This theorem states that the true ground state energy of a system will always be less than or equal to that predicted using a trial wave-function. Electronic wave-functions, comprised of linear combinations of atomic orbitals, obey Pauli's exclusion rule: namely that a wave-function of a fermion is antisymmetric under particle exchange.

The inter-electron repulsive term in Equation 2.2 is computationally expensive and challenging to calculate due to the pair-wise summation and indistinguishability of the electrons involved. In 1964, however, Hohenberg and Kohn [Hohenberg and Kohn [1964]] made a significant breakthrough by proving that the ground state energy, E_0, and electronic properties of a system can be uniquely described by its electron density, $n(\mathbf{r})$:

$$E[n(\mathbf{r})] = T[n\mathbf{r})] + U_{ee}[n(\mathbf{r})] + \int n(\mathbf{r})v(\mathbf{r})\mathrm{d}\mathbf{r} \tag{2.3}$$

where, as before, T and U_{ee} are functionals of the kinetic and inter-electron interaction potential energies and $v(\mathbf{r})$ is the external field, respectively. Electron density,

$n(\mathbf{r})$, and the electronic wave-function, Φ, of a system are related by:

$$\delta n(\mathbf{r}) = \int_{\mathbf{r}}^{\mathbf{r}+\delta\mathbf{r}} \Phi(\mathbf{r}', s)^* \Phi(\mathbf{r}', s) \, \mathrm{d}\mathbf{r}' \tag{2.4}$$

Equation 2.3 is the fundamental basis behind Density Functional Theory (DFT), with the electron density distribution for the ground state of a system, $n_0(\mathbf{r})$, being that which minimises E to E_0. Although precise by definition, approximations have to be made in practice in order to obtain solutions since the functional forms of T and U_{ee} are unknown. First proposed by Kohn and Sham the most fundamental simplification is to describe the system in terms of non-interacting electrons [Kohn and Sham [1965]]. In this case both the kinetic and potential energy functionals are defined and exact solutions, known as Kohn-Sham orbitals, exist. However, for real systems a correction, known as the 'exchange-correlation' functional, is needed in order to recover Equation 2.3. This term, the functional form of which is undefined, encompasses the contributions of the exchange (the energy associated with the indistinguishability of any two electrons) and correlation (the energy associated with instantaneous repulsion between electrons) energies to the total energy of the system; both normally small perturbations to the total energy.

Exchange correlation functionals can be categorised according to their form. Currently three different classes are in widespread use: Local Density Approximation (LDA), Generalised-Gradient Approximation (GGA) and hybrid. The first is a functional of $n(\mathbf{r})$ only, while the second is a functional of both $n(\mathbf{r})$ and its gradient. Hybrid functionals combine different methods, such as Hartree-Fock exchange with DFT. One of the major shortfalls of all three, however, is their description of the dispersion interactions within a system. (These interactions involve the non-local response of an electron in one region of space due to a remote instantaneous fluctuation in $n(\mathbf{r})$ and are characterised by a decay in the inter-atomic interaction energy which is approximately proportional to $-1/r^6$, where r is the distance between two atoms.) For instance, using PBE (a GGA-type exchange-correlation functional), π stacking of two benzene rings is predicted to be unfavourable [Rosa et al. [2012]].

In recent years new DFT functionals which better account for dispersion have been derived. These functionals can be broadly classed into 3 categories (See Reference [Klimes and Michaelides [2012]] for a recent review). Briefly:

1. **DFT-D methods:** In this class of DFT functionals, dispersion is reproduced by the additive correction:

$$E_{disp} = -\sum_{A,B} \frac{C_6^{AB}}{r_{AB}^6} \tag{2.5}$$

where the exact form of C_6^{AB}, the dispersion coefficient, differs from method to method. In the simplest case, DFT-D, C_6^{AB} are tabulated and are invariant to the particular environment or chemical state of the atoms in question. More complex schemes aiming to reproduce these effects have been derived. For example, using the DFT-D3 method, the C_6 value for each specific atom within a system is dependent on its co-ordination number [Grimme et al. [2010]], while for the vdW(TS) functional reference polarisabilities are used alongside C_6 coefficients in order to distribute the electron density of a molecule between its constituent atoms depending on their effective volumes [Tkatchenko and Scheffler [2009]]. Although the accuracy of some DFT-D functionals, relative to experiment, is reasonably high for specific test sets of molecules, the empirical nature of the dispersion correction limits their transferability to systems outside of those used in the parameterisation procedure.

2. **Long-range methods:** Unlike, DFT-D, dispersion interactions are calculated from the electron density directly, without the need for external parameters, by this set of functionals. All are variants of the vdW-DF method [Dion et al. [2004]; Thonhauser et al. [2007]]. Within this scheme the exchange-correlation functional is broken down into:

$$E_{XC}[n(\mathbf{r})] = E_X^{GGA}[n(\mathbf{r})] + E_C^{l}[n(\mathbf{r})] + E_C^{nl}[n(\mathbf{r})] \tag{2.6}$$

where E_X^{GGA}, the exchange functional, is modelled by the GGA; E_C^l, the local contribution to the correlation energy, is modelled by the LDA; and E_C^{nl}, the non-local contribution to the correlation energy, is estimated from $n(\mathbf{r})$ iteratively. In the case of vdW-DF, revPBE [Hammer et al. [1999]] is employed as the GGA. The strength of dispersion interactions, relative to experimental and high order *ab initio* calculations, in a range of different systems can be accurately reproduced by vdW-DF [Thonhauser et al. [2007]; Roman-Perez and Soler [2009]; Langreth et al. [2009]; Toyoda et al. [2009]; Li et al. [2012]; Wellendorff et al. [2010]; Rosa et al. [2012]]. There are however, two main deficiencies of this functional which are known: overestimation of equilibrium separations and underestimation of hydrogen bond strength. Improved variants of vdW-DF attempt to address these issues by making changes to E_X^{GGA}

and E_C^{nl} [Lee et al. [2010]] and/or using an alternative the GGA [Klimes et al. [2011]].

3. **Higher-order methods:** Using either DFT-D or long-range functionals, dispersion is treated as a pair-wise interaction, with the exact nature of the medium separating the two atoms neglected. Methods which go beyond this approach are currently being developed. Examples include those based on the atom-centred Axilrod-Teller-Muto functional which incorporates three-, as well as two-, bodied dispersion interactions [Axilrod and Teller [1943]; Muto [1943]]; and double hybrid functionals, in which two corrections are added to the GGA: one from the Hartree-Fock exchange energy and the other from the MP2 dispersion energy [Schwabe and Grimme [2007]].

In addition to the fundamental approximation of DFT, the exchange-correlation functional, further simplifications to Equation 2.3 must be made in practice. First, wave-functions describing $n(\mathbf{r})$ are constructed as a linear combination of functions from a basis set. The form of these functions which represent the Kohn-Sham atomic orbitals is chosen for computational efficiency, as well as physical and chemical accuracy. Atom-centred basis sets require only a small number of basis functions, such as Gaussians, to define an orbital and are intrinsically suited to isolated molecules. Plane-wave (PW) basis sets, on the other hand, are more apt for periodic calculations and thus have been used in Chapter 6 to calculate the interaction energy between small molecules and planar gold surfaces. Another advantage of using PW-DFT is that the accuracy of the calculation can be systematically improved by increasing the maximum wave vector, normally quoted as the PW cut-off energy.

The wave-functions of core electrons can only accurately be reproduced by PW-DFT using a very large number of wave vectors due to the steep change in potential energy close to the nucleus. However, the contribution of these electrons to binding is small in comparison to those in valence orbitals. Thus to reduce the computational expense of a calculation, core electrons are commonly modelled by pseudopotentials: functions which are smooth within a core region but reproduce contribution of core electrons to the overall electron density of an atom for distances outside this limit.

For periodic systems, integration in reciprocal space only needs to be carried out over a selected number of points, the Brillouin-Zone: the smallest volume with the same symmetry as the system as a whole which can be periodically repeated to fill wave-vector reciprocal space. In practice, this is normally done on a 3D grid–the Monkhorst-Pack grid. The number of grid points required for convergence varies inversely with cell dimension. Finer grids are needed for metallic systems, in order to reproduce the Fermi surface. A further consideration for periodic calculations is the size of the system in real space. A compromise is sought between computational expense (which increases with cell volume) and accuracy–increased by minimising interactions between periodic images of the cell.

2.1.2 Atomistic Force-Fields

A force-field (FF) is a set of empirical equations describing the forces acting between particles within a system. In an atomistic FF, the fundamental particles are atoms, normally represented by point masses of fixed charge. All FF-based simulations reported in this thesis (unless otherwise stated) have employed FFs compatible with CHARMM [MacKerell et al. [1998, 2004]; Piana et al. [2011]]. In CHARMM inter-atomic interactions are divided into two groups: bonded, V^{B}, involving atoms connected by 3 or fewer covalent bonds; and non-bonded, V^{NB}. Intra-molecular non-bonded interactions between atoms separated by 2 or fewer bonds are excluded.

The functional form of intra-molecular bonded interactions involving 2, 3 and 4 atoms (V^{bond}, V^{angle} and $\mathrm{V}^{dihedral}$, respectively) are given below in Equations 2.7, 2.8 and 2.9. In CHARMM a simple harmonic potential is used to model bond stretching:

$$V_{bonds}(r_{ij}) = \frac{1}{2}k^{b}_{ij}(r_{ij} - b_{ij})^2 \tag{2.7}$$

where r_{ij} is the distance between two atoms, k^b the bond stretching force constant and b_{ij} the equilibrium bond length. Whilst computationally efficient, representing covalent bonds in this form excludes the possibility of chemical reactions (*i.e.* the dynamic fission and reformation of a chemical bond).

Three-bodied bonded interactions, angle bending, are modelled by the Urey-Bradley potential which comprises 3-bodied and 2-bodied harmonic terms:

$$V^{angle}(\theta_{ijk}) = \frac{1}{2}k^{\theta}_{ijk}(\theta_{ijk} - \theta^{0}_{ijk})^2 + \frac{1}{2}k^{UB}_{ijk}(r_{ik} - r^{0}_{ik})^2 \tag{2.8}$$

where θ_{ijk} is the angle between the three atoms, k^{θ} the angle bending force constant and θ^0 the equilibrium angle, while k^{UB} and r^0_{ik} are the force constant and equilibrium bond length for the fictitious 1-3 bond, r_{ik}, respectively.

Dihedrals are divided into two groups- proper and improper. While the four atoms i, j, k and l of a proper dihedral are all connected by covalent bonds, the same is not true of improper dihedrals; the latter are used to model the out-of-plane motion of planar groups such as benzene rings and carbonyls. Proper dihedrals can be represented by more than one periodic cosine function of the form:

$$V^{dihedral}(\phi_{ijkl}) = k^{\phi}_{ijkl}(1 + cos(n\phi_{ijkl} - \phi^{0}_{ijkl})) \tag{2.9}$$

where k^{ϕ}, ϕ^0 and n are the dihedral torsion force constant, equilibrium angle and multiplicity, respectively, of the dihedral ϕ_{ijkl}. Improper dihedrals, on the other hand, are modelled by the simple harmonic:

$$V^{improper}(\eta_{ijkl}) = \frac{1}{2}k^{\eta}(\eta_{ijkl} - \eta^{0}_{ijlk})^2 \tag{2.10}$$

where k^{η} and η^0 are the force constant and equilibrium angle for the improper dihedral η_{ijkl}, respectively.

Non-bonded interactions comprise of two components- van der Waals and Coulomb electrostatics. The former, van der Waals, are modelled by the 12-6 Lennard Jones (LJ) potential, V^{LJ} in CHARMM:

$$V^{LJ}(r_{ij}) = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \tag{2.11}$$

where ε_{ij} and σ_{ij} are the van der Waals interaction energy and spatial extent of atoms i and j, respectively. For CHARMM, ε_{ij} and σ_{ij} are obtained using the Lorentz-Borthelot mixing rules [Allen and Tildesley [1987]]. Specifically, σ_{ij} is the arithmetic mean of σ_{ii} and σ_{jj} while ε_{ij} is the geometric mean of ε_{ii} and ε_{jj}.

The Coulombic electrostatic interaction, V^{Coulomb} between any pair of atoms is reproduced by the function:

$$V^{Coulomb}(r_{ij}) = C\frac{q_i q_j}{r_{ij}} \tag{2.12}$$

where q_i and q_j are the partial charge on atoms i and j and C is the constant $1/4\pi\epsilon$ (ϵ is the electrical permittivity of the medium). Unlike bonded interactions, non-bonded ones must be calculated over every pair of atoms within a system and hence constitute significantly to the computational expense of a FF-based simulation. It is hence advantageous to truncate both interactions using a spherical cut-off at some radius, r_c, for which V$^{NB}(r_c) \to 0$ (NB= LJ or $Coulomb$). In addition to being computationally efficient, by setting $r_c < L$, where L is the shortest box vector, artifacts arising from the finite size and periodicity of the system can be minimised. While V^{LJ} is convergent for $r_{ij} \to \infty$, the same is not true of V^{Coulomb}. The sum of the partial charges on all atoms within a cut-off distance of atom i will change from time-frame to time-frame as atoms move in and out of the cut-off sphere. This is demonstrated pictorially in Figure 2.1.

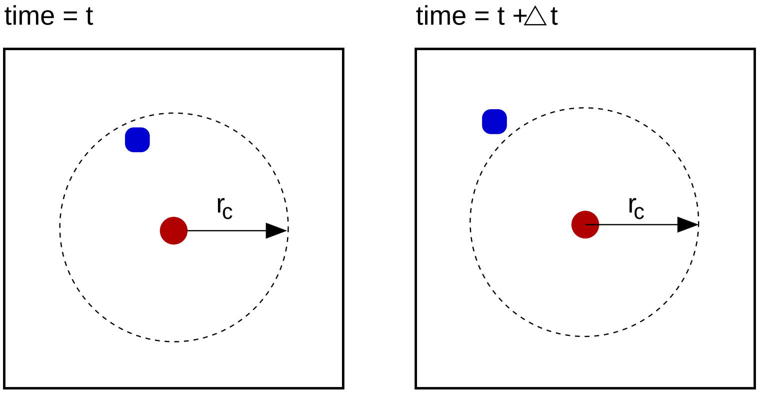


Figure 2.1: Schematic diagram to depict the movement of a second partial charge (blue) out of a sphere of radius r_c surrounding the first charge (red) from time t to time $t + \Delta t$.

However, Ewald summation of electrostatic terms, which is convergent, was used in all simulations performed in this thesis. Briefly using this method, for every point charge, a Gaussian of opposite charge is added; both point and disperse charges are centred at the same position. The sum of the two is convergent in real space. To retain the original distribution, a second Gaussian–a compensatory charge–is added. The sum of the compensatory charges can be shown to be convergent in reciprocal space [Ewald [1921]]. A third and final contribution to an Ewald sum is a correction to remove the artificial interaction between point and compensatory charges. The exact form of Ewald summation used here was Particle Mesh Ewald (PME) [Darden et al. [1993]]–a computationally efficient way in which to implement Ewald summation in reciprocal space where compensatory Gaussian charge clouds are discretised on a grid.

Simulations using any of the standard biomolecular FFs are subject to two main limitations: the inability to capture chemical reactions and atomic polarisation. ReaxFF [van Duin et al. [2001]] and AMOEBA [Ren and Ponder [2003]] are examples of specialist FFs which address these issues. (Polarisable versions of CHARMM [Anisimov et al. [2005]; Patel and Brooks [2006]; Xie et al. [2007]] also exist.) In ReaxFF [van Duin et al. [2001]], *all* pairs of atoms are subject to non-bonded interactions, described by Coulombic (Equation 2.12) and Morse potentials. Inter-atomic distance is used to define the order of a bond and it is bond order which is employed to determine the extent of corrections required to bond, angle and dihedral terms in order to reproduce the physical properties of a molecule. In the AMOEBA FF atomic polarisability is captured by a combination of permanent mono-, di- and quadra-poles added to each atom and polarisable dipoles located in specific positions. Both fixed and induced multipoles surrounding a polarisable site contribute to the strength of the electric field experienced at that location. Thus, induced dipoles must be determined iteratively. Both ReaxFF and AMOEBA are not commonly employed in large scale molecular simulations at present due to the extra computational expense associated with their implementation [Liang and Walsh [2006]].

2.2 Molecular Dynamics

In a Molecular Dynamics (MD) simulation the position and momenta of all constituent particles are propagated in time. Only classical MD simulations, in which trajectories are generated by integrating Newton's equation's of motion (Section 2.2.1), are reported in this work. The forces acting within each system here were either described by atomistic FFs (Section 2.1.2) or a DFT exchange-correlation functional (Section 2.1.1). MD is one of the most widely used computational simulation techniques. Its scope arises from the ergodicity principle. Namely, that the time-averaged and ensemble-averaged properties of a system are equal. Thus, macro- and micro-scopic observables, which can be defined using statistical mechanics in terms of the position and momenta of the microscopic components of a system, can be calculated.
There are three main components to performing an MD simulation:

1. System setup: encompassing the initial choice of co-ordinates and description of PEL.
2. Generation of the trajectory.
3. Analysis.

The steps taken to setup a simulation are system dependent but typically include energy minimisation–to remove unfavourable steric contacts–and multiple initial system geometries–to aid conformational sampling of the PEL. The two most commonly used ensembles for MD simulation–canonical (constant number, volume and temperature, NVT) and isothermal-isobaric (constant number, pressure and temperature, NPT)–have been used in the simulations reported hereafter; the algorithms used to implement them, along with those required for propagating a trajectory are described below in Sections 2.2.1, 2.2.2 and 2.2.3. Analysis common to a number of the studies detailed in Chapters 3 to 9 is outlined in Section 2.5.

2.2.1 Time Evolution

It is not possible analytically to solve Newton's equations of motion for all but the most simple systems. Instead, time-reversible numerical algorithms based on two Taylor expansions of the positions of each particle (**r**)–one forward and one backward in time–are used to carry out the integration. Specifically, the leapfrog algorithm [Hockney et al. [1974]] was used in all simulations reported in this thesis. Using this method, the integration error is on the order of Δt^4, where Δt is the time-step of the simulation. Velocities and positions are updated sequentially according to:

$$v(t + \frac{1}{2}\Delta t) = v(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m}F(t) \tag{2.13}$$

$$r(t + \Delta t) = r(t) + \Delta t v(t + \frac{1}{2}\Delta t) \tag{2.14}$$

where F(t) is the force acting on a particle at time t, determined from knowledge of the positions of all the other particles within the system at the same point in time. For numerical stability the time-step, Δt must be shorter than the period of fastest vibration within a system.

2.2.2 Thermostats

In order to simulate a system in either the NVT or NPT ensemble, temperature must be maintained using a heat bath. The principles behind three different thermostats–Berendsen, Anderson and Nosé Hoover–are discussed below; only the latter has been implemented in this work.

Berendsen thermostat

Using the simplest thermostat to implement, Berendsen, the system is weakly coupled to an external heat bath at the desired temperature, T_0, by first order kinetics:

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{T_0 - T}{\tau} \tag{2.15}$$

where deviations in temperature, T, from T_0 decay exponentially with time constant τ. Fluctuations in the kinetic energy of the system are suppressed meaning that the Berendsen thermostat does not rigorously generate the correct statistical ensemble.

Andersen thermostat

Unlike the Berendsen thermostat, simulations employing either the Andersen or Nosé Hoover sample the correct Boltzmann weighted ensemble of states. Maintenance of temperature is achieved in the former by randomly selecting particles at each time-step to undergo a collision. The probability with which a particle is selected governs the strength of coupling between the system and the heat bath. Particles selected for collision are assigned new velocities chosen from the Maxwell Boltzmann distribution at T_0. All remaining particles are unaffected.

Nosé Hoover

Opposed to the stochastic nature of the Andersen thermostat, the Nosé Hoover thermostat [Nosé [1984]; Hoover [1985]] is a deterministic method for maintaining temperature in a MD simulation. The external heat bath, at the desired temperature, T_0, is represented by a frictional entity, b. A Lagrangian restraint is added to the equation of motion of the system:

$$\frac{\mathrm{d}^2 r_i}{\mathrm{d}t^2} = \frac{F_i}{m_i} - \frac{f_b}{Q}\frac{\mathrm{d}r_i}{\mathrm{d}t} \tag{2.16}$$

where F_i is the force acting on the ith particle, of mass m_i and position r_i; Q is a constant used to describe the strength of the coupling between the system and a heat bath; and f_b is the momentum of the frictional parameter of the reservoir, described by:

$$\frac{\mathrm{d}f_b}{\mathrm{d}t} = (T - T_0) \tag{2.17}$$

where T is the instantaneous temp of system. Fluctuations in temperature are subject to oscillatory relaxation using the Nosé Hoover thermostat. The constant Q, which has dimension of mass, is related to the period of oscillation τ_T by:

$$\tau_T = \sqrt{\frac{4Q\pi^2}{T_0}} \tag{2.18}$$

2.2.3 Barostats

Algorithms analogous to those described above for thermostats exist for maintaining the pressure in a simulation by dynamically scaling the box dimensions. The Parrinello-Rahman barostat [Parrinello and Rahman [1981]; Nosé and Klein [1983]], which follows the same principles as the Nosé Hoover thermostat, has been used in all isobaric simulations carried out here. Specifically, the NPT ensemble was implemented by employing both Nosé Hoover thermostat and Parrinello-Rahman barostat.

2.2.4 First-Principles MD

All the concepts of MD discussed above are independent of the method used to calculate the forces acting on a particle in the system. In the case of 'first-principles' MD (by which I refer to the method used to calculate forces, rather than the *evolution* of time through a trajectory), two schemes exist: Born-Oppenheimer MD (BOMD) and Car Parrinello MP (CPMD).

The first, BOMD, is a straight combination of first-principles (normally DFT) calculations to identify the ground state wave-function (Section 2.1.1) and integration of Newton's classical equations of motion to evolve the simulation (Section 2.2.1). The forces, required in the second step, can be derived from the wave-function, ϕ, using Hellman-Feynman theorem which states that the derivative of the energy with respect to position (*i.e.* force, F) is equal to the expectation of the derivative of the Hamiltonian, $\hat{H}$:

$$\mathbf{F}(\mathbf{r}) = -\int \phi^*(\mathbf{r}')\frac{\delta \hat{H}}{\delta \mathbf{r}}\phi(\mathbf{r}')\mathrm{d}\mathbf{r}' \tag{2.19}$$

This equation only strictly holds if ϕ is an *exact* solution to Schrödinger's equation. Using variational principle, wave-functions closer to the exact solution can be determined iteratively by reducing the tolerance in the ground state energy. This is done at the expense of increased computation. Hence a compromise must be sought between accuracy and efficiency. It is also noted, that in the case of DFT, the *exact* form of the exchange-correlation functional is unknown meaning that the principle of variation no longer holds. Thus a trial wave function will not necessarily provide an upper estimate of the ground state energy.

A second approach to first-principles MD simulation, CPMD, was first formulated by Car and Parrinello in 1985 [Car and Parrinello [1985]] in an attempt to remove the need for expensive wave-function optimisation calculations at every time-step. In this scheme, electronic wave-functions, $\{\phi_i\}$, as well as atomic nuclei, $\mathbf{R}^N$, are treated as time-dependent properties of the system. The conserved energy is thus:

$$E_{cons} = \sum_i \frac{1}{2}\mu \int |\phi_i|^2 \mathrm{d}\mathbf{r} + \sum_I M_I \dot{\mathbf{R}}_I^2 + E[\{\phi_i\}, \mathbf{R}^N] \tag{2.20}$$

where μ is the fictitious 'mass' of an electron, M_I is the real mass of the Ith nucleus and E is the ground state energy of the system. If a very low temperature is used to evolve ϕ then the electronic energy can be kept close to the ground state. Thus, a simulation can be performed with only a single wave-function optimisation calculation: at the first step. The motion of electrons and nuclei are decoupled to prevent thermal transport between the two by careful choice of μ. Very small fictitious electron masses prevent the electronic system from deviating from the Born-Oppenheimer surface at the computational cost of reducing the integration time-step.

CPMD simulations of ammonium and ethanoate at the aqueous (100) α-Quartz interface are reported in Chapter 4. Traditionally, due to reduced frequency with which the wave-function must be optimised during a simulation, CPMD allows longer time scales to be reached (and for simulation of larger systems) than BOMD, for a given computational resource. Today, however, the computational efficiency of the two methods is more code-dependent: improvements in algorithms for wave-function optimisation can make BOMD competitive with CPMD since time-steps an order of magnitude smaller are needed in the latter.

2.3 Replica Exchange Methods

As mentioned in Section 1.2.4, one of the biggest challenges facing atomistic biointerfacial simulation is adequate exploration of phase space. Multiple minima exist in the PEL of these complex systems and, depending on their depth and the height of the barriers between them, a simulation may remain kinetically trapped in a single state within the time-scales currently accessible to molecular simulation. In recent years a number of advanced sampling techniques based on Replica Exchange Molecular Dynamics (REMD) [Sugita and Okamoto [1999]] have been developed for aqueous peptide systems in an attempt to overcome these problems [Sugita et al. [2000]; Liu et al. [2005]; Li et al. [2007, 2011]; Li and Latour [2011]; Rick [2007]; Lee and Rick [2009]; Ballard and Jarzynski [2009, 2012]; Moors et al. [2011]; Wang et al. [2011]; Terakawa et al. [2011]]. The methods fall into two broad categories: temperature and Hamiltonian REMD (T-REMD, Section 2.3.1, and H-REMD, Section 2.3.2, respectively). The principles behind both are the same: namely, a number of identical copies of the system (replicas) are simulated in parallel with exchanges being attempted between replicas at regular intervals. By incrementally scaling the temperature or Hamiltonian employed in each ensemble across replica space, the system is able to escape from deep potential energy minima. To maintain detailed balance (*i.e.* ensuring that $\mathrm{P}_{ij} = \mathrm{P}_{ji}$, where P_{ij} is the probability of the exchange $\vec{ij}$), exchanges are accepted using the Metropolis criterion:

$$P_{ij} = 1, \Delta_{ij} \leq 0 \tag{2.21}$$

where, Δ_{ij} is the difference in potential energy of the two replicas before and after the attempted swap. However, the transferability and performance of these methods to peptide-surface interface simulations is yet to be fully investigated. This is addressed in Chapter 5.

2.3.1 T-REMD

In T-REMD, the temperature of the highest replica is ideally chosen such that the system can cross even the highest barriers between minima on the PEL. The number of replicas required for T-REMD scales on the order of $f^{1/2}$, where f is the number of degrees of freedom of the entire system. Not only does this become computationally expensive for large complex systems, such as an aqueous peptide-surface interface, but the increased number of replicas prolongs the time taken for a replica to traverse the entire temperature ladder, degrading the performance of the method.

Different methods to improve the efficiency of T-REMD by reducing the number of replicas required to span temperature space have been proposed [Li et al. [2007, 2011]; Li and Latour [2011]; Rick [2007]; Lee and Rick [2009]; Ballard and Jarzynski [2009, 2012]]. Latour and *co-workers* developed the Temperature Intervals with Global Exchange of Replicas (TIGER) algorithm [Li et al. [2007, 2011]; Li and Latour [2011]], and subsequent variants, in which all replicas are quenched to a baseline temperature before calculation of a Metropolis-like exchange criterion. Replica Exchange with Nonequilibrium Switches (RENS) follows a similar methodology but instead of quenching all replicas to a baseline temperature prior to the attempted exchange A↔B, the temperature of A is increased to that of ensemble B and the temperature of B decreased to that of ensemble A in short ‘switching simulations’ [Ballard and Jarzynski [2009]]. The principles behind this approach can equally be applied to H-REMD with the potential A/B being increased/decreased to that of ensemble B/A respectively [Ballard and Jarzynski [2012]]. A third method that has been proposed for reducing the computational expense of a T-REMD simulation is Replica Exchange with Dynanical/Driven Scaling (REDS/REDS2) [Rick [2007]; Lee and Rick [2009]]. In REDS, a replica which is transient in temperature space is placed between each pair of conventional replicas. The latter are constant temperature ensembles widely separated in temperature space. The potential energy of the former, transient, replica is scaled between 0 and 1 by a parameter λ which is either treated as a dynamic variable within the simulation [Rick [2007]] or follows a fixed trajectory [Lee and Rick [2009]]. Both RENS and REDS have so far, however, been subjected to only limited testing.

2.3.2 H-REMD

H-REMD has bright prospects for reducing the computational cost of REMD for large, complex systems. A number of different variations of H-REMD have been reported in the literature so far that focus on addressing specific physio-chemical challenges [Fukunishi et al. [2002]; Liu et al. [2006]; Hritz and Oostenbrink [2008]; Meng et al. [2011]; Liu et al. [2005]; Itoh et al. [2010]; Itoh and Okumura [2013]]. In addition, T- and H-REMD can be combined in multidimensional REMD, where exchanges could be attempted between replicas at different temperatures and potential alternately [Sugita et al. [2000]].

Recently, promising results have been reported for Solute Tempering [Liu et al. [2005]; Moors et al. [2011]; Terakawa et al. [2011]; Wang et al. [2011]; Itoh et al. [2010]; Itoh and Okumura [2013]], a variant of H-REMD in which in the Hamiltonian the potential energy of the system, $V(X)$, is divided into two groups: one that comprises the 'solute' (denoted 'p')–typically the peptide/protein or portion thereof–and the other the 'solvent'–all remaining molecules/atoms (denoted 's'):

$$U(X) = U_{pp}(X) + U_{ps}(X) + U_{ss}(X) \tag{2.22}$$

By only scaling those parts of the Hamiltonian related to the solute, U_{pp} and U_{ps}, the number of replicas required scales on the order of $f_p^{1/2}$, where f_p is the number of degrees of freedom of the solute and not the entire system. Unlike T-REMD, only the zeroth replica, with a unmodified potential, samples the correct Boltzmann-weighted ensemble in this method. However, it is possible to extract unbiased data from the other replicas using reweighting techniques [Itoh et al. [2010]; Itoh and Okumura [2013]]. Solute Tempering has additional advantages for biointerfacial simulations since it offers a way in which to circumvent the issue of the substrate (included in the 'solvent' group) melting or dissolving at high temperatures–a potential pitfall of T-REMD.

In the most general form of solute tempering H-REMD, referred to herein as 'REST', *all* intra-solute interactions are incrementally scaled by an 'effective' temperature. Alternative versions have been proposed, namely: van der Waals REM [Itoh et al. [2010]] and Couloumb REM [Itoh and Okumura [2013]], in which only the LJ extent and partial charges of solute atoms are scaled, respectively. Only the REST method has been employed in this thesis. In its original formulation [Liu et al. [2005]], REST was found to perform badly for biomolecular systems which underwent large conformational changes [Huang et al. [2007]]. Several groups concurrently proposed an alternative scaling factor for the solute–solvent interactions [Moors et al. [2011]; Terakawa et al. [2011]; Wang et al. [2011]] within the Hamiltonian which not only improved replica mobility through effective temperature space and conformational sampling but also could be implemented within the MD software code Gromacs [Hess et al. [2008]]. It is this second formulation of REST that has been used here and is outlined in Section 2.3.3].

2.3.3 Replica Exchange with Solute Tempering

All REST simulations were implemented according to the method outlined in Terakawa *et al.* [Terakawa et al. [2011]] using the replica exchange and free energy perturbation codes within the standard Gromacs package [Hess et al. [2008]]. Specifically, in the biointerfacial REST simulations reported in Chapters 5, 7 and 8, the potential energy of replica j was scaled according to:

$$U_j(X) = \frac{\beta_j}{\beta} U_{pp}(X) + \sqrt{\frac{\beta_j}{\beta}} U_{ps}(X) + U_{ss}(X) \tag{2.23}$$

where U_{pp} is the intra-peptide, U_{ps} the peptide-water and peptide-surface, and U_{ss} the water-water, water-surface and surface-surface potential energies of the system X, respectively. β and β_j are the inverse of the system and 'effective' temperatures (*i.e.* β=1/T, T=system temperature) of replica j and are related via λ_j:

$$\beta_j = \beta(1 - \lambda_j) + \beta_H \lambda_j \tag{2.24}$$

where $0 \leq \lambda_j \leq 1$ and β_H is the highest effective temperature. To recover Equation 2.23 intramolecular peptide bonded (dihedral torsion force constants only in the case of Chapter 5 but both bond stretching and dihedral torsion force constants for the simulations reported in Chapters 7 and 8) and non-bonded LJ interaction energy parameters were scaled in each replica by $\lambda\gamma$ (where γ=β_H/β), while atomic partial charges were scaled by $\lambda\sqrt{\gamma}$. Wang *et al.* reported that out of the intramolecular bonded interactions, it was only essential to scale dihedral torsion force constants in order to achieve optimal conformational sampling of a peptide [Wang et al. [2011]]. Prior to production simulation, the 'effective' temperature of each replica was iteratively derived from a series of short test simulations to ensure uniform acceptance probabilities across replica space using the formula proposed by Terakawa *et al.* [Terakawa et al. [2011]]:

$$\lambda_i^{new} = \lambda_{i-1}^{new} + (\lambda_i^{old} - \lambda_{i-1}^{old})(a_i / < a >)^{\frac{1}{6}} \tag{2.25}$$

where a_i and $< a >$ are the acceptance probability of the exchange $(i-1) \rightarrow i$ and the average value across all replicas, respectively.

2.4 Free Energy Methods

Using the principle of ergodicity, the free energy change associated with moving from state i to state j, $\Delta\mathrm{G}_{ij}$, can be approximated from a MD simulation using the following relationship:

$$\Delta G_{ij} = -k_B T ln(P_{ij}) \tag{2.26}$$

where P_{ij} is the ratio of the probability of the system being in state j to the probability of it being in state i taken over the length of the simulation; T the temperature of the system; and k_B the Boltzmann constant. For Equation 2.26 to hold it is essential that both states i and j are adequately sampled, something which is difficult to achieve in large simulations of systems with complex PEL, as mentioned previously (Section 1.2.4).

In the case of biointerfacial systems, the free energy change of adsorption is of particular interest. In order to calculate this from molecular simulation for systems in which the adsorbate has high affinity for the interface, advanced sampling techniques are often used. For example, Latour and *co-workers* derived a scheme for calculating the free energy of adsorption of peptides which incorporated both T-REMD–to facilitate sampling of internal peptide conformational space–and a biasing potential–to ensure that the peptide visits both adsorbed and desorbed states [Wang et al. [2008]]. This method has been successfully applied to the adsorption of host-guest TGTG-X-TGTG peptides at a range of interfaces [Vellore et al. [2010]; Snyder et al. [2012]].

A number of alternative free energy approaches exist, two of which have been employed in the work carried out for this thesis (Chapters 3 and 8) and described in detail below–the Potential of Mean constraint Force (PMF) method and metadynamics (Sections 2.4.1 and 2.4.2, respectively).

2.4.1 Potential of Mean Force

Using a suitable reaction co-ordinate (z') which connects the two states of interest, the free energy landscape of the system along the same reaction co-ordinate can be calculated using [Trzesniak et al. [2007]]:

$$\Delta G(z) = \int_0^z < F(z') > \mathrm{d}z' \tag{2.27}$$

where ΔG(z) is the free energy change associated with the transition z'=0 → z'=z and $< F(z') >$ is the average force required to constrain the system at z'. In order to calculate the free energy of adsorption a small molecule at an interface, a suitable reaction co-ordinate is the vertical distance between the surface and the centre of mass (*c.o.m.*) of the adsorbate. The PMF method, using this reaction co-ordinate, has in recent years been used to quantify the adsorption of amino-acids (or their analogues) at the Quartz [Notman and Walsh [2009]; Wright and Walsh [2012a]], rutile [Monti and Walsh [2010]] and gold [Hoefling et al. [2010a]] aqueous interfaces.

The calculations reported in Chapter 3 were performed using the Gromacs pull code, to calculate the force required to constrain the reaction co-ordinate at each time step. Simulations at each adsorbate-surface separation were run for a minimum of 2 ns or until convergence of force block averages [Hess [2002]] was achieved (up to a maximum of 14 ns). The average force was calculated over the last 1 ns for large adsorbate-surface separations and over the last 2 or 4 ns for adsorbates close to the surface, where the force block averages converged more slowly. Between 50 and 60 MD runs per adsorbate were performed with the separation between the adsorbate and the surface being varied systematically from around 1.0 Å to 14 Å. A block averaging method [Hess [2002]] was used to estimate errors in the average forces. Average forces were integrated according to Equation 2.27. Errors in the PMF free energy were calculated by propagation of the average force errors in the integration.

Whilst by performing non-equilibrium simulations the PMF method ensures that all states of the system that lie along the reaction co-ordinate are adequately sampled, its ability to probe all possible configurations along orthogonal directions on the PEL is not enhanced [Jambeck and Lyubartsev [2013]]. For example, in the case of bio-interfacial PMF calculations, additional measures must be taken to ensure different internal adsorbate and adsorbate-surface geometries are sampled. Small molecules, such as those used in Chapter 6, have few internal degrees of freedom and adequate adsorbate-substrate sampling can be achieved using a small number (2 or 3) of different starting geometries for each restrained simulation. For larger adsorbates, such as peptides, however, the extra computational expense of ensuring adequate conformational sampling makes the PMF method for calculating the free energy of adsorption prohibitive.

2.4.2 Metadynamics

Metadynamics (metaD) is a method in which the free energy profile of a system along (a) given reaction co-ordinate(s) can be reconstructed from the bias added on-the-fly to a non-equilibrium MD simulation [Laio and Parrinello [2002]]. Briefly, reaction co-ordinates, s_i, (more commonly, in the context of metaD, referred to as 'collective variables') which connect significant minima in the PEL must be chosen. Typically Gaussians of width σ_i and height w are added with a frequency of $1/\tau$ to the PEL at the co-ordinates of each collective variable $\mathrm{s}_i(\mathrm{X(t)})$ (where X is the co-ordinates of the system at time t). The simulation therefore progresses on the time-dependent energy landscape given by the unbiased PEL of the system, U(X), and the metaD bias, V(X,t):

$$V(X,t) = w\Sigma_{t'<t}\Pi_i exp\left[-\frac{(s_i(X(t)) - s_i(X(t')))^2}{2\sigma_i^2}\right]. \tag{2.28}$$

As the bias is built up, barriers between low energy states of the system are reduced thus enabling the simulation to escape from deep minima in the unbiased PEL. In the limit t→ ∞:

$$V(X,t) \rightarrow -G(X) \tag{2.29}$$

where G(X,t) is the free energy of the system. In the case of bio-interfacial simulation, a suitable collective variable for determining the free energy of adsorption is the distance between the centre of mass (*c.o.m.*) of an adsorbate and the surface, in the direction of the surface normal. In Chapter 8 the adsorption of a gold binding peptide, AuBP-1 [Hnilova et al. [2008]] onto different crystallographic planes of gold is simulated using metaD. As with PMF calculations, the issues associated with adequately sampling phase-space orthogonal to a collective variable [Jambeck and Lyubartsev [2013]] remain. Hence in the work reported here, the REST method–to augment peptide conformational sampling–has been used in combination with metaD [Camilloni et al. [2008]; Schneider and Ciacchi [2012]]. The same principles lie behind REST metaD and the biased T-REMD method developed by Latour and *co-workers* [Wang et al. [2008]; Vellore et al. [2010]; Snyder et al. [2012]]. In the former, REST metaD, the PEL is biased on-the-fly, while in the latter, expensive umbrella-sampling calculations to determine the form the biasing potential should take, prior to the T-REMD simulation, are required. ‘Umbrella-sampling’ is a free energy method, analogous to PMF but using a harmonic constraint rather than a fixed restraint on the reaction co-ordinate [Trzesniak et al. [2007]].

When implementing metaD with REST, the metaD bias added to the two replicas i and j involved in a trial exchange must also be included in the potential energy difference, Δ_{ij}, used to determine the probability with which the exchange is accepted (Equations 2.21):

$$\Delta_{ij} = \{U_i(X_i) + V_i(X_i) + U_j(X_j) + V_j(X_j)\} - \{U_i(X_j) + V_i(X_j) + U_j(X_i) + V_j(X_i)\} \tag{2.30}$$

where, U(X) and V(X) refer to the potential energy of the system in the absence of the metaD bias (Equation 2.23) and the metaD bias itself (Equation 2.28), respectively. It is noted that there are two principal differences between Hamiltonian replica-exchange metaD (such as REST metaD) and Temperature replica exchange metaD reported by Bussi *et al.* [Bussi et al. [2006]] in terms of how the two are implemented:

1. The momenta of particles following a successful exchange attempt move are not rescaled;
2. Gaussians of the same height are added to the biasing potential in all replicas.

Both of these features arise from the fact that system temperature is invariant in replica space in the former, Hamiltonian replica exchange metaD.

Equilibrium, but not time dependent, properties of a system can be recovered by suitable reweighting of data from non-equilibrium simulations. In general, the unbiased probability of an observable having a specific value $P(Obs)$ can be derived from the respective biased probability, $P_{biased}(Obs)$, using:

$$P(Obs) = \frac{< P_{biased}(Obs) exp\{\frac{W}{kT}\} >}{< exp\{\frac{W}{kT}\} >} \tag{2.31}$$

where W is the weight factor. For REST metaD simulations specifically, W is comprised of two components: the time dependent metaD bias and the bias due to the scaling of solute interactions by an effective temperature, $Y(X(t))$ (derived from Equation 2.23):

$$Y(X(t)) = (1 - \frac{\beta_j}{\beta})U_{pp}(X(t)) + (1 - \sqrt{\frac{\beta_j}{\beta}})U_{ps}(X(t)) \tag{2.32}$$

The simplest approach, analysis of the reference replica ($\lambda = 0$, Equation 2.24) has been adopted only herein–both for REST MD simulations reported in Chapters 5 and 7, and REST metaD simulations detailed in Chapter 8. Incorporating data obtained from all replicas in a REST simulation would be one way to improve statistical sampling in the future.

The factor by which the time dependent metaD bias to the reference replica of a REST metaD simulation must be re-weighted to obtain the unbiased ensemble of states also requires careful consideration. While using $V(X(t), t)$, as defined in Equation 2.28, directly is formally correct, it will exponentially weight frames from the end of the simulation more heavily than those in the initial stages–a consequence of $V(X(t), t)$ growing continuously with time. Alternative reweighting strategies are discussed in Section 8.3.

2.5 Analysis

In the analysis of all interfacial simulations carried out in this work it was important that adsorbate-surface separation was defined in a consistent manner. In the case of the aqueous α-Quartz interface, the average z co-ordinate of Si atoms in the top layer of silica was chosen as the reference point. The same definition has been used in previous studies [Notman and Walsh [2009]; Oren et al. [2010]]. Similarly, binding distances were measured from the z co-ordinate of gold atoms in the uppermost layer of the three gold surfaces. For the reconstructed Au(100) interface this was an average co-ordinate due to undulations in the surface.

2.5.1 Hydrogen Bonding

The geometric criteria defined by Jedlovszky *et al.* to determine the existence of a hydrogen bond in a box of TIP3P water were employed in the analysis of most simulations reported in this thesis [Jedlovszky et al. [1998]]. (Hydrogen bond analysis presented in Chapters 5 and 6 was carried out slightly differently for the reasons outlined in the text.) Specifically, a hydrogen bond was said to be present on the simultaneous satisfaction of two distance metrics–O/N$\cdots$O and H$\cdots$O–and one angle–H$\cdots$(O/N)$\cdots$O. In Chapters 3 and 4, cut-off distances were fine-tuned for each specific type of hydrogen bond using the position of the first minimum in the appropriate radial distribution function (RDF), following the precedent set in reference Jedlovszky et al. [1998]. Distances derived for intermolecular hydrogen-bonding in TIP3P water were used in all other cases. An angular cut-off of 30° was used throughout to ensure a near linear arrangement of the three atoms involved

Jedlovszky et al. [1998]; Luzar and Chadler [1996, 1993].

The persistence of a hydrogen bond over two different time-scales (ps and ns) can be probed by calculation of the continuous $S(t)$ and instantaneous $C(t)$ hydrogen-bond time autocorrelation functions respectively. These are defined as:

$$X(t) = \frac{< h(0) \cdot h(t) >}{< h >} \tag{2.33}$$

where $< h >$ is the number of hydrogen bonds at t=0 averaged over a number of different starting points throughout the simulation. $h(t)$ is 1 if a hydrogen bond present at time 0 is also present at time t and 0 otherwise; for $X(t) = S(t)$ this holds only if the particular hydrogen-bond exists between times 0 and t, whilst for $X(t) = C(t)$ it is independent of this condition. For short time-scale simulations, such as the CPMD simulations reported in Chapter 4, the continuous hydrogen-bond time-autocorrelation function, $S(t)$, has been reported to be a better measure of hydrogen-bond lifetime [Rosenfeld and Schmuttenmaer [2011]; Paul and Chandra [2004]] because it normally displays exponential decay within the timescales probed and is not dependent on the time at which the hydrogen bond was formed. In addition, for small length-scale systems it is inevitable that hydrogen bonds between specific atoms will break and reform in time due to the limited number of atoms. This would potentially lead to an over-estimation in the characteristic hydrogen-bond lifetime if the instantaneous, rather continuous time autocorrelation function was used. On the other hand, for large systems and larger time-steps, the instantaneous correlation function is more appropriate since it is independent of sampling frequency apart from for very short life-times. It must be noted that in this study hydrogen bonds between individual *atoms* rather than between *molecules* have been identified in all analyses. (For example, hydrogen bonds $H_aO_c{-}H_b \cdots O_dH_2$ and $H_bO_c{-}H_a \cdots O_dH_2$ were classed as distinct.)

2.5.2 Peptide Structure Clustering

To identify the most likely structure a peptide or protein sequence is likely to adopt in a particular environment it is often useful to group similar structures, generated during a simulation, together into clusters. Not only this, but cluster analysis can also:

- give an indication of the amount of peptide conformational space explored by a system. (For example a peptide sequence which samples a larger number of clusters than another is likely to be conformationally more lable, whilst when benchmarking two methods using the same system, the method which enhances conformational sampling the most will give rise to the largest total number of clusters.)

- be used as a metric for equilibration of a simulation by monitoring convergence in the total number of clusters identified.

Throughout this thesis (Chapters 5, 7 and 8), the Daura clustering algorithm was used to classify peptide (QBP-1 [Oren et al. [2007]] and AuBP-1 [Hnilova et al. [2008]]) structures into clusters [Daura et al. [1999]]. Briefly, the protocol for this procedure was as follows:

1. Calculate the root mean squared deviation (RMSD) of a subset of peptide atoms, after least-squares fitting, between all pairs of structures within the pool of structures generated during a trajectory.
2. Determine the number of neighbours each structure has. A 'neighbour' is defined as another structure for which the RMSD between the pair is less than a cut-off value.
3. Assign the structure with the largest number of neighbours (labelled the 'centroid' structure of the cluster), along with all its neighbouring structures to a cluster.
4. Remove all structures assigned to the cluster from the overall pool of structures.
5. Repeat steps 1-4 until all structures have been assigned to a cluster.

Unless stated otherwise, peptide structures were clustered according to their backbone atom positions, with a RMSD cut-off of 2 Å. The same procedure and cut-off were used to assign structures of QBP-1 in solution [Notman et al. [2010]]. Here a cluster was defined as being 'significant' based on its size. Specifically, the total number of 'significant' clusters was that required to account for 95% of the total structure population.

To investigate whether similarities existed between different simulations (Chapter 5) or environments (Chapter 7 and 8), in the structures most likely to be adopted by a given peptide, additional RSMD calculations were performed. In this case only the centroid structures of the most populated clusters from each environment/simulation were chosen. Clusters, and the structures within them, were said to be 'identical' if the RMSD was less than the cut-off used in the original cluster analysis (2 Å) and 'similar' if it fell within the range 2-3 Å.

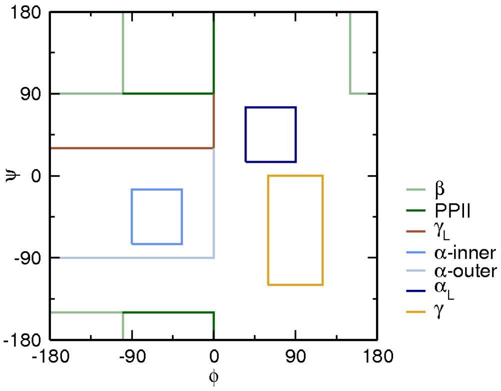


Figure 2.2: Boundaries in ϕ/ψ space demarking the principal regions in a Ramachandran plot for analysis of QBP-1 structure.

2.6 Peptide Secondary Structure

A Ramachandran plot, which depicts the distribution of peptide backbone ϕ-ψ dihedral angles sampled over a trajectory, is often used to give a visual indication of the secondary structural characteristics of a peptide. Specific combinations of the torsion angles are representative of different peptide secondary structures, such as α-helix (-60,-45), β-sheet (-135,-135), and PPII (-75, 150). As with peptide cluster analysis, it is interesting to quantify the similarities and/or differences in the secondary structural characteristics of a given peptide either in different environments or in the same environment (Chapter 7 and 8), but simulated using a different protocol (Chapter 5). For this purpose ϕ-ψ phase-space was subdivided into bins centred around the characteristic range of values for each type of secondary structure region (see Figures 2.6 and 2.6). The percentage of dihedral angles found to lie within each bin was then calculated.

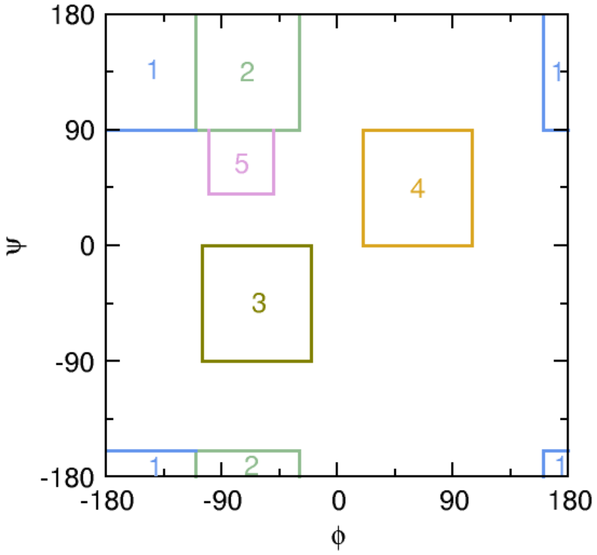


Figure 2.3: Boundaries in ϕ/ψ space demarking the principal regions in a Ramachandran plot (1 β, 2 PPII, 3 α, 4 α_L and 5 γ_L) for analysis of AuBP-1 structure.

Chapter 3

Free Energy Calculations of Amino-Acid Analogue Adsorption at Aqueous α-Quartz Interfaces.

3.1 Introduction

In this Chapter, the free energy of adsorption of amino-acid analogues at three different aqueous α-Quartz interfaces–(100), (011) and (001)–have been calculated using the Potential of Mean constraint Force method. The motivation for carrying out this work was two-fold. First the relative binding affinities of the functional groups present in the side-chains of amino-acids to α-Quartz gives an indication of the intrinsic contribution of a residue, within a peptide, to its overall adsorption. Such binding energies are independent of the influence of peptide conformation. Secondly, the affinity of a given adsorbate for one interface over another allows, for the first time, the hypothesis that peptide binding could be used to discriminate between the different crystallographic planes of α-Quartz to be probed.

This work is an extension of that previously reported by Notman and Walsh who calculated the free energy of adsorption of water, methane and methanol at the aqueous (100) α-Quartz interface [Notman and Walsh [2009]]. For the reasons outlined in Section 1.2.2, the adsorption of simple small molecules which feature the same functional groups as those present in a peptide has been modelled, rather than the amino-acids themselves. This was done in order to de-convolute side-chain from N-/C-terminus driven adsorption; the former being more relevant for peptide interfacial binding. The 'building block' approach of using the surface binding affinity of each functional group present in an amino-acid in isolation, to estimate its overall attraction to an interface, has recently been validated by first-principles calculations of glycine adsorption on to chromium oxide [Garrain et. al. [2011]]. Seven adsorbates were chosen for the study–water, methane, methanol, benzene, ethanoate, ammonium and butyl ammonium–analogues of the amino-acids alanine/leucine/valine/isoleucine, serine/threonine, phenylalanine/tryptophan/tyrosine, glutamate/aspartate and lysine, respectively. Two analogues of lysine were modelled to investigate the possible role that post-translational alklylation of the residue in silaffins [Kroger et al. [1999, 2001]]–proteins known to aid silica precipitation in diatoms–might play in silica binding. The results for five adsorbates–methane, ammonium, butyl ammonium, benzene and ethanoate–only are presented hereafter. For further details concerning water and methanol adsorption see reference [Wright and Walsh [2012a]].

3.2 Methods

MD simulations of the adsorption of amino-acid analogues at the aqueous (100), (001) and (011) α-quartz interfaces, were carried out using the Gromacs 3.3.3 software package [Lindahl et al. [2001]]. The (100) (α termination), (001) and (011) surfaces had areas of 544, 837 and 1359 Å^2 and silanol densities 0.0588, 0.0765 and 0.0471 OH Å^{-2}, respectively (Figure 1.2). (In order to satisfy the partial charge distribution in LFF [Lopes et al. [2006]], the FF used to model quartz here, the opposite surface of each slab was hydrogenated. The same slab set-up–hydroxylation of the interface at which adsorption occurs and hydrogenation of the other–has been used throughout this thesis, in all simulations of aqueous Quartz interfaces (Chapters 3, 4, 5, 7).) Three-dimensional periodic boundary conditions were imposed such that the average inter-slab (vertical) spacing was around 33.2 - 36.7 Å, amounting to a total of approximately 580, 1000 and 1450 explicit water molecules located between neighbouring slabs for the (100), (001) and (011) cases respectively.

All simulations were performed using the *NPT* ensemble at a temperature of 300 K using a Nosé-Hoover thermostat, and at a pressure of 1 bar with pressure coupled to a Parrinello-Rahman barostat (in the direction normal to the surface only). Newton's equations of motion were solved using the leapfrog algorithm [Hockney et al. [1974]] with an integration time step of 1 fs. Cut-offs for non-bonded Lenard-Jones and Particle Mesh Ewald (PME) electrostatic interactions were 10 Å. Unlike previously [Notman and Walsh [2009]], no constraints were applied to any of the silanol O-H bonds, so that all α-Quartz atoms in the slab were free to move in these simulations. Bond lengths within the water molecules were constrained using the SETTLE algorithm [Miyamoto and Kollman [1992]]. Each simulation was prepared by a steepest decent energy minimisation followed by a 1 ns MD run to equilibrate the water molecules. For larger adsorbates, butyl ammonium, benzene and ethanoate, several different starting conformations relative to the surface were used for the constrained runs at small adsorbate-surface separations.

Since this work was a prelude to investigations into the interactions between peptides and α-Quartz surfaces, both the quartz (LFF [Lopes et al. [2006]]) and water (modified TIP3P [Jorgensen et al. [1983b]; Neria et al. [1996]]) FFs were chosen because they are compatible with the bio-organic FF, CHARMM [MacKerell et al. [1998]]. Amino-acid analogues were modelled using the CHARMM parameters for the side-chain functional groups of their respective amino-acids. Partial charges were adjusted slightly after hydrogen termination to ensure that the charge state of each adsorbate was correct for pH 7. Counter-ions (either Na^+ or Cl^-) were added where required.

3.2.1 Analysis

The average number of hydrogen bonds per time frame and their average lifetime, made by ammonium and ethanoate to either the α-Quartz surface or water at their PMF minima distance from the surface, were calculated over the last 0.5 ns of their respective constrained trajectories. For this analysis, coordinates were saved more frequently, every 0.01 ps. The characteristic lifetime of a hydrogen bond was defined as the time taken for $C(t)$ (hydrogen bond time autocorrelation function, see Section 2.5.1 for definition) to decay to $\frac{1}{e}$.

To reveal whether adsorbates were anisotropically oriented at their PMF minimum separations, adsorbed conformations were classified by the angles made by key bonds (or in the case of benzene, plane) with respect to the surface normal. Analysis was done using the last 2 ns of the constrained MD runs for all molecules. Details of the specific angles used for each classification are given with the corresponding Figures (Figures A.3, A.4 and A.5).

3.3 Results

The optimal free energy of adsorption for each adsorbate, and the surface-separations at which these occur are summarised in Table 3.1. In the following, the results for the five adsorbates, methane, ammonium, butyl ammonium, benzene and ethanoate, are presented in turn.

	(100)		(001)		(011)	
	kJ mol^{-1}	Å	kJ mol^{-1}	Å	kJ mol^{-1}	Å
methane	-5.17 ± 0.49	1.4	-4.48 ± 0.37	3.2	-6.71 ± 0.38	3.0
ammonium	1.26 ± 0.36	4.0	3.04 ± 0.37	4.1	5.98 ± 0.25	5.0
butyl ammonium	-2.82 ± 0.56	3.7	-1.94 ± 0.42	4.6	-4.39 ± 0.36	4.4
benzene	-8.75 ± 0.47	3.1	-7.16 ± 0.35	4.7	-11.89 ± 0.35	4.2
ethanoate	-5.51 ± 0.45	3.7	-6.76 ± 0.39	4.1	-11.96 ± 0.42	4.0

Table 3.1: Optimal PMF free energy of adsorption, and vertical position of the PMF minimum for amino-acid analogues on (100), (001) and (011) α-Quartz surfaces, under aqueous conditions. Errors were obtained via block averaging as outlined in Section 2.4.1.

3.3.1 Methane

Methane adsorbed favourably at all three aqueous α-Quartz interfaces with optimum binding strength increasing with decreasing surface-silanol density (Table 3.1). The binding of methane to the (011) surface was significantly more favourable than that to either the (100) or (001) surfaces, indicating that the binding of methane to α-quartz is facet specific (Table 3.1). As well as showing energetic discrimination over crystallographic termination, there was also a distinct spatial difference (in terms of the optimal methane-surface separation (Table 3.1)) between (100) and the (001) and (011) surfaces.

The (100) methane free energy profile agrees with previous work [Notman and Walsh [2009]] (Figure A.1a). The PMF minimum at 1.4 Å coincides with the PMF local minimum of water within interstitial surface sites, where water density is approximately half that of bulk. This suggests that the unfavourably close proximity of methane to surface silanols is more than outweighed energetically by the ability of the interface hydrophobically to shield the adsorbate. Herein, surface-mediated ‘hydrophobic shielding’ refers to both the direct–a reduction in the solvent accessible surface area of a bound species–and indirect–regions of low density being present at an interface which perturbs the structure of water–influence of adsorption. Although the (011) and (001) surfaces have no interstitial sites, orientational analysis [Wright and Walsh [2012a]] revealed that hydrophobic screening could be used to explain the trend in binding affinities of methane at all three aqueous interfaces. Specifically, preferential adsorption to (011) is consistent with this being the only interface at which the most likely adsorbed geometry was ‘tripod’-like (Figure A.1d)–an orientation which minimises solvent accessible surface area of the hydrophobic molecule.

Whilst at odds with the highly charged nature of silica-binding peptides (Table 1.1), the favourable adsorption of methane tallies with the hydrophobic content of those sequences thought to have high affinity specifically for α-Quartz [Tamerler et al. [2007]; Oren et al. [2007]]. It also raises the possibility that it is the post-translational *alkylation* of lysines in silaffins (which enriches their hydrophobic content) that promotes silica-binding, rather than the presence of the charged ammonium groups.

3.3.2 Ammonium

Unlike methane, the free energy of adsorption of ammonium at all three aqueous interfaces was unfavourable at all distances considered (Figure 3.1, inset, Table 3.1). Only for (100) and (001) was there a local minimum in the PMF profile, at a surface separation of approximately 4.0 Å; a distance commensurate with the second minimum in the PMF profile of methane on (100) (Figure A.1). These results were in sharp contrast to the (110) rutile titania surface where calculations suggest that ammonium adsorbs relatively strongly while methane does not [Monti and Walsh [2010]].

Adsorption of ammonium to α-Quartz can only be mediated by hydrogen bonding and/or electrostatic interactions. The very small number of quartz-ammonium hydrogen bonds formed, either at the PMF minima or when closer to the surface, can be accounted for by the small partial charge on the silanol oxygens within the FF (Tables 3.2 and A.1). Hence moving ammonium from bulk water into interstitial sites on the (100) surface, where the water density is approximately half that of bulk, is unfavourable. Equally, for (001) the ammonium PMF local minimum was located in a water-dense region, so moving closer to the interface, into water-depleted region, is unfavourable. However, the degree of structuring of the first water layer cannot be the only reason for the lack of adsorption since the free energy change of bringing ammonium from bulk water to the (011) surface was the least favourable of the three, despite the first water layer being most dense at this aqueous interface [Notman and Walsh [2009]; Wright and Walsh [2012a]]. This suggests that the adsorption of ammonium is also dependent on surface silanol density and topology.

The repulsive nature of the quartz-ammonium interaction suggested by these PMF calculations is at odds with the primary sequences of silica binding peptides (Table 1.1). However, the results of previous experimental and computational studies which probe this phenomenon [Churchill et al. [2004]; Vlasova and Golovkova

	(100)	(001)	(011)
quartz-ammonium	0.14	0.07	0.05
water-ammonium	1.82	1.93	2.15
quartz-ethanoate	17.34	>100	27.67
water-ethanoate	6.54	7.74	6.26

Table 3.2: Hydrogen bond correlation life-time (ps) formed by the adsorbates ammonium and ethanoate at the PMF minimum distance on (100), (001) and (011) α-Quartz.

[2004]; Zimmerman et al. [2004]; Kitadai et al. [2009]; Holinga et al. [2011]; Ben Shir et al. [2010, 2012]; Nonella and Seeger [2008]; Costa et al. [2008]; Zhao et al. [2011]; Rimola et al. [2009]; Folliet et. al. [2013]], against which the findings could be benchmarked, are ambiguous (see Sections 1.3.1 and 4.4 for a longer discussion). The possibility that a deficiency exists in the FFs used to model the system, therefore, could not be either verified or ruled out and hence warranted further study.

First, the FF used to model the adsorbate–CHARMM–was considered. Previous reports revealed that the attraction between a terminal NH_3^+ peptide group and non-modified TIP3P water is overestimated by CHARMM for solution-based simulations [Liang and Walsh [2007]]. Since in the work of Liang and Walsh [Liang and Walsh [2007]] closer agreement to first-principles was obtained for simulations carried out with AMBER than CHARMM, in terms of the NH_3^+–H_w RDF profile, the adsorption free energy profile of ammonium on the (100) α-Quartz surface was also calculated using AMBER parameters for NH_4^+. However, the PMF profiles for both FF were in agreement–both indicating that adsorption of ammonium on to α-Quartz was unfavourable (Figure A.2). In the bio-interfacial context, biomolecule-surface interactions modelled by CHARMM at the aqueous ammonium-terminated SAM interface were found to be underestimated [Vellore et al. [2010]]. Both the exaggerated ammonium-water interaction, making displacement of waters from the hydration shell of the molecule on surface adsorption unfavourable, and the underestimated attraction for the interface itself could account for the unfavourable ammonium-Quartz binding affinity calculated here.

Second, first-principles simulations of ammonium adsorption at the aqueous (100) α-Quartz interface were carried out to investigate the role of the silica FF. These calculations are reported in Chapter 4.

3.3.3 Butyl ammonium

Unlike ammonium, butyl ammonium bound to all three surfaces although not as strongly as the hydrophobic adsorbates methane and benzene (Table 3.1, Figure 3.1a). Within error, the adsorption free energy was the same on (100) and (001), but stronger on (011); this concurs with the binding trend of methane. However, there were notable differences in the PMF adsorption profiles between the (100) and the (001)/(011) surfaces arising from their surface topologies. For (100) the PMF minimum surface-separation was 3.7 Å. This was followed by a plateau in the adsorption profile at a distance which was approximately commensurate with the PMF adsorption minima on the other two surfaces. This can be explained by considering the positions of the individual atoms within butyl ammonium at each surface-*c.o.m.* separation in relation to the PMF profiles of methane and ammonium. For all three surfaces, the favourable adsorption free energies correspond to the closest approach of the *c.o.m.* for which the NH_3^+ group is positioned at distances commensurate with local minima in the corresponding ammonium PMF profile. The hydrophobic part of the molecule can then adopt average conformations for which individual carbon atoms are at separations roughly commensurate with minima in the methane PMF profile. On the (100) surface two minima in the methane PMF profile were accessible whereas on the other two surfaces there was only one minimum. Hence the surface topology directly gives rise to the distinctly shaped PMF profile of butyl ammonium on the (100) surface.

Visual inspection of the constrained trajectories at the PMF minima revealed that the mobility and conformational freedom of butyl ammonium was slightly reduced on the (100) surface relative to the other two surfaces as one/two carbons were trapped within the interstitial sites, preventing lateral movement–motion observed in the [$\bar{1}$20] and [110], and [$\bar{2}$10] directions (Figure 1.2) on the (001) and (011) surfaces, respectively. Figure 3.1b-d depicts typical geometries of butyl ammonium adsorbed at the PMF minimum separation at each interface; in every case, the NH_3^+ group protruded into bulk water. However, for the (001) surface it was difficult to identify a ‘typical’ conformation as the adsorbate retained full conformational freedom upon adsorption. Orientational analysis confirmed these visual observations, revealing on (100) a preferential adoption of an ‘N-up’ conformation at the PMF minimum surface separation; on (001) ‘flat’ conformations were most common, whilst on (011) ‘flat’ and ‘N-up’ were equally likely (Figure A.3).

To probe whether the effect on the adsorption free energy of adding an alkyl chain to ammonium is additive, the average height above the surface for each heavy atom in butyl ammonium was found from the last 2 ns of the constrained trajectories

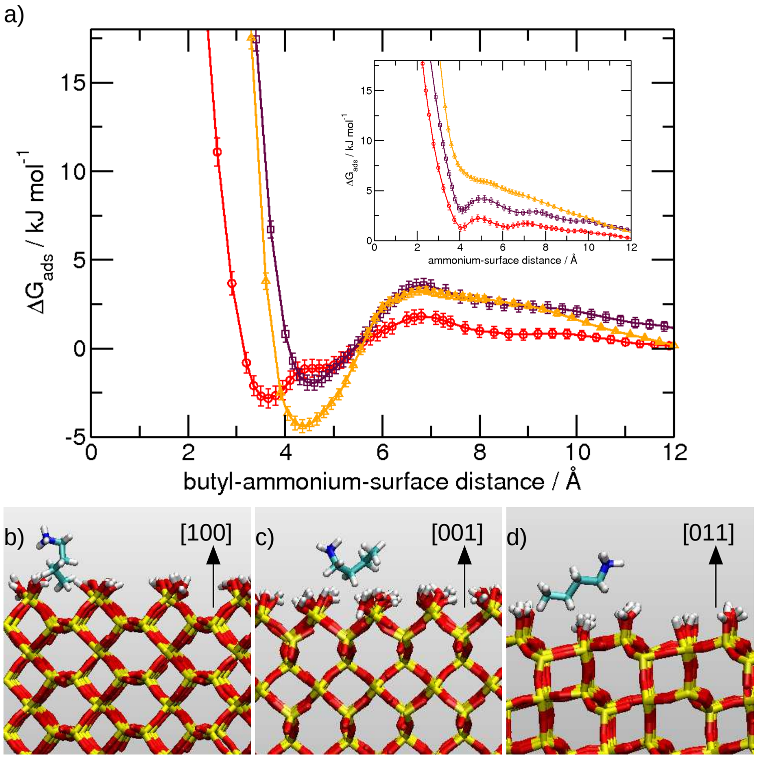


Figure 3.1: a) PMF free energy of adsorption butyl ammonium on the (100), (001) and (011)) α-quartz surfaces as a function of surface separation. Typical configurations of butyl ammonium adsorbed onto the b) (100) (‘N-up’), c) (001) (‘flat’) and d) (011) (‘flat’) surfaces. Inset: PMF free energy of adsorption of ammonium on the (100), (001) and (011) α-quartz surfaces as a function of vertical surface separation. Data in red, purple and yellow correspond with (100), (001) and (011) respectively.

at its PMF minimum. Each atom was then assigned an adsorption energy contribution, corresponding to the free energy of adsorption of methane or ammonium at same height on that surface. A prediction of the adsorption free energy of butyl ammonium was then made by summation of these contributions from each heavy atom. The additive free energies (in $\mathrm{kJ\,mol^{-1}}$) were -4.5 ± 2.25, -10.3 ± 4.09 and -13.81 ± 1.74 for (100), (001) and (011) respectively, lying outside the error boundaries of the actual PMF free energies calculated here for the latter two. The difference between the additive and calculated free energies can in part be accounted for by the size differences between the CH_2/CH_3 groups and methane. It would therefore be expected that the surface would provide each CH_2 or CH_3 group with less hydrophobic shielding than methane and hence their associated free energy of adsorption would be less. The closer agreement between the additive and calculated free energies for (100), relative to the other two surfaces, could be a consequence of the lack of lateral motion on the former at its PMF minimum, resulting in a more stable 'average conformation'. However, these data suggest that for a conformationally-restricted molecule adsorbed onto α-Quartz, a qualitative estimate of its adsorption free energy can be obtained from additive contributions of its functional groups. This could be a powerful tool in correlating the work here to AFM experimental data for small molecule adsorption on α-Quartz surfaces, where functional groups are typically attached to an AFM tips by a hydrocarbon chain.

3.3.4 Benzene

Of all the adsorbates tested, benzene adsorbed the most strongly onto the (100) and (001) surfaces (Figure 3.2, Table 3.1), showing statistically significant facet specificity in binding energy, between all three surfaces. The PMF adsorption profiles also display the same surface topological dependence as butyl ammonium, in which the (100) surface profile has a minimum followed by a plateau, whereas on (001) and (011), there is one minimum only.

Orientational analysis (Figure A.4) shows that for all three surfaces at its PMF minimum, benzene preferentially adopts an 'upright' conformation, whilst the 'flat' conformation is unfavourable. This is in contrast to butyl ammonium where on adsorption different orientations on the (100) and (001) surfaces were preferentially adopted. Visual inspection indicated that hydrophobic shielding of benzene by the (100) surface at the PMF minimum arose from the penetration of some of its hydrogen atoms into the interstitial sites when in an 'upright' conformation. On (001) and (011), it might be expected that maximal surface hydrophobic shielding could be obtained by benzene in a 'flat' conformation in the absence of large in-

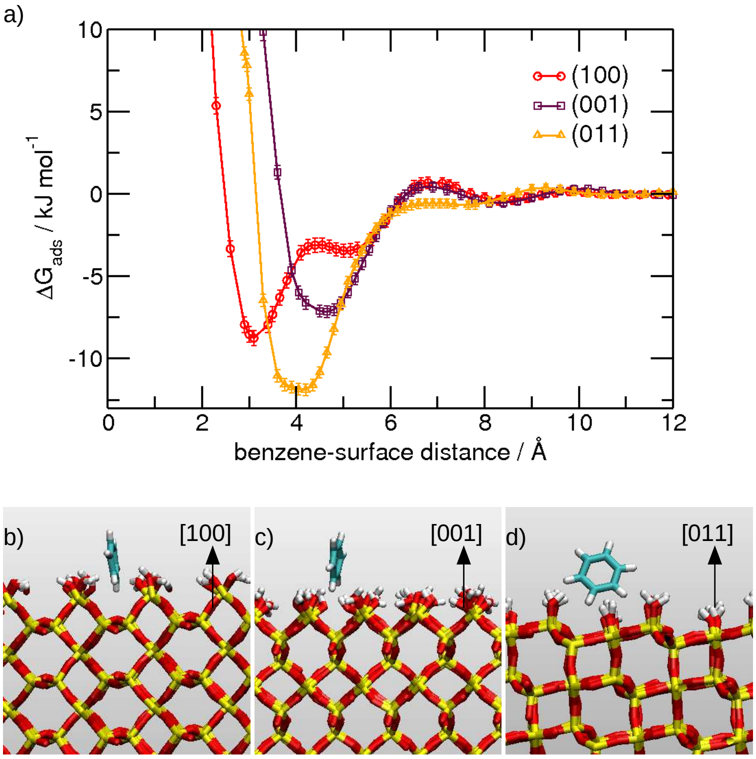


Figure 3.2: a) PMF free energy of adsorption benzene on (100), (001) and (011) α-Quartz surfaces as a function of surface separation. Typical configurations of benzene adsorbed onto the b) (100) ('upright'), c) (001) ('upright') and d) (011) ('upright') surfaces.

terstitial sites at either interface. However, calculation of the average height above the surface of the hydrogen atoms (using data from the last 2 ns of the constrained trajectories at the PMF minimum) revealed that on average four lay within water depleted regions at all three aqueous interfaces. This highlights the importance of water structuring above α-quartz surfaces on the adsorption of aromatic residues.

All the α-Quartz surfaces studied feature interstitial grooves. To test whether benzene showed preference over any of these grooves at the PMF minimum surface separation, ring plane lateral orientational analysis was carried out. This revealed that plane orientation was isotropic at this separation in all three cases. Lateral orientation became anisotropic at very small surface separations with alignment to surface grooves being preferred. This confirms visual inspection of the trajectories, in which benzene was observed to spin and tumble at the PMF minimum, but move laterally above the (001) (in the $[\bar{1}20]$ and [110] directions, Figure 1.2) and (011) (in the $[\bar{2}10]$ direction, very slowly, Figure 1.2) surfaces only.

Tryptophan is one of the residues prevalent in the sequence motifs identified by Oren *et al.* to be enriched in quartz-binding peptide sequences [Oren et al. [2010]]. They also reported that, during MD simulations of strong quartz-binding peptide sequences, the residue made close contact with the (100) α-quartz surface suggesting that tryptophan has a direct role in peptide adsorption. The same result was found from the simulations carried out in Chapters 5 and 7. However, it is expected that on adsorption a tryptophan residue within a peptide would have less conformational freedom than benzene due to the longer aspect ratio of the indole ring and the constraints imposed by the peptide backbone on the adsorption geometries.

3.3.5 Ethanoate

Ethanoate was the amino-acid analogue which bound most strongly to the (011) surface (Table 3.1, Figure 3.3a). It also showed substantial affinity for the (100) and (001) surfaces, where it adsorbed more strongly than methane. As with the other adsorbates, ethanoate also displayed spatial facet specificity. Contrary to benzene, methane and butyl ammonium, it bound more strongly to the (001) surface than (100), indicating that the forces driving ethanoate adsorption are distinct from those for the other three molecules.

There are two competing mechanisms for ethanoate adsorption onto α-quartz: hydrophobic shielding of the methyl group, and carboxylate-silanol hydrogen bonding. On the (100) surface two significant minima existed in the PMF profile. At the local minimum (3 Å) hydrophobic shielding of the methyl group may drive adsorption. This can be inferred from both visual inspection of the trajectory, where the

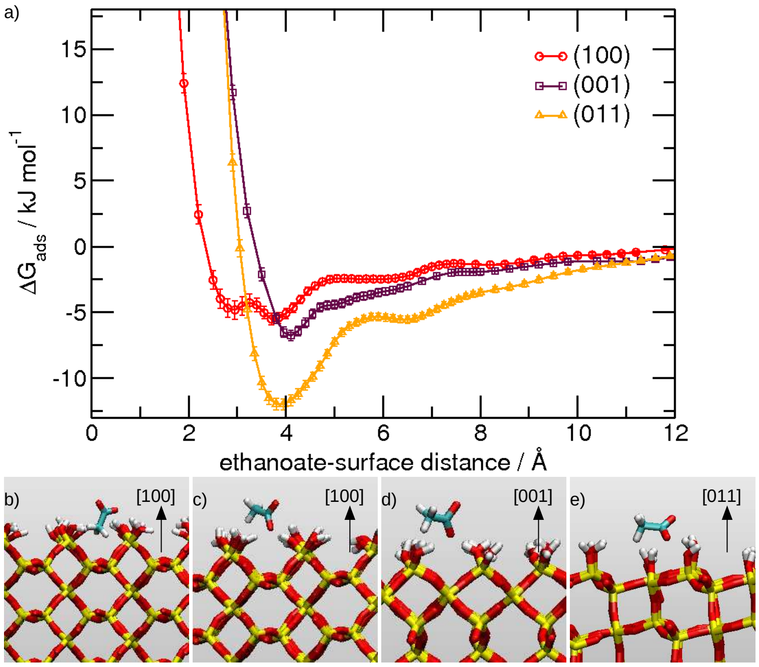


Figure 3.3: a) PMF free energy of adsorption ethanoate on (100), (001) and (011) α-Quartz surfaces as a function of surface separation. b)-d) Typical configurations of ethanoate on (100) at b) a separation of 3.0 Å ('carboxylate-up') and c) PMF minimum ('methyl-up'), d) on (001) ('flat'), and e) on (011) ('flat').

methyl group penetrates the interstitial space (Figure 3.3b), and by orientational analysis, where 'carboxylate-up' configurations are preferred (Figure A.5). However, hydrogen bond analysis identified that ethanoate-silanol hydrogen bonds also have an important role in driving its adsorption at this surface separation, having a very long characteristic lifetime (57.3 ps). This would also account for the average height of methyl carbon (when the $c.o.m$ of ethanoate is at this local minimum) being closer to the surface than that which would be energetically favourable for the adsorption of methane. On the other hand, at the global PMF minimum carboxylate-silanol hydrogen bonding drives ethanoate adsorption. Orientational analysis at this surface separation showed that 'flat' and 'methyl-up' configurations were almost equally likely (Figures 3.3 and A.5). The former configuration balances partial hydrophobic shielding of the methyl group with carboxylate-silanol hydrogen bonding, whilst in the latter, hydrogen bonding is maximised at the expense of exposing the methyl group to bulk water.

As noted previously, differences in surface topology, relative to the (100) surface, means that the PMF profiles of ethanoate adsorption onto (001) and (011) have only one significant minimum. Consistent with butyl ammonium, ethanoate also preferentially adopts a 'flat' configuration when bound to the latter two interfaces (Figures 3.3 and A.5). Both partial hydrophobic shielding of the methyl group and ethanoate-silanol hydrogen bonding drive adsorption. Although there were no statistically significant differences between the average numbers of ethanoate-silanol and ethanoate-water hydrogen bonds at the three interfaces, there was significant variation in the characteristic lifetimes of the ethanoate-silanol hydrogen bonds (Tables 3.2 and A.1). The longest lifetimes for both ethanoate-silanol and ethanoate-water hydrogen bonds were seen for (001), the surface with the highest silanol density, and on which ethanoate displayed the least lateral motion. The greater persistence of ethanoate-water hydrogen bonding at this interface is consistent with the previous study, in which the lateral diffusion of water molecules was found to be slower above the (001) surface than the (100) and (011) [Notman and Walsh [2009]]. However, neither the characteristic lifetimes of ethanoate-silanol hydrogen bonds nor the adsorption energy onto the three surfaces correlate with surface silanol density, suggesting that surface topology is also important. The fact that ethanoate adsorption onto the (011) surface is significantly stronger than onto the other two surfaces is likely, in part, to be due to ethanoate-water hydrogen bonding in the dense water layer above this surface.

Experimentally, the role of negatively-charged residues on the silica binding characteristics of peptides has not been probed extensively. Sano *et al.* examined the influence of residue D5 in TPB-1 (RKLPDA) on silica binding and silification by alanine substitution [Sano et al. [2005]]. They concluded that aspartate had a slight influence on the silica binding characteristics of the peptide, but not on silica precipitation. However, phage-displayed, rather than isolated, peptides were used in the surface binding experiments, where the presence of the phage could significantly influence peptide conformation and hence binding. The adsorption of aspartic and glutamic acid on α-quartz surfaces was found to be negligible by Churchill *et al.* [Churchill et al. [2004]], whose findings indicated that amino-acid–quartz adsorption was driven by electrostatic interactions.

Like experiment, the seeming inconsistency between the strong binding affinity of ethanoate calculated here and the relatively minor presence of the amino-acids aspartate and glutamate in strong silica-binding sequences (Table 1.1) cannot be readily understood by previous computational studies (see Section 4.4 for a detailed discussion). Thus, similar to ammonium, in order to rule out the possibility that the findings from the PMF calculations carried out here were purely due to a deficiency of the FF used to model the silica interface, first-principles simulations of ethanoate adsorption at the aqueous (100) α-Quartz interface were conducted; the results of which are presented in Chapter 4.

3.4 Discussion

If the calculated adsorption strength of these amino-acid analogues is to be useful in the design of quartz-binding peptides with predictable properties, the lowest energy adsorbate-surface configurations found here must be taken into account. For most residues the alkyl content of the side-chains will be restrained away from the surface by the peptide backbone, unable to contribute to surface binding unless flanked by residues which orient it such that it has close surface contact. This is particularly important when extrapolating the adsorption free energies of ethanoate and butyl ammonium calculated in this study to those of aspartate/glutamate and lysine within a peptide, since the hydrophobic portions of these amino-acids would not be able to interact to the same degree with a α-Quartz surface. More faithfully to calculate the free energy of adsorption of charged (and polar) amino-acids onto surfaces in future, the possible adsorbed configurations of their analogues could be restricted. For example, for the lysine analogue (butyl ammonium), the terminal CH_3 group could be restrained at a minimum surface separation. However, in the context of understanding the role of post-translationally modified lysines in silaffin-silica adsorption, the length of the altered side-chain is substantial. Direct contact between these alkyl moieties and the surface could increase the binding affinity of silaffins, as well as aid their aggregation in solution; a role identified by previous work [Kroger et al. [2002]; Sumper and Brunner [2008]].

Alongside accounting for differences in the availability (for surface adsorption) of a given residue side-chain compared to its analogue, the influence of peptide conformation must also be considered. For example, at first sight these data suggest that a peptide rich in alanine might have lower affinity for the (100) surface than one enriched in tryptophan. However, if the bulkier nature of tryptophan restricts the conformation of the latter peptide such that fewer residues are able to make direct surface contact, then the combined adsorption of a larger number of alanines in the former sequence may result in it having higher overall surface affinity. The influence of peptide conformation on its surface adsorption is complex, and discussed more in Chapters 5, 7 and 8.

3.5 Conclusions

Potential of Mean constraint Force free energy calculations to quantify the adsorption strength of amino-acid analogues on α-quartz surfaces under aqueous conditions suggest the following trends in binding: non-polar aromatic > negatively-charged > non-polar aliphatic > positively-charged. Hydrophobic shielding by the surface, of both charged and uncharged molecules, had a major role in driving binding. This was exemplified by the more favourable free energy of adsorption of butyl ammonium than ammonium. In addition, closer analysis revealed that molecules with hydrophobic content tended preferentially to adopt configurations for which maximal hydrophobic screening by the surface could be achieved. All adsorbates showed energetic facet specificity in binding, and most (ammonium being the exception), displayed spatial and configurational surface discrimination. Differences in the preferred adsorbate orientation between the three surfaces were mainly due to surface topology rather than silanol density. Namely, molecules adsorbed on the (100) surface tended to adopt upright configurations with hydrophobic content penetrating interstitial spaces; while for (001) and (011) the lack of penetrable interstitial spaces meant that flat configurations relative to the interface were more frequently observed.

The two seemingly anomalous results–the weak/strong affinities of ammonium/ethanoate for the aqueous α-Quartz interface–are addressed in Chapter 4.

Chapter 4

First-Principles Molecular Dynamics Simulations of Ammonium and Ethanoate Adsorption at the Aqueous (100) α-Quartz Interface.

4.1 Introduction

The relative binding affinities of ammonium and ethanoate at the aqueous α-Quartz interface calculated in Chapter 3 were contrary to what might be expected based on the prevalence of positive and negatively charged residues, respectively, in silica-binding peptide sequences (Table 1.1). One explanation for these seemingly anomalous results could be a deficiency in the FF used to model silica-adsorbate interactions, LFF [Lopes et al. [2006]]; a possibility which is addressed in this chapter by CPMD simulations of the two amino-acid analogues adsorbed onto the (100) α-Quartz surface under aqueous conditions. The functional groups of ammonium and ethanoate–NH_3^+ and COO^-–are prevalent in peptides, featured by both peptide and amino-acid termini in solution and the side-chains of aspartate, glutamate and lysine residues at pH 7. It is therefore important for future FF-based MD, simulations employing LFF, that the surface adsorption of these two moieties can be reasonably reproduced.

The first-principles simulations reported herein were necessary to check the accuracy of LFF due to a lack of consensus in the literature concerning ammonium and carboxylate adsorption at aqueous silica interfaces (see Sections 1.3.1 and 3.3.5). For example, experimental studies to date have indicated that lysine binds both strongly [Churchill et al. [2004]; Kitadai et al. [2009]] and weakly [Zimmerman et al. [2004]] to silica. Previous theoretical studies are equally ambiguous and have only modelled the *co*-adsorption of the two functional groups–NH_3^+ and COO^-–at this interface. (By co-adsorption I refer to the surface binding of the N- and C-termini of the simple amino-acids glycine or alanine. In this case the connectivity of the two functional groups means that the affinity of one moiety for the interface cannot be probed in isolation from the other.) Specifically, again like experiment, these amino-acid–silica DFT calculations have suggested that either direct ammonium-silanol interactions are stronger [Nonella and Seeger [2008]] or weaker [Costa et al. [2008]; Zhao et al. [2011]; Folliet et. al. [2013]] than COO^--silanol interactions, and/or, that both types of interaction are indirectly mediated through interfacial waters [Nonella and Seeger [2008]; Rimola et al. [2009]].

4.2 Methods

Due to the computational expense of using CPMD to model an interface that includes liquid water, and to maintain a like-for-like comparison between first-principles and FF-based MD simulations, both were of modest spatial- and time-scales. Specifically, each system was comprised of 95 waters, a slab of α-Quartz ($Si_{52}O_{104}H_{16}$), 1 adsorbate (NH_4^+ or CH_3COO^-) and 1 counter-ion (Cl^-, and Na^+ (for FF-based MD ethanoate simulation) or H_3O^+ (for the CPMD ethanoate simulation)). As in Chapters 3, 5 and 7, adsorption onto the fully hydroxylated α termination of the (100) surface, featuring geminal silanol groups, has been modelled (Figure 1.2a). Counter-ions were inserted into the cell as far from the adsorbate as possible (equal to or greater than a distance of approximately 23 Å, based on the thickness of the inter-slab water layer–see below), adjacent to the hydrogenated quartz facet. They were not observed to interfere with or approach the relevant adsorbate at any point during the simulations. Simulations of 10 ps duration were carried out, with two runs for each adsorbate performed.

The same combination of FFs used in Chapter 3–LFF [Lopes et al. [2006]], CHARMM [MacKerell et al. [1998]] and modified TIP3P [Jorgensen et al. [1983b]; Neria et al. [1996]]–were employed in the FF-based MD simulations detailed below. (A comparison between CPMD and an alternative FF for silica, CLAYFF [Cygan et al. [2004]], was also carried out. The results of this study can be found in Reference [Wright and Walsh [2012b]].) Exact, direct comparison between the CPMD and FF-based MD approaches was impossible due to the different PEL of the DFT functional and FF; hence the latter simulations were set up in a couple of different ways (*e.g.* LFF-run1 and LFF-run2; see details below).

4.2.1 Car-Parrinello Molecular Dynamics Simulations

All CPMD simulations were carried out using the CPMD software package, version 3.13.2 [CPMD [1997-2001]]. The PBE functional [Perdew et al. [1996]], with a cut-off of 340 eV was used in combination with Vanderbilt ultrasoft pseudopotentials [Vanderbilt [1990]]. Prior to MD simulation, each simulation cell was geometry optimised under these conditions (to a threshold of 2.57×10^{-02} eV Å^{-1}). The resulting orthorhombic unit cells had dimensions of 10.710 × 12.470 × 40.501 Å^3 and 10.708 × 12.470 × 40.900 Å^3, for ammonium and ethanoate respectively. A total of 95 water molecules were inserted into the space between the slab and its periodic image in the z-direction, with a resulting inter-slab water layer thickness of ~23 Å. This water layer was of sufficient thickness to minimise perturbations in the water density at the top surface of the slab arising from the presence of the adjacent periodic image of the slab in the z-direction. The water layer thickness was comparable to that used in recent first-principles simulations of the quartz $(10\bar{1}1)$-water interface reported by Skelton *et al.* [Skelton et al. [2011a]], and in excess of that employed in recent first-principles simulations of the interface between the quartz (0001) surface and liquid water [Sulpizi et. al. [2012]; Gaigeot et al. [2012]].

All atoms were free to move in the subsequent MD simulations, which were conducted in the NVT ensemble and at a temperature of 300 K using a Nosé-Hoover thermostat. Although deviating from experimental conditions (which would be better represented by the isobaric-isothermal ensemble), the use of the NVT ensemble was a practical necessity–an approximation routinely used in first-principles simulations of mineral–water interfaces [Nonella and Seeger [2008]; Zhao et al. [2011]; Skelton et al. [2011a]; Tilocca and Cormack [2011]; Sulpizi et. al. [2012]]. A time-step of 0.09976 fs and a fictitious electron mass of 0.219 a.m.u. were used, and k-point sampling included the Γ point only. For both ammonium and ethanoate, two independent CPMD simulations (denoted 'CPMD run-1' and 'CPMD run-2') were initiated from the same starting coordinates (with the centre of mass of ammonium and ethanoate located 1.818 and 2.485 Å above the quartz surface respectively), making a total of 20 ps of dynamics for each adsorbate, and 40 ps of CPMD simulation time in total. Only one adsorbate-surface initial configuration, per adsorbate, was used in this study, due to the extreme computational expense of the CPMD simulations. While the initial configuration between the two runs for each adsorbate was the same, different initial velocities were assigned to the atoms in runs 1 and 2.

Although using only one starting configuration per adsorbate may impose limitations on the interpretation of the results, exhaustive adsorbate-surface conformational sampling is currently not computationally viable using CPMD. However, in comparison with previous work modelling the adsorption of whole amino-acids [Nonella and Seeger [2008]; Costa et al. [2008]; Zhao et al. [2011]; Folliet et. al. [2013]], the ability of each functional group to explore different surface binding geometries *during* a simulation is not limited here by the added complexity of the connectivity to the other terminus. To probe the dependence on initial conditions, the lateral motion of each adsorbate across the surface was monitored during the CPMD simulations. These results showed that neither adsorbate remained laterally trapped in its initial geometry or location (Figure B.1).

4.2.2 Force-field MD Simulations

Although LFF was extensively parametrised following the CHARMM methodology [Lopes et al. [2006]] and contains both bonding and non bonding parameters for silica atoms, discrepancies between the dimensions of a fully-flexible quartz slab modelled with LFF and that modelled with first principles techniques exist. For this reason direct, like-for-like comparison between LFF and CPMD (in terms of the relaxation state of the initial co-ordinates) is not possible. LFF simulations using both the PBE (run-1) and LFF (run-2) optimised geometries for quartz were therefore carried out. In LFF run-1 silica atoms were restrained to their initial PBE positions for the entire simulation, whilst in run 2 they were free to move within the constraints of the LFF potential. The initial adsorbate-quartz configurations in run-1 were identical to those in the CPMD trajectories, whilst for run-2 the CPMD adsorbate-quartz conformations were relaxed using LFF-CHARMM potential prior to initiation.

All LFF simulations were carried out using GROMACS 4.0.7 in the NVT ensemble (to maintain as much similarity with CPMD as possible). Identical to the CPMD simulations, the temperature was maintained at 300 K using a Nosé-Hoover thermostat and a time-step of 0.1 fs was used. Cut-offs for non-bonded LJ and PME electrostatic interactions were 5 Å. The unusually small cut-off distances arose due to the small system size of the CPMD simulations. To investigate whether the finite size, together with the periodic boundary conditions employed in both the LFF and CPMD simulations, introduced artifacts in to the results, larger scale simulations (2×2 super-cell of the original CPMD-sized cell) were performed. In the spatially extended runs a cut-off distance of 10 Å was used. Since comprehensive analysis revealed that only negligible differences existed between the large- and small-scale simulations, only the results from the latter set (those of comparable size to CPMD) are presented herein. Water was modelled using the modified TIP3P potential (herein referred to as TIP3P), with the O–H bond lengths restrained by the SETTLE algorithm [Miyamoto and Kollman [1992]]. The same CHARMM parameters used for ammonium and ethanoate in Chapter 3 were employed in this study.

4.2.3 Analysis

As a measure of adsorption, both the height of the adsorbate centre of mass (*c.o.m.*) in the z direction relative to the top of the quartz surface and the shortest possible 'hydrogen-bond like' distance (either $\mathrm{NH}\cdots\mathrm{O}_{Si}$ or $\mathrm{COO}^{-}\cdots\mathrm{H}_{Osi}$) were monitored as a function of time in all simulations. This 'hydrogen-bond like' distance is not intended formally to describe a hydrogen-bond (see below), but was included for the purposes of comparison with previous simulation studies [Zhao et al. [2011]].

In addition to the 'hydrogen-bond like' distance described above, hydrogen-bonds between the surface/water and the adsorbate were formally defined according to Section 2.5.1. The specific distance metrics used (Table B.1) were derived from adsorbate-water RDFs (Figure B.2). The dependence of the results on these RDF-based geometric criteria, instead of the 'standard' atom-atom values for typical hydrogen bonds in TIP3P liquid water ($\mathrm{O}\cdots\mathrm{O} < 3.5\,\text{Å}$, $\mathrm{O}\cdots\mathrm{HO} < 2.45\,\text{Å}$ and angle of 30°), was tested; the difference between the two sets of results was negligible and so only the results based on the first (RDF-based) criteria are reported herein. The characteristic hydrogen-bond lifetime, τ_{HB}, of both adsorbate–silica and adsorbate–water hydrogen-bonds was determined from $S(t)$, the continuous hydrogen-bond time autocorrelation function (Section 2.5.1) using the method outlined by Rosenfeld and Schmuttenmaer [Rosenfeld and Schmuttenmaer [2011]]. Namely, for a hydrogen-bond for which $S(t) \neq 0$ at $t = 4\,\mathrm{ps}$ then:

$$\tau_{HB} = \frac{-t_{\delta}}{lnS(t_{\delta})} \tag{4.1}$$

where t_{δ} is the longest time period over which $S(t)$ was calculated (4 ps). For less persistent hydrogen-bonds, where $S(t) = 0$ at $t = 4\,\mathrm{ps}$, τ_{HB} could be determined by numerically integrating $S(t)$ over the time period investigated:

$$\tau_{HB} = \int_{0}^{\infty} S(t)\,\mathrm{d}t \tag{4.2}$$

4.3 Results

4.3.1 Ethanoate

Ethanoate remains very close to the α-Quartz surface in all runs, both CPMD and LFF, for the whole timescale probed (Figures 4.1a, 4.2a). For both CPMD runs, the ethanoate centre of mass remains within 3 Å from the surface, maintaining a (CO)$O^{-}\cdots$HOSi separation of 1-2 Å throughout the entire trajectory. For the LFF simulations, the ethanoate-surface separation and (CO)$O^{-}\cdots$HOSi distance were slightly greater. These discrepancies, however, were small and arise in part from differences in the dimensions of the quartz slab for the two approaches (Section 4.2.2).

In Chapter 3 the optimal surface-separation of ethanoate bound to α-Quartz was found to be 3.70 Å with an additional local minimum present in the adsorption free energy profile at 2.95 Å [Wright and Walsh [2012a]]. Whilst it is inappropriate to compare these adsorption distances to the CPMD trajectories (due to differences in between PBE and LFF in the description of the PEL) it is noted that in the case of the LFF simulations, ethanoate remains kinetically-trapped within the local minimum on the timescales probed. The slow diffusion of the adsorbate away from the interface, if it was merely an "indifferent" binder to α-Quartz, could thus potentially be overseen by the short MD runs. Hence it was necessary formally to quantify adsorbate-surface hydrogen-bonding to help discriminate between the two states: "strong" and "indifferent" adsorption. (The short 'hydrogen-bond like' distance of ethanoate from the surface does not preclude non-linear geometries, indicative of the latter state.)

Careful analysis revealed, however, that extensive hydrogen-bonding between ethanoate and the surface was in fact responsible for driving persistent close contact. Within error, approximately two ethanoate-silanol hydrogen-bonds were present, on average, per time frame during both the CPMD and LFF trajectories (Table B.2). Differences between the two approaches existed only in the overall hydration number of COO^{-}, with the total number of hydrogen-bonds made by ethanoate being ∼4 on average for CPMD and ∼6 for LFF. The over-structuring of water around COO^{-} using CHARMM relative to CPMD has been previously noted in other studies [Hugosson et al. [2006]; Liang and Walsh [2006, 2007]]. The average ethanoate-silanol and ethanoate-water hydrogen-bond lengths were comparable, within error, both between each other and between CPMD and LFF. Using the persistence of a hydrogen-bond as measure of strength, it is clear from the hydrogen-bond time-autocorrelation functions that ethanoate-silanol (eq) hydrogen

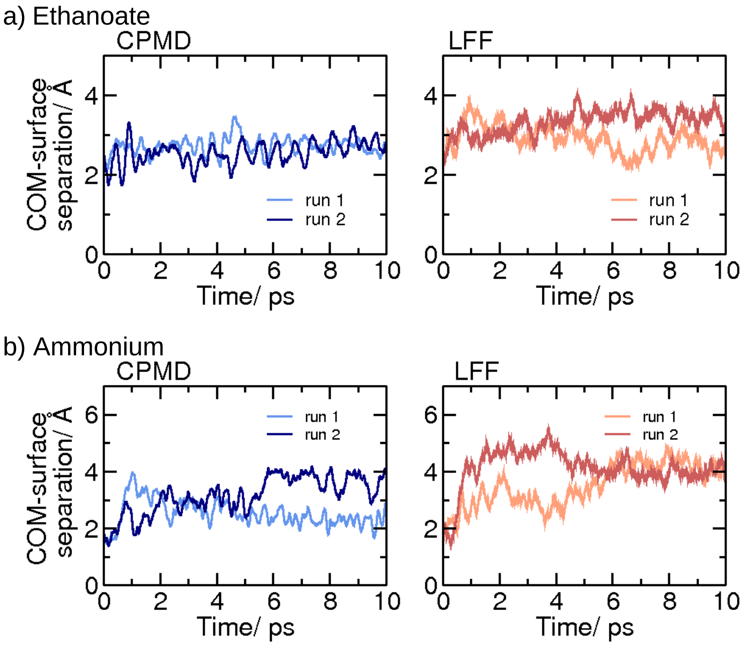


Figure 4.1: Vertical separation between the quartz surface and centre of mass of a) ethanoate, and b) ammonium, taken from the CPMD and FMD simulations.

bonds were significantly more persistent than ethanoate-water (ew) ones, with for CPMD τ_{HB}^{eq} being in excess of 2 ps whilst τ_{HB}^{ew} was an order of magnitude smaller, 0.1-0.2 ps (Figure 4.3a). Again, in terms of persistence, LFF is seen slightly to over-estimate ethanoate-water hydrogen-bonding relative to CPMD. Despite this, LFF can reasonably recover the main characteristics governing ethanoate adsorption onto α-Quartz relative to first-principles. In addition, these simulations suggest that ethanoate is strongly attracted to the (100) α-Quartz surface under aqueous conditions.

4.3.2 Ammonium

In sharp contrast to the strong adsorption of ethanoate, ammonium maintains close surface contact (defined as consistently sustaining a surface separation of less than 3 Å after an initial repulsion) for the entire timescale probed in only *one* out of the two CPMD trajectories (CPMD run-1) (Figure 4.1b). During CPMD-run2 the molecule moved to ∼4 Å away from the surface after 6 ps and did not return to make consistent surface contact for the remainder of the trajectory. LFF simulations in Chapter 3 revealed a (repulsive) local minimum in the PMF profile of ammonium adsorbed onto the (100) surface at a separation of 4 Å. Unlike ethanoate, the short timescale LFF simulations reported here for the more loosely bound ammonium corroborate this, with the ammonium-surface separation fluctuating around 4 Å in LFF runs 1 and 2. However, both LFF simulations display weaker ammonium-quartz binding than CPMD run-2, with desorption being almost immediate in the former, but not the latter.

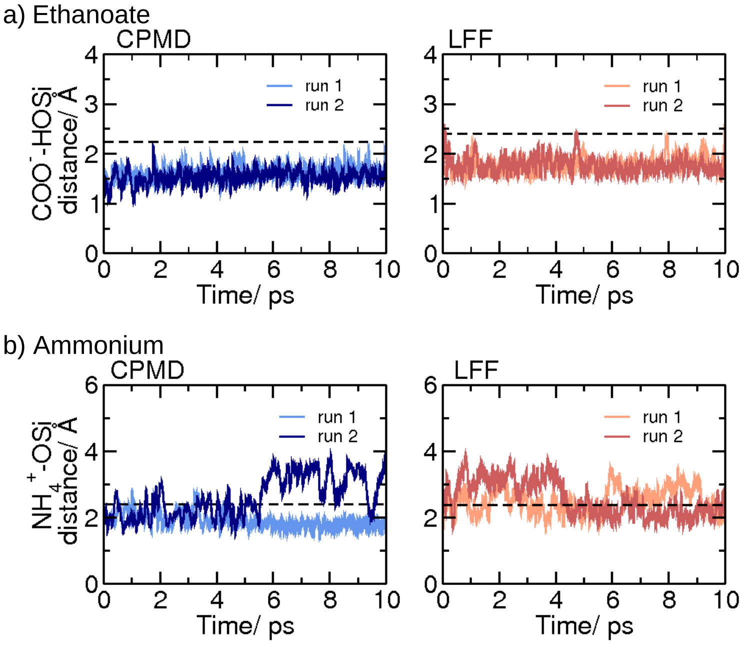
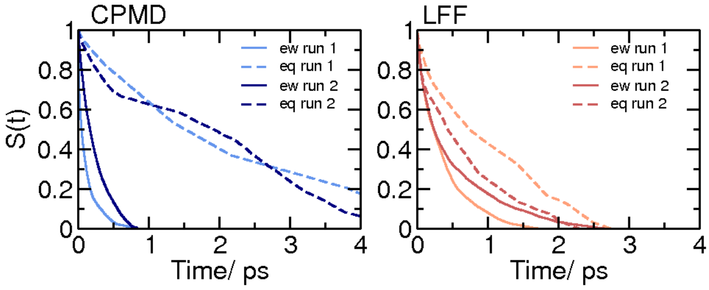


Figure 4.2: 'Hydrogen-bond like' distances between the quartz surface and a) ethanoate ((CO)$O^{-}\cdots$H(OSi) distance), and b) ammonium (NH$\cdots$O(H)Si distance), taken from the CPMD and LFF simulations. In both plots the black-dashed line denotes the RDF-determined cut-off used in the criteria to define each type of hydrogen-bond.

Consistent with trends in ammonium-surface separations, only for CPMD run-1 was significant adsorbate-surface hydrogen-bonding observed (Table B.3), with the trajectory average being 1.4 hydrogen-bonds per time frame. This was significantly greater than LFF simulations, in which the average number of hydrogen-bonds per time frame fell in the range of 0.0-0.4. The total number of hydrogen-bonds made by ammonium (to either the surface or water) and the lengths of these bonds on average were found to be approximately 3 and 1.9 Å respectively for both CPMD and LFF simulations. This suggests that, in contrast to ethanoate, ammonium was hydrated to a similar degree in each approach. Persistence of both water- and silanol- ammonium hydrogen-bonding, qualified by their continuous hydrogen-bond time-autocorrelation functions, is shown in Figure 4.3b. These data reveal that not only were there more hydrogen-bonds made between the surface and ammonium during CPMD simulations than LFF on average, but that they were also significantly more persistent. Water-ammonium hydrogen bonds were also slightly more persistent in the case of CPMD than LFF. However, like ethanoate, LFF correctly recovers the trend in the persistance between the two types of hydrogen-bond (surface and water) relative to first-principles.

a) Ethanoate



b) Ammonium

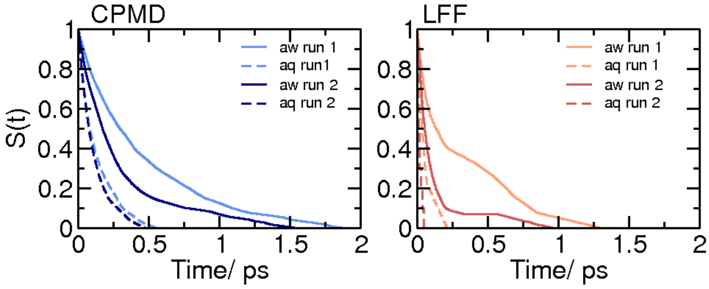


Figure 4.3: Continuous hydrogen-bond time-autocorrelation functions for: a) ethanoate–quartz (eq) and ethanoate–water (ew) hydrogen-bonds, and, b) ammonium–quartz (aq) and ammonium–water (aw) hydrogen-bonds, taken from the CPMD and LFF simulations.

Closer analysis of ammonium-silanol hydrogen-bonding featured in CPMD run-1, where ammonium maintains the most consistent surface contact, indicates that even here ammonium-silanol hydrogen bonds were relatively weak. Although in this run the adsorbate remains within hydrogen-bonding distance of the surface at all times, its adsorbed N–H$\cdots$O$_{Si}$ geometry is not always linear, meaning that the actual number of formal interfacial hydrogen-bonds formed was much less than that which might be expected from inspection of Figure 4.2b. This highlights the need for more than one distance criterion in the geometric definition of a hydrogen-bond. Hence, considering all the evidence from both CPMD runs, the short ammonium-surface separation in CPMD run-1 was more likely due to the “indifferent” nature of ammonium-quartz interaction than the adsorbate being proactively attracted to the aqueous interface. Ideally, this hypothesis could be tested by performing a series of much longer CPMD simulations, starting from a wide number of different initial configurations and surface separations. However, such a study is outside the scope of what is currently feasible computationally.

Although these simulations suggest that LFF underestimates the attraction between ammonium and the (100) quartz surface relative to CPMD, the FF still captures most of the essential features of the ammonium-quartz interaction. First, both CPMD runs indicate a lack of compelling driving force for ammonium to remain adsorbed at the interface. This is mimicked by LFF, although the LFF ammonium-quartz interaction appears to be slightly repulsive, rather than merely "indifferent". Second, water-ammonium hydrogen-bonds are more persistent on average than silanol-ammonium hydrogen-bonds for both CPMD and LFF. The characteristics of these hydrogen-bonds are pivotal to the description of ammonium-quartz adsorption; that LFF reproduces the CPMD trend for ammonium-quartz hydrogen bonds to be weaker than that of water-ammonium is essential if the FF is to be used in future biointerfacial studies.

4.4 Discussion

The first-principles simulations reported in this Chapter tally with a number of other studies; specifically: the very recent work of Folliet *et al.* [Folliet et. al. [2013]] who investigated the adsorption of glycine to two different sites (vicinal and nested geminal) on an amorphous silica surface under microsolvated conditions using PBE with the DFT-D2 dispersion correction [Grimme [2006]]; Zhao *et al.* [Zhao et al. [2011]] who used MD to model the adsorption of zwitterionic glycine in liquid water on the (100) edingtonite surface with a self-consistent charge density functional tight-binding approach; and Costa *et al.* [Costa et al. [2008]] who studied the adsorption of microsolvated zwitterionic glycine onto an amorphous silica surface using first-priniciples MD. In the first, published after the work reported here [Wright and Walsh [2012b]], short CPMD simulations in the NVE ensemble were carried out to identify conformations of glycine at the interface which were stable for 0.5 ps or more. These glycine-silica geometries were subsequently optimised in the absence and presence of water (up to 3 molecules) in order to identify the lowest energy state of the system. The authors found that in its most favourable conformation, the C-terminus of glycine was directly hydrogen-bonded to a silanol group while a water molecule bridged between the ammonium moiety and the surface. Consistent results, in terms of direct carboxylate- and water-mediated ammonium-silica binding, were reported in the earlier MD study [Costa et al. [2008]] which employed fully solvated conditions. Similarly, the DFTB simulations of Zhao *et al.*, which were much longer in duration (20 ps compared to only 1.5 ps at most for those of Costa *et al.*), suggested that N-terminal contact between glycine and edingtonite was transient whilst carboxylate binding persisted for the entire length of their trajectories [Zhao et al. [2011]]. In this latter work a single distance criterion (equivalent to the 'hydrogen-bond like' distance monitored in this Chapter) was used to identify adsorbate-surface hydrogen bonding, making it difficult to establish the exact nature ("strong", "indifferent" or "weak") of the interaction between the N-terminus of glycine and the interface.

Only a first-principles MD study modelling the adsorption of alanine to the hydroxylated (001) edingtonite surface [Nonella and Seeger [2008]] is at odds with the CPMD simulations carried out here (and the other works detailed above). In this last study a number of different first-principles simulations were run- each employing slightly different conditions. In the most relevant simulation with which to compare, a zwitterionic alanine molecule was initially placed with its centre of mass about 5.5 Å from the top of the surface (fully protonated, carrying no overall charge). The system was simulated for 15.5 ps, the first 2.5 ps equilibration period being conducted at 100 K before the temperature was raised to 300 K. Neither the carboxylate nor ammonium groups made direct contact with the surface during this time period. However, diffusion of alanine to the interface and its subsequent reorientation to facilitate strong binding would not necessarily occur on the timescale probed–most likely the reason for the discrepancy between the conclusions of Nonella and Seeger and other works. Additionally, when making comparison between the CPMD simulations carried out here to the other reports mentioned above, the connectivity of the N- and C-termini in an amino-acid makes it possible that the strong adsorption of one terminus (*eg.* the carboxylate group) inhibited adsorption of the other by restricting the range of functional group-surface geometries accessible.

The stronger affinity of ethanoate to the (100) α-Quartz surface compared to that of ammonium may arise, not only from the opposing charge states of the two adsorbates, but also the differing adsorbate geometries. In particular, the spatial match between surface silanols and hydrogen bond donor or acceptor atoms in each molecule impacts on the preferred adsorbate-surface binding mode and hence may also limit the number of adsorbate-surface hydrogen bonds that can form. The CPMD trajectories chiefly featured ethanoate bound to the surface in a bidentate manner, bridging two silanol groups (not part of the same geminal pair) (Figure 4.4a). Around 87% and 70% of all ethanoate-surface hydrogen-bonds arose from

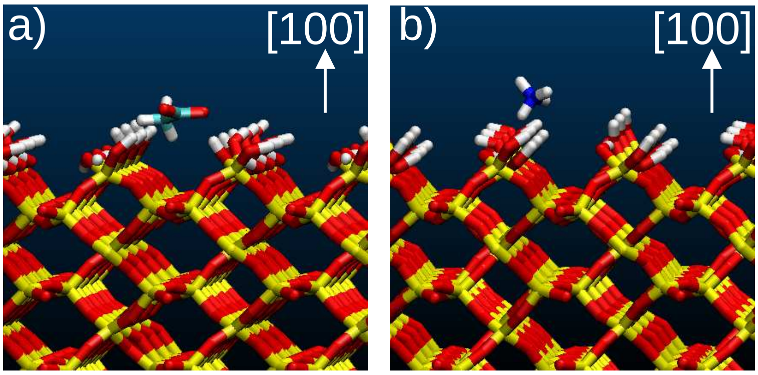


Figure 4.4: Snapshots of typical configurations of a) ethanoate and b) ammonium at the aqueous (100) α-quartz interface (taken from ethanoate CPMD-run1 and ammonium CPMD-run2 respectively). a) Shows ethanoate bound in a bidentate orientation whilst in b) ammonium is bound monodentate to the surface. Silicon atoms are shown as yellow, oxygen as red, hydrogen as white, nitrogen as blue and carbon as green. Water is omitted for clarity.

this bidentate geometry in CPMD run-1 and CPMD run-2, respectively. Although a similar bidentate geometry was possible for ammonium adsorption, ammonium was also observed to bind in a monodentate fashion for a significant proportion of the time (Figure 4.4b). Consistent with differences in surface separation between the two CPMD trajectories in the case of ammonium, the preferred mode of binding varied between runs. Overall, differences in the hydrogen-bonded configurations of ammonium and ethanoate are a consequence of the longer $O\cdots O$ span of a COO^- group being more comensurate with (100) α-quartz silanol-silanol spacings, compared to the corresponding $H\cdots H$ span of NH_4^+. The specific mode of quartz-binding by each functional group could therefore differ between crystallographic planes depending on surface topology. In addition, at the (100) interface a detailed comparison of adsorbate binding modes between CPMD and LFF however, is not appropriate due to subtle differences in the geometries of the underlying silica surfaces used for both approaches.

Finally, it is worth considering possible artifacts that might have been introduced into the CPMD simulations carried out in this thesis by use of the PBE exchange correlation functional. As noted in Section 2.1.1, deficiencies in the description of dispersion interactions by PBE are well known. However, Skelton *et al.* compared first-principles simulations of the (100) α-Quartz interface to experimental X-ray data, finding good agreement between the two [Skelton et al. [2011b]]. To date, equivalent benchmarks of dispersion corrected functionals at this interface have not been carried out and the use of such functionals for MD simulation of aqueous interfaces remains rare in the literature. Furthermore, static calculations of glycine adsorption onto Cr_2O_3 [Garrain et. al. [2011]] and silica [Folliet et. al. [2013]] surfaces performed with and without dispersion corrections have shown that dispersion only altered the absolute binding strength, not the trend in the relative affinities.

4.5 Conclusions

In summary, CPMD simulations, the first to model the adsorption of charged biomolecule fragments at the aqueous (100) α-Quartz interface, indicate that ethanoate has a strong affinity for the surface. On the other hand, the evidence suggested that ammonium-quartz interactions were more "indifferent" in nature, being neither strongly attractive nor repulsive. The relative adsorption strengths of the two molecules was inferred by not only monitoring binding distances but also by quantifying adsorbate-silanol hydrogen bonding. In particular, ethanoate-silanol hydrogen bonding was found to be both more extensive and more persistent than that of ammonium-silanol. These results corroborate those presented in Chapter 3 suggesting that the seemingly anomalous finding that the more favourable free energy of adsorption at the aqueous α-Quartz interface of ethanoate than ammonium was not due to the FF used. In fact, in this Chapter it has been shown that, relative to CPMD, LFF was able to model the key features of ammonium and ethanoate-quartz adsorption. For ammonium however, this work also suggests that the LFF description of adsorbate-quartz interactions is slightly too repulsive in nature and thus there is scope for future improvements to the potential.

Chapter 5

Benchmarking Replica Exchange with Solute Tempering for Bio-interfacial Systems: QBP-1 Adsorption onto the (100) α-Quartz Surface.

5.1 Introduction

In addition to the direct contribution of each residue within a peptide to contribute to its overall affinity for an interface, the role of peptide conformation, also imparted by its primary sequence, must be considered (Section 3.4, [Skelton et al. [2009]; Verde et al. [2011]; Feng et al. [2012]]). Use of regular MD to identify trends in the conformations of peptides with strong affinity for a given interface is hampered however by multiple minima which exist in the PEL of these large complex systems (Section 2.3). While experimental techniques to probe interfacial peptide conformation remain in their infancy (Section 1.2.3), it is important to investigate the suitability of advanced sampling simulation methods, designed for solution-based biomolecular studies, to the bio-interfacial regime. In theory the structural data required to aid the rational design of strong material-binding peptide sequences able to discriminate between either substrates of different composition or those presenting a different crystallographic plane of the same material could be obtained from such studies.

In this Chapter, the performance of two methods based on REMD–T-REMD

(Section 2.3.1) and REST (Section 2.3.3)–in terms of sampling different adsorbed conformations of a peptide at an aqueous inorganic interface are evaluated. The aim of the work is to identify a REMD variant which maximises the amount of peptide phase space sampled for a given computational resource. At first glance, the reduced number of replicas required for a REST simulation compared to conventional T-REMD makes it appear favourable. However, without a study such as this to benchmark the two techniques, the possibility that the results of the former method, REST, are in some way compromised can not be ruled out. In addition to comparing REST with T-REMD, the influence of the initial replica population on the results and of combining of solvent/side-chain mass scaling [Bennett [1975]; Lin and Tuckerman [2010]] with REMD are investigated here. First proposed by Bennett in 1975, scaling of atomic masses can speed up/slow down low/high frequency dynamics without altering the equilibrium properties of the system [Bennett [1975]]. More recently this concept has been implemented by Lin and Tuckerman to enhance the conformational sampling of a peptide under aqueous conditions by scaling solvent and side-chain atomic masses [Lin and Tuckerman [2010]]. The concept of combining mass scaling with T-REMD and REST to a bio-interfacial system had not been used prior to this work [Wright and Walsh [2013]].

Adsorption of the QBP-1 peptide [Oren et al. [2007]] (also referred to as 'S1') onto the (100) α-quartz surface was used as exemplar system for this study. Previously reported MD [Oren et al. [2010]] and T-REMD [Notman et al. [2010]] simulations of QBP-1 adsorbed at the same interface and in solution, respectively, provide theoretical benchmarks against which the REMD-based simulations carried out in this Chapter can be compared. In addition, the physical reliability of the results can be interrogated by comparison with experimental solution CD structural and motif enrichment data [Oren et al. [2010]]; the latter derived from bioinformatics approaches [Oren et al. [2010]].

5.2 Methods

5.2.1 System Details

All variants of the Replica Exchange Molecular Dynamics (REMD) simulations carried out here have modelled the quartz binding peptide QBP-1 (PPPWLPYMPPWS) [Oren et al. [2007]] adsorbed at the fully hydroxylated aqueous (100) α-quartz interface (Figure 1.2a) using the Gromacs (version 4 and above) software package [Hess et al. [2008]]. As throughout this thesis, the α termination of the (100) surface, which features geminal silanols of surface density $0.0588\,\text{Å}^{-2}$, has been used.

Here the quartz slab had dimensions 74.6 x 65.6 x 21.3 Å^3. A total of 7300 water molecules were inserted between neighbouring slabs in the z-direction with periodic boundary conditions being applied to all three dimensions. All quartz atoms were free to move within the simulations, whilst the O-H bond length of water was constrained using the SETTLE algorithm [Miyamoto and Kollman [1992]].

Simulations were performed in the NPT ensemble at a pressure of 1 bar coupled to a Parrinello-Rahman barostat (in the z dimension only), with the desired temperature of each replica being maintained using a Nosé-Hoover thermostat. Newton's equations of motions were solved using the leapfrog algorithm [Hockney et al. [1974]], with an integration time step of either 0.5 fs (for simulations involving mass scaling) or 1 fs (for all other simulations). PME electrostatic summation was cut-off at 13 Å, whereas a force-switched cut-off starting at 10 Å and ending at 12 Å was used for LJ non-bonded interactions.

Like in Chapters 3 and 4, silica was modelled using LFF [Lopes et al. [2006]], while CHARMM27 [MacKerell et al. [1998, 2004]] and the modified TIP3P potential [Jorgensen et al. [1983b]; Neria et al. [1996]] were used to model the peptide and water, respectively. Although the first-principles simulations reported in Chapter 4 and reference [Wright and Walsh [2012b]] suggested that LFF slightly underestimates the strength of attraction between ammonium functional groups and α-Quartz, it was thought to be the most appropriate biointerfacial silica FF for studying QBP-1–quartz adsorption at the time at which the simulations reported here were performed. QBP-1 is a highly hydrophobic sequence featuring only one ammonium group at the N-terminus.

5.2.2 Temperature-REMD

T-REMD simulations were conducted by simulating a number of independent replicas in parallel, each at a different temperature. A total of 36 replicas were used in each simulation, at temperatures ranging from 295.0 K to 392.29 K (295.00, 297.49, 300.00, 302.52, 305.06, 307.61, 310.18, 312.77, 315.37, 317.99, 320.63, 323.28, 325.95, 328.61, 331.31, 334.03, 336.77, 339.53, 342.31, 345.10, 347.91, 350.74, 353.58, 356.45, 359.32, 362.23, 365.15, 368.11, 371.07, 374.05, 377.04, 380.05, 383.09, 386.14, 389.22, 392.29). Temperatures were generated using the online temperature generator for T-REMD simulations [Patriksson and van der Spool [2008]]. Each replica was populated with a different initial peptide-surface configuration to maximise the amount of peptide conformational phase space sampled, and thus, prevent the system from being kinetically-trapped within a deep local minimum, as might occur if the same initial peptide configuration was used in each replica. Therefore, initial peptide conformations were chosen randomly to be one of three things: 1) QBP-1 structures specifically featuring the most common folded peptide secondary structure configurations constructed by hand (namely, α-helix, α-turns, β-sheet, β-turns or PPII), 2) the centroid structure of one of the four most dominant clusters of QBP-1 in solution [Notman et al. [2010]], or 3) snapshots taken from a 1 ns regular MD simulation of QBP-1 in aqueous solution at 400 K. Prior to T-REMD, each replica was equilibrated at its target temperature for 1 ns (with no exchange moves attempted). This set of equilibrated replicas was then used as the starting point for each of the different variants of T-REMD simulations subsequently conducted. Smaller subsets of these equilibrated replicas were used in the REST simulations (see below).

Exchanges were attempted between replicas of neighbouring temperatures every 1000 MD steps [Okabe et al. [2001]]. (Specifically, as implemented by Gromacs [Hess et al. [2008]], the probability of exchange between odd (1↔2, 3↔4, *etc*) and even (0↔1, 2↔3, *etc*) pairs of replicas was determined alternately with the frequency of exchange between any given pair (*e.g.* 0↔1) being 2000 MD steps.) Exchanges were accepted with an average probability of approximately 0.2. All T-REMD simulations were run for $15\text{x}10^6$ MD steps making a total simulation of 540×10^6 MD steps per method tested.

One T-REMD simulation using regular, unscaled, atomic masses was carried out to provide a benchmark against which the performance of all the other REMD-based methods could be based. Three additional T-REMD simulations were also conducted, each with different degrees of mass scaling. The scaling constants chosen were those reported by Lin and Tuckerman [Lin and Tuckerman [2010]]. In the first case, denoted herein as ‘light’, the masses of the water atoms were scaled by 0.1. The remaining two T-REMD simulations carried out combined mass scaling of the water atoms (0.1) with mass scaling of the peptide side-chain atoms, by constants of 0.6 and 0.8 respectively (herein denoted as ‘light sc 0.6’ and ‘light sc 0.8’).

5.2.3 Replica Exchange with Solute Tempering

A number of different REST simulations of QBP-1 adsorbed at the aqueous (100) α-Quartz interface were carried out to establish parameters which optimised the performance of this method. All simulations were implemented as described in Section 2.3.3. Each replica was simulated in the same NPT ensemble, at a temperature of 300 K. The effective temperatures of the 11 replicas (300.00, 307.61, 315.20, 323.61, 331.95, 341.32, 350.99, 361.05, 370.44, 379.59, 392.29 K) were derived iteratively. The highest ‘effective’ temperature was chosen to be 392.29 K, to make the effective and actual temperature windows of the REST and T-REMD simulations approximately equivalent. The number of replicas used, 11, was based on comparison with previous works in which the version of REST employed here was also used. Specifically, Terakawa *et al.*, modelling a 13 residue peptide [Terakawa et al. [2011]], and Wang *et al.*, simulating the peptide trpcage (20 residues) [Wang et al. [2011]], both used 10 replicas to span an effective temperature ranges of 300-600 K and 300-572 K, respectively. Both studies were of peptides in aqueous solution. The ratio of number of replicas to effective-temperature window was increased in this work to account for the added complexity of simulating an interfacial system. (The number of replicas and effective temperature range was increased in the REST simulations reported in Chapters 7 and 8 in order to optimise the method to the computational resource available.) Before initiating REST, starting configurations were equilibrated for 1 ns at their target potentials with no exchange moves attempted. Preliminary simulations, detailed in Reference [Wright and Walsh [2013]], indicated an optimal exchange attempt frequency for REST of 1000 MD steps. This was adopted for all simulations reported hereafter. In this Chapter, the average probability of an exchange attempt being accepted was approximately 0.7. Simulations were run for 10×10^6 MD steps.

In total 6 REST simulations were performed. Firstly, the influence of initial replica structure population on the performance of REST, in terms of conformational sampling, was investigated. The specific focus was on the inclusion of helical peptide structures and initial peptide-surface separation. Previous MD studies and solution-based CD measurements suggested that QBP-1 does not exhibit significant α-helix secondary structure [Oren et al. [2010]; Notman et al. [2010]]. However, the findings from these former simulations may have been biased by the absence of helices in the starting configurations tested due to the high potential energy barrier involved in α-helix formation. Conversely, preliminary results from the interfacial REST simulation performed here which included α-helices as part of the initial replica population (herein labelled ‘REST 1’) showed that such structures were persistent. The findings from this initial run could have arisen from a deficiency in the FF used to model the peptide, although the evidence in the literature supporting a helical bias in CHARMM27 is conflicting. For example, the benchmark biointerfacial study carried out by Latour and *co-workers* modelling the LKβ7 and LKα14 peptide adsorption onto SAMs reported no bias [Collier et al. [2012]], contrary to the results of solution-based simulations which suggested there was [Piana et al. [2011]]. Therefore, in order to investigate whether the presence/absence of α-helical structures of QBP-1 in this work was solely due to the initial replica population, a second REST simulation in which none of the starting configurations featured a helix was carried out (labelled ‘REST no h’). In a similar vein, visual inspection of the ‘REST 1’ simulation revealed that in some replicas QBP-1 remained strongly adsorbed to the surface, ‘locked’ in a particular conformation by penetration of certain residues into interstitial surface sites. To investigate the extent to which this hindered sampling, another REST simulation for which the initial replica population differed from ‘REST 1’ only in the displacement of the peptide further away from the quartz surface in two out of the 11 replicas, was carried out (labelled ‘REST 2 off’).

Secondly, the influence of combining mass scaling with REST was investigated for each of the three sets of initial replicas described above: ‘REST 1’ (labelled ‘REST light’), ‘REST no h’ (labelled ‘REST light no h’) and ‘REST 2 off’ (labelled ‘REST light 2 off’). Only scaling of the water atomic masses (by 0.1) was implemented during these simulations.

5.2.4 Analysis

In this Chapter detailed analysis of the constant-ensemble run at a(n) (effective) temperature of 300 K for each variant of the REMD simulations was carried out only. To assess how well each variant performed in terms of conformational sampling, two principal benchmarks were used: one based on clustering of the trajectory (Section 2.5.2), and the other based on peptide secondary structure (Section 2.6). For all analyses (unless stated otherwise), data from the last $4{\times}10^6$ MD steps was used only. In these later stages, each simulation was observed to have reached a steady state (Figure 5.1). It must be noted that since the aim of this work was to identify a REMD-based method which optimised conformational sampling of the peptide for a given CPU resource, throughout this Chapter MD steps, rather than time, was used as a monitor of simulation progression. This was done in order to place simulations employing ‘normal’ and ‘light’ water (which differ in terms of integration time step) on a level-footing.

A number of peptide-surface binding properties were also calculated, to allow comparison with previous simulation and experimental studies [Oren et al. [2010]; Notman et al. [2010]] of QBP-1. Hydrogen bonding (intra-peptide, peptide-quartz and peptide-water) was quantified using just two of the geometric criteria outlined in Section 2.5.1. Namely, a hydrogen bond was defined as being present on the simultaneous satisfaction of a heavy atom donor–acceptor distance of less than 3.5 Å and a hydrogen-donor-acceptor angle of less than 30°. For each simulation residues bound to the Quartz surface for structures belonging to the three most populated clusters were identified using a closest-atom residue-surface cut-off distance of 3 Å.

The structures themselves were identified by cluster analysis conducted over the final 4×10^6 MD steps of a simulation. Binding motifs–dyads and triads–were identified using the same closest atom-surface criterion applied to two/three sequential residues. In this case analysis was conducted over the ensemble of *all* peptide structures sampled in the last 2×10^6 MD steps of each trajectory, not only those belonging to the most stable clusters.

To investigate whether the most populated clusters identified by the REMD-based simulations represented low-energy peptide-quartz configurations, the relative potential energies of a number of systems were probed by carrying out several additional standard MD simulations of 5 ns duration. This was particularly important to help identify whether helical structures were in fact low-energy adsorbed states of QBP-1, or just an artefact of the initial replica population chosen. Hence, six different snapshots of QBP-1 adsorbed onto the (100) α-quartz surface, representative of both extended, helical, and solution clusters 1 to 4 [Notman et al. [2010]], were taken from different REMD simulations. Each system was otherwise identical (*e.g.* in terms of periodic cell dimensions and chemical composition). Simulations were conducted in the NVT ensemble, at a temperature of 300 K. The average potential energy of the entire system, PE_{sys}, was calculated over the last 2 ns of the trajectory with a block averaging method used to estimate the errors [Hess [2002]] in each case. All results were reported relative to the lowest energy configuration (extended, *ext*):

$$\Delta PE_{sys-ext} = PE_{sys} - PE_{ext} \tag{5.1}$$

where sys = helix, Sol-1, Sol-2, Sol-3 or Sol-4.

5.3 Results

5.3.1 Replica Mobility

For any T/H-REMD simulation efficiently to sample as much of peptide conformational space as possible, all replicas must be mobile throughout the entire temperature (potential) window. This is one of the main prohibiting factors for the use of T-REMD to simulate large systems, since replica mobility is reduced if one increases the number of replicas used, and/or decreases the exchange acceptance probability. Long total simulation times are therefore required for a given replica, initially at the lowest temperature, to migrate through the temperature ladder. The version of REST used here, was reported to show improvement in this regard for test cases of biomolecules in solution [Wang et al. [2011]]. However, the effect on replica mobility of increasing system complexity by adding a mineral interface onto which the peptide adsorbs, for both T-REMD and REST methods, had not been considered prior to this work.

Representative mobilities of four replicas (lowest, highest and two intermediate (effective) temperatures), taken from the simulations, of for each of the different variants of T-REMD and REST are shown in Figures C.1 and C.2, respectively. As expected, the extent of replica mobility for T-REMD is much less than that for REST, but in both cases replicas initially at the lowest/highest (effective) temperature were able to visit all the (effective) temperatures within the simulation time.

5.3.2 Clustering Analysis

The number of clusters representing different backbone structures of QBP-1 as a function of MD steps for each of the REMD-based simulations performed in this Chapter is shown in Figure 5.1. The total number of unique peptide structures sampled by QBP-1 in all T-REMD simulations of the peptide adsorbed at the aqueous α-Quartz interface was less than that found by REMD-based simulations of QBP-1 in solution (37-41 clusters for each of the T-REMD methods used here, compared to both 61 for the previous T-REMD study [Notman et al. [2010]] and 106 for the solution-based REST QBP-1 simulation reported in Chapter 7, all after 10×10^6 MD steps). Taken together with the fact that the number of clusters found in the interfacial simulations appears to have plateaued after 15×10^6 MD steps, these results suggest that upon adsorption to quartz the conformation of QBP-1 is more restricted than it is in solution, as might be expected.

When comparing the number of clusters found by the different variants of T-REMD used, the influence of mass scaling on these results appears to be slight, with the performance of each effectively equivalent to that of the normal T-REMD simulation. However, the slight trend in these 'light' T-REMD results was in agreement with Lin and Tuckerman [Lin and Tuckerman [2010]] with optimal conformational sampling achieved over the majority of the trajectory by the simulation in which mass scaling parameters of 0.1 for water and 0.6 for side-chain atoms were used. Previous work has shown that QBP-1 is an exceptionally stiff peptide due to its high proline content [Notman et al. [2010]]. Hence it might be the case that the ability of mass scaling significantly to enhance conformational sampling of the system was impeded by the rigid nature of the peptide studied here. This does not rule out the possibility that combining the technique with REMD for more flexible sequences in the future may yield more positive results.

Fewer clusters were found by the REST simulations than the T-REMD ones within the same number of MD steps (Figure 5.1b). This was due to the reduced number of replicas used in the REST method than T-REMD, effectively seeding the REST simulations with fewer clusters. It is noted, however, that the total number of clusters found by a method is not necessarily an absolute measure of its peptide conformational sampling performance, *e.g.* if only large numbers of sparsely populated (minority), high-energy clusters are being found and the low energy adsorbed states of the peptide are not being adequately sampled. For this reason, both the total number and number of 'significant' (see Section 2.5.2) backbone clusters over the final stages of each simulation, after the systems had equilibrated, were calculated (Table 5.1). Using these refined metrics of conformational sampling performance, the difference between T-REMD and REST was less pronounced. Hence, whilst clustering analysis can be a useful tool for the comparison of the different T-REMD and REST variants, it is of limited use as a stand-alone assessment of performance and should be used with care.

method	total	signif.	1 /(%)	2 /(%)	3 /(%)
T-REMD normal	21	12	31.7	19.9	14.3
T-REMD light water	21	11	45.1	11.2	9.9
T-REMD light 0.6 sc	19	11	33.5	12.9	10.9
T-REMD light 0.8 sc	22	11	21.1	20.0	17.8
REST 1	15	6	44.7	15.8	10.4
REST no h	24	10	21.9	19.2	16.7
REST 2 off	18	9	29.0	13.8	13.8
REST light	18	9	26.2	21.6	18.7
REST light no h	21	8	31.5	30.2	8.9
REST light 2 off	16	7	34.9	18.8	17.5

Table 5.1: Total number and number of 'significant' clusters identified for each REMD-based simulation. (Cluster analysis carried out over the last 4×10^6 MD steps only.) The % of total structure population belonging to each of the three most populated clusters identified over the same time period is also recorded.

As with T-REMD, data pointing to the benefits of combining mass scaling with REST are inconclusive here, with the results being highly dependent on the initial replica population. As anticipated, the variant 'REST 2 off', in which the peptide was displaced from the quartz surface initially in 2 out of the 11 replicas, showed slightly enhanced conformational sampling, in terms of the number of clusters found, compared to 'REST 1'. This confirms the hypothesis that for some peptide-surface geometries, strong interfacial adsorption can hinder conformational sampling of the peptide. It is anticipated that this phenomenon would be even more

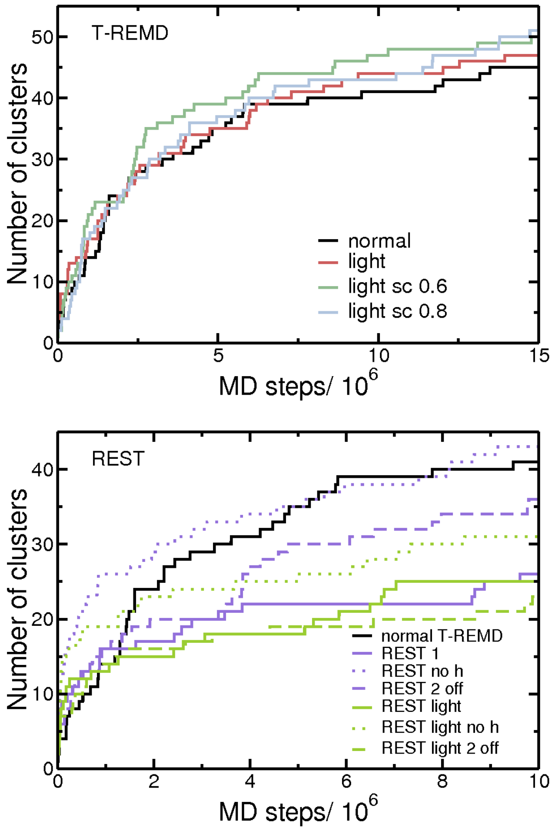


Figure 5.1: Number of clusters of QBP-1 backbone structure found from a) T-REMD and b) REST simulations as a function of MD steps.

pronounced for simulations of other material-binding peptides at their respective interfaces, since in comparison the interactions between quartz-binding peptides and quartz are thought to be relatively weak [Seker and Demir [2011]]. Interestingly, however, combining mass scaling of water with displacing the peptide from the surface in two replicas ('REST light 2 off') was not so favourable. In these simulations the light water, with reduced viscosity, enhanced the rate of peptide diffusion to the quartz surface. The time taken for peptide adsorption was just $4{\times}10^6$ MD steps in both replicas initially displaced from the interface in the 'REST light 2 off' simulation, compared to more than twice as long in the 'REST 2 off' one (employing 'normal' water). Hence, for 'REST light 2 off', the benefits (defined in terms of conformational sampling) of displacing the peptide from the surface initially in two replicas were cancelled out, in this case, by the reduced viscosity of the 'light water'.

Omission of helices from the initial replica population ('REST no h') enhanced the number of clusters found by the REST method to a greater degree than peptide displacement ('REST 2 off'). In fact, the total number of clusters found by 'REST no h' after $10{\times}10^6$ MD steps was roughly equivalent to that found by normal T-REMD after the same number of MD steps, despite the former being seeded with less than 1/3rd of the number of initial structures. However, in light of previous conflicting reports of a possible helical bias in CHARMM27 (Section 5.2.3), even though the omission of helices may increase the total number of clusters identified and enhance peptide conformational sampling, it still may be still essential to incorporate them in the initial replica population of a T-REMD or REST simulation fully to capture the physical characteristics of the system. To probe this, the relative potential energies of systems containing extended and helical forms of QBP-1 adsorbed onto the (100) surface were calculated (Section 5.3.6). Like 'REST 2 off', use of 'light water' with 'REST no h' ('REST light no h') did not improve conformational sampling, in terms of the number of clusters identified.

5.3.3 Peptide Structures and Ramachandran Analysis

Extended structures of QBP-1 dominated the three most populated clusters identified during backbone clustering analysis over the last $4{\times}10^6$ MD steps, for each of the REMD-based methods used. RMSD comparison between the centroid structures of each cluster from the different variants quantified visual inspection that indeed very similar structures were being frequently sampled in all of the T-REMD and REST simulations. This verified that none of the variations to the standard T-REMD procedure employed here had a significantly biasing effect. The three most populated clusters, which taken together typically encompassed up to 70% of the total population of structures (Table 5.1), could be broadly described as ‘fully extended’, ‘extended with kinks’ and ‘C-shaped’ (Figure 5.2). Helical structures of QBP-1 on adsorption (Figure 5.2d) formed minority clusters in most variants (accounting for 10% or less of the total structure population in the latter stages of the simulations (Table C.1)) apart from ‘REST 2 off’, ‘REST light’ and ‘REST light 2 off’. In these three simulations, helical clusters (defined as a cluster for which the RMSD of backbone atom positions of the centroid structure to that of an ideal QBP-1 α-helix was less than 2 Å) formed the 2nd or 3rd most significant clusters in analysis carried out either over the entire trajectory or only over final $4{\times}10^6$ MD, steps and accounted for up to 24% of the total structure population. Not only this, but their populations were stable with time, neither significantly growing nor decreasing in size as a proportion of the total number of structures (Figure C.3). The evidence is suggestive that the presence of helical clusters in this subset of REST simulations was not just a consequence of the initial conditions used. Instead, it indicates that these three variants of REST, all of which incorporated helical structures in the initial population of peptide structures, were either more or less accurately capturing the physical characteristics of the system; an issue addressed by energetic analysis in Section 5.3.6.

As well as there being structural overlap between structures in the three most populated clusters in each of the REMD-based methods, there was also structural overlap between the findings from these interfacial simulation and the four most populated clusters of QBP-1 in solution, identified by a previous study [Notman et al. [2010]] (Table C.2). Structures similar to the extended solution clusters 2 and 4 were found more frequently when QBP-1 was adsorbed on the α-Quartz surface than those similar to the ‘C’-shaped cluster 1 and ‘n’-shaped cluster 3. This RMSD analysis suggests that in terms of the most populated clusters, QBP-1 undergoes relatively small structural changes upon adsorption to the (100) α-Quartz surface from bulk water, perhaps due to the rigid nature of its backbone imparted by the large proportion of proline present in the sequence. This behaviour is in contrast to a number of other material binding sequences [Capriotti et al. [2007]; Collino and Evans [2008]].

Ramachandran plots generated over the last $4{\times}10^6$ MD steps of the trajectories are shown in Figure 5.3. In each, four regions of ϕ-ψ space corresponding to PPII (-75, 160°), β (-135, 135°), α (-60, -45°) and α_L (60, 45°) have been comprehensively sampled. This is in stark contrast to the Ramachandran plot of QBP-1 in solution ([Notman et al. [2010]] and Figure 7.3), where only the PPII and α regions were sampled. However, akin to in solution, the most dominant secondary structure

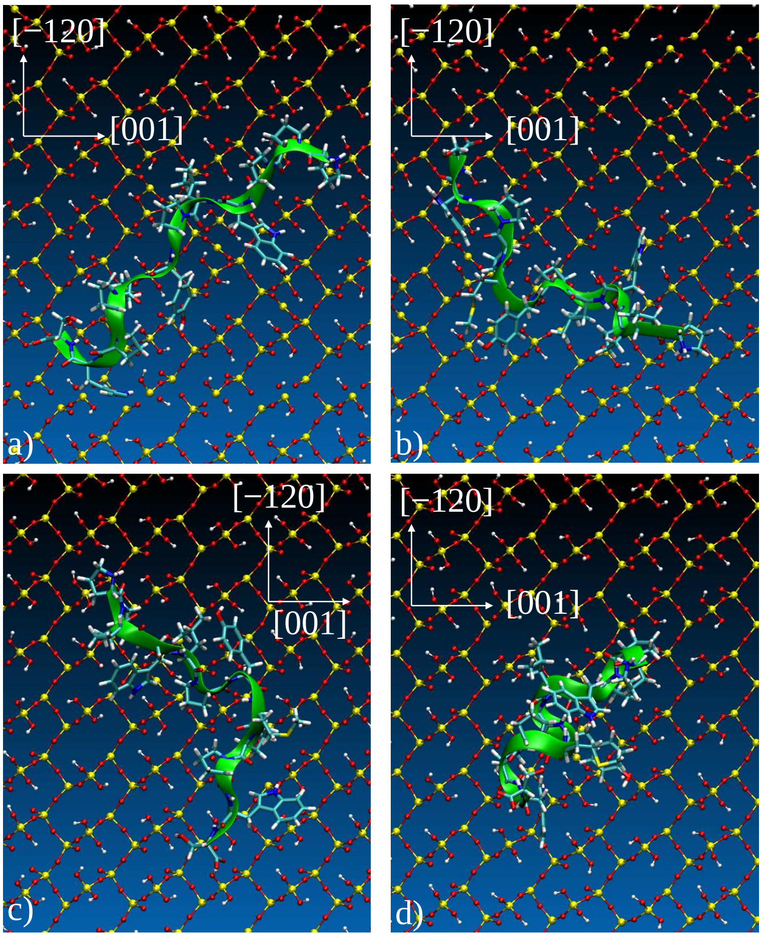


Figure 5.2: Commonly observed structures of QBP-1 in the most populated clusters after REMD simulations of the peptide adsorbed onto the (100) α-quartz surface: a) extended, b) extended with kinks (akin to solution clusters 2 and 4 Reference [Notman et al. [2010]]), c) 'C' shaped (akin to solution cluster 1), and d) helix (found as a top-3 populated cluster using the 'REST 2 off', 'REST light' and 'REST light 2 off' methods). Water has been omitted for clarity.

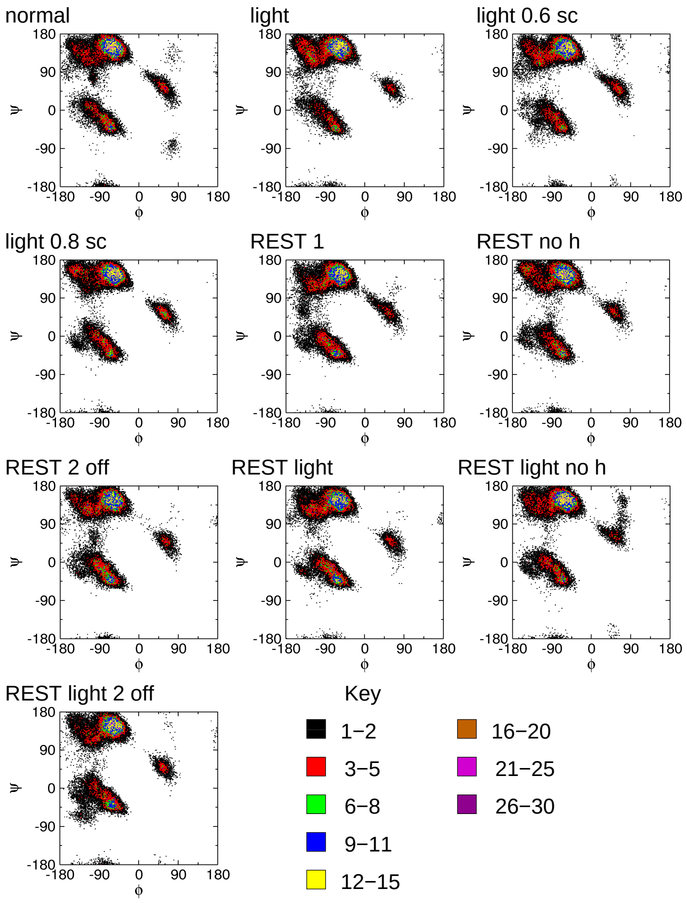


Figure 5.3: Ramachandran plots for QBP-1 adsorbed on the quartz surface for each RE-based variant.

characteristic of QBP-1 is PPII when adsorbed onto the quartz surface, with typically 50% or more of the dihedral angles sampled in all variants of T-REMD and REST simulations falling within this signature region. It is noted that the shape of the α region sampled by the solution and adsorbed peptide is markedly different, with only the latter showing the characteristic elliptical footprint of an α-helix. Hence, despite the RMSD analysis revealing strong structural similarities between the *most populated* clusters of QBP-1 in solution and those adsorbed at the aqueous α-Quartz interface, the presence of the surface does bring about differences in an ensemble containing *all* QBP-1 structures upon adsorption.

Comparison of Ramachandran plots revealed only small differences between methods, mainly in the sparsely sampled regions of ϕ-ψ phase space, corresponding to structures thought to have high energy. For example the γ region was only visited in the normal T-REMD simulation, whilst the γ_L and ϵ regions were more developed in some variants compared with others.

More detailed analysis of the Ramachandran plots was carried out by subdividing each plot into bins corresponding to regions of different secondary structure (Figure 2.6). The results revealed slight differences between the REMD variants in the relative frequencies with which each region was sampled. However, to interpret these data in a more meaningful way, regions were ranked according to the frequency with which they were sampled; the correlation of these ranks with those for normal T-REMD was quantified by a simple statistical parameter, Spearmans' Rank Correlation Coefficient [Zar [1972]]. This showed that the relative frequency with which each secondary structure region was sampled was in fact correlated in a statistically significant manner with the normal T-REMD simulation for all the other methods, with calculated Spearman's Rank Correlation Coefficients ranging from 0.90 to 1 (corresponding to significances of 95-100%, respectively). Not only this, but visual inspection of the evolution of the Ramachandran plots with number of MD steps revealed strong similarities, with the PPII, α and α_L regions being the first to be sampled across all the variants.

Special consideration of the α region in the Ramachandran plots was carried out in light of the difficulties described above in determining the biasing influence of the inclusion/omission of α helices in the initial replica populations. First, it is noted that the region was present and developed in the 'REST no h' and 'REST light no h' simulations, suggesting that the α signature cannot solely have arisen as a result of the initial replica population. Analysis similar to that above, ranking the contribution of each residue to sampling within the α region and comparing the ranks of each method with those obtained from the normal T-REMD simulation, did not reveal any significant correlation. However, the calculated α propensities of the amino acids present in QBP-1 (intrinsic α propensities of an amino acid when not part of a well defined secondary structure such as a helix or β-sheet) are perhaps too similar (ranging from 0.85 for Tyr to 1.17 for Trp [Swindells et al. [1995]]) for the results of this type of analysis to be significant.

5.3.4 Hydrogen Bonding

In all of the T/H-REMD-based simulations carried out the general trend in the number of hydrogen bonds formed by QBP-1 was as follows: intra-peptide < peptide-quartz < peptide-water (Table C.1). This trend was also in agreement with previous MD simulations of QBP-1 adsorbed onto the (100) α-quartz surface [Oren et al. [2010]], although the absolute number of each type of hydrogen bond formed differed slightly. These discrepancies were most pronounced for the average number of intra-peptide hydrogen bonds formed, with QBP-1 in 'REST no h', and the previous MD simulations exhibiting negligible intra-peptide hydrogen bonding compared to the other simulations. The most probable reason for this was the presence of helical-type structures being sampled, which are stabilised by backbone hydrogen-bonded bridges between residues. In particular, 'REST light' and 'REST light 2 off', simulations in which 18.8% and 17.5% of structures sampled in the final stages were helical, respectively, featured more intra-peptide hydrogen bonding on average than the other REMD variants.

5.3.5 Binding Residues and Motifs

Identifying the residues in the three most populated backbone clusters which were on average bound to the quartz surface, revealed once more that there were strong similarities between the different REMD variants (Table C.3). In general these results showed that surface adsorption of QBP-1 was primarily mediated through residues W4, L5, P6, Y7, M8, P9 and W11. The high affinity of these hydrophobic residues for the quartz interface concurs with both previous MD simulations of QBP-1 adsorbed at the aqueous (100) α-Quartz interface [Oren et al. [2010]] and the findings of Chapter 3, where the free energy of the hydrophobic amino-acid analogues, methane and benzene, was found to be very favourable [Wright and Walsh [2012a]]. In particular, it is interesting to note that analogous to preferred geometry of benzene adsorbed at this interface (Section 3.3.4, Figure 3.2b), the tryptophan and tyrosine residues of QBP-1 were found preferentially to bind such that the six-membered aromatic ring was perpendicular to the surface plane. Despite the caveats highlighted in Sections 3.3.4 and 3.4, this demonstrates that amino-acid analogues can adopt similar configurations on adsorption to their respective residue side-chains within a peptide, substantiating the use of analogues in future simple computational models of peptide adsorption simulations.

A recent bioinformatics study identified a number of dyad (2 residue), triad (3 residue) and tetrad (4 residue) motifs that are enriched in peptide sequences with high affinity to quartz, relative to the biocombinatorial library (Ph.D.-12) used in the peptide screening process [Oren et al. [2010]]. A number of these dyads and triads are also present in QBP-1 and so it is interesting to investigate whether these enriched motifs are directly responsible for binding QBP-1 to quartz or only contribute indirectly, in terms of either overall conformation and/or hydrophobicity of the sequence. A similar analysis conducted after regular MD simulations of QBP-1 adsorbed on the quartz surface concluded that they were binding motifs [Oren et al. [2010]]. However, in light of potentially inferior peptide conformational sampling accessible from regular MD simulations (Section 5.3.6), a similar analysis was also carried out here. Not only does this quantify QBP-1 surface binding, but it also provides another criterion against which the different REMD-based methods can be compared.

Of the enriched quartz-binding dyads present in the QBP-1 sequence, PW, WL and LP were commonly found also to be binding dyads in most REMD simulations (Table C.4). The enriched dyad PP was also identified as a binding dyad in one of the ten simulations. The relative frequencies with which each of these enriched dyads was identified, across the different REMD simulations, as being a binding dyad followed a similar trend to the ranking of their Motif Enrichment Factors (MEFs) [Oren et al. [2010]], with PW and WL being more prevalent than LP and PP. In a similar vein, dyads WS and MP, with small MEFs, were not identified as significant binding dyads in the simulations of adsorbed QBP-1. Akin to binding dyad motifs, the enriched triads PPW, PWL, WLP were also commonly found to contribute to the adsorption of QBP-1 onto the quartz surface (Table C.5).

Also consistent with previously [Oren et al. [2010]], tyrosine played an important role in QBP-1 adsorption in all simulations performed here, in terms of it being a bound residue both in its own right and within a motif (especially noted the binding dyads PY and YM). Hence although tyrosine and its motifs are not statistically enriched in the experimentally-identified ensemble of quartz binding peptide sequences, possibly as a result of a bias in the specific peptide libraries screened, or that introduced by the elution conditions [Puddu and Perry [2012]], when surrounded by the appropriate flanking residues, as in QBP-1, it can mediate strong surface adsorption.

Although the absolute contributions from each motif to QBP-1 adsorption differed from method to method, the broad findings from this motif analysis for the different REMD simulations were in qualitative agreement, and more importantly concurred with experimental data [Oren et al. [2007, 2010]] (which is independent of the force-field and model used).

5.3.6 Energetics

The total potential energies of otherwise identical systems containing the QBP-1 peptide adsorbed onto the (100) α-quartz surface indicate that out of the six different configurations tested, the extended conformation was the most energetically favourable. This concurs with the results of the majority of the T-REMD/REST simulations, where the most populated clusters featured extended QBP-1 structures. Systems containing QBP-1 in adsorbed configurations which had strong similarities to solution clusters 2 and 4 [Notman et al. [2010]] were next lowest energy, followed by solution clusters 1 and 3 (in that order) [Notman et al. [2010]] and then the helical structure (Figure 5.4). This energetic trend, taken over all of the replica exchange variant simulations, correlates with the number of highly populated clusters of the same structures. Qualitatively, therefore, these data suggest that QBP-1 conformations with a greater degree of chain extension are favoured by the peptide upon adsorption to the aqueous (100)-αQuartz interface.

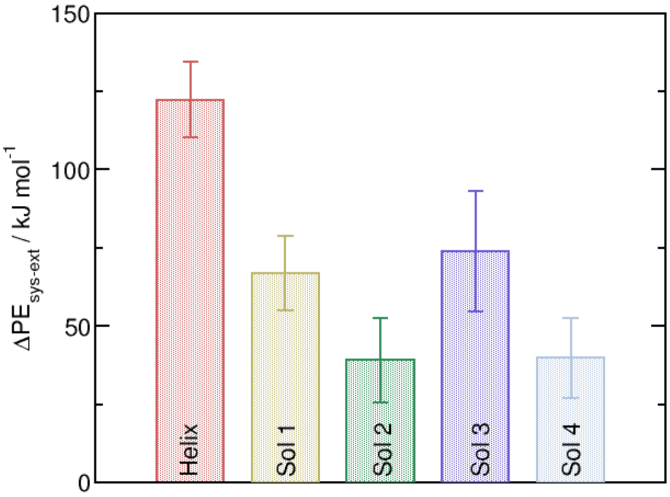


Figure 5.4: Potential Energy difference, $\Delta PE_{sys-ext}$, Equation 5.1.

Particularly important to note is that the helical structure at the interface is significantly disfavoured in terms of energy compared to the other adsorbed configurations tested. Hence, in the case of QBP-1, the presence of helical structures being regularly sampled during REMD simulations, even after an initial equilibration period, is possibly the result of their inclusion in the initial replica population, as well as possible helical bias in the CHARMM27 FF [Piana et al. [2011]]. Since conformational sampling, in terms of the number of different backbone clusters found, was greatest for 'REST no h', this could be particularly advantageous for future simulations of the QBP-1 peptide: it suggests that the performance of the REST method can be enhanced by the exclusion of helices from the initial replica population, without biasing the results. This conclusion cannot necessarily be extrapolated to other peptide sequences though, without a similar energetic analysis. On the other hand, the fact that one of the more populated clusters identified by the 'REST 2 off', 'REST light' and 'REST light 2 off' methods featured a helical structure opens a number of questions, such as whether the use of the CHARMM27 FF, and/or solvent mass scaling or the REST method itself, may have contributed to this effect. Thus in Chapter 7 the first of these factors was tested by using a newer version of CHARMM (CHARMM22*), which features modified torsion potentials [Piana et al. [2011]], to model QBP-1–quartz adsorption and comparing the results with those in this Chapter, where CHARMM27 [MacKerell et al. [1998, 2004]] was used.

Finally, visual inspection revealed that in each of these regular MD simulations, no change in peptide conformation or lateral motion of the peptide across the surface occurred during the 5 ns time period. This again highlights the need for the development of enhanced sampling methods for modelling complex systems such as peptides adsorbed at interfaces.

5.4 Discussion

These simulations demonstrate that REST can deliver results of comparable quality to T-REMD. However, extrapolation to REMD-based simulations of other material binding peptides at their respective interfaces must be done with caution for a number of reasons mentioned above. In particular, QBP-1 is a stiff peptide [Notman et al. [2010]] with high proline content, and so it is possible that techniques such as mass scaling, used alongside T-REMD or REST, may have a more profound effect on conformational sampling when applied to more flexible peptides sequences. In addition, for the specific case of employing 'light water' in a simulation, the possible effect of reduced solvent viscosity on facilitating more rapid surface adsorption, as observed in the ‘REST light 2 off’ trajectory must be considered. Whether this was an anomalous result, borne out by the particular set of random initial velocities used, can only be verified by further extensive testing.

However, even taking this into account it is unlikely that T-REMD with mass scaling could out-perform REST in terms of the computational resource required to sample the equivalent amount of phase-space. As noted by Wang *et al.*, not only is the REST method less computationally intensive compared with T-REMD, in that significantly fewer replicas are required, it is also computationally less expensive, in that a given number of MD steps can be completed in fewer CPU hours [Wang et al. [2011]]. For the system simulated in this study, the REST simulations took roughly 50% less CPU time for the same number of steps using the same number of processors per replica. Combined with the reduced number of replicas required for the REST simulations, T-REMD was up to six times more computationally expensive. Hence for T-REMD with mass scaling to be competitive with REST, then it must sample the same amount of phase-space in less than 1/6th of the number of MD steps–a feat not achieved by the simulations performed in this Chapter.

The choice of structures used in the initial replica population can influence both the results (*i.e.* of the inclusion or omission of helices) and the opportunity for the peptide optimally to sample conformational space (in the case of displacing the peptide from the quartz surface in 2 of the 11 replicas). Although the findings of this study suggest that helical folding and unfolding is unlikely to be observed within the time-scale accessible to a REMD-based simulation with modest temperature window, the question of whether helices should be included in the initial population of structures can at present only be addressed on a case-by-case basis. Energetic analysis, similar to that carried out here (Section 5.3.6), is one way in which this choice could be made. However, for flexible peptide sequences, such considerations may yet prove unnecessary, if the energetic barrier to the formation and destruction of helices is significantly smaller than that for QBP-1.

Strong peptide-surface adsorption can in itself also hinder exploration of phase-space during REMD simulations. In the case of QBP-1, the number of different backbone clusters sampled was limited when the peptide was initially positioned close to the quartz surface in all replicas. However, this effect can be more pronounced for more strongly adsorbing material binding peptides, such as the gold binding peptide, AuBP-1 [Hnilova et al. [2008]], at the aqueous gold interface (Chapter 8). Initially displacing the peptide from the surface in a number of the replicas may not be sufficient to overcome this issue for starting configurations in which the peptide is moderately displaced from the interface, rapid adsorption may occur; larger initial surface separations, on the other hand, risk sampling the solution, rather than, adsorbed-state of the peptide. Use of biasing potentials, either predetermined [Wang et al. [2008]; Vellore et al. [2010]; Snyder et al. [2012]] or calculated on the fly (*e.g.* metadynamics [Schneider and Ciacchi [2012]]), combined with replica exchange based methods, are possible solutions to the problem.

When considering the influence of initial replica population on the performance of REST, then it is important also to remember that if a bio-interfacial system has fully equilibrated during a REMD-based simulation then the reference replica would sample the canonical/isobaric isothermal ensemble (dependent on the simulation conditions used) of conformational states of the adsorbed peptide; this ensemble is, of course, *independent* of the starting conditions. One could estimate the simulation time required to reach this state from two key metrics: the characteristic time taken by a replica (by 'replica' I mean the continuous trajectory of co-ordinates) to undergo significant conformational change, and the average time period for a replica to cycle through all (effective) temperatures. Whilst, as has been shown here, REMD-based methods can vastly improve exploration of phase space in complex systems relative to regular MD, reaching the fully equilibrated state requires extensive simulation times and remains computationally unviable for most systems. Until alternative methods are perfected, therefore, REST, when compared to T-REMD, offers an efficient way to sample as much of peptide conformational space in a bio-interfacial simulation as possible for a given CPU resource.

5.5 Conclusions

The simulations performed in this Chapter demonstrate that REST-based approaches can accomplish enhanced conformational sampling of peptides adsorbed at aqueous interfaces, for around half the computational time and one third of the total amount of processors, compared with temperature-based REMD approaches, while delivering comparable performance in a range of key metrics. Use of mass-scaling, for either the solvent or the solvent and solute side-chains, was found to deliver little benefit when implemented for the QBP-1 peptide adsorbed onto quartz in either approach, although further testing with more flexible sequences is necessary before strong conclusions about this technique can be reached. In agreement with experiment and bioinformatics studies, in its most likely conformation the QBP-1 peptide was adsorbed to the α-Quartz surface by motifs featuring proline, tryptophan and leucine. Comparison with previous T-REMD simulations of the QBP-1 in solution [Notman et al. [2010]] and those carried out here (Chapter 7) suggests it does not undergo substantial structural change upon adsorption, but remains predominantly extended with pronounced PPII character. The data in this Chapter suggest that the overall ensemble of structures bound at the interface is, however, different to the corresponding ensemble in solution.

The possible influence of the FF to favour helical conformations of QBP-1 is probed in Chapter 7, whilst it would be interesting to investigate the effect of combining REST with mass scaling for alternative bio-interfacial systems with more flexible peptide sequences in the future. In addition, REST simulations of QBP-1 adsorption at the (001) and (011) aqueous α-Quartz interfaces would be one way to investigate whether the ability of amino-acid analogues to discriminate between different crystallographic planes of quartz on binding observed in Chapter 3 can be extrapolated to peptide adsorption.

Chapter 6

GolP-CHARMM: A First-Principles Based Force-field for modelling protein adsorption to Au(111), Au(100)(1×1) and Au(100)(5×1).

6.1 Introduction

The potential range of nanotechnological applications based on AuNPs is vast due to their unique optical and chemical properties (Section 1.4). Size and shape selective AuNP synthesis directed by the presence of a peptide which preferentially binds to one crystallographic plane of gold over another would be an important step in realising these goals. Unlike platinum [Chiu et al. [2011]; Ruan et al. [2013]] and palladium [Peng et al. [2013]], facet selective sequences for gold have yet to be identified, however. The use of simulations carried out at the aqueous biomolecule–gold interface, in partnership with experiment, to guide the design of such peptides is hindered at present by the lack of a suitable FF.

Prior to commencing the work reported in this Chapter, two bio-interfacial FFs for gold–CHARMM-METAL [Heinz et al. [2008]; Feng et al. [2011]; Pandey et al. [2009]] and GolP [Iori et al. [2009]; Iori and Corni [2008]]–were commonly in use.

Each has different shortcomings, discussed in depth in Section 1.4.2. Briefly, while GolP has a number of advantages over CHARMM-METAL–namely that it includes terms to account for the dynamic polarisation of gold atoms and was specifically parameterised to reproduce differences in the nature of the interactions between chemi- and physi-sorbing functional groups–it was only derived for modelling adsorption at the aqueous Au(111) interface. The protocol used to parameterise GolP has therefore been improved and extended here. In this FF parameters compatible with the bio-organic FF CHARMM [MacKerell et al. [1998]] for modelling peptide adsorption onto the Au(111), Au(100)(1x1) and Au(100)(5x1) planar surfaces under aqueous conditions have been derived. Whilst peptide adsorption to AuNP edge and vertex sites cannot be modelled using the new FF, GolP-CHARMM [Wright et al. [2013b,c]], the relative length scales of the sequences [Brown [1997]; Naik et al. [2002b]; Slocik et al. [2005b]; Hnilova et al. [2008]; Causa et al. [2013]] and AuNPs currently used in experimental work [Sethi and Knecht [2009, 2010]; Causa et al. [2013]] suggest that AuNP facets can often be acceptably approximated by planar surfaces [Brancolini et al. [2012]]. It was appropriate to include terms for both the native and reconstructed forms of the Au(100) surface in GolP-CHARMM since the exact structure of a Au(100) facet featured by a AuNP is presently unknown [Zehner et al. [1975]; Shi and Stampfl [2008]; Li et al. [2008]; Wang and Palmer [2012]; Surrey et al. [2012]].

GolP-CHARMM has been primarily fit to quantum mechanical, rather than experimental, data. The latter was used however to benchmark the dispersion corrected functional, vdW-DF [Dion et al. [2004]; Thonhauser et al. [2007]], used to carry out the PW-DFT calculations. Empirical bio-interfacial FFs parameterised against first-principles calculations can often be more accurate than their experimental counterparts principally because both adsorbate conformation and binding affinity can be used in the fitting process [Herbers et al. [2013]]. Experimental interaction energies are normally ensemble-based and hence offer little information about preferred adsorbate-surface geometry.

6.2 Methods

In GolP-CHARMM, the potential energy function for the interaction between the gold surface and a biomolecule is [Iori et al. [2009]]:

$$V^{tot}_{Au-X} = V^{im}_{Au-X} + V^{vdW}_{Au-X} + V^{chemisorb}_{Au-X} + V^{\pi}_{Au-X} \tag{6.1}$$

where V^{im}_{Au-X} is the Coulombic electrostatic interaction between the biomolecule and the image charges it induces in the gold upon adsorption, and V^{vdW}_{Au-X}, $\mathrm{V}^{chemisorb}_{Au-X}$ and V^{π}_{Au-X} are the van der Waals, chemisorption and π electron interactions between the biomolecule and gold. A dynamic ensemble of rigid rods (described elsewhere [Iori and Corni [2008]]), each with small dipole, has been used to capture image charge effects, whilst van der Waals interactions are described by the 12-6 LJ potential (Equation 2.11). LJ parameters for gold atoms (σ_{Au-Au}, ϵ_{Au-Au}), when combined with LJ parameters from biomolecule atom types, were fitted to reproduce either experimental or PW-DFT interaction energies. Since the gold atoms themselves are immobile in the FF, the Au-Au terms are not needed to model the gold directly, but only to define cross-terms with atoms from the interacting biomolecules. In accordance with CHARMM, the Lorentz-Berthelot mixing rules (Section 2.1.2) were used here [Allen and Tildesley [1987]].

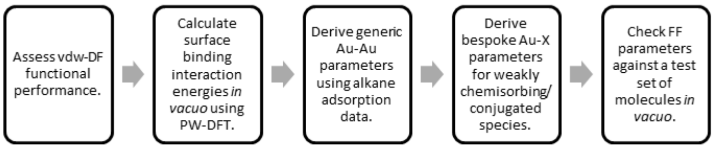

Assess vdw-DF functional performance.
Calculate surface binding interaction energies *in vacuo* using PW-DFT.
Derive generic Au-Au parameters using alkane adsorption data.
Derive bespoke Au-X parameters for weakly chemisorbing/ conjugated species.
Check FF parameters against a test set of molecules *in vacuo*.

Figure 6.1: Scheme depicting the general steps taken to derive GolP-CHARMM FF parameters for the Au(111) and Au(100) surfaces.

Analogous to GolP [Iori et al. [2009]] a multi-scale approach has been adopted (Figure 6.1), similar to that used by delle Site and *co workers* [Schravendijk et al. [2007]; Ghiringhelli and delle Site [2008]]. Generic LJ parameters for gold (denoted 'Au–Au') were fit to reproduce alkane adsorption energies. Specific Au-X (X = S, N, O, H(-O/S) and C in aromatic ring systems) LJ terms have then been derived for weakly-chemisorbing species and aromatic sp^2 hybridised carbon systems, using small molecules containing the same functional groups as those present in peptides. Whilst in the original GolP these terms were added for Au–S and Au–N interactions only, this has been extended in GolP-CHARMM also to include Au–O and Au–H(O). Here the FF was fitted to reproduce the adsorbed geometry and energetics of single water molecules using PW-DFT data. This contrasts with both GolP and CHARMM-METAL, neither of which were parametrised specifically against gold-water interactions.

Experimental data for alkane adsorption, from which LJ parameters can be derived, is only available for the Au(111) surface. Hence, in order to extend GolP to gold surfaces other than Au(111), interaction energies have been calculated using PW-DF. Here the long-range, dispersion corrected, van der Waals Density Functional (vdW-DF) (Section 2.1.1) [Dion et al. [2004]; Thonhauser et al. [2007]; Roman-Perez and Soler [2009]], combined with the revised Perdew-Burke-Ernzerhof (revPBE) [Hammer et al. [1999]] functional, was employed. First, in this Chapter, the suitability of this functional for describing small molecule-gold adsorption was assessed by comparing experimental and calculated interaction energies for a number of different functional groups on Au(111) (6.3.1). A combination of experimental and calculated reference adsorption energies was used to fit FF LJ parameters for the Au(111) surface. In the absence of appropriate experimental data, LJ terms for Au(100)–both native and reconstructed forms–were fitted to PW-DFT data calculated here.

It is important to stress that, unlike GolP [Iori et al. [2009]], both the energy *and geometry* of the adsorbed molecules were used to refine the FF when fitting parameters for the *native* forms of the two gold surfaces–Au(111) and Au(100). However, several groups have shown that vdW-DF overestimates equilibrium intermolecular [Dion et al. [2004]; Thonhauser et al. [2007]; Lee et al. [2010]] and molecule-surface [Toyoda et al. [2009]; Romaner et al. [2009]; Lee et al. [2010]] separation distances. Therefore, it was not appropriate to fit FF parameters to match these calculated distances exactly. After reviewing the literature, a tolerance of ∼0.2–0.3 Å was adopted to parameterise the FF; the *trend* in equilibrium separations from the Au(111) and Au(100)(1×1) surfaces predicted by GolP-CHARMM *in vacuo* is the same as that found from vdW-DF calculations. The more complex nature of adsorption onto the reconstructed Au(100) surface–chiefly due to the spatial and energetic heterogeneity of the interface–made explicitly incorporating equilibrium adsorbate-gold separation distances in the parameterisation strategy unfeasible; molecular-surface geometry was qualitatively accounted for, however.

Since the gold atoms themselves are immobile in GolP-CHARMM, the structure of the interfacial slab must be carefully considered (Figure 1.3). In the case of the native surfaces–Au(111) and Au(100)–gold atoms were arranged ideally using an Au-Au separation distance of 2.93 Å [Iori et al. [2009]]. In contrast, the PW-DFT optimised, rather than ideal, Au(100)(5×1) slab geometry was used as the basic structural unit for making larger slabs in this case. In its relaxed structure the reconstructed Au(100) interface is buckled due to the spatial mismatch between surface and underlying Au atoms. In contrast, optimisation of the two native surfaces brought about minimal changes in slab geometry–the Au(111) surface remained topologically flat despite sharing a similar hexagonal-like spatial arrangement of surface Au atoms with Au(100)(5×1). The degree of surface corrugation of

Au(100)(5×1) modelled by vdW-DF was in good agreement with experiment (See Section 6.3.1). Since the rumpled topography of the Au(100)-*hex* surface may play a role in facet-selective binding [Phan et al. [2012]], it was essential to incorporate this structural feature in GolP-CHARMM. (The bottom of the Au(100)(5×1) gold slab used to parametrise the FF featured the Au(100)(1×1) surface. In the fitting procedure for the reconstructed interface, FF parameters derived for the native surface were used to model atoms in this bottom facet and those in the bulk of the slab.)

A feature of GolP is the presence of virtual interaction sites on the Au(111) surface to direct 'atop' Au atom adsorption (by 'atop' I mean that the heteroatom is preferentially adsorbed directly above a real Au atom in the surface, rather than above a hollow or bridge surface site) as observed in first-principles calculations [Iori et al. [2009, 2008]; Hong et al. [2009]]. The use of non atom-centred interaction sites is an established practice in FFs (*e.g.* the TIP4P [Jorgensen et al. [1983a]] or TIP5P [Mahoney and Jorgensen [2000]] potentials). In GolP-CHARMM it is the virtual interaction sites, and not the gold surface atoms, that have spatial extent and interact with adsorbing species via the LJ terms; the dipolar rods which capture the image charge contribution to this interaction are, however, located on the Au atoms. The combined use of atom centred/off-centred dipolar rods/virtual interaction sites in GolP-CHARMM is an empirical tool for describing the delocalised electron density in the plane of the metal surface.

All three gold interfaces parameterised in this Chapter therefore feature virtual interaction sites sited in surface hollows (Figure 6.2). For the Au(111) surface this is an identical geometry to that used previously [Iori et al. [2009]]. Four different geometries were tested in the case of Au(100)(1×1) [Wright et al. [2013b]] with the hollow site found to be the most favourable. The reconstructed Au(100) FF is more complex. In order for the FF to reproduce the spatial heterogeneity of adsorption observed in the first-principles calculations (Table D.3), two sets of virtual interaction sites were necessary. Those located in the hollow sites (denoted 'AUIU'), were homogeneous, whereas the second set (denoted 'SX', where X=0, 1, 2, 3), located in bridge sites running perpendicular to the reconstruction (in the [011] direction), were not. For all three interfaces the virtual interaction sites lie within the plane of the surface, with the z co-ordinate being the arithmetic mean of the two/three Au atoms from which the bridge/hollow site is comprised.

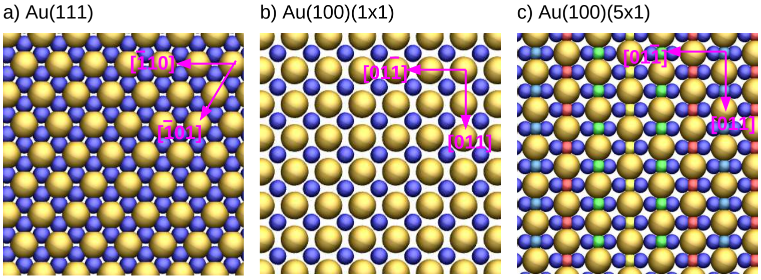


Figure 6.2: Arrangement of virtual sites on a) the Au(111), b) the Au(100)(1×1) and c) Au(100)(5×1) surfaces. Real gold surface atoms are shown in gold and hollow virtual interaction sites in blue. The same colour scheme has been used for bridge virtual interaction sites on the Au(100)(5×1) surface as that in Figure 6.4.

Plane Wave DFT Calculations

PW-DFT calculations have been carried out for the molecules listed in Table 6.1 adsorbed onto the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces *in vacuo*, using Quantum Espresso (versions 4.3.x) [Giannozzi et al. [2009]]. Three-dimensional periodic boundary conditions were employed, with supercell dimensions (Au(111): $2\sqrt{3}\times 3$ for small molecules and $3\times 3\sqrt{3}$ or $2\sqrt{3}\times 4$ for larger ones; Au(100)(1×1): 3×3/4×4 for small/large adsorbates; Au(100)(5×1): 1×4 in all calculations) and vacuum thickness in the direction normal to the surface chosen to ensure interactions between periodic images were insignificant [Iori et al. [2009]]. (An Au-Au lattice parameter of 2.93 was used throughout.) Gold slabs 4/5/5 atomic layers thick were used for the Au(111)/Au(100)(1×1)/Au(100)(5×1) surfaces. In the case of the reconstructed interface only, Au atoms in the lower-most layer were frozen in their ideal Au(100)(1×1) positions; all other Au atoms were relaxed. Calculations were performed using vdW-DF [Dion et al. [2004]; Thonhauser et al. [2007]; Roman-Perez and Soler [2009]] with the revPBE exchange-correlation functional [Hammer et al. [1999]] and ultrasoft pseudopotentials [Vanderbilt [1990]] (derived using the PBE exchange-correlation functional [Perdew et al. [1996]; Rosa et al. [2012]]) to describe electron-ion interactions. Plane wave kinetic energies and electron densities were truncated at 25 Ry and 200 Ry, respectively. Brillouin zone integration was carried out using a Gaussian smearing technique [Fu and Ho [1983]] with width 0.05 Ry.

Optimal adsorbate-gold geometries (Figures D.2, D.3, D.4 and D.5) were obtained by relaxing the structure of each small molecule when in close proximity to the surface (to a convergence criterion of 0.026 eV/Å). In the case of the Au(111) surface, favourable configurations found from the previous study [Iori et al. [2009]]

molecule	Au(111)		Au(100)(1×1)		Au(100)(5×1)	
	kJ mol^{-1}	Å	kJ mol^{-1}	Å	kJ mol^{-1}	Å
acetone	−40.8	3.21	n/a	n/a	n/a	n/a
benzene	n/a	n/a	−57.9	3.39	−59.0	3.27
butane	−42.3	3.90	−43.2	3.72	−45.5	3.72
cyclohexane	−50.6	3.78	n/a	n/a	n/a	n/a
diethyl sulphide	−76.6	3.00	−85.3	2.87	−85.5	2.73
dimethyl sulphide	n/a	n/a	−69.1	2.87	−71.4	2.70
ethane	−24.7	3.84	n/a	n/a	n/a	n/a
ethanethiol	n/a	n/a	−55.1	2.77	−57.7	2.72
imidazole	−54.1	2.51	−67.2	2.43	−65.5	2.35
indole	n/a	n/a	−87.9	3.27	−90.4	3.29
hexane	−60.3	3.88	n/a	n/a	n/a	n/a
methane	−16.5	3.77	−16.3	3.69	−18.0	3.83
methanethiol	−45.3	2.88	−51.8	2.78	−52.8	2.74
methanoic acid	−29.2	3.06	−29.5	2.94	−32.3	3.05
methanol	−30.5	2.93	−31.7	2.85	−34.1	2.83
methyl amide	−34.2	2.97	−34.9	2.72	−40.8	2.74
methyl amine	−54.8	2.57	−61.3	2.48	−62.3	2.46
phenol	n/a	n/a	−66.3	3.33	−67.6	3.40
toluene	n/a	n/a	−68.8	3.21	−69.6	3.45
water	−18.3	3.03	−20.8	2.90	−20.5	2.97

Table 6.1: Optimal PW-DFT calculated interaction energies of small molecules adsorbed onto the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces *in vacuo* (kJ mol^{-1}) and equilibrium separations in direction normal to the surface (Å). 'n/a' refers to a calculation which was not performed.

(which employed the PBE functional only) were used as the starting point for the work here. For Au(100)(1×1) a number of initial geometries were tested, differing in both molecular orientation in space (*e.g.* with plane or key bonds oriented either perpendicular or parallel to the surface normal) and position relative to the underlying gold lattice (*e.g.* heteroatoms placed above top, hollow and bridge Au atom sites). A vacuum thickness of 10 Å and a Monkhorst-Pack *k*-point mesh of 4×4×1 were used in all geometry optimisation calculations at these two interfaces.

The large size (104 Au atoms) and low symmetry of the gold slab needed to model the Au(100)-*hex* reconstruction, however, made full exploration of adsorbate-gold conformational space (*i.e.* positioning and orientation of the adsorbate relative to the underlying gold surface, as well as internal adsorbate conformation) computationally prohibitive at the first-principles level of theory in this case. Hence the extended, four-step systematic procedure described below was used to identify favourable adsorbate-gold geometries at the Au(100)(5×1) interface (Figure 6.3).

(A slightly coarser Monkhorst-Pack k-point mesh of 2×2×1 was used in these calculations.)

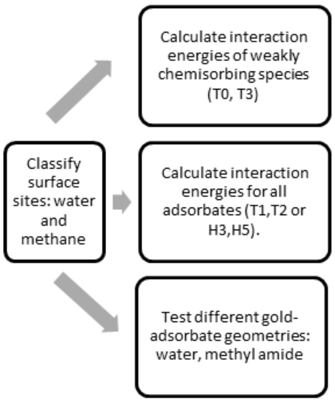

Classify surface sites: water and methane
Calculate interaction energies of weakly chemisorbing species (T0, T3)
Calculate interaction energies for all adsorbates (T1,T2 or H3,H5).
Test different gold-adsorbate geometries: water, methyl amide

Figure 6.3: Scheme depicting the steps taken to identify optimal adsorbate-gold geometries on the Au(100)(5×1) surface. See Figure 6.4 for description of adsorption site labels.

1. First, the bare Au(100)(5×1) surface was relaxed from an initial structure in which the surface Au atoms were arranged in an ideal hexagonal geometry [Havu et al. [2010]; Shi and Stampfl [2008]; Venkatachalam et al. [2008]; Jacob [2007]; Feng et al. [2005]; Baber et al. [2012]] above an Au(100) lattice (Au-Au distance 2.93 Å). After relaxation, six distinct spatial environments for Au atoms in the Au(100)(5×1) surface were identified (Figure 6.4, labelled T0 to T5).

2. Therefore, second, it was necessary to probe the spatial dependence of adsorption energies on this surface. Water and methane were chosen as test molecules. Water, a representative polar molecule, was selected due to is prevalence in biomolecule-gold simulations carried out under aqueous conditions, for which the FF developed herein is designed. Methane is a representative non-polar molecule, chosen for these test calculations since it is the simplest analogue of the hydrophobic moieties present in all peptides. In addition, the methane–gold calculated and experimental interaction energies at the Au(111) interface *in vacuo* (16.5 and 14.5 $\mathrm{kJ\,mol^{-1}}$ [Wetterer et al. [1998]], Table D.7) were found to agree well using the vdW-DF functional [Wright et al. [2013b]]. For water, however, suitable experimental data against which the calculated water-Au(111) interaction could be benchmarked was not available. On both

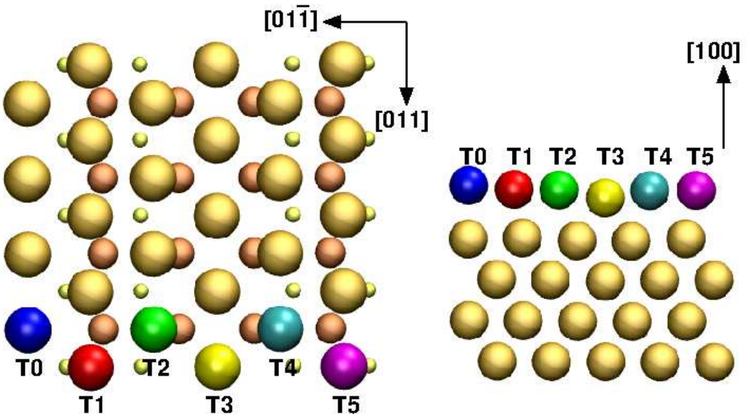


Figure 6.4: The Au(100)(5×1) surface after relaxation using vdW-DF [Dion et al. [2004]; Thonhauser et al. [2007]] viewed a) from above and b) from the side. For the plan view (left) surface layer Au atoms are depicted as largest and in medium gold, 2nd layer Au atoms dark gold and medium size, while the 3rd layer of Au atoms are the smallest size and light gold in colour.

the Au(111) and Au(100)(1×1) surfaces a single water molecule was found preferentially to adsorb atop Au atoms, while methane adsorbed above hollow sites. Hence, water and methane were used here to investigate the relative adsorption energies of the different top (Figure 6.4) and hollow (Figure D.1) site locations, respectively, on the Au(100)(5×1) surface.

3. Third, optimal adsorbate-gold geometries at the Au(100)(5×1) interface for the other small molecules listed in 6.1 were then obtained by performing at least two different geometry optimisation calculations (each initiated in a different conformation) per adsorbate. Since the preferred orientation and positioning (above a top, hollow or bridge surface site) relative to the underlying gold surface for each molecule was the same on adsorption to the Au(111) as it was on the Au(100)(1×1) surface (Figures D.2 and D.3) [Wright et al. [2013b]], this favourable orientation was used to construct starting conformations for calculations at the Au(100)(5×1) interface. The most and least binding top (T2 and T1, respectively, Figure 6.4) and hollow (H5 and H3, respectively, Figure D.1) sites, identified by the water and methane calculations, were also used in this process. For example, benzene was found preferentially to adsorb with its plane parallel to the surface, centred above a hollow site on both the Au(111) and Au(100)(1×1) surfaces (Figures D.2 and D.3). Therefore the two initial conformations of the adsorbate-Au(100)(5×1) system relaxed were: one

with the benzene ring centred above an H3 site, and the other with it centred above a H5 site—in each geometry the plane of the molecule was parallel to the plane of the surface. In contrast, imidazole preferentially adsorbed with its plane perpendicular to both the Au(111) and Au(100)(1×1) surfaces with its unprotonated nitrogen atop a surface Au atom (Figures D.2 and D.3). Hence, in this case the two imidazole–Au(100)(5×1) configurations relaxed featured the same imidazole–gold geometry, but with the unprotonated nitrogen adsorbed to T1 and T2 surface sites, respectively.

To test the hypothesis that the most favourable adsorption site and geometry identified for each molecule on both the Au(111) and Au(100)(1×1) surfaces were also optimal at the Au(100)(5×1) interface, additional calculations were performed for water, methane and methyl amide. The last was chosen to ensure that the new FF could acceptably reproduce peptide backbone adsorption. Both the orientation of water relative to the gold surface ('O-down', 'H-down' or 'flat') and adsorption to hollow sites were investigated; for methane, adsorption to selected top sites was probed. Three supplementary optimisations of the methyl amide–gold system were performed: one in which the plane of the molecule was initially parallel, rather than perpendicular, to the surface; one with the molecule initially above surface hollow, rather than top, sites; and one in which the orientation of the molecule with respect to the direction of the reconstruction was tested. While methyl amide adsorption to hollow sites and oriented with its plane parallel to the surface were both found to be unfavourable, the calculations did identify that it was more favourable for the normal to the plane of the amide group to be oriented along the [011] vector, *i.e.* perpendicular, rather than parallel, to the direction of the surface reconstruction. In other words, a structure in which the O atom of methyl amide was above a T2 site while the N atom was approximately atop a T3 site, rather than both atoms being atop T2 sites. This favourable orientation was subsequently found to be low-energy for methanoic acid, where the carboxylate O was atop T2 and the hydroxyl O atop T3.

4. A fourth step was only required for those molecules which weakly chemisorb to gold. The interaction energies of species that interact with gold via physisorption only were found to vary minimally with adsorption site on the Au(100)(5×1) surface. However, there was a strong spatial dependence, of up to almost $20\,\mathrm{kJ\,mol^{-1}}$, for the weakly chemisorbing molecules such as methyl amine, imidazole, thiols (methane- and ethanethiol) and dialkyl sulphides

(dimethyl and diethyl sulphide). For the FF to reproduce this surface heterogeneity in adsorption energy, parameters must be fit to each of the six different top site locations on the Au(100)(5×1) surface. However, T1 and T5 atoms are almost spatially equivalent, in terms of both their positioning in the plane of the surface, relative to the second layer of Au atoms, and perpendicular to it. The same is true of T2 and T4 Au atom environments. Hence, only two further geometry optimisation calculations were carried out for these weakly chemisorbing molecules: one with the adsorbate initially placed above a T0 and one in which it was placed above a T3 site.

Once optimal adsorbate-gold geometries had been identified at each surface–Au(111), Au(100)(1×1) and Au(100)(5×1)–interaction energies were calculated as follows:

$$\Delta E_{mol\_Au} = E_{mol\_Au} - E_{mol} - E_{Au} \tag{6.2}$$

where E_{mol_Au}, E_{mol} and E_{Au} are the total energies of systems describing the small molecule adsorbed at the interface, the adsorbate only and the gold slab only, respectively. All three cells were relaxed, as described above, with optimal geometries being used in a single point calculation to determine the total energy. A thicker vacuum layer of 25 Å and a finer Monkhorst-Pack *k*-point mesh of 6×8×1/8×8×1/4×4×1 (Au(111)/Au(100)(1×1)/Au(100)(5×1)) were employed in these single-point calculations.

6.2.1 *in Vacuo* Force-Field Calculations

FF-based calculations were performed using the MD simulation package, Gromacs version 4.5.X [Hess et al. [2008]]. All simulations were carried out in the *NVT* ensemble, implemented with a Nosé-Hoover thermostat and three dimensional periodic boundary conditions. The leap-frog algorithm [Hockney et al. [1974]] with a timestep of 1 fs was employed. PME electrostatic summation was used with a real-space cut-off at 11 Å, whereas a force-switched cut-off starting at 9 Å and ending at 10 Å was used for LJ non-bonded interactions. The rigid-rod-dipole method for gold atom polarisation was implemented following an identical method to the original GolP FF [Iori et al. [2009]; Iori and Corni [2008]] with random initial dipole orientations used throughout.

Gold slabs of surface areas 29.3×30.4 Å ^{2}, 29.3×29.3 Å ^{2} and 29.3× 23.44 Å ^{2} (5 layers thick in all cases) were used for the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces respectively. (For Au(100)(5×1) this was a 2×2 supercell of the PW-DFT optimised slab.) The z dimension of the simulation cell was 70 Å, giving a vacuum gap between the adsorbed molecule and the nearest periodic image of the slab of approximately 60 Å.

To identify the minimum energy structure of each small molecule when adsorbed, short simulated annealing MD simulations of up to 40 ps were carried out. The temperature of the gold dipoles was maintained at 300 K throughout whilst that of the adsorbate was incrementally decreased to 1 K. Initial adsorbate temperatures were varied, to ensure adsorption, from 300 K to 100 K, depending on the size of the molecule. In all cases simulations were started with the adsorbate displaced from the interface. At least 5 different annealing cycles were performed per adsorbate, starting from different initial adsorbate conformations.

Akin to PW-DFT, this protocol for identifying optimal adsorbate-gold FF geometries was modified slightly for the reconstructed Au(100) interface. Briefly, the simulated annealing simulations performed fell into two categories: 1) annealing from vacuum–analogous to the the annealing procedure used for the Au(111) and native Au(100) surfaces; and 2) site-specific annealing. The first procedure was used to fit FF parameters for those molecules which only physisorbed onto the gold surface, whilst the second was necessary for chemisorbing species. In the latter case, six independent simulations were initiated per adsorbate: in each the molecule was placed close to the surface above a different top site. The optimal PW-DFT orientation of the molecule relative the underlying gold surface was used as the starting configuration. To prevent diffusion away from weak interaction sites, stimulated annealing runs of 1 ps duration only, in which the temperature was reduced from 5 K to 1 K, were performed. As a test of FF parameters derived for this surface, the 'site-specific' annealing protocol was also carried out for a number of physisorbing molecules, whilst the 'vacuum' annealing procedure was applied to chemisorbing species.

Classical interaction energies were calculated following two 50 ps constant-temperature (300 K) MD simulations: the first with the molecule midway between the gold slab and its periodic image, and the second with the molecule adsorbed on the surface. During both, all atoms were kept fixed in their initial relaxed positions with only the gold dipoles able freely to rotate. The non-electrostatic contribution to the interaction energy is the difference between the bond, angle, dihedral and LJ components of the potential energy of the bound and free molecule systems. The *free*-energy of the electrostatic interaction between the adsorbed molecule and the ensemble of dipoles (F_{Coul}) was approximated from the difference between the average electrostatic energies (E_{Coul}) using 6.3, which was derived previously [Iori and Corni [2008]].

$$F_{Coul} = \frac{E_{Coul}}{2} \tag{6.3}$$

6.2.2 Force-field MD Simulations of Aqueous Gold Interfaces

MD simulations of the aqueous Au(111), Au(100)(1×1) and Au(100)(5×1) interfaces were performed to investigate facet-selective differences in water structuring above each surface using the newly derived FF parameters. Gold slabs, of surface areas $43.95 \times 40.60\,\text{Å}^2/29.30 \times 29.30\,\text{Å}^2/29.3\times 23.44\,\text{Å}^2$ (Au(111)/Au(100)(1×1)/Au(100)(5×1)) were solvated with 2040/1425/1354 modified TIP3P [Jorgensen et al. [1983b]; Neria et al. [1996]] water molecules. The cell dimension in the direction normal to the surface was adjusted prior to the simulation such that density of water midway between opposite gold facets was correct for liquid TIP3P water in bulk. All other conditions were the same as those in Section 6.2.1 other than the cut-offs used in conjunction with non-bonded interactions were slightly modified in the case of the Au(111) and Au(100)(1×1) simulations. (Specifically, PME summation was cut-off at 13 Å whilst a force-switched cut-off starting at 10 Å and ending at 11 Å were used for LJ non-bonded interactions.)

6.2.3 Analysis

The structure of water at all three interfaces was analysed using data from the final 2 ns of the respective simulations. The mass density profile of water above the surface was used to define the thickness of the first adsorbed water layer. Specifically, this layer of water was defined as those molecules residing between the gold surface and the z co-ordinate corresponding to the position of the first minimum in the water density profile. In contrast to analysis carried out in other Chapters, here hydrogen bonds were defined using the criteria of Cicero *et al.* to facilitate comparison between the two studies [Cicero et al. [2011]]. Namely, in this Chapter, a hydrogen bond was present on the simultaneous satisfaction of an O$\cdots$O distance of less than 3.5 Å and an O$\cdots$HO angle greater than 140°.

The orientation of a water molecule with respect to the surface was monitored using its dipole moment. Specifically, $< cos(\theta) >$ was calculated as a function of the perpendicular distance from the interface, where θ was defined by the angle between the water dipole moment and the surface normal. Using this definition, $cos(\theta)$ can vary from -1 (corresponding to both H atoms pointing towards the surface) to $+1$ (corresponding to both H atoms directed away from the surface); a water molecule with its dipole moment oriented perpendicular to the surface normal will have $cos(\theta)=0$.

6.3 Results

6.3.1 Evaluation of vdW-DFT

In recent years the vdW-DF functional has been successfully used to model small molecule-Au(111) adsorption [Toyoda et al. [2009]; Li et al. [2012]; Wellendorff et al. [2010]; Rosa et al. [2012]]. However, these studies are limited in number and mainly consider aromatic absorbates only. Moreover, the structure of the gold itself, and in particular the reconstruction of the Au(100) surface, has not been probed in depth. As mentioned in Section 2.1.1, whilst it has been shown that vdW-DF can accurately reproduce van der Waals interactions at minimum energy separations, a number of limitations have also been identified [Lee et al. [2010]], chiefly: 1) equilibrium separations can be overestimated; 2) dispersion interactions can be overestimated at intermediate separations; and 3) hydrogen-bond strengths can be underestimated. In light of these factors it was necessary to evaluate the suitability of using vdW-DF/revPBE here. This was accomplished by first comparing experimental [Wetterer et al. [1998]; Syomin and Koel [2002a]; Lavrich et al. [1998]; Syomin and Koel [2002b]] and calculated interaction energies for a range of adsorbates on Au(111). Second, the structure of the PW-DFT optimised Au(100)(5×1) surface was benchmarked against experimental data [Ocko et al. [1991]] and previous first-principles calculations [Jacob [2007]; Shi and Stampfl [2008]; Havu et al. [2010]].

in vacuo Small Molecule Adsorption Energies on Au(111)

Figure 6.5 shows that vdW-DF is, in fact, able to reproduce experimentally-determined interaction energies for small-molecule gold adsorption reasonably well. The slight tendency for PW-DFT interaction energies of unbranched molecules of long aspect ratio to be larger than the experimentally reported values is consistent with a recent MD study [Fichthorn et al. [2007]]. Specifically, Fichthorn *et al.* found that the pre-exponential factor used in the method to calculate alkane interaction energies from raw experimental data increased with chain length, a fact that had not been accounted for [Wetterer et al. [1998]]. As a result, the experimental interaction energy of dodecane with Au(111) was thought to be underestimated by as much as $24\,\mathrm{kJ\,mol^{-1}}$ [Fichthorn et al. [2007]]. Taking this into account, the evidence suggests that vdW-DF is a reliable, computationally-accessible first-principles method, that accounts for dispersion. The functional is hence appropriate for use in determining the small molecule-gold interaction energies needed herein to derive FF parameters.

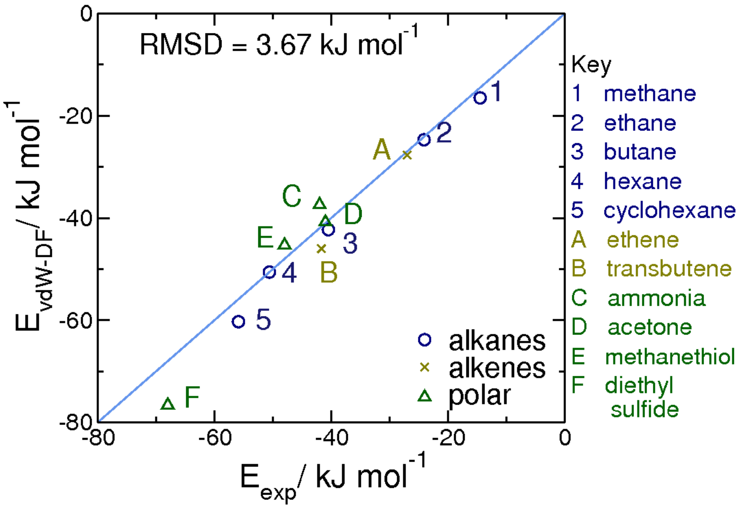


Figure 6.5: Correlation between experimental interaction energies for small molecules adsorbed onto the Au(111) surface and those calculated with vdW-DF [Dion et al. [2004]; Thonhauser et al. [2007]; Roman-Perez and Soler [2009]].

The Au(100)(5×1) surface

After relaxing the Au(100)(5×1) surface from an initial structure–featuring surface Au atoms arranged in an ideal hexagonal over-layer above an ideal bulk Au(100)(1×1) lattice–using vdW-DF, significant undulations were present in the uppermost layer (Table D.1). In addition, expansion of the Au-Au interlayer distance by 20% at the interface, also due to the spatial mismatch between surface and underlying Au atoms, was observed (Table D.2). To compare the structure of the reconstructed Au(100) surface relaxed using vdW-DF with previous work at this interface [Ocko et al. [1991]; Jacob [2007]; Shi and Stampfl [2008]; Havu et al. [2010]] both effects–surface buckling and interlayer expansion–were quantified using the order parameters defined by Havu *et al.* [Havu et al. [2010]] (Tables D.1 and D.2). Importantly the results for vdW-DF functional, together with the (5×1) approximation to the Au(100)-*hex* reconstruction, were in good agreement with experimental data [Ocko et al. [1991]].

Six different surface Au atom environments exist along the $[01\bar{1}]$ direction (herein referred to as the direction of the reconstruction) on the relaxed Au(100)(5×1) surface (Figure 6.4). Top sites T0, T2 and T4 all have Au-Au-Au angles greater than 180° (190.7°, 189.7° and 189.4° respectively), making them slightly more electron deficient than top sites T1, T3 and T5, all of which have Au-Au-Au angles less than 180° (175.1°, 160.3° and 174.9° respectively, Figure D.6). Consistent with this trend, T2 was identified by geometry optimization calculations of the water-Au(100)(5×1) system to be the most strongly adsorbing of all the interaction sites; top site T1 was the least adsorbing. The difference in energy between the two interaction sites was, however, just $2.7\,\mathrm{kJ\,mol^{-1}}$ (Table D.3). Similarly, geometry optimization calculations of the methane-Au(100)(5×1) system in which the binding affinity of different hollow sites was probed revealed a maximum difference of $1.4\,\mathrm{kJ\,mol^{-1}}$ between the sites (Table D.3, Figure D.1). Weakly chemisorbing gold species (thiols, dialkyl sulphides and nitrogen moieties possessing a lone pair of electrons), on the other hand, showed a significant energetic preference on adsorption, for protruding top sites T0 and T2 (Table D.3).

6.3.2 Force-field Parameterisation

Implicit in the derivation of all FFs is the balance between chemical and physical accuracy and practical implementation. At the first principles level of theory it can be shown that the interaction between an atom in an adsorbing molecule and gold is unique to the specific chemical environment of both the adsorbate and gold. For instance, the presence of the electron withdrawing carbonyl group in an amide, compared to an alkyl moiety in an amine, means that the overlap of N and Au atomic orbitals on adsorption to a gold surface is different for the two nitrogen atoms. Similarly, the spatial arrangement of gold atoms, and therefore the surface electron density distribution, differs from one crystallographic plane of gold to another, thus requiring each surface to be parametrized separately. On the other hand, a FF parametrized to contain a unique Au-X (X= any atom) parameter for each atom in each molecule would, most likely, be overparametrized.

In the FF derived herein, parametrized to describe the interaction of peptides with the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces, only a small number of different Au-X parameters were in fact necessary for each interface. Cross-terms fit for one species adsorbed at a given surface were often transferable to similar chemical moieities bound to the same crystallographic plane. However, unlike CHARMM/CVFF-METAL [Feng et al. [2011]; Pandey et al. [2009]; Heinz et al. [2009]], parameters were not transferrable *between* surfaces. Tests were carried out to see if parameters derived to capture peptide adsorption at the aqueous

Au(111) interface could be directly transferred, unaltered, to the Au(100)(5×1) surface due to strong similarities in the spatial arrangement of surface Au atoms between the two. Not only was the fit, relative to the reference PW-DFT data, of the Au(100)(5×1) surface FF compromised by adopting this approach, but the Au(111) FF terms could not reproduce the coupled energetic/spatial heterogeneity of adsorption of weakly chemisorbing species onto the Au(100)(5×1) surface. Hence, in GolP-CHARMM, a separate set of parameters for each gold surface has been derived. Importantly though, the three interfaces–Au(111), Au(100)(1×1) and Au(100)(5×1)–were parametrized in a consistent manner to enable the FF to be used to model direct facet-dependent differences in adsorption.

Optimal GolP-CHARMM parameters for the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces are presented in Tables D.4, D.5 and D.6, respectively. Data used in the fitting process is presented in Tables D.7, D.10, D.12, D.13 and D.14. In what follows, the parameters for each interface are discussed in turn.

Au(111)

Following the GolP methodology, and for reasons outlined in Section 6.2, virtual interaction sites exist in GolP-CHARMM to enable the FF to reproduce the favourable geometries of small molecules relative to the underlying gold surface identified by the vdW-DF calculations. For the Au(111) surface there are two virtual sites (denoted AUI) for every real gold surface atom (denoted AUS), located in the two hollow sites of each surface unit cell (Figure 6.2a). It has previously been shown that the same LJ σ_{Au-Au} and ϵ_{Au-Au} parameters can be applied to both the surface virtual interaction sites and gold atoms within the bulk (denoted AUB) without loss of accuracy [Iori et al. [2009]].
Generic Au–Au parameters Using the generic Au–Au LJ parameters (to be combined with all CHARMM atom types other than the special cases described below) the RMSD between experimental [Wetterer et al. [1998]] and GolP-CHARMM interaction energies for alkane adsorption was 1.88 kJ mol^{-1} (Figure D.7). The marked difference between these GolP-CHARMM parameters and those used in the existing GolP FF is a direct consequence of both energetic *and spatial* data being used in the fitting procedure. In fact, choosing LJ parameters for GolP-CHARMM that were closer to the OPLS-AA values (σ_{Au-Au}=3.2 Å, ϵ_{Au-Au}=0.75 kJ mol^{-1} GolP-CHARMM; σ_{Au-Au}=3.2 Å, ϵ_{Au-Au}=0.65 kJ mol^{-1} OPLS-AA) also gave a reasonably good fit to the adsorption energy for alkanes (RMSD = 2.92 kJ mol^{-1}) but resulted in adsorption geometries for which the molecules were too close to the surface.
Parameters for sp^2 hybridised carbon Following the procedure originally developed

for GolP [Iori et al. [2009]], a refined set of Au–Au LJ parameters, to be used in conjunction with existing CHARMM LJ parameters for sp^2 hybridized carbon atoms (*eg.* carbon atoms either within aromatic rings or those in isolated/conjugated C=C bonds), were needed. In the absence of these customised terms, the affinities of ethene [Wetterer et al. [1998]], benzene [Wetterer et al. [1998]; Syomin et al. [2001]] and 1,3-butadiene [Wetterer et al. [1998]] for the Au(111) surface were underestimated to a similar degree in GolP-CHARMM as that reported previously for OPLS-AA [Iori et al. [2009]]. This refinement of the generic Au–Au LJ parameters for aromatic C was also required for the two Au(100) surfaces.

Parameters for chemisorbing species Sulphur containing moieties and functional groups featuring a nitrogen atom with a lone pair of electrons were shown previously weakly to chemisorb onto the Au(111) facet [Iori et al. [2009]]. PW-DFT calculations, performed here using vdW-DF, confirmed these findings. It was not possible to represent this weak chemisorption using standard mixing rules to combine CHARMM potentials with the generic Au–Au potential described so far: the resulting FF substantially overestimated the binding distance, and underestimated the energy of interaction. Therefore specific LJ cross-terms for Au–S (fitted to methanethiol and methyl sulphide adsorption) and Au–N interactions were introduced. For nitrogen it was also essential to differentiate between N atom types in aliphatic and aromatic groups when fitting these special interaction parameters. This perhaps was not surprising considering differences in the electron density distribution surrounding N atoms in the two chemical environments; a phenomenon which not only influences the magnitude of the image charge induced in the metal slab, but also the degree of orbital overlap between N and surface Au atoms. Therefore separate Au–N cross-terms for unprotonated histidine residues (to reproduce imidazole-Au(111) adsorption (atom type 'NR2')) and for unprotonated amino-acid/peptide N-terminal groups and lysine residues were fitted. This second set was fitted to methylamine-Au(111) adsorption. Methylamine, featuring an alkyl group, is a better analogue of the unprotonated N-terminus than ammonia, originally used to fit FF parameters in GolP [Iori et al. [2009]]. Since polar hydrogens bonded to N already have a small spatial extent in CHARMM, unlike OPLS-AA, no additional Au–H(N) parameters for the *amine* H atoms were required. The same procedure–involving fitting *two* sets of LJ cross-terms for N–was also found to be necessary for the native and reconstructed Au(100) interfaces.

Parameters for Oxygen The issues with the heteroatoms N and S also extended to organic oxygen. In particular, using only the terms described above, the FF was able to reproduce the correct interaction energy for methylamide adsorbed onto the

Au(111) surface, but gave the wrong geometry; the Au–O interaction was too weak in comparison to the Au–N. In fact, the affinity of all oxygen containing moieties to Au(111) was consistently underestimated relative to PW-DFT. Therefore the FF was extended by the inclusion of a specific Au–O cross-term. It is important to note that for GolP-CHARMM correctly to describe both the binding energy and geometry of *amide* group adsorption onto gold (at all three surfaces–Au(111), Au(100)(1×1) and Au(100)(5×1)) a new atom type must be introduced into CHARMM to differentiate between amide (labelled 'NH2') and unprotonated amine N (herein labelled 'NHT2'). The refined Au–N cross terms derived for each surface should be applied only to the interaction of the former, the unprotonated *amine* nitrogen atoms, with gold (*i.e.* Au–NHT2 intearations).

In re-parametrising the Au(111) surface to be compatible with CHARMM, a specific aspect of the description of the Au(111)-liquid water interface that remained previously unresolved in the two existing classical FFs (Section 1.4.2) has been addressed. Recent CPMD simulations using the PBE functional revealed that interfacial water molecules have an enhanced tendency to act as hydrogen bond donors, relative to those molecules in bulk liquid water [Cicero et al. [2011]]; a finding subsequently confirmed by BOMD simulations of the same interface by Nadler and Sanz using dispersion corrected functionals [Nadler and Sanz [2012]]. This behaviour is not captured by the original GolP/SPC and CVFF-METAL/SPC models [Cicero et al. [2011]]. A correction for GolP/SPC (*i.e.* to add a small Au–H(O) LJ interaction) was identified in Cicero *et al.* [Cicero et al. [2011]]. Here, to reproduce the first-principles behaviour using GolP-CHARMM, it was critical to ensure that the vdW-DF energetics and geometry of a single water molecule adsorbed onto Au(111) in vacuum could be reproduced. When using the generic GolP-CHARMM Au–Au LJ terms in combination with water atoms, O and H, not only was the affinity of a TIP3P water molecule for Au(111) significantly underestimated ($-13.4\,\mathrm{kJ\,mol^{-1}}$ GolP-CHARMM *v*.s $-18.3\,\mathrm{kJ\,mol^{-1}}$ vdW-DF), but it was also mis-oriented with one of its H atoms pointing towards the gold surface (Figure 6.6a). Configurations similar to this were sometimes found with SPC waters at the (uncorrected) GolP Au(111)-water interface [Cicero et al. [2011]]. This orientation hinders the molecule from acting as a hydrogen bond donor to surrounding waters. In the current work it was necessary both to increase the strength of the interaction between the water oxygen atom and the gold surface, and to enlarge the spatial extent of its hydrogen atoms. Whilst it was not possible to make the preferred TIP3P-Au(111) conformation identical to that found in vdW-DF calculations (Figure 6.6b), the optimal configuration for a single water molecule adsorbed at the interface (Figure 6.6c) at least does not impair its potential to act as a hydrogen bond donor when surrounded by liquid water.

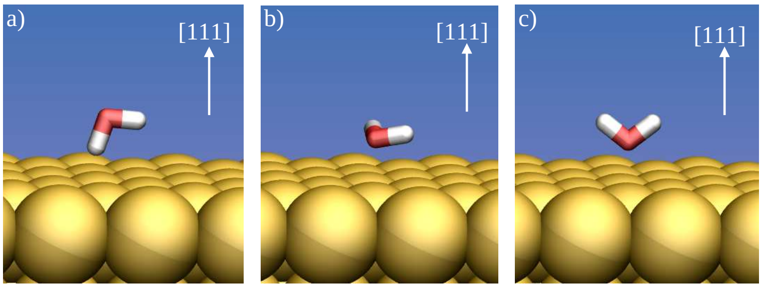


Figure 6.6: The most favourable orientation of a single water molecule adsorbed onto Au(111) a) using the FF without extra Au–O LJ parameters, b) using vdW-DF and c) using the new FF with Au–O LJ parameters.

Au–O and Au–H LJ cross parameters that were fitted to water adsorption were found subsequently to be transferable to other small molecules. Only for methanol was the affinity for gold still underestimated by the FF, although this discrepancy was reduced from 8.7 kJ mol^{-1} to 5.5 kJ mol^{-1} by the inclusion of Au–O and Au–H parameters.

in vacuo FF validation As an initial validation of the Au(111) GolP-CHARMM parameter set, the *in vacuo* binding energies of a number of small molecules that had not been used in the fitting process were calculated and compared with experiment [Wetterer et al. [1998]; Syomin and Koel [2002a,b]] (Table D.8). The test set was chosen to benchmark the interaction of a range of different functional groups, both independently and in combination, with Au(111). For example, the transferability of the generic Au–Au LJ parameters derived using linear alkanes to cyclic systems was tested with cyclohexane, whilst the ability of the new FF to describe the adsorption of a molecule with more than one independently parametrized functional group tested with non-1-ene, trans-2-butene, ethanethiol, dibutyl sulphide and toluene. The agreement between GolP-CHARMM and experiment was comparable to that of GolP (RMSD of interaction energy of 4.15 kJ mol^{-1} for GolP-CHARMM; *cf.* 4.97 kJ mol^{-1} for GolP [Iori et al. [2009]] for a similar set of test molecules, see Table D.8 for details).

Ideally, FF parameters for all of the functional groups present in a peptide would be benchmarked against experimental results obtained in a consistent manner, but such data are not available. Therefore the transferability of the Au–O and Au–H parameters, which were not probed by the experimental test set (Table D.8), was tested using vdW-DF calculations (Table D.9). For charged molecules, extensive calculations carried out as part of this work revealed that the PW-DFT interaction energies depended on the simulation cell dimensions–far more strongly than would be expected for electrostatic interactions alone [Makov and Payne [1995]]. Thus it was not possible to obtain reliable *in vacuo* gold-binding affinities for these species using vdW-DF (for any of the three surfaces considered here), against which the FF could be compared. However, NH_3^+ and COO^- moieties, in particular, are prevalent in biomolecules, being featured by all uncapped peptide termini under neutral pH conditions. Hence it was essential for future meaningful bio-interfacial-gold simulations to be sure that the new CHARMM FF parameters were also transferable to these moieties. The potential energies of adsorption at the aqueous Au(111) interface for 5 amino-acids–Ala, Asp, Arg, Lys, His–were calculated from MD simulation. The trend in the binding affinities of these residues (reported in Reference Wright et al. [2013b]), which included both positively and negatively charged functional groups, compared favourably with experimental evidence [Cohavi et al. [2011]; Peelle et al. [2005]; Hnilova et al. [2008]].

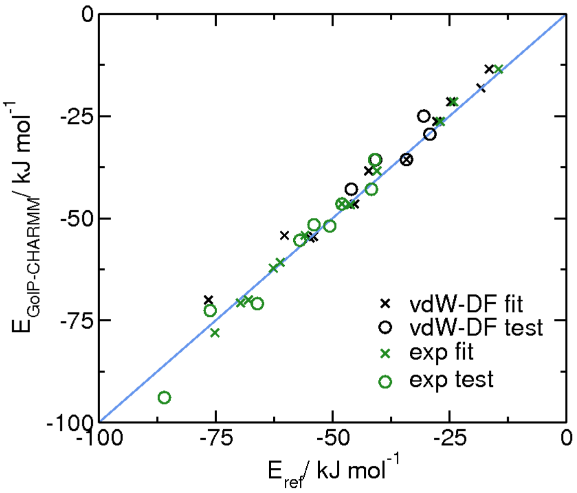


Figure 6.7: Comparison between the interaction energies of small molecules with Au(111) *in vacuo* calculated using GolP-CHARMM FF and either experimental data (green) or those calculated using the vdW-DF exchange functional (black). Both molecules used to derive FF parameters (crosses) and those in the test set (open circles) are shown. The solid blue line denotes 1:1 correlation.

Au(100)(1×1)

Generic Au–Au parameters For the Au(100) surface–both native and reconstructed forms–in the absence of appropriate experimental data, all reference small-molecule interation energies used in the FF fitting and validation process were obtained from vdW-DF calculations. Specifically, for the two Au(100) interfaces, the generic Au–Au LJ parameters (for use with all CHARMM atom types, other than those individually parametrized below) were fitted to the interaction energies of methane and butane, with the respective facet, only. Larger alkanes were not used since hydrocarbon chains of longer lengths are not present in peptides.

Parameters for sp^2 hybridised carbon As with Au(111), a number of gold-organic cross-terms were refined, including those between aromatic C and Au atoms, satisfactorily to fit the FF for the Au(100)(1×1) surface. In general, the transferability of the cross-terms, between atoms in similar chemical environments, was slightly less effective at this interface, however. This is best exemplified by the adsorption of hydrocarbon molecules with sp^2 hybridized carbon atoms. Specifically, unlike Au(111), it was not possible to derive a single set of refined Au–Au LJ parameters for the native Au(100) surface that was transferable between different chemical environments (*eg.* aromatic rings, isolated C=C and conjugated C=C bonds). The computational expense of the PW-DFT calculations necessary to derive and test modified Au–Au parameters for different types of C=C bonds (isolated and conjugated) was beyond the scope of this work (to present a FF for polypeptide-gold adsorption where isolated C=C bonds are not present) and so was not carried out. As a repercussion, the customised Au–Au parameters present in Table D.5, fitted to benzene adsorption, are for use with sp^2 hybridised carbon atoms in *aromatic* systems only. These terms, however, transferred well to other small molecules featuring aromatic rings (*e.g.* phenol, indole and toluene (Table D.11)).

Parameters for chemisorbing species A greater dispersity of interaction, relative to Au(111), was also seen for sulphur with the Au(100)(1×1) surface. In particular, the gold-binding affinities of thiols and sulphides were quite different–a characteristic common to both forms of the Au(100) surfaces. Therefore, two sets of Au–S cross-terms were defined, parameterized using methanethiol and dimethyl sulphide, respectively. Like nitrogen, a new atom type was introduced into CHARMM to distinguish between thiol (herein labelled ‘ST’) and dialkyl sulphide (labelled ‘S’) sulphur atoms. The organic interactions associated with these two S atom types are identical and it is only the LJ gold cross-terms which differ. Like Au(111), two sets of Au–N LJ cross-terms (Au–NR2 and Au–NHT2) were fit for the Au(100)(1×1) surface.

Parameters for Oxygen In a manner consistent with the Au(111), Au–O and Au–H cross-terms were fit to PW-DFT data–energetics and adsorbed geometry–for a single water molecule at the Au(100)(1×1) interface. However, in the absence of experimental or first principles MD data, validation was difficult. Specifically, unlike Au(111) there is no evidence to suggest what the favourable water-gold orientation is at the gold-liquid water Au(100) (either native or reconstructed) interface. Instead, FF parameters were checked against first-principles data for the adsorption of a single water molecule *in vacuo*. Specifically, the energetic trend in interaction energies for binding to different surface sites (bridge, hollow, top) and/or different water-gold geometries ('O down', '1H down', '2H down' and 'flat') was considered. Agreement between FF and vdW-DF, at least at this single molecule level, was found to be favourable.

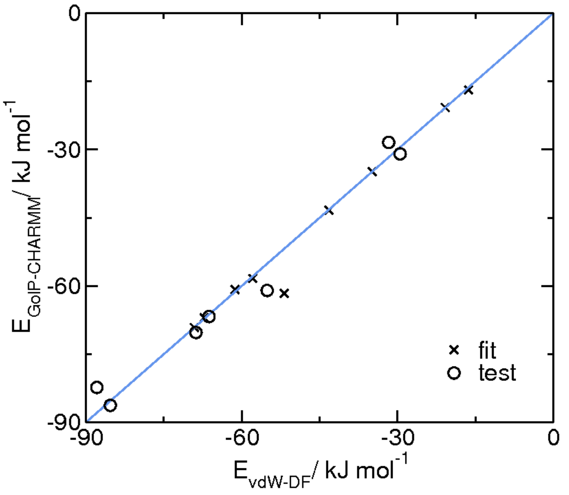


Figure 6.8: Comparison between the interaction energies of small molecules with Au(100) *in vacuo* calculated using GolP-CHARMM to those calculated using the vdW-DF exchange functional. Both molecules used to derive FF parameters (crosses) and those in the test set (open circles) are shown. The solid blue line denotes 1:1 correlation.

Unlike the Au(111) interface, in the absence of specific Au–O parameters, only the binding affinities of water and methanol were underestimated relative to vdW-DF. Therefore, these cross-terms are only to be used with water and hydroxyl oxygen atoms attached to aliphatic groups (herein labelled 'OHM1'); they are not appropriate for -OH attached to aromatic or acid groups such as those found in tyrosine and glutamic acid, respectively (these retain the default atom type of 'OH1'). A second set of Au–O LJ parameters were needed for carbonyl oxygen, however, in order for the FF correctly to reproduce the first-principles trend in equilibrium separations of methanoic acid and methyl amide adsorbed onto the Au(111) and Au(100)(1×1) surfaces.

in vacuo FF validation As an initial test of the Au(100) FF parameters, the interaction energies of seven small molecules (methanol, indole, methanoic acid, phenol, ethanethiol, diethyl sulphide, toluene) that had not been used in the fitting procedure were used to benchmark GolP-CHARMM against the results of the PW-DFT calculations (Table D.11). Most of the different functional groups parametrized above were represented by this group, either independently or in combination. The agreement between FF and vdW-DF interaction energies for the test set was good, with an RMSD in the interaction energies of $3.64\,\mathrm{kJ\,mol^{-1}}$ (Figure 6.8). The largest outlier was ethanethiol, where the strength of Au(100)(1×1) binding was overestimated by approximately 12%.

Au(100)(5×1)

As stated in Section 6.2, the spatial and energetic heterogeneity of the reconstructed Au(100) surface necessitates the use of two sets of virtual interaction sites in the FF (Figure 6.2c). Parameters (generic Au–Au, refined Au–Au for sp^2 hybridised carbon and special Au–O cross terms) for physisorbing species were derived in a manner consistent with the Au(111) and Au(100)(1×1) surfaces [Wright et al. [2013b]]. The derivation of these LJ terms, which apply only to virtual interaction sites located in the surface hollows, AUIU, is outlined in detail in reference Wright et al. [2013c]. In this Chapter, the Au–N and Au–S cross terms, defined for both types of virtual interaction site–AUIU and Sx (where 'x'$=0, 1, 2, 3$)–are discussed only.

The first stage of this fitting procedure was to test whether the significant differences (of up to $20\,\mathrm{kJ\,mol^{-1}}$, Table D.3), identified by vdW-DF calculations, in adsorbate interaction energies at the different top site locations on the Au(100)(5×1) surface could be reproduced by the FF using just a homogenous set of hollow site Au–N/Au–S parameters. That these tests were negative suggests the energetic heterogeneity of the surface found in PW-DFT calculations (for example, the preferential adsorption of methanethiol to T2 surface sites over T3) did not solely arise from the geometrical arrangement of Au atoms; this is identical in both FF and first-principles slab models (Section 6.2). Instead I propose that electron density is unevenly distributed across the Au(100)(5×1) surface, and it is this which can direct the adsorption of species with lone pairs of electrons to electron-deficient sites.

Special parameters for Nitrogen Like Au(111) and Au(100)(1×1), two sets of Au–N LJ parameters were needed for the Au(100)(5×1) surface: Au–NR2 and Au–NHT2. Both sets of refined Au–N parameters were fitted using 'site specific' annealing: Au–NHT2 to methyl amine adsorption, and Au–NR2 to imidazole adsorption (Table D.13). In each case a small Au–N LJ cross-term was added to the hollow interaction sites 'AUIU', while the σ and ϵ parameters of the bridge interaction sites were adjusted in turn to incorporate spatial heterogeneity into surface adsorption. The σ/ϵ parameters given to S0 and S2 sites were significantly smaller/larger than those assigned to S1 and S3 sites in order to represent the increased acidity of Au atoms in T0 and T2-type locations. Cross-terms were subsequently tested by annealing both molecules from vacuum to the surface to allow full relaxation of internal adsorbate and adsorbate-gold conformation (D.12). In the case of imidazole the optimal adsorbed location identified by these tests was atop a T2-type site, in agreement with vdW-DF; for methyl amine it was atop a T0 site. Whilst a preference for T2 over T0 site adsorption was found using vdW-DF for the latter adsorbate, methyl amine, the energetic difference between the two locations was slight, $2.2\,\mathrm{kJ\,mol^{-1}}$.
Special parameters for Sulphur Refined parameters for Au–S interactions–one set for thiols and another for dialkyl sulphides–with the Au(100)(5×1) interface were fit in a similar manner to that described above for nitrogen. Namely, LJ terms for both types of virtual interaction sites, hollow (AUIU) and bridge (Sx), were fit using 'site specific' annealing (thiol Au–ST to methanethiol adsorption, and dialkyl sulphide Au–S to dimethyl sulphide adsorption, respectively, Table D.14); the refined parameter sets were subsequently tested using the more extensive annealing protocol. The atom-type 'ST', introduced to differentiate between thiol and dialkyl sulphide adsorption onto the Au(100)(1×1) surface, was also used for the reconstructed interface. For methanethiol and dimethyl sulphide the energetic (D.12) and spatial agreement between FF and vdW-DF was good, with T2-type site adsorption being optimal for both.

The transferability of the refined Au–S LJ parameters from one thiol/dialkyl sulphide to another was also assessed by calculating the interaction energies of ethanethiol and diethyl sulphide with the Au(100)(5×1) surface respectively (Table D.12). Similar to Au(100)(1×1), the interaction energy of ethanethiol with Au(100)(5×1) was overestimated by approximately 12%, relative to vdW-DF, using FF parameters fit to methanethiol adsorption. Experimentally, although it has been shown that the strength of thiol physisorption to the Au(111) surface *in vacuo* increases linearly with alkyl chain length for $C_nH_{2n+1}SH$, $n \geq 2$ [Lavrich et al. [1998]], there is no evidence to suggest whether this trend can be extrapolated to methanethiol and to adsorption onto the Au(100) surface or not. Hence it is speculated that the overlap of Au and S atomic orbitals is significantly hindered on the addition of a second carbon atom to the alkyl chain of a protonated thiol, at least in the case of adsorption to the Au(100) surface. These quantum effects would account for the non-linear increase in interaction energy from methanethiol to ethanethiol Au(100) adsorption; a feature that cannot be reproduced by an additive FF such as that derived here. However, the ultimate aim of the work presented in this Chapter was to derive a FF to describe the interactions of peptides with Au(100) under aqueous conditions. Since methanethiol is a closer analogue of the cysteine side-chain than ethanethiol, the former, methanethiol, was chosen for fitting Au–ST parameters for both forms of the Au(100) surface.

Akin to ethanethiol, the adsorption of diethyl sulphide to the Au(100)(5×1) surface, after annealing from vacuum, was overestimated by the FF. There were small differences in the optimal diethyl sulphide–gold geometries identified by the FF and PW-DFT, including the angle between the surface normal and a vector from S to the mid-point between the two methylene C atoms (66.3° and 20.3° for the FF and vdW-DF, respectively). An additional PW-DFT calculation, initiated in the optimal FF geometry above the T2 top site, revealed that this conformation was not favourable at the first-principles level of theory. However, on closer inspection, the internal conformation of diethyl sulphide itself was slightly different after relaxation using vdW-DF, compared to that found by the FF. For example, the C–S–C angle was larger (102.7° compared to 97.2° for vdW-DF and the FF, respectively) and the C–C–S angles more varied (110.2° and 117.0° compared to 113.3° and 113.7° for vdW-DF and the FF, respectively). Hence, part of the discrepancy between the optimal diethyl sulphide-gold FF and vdW-DF interaction energies arose from differences in the internal configuration of the molecule described by the two methods; a feature of the CHARMM FF itself rather than the GolP-CHARMM terms fit to reproduce its interaction with the gold surface. This was supported by the much closer agreement between PW-DFT and FF interaction energies observed when the ‘site-specific’ annealing protocol was used (Table D.14); in this case the simulations were too short for the internal conformation of diethyl sulphide to be relaxed from its initial DFT optimised geometry to that of the FF. Importantly, however, GolP-CHARMM Au–S parameters can reproduce the features in diethyl sulphide binding to Au(100)(5×1) which are most relevant for describing methionine adsorption (S as the closest point of contact with the surface, with both flanking methyl groups pointing away from the surface).

6.3.3 Aqueous Au(111) and Au(100) Interfaces

It is essential for future bio-interfacial simulations that the new FFs can model liquid water at the Au(111) and Au(100) surfaces, since it has been previously shown that water structuring above an inorganic surface can strongly influence biomolecule adsorption (Section 1.2.4, [Skelton et al. [2009]; Jena and Hore [2010]; Schneider and Ciacchi [2012]]). However, few unambiguous *atomistic* descriptions of *liquid* water adsorbed onto gold exist. Only the *aqueous* Au(111) interface has been previously studied at the first-principles level of theory [Cicero et al. [2011]; Nadler and Sanz [2012]], although several groups have used BOMD to probe water bilayer adsorption onto both Au(111) [Meng et al. [2004]; Schnur and Gross [2009]] and Au(100)(1×1) [Lin and Gross [2012]] *in vacuo*. In this section GolP-CHARMM parameters for the aqueous Au(111) surface are first benchmarked against the CPMD simulations of Cicero *et al.* [Cicero et al. [2011]]. The structure of liquid water adsorbed to the three gold surfaces parameterised here is then compared to probe whether facet selective differences occur. These simulations using the new FF offer predictions which, in the future, could be confirmed or falsified by either experiment or first-principles calculations.

Au(111): Comparison between GolP-CHARMM and CPMD

In general, the structure of liquid water above Au(111) observed in the simulations performed here using GolP-CHARMM was in good overall agreement with the previous CPMD study [Cicero et al. [2011]]. Specifically, the new FF reproduced the enhanced tendency of interfacial water molecules to act as hydrogen bond donors compared with those in bulk liquid water (Figure 6.9b). The ability of the FF to capture these subtle water characteristics is critical for improving the accuracy of bio-interfacial simulations, where hydrogen bonding to first-layer adsorbed waters may enhance or facilitate biomolecule-surface binding [Schneider and Ciacchi [2012]].

Detailed analysis of water-gold conformation as a function of surface separation revealed that, for the Au(111) surface, molecules in the first layer are oriented in an ordered fashion, with a preference for the oxygen atom in a water molecule to lie closer to the interface than the hydrogen atoms (Figure 6.9c). The modal angle between the dipole moment of these waters and the surface normal was 90° however, similar to the geometry of a single water adsorbed *in vacuo* onto the same surface calculated using vdW-DF (Figure 6.6b). Conformations such as this were also observed to be favourable for the subset of molecules residing in the first adsorbed layer in the CPMD simulations [Cicero et al. [2011]].

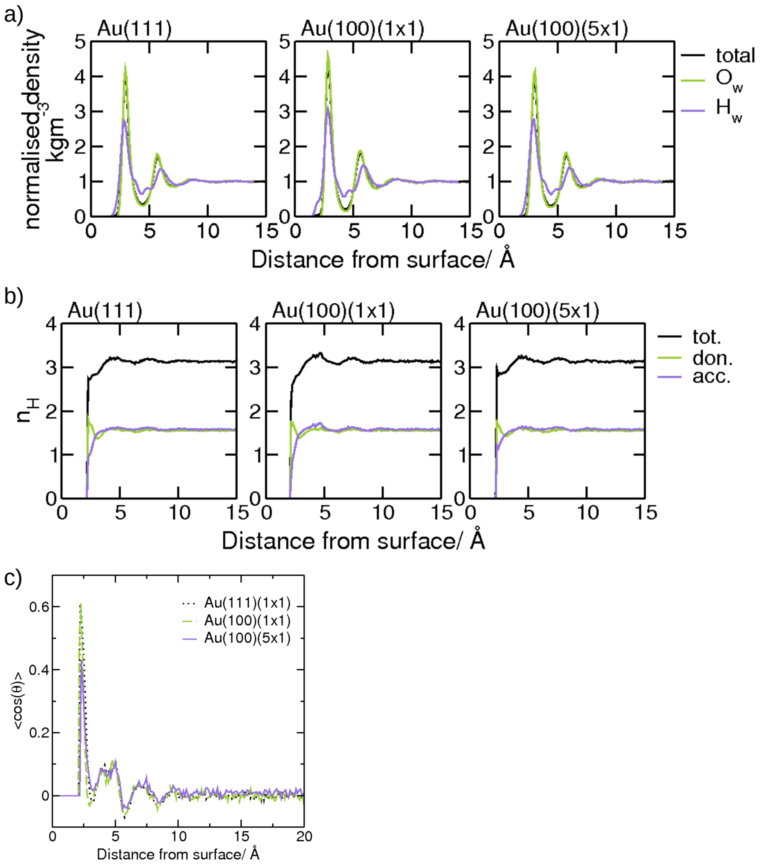


Figure 6.9: A comparison of the structure of water above the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces using three metrics: a) normalised mass density profiles, b) total number and number of donor/acceptor hydrogen bonds as a function of surface separation and c) water dipole moment.

GolP-CHARMM also shows improvement in its description of the Au(111)-water interface over both the original GolP and the CVFF/SFC FFs. Unlike GolP, GolP-CHARMM does not require any further corrections [Cicero et al. [2011]] to be able to reproduce the predominance of water molecules acting as hydrogen bond donors. Whilst, at odds with CVFF/SFC, it correctly reproduces the preference of water for top site adsorption. In addition, it must be remembered that only GolP-CHARMM was specifically fit to reproduce the *energetics* of single water molecule adsorption onto the Au(111) surface to PW-DFT data. It is worth noting that if direct comparison is made between GolP-CHARMM and its predecessor, GolP [Iori et al. [2009]], then subtle changes in the structure and affinity of water at the aqueous Au(111) interface, brought about by the additional Au–O and Au–H(O) terms in GolP-CHARMM, may impact on biomolecule adsorption. For instance, the interaction energy of a single water molecule *in vacuo* for Au(111) was $-14\,\mathrm{kJ\,mol^{-1}}$ using GolP [Iori et al. [2009]] but $-18.1\,\mathrm{kJ\,mol^{-1}}$ using GolP-CHARMM. Thus the competition between biomolecule and water adsorption on the gold surface is predicted to be less enthalpically favourable for GolP-CHARMM than the original GolP. However, at the same time, the denser, more structured first adsorbed water layer featured by GolP-CHARMM means that the entropy change brought about by displacement of these waters into solution by an adsorbate would be larger. These two effects may partially cancel each other out.

Some discrepancies between GolP-CHARMM and the CPMD simulation of Cicero *et al.* [Cicero et al. [2011]] remain, however. First, using the FF, only a single population of water molecules in the first layer exist, with no shoulder on the O_w mass density profile in the direction normal to the surface. Second, interfacial water remains relatively over-structured, with the density of the first adsorbed layer being greater using GolP-CHARMM than that found from first-principles. Despite these small differences, GolP-CHARMM can acceptably describe the aqueous Au(111)-gold interface.

Facet Dependent Water Structuring

The careful derivation of FF parameters for the Au(111) and Au(100) surfaces in a consistent manner allows detailed comparison between the structure of liquid water at each interface for the first time (Figures 6.9 and 6.10). In the past the water-Au(100)-*hex* interface has been approximated by the water-Au(111) one, since the hexagonal arrangement of Au surface atoms is similar in both cases [Lin and Gross [2012]]. This assumption was recently brought into question by FF-based MD simulations which found differences in the structure of the aqueous MgO interface existed when two different surface models–one atomically flat (akin to Au(111)) and the other corrugated (akin to Au(100)(5×1)–were employed [Phan et al. [2012]]. In the case of gold though, the simulations performed in this Chapter show that many of the properties of the planar aqueous gold interface are, in fact, common to the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces.

For all three surfaces there is a preference for interfacial waters to be oriented 'O down' (inferred from the positive dipole moment, Figure 6.9c), facilitating their enhanced ability to act as hydrogen bond donors to surrounding water molecules (Figure 6.9b). This preference is weakest in the case of the Au(100)(5×1) surface–most likely a consequence of its rumpled topography disrupting regular water packing. Unlike the native form of the surface though, there is no significant shoulder in the H_w mass density profile of water close to the Au(100)(5×1) interface to indicate the presence of a small subset of water molecules oriented in the opposite manner, 'H down' (Figure 6.9a).

The most profound differences between the three aqueous interfaces were observed in the *in-plane* distribution of water molecules, rather than structuring in the perpendicular direction (Figure 6.10). Specifically, although waters in the first adsorbed layer are localised atop Au atoms in each case, this is significantly more pronounced for the native Au(100) surface than the other two. The trend correlates loosely with that in water density of the first adsorbed layer as a whole, with the average mass of water per unit surface area being $339/331/324 \times 10^{-9}\,\mathrm{kg\,m^{-2}}$ for the Au(100)(1×1)/Au(100)(5×1)/Au(111) surfaces, respectively. (Density per surface area is a more appropriate metric than the maximum density per volume taken directly from Figure 6.9a due to undulations in the Au(100)(5×1) surface.) Interestingly, however, both observations differ slightly from the trend in the interaction energies of single water molecules adsorbed *in vacuo*, predicted by GolP-CHARMM: $-21.8/-20.9/-18.1\,\mathrm{kJ\,mol^{-1}}$ for Au(100)(5×1)/Au(100)(1×1)/Au(111).

Since both the degree to which liquid water close to the native and reconstructed Au(100) surfaces is structured and the geometrical arrangement of the interfacial molecules differ significantly, water structuring could play an important role in determining selectivity in biomolecule adsorption. Thus in order for simulation to aid the design of peptide sequences able to distinguish between Au(111) and Au(100), either the reconstructed status of the Au(100) surface must be known experimentally, or both the native and reconstructed forms of the interface should be considered in the modelling process–something which has been neglected in the past [Heinz et al. [2009, 2010]; Yu et al. [2010, 2012]]. Based on the structure of the aqueous interfaces alone, it could be suggested that the differences in the mech-

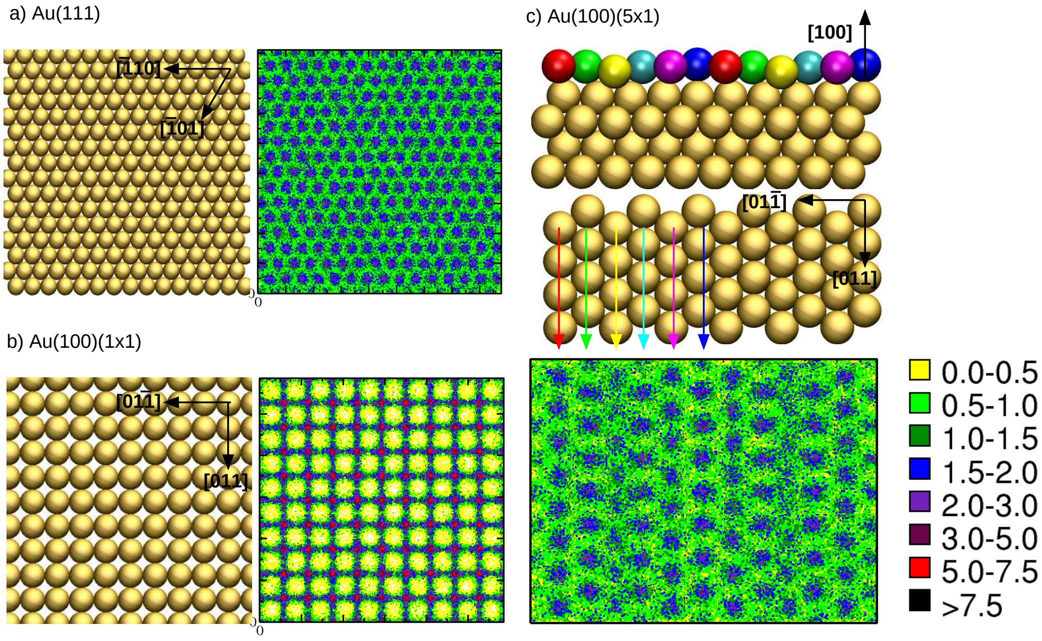


Figure 6.10: Spatial arrangement of Au atoms in the a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1) surfaces, together with the density of water in the plane of the interface, for the first adsorbed layer only.

anisms governing peptide selectivity between Au(111) and Au(100) on adsorption would be weakened by reconstruction of the latter surface.

6.4 Conclusions

In conclusion, in this Chapter a new FF for modelling peptide adsorption at the aqueous Au(111), Au(100)(1×1) and Au(100)(5×1) interfaces, GolP-CHARMM, has been derived using a combination of experimental and PW-DFT data. This FF is the first biointerfacial FF for gold to be specifically parameterised to reproduce both the energetics and geometry of an adsorbed water molecule relative to first-principles. In keeping with its predecessor, GolP [Iori et al. [2009]], refined sets of parameters for describing the interaction of a number of functional groups–sp^2 hybridised carbon atoms and the heteroatoms O, S and N–with gold were fit. These terms were needed to account for disparities in the physical and chemical binding characteristics (*e.g.* chemi- *v.s* physi-sorption) of each moiety. The energetic spatial heterogeneity of the Au(100)(5x1) surface for chemisorbing S and N species, observed in PW-DFT calculations (at this interface only), has also been accounted for.

As well as being a significant contribution to the realm of bio-interfacial simulation in its own right–by enabling selectivity in peptide adsorption at the aqueous Au(111) and Au(100) interfaces to be investigated by simulation–the protocol used to derive GolP-CHARMM can be adopted by others in the future to derive terms for other surfaces. If the same procedures are used for a range of noble metals (*e.g.* Ag, Cu, Pt, Pd) the adsorption of a specific peptide sequence at the aqueous gold and, say, silver surfaces could be investigated on a level-footing for the first time whilst accounting for the dynamic polarisation of metal atoms. The method by which surface heterogeneity in the reconstructed Au(100) facet is empirically accounted for here could also be built on to design FFs for other non-ideal metal surfaces such as those with defect sites. Differences between the aqueous Au(100)(1×1) and Au(100)(5×1) interfaces in Section 6.3.3 highlight the need for future computational models to account for the exact experimental structure of the surface, where possible.

As part of this study (but not reported in this Chapter for the sake of brevity) the adsorption of five amino acids–Ala, Arg, Asp, His (both protonated and unprotonated forms) and Lys–at each of the three aqueous gold interfaces parameterised in GolP-CHARMM were simulated [Wright et al. [2013b,c]]. Qualitatively the results suggested that, for the majority, the mode of binding and the strength of binding to the Au(111) and Au(100)(5x1) surfaces were similar. In the case of the native Au(100) surface, however, adsorption driven by consistent contact between adsorbate and gold atoms was only observed for chemisorbing species, HisA (unprotonated His). (Arginine also made direct contact with the surface in some runs.) The remaining amino acids were not able to penetrate the strong layer of interfacial waters. It was therefore interesting to investigate how the qualitative trends in adsorption between Au(111), Au(100)(1x1) and Au(100)(5x1) for small biomolecules, amino acids [Wright et al. [2013b,c]],are translated into the binding mechanism of whole peptides. The REST metaD simulations of AuBP-1–gold adsorption reported in Chapter 8 quantitatively address this phenomenon.

Chapter 7

Material Selectivity of QBP-1 Adsorption.

7.1 Disclaimer

Part of the work reported in this Chapter was carried out by Dr. P. Palafox-Hernandez. Specifically: the REST simulation of QBP-1 adsorbed at the aqueous Au(111) interface and the regular MD simulations conducted for energetic analysis (Section 7.3.3). The former included the choice of initial peptide structures and number of replicas (and their effective temperatures) used, and construction of the peptide topology file, which were also all subsequently used in the solution and quartz REST simulations. The analysis of the data from all simulations, and its presentation in this Chapter, is my own work.

7.2 Introduction

In this Chapter the adsorption of quartz-binding peptide QBP-1 [Oren et al. [2007]] is reconsidered. This time, its adsorption to two different aqueous interfaces–(100) α-Quartz and Au(111)–was studied using REST [Moors et al. [2011]; Wang et al. [2011]; Terakawa et al. [2011]], a computationally efficient REMD-based method for studying the interfacial conformation of a peptide investigated in Chapter 5. The GolP-CHARMM FF for modelling protein-gold interactions [Wright et al. [2013b]], derived in Chapter 6, was employed.

Material-binding peptide sequences able to discriminate between two substrates of differing composition are highly desirable due to their potential to enable fine control over NP assembly (Figure 1.1a, b). In recent years fusion proteins consisting of gold- and silica-binding domains have been commonly used to demonstrate the viability of these synthetic strategies [Hnilova et al. [2012a,b,c]]. Interest in these two substrates in particular stems from the potential to combine the optical or catalytic properties of AuNPs with an abundant resource, silica, to manufacture functional materials economically (Section 1.1). In all of the case studies carried out, QBP-1 was used as the silica binding domain. The prevailing opinion in the literature, based on experiment, suggests that QBP-1 preferentially adsorbs to quartz over gold [Hnilova et al. [2012b,a,c]; Tamerler et al. [2010]]. However, when the evidence is critically examined, the results are more inconclusive (Section 7.5). It is hence interesting to use simulation to investigate whether QBP-1 is able to discriminate between the two materials–quartz and gold–on binding. In addition, the possibility suggested in Section 5.3.6 that the FF used in Chapter 5 may have accounted for the anomalous occurrence of helical structures of QBP-1 at the aqueous (100) α-Quartz interface is followed up in this work. In this Chapter, CHARMM22* [Piana et al. [2011]] was used, enabling comparison between CHARMM22* and CHARMM27 [MacKerell et al. [1998, 2004]] FFs for the study of small peptides in biointerfacial systems.

7.3 Methods

7.3.1 System Details

All simulations of Quartz Binding Peptide-1, QBP-1 (PPPWLPYMPPWS) [Oren et al. [2007]], performed in this Chapter were carried out using the software package Gromacs 4.5.5 [Hess et al. [2008]]. Three different systems were considered: the isolated peptide in solution, the peptide adsorbed at the aqueous (100) α-Quartz interface (Figure 1.2a) and the peptide adsorbed at the aqueous Au(111) interface (Figure 1.3a), denoted ‘solution’, ‘quartz’ and ‘gold’, respectively. The first comprised the peptide solvated by 6605 TIP3P water molecules in a simulation cell of dimensions $58.60 \times 60.90 \times 55.50\,\text{Å}^3$. A quartz slab of dimensions identical to that used in Chapter 5 was in this case solvated by 9164 TIP3P water molecules for the second, quartz, run. The simulation cell had dimensions $74.58 \times 65.64 \times 79.00\,\text{Å}^3$, slightly larger than those used previously (Section 5.2.1). The third system, ‘gold’, comprised a gold slab 5 layers thick featuring the Au(111) surface, the QBP-1 peptide and 6605 TIP3P water molecules; the simulation cell had dimensions $58.60 \times 60.90 \times 67.60\,\text{Å}^3$. The depth of water between the top surface of the slab and the bottom surface of its image was comparable in both sets of interfacial simulations, ‘quartz’ and ‘gold’, being $\sim$50 Å.

In contrast to previous solution [Notman et al. [2010]] and bio-interfacial quartz (Chapter 5, [Wright and Walsh [2013]]) simulations, the CHARMM22* [Piana et al. [2011]] FF was used to model the peptide in this Chapter. FFs compatible with CHARMM, namely LFF [Lopes et al. [2006]] and GolP-CHARMM (Chapter 6, [Wright et al. [2013b]]), were employed to describe peptide-substrate interactions at the aqueous α-Quartz and Au(111) interfaces, respectively. Water was modelled by the modified TIP3P [Jorgensen et al. [1983b]; Neria et al. [1996]] potential. As throughout this thesis, the rigid form of water was used, with bond lengths constrained by the SETTLE algorithm [Miyamoto and Kollman [1992]]. Simulations in all three environments were carried out in the NVT ensemble with the temperature maintained at 300 K using a Nosé-Hoover thermostat [Nosé [1984]; Hoover [1985]]. Cell dimensions in the z dimension were adjusted prior to simulation to ensure that the density of liquid water far from (both) the peptide (and substrate) was consistent with that of TIP3P water simulated under the same conditions in an NPT ensemble at a pressure of 1 bar. Newton's equations of motion were solved using the Leapfrog algorithm [Hockney et al. [1974]] with an integration time-step of 1 fs. Distances consistent with those used to derive LFF [Lopes et al. [2006]] and GolP-CHARMM (Section 6.2.1, [Wright et al. [2013b]] were chosen for truncation of PME electrostatic summation and LJ non-bonded interactions. Namely, in the solution and gold simulations PME summation was truncated at 11 Å, while for the quartz run it was truncated at 12 Å with a switching potential applied at 10 Å. A force-switched cut-off starting at 9 Å and ending at 10 Å was used for LJ non-bonded interactions in all three simulations.

7.3.2 Replica Exchange with Solute Tempering

To ensure adequate sampling of peptide conformational space, the REST (Section 2.3.3, [Moors et al. [2011]; Terakawa et al. [2011]; Wang et al. [2011]] method was employed to simulate QBP-1 in the three environments–solution, quartz and gold. In this Chapter a total of 16 replicas, spanning an effective temperature window of 300-433 K was used (300.00, 305.35, 310.89, 317.25, 323.88, 331.57, 339.88, 349.11, 358.07, 367.37, 380.98, 389.80, 398.22, 406.85, 419.97, 433.00 K). The replica number was increased from 11 to 16 in this work, compared to that carried out in Chapter 5, to enable a better fit to the architecture of the computational resource available. By using more replicas, the effective temperature range could, in turn, be increased.

Following a similar procedure to Chapter 5, the initial conformation of the peptide present in each replica featured a different common folded backbone secondary structure configuration (*e.g.* α-helix, β-sheet, β-turn *etc*) constructed by hand. REST simulations of QBP-1 in the three environments all featured the same set of peptide starting conformations (Section 5.2.2). It is worth commenting here specifically on the inclusion of helices in these starting conformations. Although the previous study suggested that conformational sampling of QBP-1 could be maximised by the exclusion of helical structures from initial replica configurations without biasing the results (Chapter 5) [Wright and Walsh [2013]], these conclusions were reached from energetic analysis of the peptide adsorbed at the aqueous (100) α-Quartz interface (Section 5.3.6). The QBP-1 helix may be a favourable conformation for the peptide to adopt in solution or when adsorbed onto gold. For the reasons highlighted in Section 5.2.3–namely the high energy barrier to helix formation due to the stiff nature of the peptide–therefore, it was necessary to include the QBP-1 helix initially in the REST simulations performed in this Chapter.

Before initiating REST, starting configurations were equilibrated for 1 ns at their target potentials with no exchange moves attempted. Simulations were run for 15×10^6 MD steps, with an exchange attempted between neighbouring replicas every 1000 MD steps. In all three runs the average probability with which an exchange was accepted was $\sim$70%; in each case, all replicas were mobile throughout the whole of effective temperature space (Figure E.1).

7.3.3 Analysis

For each of the three environments–solution, quartz and gold–analysis was carried out on the constant-ensemble replica run at the reference effective temperature of 300 K. Two fundamental aspects of QBP-1 adsorption were investigated: the structure of the peptide and the manner in which substrate binding was mediated. Unless stated otherwise, only the final 5 ns of each REST trajectory was analysed. In these later stages, each simulation was observed to have reached a steady state (Figure E.2).

Unlike the closest atom–surface separation criterion used in Chapter 5 (Section 5.2.4) to identify adsorbed residues and motifs (dyads and triads), here the distribution of residue functional group-surface distances generated over the entire REST trajectory (Figures E.3 and E.4) was used to define residue specific ‘bound distance’ cut-offs for each interface (Table E.1). A dyad/triad was classed as being bound to the surface if the two/three sequential residues from which it was composed were simultaneously adsorbed.

7.4 Results

7.4.1 Peptide Conformation

The study of a peptide's conformation can be subdivided into two fundamental aspects: characterisation of the dominant structures it adopts; and analysis of the regions of conformational space sampled by the entire ensemble of its structures. In the following, cluster analysis was used to probe the first phenomenon (Section 2.5.2), while secondary structure analysis based upon Ramachandran plots was employed to investigate the latter entity (Section 2.6).

Clustering of QBP-1 structures sampled in the final stages of each trajectory (Table 7.1) suggests that the conformational freedom of the peptide was significantly restricted on adsorption to either interface, quartz or gold, relative to solution, in agreement with the findings reported in Section 5.3.2. There was little difference between the two substrates, however, with a total of 26 clusters being identified in both cases. This result was invariant to the exclusion of infrequently sampled clusters–thought to represent high energy peptide structures–with the number of 'significant' clusters of adsorbed peptide structures being almost the same (9 and 11, at the gold and quartz interfaces, respectively).

	total	significant	1/%	2/%	3/%
Solution	69	31	21.6	11.7	8.9
Quartz	26	11	39.4	16.4	10.4
Au(111)	26	9	34.4	16.1	12.2

Table 7.1: Total number of, and the number of 'significant', backbone clusters calculated over the final 5 ns. Percentage population of top 3 clusters identified by clustering analysis conducted over the same portion of the trajectory, is also given.

In all three environments, the most populated clusters of QBP-1 represented extended–either straight PPII or extended with a kink in the central region–structures (Figure 7.1). Calculation of atom backbone position RMSDs between pairs of centroid cluster structures identified in different simulations revealed that there were significant similarities between the most likely conformations of the peptide, both between in solution and at either interface and between adsorption to the two different substrates. The extended nature of these structures is broadly consistent with both experimental CD measurements of the peptide in solution [Oren et al. [2010]; Notman et al. [2010]] and previous simulation studies (Chapter 5: Figure 5.2; [Notman et al. [2010]; Wright and Walsh [2013]]) of the peptide (in solution or adsorbed at the aqueous (100) α-Quartz interface).

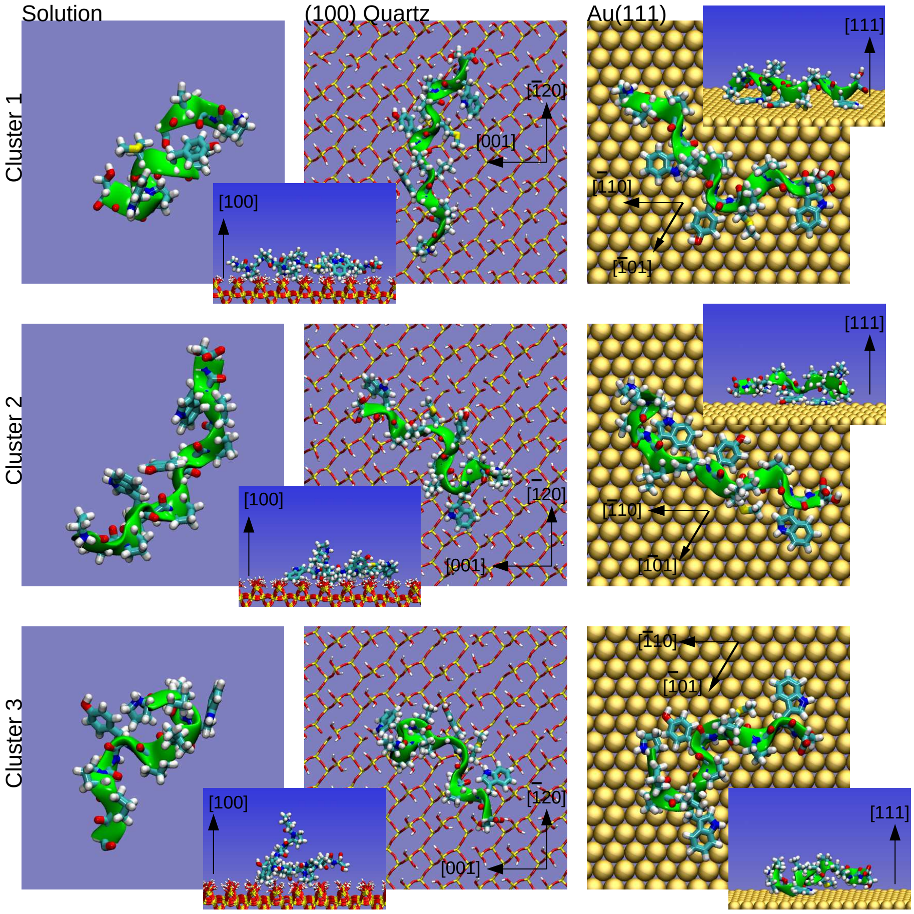

Solution
(100) Quartz
Au(111)
Cluster 1
Cluster 2
Cluster 3
[100]
[1̄20]
[001]
[111]
[1̄10]
[1̄01]

Figure 7.1: Snapshots of the centroid structure belonging to the three most populated clusters of QBP-1 in solution and adsorbed at either the aqueous α-Quartz or Au(111) interfaces. Water has been omitted for clarity.

There was one notable exception: solution cluster 1 (the cluster found to have the largest population after simulation of QBP-1 in solution), which was a helix. Helical clusters of QBP-1 were also identified at the aqueous α-Quartz and Au(111) interfaces, but had much smaller relative populations (3.1% and 10.4%, respectively, compared to 21.6% of the total population of structures in solution). While helices featured as prominent structures in a subset of the REST simulations of QBP-1 adsorbed onto the α-Quartz surface performed in Chapter 5 (Section 5.3.3), potential energy calculations revealed that they were in fact unfavourable with respect to extended conformations of the peptide on adsorption (Section 5.3.6) [Wright and Walsh [2013]]. In addition, CD spectra of QBP-1 in solution did not display any significant α signature, even in the presence of the secondary structure stabilising agent 2,2,2-trifluoroethanol [Oren et al. [2010]]. That, even using the newer CHARMM22* FF [Piana et al. [2011]]–which features modified torsional parameters designed to correct for the reported helical bias of CHARMM27 [MacKerell et al. [1998, 2004]]–for the present work, helical structures still featured, albeit with a reduced frequency, warrants further comment.

Firstly, the solution CD spectrum of QBP-1 may, in fact, not be inconsistent with the results of the solution-based REST simulation. CD is an ensemble and sequence averaged measurement of secondary structure (Section 1.2.3). Only the central region of QBP-1 in solution cluster 1 was helical–the conformation of both termini deviated substantially from a helix. In addition, even with a cluster population of ~20% the helix was a minority structure. Taken together, these two facts mean that the α-helix signature of QBP-1 may have been overwhelmed in the CD spectrum. It must also be commented that the deconvolution of CD spectra is done with reference to databases containing proteins of known structure. The suitability of these references for identifying the secondary structural propensity of an IDP–a lable class of proteins with no fixed secondary structure to which material binding peptides are thought to belong [Evans et al. [2008]]–is unknown.

Secondly, energetic analysis, carried out in an identical manner to that performed in Chapter 5, was done to determine the potential energy of each system featuring QBP-1 in the centroid structure of the most highly populated cluster (PE_{CX}), relative to that featuring QBP-1 in the centroid structure of the helical cluster (PE_{helix}):

$$\Delta PE = PE_{CX} - PE_{helix}. \tag{7.1}$$

(In the case of solution, the potential energy differences of clusters 2 and 3 relative to that of the helical cluster (solution cluster 1) were calculated.) Regular

MD simulations were performed using both CHARMM22* [Piana et al. [2011]] and CHARMM27 [MacKerell et al. [1998, 2004]].

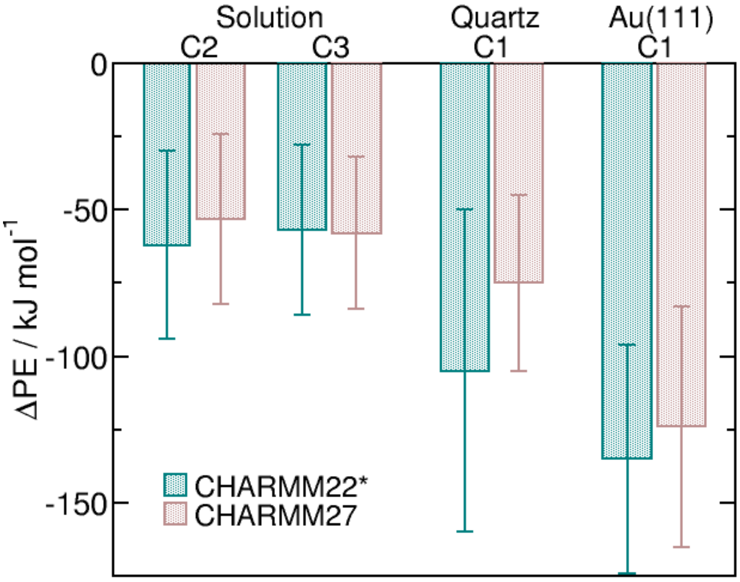


Figure 7.2: Potential energy difference between centroid structure of cluster CX and that of the cluster representing QBP-1 helical conformations.

The results of the energetic calculations are presented in Figure 7.2. In each environment, the QBP-1 helix structure was unfavourable with respect to more extended conformations of the peptide, consistent with the findings of Section 5.3.6. However, the energetic discrimination between helical and more extended structures was very small for the peptide in solution. Thus, the relatively large population of the solution helical cluster is not inconsistent with the REST simulation having reached a steady state. (At equilibrium, based on the Boltzmann distribution, the population of a cluster should be exponentially related to its free energy. The potential energy difference calculated in Equation 7.1 is only a reasonable approximation of the free energy difference between two states if the entropic contribution to the free energy is small.) In addition, these results suggest that for QBP-1, at least, there is little difference between CHARMM22* [Piana et al. [2011]] and CHARMM27 [MacKerell et al. [1998, 2004]] in terms of helical propensity. This is perhaps unsurprising considering the primary sequence of the peptide. Proline–which accounts for 50% of QBP-1–is one of only two amino-acids in CHARMM22* (Gly and Pro) for which the torsional parameters are identical to those of CHARMM27 [Piana et al. [2011]].

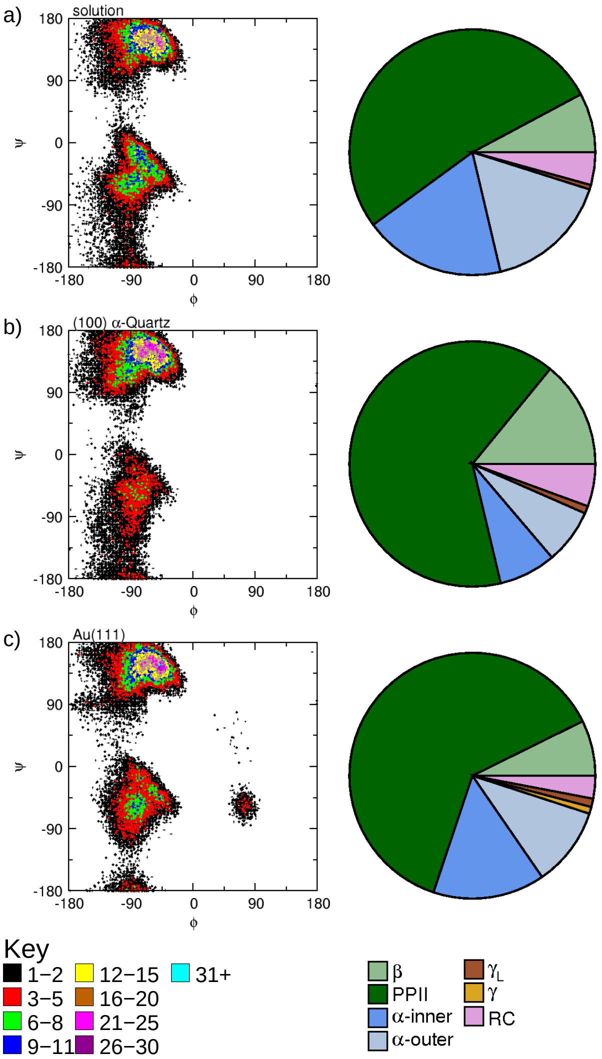


Figure 7.3: Left: Ramachadran plots of QBP-1 a) in solution and adsorbed at the aqueous, b) α-Quartz, or c) Au(111) interfaces. Right: Pie charts depicting the frequency with which each region of ϕ/ψ space was sampled.

In addition to the structural similarities between the most likely conformations of QBP-1 adsorbed onto quartz and gold, the REST simulations suggest that there is little difference in the ensemble of all structures featured at these interfaces (Figure 7.3). In solution, as well as when adsorbed, QBP-1 mainly has PPII character; structures with α, β and random coil character are also featured, to a lesser extent, by the ensemble of all peptide conformations. Consistent with the presence of a helical cluster of QBP-1 in each simulation, the three Ramachandran plots all feature an elliptical footprint in the $-180<\phi<0$, $-180<\psi<0$ quadrant–the signature associated with an α helix. Only minor variation, in regions of ϕ/ψ phase-space infrequently sampled, was observed between the different systems. The most notable of these was the existence of structures with γ character at the aqueous Au(111) interface only.

However, by making comparison with Figure 5.3, the regions of conformational space sampled by an ensemble of QBP-1 structures can be seen to depend more on the *FF* used than the *substrate* onto which the peptide is adsorbed (Figure E.5). In particular, using CHARMM27 [MacKerell et al. [1998, 2004]] the α/random coil character of QBP-1 upon adsorption to α-Quartz was more/less pronounced than that predicted by CHARMM22*. Additionally, QBP-1 was found to sample the α_L region of ϕ/ψ phase-space at this interface only whilst modelled using CHARMM27; a secondary structural signature of the peptide which was consistently identified in all of the REMD-based simulations carried out in Chapter 5, independent of the initial replica population (Figure 5.3) [Wright and Walsh [2013]]. Never-the-less, RMSD analysis revealed structural similarities between the two FFs when the backbone conformation of the centroid structures representing dominant clusters of QBP-1 adsorbed at the α-Quartz interface was considered. Hence the evidence suggests that while the alterations made to CHARMM can bring about small changes in the ensemble of all structures of small peptide sequences such as QBP-1, they do not significantly impact on the configuration of its most likely structures.

Overall, the structural analysis carried out here suggests that QBP-1 does not undergo significant conformational change on adsorption to either the α-Quartz or Au(111) surface. Moreover, the findings imply that surface induced peptide folding is unlikely to be the source of the material selectivity of QBP-1, should it exist.

7.4.2 Surface Binding

While the structure of the QBP-1 peptide itself, in its adsorbed state, may not significantly differ between the quartz and gold surfaces, the orientation of the peptide with respect to the interface could contribute to its reported selectivity. For instance, one possible explanation for weak binding to gold would be if QBP-1, in an extended conformation, was bound to the surface with its axis perpendicular to the interface; in such bound peptide geometries, surface contact would be minimal. To probe this phenomenon the degree of surface contact found in the later stages of each interfacial simulation has been quantified using data from the entire ensemble of peptide structures. Both the frequency with which individual residues (Figure 7.4, Tables E.2) and motifs of consecutive residues (Tables E.3 and E.4) were in contact with the surface was investigated.

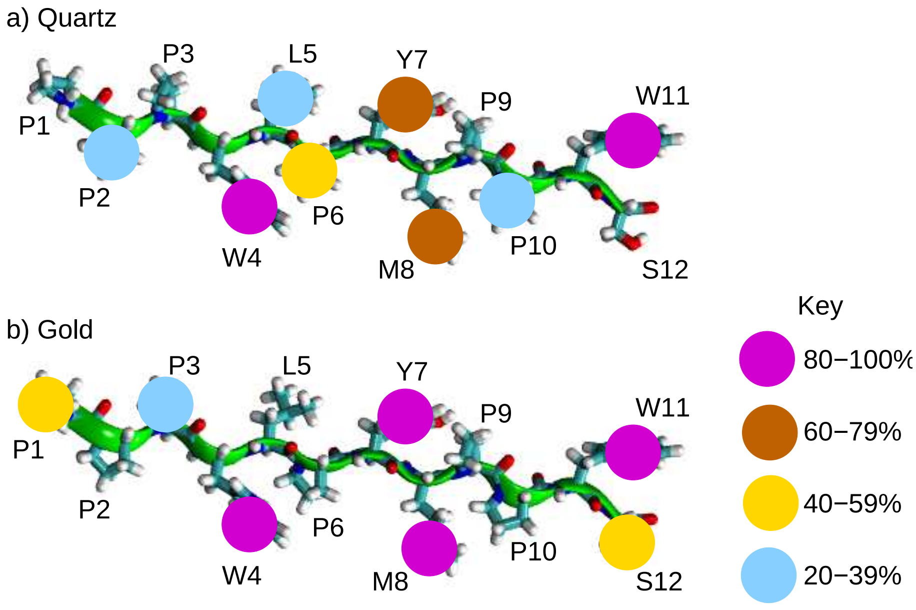


Figure 7.4: Schematic depicting the degree to which each residue contributes to surface binding at the aqueous a) α-Quartz and b) Au(111) interfaces. The key corresponds to the percentage of frames in which the residue is bound to the surface (Table E.2).

Consistent with trends in the gold binding propensities of individual amino-acids reported in the literature [Cohavi et al. [2011]; Hnilova et al. [2008]; Peelle et al. [2005]; Hoefling et al. [2010a]; Feng et al. [2011]; Wright et al. [2013b]], aromatic residues W4, Y7 and W11, along with the sulfur containing residue M8 were all bound to the Au(111) surface in the majority of frames. The same four residues were also found to be important in mediating QBP-1–quartz adsorption. Making direct comparison for each of these strong binding residues individually between the two materials, the likelihood of them being adsorbed to the gold surface was, in fact, slightly *greater* than that to quartz. However, although the terminal residues, P1 and S12, also contributed to gold-binding, albeit to a lesser extent, the majority of the sequence remained unbound. Conversely, at the aqueous α-Quartz interface the binding propensity of each residue was more consistent, especially in the central segment of the peptide, W4 to W11. This was particularly apparent when the adsorption of consecutive residues–namely, dyad and triad motifs–was considered (Tables E.3 and E.4). Namely, at the Au(111) interface, none of the 10 possible triads made significant surface contact.

Visual inspection of bound states, taken together with data analysis for the adsorption of residue functional groups, suggests that differences between gold and quartz in terms of the mode of QBP-1 adsorption exist. In its adsorbed state on the Au(111) surface, the QBP-1 peptide can be thought of as being bound to the interface through a series of 'stilts'; at the aqueous α-Quartz interface, however, QBP-1 is more likely to lie flat against the surface, making more continuous contact. The frequency with which the backbone, rather than side-chain, atoms of each residue make surface contact with the two substrates is in agreement with this hypothesis (Table E.5 and Figure E.6). Moreover, aromatic residues Trp and Tyr in QBP-1 can be seen further to discriminate between the two materials in their mode of side-chain binding. In the case of adsorption to gold, only one dominant adsorbed residue–surface geometry existed–that in which the phenyl ring was oriented parallel to the plane of the interface (7.5b). The mode of adsorption of Trp and Tyr to the quartz surface was slightly more ambivalent, however. In their most likely bound state at this interface, in contrast to gold, the aromatic ring was upright or tilted with respect to the surface (7.5a) to facilitate penetration into an interstitial site. Such bound geometries for phenyl groups at the aqueous α-Quartz interface have been reported before (Sections 3.3.4 and 5.3.5).

7.5 Discussion

There is no strong evidence from the REST simulations carried out in this Chapter qualitatively to suggest that QBP-1 preferentially adsorbs to quartz over gold. At both interfaces, the number of contact residues and structure of QBP-1 are similar. Only the mode of binding differs slightly, with more continuous contact along the whole length of the sequence and closer backbone-substrate binding being found upon adsorption to quartz than gold (Section 7.4.2). Without further simulations to calculate the free energy of adsorption, it is not possible to determine whether this slightly more concerted adsorption mode of binding to quartz is sufficient to mediate material selectivity or not. However, when the experimental data in the

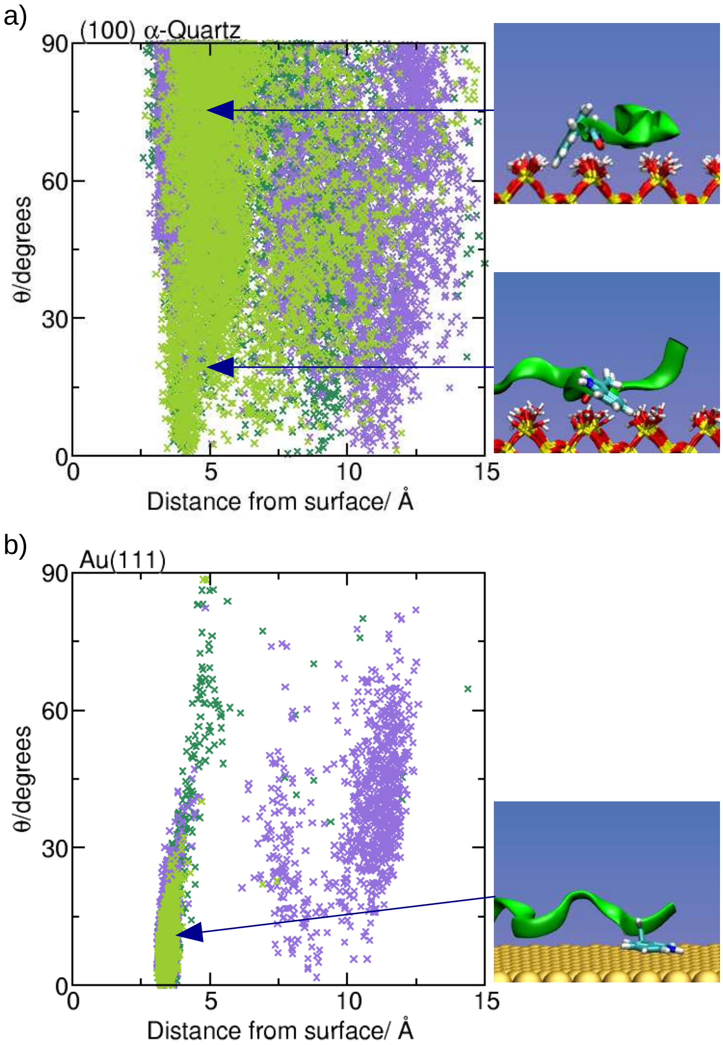


Figure 7.5: The orientation of aromatic residues with respect to a) the α-Quartz and b) the Au(111) surface. θ was defined as the angle between the normal to the aromatic plane of each residue and the z direction. Data from the tryptophan residues W4 (dark) and W11 (light) were plotted in green, whereas that from Y7 was plotted in purple. Right: Snapshots depicting a trp residue adsorbed onto the α-Quartz surface in an upright/flat conformation (top/middle) and onto the Au(111) surface in a flat conformation (bottom).

literature [Hnilova et al. [2012b,a,c]], suggestive of material selectivity, is analysed carefully, the results are open to interpretation.

In what follows, each of the experimental studies is reviewed. First, before turning to more circumstantial evidence, it is worth considering the energetic data. Hnilova *et al.* fitted SPR shifts to a Langmuir adsorption isotherm in order to calculate the free energy of adsorption of QBP-1 to silica, reporting a value of $-8\,\mathrm{kcal\,mol^{-1}}$ [Hnilova et al. [2012c]]. The raw SPR data from the equivalent experiment at the aqueous gold interface is also reported in the paper and is suggestive of weaker binding. No binding energy was calculated though from this data. More recently, Knecht and *co workers* used QCM to determine the binding affinity of QBP-1 to polycrystalline gold surfaces, reporting a value of $-8.4 \pm 0.3\,\mathrm{kcal\,mol^{-1}}$ [Tang et al.. [2013]]. The conditions (*e.g.* presence or absence of buffer, concentration of adsorbate) under which the two studies–SPR and QCM–were carried out differed, meaning that it is not possible to draw strong conclusions. However, at first glance, taken together these results indicate that in fact QBP-1 cannot energetically discriminate between the two substrates on binding; evidence supported by the simulations performed here.

Over the last couple of years the material selectivity of QBP-1 has been inferred a number of times from fusion peptide experiments. In two cases the peptide–QBP-1$\cdots$AuBP-1 [Hnilova et al. [2012b,a]], AuBP-1$\cdots$QBP-1 [Hnilova et al. [2012b]]–was stamped onto a glass surface. Assuming AuBP-1 has little affinity for quartz, then the fusion peptide is likely to adhere to the slide via its QBP-1 domain. Subsequent exposure to gold, in the form of AuNP, may have been too short (15-30 min) for significant QBP-1–quartz desorption and QBP-1–gold adsorption to be observed, especially as the evidence in this Chapter together with the two experimental binding energies suggest that the thermodynamic driving force for the process would be slight at most. Overall, however, the results from these two fusion peptide experiments [Hnilova et al. [2012b,a]] are not inconsistent with the findings here.

Only the study carried out by Hnilova *et al.* published in *Soft Matter* in 2012 is, perhaps, at odds with the simulations performed in this Chapter [Hnilova et al. [2012c]]. In this case the QBP-1, labelled with the biotinylating agent NSH-biotin, was exposed to a silica substrate patterned with gold regions. Preferential adsorption of QBP-1 to silica over gold was inferred by visualisation with streptavidin quantum dots and high resolution AFM. The whole procedure was carried out in the presence of PBS buffer, however. The presence of buffers in material-binding peptide assays has previously been noted to moderate the results (Section 1.3.1 [Brunner et al. [2004]; Lutz et al. [2005]; Nergiz et al. [2013]]). Not only this, but conjugation to the biotintylating agent may have brought about conformational changes in either QBP-1 itself or its aggregates in solution which favoured silica adsorption over gold. It is difficult to conclude, therefore, based on the results of this one study, that QBP-1 is able to discriminate between quartz and gold on binding. Further experiments, to determine the free energy of adsorption of QBP-1 to quartz and gold under identical conditions, are required.

7.6 Conclusions

In conclusion, the REST simulations reported in this Chapter suggest that while the conformation of the QBP-1 peptide itself is almost invariant to its environment–in solution or adsorbed at the aqueous (100) α-Quartz or Au(111) interfaces–its mode of adsorption to the two materials, quartz and gold, differs slightly. A significant proportion of the peptide was found to make moderate contact with the quartz surface, while it was adsorbed to gold via the side-chains only of a number of strong binding residues. In both cases, extended backbone structures were the most likely. In terms of overall binding residue contact and conformation, however, there was little evidence to support the experimental hypothesis of preferential QBP-1–quartz adsorption. Free energy calculations, like those carried out in Chapter 8, together with experimental measurements carried out in a consistent manner at both interfaces are required to further probe this phenomenon.

In addition, simulations at the aqueous α-Quartz interface performed in this work, when compared against those in Chapter 5, reveal that the differences between the CHARMM27 and CHARMM22* FFs are subtle, at least for the QBP-1 peptide.

Chapter 8

Facet Selectivity of AuBP-1 adsorption onto Gold: A Replica Exchange with Solute Tempering Metadynamics Study.

8.1 Disclaimer

Part of the work reported in this Chapter was carried out by Dr. P. Palafox-Hernandez. Specifically many of the details associated with setting up the ‘REST’ component of the simulations (number of replicas, effective temperature of each replica, initial replica peptide conformations and scaled peptide topology file) were taken from a parallel study reported in reference Tang et al.. [2013] which included the Au(111) REST MD run mentioned herein. The Au(111) REST metaD simulation was also carried out by Dr. P. Palafox-Hernandez; the latter only was under my direction–I supplied initial co-ordinate files and the modified code.

The analysis of the data from all simulations, and its presentation in this Chapter, is my own work.

8.2 Introduction

In this Chapter the FFs derived in Chapter 6 for modelling the adsorption of proteins to the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces under aqueous conditions are used in combination with the REST method [Moors et al. [2011]; Terakawa et al. [2011]; Wang et al. [2011]]–benchmarked in Chapter 5 for biointerfacial systems–to investigate whether adsorption of the AuBP-1 [Hnilova et al. [2008]] peptide could be used to differentiate between these crystallographic planes of gold. Peptide sequences able to discriminate between different gold surfaces on adsorption are highly desirable due to their potential for mediating the growth of AuNP with well-defined tunable morphologies (Section 1.1). Whilst others in the past have used simulation to investigate the facet selective adsorption of peptides onto gold–specifically between Au(111) and Au(100)–the reconstructed status of the latter surface, amongst other things (Section 1.4.2), was neglected. As well as probing the role that Au(100) surface reconstruction may have on peptide adsorption, the work in this Chapter aims to build on the qualitative findings from amino-acid MD simulations at the three aqueous interfaces reported briefly in Section 6.4 and documented in full in references Wright et al. [2013b,c].

As mentioned in Section 5.5, the affinity of AuBP-1 for gold is qualitatively stronger than that of QBP-1 for quartz. Preliminary investigations of the former system (AuBP-1–gold) revealed that, during REST MD simulations, the peptide remained tightly bound to the interface in the reference replica. The large potential energy difference between adsorbed and desorbed states led to segmentation of replicas in effective temperature space, with minimal exchange between the two regimes (Figure F.1a). The ability of REST to enhance the conformational sampling of AuBP-1 at the aqueous gold interface was thus dimished. In this Chapter, metaD (Section 2.4.2)–where a small bias was added to the peptide COM–surface separation–was used in combination with REST to ensure that AuBP-1 was both bound and unbound to the surface in each replica as the simulation progressed. The metaD method was primarily used here to enhance conformational sampling within the scope of REST.

The work in this Chapter is divided into two main segments. First, schemes for recovering the unbiased distribution of AuBP-1 characteristics at each interface from the metaD simulations are discussed (Section 8.4); before, second, the physical insights which can be inferred from this study are presented (Section 8.5).

8.3 Methods

8.3.1 System Details

Both Molecular Dynamics (MD) and metadynamics (metaD) simulations of AuBP-1 (WAGAKRLVLRRE) [Hnilova et al. [2008]], performed here were carried out using the software packages PLUMED 1.3 [Bonomi et al. [2009a]] and GROMACS 4.5.5 [Hess et al. [2008]]. The appropriate exchange criterion (Equation 2.30) was added to the PLUMED code in order to implement Hamiltonian replica exchange metaD–a procedure not present in standard packages at the time when the work was initiated. Four different systems were considered: the isolated peptide in solution and, the peptide adsorbed at the aqueous Au(111), Au(100)(1×1) and Au(100)(5×1) interfaces. The first comprised the AuBP-1 peptide solvated by 6605 TIP3P water molecules in a cubic simulation cell of length 58.28Å. Orthorhombic cells of dimensions $58.60 \times 60.90 \times 67.60\,\text{Å}^3$, $58.60 \times 58.60 \times 67.60\,\text{Å}^3$ and $58.60 \times 58.60 \times 76.51\,\text{Å}^3$ were used for the interfacial simulations carried out at the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces, respectively. These were comprised of the AuBP-1 peptide, a gold slab (5/5/9 layers thick) and 6605/6540/6355 TIP3P water molecules. The protonation state of the peptide at pH 7 was modelled, with 3 Cl^- counterions added to balance the charge. The depth of water between the top surface of the slab and the bottom surface of its image was comparable in all three sets of interfacial simulations, being $\sim$60 Å.

In all simulations, CHARMM22* [Piana et al. [2011]] was chosen to model the peptide, while the modified TIP3P [Jorgensen et al. [1983b]; Neria et al. [1996]] potential, with which the bio-organic CHARMM FF [MacKerell et al. [1998]] has been harmonised, was used to represent water. Bond lengths within water were constrained by the SETTLE algorithm [Miyamoto and Kollman [1992]]. Peptide-gold interactions at each of the different aqueous gold interfaces were described by GolP-CHARMM [Wright et al. [2013b,c]]. In this work the interfacial simulation of the Au(100)(5×1) surface used a gold slab with a lateral footprint of dimension 4×5, and approximately double thickness, compared to the basic structural unit of the surface in the FF [Wright et al. [2013c]]. Simulations were carried out in the NVT ensemble with the temperature maintained at 300 K using a Nosé-Hoover thermostat [Nosé [1984]; Hoover [1985]]. In all cases, the cell dimension in the z dimension was adjusted prior to simulation to ensure that the density of liquid water far from both the substrate and peptide was consistent with that for TIP3P water simulated under the same conditions in an NPT ensemble at a pressure of 1 bar. Newton's equations of motion were solved using the Leapfrog algorithm [Hockney et al. [1974]] with an integration time-step of 1 fs; co-ordinates were saved every 1 ps. PME electrostatic summation was truncated at 11 Å, while a force-switched cut-off starting at 9 Å and ending at 10 Å was used for LJ non-bonded interactions.

8.3.2 Replica Exchange with Solute Tempering

REST simulations were set up in a manner identical to that employed in Chapter 7. Briefly, a total of 16 replicas were used to span an effective temperature range of 300-433 K (300.00, 305.35, 310.89, 317.25, 323.88, 331.57, 339.88, 349.11, 358.07, 367.37, 380.98, 389.80, 398.22, 406.85, 419.97, 433.00 K). Each replica was initially populated with AuBP-1 present in a different conformation; peptide structures were constructed by hand to feature common folded backbone secondary structural motifs. All REST simulations featured the same set of internal-peptide and peptide-gold (orientation with respect to and distance from the gold surface) starting conformations. The peptide was initially close to the top surface of the gold slab in 13 out of the 16 replicas, in the centre of the cell in 2 and close to the nearest periodic image of the bottom face in one. Before initiating REST, starting configurations were equilibrated by a short MD simulation, of duration 1 ns, at their target potentials. No exchange moves were attempted and no bias was added to the collective variable during this time. In the subsequent production simulations, exchanges between neighbouring replicas were attempted every 1 ps.

The results of three REST metaD and two REST MD simulations are reported in this Chapter. The CV chosen for the interfacial metaD runs was the position of the *c.o.m.* of the peptide in the z dimension (the direction normal to the gold surface). Small Gaussians of width, σ=0.1 Å, and height, w=0.1 kJ mol^{-1}, were added every 0.5 ps along the direction of the CV. MetaD simulations were run for 100 ns, making a total of 1.6 μs of dynamics per interface. All replicas were mobile throughout the whole of effective temperature space (See Figure F.1b for trajectories along the effective temperature ladder of representative replicas in each simulation). The two REST MD simulations–one of AuBP-1 in solution and the other of the peptide at the aqueous Au(111) interface–were carried out for comparative purposes and were run for 15 ns only.

8.3.3 Analysis

Like Chapter 7, analysis of the simulations reported herein was oriented to help answer two key questions: 1) how does the structure of AuBP-1 differ when adsorbed to one crystallographic plane of gold compared to another and 2) is there evidence for energetic and/or spatial selectivity in adsorption between the aqueous Au(111) and Au(100) interfaces. Unless otherwise stated, the entire 100 ns REST metaD trajectory of the reference replica (the replica in which the effective and actual simulation temperatures are identical) was analysed in all three cases–Au(111), Au(100)(1×1) and Au(100)(5×1). It is recognised that both adsorbed and desorbed states of AuBP-1 are sampled in these metaD simulations. The primary aim of this work was to investigate the former (surface-bound conformations) only. This can be achieved by reweighting the bias added to the CV connecting the two states in a suitable way (Section 8.4). The advantage of this method is that it circumvents the need for an arbitrarily chosen cut-off to be used to mark the transition from the peptide being adsorbed, to free in solution. Analysis of the two REST MD simulations, where presented, was performed on the final 5 ns of each run only.

As before, a combination of secondary structural analysis based on ϕ/ψ phase-space depicted in a Ramachandran plot (Section 2.6) and structure clustering was used to probe the potential for AuBP-1 to fold in a different manner on adsorption to Au(111) and Au(100). In the latter case–structure clustering–however, it was not appropriate here to use the Daura cluster algorithm [Daura et al. [1999]] to analyse data from the metaD simulations. Each structure within the pool of structures identified over the entire trajectory (Step 1, Section 2.5.2) has a different weight, determined by the metaD bias added to the simulation at the time at which the structure was sampled. These different weights must be accounted for, both when identifying reference structures (cluster centroids, in the case of the Daura algorithm) and when assigning structures to clusters.

An alternative strategy for grouping together the similar conformations of AuBP-1 sampled during a simulation is to use a pre-defined set of reference structures. In this study 9 distinct conformations of AuBP-1 were identified: 8 from the REST MD runs and 1 unique structure–the latter found only to feature significantly after the more extensive metaD simulations. Specifically, the Daura algorithm was used to cluster structures sampled in the final stages of the REST MD simulations of AuBP-1 in solution and at the Au(111) interface. As in Chapters 5 and 7, data from the unbiased reference replica only was used, meaning that the assumption that all structures had a uniform weight was justified. A RMSD peptide backbone position cut-off of 3 Å was employed. Across the two simulations, the centroid structures of 8 clusters–S2, S3, S4, S6 (A5), A1 (S1), A2 (S5), A3 and A4 (where Ax and Sx are the xth most populated clusters of the Au(111) and solution runs respectively)–were distinct (by which I mean that the RMSD in peptide backbone atoms between any pair was greater than the cut-off used in the clustering). Hereafter, these structures are labelled a-h (Figure 8.1). Broadly, solution derived conformations a-d are more globular than the extended structures e-h found to dominate at the Au(111) interface in the REST MD simulation. The ninth structure, i, is also extended in nature.

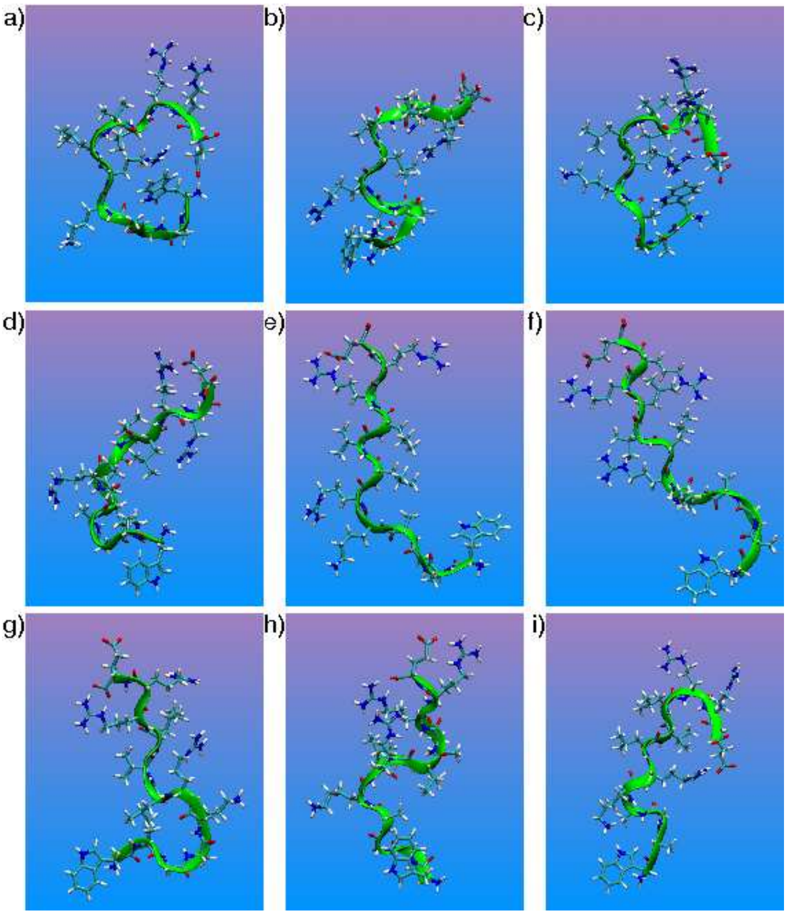


Figure 8.1: Reference structures a-i used to cluster conformations of AuBP-1 at the aqueous Au(111), Au(100)(1×1) and Au(100)(5×1) interfaces, during the REST metaD simulations.

Akin to Chapter 7, the residues and motifs (dyads and triads) which were involved in binding AuBP-1 to an interface were identified using a set of criteria based on side-chain functional group surface separations. However, since the peptide COM–gold distance was the CV biased in the metaD simulations it was not appropriate to use the trajectory-averaged distribution of side-chain–surface separations, as before, to determine cut-off distances for adsorbed states. Instead, the maximum distance for which a residue could physically be interpreted as being in direct contact with the gold atoms in the upper-most layer of each surface has been used. This distance was defined as the Au-X (X=side-chain functional group heavy atom as stated in Table F.1) van der Waals equilibrium separation plus 1 Å. The binding propensity of a residue was classed as one of three things based on its percentage likelihood of being adsorbed to an interface: 26-50% 'moderate', 51-75% 'significant' and 76-100% 'strong'.

8.4 Schemes for Reweighting Structural Data from a Metadynamics Simulation

While in recent years metaD has become a more frequently employed technique within the realm of molecular simulation, its use thus far reported in the literature has mainly been focussed on either accelerating rare event sampling (for example simulating the crystallisation of ice [Quigley and Rodger [2008]]) or for calculating free energy landscapes [Jambeck and Lyubartsev [2013]; Schneider and Ciacchi [2012]]. Analysis of the structural data generated by such simulations has largely been neglected due to the difficulties in relating the biased metaD trajectory to the properties of the system in the ensemble of interest. In this Section 6 different reweighting schemes are described and subsequently evaluated. Schemes 1-4, derived here, are based on simple manipulation of the metaD bias added to the CV, s, at the time in the simulation when a frame is sampled, $\mathrm{V}(s(X(t)), t)$ (Equation 2.28, $X(t)$ are the co-ordinates of the system at time t). The reweighting factor identified by these schemes, $\mathrm{W}(X(t), t)$, is used directly in Equation 2.31 to recover the unbiased probability distribution of an observable. Strategies 5 and 6, are methods published by Tiana and Bonomi *et al.*, respectively, for reweighting normal [Tiana [2008]] and well-tempered [Bonomi et al. [2009b]] metaD simulations.

8.4.1 Scheme 1: Total bias 'V'

In this, the simplest case, the total bias added to the CV, up until the time when a structure is sampled, is used for reweighting. *i.e.*

$$W(X(t),t) = V(s(X(t)),t) \tag{8.1}$$

where $\mathrm{V}(s(X(t)),t)$ is defined in Equation 2.28.

8.4.2 Scheme 2: Relative Bias 'ΔV'

One of the problems of reweighting data from a metaD simulation using Scheme 1 is that the bias added, $\mathrm{V}(s(X(t)),t)$, increases continuously with time even after free energy differences between important states have converged. Thus, the weight given to two identical structures, sampled at different stages during a trajectory, is not the same–the second occurrence is given more importance than the first.

In theory, after convergence, the shape of a free energy profile is independent of time. The weight of a structure can therefore be related to its corresponding relative bias, $\Delta\mathrm{V}(s(X(t)),t)$, defined as:

$$\Delta V(s(X(t)),t) = V(s(X(t)),t) - V(s(X_0(t)),t) \tag{8.2}$$

where $\mathrm{V}(s(X_0(t)),t)$ is the bias added to a reference region (corresponding to system co-ordinates X_0) along the CV. For the biointerfacial systems considered here, $\mathrm{V}(s(X_0(t)),t)$ was chosen to be the average bias added to the region of CV-space (peptide COM–gold separation) corresponding to the 5 Å either side of the midpoint between the gold slab and the periodic image of it nearest neighbour. Since the reference bias changes with time and must be defined in some discrete manner, there is a possibility that small statistical errors will be introduced into the weight given to a structure determined using this strategy.

8.4.3 Scheme 3: Relative Bias 'ΔV equilib'

For Scheme 2, data from the entire metaD trajectory was analysed using the reweighting factor $\mathrm{W}(X(t),t) = \Delta\mathrm{V}(s(X(t)),t)$. As stated above, the physical basis for using ΔV, rather than V, to reweight structures relies on convergence in the free energy profile of the system of interest. There is evidence to suggest that this criterion may not strictly have been met in any of the runs performed here, even within the 100 ns simulation time-frame probed (Section 8.5.1). Before convergence, and especially in the initial stages of a simulation, the bias $\mathrm{V}(s(X(t)),t)$ does not grow uniformly along the CV. The peptide was initially placed adjacent to the upper facet of the gold slab in the majority of replicas. Thus, the simulations sample the region of CV space corresponding to the free energy minimum close to this surface extensively at the start of each trajectory; the bias added to the reference region V$(s(X_0(t)), t)$ grows at a much slower rate during this period. As a consequence, structures sampled at the start of a simulation will be given more weight in the final results than those visited later–the opposite bias to that inherent in Scheme 1. In an attempt to overcome this shortcoming of Scheme 2, a third Scheme, 'ΔV equilib', was tested. Schemes 2 and 3 differ only in the number of frames used for analysis–data from only the second half of the trajectory (t =50-100 ns) was employed in the latter instance. From Figure F.3 it can be seen that the free energy of AuBP-1 adsorption onto all three surfaces was *approximately* stationary over the *majority* of this later time period.

8.4.4 Scheme 4: 'Time period'

To prevent frames sampled during the early/late stages of a metaD trajectory from contributing significantly more to the final results, based on time alone, then it may be more appropriate to calculate the unbiased probability of an observable, Obs, in small intervals of time; structures could be weighted and normalised within each time segment using the average bias potential for the interval. By substituting B$((X(t)))$=$exp\{W/kT\}$ into Equation 2.31, the unbiased ensemble average value of Obs, $< Obs >$, is:

$$< Obs >= \frac{< B(X(t))Obs(X(t)) >}{< B(X(t)) >} \tag{8.3}$$

which can be split up into the sum over time intervals t'=1–n:

$$< Obs >= \frac{1}{\sum_{t'=1}^{n} R_{t'}} \left(R_1 \frac{\int_0^{t_1} \int Obs(X(t'), t')\bar{B}(X)\mathrm{d}X\mathrm{d}t'}{\int_0^{t_1} \int \bar{B}(X)\mathrm{d}X\mathrm{d}t'} + R_2 \frac{\int_{t_1}^{t_2} \int Obs(X(t'), t')\bar{B}(X)\,\mathrm{d}X\mathrm{d}t'}{\int_{t_1}^{t_2} \int \bar{B}(X)\,\mathrm{d}X\mathrm{d}t'} + ... \right) \tag{8.4}$$

This scheme, labelled 'time period' (or 'tp') was implemented using $B(X(t))$=$exp\{V(s(X(t)), t)/kT\}$ and a time interval of 1 ns. Specifically the work flow was as follows:

1. Calculate the reweighted average of the observable within each segment of time $Obs_{t'}$, t=t_1 to t=t_2. To do this, the weight given to a frame at time = t was set equal to $exp\{V_{t2}(s(X(t)), t)/kT\}$ (*i.e.* the bias added to the CV being sampled at time t determined using the bias potential generated at the end

of the time segment, time t_2). $Obs_{t'}$ was normalised in each time segment to counterbalance the fact that the total bias, $V(s(X(t)), t)$, grows continuously during a simulation.

2. The contribution of the average value of the observable over an individual time period, $Obs_{t'}$, to the overall trajectory-averaged value, $< Obs >$, was determined by the range of the CV being sampled in the time interval, *t*'. Time periods during which the simulation samples the CV close to a free energy minimum extensively should be given more importance than those during which it primarily visits higher energy regions. The weight given to each time period thus requires prior knowledge of the free energy profile of the system along the direction of the biased CV, $G(s(X))$. The average weight given to a time interval, $R(t')$, was:

   $$R(t') = \frac{\sum_{t=t_1}^{t=t_2} exp(-G(s(X(t)))/kT)}{t_2 - t_1} \tag{8.5}$$

   Two variants of the reweighting Scheme 4–differing only in the definition of $G(s(X))$–have been tested here. The first, 'time period 1' (or 'tp1'), employed the free energy profile generated after 100 ns of simulation as the best estimate of $G(s(X))$, labelled $G_{f1}(s(X))$. As discussed in Section 8.5.1, for complex systems such as those considered in this study, both the slow diffusion of the peptide and convergence in sampling along directions on the PEL orthogonal to the biased CV, introduce errors into the free energy profile derived in this manner.

3. Within the 'time period' reweighting method, however, the free energy profile a system can be improved in a self-consistent manner. Using the weights, $B(X(t))$ derived from iteration i, the CV biased in a simulation can be treated as an observable. Its unbiased probability distribution, $P(s(X))$ can therefore be calculated from the simulation data. Inverting $P(s(X))$ according to the standard relationship:

   $$G = -kTln(P(X)) \tag{8.6}$$

   generates $G_{f_{i+1}}(s(X))$, an improved free energy profile which can be fed-back into step 2, above.

4. After 5 rounds of iterations, the free energy profiles of the three systems investigated in this work–AuBP-1 adsorption at the aqueous Au(111), Au(100)(1×1)

and Au(100)(5×1) interfaces–were observed to converge. The 5th generation profile, $G_{f5}(X)$, was used in the second variant of the reweighting Scheme 4, 'time period 5' (also referred to as 'tp5').

8.4.5 Scheme 5: Tiana method

In 2008 Tiana proposed a method whereby the microscopic ensemble average value of a variable y, orthogonal to the CV biased in a metaD simulation (x), can be estimated. The reweighting scheme is based on the fundamental assumption that barriers in the PEL of the system projected in the y plane are small. Between successive updates to the bias potential, the system explores all accessible conformational phase-space in the y direction, on a given range of x. The algorithm, outlined in detail in Reference Tiana [2008], requires knowledge of x, y, the potential energy of the entire system, U, and metaD bias, V, each time the bias is updated and at all time intervals in between. Briefly, it operates as follows:

- For each interval of time between successive updates of the metaD bias potential, t, the partition function of the system corresponding to the range of x sampled, Z_t, can be approximated from average Boltzmann factor of the energy term ($U+V_t$):

$$Z_t = \Omega(A_t) < exp(-(U + V_t)/kT) >_{A_t} \tag{8.7}$$

  where A_t is the largest range of connected x sampled in the time period and $\Omega(A_t)$ is the 'volume' of A_t (taken to be proportional to A_t).

- By dividing the entire range of x visited during a simulation into bins, Λ_i ($= x_{i+1} - x_i$), the overall partition function for the system, Z, can be calculated at the end of the trajectory. The product $Z_t < exp\{V_t/kT\} >_{A_t}$ contributes to Z if the range of x sampled within the time interval t, A_t, includes Λ_i (*i.e.* $t : \Lambda_i \subset A_t$) by the expression:

$$Z = \sum_i \frac{1}{n_s(\Lambda_i)} \sum_{t:\Lambda_i \subset A_t} Z_t < exp(V_t/kT) >_{A_t} \tag{8.8}$$

  where $n_s(\Lambda_i)$ is the number of unique time intervals in which the region of x-space corresponding to Λ_i is visited during a simulation.

- Using the partition functions–both partial and full versions–calculated in Equations 8.7 and 8.8 the unbiased ensemble average $< y >$ can be estimated:

$$< y >= \frac{1}{Z} \sum_i \frac{1}{n_s(\Lambda_i)} \sum_{t:\Lambda_i \subset A_t} Z_t < yexp(V_t/kT) >_{A_t} \tag{8.9}$$

Tiana tested the algorithm on two 'toy' 2-D systems–the second of which differed from the first only in the addition of a 'roughness' factor added in the y dimension–as well as using it to estimate the mean ψ dihedral angle of alanine dipeptide in vacuum from a metaD simulation in which the ϕ angle was biased [Tiana [2008]]. In the case of the simplest model system, results predicted by the reweighting scheme were compared to analytical solutions. Two conclusions were drawn: first, the accuracy of the scheme decreased with increasing temperature; and second, the algorithm failed if the length and number of points sampled in each time interval between successive updates of the metaD bias was too small. In turn, to maintain the same degree of accuracy at two different temperatures, a longer time interval had to be employed in the hotter case. To date, there are no reports in the literature of this reweighting scheme, proposed 5 years ago, being applied to more complex systems, such as those under aqueous conditions.

8.4.6 Scheme 6: Bonomi Method

Scheme 6, derived by Bonomi *et al.*, is designed for reweighting CVs orthogonal to those to which the metaD bias is added during a well-tempered metaD simulation [Bonomi et al. [2009b]]. Well-tempered metaD is a recent variant of metaD where the height, w, of a Gaussian hill added to the biased CV (Equation 2.28) incrementally decreases with time [Barducci et al. [2008]] according to:

$$w = \omega_0 \tau exp\{-V(s(s(X(t))), t)/\Delta T\} \tag{8.10}$$

where ω_0 is the initial bias deposition rate, τ the frequency with which Gaussian hills are added during a simulation and ΔT a temperature representing the well-tempered bias factor. In contrast to normal metaD, the total bias, V$(s(X(t)), t)$, thus, does not continue to grow as a simulation progresses, but converges to a time independent limit of:

$$V(s(X(t)), t \to \infty) \to -\frac{\Delta T}{\Delta T + T} G(s(X)) \tag{8.11}$$

where T is the system temperature. Assuming V$(s(X(t))$ evolves adiabatically (*i.e.* that the rate of growth of V is independent of $s(X)$) then Parrinello and *co-workers* suggested that the rate at which the biased distribution of an observable changes, $P_{biased}(Obs, t+\Delta t)$, can be related to the rate of change of V, $\dot{V}$, by the approximate expression:

$$P_{biased}(Obs, t + \Delta t) = exp\{-(\dot{V}(s(Obs), t) - < \dot{V}(s, t) >)\Delta t/kT\}P_{biased}(Obs, t) \tag{8.12}$$

From $P_{biased}(Obs, t)$ and V$(s(X(t)), t)$, the unbiased probability distribution $P_{unbiased}$ can be recovered (Equation 2.31). The algorithm needed to carry out this reweighting procedure is distributed as part of the PLUMED software package [Bonomi et al. [2009a]] and can be implemented when post-processing a well-tempered metaD simulation. Whilst this scheme has been peer-reviewed and is now an established method for reweighting well-tempered metaD simulations, some of the assumptions–foremost, V$(s(X(t))$ adiabaticity–upon which it is based, undermine its suitability for analysising the REST metaD runs carried out in this Chapter.

8.4.7 Evaluation

To evaluate each of the reweighting schemes outlined above, a number of properties of the three systems–AuBP-1 adsorbed at the aqueous Au(111), Au(100)(1×1) and Au(100)(5×1) interfaces–were calculated and the results compared. As expected, due to the inherent propensity of some schemes to exponentially weight structures sampled earlier/later in a metaD trajectory greater (for example, Schemes 2 (ΔV) and 1 (V), respectively) the conclusions which could be drawn from such findings were dependent on the method employed. In what follows, the propensity for R10 and R11 to directly bind to the Au(100)(1×1) surface has been used as an example (Table 8.1).

	raw %	1 (V) %	2 (ΔV) %	3 (ΔV equilib) %	4 tp1 %	4 tp5 %	6 (Bonomi[a]) %
R10	11.8	70.4	100.0	60.5	54.8	57.0	20.2
R11	13.9	18.0	90.3	14.2	37.5	46.9	23.0

Table 8.1: Percentage of frames in which residues R10 and R11 make direct contact with the Au(100)(1×1) surface under aqueous conditions, calculated using the different reweighting schemes outlined above. 'raw' corresponds to bias data extracted from the metaD simulation without reweighting;[a] reference Bonomi et al. [2009b]. No data is available for Scheme 5 (Tiana [Tiana [2008]]) due to the numerical instability of the algorithm.

As expected, due to the inherent propensity of Scheme 1, and by selecting frames from the second half of the trajectory for analysis only in Scheme 3, there is qualitative agreement between the results of the two methods: R10 displays significantly greater affinity for the Au(100)(1×1) surface than R11. In contrast, analysis using Scheme 2–which implicitly favours initial structures–suggests that both R10 and R11 are important in mediating strong gold binding. After inspecting the evolution ΔV with time during this simulation (Figure F.2b), it was possible to observe two phases–$t < 30$ ns and $t > 30$ ns–during the second of which ΔV was approximately constant. Consistent with this trend, the binding propensities of R10 and R11 were almost invariant when the range of time frames selected in Scheme 3 was extended from $t =$ 50-100 ns to $t =$ 30-100 ns (R10/R11 were bound to the Au(100)(1×1) surface in 58.8/13.9% of frames respectively).

Analysis of the raw data suggests differences exist in the mode of AuBP-1 adsorption to the Au(100)(1×1) surface between the two time zones, $t </> 30$ ns. The effect is two-fold. Firstly, the ratio of frames in which R10 is adsorbed and R11 desorbed to that in which R10/R11 are unbound/bound is ∼3:2 in the later 'phase'. Secondly, during this time, the peptide COM–surface separation is larger on average for the latter set of conformations (R10 desorbed and R11 adsorbed to the surface), decreasing their weighting in the analysis. It is this second binding characteristic, rather than the proportion of frames in the raw data which feature R10/R11 as the single tethering point, which differs in the initial phases. Ideally, the two modes of binding should be weighted equally, or at least not unequally in a manner biased solely on the time point in the simulation at which they were sampled. Reweighting Schemes 1-3 all fail in this regard and therefore were not considered for the final data analysis. It must also be noted that the agreement between Schemes 1 and 3 in terms of the physical results that they predicted was only reasonable in the case of the Au(100)(1×1) interface. As can be observed from Figures F.2a and c, the magnitude of ΔV remains large and transient for the majority of the metaD simulations at the Au(111) and Au(100)(5×1) surfaces. Thus, in these two cases, while Scheme 1 inherently gives more importance to frames in the final 10-20 ns of the trajectory, Scheme 3 weights conformations at the beginning of the time period considered ($t = 50$ ns) greater, leading to discrepancies.

It was not possible to implement Scheme 5, the reweighting method proposed by Tiana [Tiana [2008]], to the complex bio-interfacial systems considered here; fluctuations in the potential energy, U, on the order of 4000 kJ mol^{-1} make the algorithm numerically unstable (Equations 8.7 and 8.8). This shortcoming of the scheme has been noted previously [Bonomi et al. [2009b]]. In addition, the fundamental assumption on which the algorithm is based–that phase space orthogonal to the biased CV is extensively sampled between updates to the bias potential–was not met. In part this was due to the manner in which the simulations were conducted, with the metaD bias being added more frequently (0.5 ps) than frames being saved for analysis (1 ps). However, even had the metaD bias been evolved much more slowly, the complexity of the system, again, would have most likely prevented this criterion from being satisfied [Jambeck and Lyubartsev [2013]].

Whilst the reweighting algorithm for analysing well-tempered metaD simulations, Scheme 6 [Bonomi et al. [2009b]], is numerically stable, like Scheme 5, insufficient data between successive updates of the metaD bias potential was available (due to the way in which the simulations were conducted) to justifiably use it, in its current form. In addition, as stated previously, the method is based on the assumption of $\mathrm{V}(s(X(t))$ adiabaticity. Whilst this criterion is satisfied as $t \rightarrow \infty$ for a well-tempered metaD simulation, it is not held in the case of normal metaD prior to convergence in the free energy profile. Thus, reweighting Scheme 6 was also thought not to be suitable for physically interpreting the results of the REST metaD simulations carried out here.

Disregarding Schemes 1-3 and 5-6, leaves only the two variants of the 'Time Period' Scheme 4: tp1 and tp5. The advantage of this method is that, depending on the accuracy of the approximate free energy profile on which data from individual time intervals is weighted, the 'value' given to each peptide structure is independent of the time at which it was sampled. For modelling AuBP-1 adsorption to the aqueous Au(100)(1×1) interface, employing either the 1st or 5th generation free energy profile (G_{f1} or G_{f5}, respectively, Figure 8.2b) in the reweighting scheme has only a marginal influence on the final results (Table 8.1). For the other two surfaces, however, where the profiles at the end of the metaD simulations are less developed, the differences are more pronounced. Ideally, all three simulations would be run until $\Delta\mathrm{G}(s(X(t))$ is invariant with time (or at least, as discussed in Section 8.5.1, the profile averaged over different time intervals at the end of the simulation was stationary). Without access to unlimited computational resource, such a demand is presently not feasible.

The self-consistently derived 5th generation G_{f5} is perhaps a better estimate of the exact free energy profile of the system along the direction of the CV than G_{f1} for all three systems. By being constructed from the unbiased distribution of the CV throughout the entire trajectory, the issue associated with the region of phase space closest to the top/bottom surface of the gold slab being sampled more in the early/late stages of the trajectories, due to the way in which the simulations were initially set up, is overcome. In particular, for Au(111) and Au(100)(5×1), G_{f1}, due to this sampling issue, features a deeper/broader free energy minimum close to the bottom gold surface of the nearest periodic image of the slab (Figure

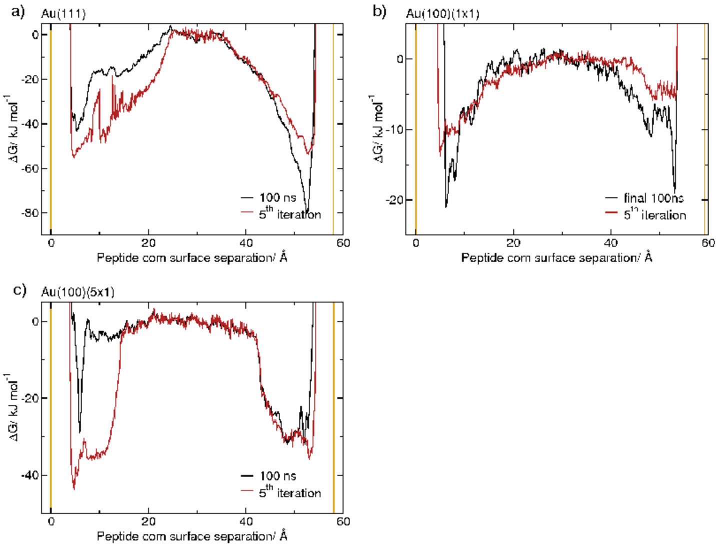


Figure 8.2: Free Energy Profiles G_{f1} and G_{f5} of AuBP-1 adsorbed at the aqueous a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1) interfaces.

8.2a, c). As a consequence, time intervals during which this region of phase-space was more extensively sampled (the majority of which fall in the second half of the trajectory) are given more importance than those during which the symmetrically equivalent region–the free energy minimum close to the top surface of the gold slab–was primarily visited. Thus, a slight bias towards frames at the end of the simulation is introduced into the reweighting scheme 'tp1'. The symmetry of the simulation cells is, on the whole, better represented by G_{f5} than G_{f1} (Figure 8.2). The time dependent unequal weighting of peptide structures in 'tp1' scheme is therefore to corrected for in scheme 'tp5'.

After careful consideration of all the reweighting options outlined in this Section, Scheme 4 'tp5' was the most favourable. It has therefore been used in all analysis of the three REST metaD simulations reported hereafter. The number of different methods for interpreting the structural information generated during a biased metaD trajectory investigated here was not exhaustive. In addition, it is anticipated that as the use of metaD as a simulation technique–either in combination with replica exchange methods or in isolation–becomes more widespread, motivation in the scientific community to develop better and/or more efficient protocols for eliciting as much information as possible from the trajectories will increase.

As well as the reweighting schemes discussed above, there are two alternative approaches which could have been adopted here:

1. **Bound frames:** The CV (peptide COM–surface separation distance) could be divided into 'adsorption zones' (*e.g.* 0-7.5 Å, 7.5-10.0 Å, 10.0-15.0 Å, 15.0-20.0 Å and bulk water) and frames from each metaD simulation binned accordingly. By analysing the physical characteristics of the peptide in each zone separately, the properties of AuBP-1 when in close contact with the Au(111) surface could be compared and contrasted with those when adjacent to Au(100)(1×1) *etc.* Using 'raw' data makes the assumption that the free energy profile of the peptide is constant within each adsorption zone–something which is only approximately true at distances far from the interface. A further problem with this method is the use of arbitrarily chosen cut-off distances to demark between the adsorbed and desorbed states of AuBP-1.

2. **Further Simulation:** The biased PEL of each system built up over the 100 ns of REST metaD simulation carried out here could be used as the initial PEL employed in subsequent REST MD runs. Data from the extended portion of the simulations, where no additional metaD bias is added, could be reweighted in a straight forward manner using Equation 2.31. Whilst this two step protocol–run a REST metaD simulation until the system samples all relevant states along the biased CV with equal likelihood and then use the resulting biased PEL in an extended REST MD simulation–is both simple and can be used to recover information about the Boltzmann weighted ensemble of states of the system, it is not computationally efficient. Not only is all the structural information generated in the first, REST metaD, simulation lost, but the second, REST MD, simulation must be of significant length to enable all parts of, what is an essentially flat PEL, to be exhaustively sampled.

8.5 Results

8.5.1 Binding Strength

While the primary aim of employing metaD in combination with REST in this study was to overcome the issues associated with strong surface adsorption hindering conformational sampling of the REST MD method (Section 5.4), rather than to accurately calculate the free energy of adsorption of AuBP-1 to the three different crystallographic planes of gold, it is none-the-less interesting to consider whether there is a trend in the relative binding affinities of the peptide to the different surfaces. First it is noted that even after 100 ns of simulation neither the shape of the adsorption free energy profile (Figure 8.3) nor the magnitude of the free energy of adsorption (Figure F.3) has ‘converged’. (By ‘converged’ I mean stationary–within the limit of small fluctuations on the order of $w = 0.1\,\mathrm{kJ\,mol^{-1}}$ (the height of a Gaussian hill)–with time.) Whether strict ‘convergence’ within computationally feasible simulation time scales, for complex systems such as those considered here, is possible, is questionable. Certainly in this instance the simulations were setup in a non-optimal manner, with the peptide initially clustered to one region of the CV in the majority of replicas. Not only must the metaD bias be built up sufficiently in all replicas for the PEL to be essentially flat, but sufficient time must have elapsed for a random walk on this landscape to disperse these clustered replicas across CV space, for the adsorption profiles to begin to converge. While progress towards satisfiying the first of these conditions was made during the simulations reported here, the time scale for the second is much longer.

As highlighted by the recent paper by Jambeck and Lyubartsev [Jambeck and Lyubartsev [2013]], the inability to adequately sample all of conformational space orthogonal to a biased CV within the time period of a simulation will also hinder convergence of a metaD run. In the case of the biointerfacial systems modelled here, it is the inter-conversion between different modes of peptide adsorption, in particular, which could impeed convergence. For instance, if the transition between two binding states of AuBP-1 is only possible by the following mechanism–complete desorption, re-folding/orientation in solution and readsorption in an alternative manner–then the time scale for this process would be slow.

As stated in Section 8.4.7, the symmetrical nature of the CV in the REST metaD simulations reported here means that two estimates of the binding affinity of AuBP-1 for the crystallographic plane in question–the top and bottom surface of the gold slab and its periodic neighbour, respectively–can be attained per run. A simple estimate of the uncertainty in the binding free energy–adopted herein–at a given interface, is thus the difference between these two values. Others in the past have calculated this error boundary as the maximum fluctuation of $\Delta \mathrm{G}_{min}^{ads}(t)$ over the final stages of a metaD run, T_f, away from its average value, $< \Delta \mathrm{G}_{min}^{ads} >_{T_f}$, taken over the same time period [Schneider and Ciacchi [2012]]. For AuBP-1–gold adsorption the magnitude of the error estimated by the two different methods was comparable.

Independent of the adsorption free energy profile used–ΔG_{f1} or ΔG_{f5}–there was the following (statistically significant) energetic trend:

$$\mathbf{Au(111) > Au(100)(5{\times}1) > Au(100)(1{\times}1)}$$

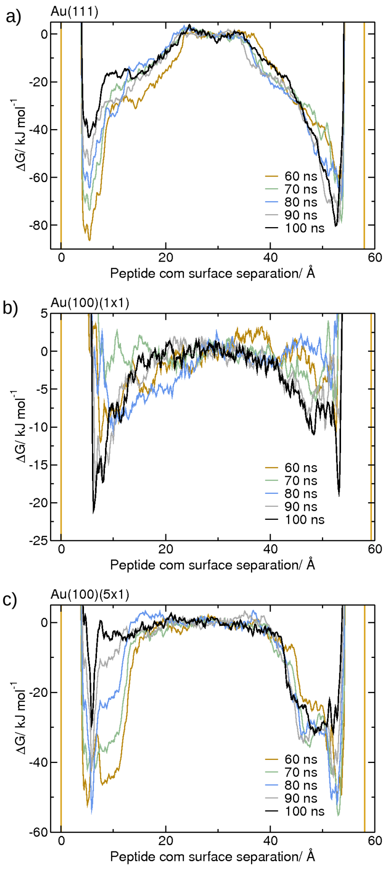


Figure 8.3: The free energy of adsorption profile for AuBP-1 binding at the aqueous a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1) interfaces. Profiles shown after 60, 70, 80, 90 and 100 ns of metaD simulation.

	ΔG_{f1}/ kJ mol^{-1}	ΔG_{f5}/ kJ mol^{-1}
Au(111)	-80.3 ± 37.1	-55.1 ± 1.5
Au(100)(1×1)	-21.0 ± 2.4	-13.5 ± 7.8
Au(100)(5×1)	-31.8 ± 2.7	-43.5 ± 7.8

Table 8.2: Free energy of adsorption of AuBP-1 at the aqueous Au(111), Au(100)(1×1) and Au(100)(5×1) interfaces estimated using both the ΔG_{f1} and ΔG_{f5} binding profiles (Section 8.4.4 and Figure 8.2).

i.e. AuBP-1 adsorbs more strongly to Au(111) than Au(100) under aqueous conditions, irrespective of the reconstructed status of the latter surface. These conclusions are consistent with the previous work of Heinz and *co-workers* [Heinz et al. [2009, 2010]; Yu et al. [2010, 2012]] who modelled peptide adsorption at the native form of the two aqueous interfaces using the CVFF/CHARMM-METAL FF [Heinz et al. [2008]; Pandey et al. [2009]; Feng et al. [2011]]. However, they also confirm the hypothesis put forward on the basis of MD simulations employing GolP-CHARMM [Wright et al. [2013b,c]] in which the characteristics of amino-acids bound to the Au(100)(1×1) and Au(100)(5×1) surface were compared, Section 6.4 [Wright et al. [2013c]]. Namely, that peptides have stronger affinity for the reconstructed form of the surface, than native, under aqueous conditions. Therefore, whilst seemingly inline with current opinion in the literature, this study–the first to investigate trends in the *free* energy of peptide–gold binding at different crystallographic planes of the metal–adds *significantly* to scientific knowledge in a number of ways:

- First, the exact structure of Au(100) faces–native or reconstructed–present on the surface of a AuNP has an important influence on the energetic selectivity of binding between Au(100) and Au(111). In the limit of planar interfaces, it is known that the Au(100)-*hex* structure is the energetically most favourable under experimental conditions. A simple model for Au(100) structure–AuNP growth might be a gradual reduction in the surface energy of the reconstructed form of the surface, relative to the native, as overall particle size increases. In terms of peptide adsorption (assuming that the biomolecule binding in itself does not alter the relative surface energies of the native and reconstructed forms of Au(100)), then Au(111)/Au(100) selectivity would decrease with NP size. This potentially could have consequences for using peptides to mediate shape-selective AuNP synthesis, especially in the realm of larger NPs.
- Second, differential binding of AuBP-1 between Au(111) and Au(100)(5×1) is contary to the 'epitaxial match' mechanism of peptide–gold adsorption proposed by Heinz *et al.* [Heinz et al. [2009]]. Following his conjecture, adsorption

to the two interfaces would be comparable since in both cases the surface atoms are arranged hexagonally.

Finally, as both a benchmark of the ‘convergence’ of the metaD simulations performed in this Chapter and the GolP-CHARMM FF derived in Chapter 6, it is interesting to compare these calculated binding affinities for AuBP-1–gold adsorption with experiment. The latter, $-8.9\,\mathrm{kcal} \pm 0.2\,\mathrm{mol}^{-1}$ ($-37.4\,\mathrm{kJ} \pm 0.8\,\mathrm{mol}^{-1}$), was determined from SPR measurements of the peptide (in phosphate buffer) bound to a polycrystalline gold surface [Hnilova et al. [2008]]. While the exact composition of the surface probed experimentally is unknown, it is reasonable to suppose that it primarily featured the Au(111) interface, the lowest energy crystallographic plane of gold. The slightly stronger affinity of AuBP-1 for Au(111) calculated here can in part be accounted for by insufficient sampling in the region of bulk water, midway between the gold slab and it nearest period neighbour, evidenced by a lack of ‘convergence’ in the corresponding free energy profile (Figures 8.3a, F.3a). Other reasons for the discrepancy between the two values–experiment and calculated–could include: the presence of surface defects and grain boundaries experimentally, not included in the computational model; contributions from binding to the Au(100) and other higher energy crystallographic planes of gold; and a higher concentration of AuBP-1 at the gold surface in experiment resulting in electrostatic interactions between adjacent bound peptides becoming significant. Despite these caveats, it it important to note that the overall agreement between experiment and theory is good, with both binding affinities being of the same order of magnitude. If the relationship between potential and free energy of adsorption is linear for both GolP-CHARMM and CHARMM-METAL, then based on the results of amino-acid MD simulations [Feng et al. [2011]; Wright et al. [2013b]], it can be predicted that the latter FF would overestimate AuBP-1 binding by an additional factor of four.

In the following analysis, the possible origin for this differential adsorption of AuBP-1 to different gold surfaces has been probed.

8.5.2 Cluster Analysis

At all three aqueous gold interfaces AuBP-1 is predominantly present in an extended structure corresponding to reference clusters e and f (Table F.2, Figure 8.1). In contrast to solution, the peptide becomes conformationally restricted upon adsorption, as evidenced by a less even distribution in the population of the 9 different clusters (Table F.2). An analogous finding for QBP-1 upon adsorption to either (100) α-Quartz or Au(111) is reported in Sections 5.3.2 and 7.4.1. These results suggest that, in terms of its most likely structure, AuBP-1 undergoes adsorption-induced chain extension when bound to any planar gold interface, independent of the exact spatial arrangement of Au atoms in the surface.

8.5.3 Secondary Structure

While the most likely interfacial conformation of AuBP-1 is independent of the specific crystallographic plane of gold to which the peptide is adsorbed, there was a slight difference in ensemble of all structures at the three surfaces (Figure 8.4). First though, like QBP-1, it is interesting to compare the secondary structural characteristics of AuBP-1 in solution predicted by this computational model–CHARMM22*/TIP3P–to the experimental CD spectrum of the peptide [Hnilova et al. [2008]]. With the caveats highlighted in Section 7.4.1 for the interpretation of CD spectra in mind, the agreement between the two–experiment and theory–is good: both suggest a propensity for AuBP-1 to adopt PPII and random coil structures. On adsorption to the Au(100)(1×1) surface the ensemble of all AuBP-1 conformations is perturbed only minimally, as might be expected due to the weak affinity of the peptide for this interface. On binding to Au(111) and Au(100)(5×1), this secondary structural analysis, in agreement with cluster data, suggests that the ensemble of all AuBP-1 structures undergoes a slight conformational change from compact configurations–with α character–to more extended ones–with increased β content.

8.5.4 Binding Mechanism

Taken together–cluster and secondary structure characteristic–analysis of AuBP-1 at each of the three aqueous gold interfaces probed in this work revealed only marginal differences existed between the surfaces. These variations were too small to reasonably support the notion of differential surface-induced folding being the mechanism responsible for the energetic selectivity in adsorption calculated in Section 8.5.1. Therefore, given the internal configuration of AuBP-1 is approximately the same at the aqueous interfaces of the three crystallographic planes of gold, the orientation of the peptide with respect to the surface and, in particular, differences in the contact made between each residue and the surface gold atoms must play an important role.

The number of residues bound to gold directly on average was calculated for each interface and was found to be 6.7, 3.1 and 6.7 for the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces respectively. The trend in this data correlates with that for the stronger adsorption of AuBP-1 to Au(111) and Au(100)(5×1) than

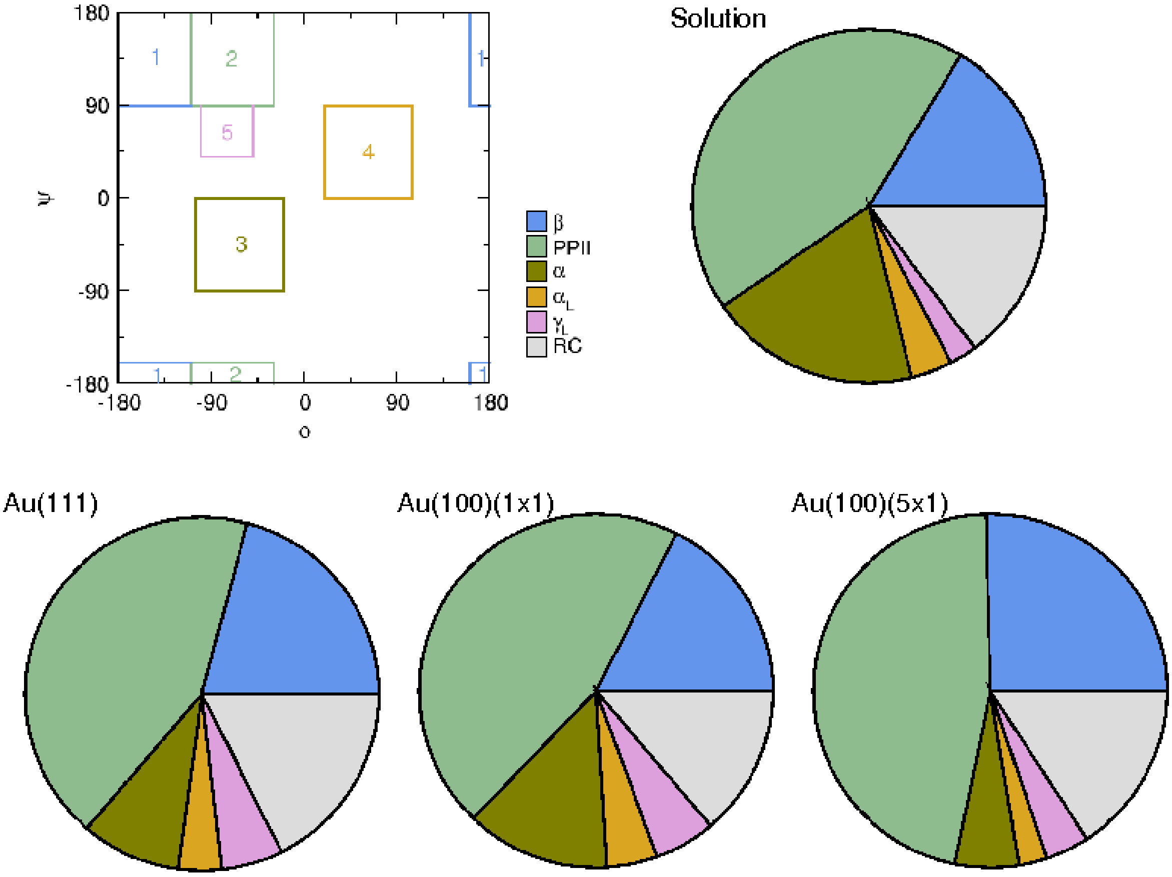


Figure 8.4: Secondary structural characteristics of AuBP-1 in solution and upon adsorption to the Au(111), Au(100)(1×1) and Au(100)(5×1) surfaces under aqueous conditions. Top left depicts the regions of ϕ/ψ phase-space on a Ramachandran plot used to define each characteristic- β, PPII, α, α_L, γ_L and random coil (RC).

Au(100)(1×1). When the binding propensity of the 12 residues is considered and ranked in turn, the results indicate strongly that the number and 'quality' of the anchoring points between AuBP-1 and gold increases concurrently with the binding affinity of the peptide as a whole. Specifically, as shown in Figure 8.5, on adsorption to Au(111)–the surface with the greatest binding affinity–4 residues were strong binders, 4 significant and 1 moderate. An intermediate number of residues mediated adsorption to Au(100)(5×1) (4 strong, 2 significant and 3 moderate); while at the aqueous Au(100)(1×1) interface the fewest made direct contact with gold (1 significant and 5 moderate). The trend in the binding affinities of individual residues is reflected in those of the corresponding sequential dyad and triad sequences (Tables F.3 and F.4).

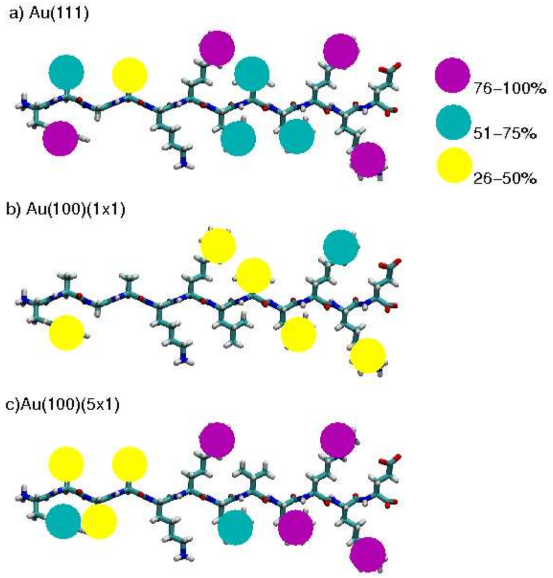


Figure 8.5: The binding propensity of individual residues within AuBP-1 at the aqueous a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1) interfaces rated according to the percentage likelihood of being in direct contact with the gold surface. Residues labelled as 'strong' binders are highlighted in magenta, 'significant' binders in 'cyan' and 'moderate' in yellow.

The significantly greater degree of direct contact made between AuBP-1 and Au atoms in the surface layer of the interface in the case of Au(111) and Au(100)(5×1), compared to Au(100)(1×1), is in agreement with observations made during MD simulations of amino acids bound to the three surfaces [Wright et al. [2013b,c]]. Consistent with the hypothesis put forward in Section 6.4, it is suggestive of two distinct modes of peptide adsorption: the first mediated by direct contact with gold and the second by the layer of interfacial water. Energetic data calculated in this Chapter indicates that the second, solvent-separated adsorption mode is quantitatively weaker than the first. In addition it is interesting to note that the relative binding affinity of AuBP-1 for the three surfaces increases with decreasing density, per unit surface area, of water molecules in the first adsorbed layer (Section 6.3.3). Thus, like as has been reported previously for metal oxide interfaces [Skelton et al. [2009]; Schneider and Ciacchi [2012]], interfacial water structuring plays a critical role in peptide–gold adsorption.

In terms of the nature of the residues anchoring AuBP-1 to the three gold surfaces, there is little difference. Arginine contributes strongly in all cases. The presence of the –RR– motif in the C-terminal portion of the peptide means that this half of AuBP-1 makes surface contact more, on average, than the N-term. On adsorption to Au(100)(1×1), where binding is weakest, AuBP-1 is effectively tethered to the surface by the R10 residue only. The planar aromatic W1, is a 'strong' binder on the atomically flat Au(111) surface, but only a 'significant' anchoring point at the rumpled Au(100)(5×1) interface. Conversely, the bulkier side-chain of the hydrophobic L9 residue, which could feasibly have a greater degree of surface area contact with an undulating surface, such as Au(100)(5×1) than an atomically flat one, contributes more to AuBP-1 adsorption on the reconstructed Au(100) surface than Au(111). However, there is insufficient evidence from this study alone for strong conclusions about the relationship between residue shape and binding propensity for the aqueous Au(111) and Au(100)(5×1) interfaces to be drawn. The conjecture that the adsorption of peptides rich in aromatic planar residues is stronger to atomically flat surfaces, such as Au(111), could be tested in the future.

8.6 Conclusions

In this Chapter it has been demonstrated that Replica Exchange with Solute Tempering can be combined with metadynamics as a simulation technique to interrogate the interfacial structure of a material-binding peptide with relatively high affinity for its respective interface. A small bias is added on-the-fly to a collective variable connecting the peptide to the surface, driving the adsorbate away from the interface and thus, ensuring that both bound and unbound states are sampled. In doing so, the issues associated with strong peptide adsorption hindering replica mobility, and hence conformational sampling, during a REST MD simulation can be overcome. The properties of the system in the Boltzmann weighted ensemble of interest can be recovered from the biased REST metaD trajectory by careful reweighting. Here, after evaluating six different reweighting schemes, the 'time-period 5' method, based on reweighting frames within short time intervals using an average bias potential (Section 8.4.4), was selected as the most appropriate.

Not only is the work in this Chapter novel in terms of reweighted-data analysis applied to the REST metaD trajectories, but it also provides a unique insight into the facet-selectivity of AuBP-1 gold adsorption. Confirming current scientific opinion (Section 1.4.2), the trend in the free energies of adsorption calculated here suggests that peptides bind more strongly to the Au(111) surface under aqueous conditions than the Au(100), irrespective of the reconstructed status of the latter crystallographic plane. However, Au(100) surface reconstruction *does* have a significant impact on both the strength and mode of AuBP-1 adsorption. Binding to Au(100)(5×1) is substantially stronger than that to Au(100)(1×1) with the adsorbed form of AuBP-1 at the Au(111) interface having more characteristics in common (secondary structural propensity and, number and nature of binding residues) with the peptide bound to the reconstructed Au(100) surface than the native. Taken together with differences in interfacial water structuring at the three gold surfaces (Section 6.3.3), these results suggest that solvation plays a critical role in pepitde-gold adsorption, a hypothesis proposed in the past for biomolecule binding to metal oxide interfaces only [Skelton et al. [2009]; Schneider and Ciacchi [2012]].

Chapter 9

Conclusions and Future Prospects

In conclusion, the simulations reported in this thesis indicate that it is indeed possible for peptides to be able to discriminate between different crystallographic planes of a given material upon adsorption to either the aqueous α-Quartz or gold interface. For the former substrate–Quartz–this was demonstrated in Chapter 3 by the energetic and spatial selectivity of small molecule adsorption. That the binding of species which form the fundamental building-blocks of peptides is facet-specific, is highly suggestive of the same differential binding characteristics being displayed by the biomolecules themselves. In the case of gold, the extensive simulations employing advanced sampling techniques at the forefront of current research reported in Chapter 8, provide direct evidence for the strength with which the gold-binding peptide AuBP-1 [Hnilova et al. [2008]] adsorbs to gold to decrease according to the following trend: Au(111) $>$ Au(100)(5$\times$1) $>$ Au(100)(1$\times$1). Continued investigations at both interfaces–quartz and gold–should in principal permit simulation to rationally design facet-selective peptide sequences; a highly desirable outcome in the realm of bionanotechnology and biomimetic materials synthesis (Section 1.1).

In addition to the main conclusion, stated above, there are a number of smaller outcomes to this work, each of which is a significant contribution to the realm of biointerfacial simulation and/or provides a novel insight into the physical system investigated. Firstly, the strong similarities between the binding behaviours of the side-chains of residues within a peptide sequence (Chapters 5 and 7) and amino-acid analogues featuring the same functional group (Chapter 3) validates the use of the latter in future work. This is highly desirable when probing the binding propensity of individual amino-acids to an interface (*e.g.* during parametrisation of a biointerfacial FF) since it enables the use of smaller, simpler systems and hence circumvents many of the problems associated with conformational sampling.

Secondly, the investigations reported in Chapters 3 to 5 at the aqueous Quartz interface all indicate that peptide binding is mediated by strong attractive interactions between hydrophobic residues such as tryptophan and tyrosine. The interaction between ammonium and quartz was found to be minimal. Whilst this later result is contrary to the prevalence of positively charged residues in silica-binding peptides identified experimentally (Section 1.3.1), it was confirmed by first-principles simulations at the aqueous, charge-neutral, fully hydroxylated α-Quartz interface, detailed in Chapter 4. This raises a number of possibilities: the exact nature of the silica surface (*e.g.* amorphous *v.s* crystalline, charge density, topology *etc*) onto which a peptide adsorbs is critical to its binding mechanism; there are two different modes of peptide-silica adsorption–the first mediated by electrostatic interactions and the second by hydrophobic ones; and that the positively charged residues enriched in experimentally identified sequences play an important conformational role, rather than being surface anchoring points.

Thirdly, critical analysis of existing data in the literature and REST simulations of QBP-1 [Oren et al. [2007]] adsorption to the (100) α-Quartz and Au(111) surfaces under aqueous conditions, both presented in Chapter 7, bring into question the specificity of this sequence for quartz. One option for resolving this matter in the future would be carry out free energy calculations, such as the REST metaD simulations reported in Chapter 8, of QBP-1 adsorbed onto the two materials. It is interesting to note, however, that neither in the case of QBP-1 adsorbed onto surfaces of differing composition (Chapter 7), nor the instance of AuBP-1 bound to alternative crystallographic planes of gold (Chapter 8), was there evidence from the simulations carried out here, for discrepancies in peptide structure being the main driving force for any reported/observed material/facet selectivity. All simulations suggested that selectivity, should it exist, was instead mediated by differences in residue–surface contact and the influence of interfacial water structuring on this phenomenon. The possibly that larger, intrinsically disordered, materials binding peptides, such as n16N [Collino and Evans [2008]] or JAK-1 [Capriotti et al. [2007]], may undergo interfacial-selective surface-induced folding remains open.

The work presented in Chapters 5 and 8, benchmarking the Replica Exchange with Solute Tempering (REST) method, is critically important for the future of biointerfacial simulations, if the issues associated with the poor conformational sampling of such complex systems are to be overcome. Although the tests carried out in Chapter 5 showed that REST could achieve a comparable degree of conformational sampling for up to $1/6^{th}$ of the computational resource of a conventional T-REMD simulation, there is still scope for further improvement. This was highlighted in Chapter 8 when the AuBP-1–gold system was considered. For strong material-binding peptides at their respective interfaces, additional measures must be taken–such as the combination of REST with the free energy technique metaD–for the method to remain favourable. There were also a number of the other questions posed by the findings of Chapter 5 which could be addressed by future work. In particular, it would be interesting to investigate the influence of scaling the mass of water and/or side-chain atoms on conformational sampling within a REST MD simulation, using peptide sequences of graded flexibility. This would help establish whether the relatively insignificant impact of mass scaling reported here primarily arose due to the conformationally stiff nature of QBP-1 imparted by its high proline content.

If the use of metadynamics is to become widespread in the future–in the realm of biointerfacial simulation or any other environment–then the issue of how best to reweight the biased trajectories generated, in order to recover properties of the system in the Boltzmann weighted ensemble of interest, must be addressed fully. The evaluation of six different reweighting schemes presented in Chapter 8 provides a foundation for others to build on. In addition to the unique physical insights which were gained from the REST metaD simulations of AuBP-1–gold adsorption, the work raised numerous issues. First and foremost is the choice of collective variable(s) for biointerfacial systems. Peptide COM–surface separation alone is insufficient for ensuring that the whole range of adsorbed and desorbed states is sampled within a computationally reasonable simulation time-frame, when the transition between different modes and/or geometries of adsorption remains slow (Section 8.5.4). Use of more than one biased collective variable, on the other hand, increases both the computational overhead of the simulation and the time taken for ‘convergence’ in the free energy landscape. The definition of ‘convergence’ itself, in the context of an adsorption free energy profile, should be debated by those within the scientific community. With only one dimension, these profiles are inherently weighted averages of multiple adsorption pathways and should strictly be interpreted as such. One approach, which may prove to be more insightful, could be to identify a handful of unique mechanisms possible for peptide binding. Carrying out a series of metaD (or metaD in combination with another advanced sampling technique such as REST) simulations in which the pathway to adsorption is restricted, could help identify the most favourable adsorption mechanism, in a computationally efficient manner.

Finally, without the FFs derived in Chapter 6, it would have been necessary to make a number of severe compromises in order to carry out the simulations at the aqueous gold interface reported later in this thesis (Chapters 7 and 8); shortcomings existed in both of the previously widely used biointerfacial gold FFs, as detailed in Section 1.4.2. Not only can the GolP-CHARMM suite of FFs be used directly by others within the simulation community, but the methodology outlined in its derivation could be adopted to parametrise other noble metal surfaces (*e.g.* Ag, Pt or Pd). If a range of metal FFs were derived using the same procedure, then peptides could be tested, by means of simulation, for selectivity in binding. In addition, the protocol developed for the Au(100)(5×1) surface, to introduce spatially-dependent energetic heterogeneity in adsorption paves the way for others to consider other non-planar and/or non-ideal surfaces, hence moving simulation closer to experiment.

One of the most interesting outcomes of Chapters 6 and 8 in investigating facet selectivity upon adsorption to gold surfaces, is the apparent importance of interfacial water structuring. Previously, water was thought to play a less fundamental role at planar metal interfaces in determining peptide binding than at metal oxide ones. The work carried out here suggests that the in- and out-of-plane structure of Au atoms in the upper-most layer of a gold slab, together with the binding affinity of a single water molecule for an interface *in vacuo*, contribute to AuBP-1 binding significantly more weakly to Au(100)(1×1) than the other two surfaces investigated–Au(111) and Au(100)(5×1). To interrogate these issues further, it would be interesting to use GolP-CHARMM to probe the impact of adsorbate shape and surface topology on binding affinities under aqueous conditions. Simple amino-acid analogues could be considered first, before moving onto peptide sequences.

GolP-CHARMM could also enable scientists to shed light on the reconstructed status of Au(100) facets featured by AuNPs. One method would be to compare the theoretical binding affinities of small biomolecules to the Au(100)(1×1) and Au(100)(5×1) surfaces with appropriate experimental data for the same species adsorbed to AuNPs. Alternatively, analogous to the experimental procedure used by Chui *et al.* [Chiu et al. [2011]], peptide sequences able to discriminate between the planar Au(100)(1×1) and Au(100)(5×1) surfaces could be designed with the aid of simulations using GolP-CHARMM. By measuring the experimental affinity of these engineered peptides for AuNPs, the nature of the Au(100) facet featured by AuNPs of a given size and shaped could be determined.

In closing, two of the grand challenges facing the realm of biointerfacial simulation–the development and benchmarking of FF parameters, and improving conformational sampling within these complex systems–have been tackled here. Two further major issues must be addressed in the future if simulation is to move closer to experiment: first, the use of non-ideal systems which include additives (*e.g.* ions or buffers) and/or surface defects; and second, regimes in which the surface concentration of the adsorbate is high enough for peptide-peptide interactions and/or aggregation to become important.

Appendix A

Additional Information: Free Energy Calculations of Amino-Acid Analogue Adsorption at Aqueous α-Quartz Interfaces.

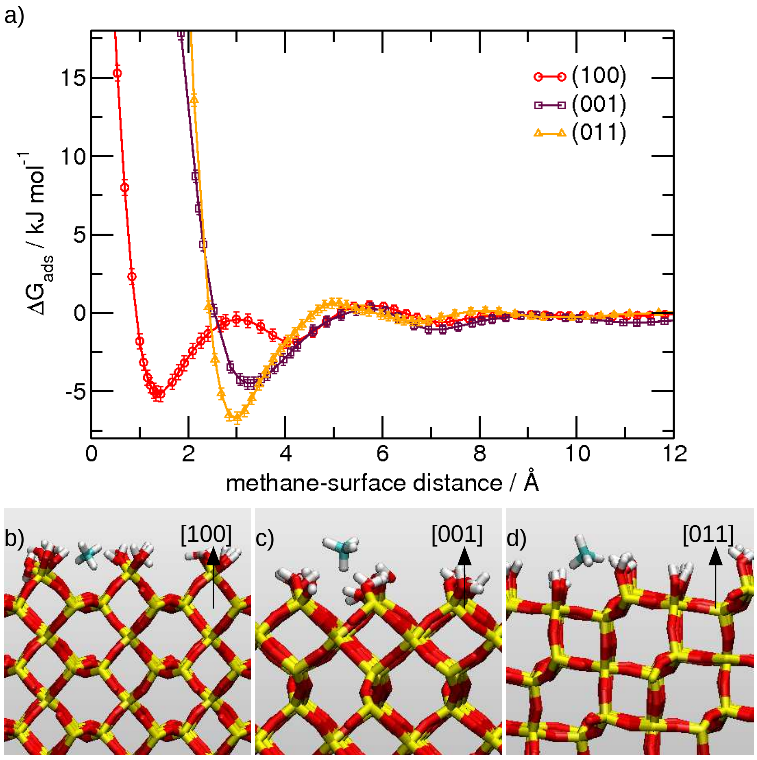


Figure A.1: a) PMF free energy of adsorption methane on (100), (001) and (011) quartz surfaces as a function of surface separation. b) Typical configurations of methane adsorbed onto the (100), c) (001) and d) (011) surfaces.

	(100)	(001)	(011)
quartz-ammonium	0.173 ± 0.391	0.083 ± 0.279	0.009 ± 0.093
water-ammonium	2.947 ± 0.817	3.239 ± 0.755	3.003 ± 0.782
quartz-ethanoate	1.619 ± 0.709	1.699 ± 0.519	1.497 ± 0.612
water-ethanoate	5.263 ± 0.953	5.229 ± 0.902	5.307 ± 0.942

Table A.1: Average number of hydrogen bonds formed by the adsorbates, methanol, ammonium and ethanoate, at the PMF minimum distance above the (100), (001) and (011) α-quartz surfaces. Hydrogen bonds to methanol were defined as the simultaneous satisfaction of the following criteria: O-O_{MeOH} distance of less than 3.46 Å; either H-O_{MeOH} distance of less than 2.50 Å or O-H_{MeOH} distance of less than 2.32 Å; and H-O-O angle of less than 30°; hydrogen bonds to ammonium as the simultaneous satisfaction of the following criteria: N-O distance of less than 3.72 Å; H-N distance of less than 2.24 Å; and H-N-O angle of less than 30°; and hydrogen bonds to ethanoate as the simultaneous satisfaction of the following criteria: O-O distance of less than 3.28 Å; H-O distance of less than 2.46 Å; and H-O-O angle of less than 30°.

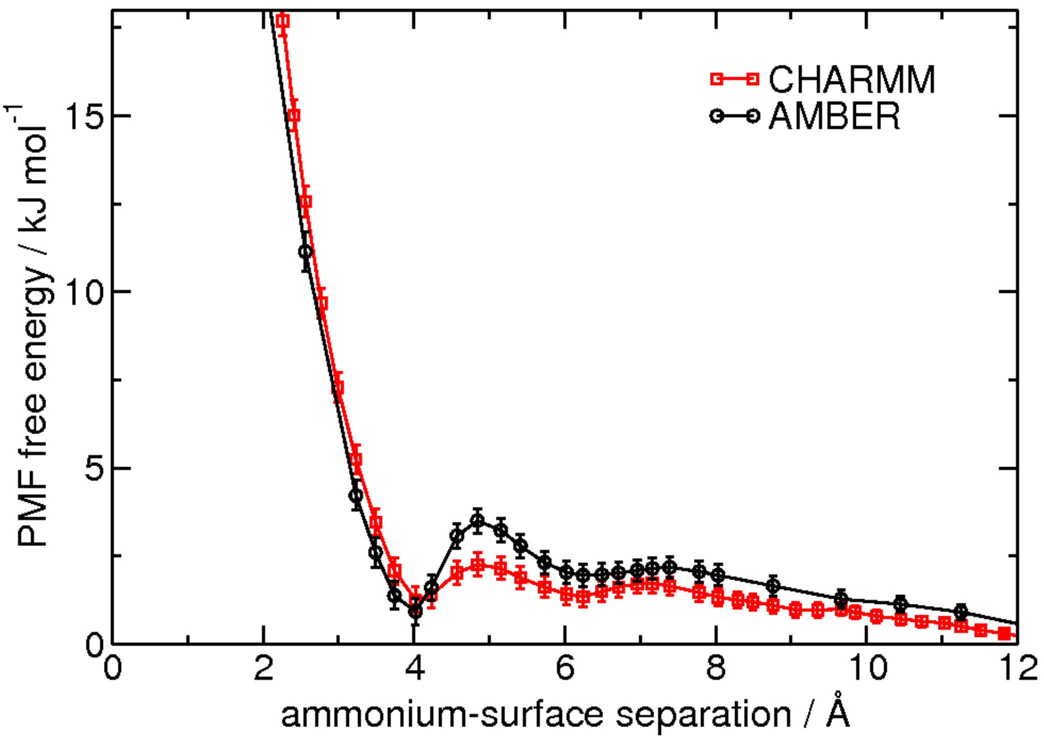


Figure A.2: PMF free energy of adsorption of ammonium on the (100) α-quartz surface as a function of surface separation calculated with CHARMM and AMBER parameters for NH_4^+.

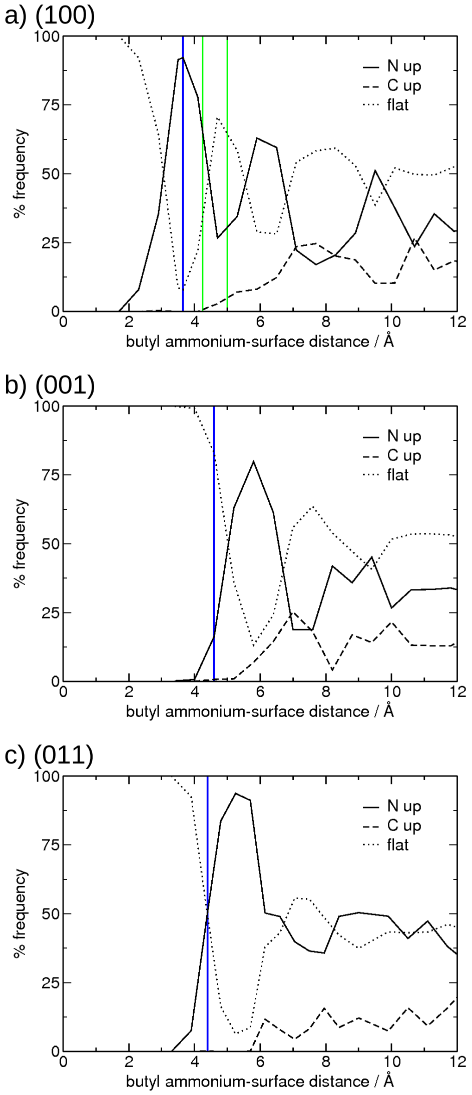


Figure A.3: Orientational analysis of butyl ammonium as a function of vertical surface separation from the a) (100), b) (001) and c) (011) α-quartz surfaces. The angle the CH_3–NH_3^+ vector of butyl ammonium made to the surface normal was calculated as a function of surface separation. Angles less than 60° were classed as 'N-up', between 60° and 120° as 'flat', and greater than 120° as 'C-up'. The surface-separation corresponding to the minimum in the PMF profile of butyl ammonium above each surface is indicated by the blue line. The green lines on the (100) orientation plot a) define the region of the plateau in the PMF profile above this surface.

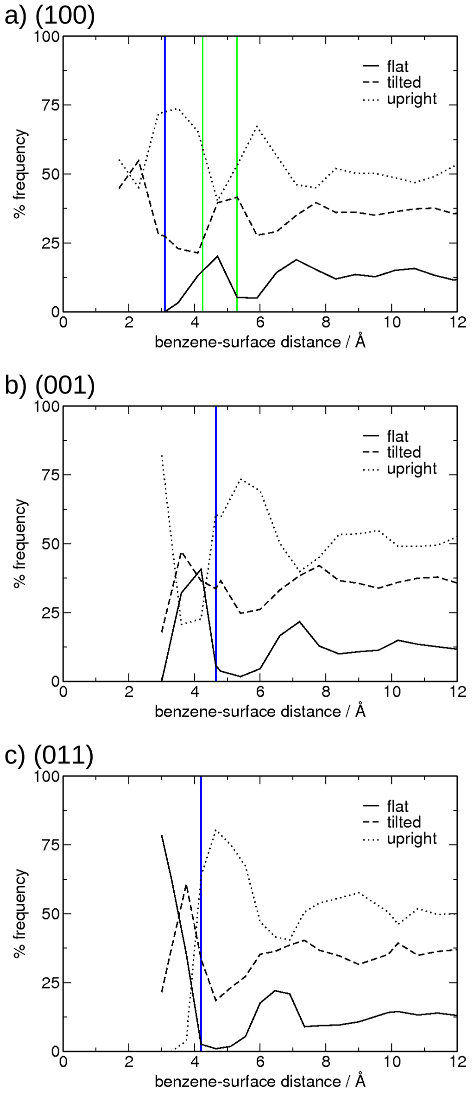
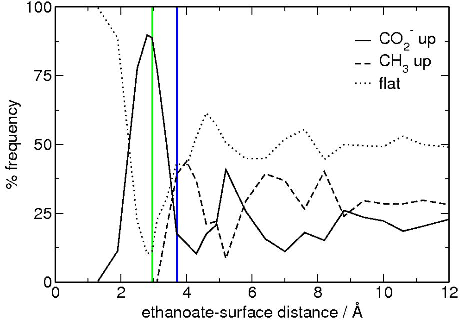
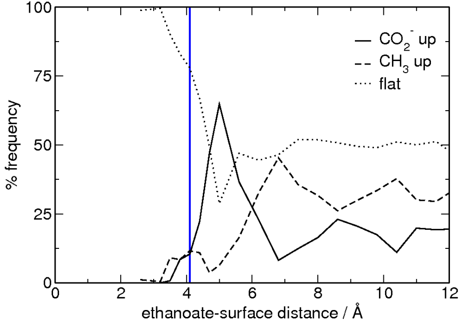


Figure A.4: Orientational analysis of benzene as a function of vertical surface separation from the a) (100), b) (001) and **c)** (011) α-quartz surfaces. The angle between the normals to the planes made by the benzene aromatic ring and the quartz surface was used to classify adsorbed structures as follows: benzene plane made to the surface normal was classified as follows: structures with angles of less than 30° or greater than 150° correspond to 'upright', those greater than 60° but less than 120° to 'flat', and all others to 'tilted'. The surface-separation corresponding to the minimum in the PMF profile of benzene above each surface is indicated by the blue line. The green lines on the (100) orientation plot a) define the region of the plateau in the PMF profile above this surface.

a) (100)



b) (001)



c) (011)

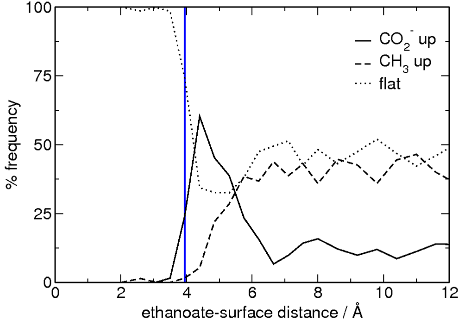


Figure A.5: Orientational analysis of ethanoate as a function of vertical surface separation from the a) (100), b) (001) and **c)** (011) α-quartz surfaces. The angle the CH_3–CO_2^- vector of ethanoate made to the surface normal was calculated as a function of surface separation. Angles less than 60° were categorised as ‘carboxylate-up’, between 60° and 120° as ‘flat’, and greater than 120° as ‘methl-up’. The surface-separation corresponding to the minimum in the PMF profile of ethanoate above each surface is indicated by the blue line. The green line on the (100) orientation plot a) corresponds to the local minimum in the PMF profile above this surface.

Appendix B

Additional Information: First-Principles Molecular Dynamics Simulations of Ammonium and Ethanoate Adsorption at the Aqueous (100) α-Quartz Interface.

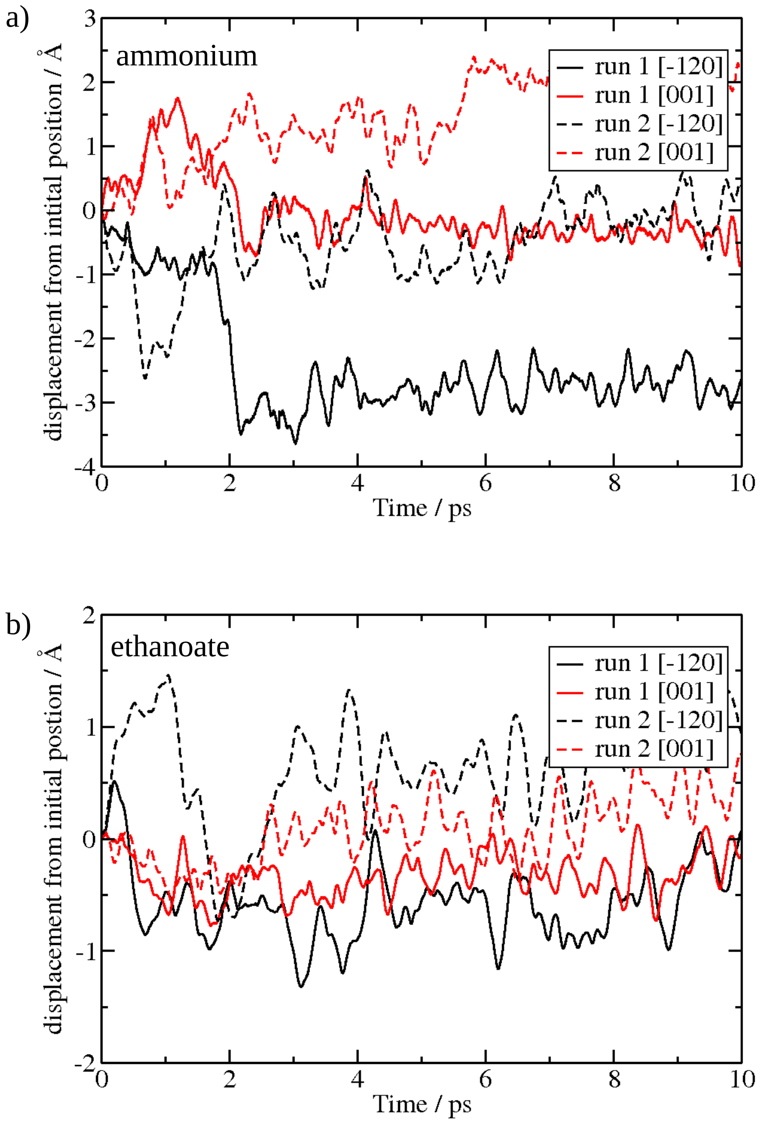


Figure B.1: Displacement of a) NH_4^+ and b) CH_3COO^- from their initial positions, during CPMD simulations, along the $[\bar{1}20]$ and [001] vectors in the (100) quartz surface. These vectors point along the major grooves in the surface.

	CPMD Å	TIP3P/CHARMM Å
$\mathbf{NH_4^+}$		
NH$\cdots$O	2.40	2.39
N$\cdots$O	3.41	3.76
Ethanaote		
OH$\cdots$O$^-$	2.24	2.40
O$\cdots$O$^-$	3.48	3.30

Table B.1: Cut-off distances for hydrogen bond definitions, taken from radial distribution functions for each approach.

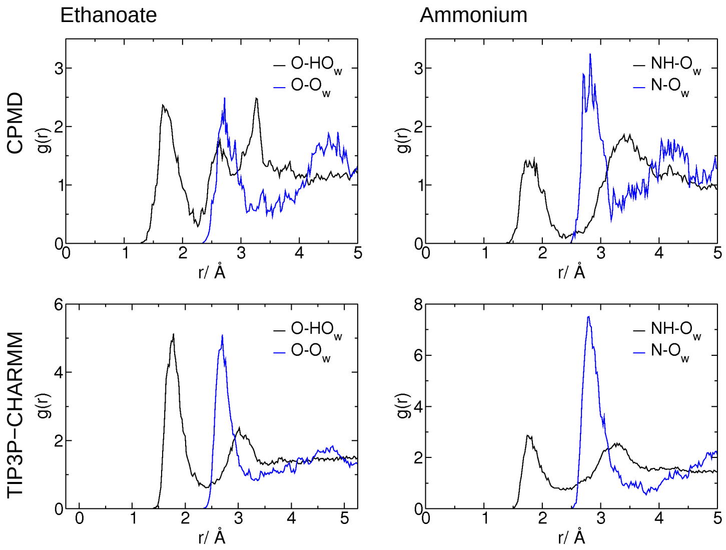


Figure B.2: Radial Distribution Function plots for water surrouding ethanoate and ammonium from which the distance criteria for hydrogen bonding were derived, for both CPMD and FMD approaches.

		number of H-bonds	length Å
quartz-CH_3COO^-			
LFF	run 1	1.898 ± 0.506	1.889 ± 0.235
LFF	run 2	1.645 ± 0.511	1.834 ± 0.180
CPMD	run 1	1.915 ± 0.315	1.721 ± 0.191
CPMD	run 2	2.417 ± 0.532	1.700 ± 0.230
water-CH_3COO^-			
LFF	run 1	3.909 ± 0.832	1.816 ± 0.172
LFF	run 2	5.062 ± 1.045	1.857 ± 0.185
CPMD	run 1	2.583 ± 0.804	1.855 ± 0.213
CPMD	run 2	1.863 ± 0.705	1.919 ± 0.233

Table B.2: Trajectory average number (per timeframe) and average length (Å) of hydrogen bonds formed by CH_3COO^- to water or the quartz surface from both CPMD and FMD simulations. Hydrogen bonds were defined as per the criteria detailed in Section 2.5.1 and defined in Table B.1.

		number of H-bonds	length Å
quartz-NH_4^+			
LFF	run 1	0.019 ± 0.165	2.057 ± 0.196
LFF	run 2	0.404 ± 0.529	2.074 ± 0.161
CPMD	run 1	1.378 ± 0.698	1.889 ± 0.148
CPMD	run 2	0.340 ± 0.530	1.952 ± 0.166
H_2O-NH_4^+			
LFF	run 1	3.230 ± 0.892	1.882 ± 0.165
LFF	run 2	2.704 ± 0.738	1.871 ± 0.161
CPMD	run 1	2.032 ± 0.581	1.858 ± 0.174
CPMD	run 2	2.122 ± 0.770	1.871 ± 0.178

Table B.3: Trajectory average number (per timeframe) and average length of hydrogen bonds formed by NH_4^+ to water or the quartz surface from both CPMD and FMD simulations. Hydrogen bonds were defined as per the criteria detailed in Section 2.5.1 and defined in Table B.1.

Appendix C

Additional Information:Benchmarking Replica Exchange with Solute Tempering for Bio-interfacial Systems: QBP-1 Adsorption onto the (100) α-Quartz Surface.

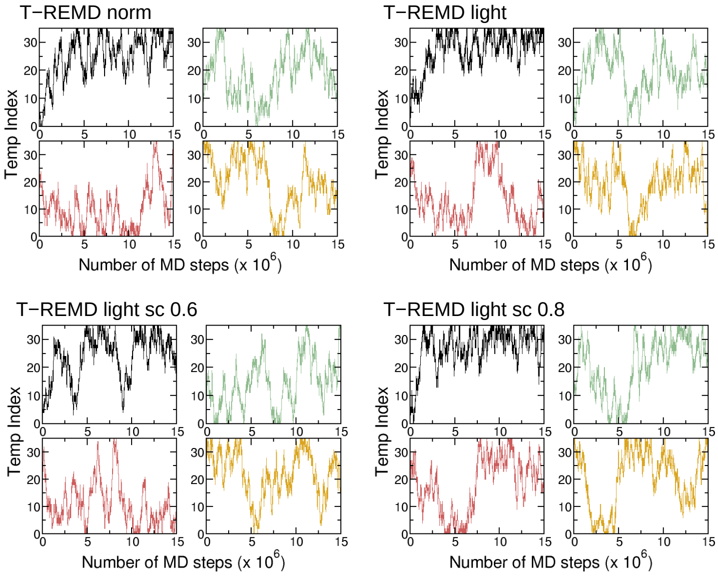


Figure C.1: Temperature mobilities for 4 out of the 36 replicas in each T-REMD simulation. Replicas shown are (clockwise from top left) 1 (black), 12 (green), 24 (red) and 36 (gold).

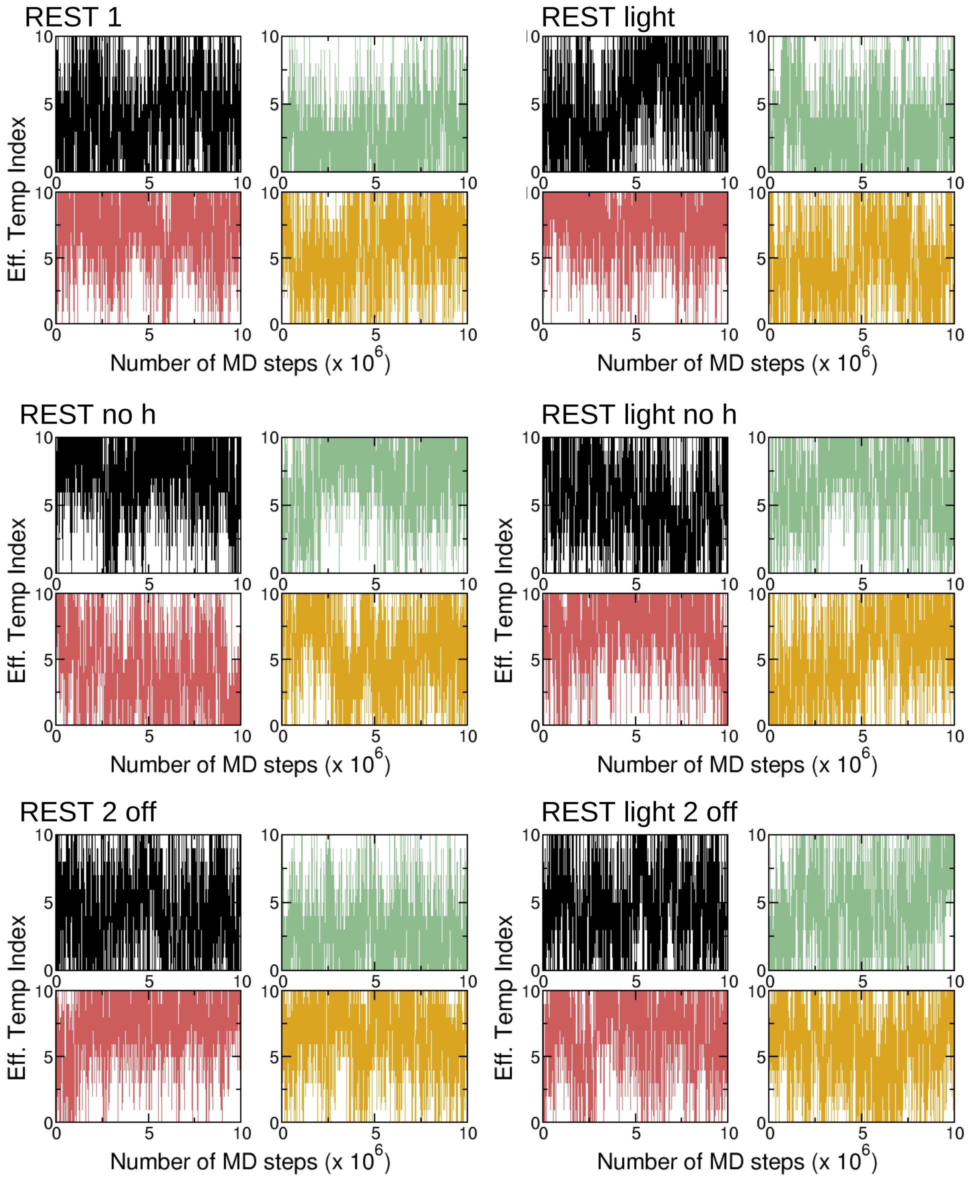
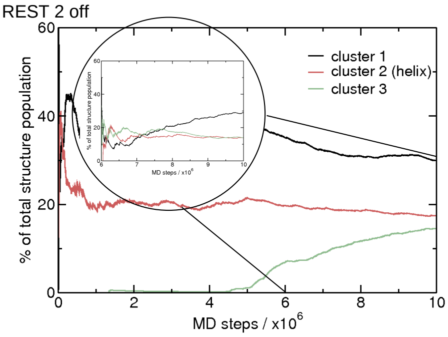
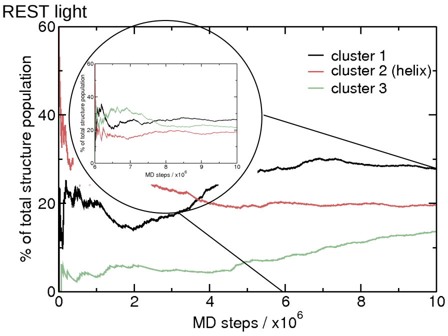


Figure C.2: Effective temperature mobilities for 4 out of the 11 replicas in each REST simulation. Replicas shown are (clockwise from top left) 1 (black), 4 (green), 8 (red) and 11 (gold).

REST 2 off



REST light



REST light 2 off

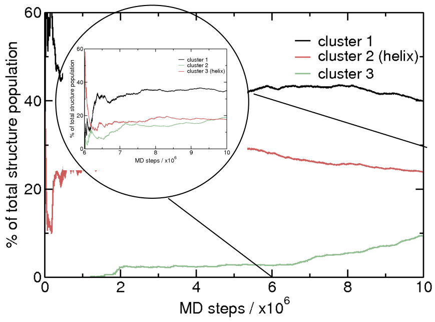


Figure C.3: Percentage of the total structure population belonging to each of the three most populated clusters as a function of MD steps shown for ‘REST 2 off’, ‘REST light’ and ‘REST light 2 off’ simulations. Data shown for cluster analysis performed over the entire trajectory and over the last 4x10^6 MD steps only (inset). Clusters comprised of helical structures are shown in red.

method	% helical-type structures	intra peptide	pept-water	pept-quartz
normal T-REMD	0.0	0.49	27.37	4.89
light T-REMD	0.0	0.23	27.84	4.95
light sc 0.6 T-REMD	0.0	0.42	28.09	4.90
light sc 0.8 T-REMD	0.0	0.27	27.53	5.49
REST1	10.4	0.68	28.16	4.77
REST no h	0.0	0.15	27.97	4.61
REST 2 off	13.9	0.76	28.39	3.80
REST light	18.8	1.01	27.24	4.56
REST light no h	0.0	0.23	27.92	5.41
REST light 2 off	17.5	1.14	27.36	4.64

Table C.1: Average numbers of hydrogen bonds per MD frame.

method	cluster	sol 1	sol 2	sol 3	sol 4
T-REMD normal	1	0.262	0.247	0.591	0.213
T-REMD normal	2	0.382	0.351	0.764	0.277
T-REMD normal	3	0.504	0.563	0.165	0.674
T-REMD light	1	0.385	0.365	0.788	0.267
T-REMD light	2	0.317	0.316	0.670	0.279
T-REMD light	3	0.280	0.246	0.437	0.442
T-REMD light 0.6 sc	1	0.388	0.366	0.786	0.274
T-REMD light 0.6 sc	2	0.296	0.320	0.652	0.273
T-REMD light 0.6 sc	3	0.352	0.344	0.716	0.292
T-REMD light 0.8 sc	1	0.352	0.329	0.764	0.213
T-REMD light 0.8 sc	2	0.195	0.291	0.582	0.328
T-REMD light 0.8 sc	3	0.330	0.318	0.658	0.272
REST1	1	0.394	0.350	0.787	0.235
REST1	2	0.341	0.322	0.668	0.281
REST1	3	0.293	0.313	0.678	0.304
REST no h	1	0.147	0.279	0.579	0.314
REST no h	2	0.323	0.163	0.644	0.198
REST no h	3	0.348	0.408	0.301	0.563
REST 2 off	1	0.348	0.328	0.702	0.262
REST 2 off	2	0.373	0.366	0.498	0.418
REST 2 off	3	0.260	0.129	0.587	0.278
REST light	1	0.337	0.312	0.696	0.279
REST light	2	0.365	0.339	0.783	0.209
REST light	3	0.364	0.350	0.511	0.395
REST light no h	1	0.372	0.316	0.760	0.199
REST light no h	2	0.175	0.336	0.506	0.391
REST light no h	3	0.462	0.528	0.225	0.621
REST light 2 off	1	0.370	0.341	0.770	0.259
REST light 2 off	2	0.377	0.318	0.729	0.276
REST light 2 off	3	0.353	0.341	0.455	0.449

Table C.2: Average RMSD of backbone atom positions of all structures in the three most populated clusters identified by each REMD-based method from the centriod structure of the four most populated solution clusters of QBP-1 [Notman et al. [2010]]. 'Identical' clusters are highlighted in red, whilst 'similar' ones are highlighted in green (see Section 2.5.2).

n	N	P1	P2	$\underline{\text{P3}}$	W4	$\underline{\text{L5}}$	P6	Y7	M8	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	S12	C
n	N	$\underline{\text{P1}}$	P2	P3	$\underline{\text{W4}}$	$\underline{\text{L5}}$	P6	$\underline{\text{Y7}}$	$\underline{\text{M8}}$	P9	P10	$\underline{\text{W11}}$	S12	C
n	N	P1	P2	$\underline{\text{P3}}$	$\underline{\text{W4}}$	L5	P6	$\underline{\text{Y7}}$	M8	P9	$\underline{\text{P10}}$	W11	S12	C
l_w	N	P1	P2	$\underline{\text{P3}}$	$\underline{\text{W4}}$	L5	$\underline{\text{P6}}$	Y7	$\underline{\text{M8}}$	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	S12	C
l_w	N	P1	$\underline{\text{P2}}$	$\underline{\text{P3}}$	W4	$\underline{\text{L5}}$	$\underline{\text{P6}}$	Y7	$\underline{\text{M8}}$	P9	P10	$\underline{\text{W11}}$	$\underline{\text{S12}}$	$\underline{\text{C}}$
l_w	N	P1	P2	P3	$\underline{\text{W4}}$	$\underline{\text{L5}}$	P6	Y7	M8	$\underline{\text{P9}}$	$\underline{\text{P10}}$	$\underline{\text{W11}}$	S12	C
l_0.6	N	P1	P2	$\underline{\text{P3}}$	$\underline{\text{W4}}$	L5	$\underline{\text{P6}}$	Y7	$\underline{\text{M8}}$	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	S12	C
l_0.6	N	P1	P2	P3	W4	L5	$\underline{\text{P6}}$	Y7	$\underline{\text{M8}}$	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	S12	C
l_0.6	N	P1	P2	$\underline{\text{P3}}$	W4	$\underline{\text{L5}}$	$\underline{\text{P6}}$	Y7	$\underline{\text{M8}}$	P9	P10	$\underline{\text{W11}}$	S12	$\underline{\text{C}}$
l_0.8	N	P1	$\underline{\text{P2}}$	P3	$\underline{\text{W4}}$	$\underline{\text{L5}}$	P6	Y7	M8	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	S12	$\underline{\text{C}}$
l_0.8	N	P1	P2	$\underline{\text{P3}}$	W4	$\underline{\text{L5}}$	$\underline{\text{P6}}$	$\underline{\text{Y7}}$	$\underline{\text{M8}}$	P9	P10	$\underline{\text{W11}}$	$\underline{\text{S12}}$	C
l_0.8	N	P1	P2	P3	$\underline{\text{W4}}$	L5	$\underline{\text{P6}}$	$\underline{\text{Y7}}$	M8	$\underline{\text{P9}}$	$\underline{\text{P10}}$	$\underline{\text{W11}}$	S12	C
re_1	N	P1	$\underline{\text{P2}}$	P3	$\underline{\text{W4}}$	$\underline{\text{L5}}$	$\underline{\text{P6}}$	$\underline{\text{Y7}}$	$\underline{\text{M8}}$	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	S12	C
re_1	N	P1	P2	P3	W4	L5	$\underline{\text{P6}}$	$\underline{\text{Y7}}$	M8	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	S12	C
re_1	N	P1	P2	P3	W4	$\underline{\text{L5}}$	P6	$\underline{\text{Y7}}$	$\underline{\text{M8}}$	P9	P10	W11	S12	C
re nh	N	P1	P2	P3	$\underline{\text{W4}}$	L5	$\underline{\text{P6}}$	$\underline{\text{Y7}}$	$\underline{\text{M8}}$	P9	P10	$\underline{\text{W11}}$	$\underline{\text{S12}}$	C
re nh	N	P1	$\underline{\text{P2}}$	P3	$\underline{\text{W4}}$	L5	P6	Y7	$\underline{\text{M8}}$	P9	$\underline{\text{P10}}$	$\underline{\text{W11}}$	$\underline{\text{S12}}$	C
re nh	N	P1	$\underline{\text{P2}}$	P3	$\underline{\text{W4}}$	$\underline{\text{L5}}$	$\underline{\text{P6}}$	Y7	M8	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	$\underline{\text{S12}}$	C
re 2 off	N	P1	P2	P3	$\underline{\text{W4}}$	L5	$\underline{\text{P6}}$	$\underline{\text{Y7}}$	$\underline{\text{M8}}$	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	S12	C
re 2 off	N	P1	P2	P3	W4	L5	$\underline{\text{P6}}$	Y7	M8	P9	$\underline{\text{P10}}$	W11	S12	C
re 2 off	N	P1	P2	P3	$\underline{\text{W4}}$	$\underline{\text{L5}}$	$\underline{\text{P6}}$	Y7	M8	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	S12	C
re_l	N	P1	P2	P3	W4	L5	$\underline{\text{P6}}$	$\underline{\text{Y7}}$	$\underline{\text{M8}}$	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	S12	C
re_l	N	P1	$\underline{\text{P2}}$	P3	$\underline{\text{W4}}$	$\underline{\text{L5}}$	P6	$\underline{\text{Y7}}$	M8	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	S12	C
re_l	N	P1	P2	P3	W4	L5	$\underline{\text{P6}}$	$\underline{\text{Y7}}$	M8	P9	$\underline{\text{P10}}$	W11	S12	C
re_l nh	N	P1	$\underline{\text{P2}}$	P3	$\underline{\text{W4}}$	L5	$\underline{\text{P6}}$	Y7	$\underline{\text{M8}}$	P9	P10	$\underline{\text{W11}}$	S12	C
re_l nh	N	P1	P2	$\underline{\text{P3}}$	$\underline{\text{W4}}$	L5	$\underline{\text{P6}}$	$\underline{\text{Y7}}$	M8	P9	P10	$\underline{\text{W11}}$	$\underline{\text{S12}}$	$\underline{\text{C}}$
re_l nh	N	P1	P2	P3	W4	L5	P6	Y7	$\underline{\text{M8}}$	P9	P10	W11	S12	C
re_l 2 off	N	P1	$\underline{\text{P2}}$	P3	$\underline{\text{W4}}$	$\underline{\text{L5}}$	P6	$\underline{\text{Y7}}$	$\underline{\text{M8}}$	$\underline{\text{P9}}$	P10	$\underline{\text{W11}}$	S12	C
re_l 2 off	N	P1	P2	P3	$\underline{\text{W4}}$	L5	$\underline{\text{P6}}$	Y7	$\underline{\text{M8}}$	P9	P10	$\underline{\text{W11}}$	S12	$\underline{\text{C}}$
re_l 2 off	N	P1	P2	P3	W4	L5	$\underline{\text{P6}}$	Y7	M8	P9	$\underline{\text{P10}}$	W11	S12	C

Table C.3: Residues underlined are those which are on average bound to the quartz surface. Results are shown for the three most populated backbone clusters identified by clustering analysis carried out over the last 4x10^6 MD steps of each simulation. (Index: 'n'= normal T-REMD, 'l_w'= light T-REMD, 'l_0.6'= light sc 0.6 T-REMD, 'l_0.8'= light sc 0.8 T-REMD, 're_1'= REST1, 're nh'= REST no h, 're 2 off'= REST 2 off, 're_l'= REST light, 're_l nh'= REST light no h,'re_l 2 off'= REST light 2 off.)

	1	2	3	4
normal T-REMD	4,5 (38.4)	7,8 (27.2)	6,7 (22.6)	10,11 (18.9)
light T-REMD	3,4 (34.5)	5,6 (29.9)	6,7 (29.3)	4,5 (22.1)
light 0.6 sc T-REMD	5,6 (33.3)	6,7 (25.9)	4,5 (24.3)	3,4 (22.9)
light 0.8 sc T-REMD	10,11 (40.5)	4,5 (37.6)	6,7 (25.1)	7,8 (18.2)
REST1	4,5 (32.3)	7,8 (25.2)	10,11 (18.5)	6,7 (16.6)
REST no h	10,11 (36.4)	4,5 (32.8)	5,6 (26.3)	7,8 (23.6)
REST 2 off	6,7 (30.1)	10,11 (12.8)	4,5 (11.1)	5,6 (8.0)
REST light	4,5 (38.8)	6,7 (38.0)	10,11 (8.7)	9,10 (7.4)
REST light no h	10,11 (57.3)	6,7 (51.6)	3,4 (41.4)	7,8 (39.1)
REST light 2 off	4,5 (30.7)	7,8 (22.1)	10,11 (16.6)	6,7 (14.5)

Table C.4: Top 4 binding dyad motifs in the REMD-based simulations, calculated over the last 2x10^6MD steps of each trajectory. Percentage of trajectory for which the dyad makes binding contact is given in parenthesises. Dyads present in S1 which were found to be enriched dyads in quartz binding peptides [Oren et al. [2010]] are highlighted as follows: PW, WL, PP, LP, (given in order of reducing Motif Enrichment Factor).

	1	2	3	4
normal T-REMD	4,5,6 (15.6)	10,11,12 (14.0)	5,6,7 (12.2)	3,4,5 (8.2)
light T-REMD	6,7,8 (13.9)	5,6,7 (13.0)	10,11,12 (11.8)	7,8,9 (8.5)
light 0.6 sc T-REMD	2,3,4 (17.3)	4,5,6 (12.9)	1,2,3 (11.8)	10,11,12 (9.9)
light 0.8 sc T-REMD	10,11,12 (14.6)	6,7,8 (10.8)	9,10,11 (6.5)	4,5,6 (5.4)
REST1	4,5,6 (10.7)	2,3,4 (6.5)	3,4,5 (5.4)	1,2,3 (5.2)
REST no h	10,11,12 (26.4)	4,5,6 (18.9)	2,3,4 (11,4)	6,7,8 (10.7)
REST 2 off	4,5,6 (5.9)	9,10,11 (4.5)	10,11,12 (4.3)	6,7,8 (2.5)
REST light	3,4,5 (14.9)	2,3,4 (8.1)	10,11,12 (5.6)	6,7,8 (2.9)
REST light no h	6,7,8 (40.3)	10,11,12 (33.5)	2,3,4 (9.5)	4,5,6 (4.7)
REST light 2 off	10,11,12 (8.2)	4,5,6 (6.1)	6,7,8 (5.0)	3,4,5 (4.0)

Table C.5: Top 4 binding triad motifs in the REMD-based simulations, calculated over the last 2x10^6MD steps of each trajectory. Percentage of trajectory for which the dyad makes binding contact is given in parentheses. Dyads present in S1 which were found to be enriched dyads in quartz binding peptides [Oren et al. [2010]] are highlighted as follows: PPW, PWL and WLP (given in order of reducing Motif Enrichment Factor).

Appendix D

Additional Information:GolP-CHARMM: A First-Principles Based Force-field for modelling protein adsorption to Au(111), Au(100)(1×1) and Au(100)(5×1).

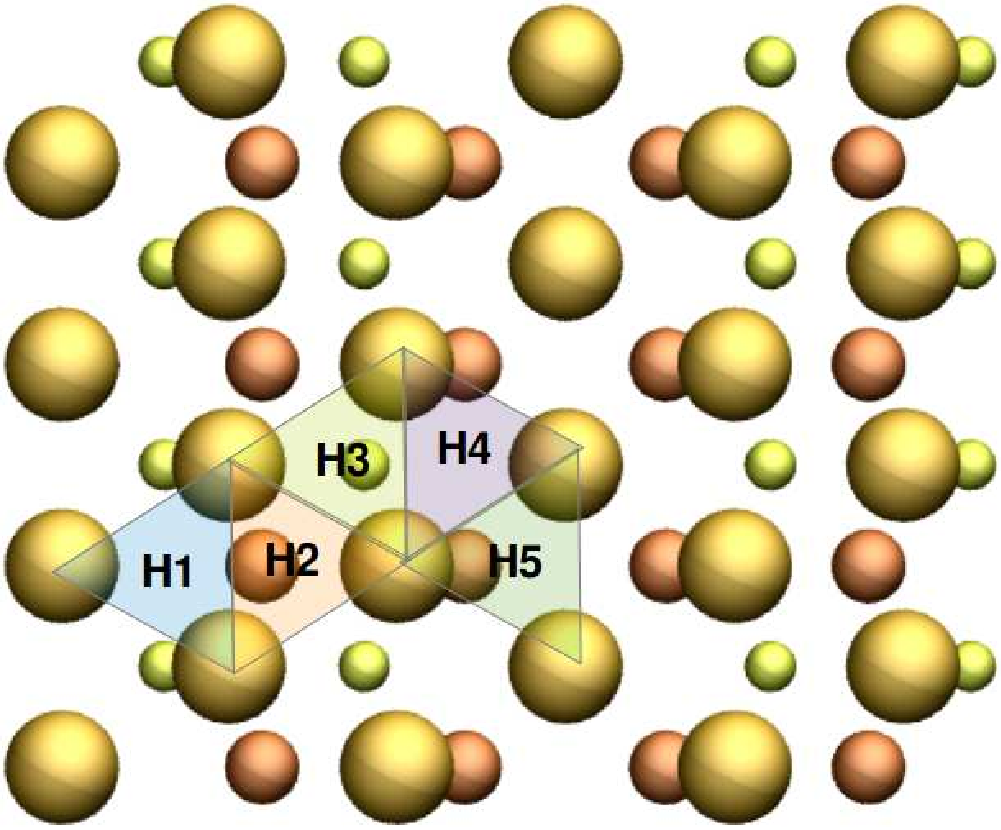


Figure D.1: Hollow sites surrounding top sites T1 and T2 tested in methane–Au(100)(5×1) PW-DFT calculations. Surface layer Au atoms are depicted as largest and in medium gold, 2nd layer Au atoms dark gold and medium size, while the 3rd layer of Au atoms are the smallest size and light gold in color.

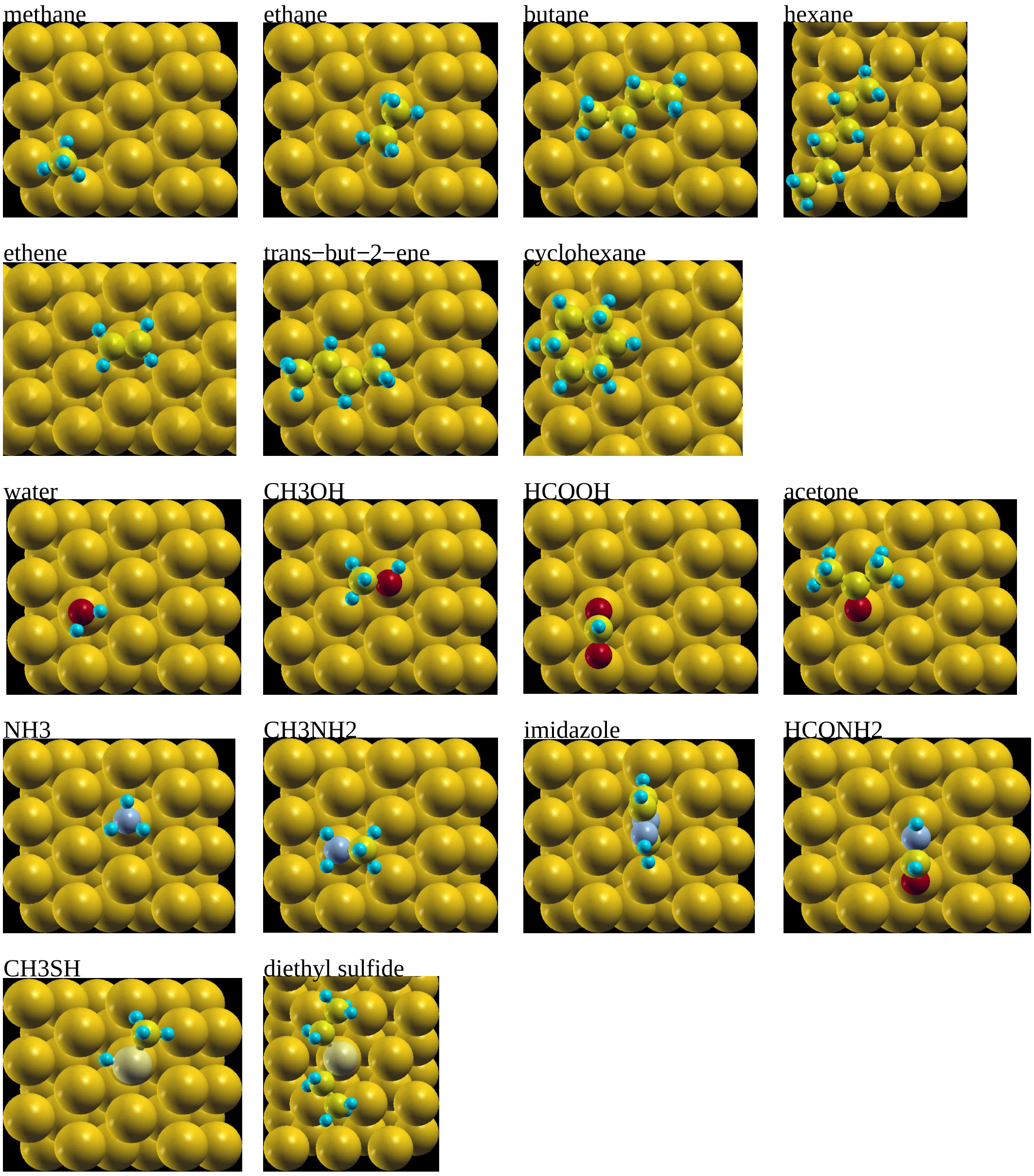

methane
ethane
butane
hexane
ethene
trans−but−2−ene
cyclohexane
water
CH3OH
HCOOH
acetone
NH3
CH3NH2
imidazole
HCONH2
CH3SH
diethyl sulfide

Figure D.2: Optimal geometries of a number of small molecules adsorbed onto Au(111) *in vacuo* obtained from PW-DFT calculations using the vdW-DF exchange functional. Images rendered using XCrySDen. Gold atoms gold, hydrogen cyan, oxygen red, nitrogen blue, carbon yellow and sulphur cream.

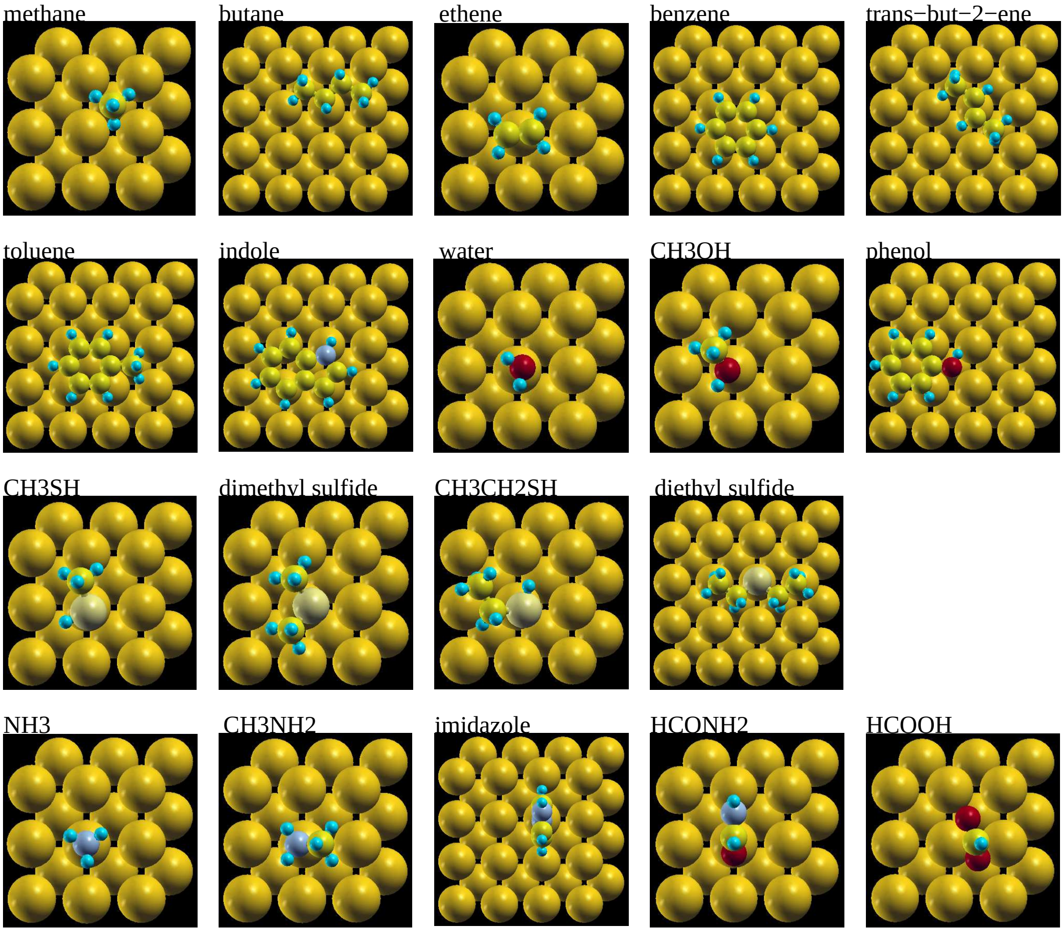

methane
butane
ethene
benzene
trans−but−2−ene
toluene
indole
water
CH3OH
phenol
CH3SH
dimethyl sulfide
CH3CH2SH
diethyl sulfide
NH3
CH3NH2
imidazole
HCONH2
HCOOH

Figure D.3: Optimal geometries of a number of small molecules adsorbed onto Au(100)(1×1) *in vacuo* obtained from PW-DFT calculations using the vdW-DF exchange functional. Images rendered using XCrySDen. Gold atoms gold, hydrogen cyan, oxygen red, nitrogen blue, carbon yellow and sulphur cream.

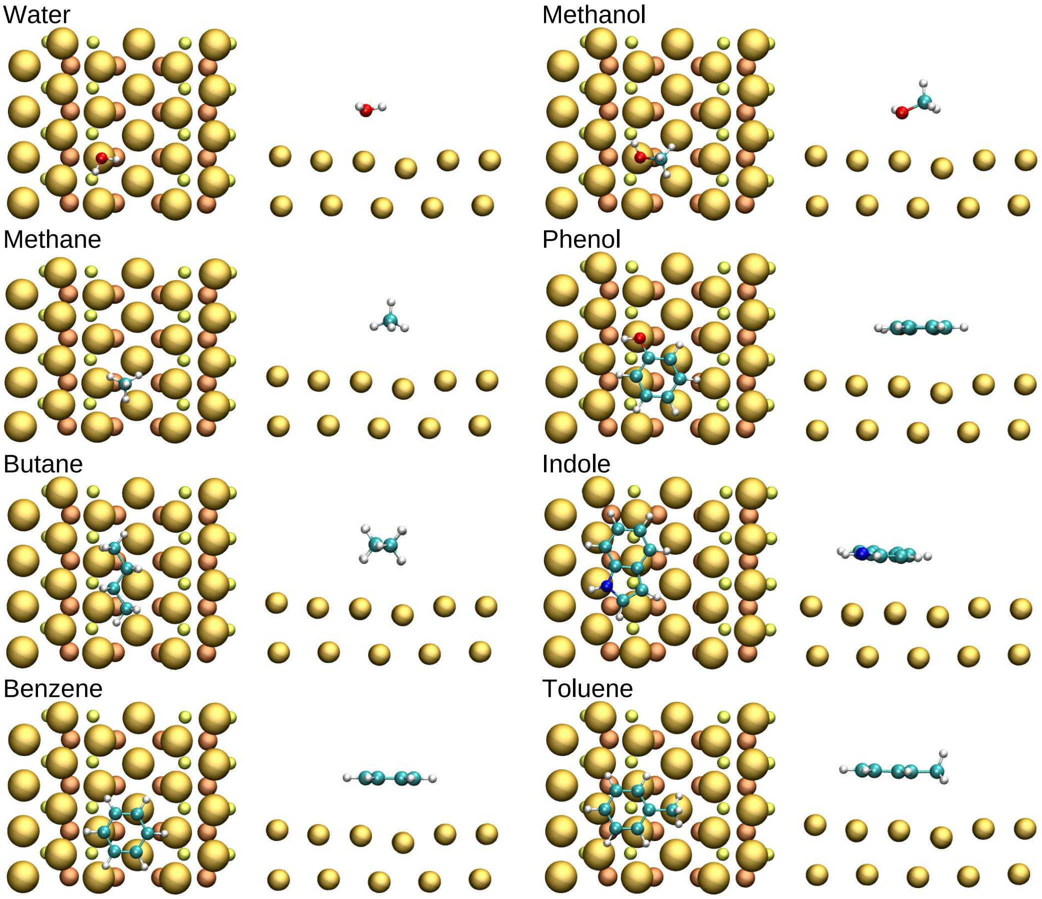

Water
Methanol
Methane
Phenol
Butane
Indole
Benzene
Toluene

Figure D.4: Most favorable geometry of each small molecule adsorbed onto the Au(100)(5×1) surface found from PW-DFT calculations using the vdW-DF functional. Left depicts a view from above the surface whilst right the system is shown from the side. Gold atoms are shown in gold (for the plan view only: surface layer are largest in medium gold, 2nd layer are dark gold and 3rd layer are smallest and light gold), carbon in green, hydrogen in white, oxygen in red, nitrogen in blue and sulfur in yellow.

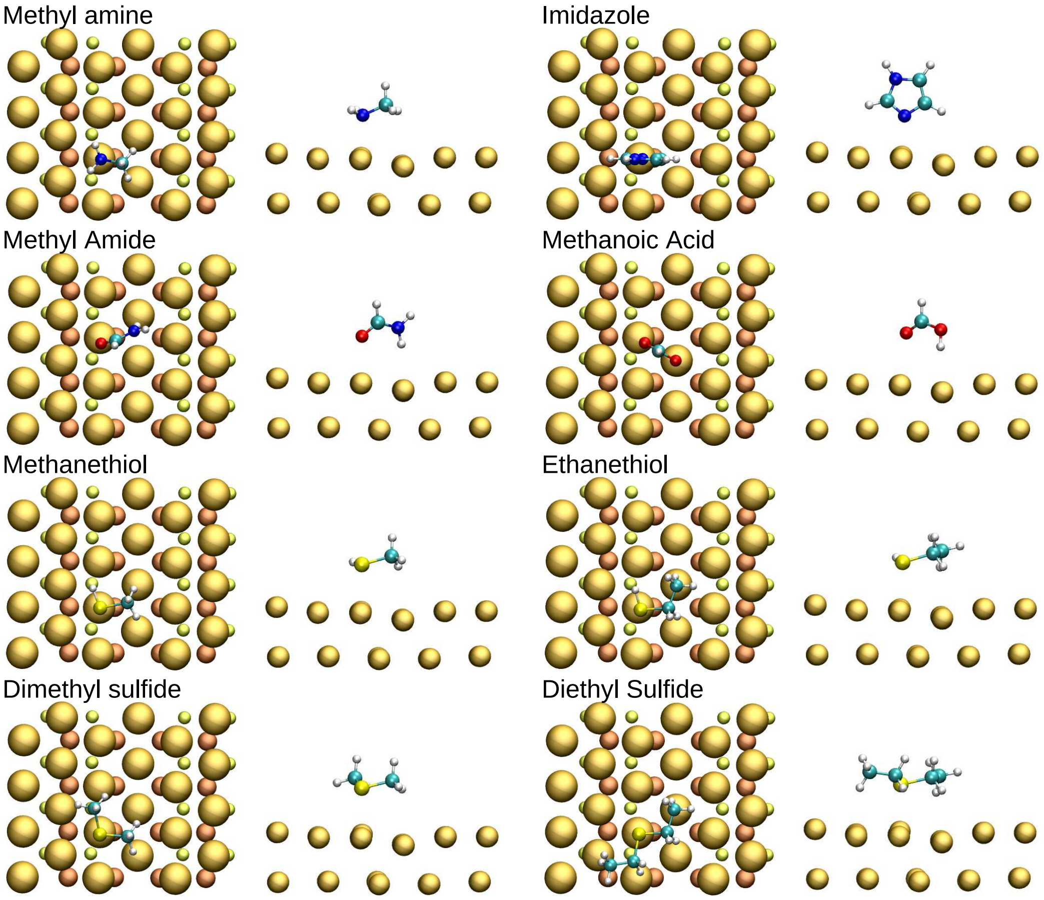

Methyl amine
Imidazole
Methyl Amide
Methanoic Acid
Methanethiol
Ethanethiol
Dimethyl sulfide
Diethyl Sulfide

Figure D.5: Most favorable geometry of each small molecule adsorbed onto the Au(100)(5×1) surface found from PW-DFT calculations using the vdW-DF functional. Left depicts a view from above the surface whilst right the system is shown from the side. Gold atoms are shown in gold (for the plan view only: surface layer are largest in medium gold, 2nd layer are dark gold and 3rd layer are smallest and light gold), carbon in green, hydrogen in white, oxygen in red, nitrogen in blue and sulfur in yellow.

	$(5\times20)^a$	$(5\times1)^b$	$(7\times1)^b$	$(5\times1)^\dagger$	$(5\times1)^\star$	Exp^c
b_1/d_{bulk}	0.32 (0.32)	0.384	0.405	0.303	0.307	0.275
b_2/d_{bulk}	0.044 (0.044)	n/a	n/a	0.066	0.065	0.069

Table D.1: Buckling amplitudes b_1 and b_2 relative to bulk d_{bulk}. Values given in parentheses from Havu *et al.* [Havu et al. [2010]] are those obtained using the PBE functional, those not in parentheses were obtained using LDA only. Data to calculate b_2 from Jacob [Jacob [2007]] is not available. ^{a} reference Havu et al. [2010]; ^{b} reference Jacob [2007]; ^{c} reference Ocko et al. [1991]; † are values calculated in this work using a k-point mesh of 2×2×1 and * are values calculated in this work using a k-point mesh of 4×4×1.

	$(5\times20)^a$	$(5\times1)^b$	$(5\times1)^c$	$(7\times1)^c$	$(5\times1)^\dagger$	$(5\times1)^\star$	Exp^d
d_{12}/d_{bulk}	1.21 (1.22)	1.20	1.19	1.16	1.19	1.20	1.20
d_{23}/d_{bulk}	0.99 (0.98)	0.992	0.988	0.990	0.998	0.998	n/a

Table D.2: Average interlayer distances d_{12} and d_{23} relative to bulk d_{bulk}. Values given in parentheses from Havu *et al.* [Havu et al. [2010]] are those obtained using the PBE functional, those not in parentheses were obtained using LDA only. [a] reference Havu et al. [2010]; [b] reference Shi and Stampfl [2008]; [c] reference Jacob [2007]; [d] reference Ocko et al. [1991]; † are values calculated in this work using a k-point mesh of 2×2×1 and ⋆ are values calculated in this work using a k-point mesh of 4×4×1.

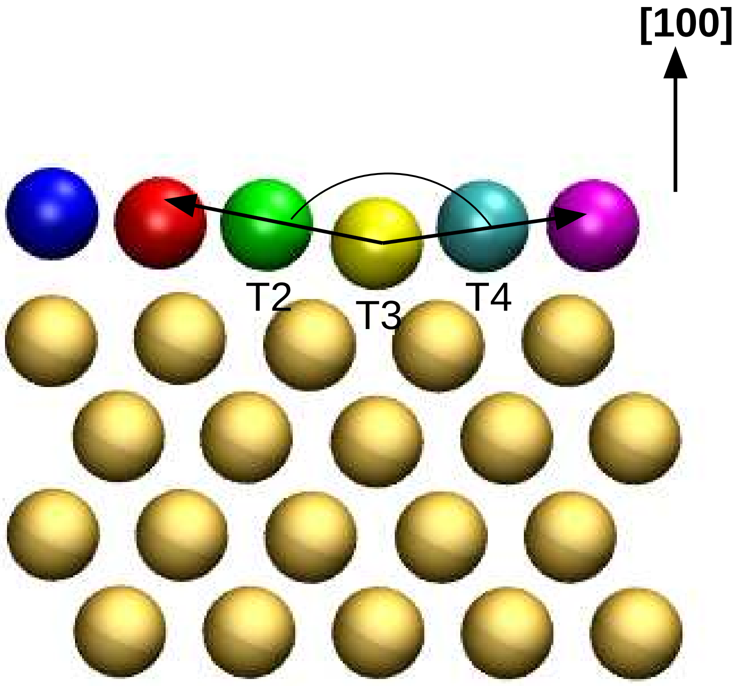


Figure D.6: Side view of the gold slab highlighting the Au-Au-Au angle about the T3 top site.

Adsorbate	H3		H5	
	E^{vdW}	d^{vdW}	E^{vdW}	d^{vdW}
	kJ mol^{-1}	Å	kJ mol^{-1}	Å
methane	−16.6	3.74	−18.0	3.83
butane	−43.5	3.90	−45.5	3.72
benzene	−57.8	3.45	−59.0	3.27
toluene	−69.6	3.45	−68.2	3.29

Adsorbate	T1		T2	
	E^{vdW}	d^{vdW}	E^{vdW}	d^{vdW}
	kJ mol^{-1}	Å	kJ mol^{-1}	Å
indole	−90.4	3.29	−89.7	3.38
phenol	−67.4	3.37	−67.6	3.40
water	−17.8	3.10	−20.5	2.97
methanol	−29.4	3.04	−34.1	2.83

Adsorbate	T1		T2		T23	
	E^{vdW}	d^{vdW}	E^{vdW}	d^{vdW}	E^{vdW}	d^{vdW}
	kJ mol^{-1}	Å	kJ mol^{-1}	Å	kJ mol^{-1}	Å
methanoic acid	−29.6	3.23	−30.4	3.02	−32.3	3.05
methyl amide	−34.9	3.02	−38.0	2.73	−40.8	2.74

Adsorbate	T0		T1		T2		T3	
	E^{vdW}	d^{vdW}	E^{vdW}	d^{vdW}	E^{vdW}	d^{vdW}	E^{vdW}	d^{vdW}
	kJ mol^{-1}	Å	kJ mol^{-1}	Å	kJ mol^{-1}	Å	kJ mol^{-1}	Å
methanethiol	−49.7	2.74	−41.3	3.02	−52.8	2.74	−39.8	3.19
ethanthiol	−54.1	2.72	−43.0	3.75	−57.7	2.72	−44.3	3.25
dimethyl sulfide	−67.5	2.67	−57.8	2.85	−71.4	2.70	−55.5	3.00
diethyl sulfide	−82.6	2.67	−71.1	3.00	−85.5	2.73	−68.7	3.13
methyl amine	−60.1	2.44	−48.7	2.62	−62.3	2.46	−45.0	2.70
imidazole	−62.9	2.34	−49.5	2.53	−65.5	2.35	−46.5	2.66

Table D.3: Interaction energies and equilibrium separations of all the small molecules used to fit and test FF parameters for the Au(100)(5×1) surface found from PW-DFT calculations using the vdW-DF functional. ‘Tx’ and ‘Hx’ refer to the initial location of the molecule on the surface; ‘T23’ denotes a geometry in which the two heteroatoms in methanoic acid and methyl amide are adsorbed atop different sites (see Section 6.2).

	AUI		AUB	
	ϵ kJ mol^{-1}	σ Å	ϵ kJ mol^{-1}	σ Å
Au–Au	0.48	3.80	0.48	3.80
Au–Au (π-C)	1.30	3.20	0.48	3.80
Au–N (imidazole)	1.60	2.85	1.60	2.85
Au–N (unprotonated N termini)	0.90	2.90	0.90	2.90
Au–S	3.20	2.85	3.20	2.85
Au–O	0.70	3.10	0.70	3.10
Au–H	0.28	2.70	0.28	2.70

Table D.4: GolP-CHARMM LJ parameters for the Au(111) surface. Interaction strengths ϵ (kJ mol^{-1}) and distances σ (Å) are given for both surface virtual sites (AUI) and bulk gold atoms (AUB). Note the Au–H interaction is for hydroxyl and amide H only.

	AUI		AUB	
	ϵ kJ mol^{-1}	σ Å	ϵ kJ mol^{-1}	σ Å
Au–Au	2.10	4.00	2.10	4.00
Au–Au (aromatic C)	2.80	3.40	2.10	4.00
Au–N (imidazole)	4.40	3.00	4.40	3.00
Au–N (unprotonated N termini)	5.40	3.00	5.40	3.00
Au–S (sulfide)	6.90	3.00	6.90	3.00
Au–S (thiol)	5.20	3.00	5.20	3.00
Au–O(=C) (carbonyl)	0.60	3.35	0.60	3.35
Au–O(-H) (water/methanol)	1.50	3.275	1.50	3.275
Au–H	0.60	2.725	0.60	2.725

Table D.5: GolP-CHARMM LJ parameters for the Au(100)(1×1) surface. Interaction strengths ϵ (kJ mol^{-1}) and distances σ (Å) are given for both surface virtual sites (AUI) and bulk gold atoms (AUB). Note the Au–H interaction is for hydroxyl, amide and thiol H only.

	σ kJ mol^{-1}	ϵ Å
AUIU AUIU	3.50	0.575
AUIU AUIU (aromatic C)	3.4	0.70
AUIU O	3.15	0.800
AUIU H(-O/-S)	2.70	0.28
AUIU NHT2	2.90	0.70
AUIU NR2	2.85	0.30
S0 NR2	2.50	7.60
S1 NR2	3.25	0.25
S2 NR2	2.50	9.00
S3 NR2	3.50	0.25
S0 NHT2	2.50	8.00
S1 NHT2	2.70	1.50
S2 NHT2	2.50	10.75
S3 NHT2	3.53	0.25
AUIU S	2.85	1.20
AUIU ST	2.85	1.20
S0 ST	2.60	7.50
S1 ST	3.00	1.00
S2 ST	2.60	9.00
S3 ST	3.75	0.40
S0 S	2.60	12.20
S1 S	3.15	0.525
S2 S	2.55	12.175
S3 S	3.8265	0.25

Table D.6: GolP-CHARMM LJ parameters for the Au(100)(5×1) surface. Interaction strengths ϵ (kJ mol^{-1}) and distances σ (Å) are given for both hollow (AUIU) and bridge (S0, S1, S2, S3) surface interaction sites. 'NR2' corresponds to an unprotonated N in imidazole, 'NHT2' to N in an unprotonated N-terminus or lysine side-chain, 'S' to S atoms in dialkyl sulfide moieties, 'ST' to S atoms in thiols. Note the Au–H interaction is for hydroxyl, thiol and amide H only. All cross terms involving bridge interaction sites that have not been explicitly given above are zero.

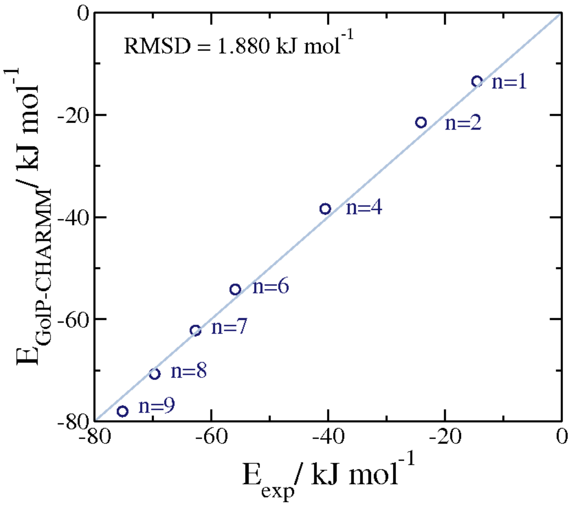


Figure D.7: Correlation between experimentally-determined interaction energies of alkanes of varying chain length (C_nH_{2n+2}) [Wetterer et al. [1998]] with Au(111), and those calculated using GolP-CHARMM parameters for gold.

molecule	E_{ref} kJ mol^{-1}	$E_{GolP-CHARMM}$ kJ mol^{-1}	d_{vdW-DF} Å	$d_{GolP-CHARMM}$ Å
methane	-16.5^{a}, -14.5^{b}	-13.5	3.77	3.53
ethane	-24.7^{a}, -24.1^{b}	-21.5	3.84	3.60
butane	-42.3^{a}, -40.5^{b}	-38.4	3.90	3.68
hexane	-60.3^{a}, -55.9^{b}	-54.2	3.88	3.70
heptane	-62.7^{b}	-62.2	n/a	3.71
octane	-69.7^{b}	-70.7	n/a	3.72
nonane	-75.2^{b}	-78.0	n/a	3.73
ethene	-27.7^{a}, -27.0^{b}	-26.3	3.35	3.27
1,3 butadiene	-46.2^{b}	-46.7	n/a	3.27
benzene	-57.9^{b}, -61.5^{c}	-60.8	n/a	3.25
CH_3SH	-45.3^{a}, -48^{d}, -58^{e}	-46.5	2.88	2.62
diethyl sulfide	-76.6^{a}, -68^{b}	-70.0	3.00	2.64
methylamine	-54.8^{a}	-54.8	2.57	2.26
imidazole	-54.1^{a}	-54.5	2.51	2.32
water	-18.3^{a}	-18.1	3.03	2.54

Table D.7: Interaction energies and equilibrium separations of set of molecules used to fit GolP-CHARMM parameters for the Au(111) surface. Energies in kJ mol^{-1} and distances in Å. ^{a} PW-DFT energy calculated from this work; ^{b} Experimental interaction energy taken from reference [Wetterer et al. [1998]]; ^{c} Experimental interaction energy taken from reference [Syomin et al. [2001]]; ^{d} Experimental interaction energy taken from reference [Lavrich et al. [1998]]; ^{e} Experimental interaction energy taken from reference [Nuzzo et al. [1987]].

molecule	E_{exp} kJ mol^{-1}	$E_{GolP-CHARMM}$ kJ mol^{-1}
acetone	-41^{a}	-35.69
cyclohexane	-50.6^{b}, 50^{c}	-51.9
cyclohexene	-54^{c}	-51.6
non-1-ene	-76.2^{b}	-72.6
toluene	-66.1^{b}	-70.9
trans-but-2-ene	-41.7^{b}	-42.9
ethanethiol	-57^{b}	-55.4
dibutyl sulfide	-86^{b}	-93.9

Table D.8: Comparison between experimental interaction energies (kJ mol^{-1}) and those obtained using GolP-CHARMM parameters for the Au(111) surface. ^{a} Experimental interaction energy taken from reference [Syomin and Koel [2002a]] ; ^{b} Experimental interaction energy taken from reference [Wetterer et al. [1998]]; ^{c} Experimental interaction energy taken from reference [Syomin and Koel [2002b]]. RMSD between GolP-CHARMM and experiment was 4.15 kJ mol^{-1} using the above 'test set' whereas the RMSD between GolP and experiment was 4.97 kJ mol^{-1} [Iori et al. [2009]] using a 'test set' of molecules that comprised of cyclohexane, cyclohexene, cycloctane, toluene, non-1-ene, acetone, trans-but-2-ene, diethyl sulfide and undec-1-ene.

molecule	E_{vdW-DF} kJ mol^{-1}	$E_{GolP-CHARMM}$ kJ mol^{-1}
CH_3OH	−30.5	−24.99
HCOOH	−29.2	−29.43
$CHONH_2$	−34.2	−35.62

Table D.9: Interaction energies for test set of molecules adsorbed onto the Au(111) surface. Energies in kJ mol^{-1}.

molecule	E_{ref} kJ mol^{-1}	$E_{GolP-CHARMM}$ kJ mol^{-1}	d_{vdW-DF} Å	$d_{GolP-CHARMM}$ Å
methane	−16.3	−16.9	3.69	3.47
butane	−43.2	−43.3	3.72	3.66
benzene	−57.9	−58.4	3.39	3.26
CH_3SH	−51.8	−51.6	2.78	2.51
CH_3SCH_3	−69.1	−69.2	2.87	2.56
methylamine	−61.3	−60.8	2.48	2.29
imidazole	−67.2	−67.0	2.43	2.25
water	−20.8	−20.9	2.90	2.50
$HCONH_2$	−34.9	−34.7	2.72	2.43

Table D.10: Interaction energies and equilibrium separations of the set of molecules used to fit GolP-CHARMM parameters for the Au(100)(1×1) surface. Energies in kJ mol^{-1} and distances in Å.

molecule	E_{vdW-DF} $\mathrm{kJ\,mol^{-1}}$	$E_{GolP-CHARMM}$ $\mathrm{kJ\,mol^{-1}}$
toluene	−68.8	−70.3
indole	−87.9	−82.3
CH_3OH	−31.7	−28.7
phenol	−66.3	−69.0
HCOOH	−29.5	−30.4
diethyl-sulfide	−85.3	−86.3
ethanethiol	−55.1	−61.5

Table D.11: Interaction energies for the test set of molecules adsorbed onto the Au(100)(1×1) surface. Energies in $\mathrm{kJ\,mol^{-1}}$.

	E_{vdW-DF} kJ mol^{-1}	$E_{GolP-CHARMM}$ kJ mol^{-1}	d_{vdW-DF} Å	$d_{GolP-CHARMM}$ Å
methane	−18.0	−17.5	3.83	3.54
butane	−45.5	−45.4	3.72	3.72
benzene	−59.0	−58.5	3.27	3.22
toluene	−69.6	−70.5	3.45	3.23
indole	−90.4	−84.6	3.29	3.11
water	−20.5	−21.8	2.97	2.55
methanol	−34.1	−30.5	2.83	2.60
methanoic acid	−32.3	−36.5	3.05	2.60
methyl amide	−40.8	−38.1	2.74	2.57
phenol	−67.6	−69.2	3.40	3.14
methyl amine	−62.3	−61.7	2.46	2.21
imidazole	−65.5	−66.0	2.35	2.18
methanethiol	−52.8	−51.5	2.74	2.50
dimethyl sulfide	−71.4	−71.4	2.70	2.49
ethanethiol	−57.7	−64.6	2.72	2.46
diethyl sulfide	−85.5	−91.2	2.73	2.46

Table D.12: Optimal interaction energies (kJ mol^{-1}) and equilibrium closest surface separations (Å) for all small molecules investigated adsorbed onto Au(100)(5×1), calculated using vdW-DF and optimal FF parameters given above.

	E_{vdW-DF} kJ mol^{-1}	$E_{GolP-CHARMM}$ kJ mol^{-1}	d_{vdW-DF} Å	$d_{GolP-CHARMM}$ Å
methyl amine T0	−60.1	−60.7	2.44	2.22
methyl amine T1	−48.7	−46.9	2.62	2.31
methyl amine T2	−62.3	−59.3	2.46	2.25
methyl amine T3	−45.0	−41.2	2.70	2.67
imidazole T0	−62.9	−65.0	2.34	2.16
imidazole T1	−49.5	−50.0	2.53	2.33
imidazole T2	−65.5	−65.6	2.35	2.19
imidazole T3	−46.5	−46.6	2.66	2.59

Table D.13: Interaction energies (kJ mol^{-1}) and equilibrium closest surface separations (Å) for molecules containing an unprotonated N atom adsorbed onto Au(100)(5×1) *in vacuo*, calculated using vdW-DF and optimal FF parameters given above. Labels 'Tx' refer to the top site location 0-3 on the Au(100)(5×1) surface.

	E_{vdW-DF} kJ mol^{-1}	$E_{GolP-CHARMM}$ kJ mol^{-1}	d_{vdW-DF} Å	$d_{GolP-CHARMM}$ Å
methanethiol T0	−49.7	−52.1	2.74	2.46
methanethiol T1	−41.3	−39.7	3.02	2.62
methanethiol T2	−52.8	−53.6	2.74	2.46
methanethiol T3	−39.8	−40.8	3.19	3.08
ethanethiol T0	−54.1	−54.0	2.72	2.45
ethanethiol T1	−43.0	−52.3	3.75	2.59
ethanethiol T2	−57.7	−64.4	2.72	2.48
ethanethiol T3	−44.3	−41.8	3.25	3.04
dimethyl sulfide T0	−67.5	−72.0	2.67	2.52
dimethyl sulfide T1	−57.8	−60.2	2.85	2.65
dimethyl sulfide T2	−71.4	−71.4	2.70	2.49
dimethyl sulfide T3	−55.5	−60.7	3.00	3.00
diethyl sulfide T0	−82.6	−78.6	2.67	2.54
diethyl sulfide T1	−71.1	−68.4	3.00	2.70
diethyl sulfide T2	−85.5	−86.3	2.73	2.51
diethyl sulfide T3	−68.7	−67.4	3.13	3.15

Table D.14: Interaction energies (kJ mol^{-1}) and equilibrium closest surface separations (Å) for sulfur containing molecules adsorbed onto Au(100)(5×1) *in vacuo*, calculated using vdW-DF and optimal FF parameters given above. Labels 'Tx' refer to the top site location 0-3 on the Au(100)(5×1) surface.

Appendix E

Additional Information:Material Selectivity of QBP-1 Adsorption.

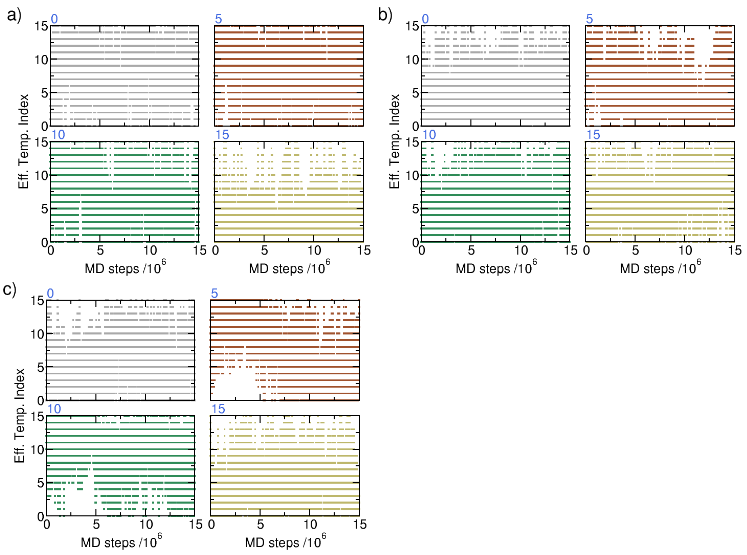


Figure E.1: Trajectories of four representative replicas (0, 5, 10, 15) through effecive temperature space in simulations a) ‘sol’, b) ‘quartz’ and c) ‘gold’.

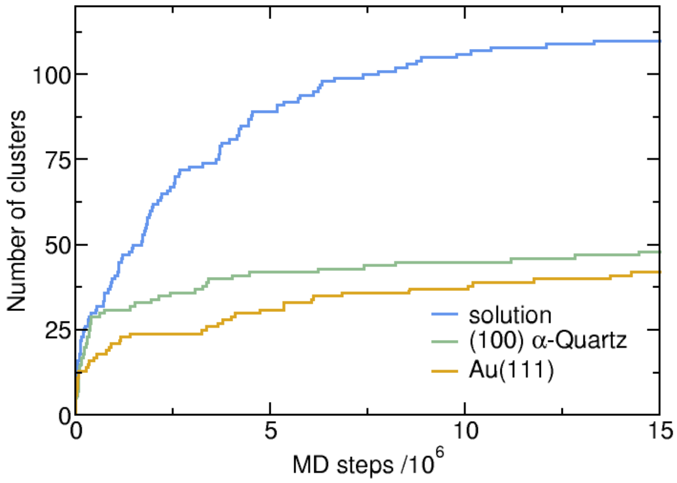


Figure E.2: Number of clusters of QBP-1 structure as a function of simulation time.

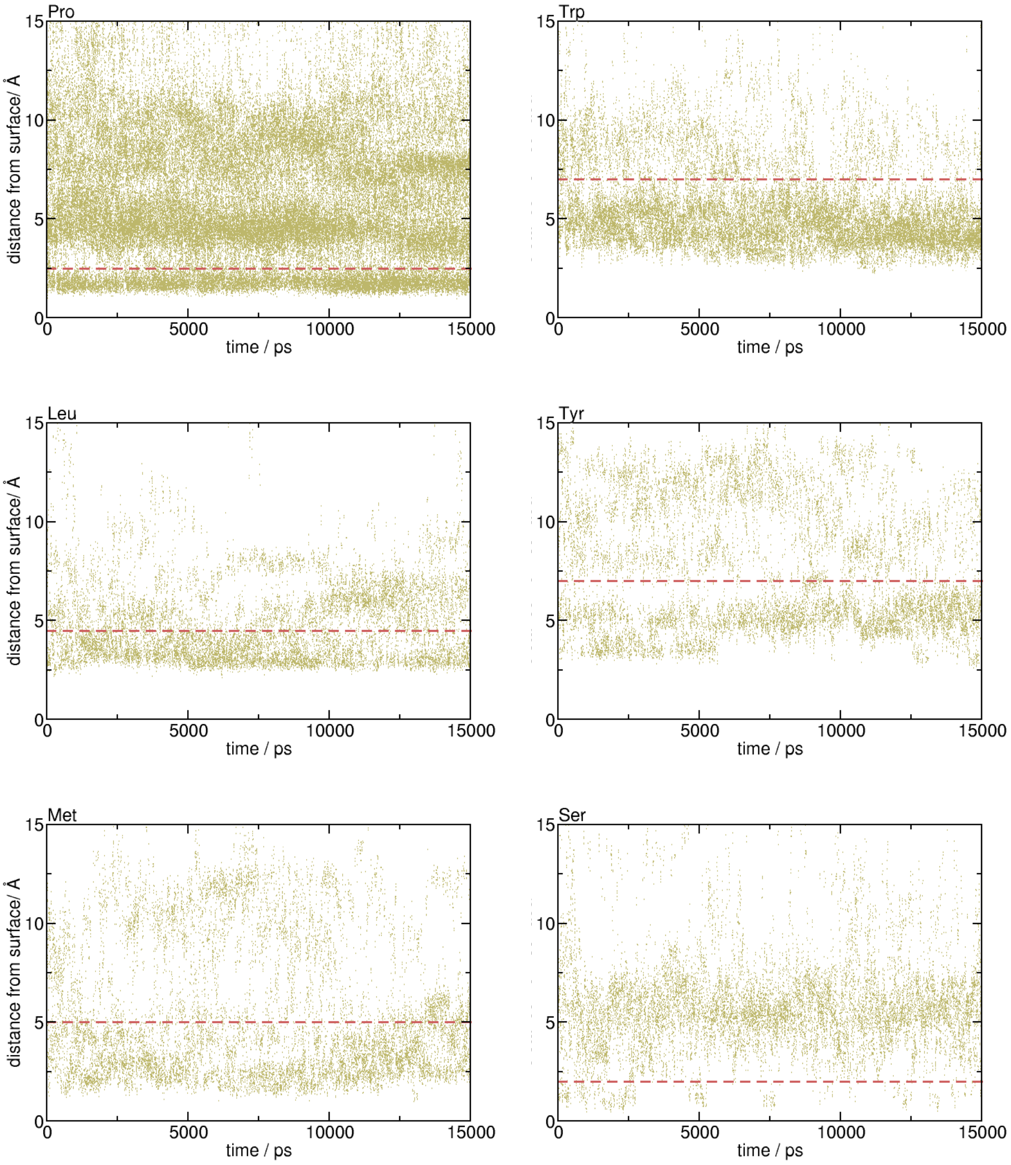


Figure E.3: Distribution of functional group—gold distances used to define residue specific adsorption cut-offs at the aqueous (100) α-Quartz interface.

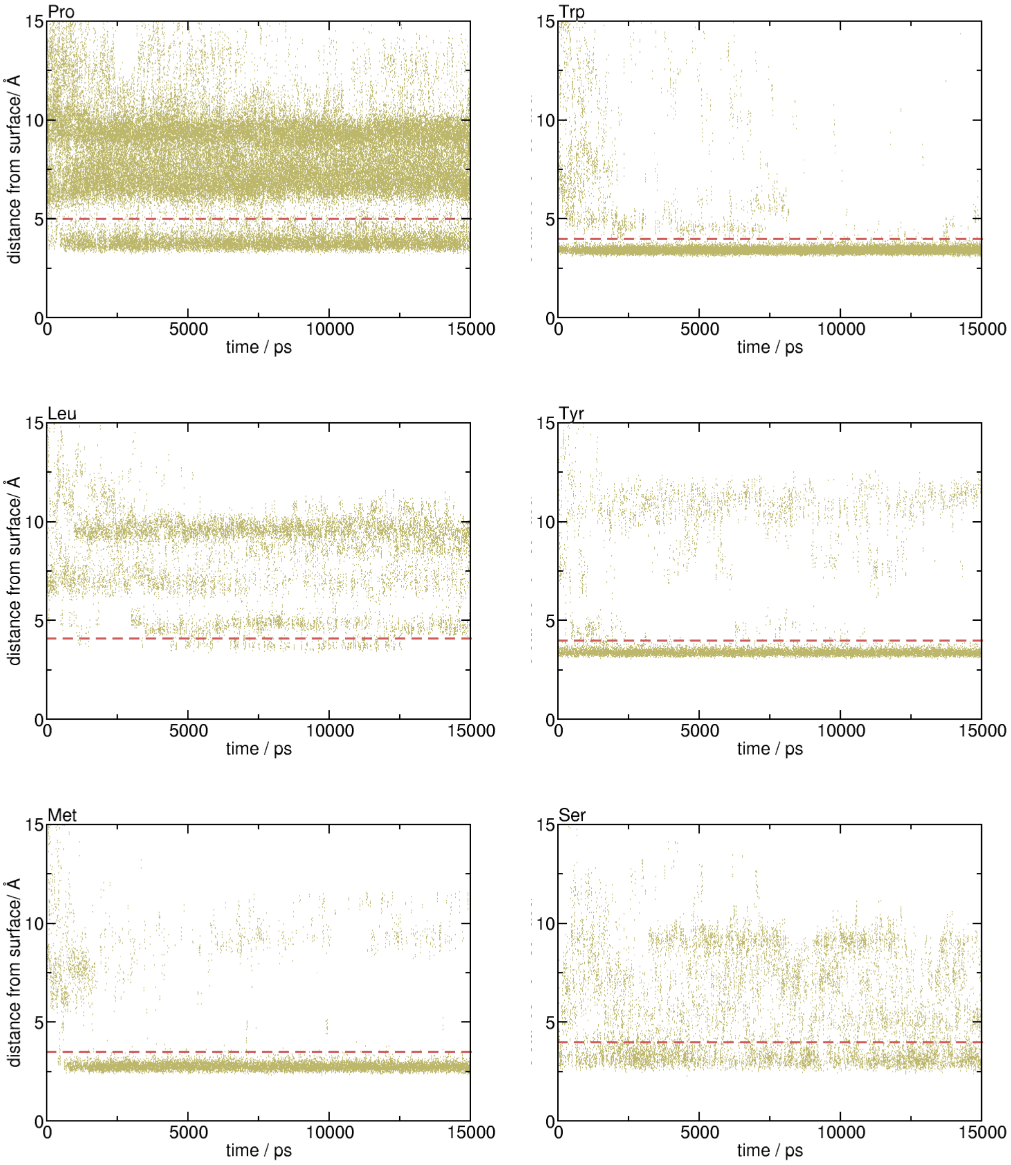


Figure E.4: Distribution of functional group—gold distances used to define residue specific adsorption cut-offs at the aqueous Au(111) interface.

	func. group	Quartz Å	Au(111) Å
Pro	C_γ	2.5	5.0
Trp	$C_{\delta 2}$ and $C_{\epsilon 2}$	7.0	4.0
Leu	C_γ	4.5	4.1
Tyr	phenyl ring	7.0	4.0
Met	S	5.0	3.5
Ser	O(-H)	2.0	4.0

Table E.1: Atoms or groups of atoms used to calculate residue–surface separations and the cut-off distances defining the adsorbed state of the residue at each interface. (In the case of more than one atom being used to define a functional group, the binding distance cut-off refers to the surface-separation of the centre of mass of the atoms involved.)

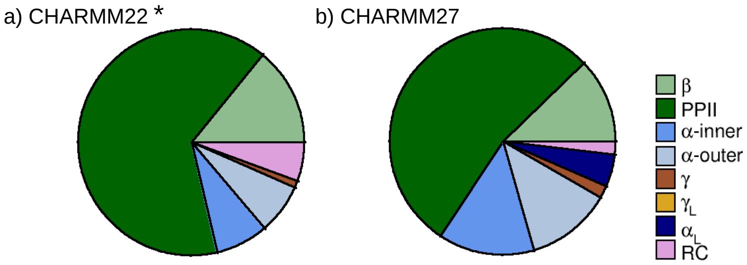


Figure E.5: Pie charts depicting the frequency with which each region of ϕ/ψ space was sampled in REST simulations of QBP-1 adsorbed at the aqueous (100) α-Quartz interface modelled using a) CHARMM22* and b) CHARMM27.

	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12
quartz	0.6	29.3	15.1	92.2	37.8	40.0	67.0	69.2	11.8	24.9	84.4	2.6
gold	43.0	12.6	25.2	97.6	4.1	15.0	84.0	93.7	15.4	0.6	99.8	44.7

Table E.2: Percentage of frames in the last 5 ns of the trajectory in which each residue is adsorbed to the quartz/gold surface. The four residues with the greatest degree of interfacial contact have been highlighted in red.

	1,2	2,3	3,4	4,5	5,6	6,7	7,8	8,9	9,10	10,11	11,12
quartz	0.3	0.0	12.6	35.0	2.5	25.8	47.8	6.8	0.0	24.3	2.2
gold	7.9	1.0	24.2	4.1	4.0	0.0	77.8	9.5	0.6	0.5	44.6

Table E.3: Percentage of frames in the last 5 ns of the trajectory in which each diad is adsorbed to the quartz/gold surface. Diads which are statistically enriched in quartz binding peptides, compared to the Ph.D.-12 peptide library as a whole are PW, WL, PP, LP [Oren et al. [2010]]. Motifs found to have significant interfacial contact during the simulations carried out in this work have been highlighted according to their sequence using the same colour scheme or underlined.

	1,2,3	2,3,4	3,4,5	4,5,6	5,6,7	6,7,8	7,8,9	8,9,10	9,10,11	10,11,12
quartz	0.0	0.0	2.9	2.5	0.0	21.0	4.6	0.0	0.0	1.0
gold	0.0	1.0	0.0	4.0	0.0	0.0	0.0	0.0	0.5	0.3

Table E.4: Percentage of frames in the last 5 ns of the trajectory in which each triad is adsorbed to the quartz/gold surface. Triads which are statistically enriched in quartz binding peptides, compared to the Ph.D.-12 peptide library as a whole are PPW, PWL and WLP [Oren et al. [2010]]. Motifs found to have significant interfacial contact during the simulations carried out in this work have been highlighted according to their sequence using the same colour scheme or underlined.

	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12
Quartz	12.2	38.2	65.7	56.2	60.4	73.1	39.8	45.6	37.1	47.7	51.6	55.1
Gold	0.0	0.0	1.8	24.3	0.6	0.7	1.8	0.0	0.1	0.9	23.4	1.5

Table E.5: Percentage of frames in the last 5 ns of the trajectory in which each the backbone of each residue is adsorbed to the quartz/gold surface. Surface-separation cut-off distances of 5.1 Å and 4.4 Å were used to define the adsorbed state of a residues' backbone atoms (defined as the centre of mass of the amide group) at the (100) α-Quartz and Au(111) interfaces respectively.

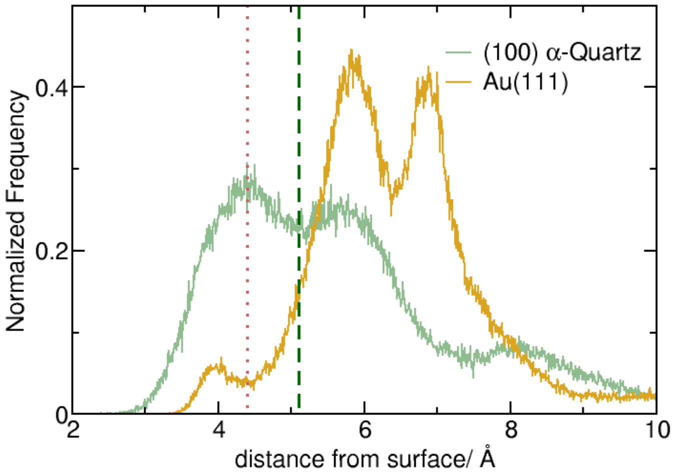


Figure E.6: Normalized frequency distribution of residue backbone atoms above the (100) α-Quartz and Au(111) surfaces. Cut-offs for backbone group surface binding are demarked by dotted lines (dark green- quartz and red- gold).

Appendix F

Additional Information:Facet Selectivity of AuBP-1 adsorption onto Gold: A Replica Exchange with Solute Tempering Metadynamics Study.

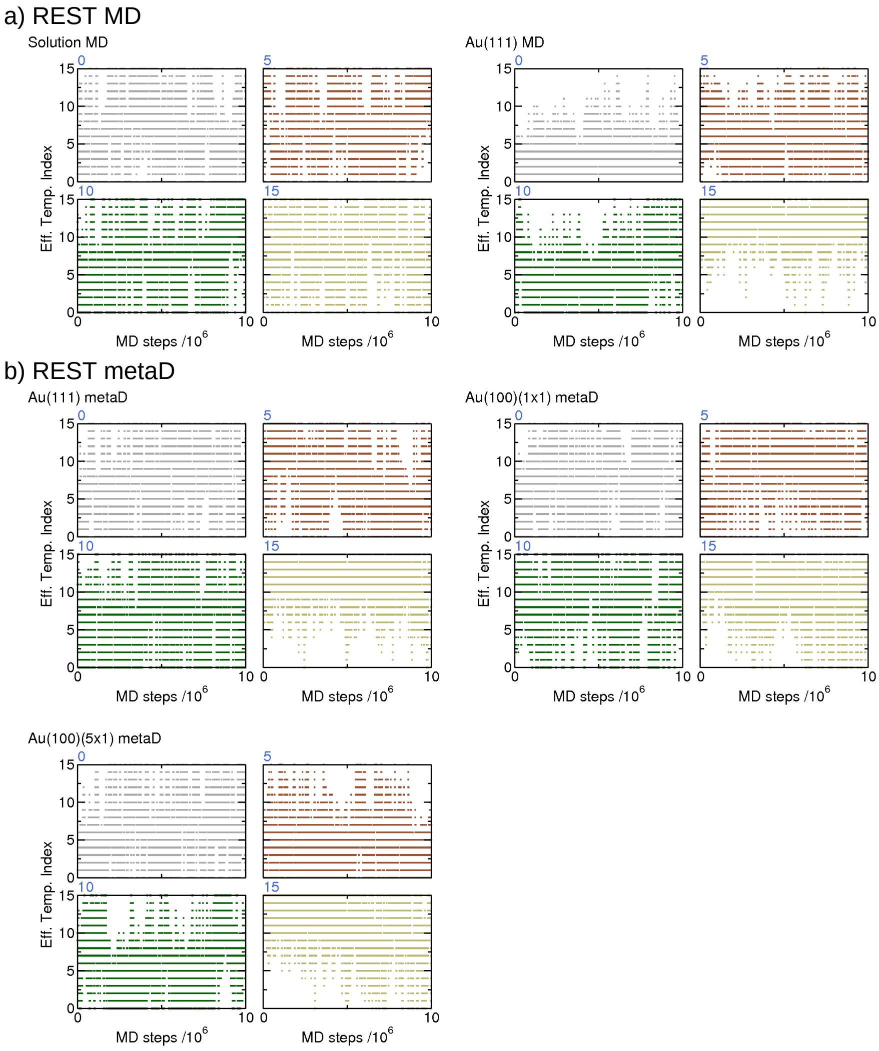


Figure F.1: Trajectories of representative replicas (0, 5, 10, 15) through effective temperature space during a) REST MD simulations (of AuBP-1 in solution and at the aqueous Au(111) interface) and during b) REST metaD simulations (at the aqueous Au(111), Au(100)(1×1) and Au(100)(5×1) interfaces). Data shown for first 10 ns of all simulations only.

	funct. group	$\mathrm{Dir}_{Au(111)}$ / Å	$\mathrm{Dir}_{Au(100)(1\times1)}$ / Å	$\mathrm{Dir}_{Au(100)(5\times1)}$ / Å
ALA	C_β	4.7	4.8	4.6
ARG	guanidium group	4.5	4.6	4.4
GLU	$\mathrm{CO}\mathbf{O}^-$	4.1	4.4	4.2
GLY	$C_\alpha\mathbf{H}_2$	3.1	3.2	3.0
LEU	C_δ	4.7	4.8	4.6
LYS	$-\mathbf{N}\mathrm{H}_3^+$	4.5	4.6	4.4
TRP	phenyl ring	4.4	4.4	4.4
VAL	C_γ	4.7	4.8	4.6

Table F.1: Amino-acid functional group gold distances used as cut-off to define direct gold adsorption at the aqueous Au(111), Au(100)(1×1) and Au(100)(5×1) interfaces. In the case of Arg and Trp, the *c.o.m.* of the heavy atoms in the functional group described above was used in determining residue side-chain surface separation, whereas for Glu, Gly, Leu and Val it was the *closest* carboxylate oxygen atom, hydrogen or methyl carbon, respectively.

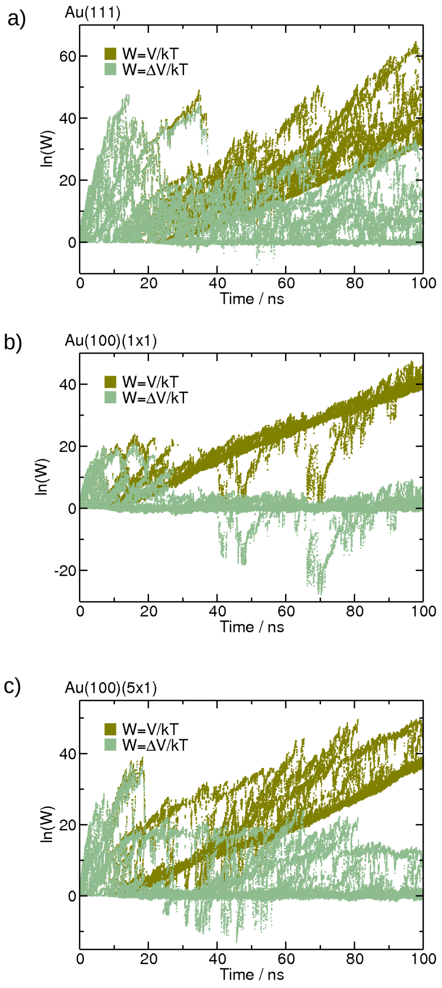


Figure F.2: Evolution of ln(W) (Equation 2.31) as a function of simulation time at the aqueous a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1) interfaces.

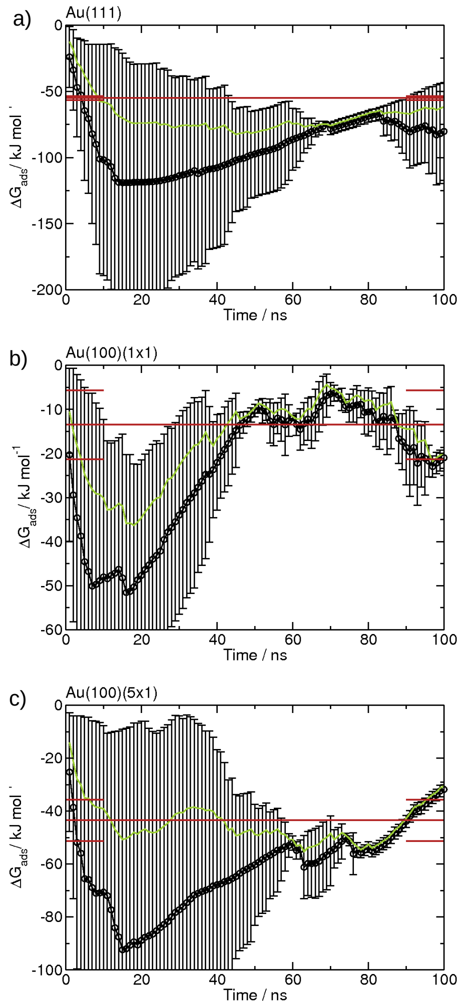


Figure F.3: Time evolution of the free energy of adsorption of AuBP-1 to a) Au(111), b) Au(100)(1×1) and c) Au(100)(5×1). Plotted in black is the minimum energy based on the peptide binding to either the top or bottom surface of the gold slab; error bars represent the energetic differences between the two interfaces. The average energy of adsorption to the two identical gold surfaces present in each simulation cell is demarked by a green line, while the free energy of adsorption, with its associated error, calculated self consistently using the 'time period' reweighting method after five iterations is plotted in brown.

	a %	b %	c %	d %	e %	f %	g %	h %	i %	unass. %
Solution	15.8	12.4	6.3	14.1	14.6	11.2	5.0	3.6	5.5	11.4
Au(111) MD	0.0	2.8	0.0	12.5	26.6	36.1	9.0	6.4	4.5	2.1
Au(111)	0.3	0.2	1.3	11.2	47.2	25.1	1.4	2.9	8.9	1.4
Au(100)(1×1)	1.5	1.2	0.3	16.3	42.0	34.2	0.2	0.0	0.2	4.2
Au(100)(5×1)	3.5	0.0	2.1	11.1	20.9	43.1	1.8	6.2	10.3	1.0

Table F.2: Percentage population of each reference cluster of AuBP-1. The most populated clusters have been highlighted as follows: C1, C2, C3.

	WA %	AG %	GA %	AK %	KR %	RL %	LV %	VL %	LR %	RR %	RE
Au(111)	51.5	8.9	2.3	4.5	14.7	46.0	35.5	39.4	52.5	76.0	13.7
Au(100)(1×1)	10.7	1.9	0.4	0.4	11.0	8.9	17.3	29.7	30.7	40.2	1.9
Au(100)(5×1)	40.3	16.7	20.9	3.5	17.0	55.0	9.3	12.2	86.5	82.2	15.6

Table F.3: Percentage likelihood for each potential binding diad to be in direct contact with the gold surface. The binding characteristics of a dyad have been highlighted following the scheme: strong 76-100%; significant 51-75%; and moderate 26-50%.

	WAG %	AGA %	GAK %	AKR %	KRL %	RLV %	LVL %	VLR %	LRR %	RRE
Au(111)	8.8	1.6	0.3	4.3	7.0	29.5	19.4	38.3	43.3	13.0
Au(100)(1×1)	1.8	0.1	0.0	0.3	0.5	7.8	17.2	29.7	30.3	1.3
Au(100)(5×1)	13.4	10.9	1.0	3.5	6.8	9.1	5.6	12.2	73.1	15.3

Table F.4: Percentage likelihood for each potential binding triad to be in direct contact with the gold surface. The binding characteristics of a triad have been highlighted following the scheme: strong 76-100%; significant 51-75%; and moderate 26-50%.

Bibliography

S. Alaeddine and H. Nygren. The adsorption of water and amino acids onto hydrophobic quartz surfaces. *Colliods and Surf. B*, 6:71, 1996.

M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, Oxford, 1987.

F. F. Amos, C. B. Ponce, and J. S. Evans. Formation of framework nacre polypeptide supramolecular assemblies that nucleate polymorphs. *Biomacromolecules*, 12: 1883, 2011.

V. M. Anisimov, G. Lamoureux, I. V. Vorobyov, N. Huang, B. Roux, and A. D. MavKerellJr. Determination of electrostatic parameters for a polarizable force-field based classical drude oscillator. *J. Chem. Theory Comput.*, 1:153, 2005.

B. M. Axilrod and E. Teller. Interaction of the van der Waals type between three atoms. *J. Chem. Phys.*, 11(6):299, 1943.

A. B. Baber, D. Torres, K. Miller, M. Nazzarro, P. Liu, D. E. Sharr, and D. J. Stacchiola. Reactivity and morphology of O-modified Au surfaces. *J. Phys. Chem. C.*, 116:18292, 2012.

A. J. Ballard and C. Jarzynski. Replica exchange with nonequilibrium switches. *Proc. Natl. Acad. Sci. USA*, 106:12224, 2009.

A. J. Ballard and C. Jarzynski. Replica exchange with nonequilibrium switches: Enhancing equilibrium sampling by increasing replica overlap. *J. Chem. Phys.*, 136:194101, 2012.

A. Barducci, G. Bussi and M. Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100:020603, 2008.

C. H. Bennett. Mass tensor molecular dynamics. *J. Comput. Phys.*, 19:267, 1975.

I. Ben Shir, S. Kabaya, T. Amitay-Rosen, Y. S. Balazs, and A. Schmidt. Molecular level characterization of the inorganic-bioorganic interface by solid state NMR: Alanine on a silica surface, a case study. *J. Phys. Chem. B*, 114(18):5989, 2010.

I. Ben Shir, S. Kababya, and A. Schmidt. Binding specificity of amino acids to amophous silica surfaces: Solid-state NMR of glycine. *J. Phys. Chem. C*, 116: 9691, 2012.

K. C. L. Black, J. Yi, J. G. Rivera, D. C. Zelasko-Leon, and P. B. Messersmith. Polydopamine-enabled surface functionalization of gold nanorods for cancer cell-targetted imaging and photothermal therapy. *Nanomedicine-UK*, 8(1):17, 2013.

M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Rateri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia and M. Parrinello. PLUMED: a portable plugin for free energy calculations with molecular dynamics. *Comp. Phys. Comm.*, 180:1961, 2009a.

M. Bonomi, A. Barducci and M. Parrinello. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.*, 30: 1615, 2009b.

G. Brancolini, D. B. Kokh, R. C. Wade, and S. Corni. Docking of ubiquitin to gold nanoparticles. *ACS Nano*, 69863, 2012.

B. D. Briggs and M. R. Knecht. Nanotechnology meets biology: Peptide-based methods for the fabrication of functional materials. *J. Phys. Chem. Lett.*, 3:405, 2012.

A. Brodka and T. W. Zerda. Properities of liquid acetone in silica pores: Molecular dynamics simulation. *J. Chem. Phys.* , 104(16):6319, 1996.

L. L. Brott, R. R. Naik, D. I. Pikas, S. M. Kirkpatrick, D. W. Tomlin, P. W. Whitlock, S. J. Clarson, and M. O. Stone. Ultrafast holographic nanopatterning of biocatalytically formed silica. *Nature*, 413:291, 2001.

S. Brown. Metal-recognition by polypeptides. *Nat. Biotechnol.*, 15:269, 1997.

E. Brunner, K. Lutz, and M. Sumper. Biomimetic synthesis of silica nanospheres depends on the aggregation and phase separation of polyamines in aqueous solution. *Phys. Chem. Chem. Phys.* , 6:854, 2004.

G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello. Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.*, 128:13435, 2006.

A. Butenuth, G. Moras, J. Schneider, M. Koleini, S. Köppen, R. Meissner, L. B. Wright, T. R. Walsh, and L. Colombi Ciacchi. Ab initio derived force-field parameters for molecular dynamics simulations of deprotonated amorphous SiO_2/water interfaces. *Phys. Status Solidi B*, 249:292, 2012.

L. Calzolai, F. Franchini, D. Gilliland, and F. Rosi. Protein-nanoparticle interaction: identification of the ubiquitin-gold nanoparticle interaction site. *Nano Lett.*, 10: 3101, 2010.

C. Camilloni, D. Provasi, G. Tiana, and R. A. Broglia. Exploring the protein G helix free-energy surface by solute tempering metadynamics. *Proteins*, 71(4):1647, 2008.

L. A. Capriotti, T. P. Beebe (Jr.), and J. P. Schneider. Hydroxyapatite surface-induced peptide folding. *J. Am. Chem. Soc.*, 129(16):5281, 2007.

R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55(22):2471, 1985.

F. Causa, R. della Modlie, E. Laccino, S. Mimmi, D. Marasco, P. L. Scognamiglio, E. Battista, C. Palmieri, C. Cosenza, L. Sanguigno, I. Quinto, G. Scala, and P. A. Netti. Evolutionary screening and adsorption behaviours of engineered M13 bacteriophage and derived dodecapeptide for the selective decoration of gold interfaces. *J. Colloid Interf. Sci.*, 389:220, 2013.

M. Chakraborty, S. Jain, and V. Rani. Nanotachnology: Emerging tool for diagonistics and therapuetics. *Appl. Biochem. Biotechnol.*, 165:1178, 2011.

H. Chen, X. Su, K.-G. Neoh, and W.-S. Choe. QCM-D analysis of binding mechanism of phage particles displaying a constrained heptapetide with specific affintiy to SiO_2 and TiO_2. *Anal. Chem.*, 78:4872, 2006.

C.-Y. Chiu, Y. Li, L. Ruan, X. Ye, C. B. Murray, and Y. Huang. Platinum nanocrystals selectivity shaped using facet-specific peptide sequences. *Nature Chem.*, 3: 393, 2011.

D. Chiu, W. Zhou, S. Kitayaporn, D. T. Schwartz, K. Murali-Krishna, T. J. Kavanagh, and F. Baneyx. Biomineralisation and size control of stable calcium phosphate core-protein shell nanoparticless: Potential for vaccine applications. *Bioconjug. Chem.*, 23610, 2012.

H. Churchill, H. Teng, and R. M. Hazen. Correlation of pH-dependent surface interaction forces to amino acid adsorption: Implications for the origins of life. *Am. Mineral.*, 89(7):1048, 2004.

G. Cicero, A. Calzolari, S. Corni, and A. Catellani. Anomalous wetting layer at the Au(111) surface. *J. Phys. Chem. Lett.*, 2:2582, 2011.

E. Cino, W.-Y. Choy, and M. Karttunen. Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations. *J. Chem. Theory Comput.*, 8:2725, 2012.

O. Cohavi, D. Reichmann, R. Abramovich, A. B. Tesler, G. Bellapadrona, D. B. Kokh, R. C. Wade, A. Vaskevich, I. Rubinstein, and G. Schreiber. A qualitative, real-time assessment of binding of peptides and protiens to gold surfaces. *J. Chem. Eur.*, 17:1327, 2011.

G. Collier, N. A. Vellore, J. A. Yancey, S. J. Stuart, and R. A. Latour. Comparison between empirical protein force fields for the simulation of the adsorption behaviuor of structured LK peptides on functionalised surfaces. *Biointerphases*, 7:24, 2012.

S. Collino and J. S. Evans. Molecular specifications of a mineral modulation sequence derived from the aragonite-promoting protein N16. *Biomacromol.*, 9:1909, 2008.

M. Colombo, S. Mazzucchelli, V. Collico, S. Avvakumova, L. Pandolfi, F. Corsi, F. Porta, and D. Prosperi. Protein-assisted one-pot synthesis and biofunctionalization of spherical gold nanoparticles for selective targeting of cancer cells. *Angew. Chem. Int. Ed.*, 51:9272, 2012.

R. Coppage, J. M. Slocik, B. D. Briggs, A. I. Frenkel, H. Heinz, R. R. Naik, and M. R. Knecht. Crystallographic recognition controls peptide binding for bio-based nanomaterials. *J. Am. Chem. Soc.* , 133:12346, 2011.

R. Coppage, J. M. Slocik, B. D. Briggs, A. I. Frenkel, R. R. Naik, and M. R. Knecht. Determining peptide sequence effects that control the size, structure and function of nanoparticles. *ACS Nano*, 61625, 2012.

D. Costa, A. Tougerti, F. Tielens, C. Gervais, L. Stievano, and J. F. Lambert. DFT study of the adsorption of microsolvated glycine on a hydrophilic silica surface. *Phys. Chem. Chem. Phys.*, 10:6360, 2008.

CPMD copyright MPI für Festkörperforschung Stuttgart 1997-2001. *copyright IBM Corp 1990-2008.*

R. T. Cygan, J. J. Liang, and A. G. Kalinichev. Molecular models of hydroxide, oxyhydroxide, and clay phases and the development of a general force field. *J. Phys. Chem. B*, 108:1255, 2004.

B. R. Danger, D. Fan, J. P. Vivek, and I. J. Burgess. Electrochemical studies of capping agent adsorption provide insight into the formation of anisotropic gold nanocrystals. *ACS Nano*, 611018, 2012.

T. A. Darden, D. York, and L. Pedersen. PME: an N log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089, 1993.

X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark. Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.*, 38:236, 1999.

S. De and S Mandal. Surfacant-assisted shape control of copper nanostructures. *Colloid Surface A*, 421:72, 2013.

R. di Felice and S. Corni. Simulation of peptide-surface recognition. *J. Phys. Chem. Lett.*, 2:1510, 2011.

M. Dion, H. Rydberg, E. Schroder, D. C. Langreth, and B. I. Lundqvist. Van der Waals density functional for general geometries. *Phys. Rev. Lett.*, 92:246401–1, 2004.

H. J. Dyson and P. E. Wright. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, 6:197, 2005.

C. M. Eggleston and G. Jordan. A new approach to pH of point zero charge measurement: Crystal-face specificity by scanning force microscopy (SFM). *Geochim. Cosmochim. Acta*, 62:1919, 1998.

E. Eteshola, L. J. Brillson, and S. C. Lee. Selection and characteristics of peptides that bind thermally grown silicon dioxide films. *Biomol. Eng.*, 22:201, 2005.

J. S. Evans, R. Samudrala, T. R. Walsh, E. E. Oren, and C. Tamerler. Molecular design of inorganic binding polypeptides. *MRS Bull*, 33(5):514, 2008.

P. P. Ewald. Die berechnung optischer und electrostatischer gitterpotentiale. *Ann. Phys.*, 64:253, 1921.

J. Feng, R. B. Pandey, R. J. Berry, B. L. Farmer, R. R. Naik, and H. Heinz. Adsorption mechanism of single amino acid and surfactant molecules to Au{111} surfaces in aqueous solution: Design rules for metal-binding molecules. *Soft Matter*, 7: 2113, 2011.

J. Feng, J. M. Slocik, M. Sarikaya, R. R. Naik, B. L. Farmer, and H. Heinz. Influence of the shape of nanostructured metal surfaces on adsorption of single peptide molecules in aqueous solution. *Small*, 8:1049, 2012.

Y. J. Feng, K. P. Bohnen, and C. T. Chan. First-principles studies of Au(100)-hex reconstruction in an electrochemical environment. *Phys. Rev. B*, 72:125401, 2005.

K. A. Fichthorn, K. E. Becker, and R. A. Miron. Molecular simulation of temperature-programmed desorption. *Catalysis Today*, 123:71, 2007.

N. Folliet, C. Gervais, D. Costa, G. Laurent, F. Babonneau, L. Stievano, J.-F. Lambert, F. Tielens. A molecular picture of the adsorption of glycine in mesoporous silica though NMR experiments combined with DFT-D calculations. *J. Phys. Chem. C*, 117:4104, 2013.

C. L. Freeman, J. H. Harding, D. J. Cooke, J. A. Elliot, J. S. Lardge, and D. M. Duffy. New forcefields for modelling biomineralization processes. *J. Phys. Chem. C*, 11111943, 2007.

C.-L. Fu and K.-M. Ho. 1st-principles calculation of the equilibrium ground state properties of transition metals: Applications to Nb and Mo. *Phys. Rev. B*, 28 (10):5480, 1983.

G. Fu, S. Valiyaveettil, B. Wopenka, and D. E. Morse. $CaCO_3$ biomineralisation: Acidic 8-kDa proteins isolated from aragonite abalone shell nacre can specifically modify calcite crystal morphology. *Biomacromolecules*, 6(3):1289, 2005.

H. Fukunishi, O. Watanabe, and S. Takada. On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction. *J. Chem. Phys.*, 116:9058, 2002.

M.-P. Gaigeot, M. Sulpizi and M. Sprik. Oxide/water interfaces: How the surface chemistry modifies interfacial water properties. *J. Phys. Condens. Matter.*, 24: 124106, 2012.

Q. Gao, W. J. Xu, Y. Xu, D. Wu, Y. H. Sun, F. Deng, and W. L. Shen. Amino acid adsorption on mesoporous materials: Influence of types of amino acids, modification of mesoporous materials, and solution conditions. *J. Phys. Chem. B*, 112: 2261, 2008.

P.-A. Garrain, D. Costa and P. Marcus. Biomaterial-biomolecule interaction: DFT-D study of glycine adsorption on Cr_2O_3. *J. Phys. Chem.*, 115:719, 2011.

L. M. Ghiringhelli and L. delle Site. Phenylalanine near inorganic surfaces: conformational statistics vs specific chemistry. *J. Am. Chem. Soc.*, 130:2634, 2008.

P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovicth. QE: A module and open source software project for quantum simulations of materials. *J. Phys.: Condens. Matter.*, 21: 395502, 2009.

D. Gibbs, B. M. Ocko, D. M. Zehner, and S. G. J. Mochrie. Structure and phases of the Au(001) surface: Inplane structure. *Phys. Rev. B*, 42:7330, 1990.

M. Gkikas, J. Timonen, J. Ruokolainen, P. Alexandrilis, and H. Iatrou. Facile aqueous synthesis and stabilisaztion of nearly monodispersed gold nanospheres by poly(L-proline). *J. Polym. Sci., Part A: Poylm. Chem.*, 51:1448, 2013.

S. Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27:1787, 2006.

S. Grimme, J. Antony, S. Ehrlich, and H. Kreig. A consistent and accurate ab initio parameterization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132:154104, 2010.

B. Hammer, L. B. Hansen, and J. K. Norskov. Improved adsorption energetics with density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys. Rev. B*, 59:7413, 1999.

J. W. Han and D. S. Sholl. Enantiospecific adsorption of amino acids on hydroxylated quartz (0001). *Langmuir* , 25:10737, 2009.

J. W. Han and D. S. Sholl. Enantiospecific adsorption of amino acids on hydroxylated quartz ($10\bar{1}0$). *Phys. Chem. Chem. Phys.* , 12:8024, 2010.

J. W. Han, J. N. James, and D. S. Sholl. First principles calculations of methylamine and methanol adsorption on hydroxylated quartz (0001). *Surf. Sci.*, 602:2478, 2008.

J. Hanes and A. Pluückthun. In vitro selection and evolution of functional proteins by using ribsome display. *Prot. Natl. Acad. Sci.*, 94:4937, 1997.

U. Harten, A. M. Lahee, J. Peter Toennies, and Ch. Woll. Observation of a soliton reconstruction of Au(111) by high resolution helium-atom diffraction. *Phys. Rev. Lett.*, 54(24):2619, 1985.

A. A. Hassanali and S. J. Singer. Static and dynamic properties of the water/amorphous silica interface: A model for the undissociated surface. *J. Computer-Aided Mater. Des.*, 14:53, 2007.

A. A. Hassanali, H. Zhang, C. Knight, Y. K. Shin, and S. J. Singer. The dissociated amorphous silica surface: Model development and evaluation. *J. Chem. Theory Comput.*, 6:3456, 2010.

P. Havu, V. Blum, V. Havu, P. Rinke, and M. Scheffler. Large-scale surface reconstruction energetics of Pt(100) and Au(100) by all-electron density functional theory. *Phys. Rev. B*, 82:161418, 2010.

T. Hayashi, K. I. Sano, K. Shiba, K. Iwahori, I. Yamashita, and M. Hara. Critical amino acid residues for the specific binding of the Ti-recognising recombinant ferritin with oxide surfaces of titanium and silicon. *Langmuir* , 25:10901, 2009.

M. He and M. J. Taussig. Antibody-ribosome-mRNA (ARM) complexes efficient selection particles for in vitro display and evolution of antibody combining sites. *Nucleic Acids Res.*, 255132, 1997.

H. Heinz, R. A. Vaia, B. L. Farmer, and R. R. Naik. Accurate simulation of surfaces and interfaces of fcc metals using 12-6 and 9-6 LJ potentials. *J. Phys. Chem. C*, 112:17281, 2008.

H. Heinz, B. L. Farmer, R. B. Pandey, J. M. Slocik, S. S. Patnaik, R. Pachter, and R. R. Naik. Nature of molecular interactions of peptides with gold palladium and Pd-Au bimetallic surfaces in aqueous solution. *J. Am. Chem. Soc.*, 131:9704, 2009.

H. Heinz, K. C. Jha, J. Luettmer-Strathmann, B. L. Farmer, and R. R. Naik. Polarisation at metal-biomolecular interfaces in solution. *J. R. Soc. Inter.*, 8:220, 2010.

C. R. Herbers, C. Li, and N. F. A. van der Vegt. Grand challenges in quantum-classical modelling of molecule-surface interactions. *J. Comput. Chem.*, 34:1177, 2013.

D. Herzbach, K. Binder, and M. H. Muser. Comparison of model potentials for molecular-dynamics simulations of silica. *J. Chem. Phys.*, 123:124711, 2005.

B. Hess. Determining the shear viscocity of model liquids from molecular dynamics simulations. *J. Chem. Phys.* , 116:209, 2002.

B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load balanced, and scable molecular simulation. *J. Chem. Theory Comp.*, 4435, 2008.

M. Hnilova, E. E. Oren, U. O. S. Seker, B. R. Wilson, S. Collino, J. S. Evans, C. Tamerler, and M. Sarikaya. Effect of molecular conformations on adsorption behavior of gold-binding peptides. *Langmuir* , 24:12440, 2008.

M. Hnilova, B. T. Karaca, J. Park, C. Jia, B. R. Wilson, M. Sarikaya, and C. Tamerler. Fabrication of hieracrhical hybrid structures using bio-enabled layer-by-layer self-assembly. *Biotechnol. Bioeng.*, 1091120, 2012a.

M. Hnilova, D. Khatayevich, D. Carlson, E. E. Oren, C. Gresswell, S. Zheng, F. Ohuchi, M. Sarikaya, and C. Tamerler. Single-step fabrication of patterned gold film array by an engineered multi-functional peptide. *J. Colloid Interface Sci.*, 26597, 2012b.

M. Hnilova, C. R. So, E. E. Oren, B. R. Wilson, T. Kacar, C. Tamerler, and M. Sarikaya. Peptide-directed co-assemebly of nanoprobes on mutlimaterial patterned solid surfaces. *Soft Matter*, 8:4327, 2012c.

R. W. Hockney, S. P. Goel, and J. Eastwood. Quiet high-resolution computer models of a plasma. *J. Comp. Phys.*, 14:148, 1974.

M. Hoefling, F. Iori, S. Corni, and K.-E. Gottschalk. Interaction of amino acids with the Au(111) surface: Adsorption free energies from molecular dynamics simulations. *Langmuir* , 26:8347, 2010a.

M. Hoefling, F. Iori, S. Corni, and K.-E. Gottschalk. The conformations of amino acids on the Au(111) surface. *Chem. Phys. Chem.*, 11:1763, 2010b.

M. Hoefling, S. Monti, S. Corni, and K. E. Gottschalk. Interaction of beta-sheet folds with a gold surface. *PLoS One*, 620925, 2011.

P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864, 1964.

G. J. Holinga, R. L. York, R. M. Onorato, C. M. Thompson, N. E. Webb, A. P. Yoon, and G. A. Somorjai. An SFG study of interfacial amino acids at the hydrophilic SiO_2 and hydrophobic deuterated polystrene surfaces. *J. Am. Chem. Soc.*, 133: 6243, 2011.

G. Hong, H. Heinz, R. R. Naik, B. L. Farmer, and R. Pachter. Toward understanding amino acid adsorption at metallic interfaces: A density functional theory study. *ACS Appl. Mater. and Inter.*, 1:388, 2009.

W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695, 1985.

J. Hritz and C. Oostenbrink. Hamiltonian replica exchange molecular dynamics using soft-core interactions. *J. Chem. Phys.*, 128:144121, 2008.

X. Huang, M. Hagen, B. Kim, R. A. Friesner, R. Zhuo, and B. J. Berne. Replica exchange with solute tempering: Efficiency in large scale systems. *J. Phys. Chem. B*, 111:5405, 2007.

Y. Huang, C.-Y. Chiang, S. K. Lee, Y. Gao, E. L. Hu, J. de Yoreo, and A. M. Belcher. Programmable assembly of nanoarchitectures using genetically engineered viruses. *Nano Lett.*, 51429, 2005.

H. W. Hugosson, A. Laio, P. Maurer and U. Rothlisberger. A comparative theoretical study of dipeptide solvation in water. *J. Comput. Chem*, 27:672, 2006.

F. Iori and S. Corni. Including image charge effects in the molecular dynamics simulations of molecules on metal surfaces. *J. Comput. Chem.*, 29:1656, 2008.

F. Iori, S. Corni, and R. di Felice. Unravelling the interaction between histidine side chain and the Au(111) surface: a DFT study. *J. Phys. Chem. C* , 112:13540, 2008.

F. Iori, R. di Felice, E. Molinari, and S. Corni. GolP: An atomistic force-field to describe the interaction of protiens with Au(111) surfaces in water. *J. Comp. Chem.* , 30:1465, 2009.

S. G. Itoh and H. Okumura. Coulomb replica-exchange method: Handling electrostatic attractive and repulsive forces for biomolecules. *J. Comput. Chem.*, 34:622, 2013.

S. G. Itoh, H. Okumura, and Y. Okamoto. Replica exchange method in van der Waals radius space: Overcoming steric restrictions for biomolecules. *J. Chem. Phys.*, 132:134105, 2010.

T. Jacob. Potential-induced lifting of the Au(100)-surface reconstruction studied with DFT. *Electrochim. Acta*, 52:2229, 2007.

J. P. M. Jambeck and A. P. Lyubartsev. Exploring the free energy landscape of solutes embedded in bilayers. *J. Phys. Chem. Lett.*, 4:1781, 2013.

P. Jedlovszky, J. P. Brodholt, F. Bruni, M. A. Ricci, E. U. Soper and R. Vallauri Analysis of the hydrogen-bonded structure of water from ambient to supercritical conditions *J. Chem. Phys.*, 108:8528, 1998

K. C. Jena and D. K. Hore. Water structure at solid surfaces and its implications for biomolecule adsorption. *Phys. Chem. Chem. Phys.* , 12:14383, 2010.

L. Jiang, Y. Tang, C. Liow, J. Wu, Y. Sun, Y. Jiang, Z. Dong, S. Li, V. P. Dravid, and X. Chen. Synthesis of fivefold stellate polyhedral gold nanoparticles with {110}-facets via seed-mediated growth method. *Small*, 9(5):705, 2013.

W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the OPLS all-atom forcefield on conformational energetic and properties of organic liquids. *J. Am. Chem. Soc.*, 118:11225, 1996.

W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* , 79:926, 1983a.

W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926, 1983.

B. D. Kay, K. R. Lykke, J. R. Creighton, and S. J. Ward. The influence of adsorbate-adsorbate hydrogen bonding in molecular chemisorption: NH_3, HF, H_2O on Au111. *J. Chem. Phys.*, 91:5120, 1989.

J. Kim, Y. Rheem, B. Yoo, Y. Chong, K. N. Bozhilov, D. Kim, M. J. Sadowsky, H.-G. Hur, and N. V. Myung. Peptide-mediated shape- and size-tunable synthesis of gold nanostructures. *Acta Biomater.*, 6:2681, 2010.

N. Kitadai, T. Yokoyama, and S. Nakashima. ATR-IR spectroscopic study of L-lysine adsorption on amorphous silica. *J. Col. Inter. Sci.*, 329:31, 2009.

J. Klimes and A. Michaelides. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.*, 137: 120901, 2012.

J. Klimes, D. R. Bowler, and A. Michaelides. Van der Waals density functionals applied to solids. *Phys. Rev. B*, 83:195131, 2011.

W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):1133, 1965.

D. M. Kolb. Reconstruction phenomena at metal-electrolyte interfaces. *Prog. Surf. Sci.*, 51109, 1996.

M. Kosmulski. The pH dependent surface charging and the points of zero charge. *J. Col. Inter. Sci.*, 253:77, 2002.

N. Kroger, R. Deutzmann, and M. Sumper. Polycationic peptides from diatom biosilica that direct silica nanosphere formation. *Science*, 286:1129, 1999.

N. Kroger, R. Deutzmann, and M. Sumper. Silica-precipitating peptides from diatoms- the chemical structure of silaffin-1a from cylindrotheca fusiformis. *J. Biol. Chem.* , 276:26066, 2001.

N. Kroger, S. Lorenz, E. Brunner, and M. Sumper. Self-assembly of highly phosphorylated silaffins and their function in biosilica morphogenesis. *Science*, 298: 584, 2002.

J. L. Kulp (III), M. Sarikaya, and J. S. Evans. Molecular charaterisation of a prokaryotic polypeptide sequence that catalyzes Au crystal formation. *J. Mater. Chem.*, 14:2325, 2004.

A. Kumar, H. Ma, X. Zhang, K. Huang, S. Jin, J. Liu, T. Wei, W. Cao, G. Zou, and X.-J. Liang. Gold nanoparticles functionalised with therapeutic and targeted peptids for cancer treatment. *Biomaterials*, 33:1180, 2012.

A. Laio and M. Parrinello. Escaping free-energy minima. *Prot. Natl. Acad. Sci.*, 99 12562, 2002.

D. C. Langreth, B. I. Lundqvist, S. D. Chakarova-Kack, V. R. Cooper, M. Dion, P. Hyldgaard, A. Kelkkanen, J. Kleis, L. Kong, S. Li, P. G. Moses, E. Murray, A. Puzder, H. Rydberg, E. Schroder, and T. Thonhauser. A density functional for sparse matter. *J. Phys: Condens. Matter.*, 21:084203, 2009.

R. A. Latour. Molecular simulation of protein-surface interactions: benefits, problems, solutions and future directions. *Biointerphases*, 3FC3, 2008.

D. J. Lavrich, S. M. Wetterer, S. L. Bernasek, and G. Scoles. Physisorption and chemisorption of alkanethiols and alkyl sulfides on Au(111). *J. Phys. Chem. B*, 102:3456, 1998.

A. J. Lee and S. W. Rick. Improving replica exchange using driven scaling. *J. Chem. Phys.*, 131:174113, 2009.

K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth. Higher-accuracy van der Waals density functional. *Phys. Rev. B*, 82:081101, 2010.

G. Li, I. Tamblyn, V. R. Cooper, H-J Gao, and J. B. Neaton. Molecular adsorption on metal surfaces with van der Waals density functional. *Phys. Rev. B*, 85:121409, 2012.

H.-Q. Li, A. Chen, S. G. Roscoe, and J. Lipkowski. Electrochemical and FTIR studies of L-phenylalanine adsorption at the Au(111) electrode. *J. Electroanal. Chem.*, 500:299, 2001.

X. Li and R. A. Latour. The temperature intervals with global exchange of replicas empirical accelerated sampling method: Parameter sensitivity and extension to a complex system. *J. Comput. Chem.*, 32:1091, 2011.

X. Li, C. P. O'Brien, G. Collier, N. A. Vellore, F. Wang, R. A. Latour, D. A. Bruce, and S. J. Stuart. An improved replica-exchange sampling method: Temperature intervals with global energy reassignment (TIGER). *J. Chem. Phys.*, 127:164116, 2007.

X. Li, R. A. Latour, and S. J. Stuart. TIGER 2: An improved algorithm for temperature intervals with global exchange of replicas. *J. Chem. Phys.*, 130: 174106, 2011.

Z. Y. Li, N. P. Young, M. Di Vece, S. Palomba, R. E. Palmer, A. L. Bleloch, B. C. Curley, R. L. Johnston, J. Jiang, and J. Yuan. Three-dimensional atomic-scale structure of size-selected gold nanoclusters. *Nature*, 451:46, 2008.

T. Liang and T. R. Walsh. Molecular dynamics simulations of peptide carboxylate hydration. *Phys. Chem. Chem. Phys.*, 8:4410, 2006.

T. Liang and T. R. Walsh. Simulation of the hydration structure of glycyl-alanine. *Mol. Simulat.*, 33:337, 2007.

Y.-H. Liao, Y.-J. Chang, Y. Yoshiike, Y.-Chorng Chang, and Y.-R. Chen. Negatively charged gold nanoparticles inhibit Alzheimer's amyloid-beta fibrillization, induce fibril dissociation and mitigate neutoxicity. *Small*, 83631, 2012.

I.-C. Lin and M. E. Tuckerman. Enhanced conformational sampling of peptides via reduced side-chain and solvent masses. *J. Phys. Chem. B*, 114:15935, 2010.

X. Lin and A. Gross. First-prinicples study of the water structure on flat and stepped gold surfaces. *Surf. Sci.*, 606:886, 2012.

E. Lindahl, B. Hess and D. van der Spoel. GROMACS 3.0: A package for molecular simulation and trajectory analysis *J. Mol. Mod.* , 7:306, 2001.

K. Lindorff-Larson, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw. Systematic validation of protein force fileds against experimental data. *PLoS One*, 7e32131, 2012.

P. Liu, B. Kim, R. A. Friesner, and B. J. Berne. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Nat. Acad. Sci.*, 102:13749, 2005.

P. Liu, X. Huang, R. Zhou, and B. J. Berne. Hydrophobic aided replica exchange: An efficient algorithm for protein folding in explicit solvent. *J. Phys. Chem. B*, 110:19018, 2006.

I. Lopes, L. Piao, L. Stievano, and J.-F. Lambert. Adsorption of amino acids on oxide supports: A solid-state NMR study of glycine adsorption on silica and alumina. *J. Phys. Chem. C*, 113(42):18163, 2009.

P. E. M. Lopes, V. Murashov, M. Tazi, E. Demchuk, and A. D. MacKerell. Development of an empirical force field for silica. Application to the quartz-water interface. *J. Phys. Chem. B* , 110:2782, 2006.

C. D. Lorenz, P. S. Crozier, J. A. Anderson, and A. Travesset. Molecular dynamics of ionic transport and electrokinetic effects in realistic silica channels. *J. Phys. Chem. C* , 112:10222, 2008.

A. Luzar and D. Chandler. Structure and hydrogen-bond dynamics of water-dimethyl sulfoxide mixtures by computer-simulations. *J. Chem. Phys.*, 98:8160, 1993.

A. Luzar and D. Chandler. Hydrogen-bond kinetics in liquid water. *Nature*, 379:55, 1996.

K. Lutz, C. Groger, M. Sumper, and E. Brunner. Biomimetic silica formation: Analysis of the phosphate-induced self-assembly of polyamines. *Phys. Chem. Chem. Phys.* , 7:2812, 2005.

A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All atom empirical potential for molecular modelling and dynamics studies of protiens. *J. Phys. Chem. B*, 102:3586, 1998.

A. D. MacKerell (Jr), M. Feig, and C. L. Brooks (III). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.*, 25:1400, 2004.

M. Mahmood, D. Casciano, Y. Xu, and A. S. Biris. Engineered nanostructural materials for application in cancer biology and medicine. *J. Appl. Toxicol.*, 32:10, 2012.

G. Makov and M. C. Payne. Periodic boundary conditions in ab initio calculations. *Phys. Rev. B*, 51:4014, 1995.

M. W. Mahoney and W. L. Jorgensen A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, 112:8910, 2000.

L. D. Marks. Experimental studies of small particle structures. *Rep. Prog. Phys.*, 57:603, 1994.

M. Meng, L. Stievano, and J-F. Lambert. Adsorption and thermal condensation mechanisms of amino acids on oxide supports. 1. glycine on silica. *Langmuir* , 20: 914, 2004.

Y. Meng, D. S. Dashti, and A. E. Roitberg. Computing alchemical free energy differences with hamiltonian replica exchange molecular dynamics (H-REMD) simulations. *J. Chem. Theory Comput.*, 7:2721, 2011.

R. A. Metzler, J. S. Evans, C. E. Killan, D. Zhou, T. H. Churchill, N. P. Appathurai, S. N. Coppersmith, and P. U. P. A. Gilbert. Nacre protein fragment templates lamellar aragonite growth. *J. Am. Chem. Soc.* , 132:6329, 2010.

P. A. Mirau, R. R. Naik, and P. Gehring. Structure of peptides on metal oxide surfaces probed by NMR. *J. Am. Chem. Soc.*, 133:18243, 2011.

S. Miyamoto and P. A. Kollman. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithms for Rigid Water Models. *J. Comp. Chem.* 13:952, 1992.

S. G. J. Mochrie, D. M. Zehner, B. M. Ocko, and D. Gibbs. Structure and phases of the Au(001) surface: X-ray scattering measurements. *Phys. Rev. Lett.*, 64(24): 2925, 1990.

S. Monti and T. R. Walsh. Free energy calculations of the adsorption of amino acid analogues at the aqueous titania interface. *J. Phys. Chem. C* , 114:22197, 2010.

S. L. C. Moors, S. Michielssens, and A. Ceulemans. Improved replica exchange method for native-state protein sampling. *J. Chem. Theory Comput.*, 7:231, 2011.

Y. Muto. Force between nonpolar molecules. *J. Phys. Math. Soc. Jp*, 17:629, 1943.

R. Nadler and J. F. Sanz. Effect of dispersion correction on the Au(111)-H_2O interface: A first-prinicples study. *J. Chem. Phys.*, 137:114709, 2012.

R. R. Naik, L. L. Brott, S. J. Clarson, and M. O. Stone. Silica-precipitating peptides isolated from a combinatorial phage display library. *J. Nanosci. Nanotechnol.*, 2: 95, 2002a.

R. R. Naik, S. J. Stringer, G. A. Agarwal, S. E. Jones, and M. O. Stone. Biomimetic synthesis and patterning of silver nanoparticles. *Nat. Mater.*, 1:169, 2002b.

K. T. Nam, D.-W. Kim, P. J. Yoo, C.-Y. Chiang, N. Meethong, P. T. Hammond, Y.-M. Chiang, and A. M. Belcher. Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes. *Science*, 312:885, 2006.

S. Z. Nergiz, J. M. Slocik, R. R. Naik and S. Singamaneni. Surface defect sites facilitate fibrillation: an insight into adsorption of gold-binding peptides on Au(111). *Phys. Chem. Chem. Phys.* 15:11629, 2013.

E. Neria, S. Fischer and M. Karplus Simulation of activation free energies in molecular systems. *J. Chem. Phys.* 105:1902, 1996.

M. Nonella and S. Seeger. Investigating alanine-silica interaction by means of first-principles molecular-dynamics simulations. *Chem. Phys. Chem.*, 9:414, 2008.

S. Nosé. A molecular-dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 52:255, 1984.

S. Nosé and M. L. Klein. Constant pressure molecular-dynamics for molecular-systems. *Mol. Phys.*, 50:1055, 1983.

R. Notman and T. R. Walsh. Molecular dynamics studies of the interactions of water and amino acid analogues with quartz surfaces. *Langmuir* , 25:1638, 2009.

R. Notman, E. E. Oren, C. Tamerler, M. Sarikaya, R. Samudrala, and T. R. Walsh. Solution study of engineered quartz-binding peptides using replica exchange molecular dynamics. *Biomacromolecules*, 11:3266, 2010.

R. G. Nuzzo, B. R. Zegarski, and L. H. Dubois. Fundemental studies of the chemisorption of organosulfur compounds on Au(111). Implications for molecular self-assembly on gold surfaces. *J. Am. Chem. Soc.*, 109:733, 1987.

B. M. Ocko, D. Gibbs, K. G. Huang, D. M. Zehner, and S. G. J. Mochrie. Structure and phases of the Au(001) surface: Absolute X-ray reflectivity. *Phys. Rev. B*, 44: 6429, 1991.

T. Okabe, M. Kawata, Y. Okamoto, and M. Mikami. Replica-exchange Monte Carlo method for the isobaric-isothermal ensemble. *Chem. Phys. Lett.*, 335:435, 2001.

E. E. Oren, C. Tamerler, D. Sahin, M. Hnilova, U. O. S. Seker, M. Sarikaya, and R. Samudrala. A novel knowledge-based approach to design inorganic-binding peptides. *Bioinformatics* , 23:2816, 2007.

E. E. Oren, R. Notman, , I. W. Kim, J. S. Evans, T. R. Walsh, R. Samudrala, C. Tamerler, and M. Sarikaya. Probing the molecular mechanisms of quartz-binding peptides. *Langmuir* , 26:11003, 2010.

R. B. Pandey, H. Heinz, J. Feng, B. L. Farmer, J. M. Slocik, L. F. Drummy, and R. R. Naik. Adsorption of peptides (A3, Flg, Pd2, Pd4) on gold and palladium surfaces by a coarse-grained Monte Carlo simulation. *Phys. Chem. Chem. Phys.*, 11:1989, 2009.

M. Parrinello and A. Rahman. Polymorphic transitions in single-crystals- a new molecular dynamics method. *J. Appl. Phys.*, 52:7182, 1981.

S. Patel and C. L. Brooks (III). Fluctuating charge force fields: Recent developments and applications from small molecules to macromolecular biological systems. *Mol. Simul.*, 32(3-4):231, 2006.

A. Patriksson and D. van der Spool. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.*, 10:2073, 2008.

S. V. Patwardhan. Biomimetic and bioinspired silica: Recent developments and applications. *Chem. Commun.*, 47:7567, 2011.

S. V. Patwardhan and S. J. Clarson. Silification and biosilification. *Silicon Chemistry*, 1:207, 2002.

S. V. Patwardhan, F. S. Emami, R. J. Berry, S. E. Jones, R. R. Naik, O. Deschaume, H. Heinz, and C. C. Perry. Chemistry of aqueous silica nanoparticle surfaces and the mechanism of selective peptide adsorption. *J. Am. Chem. Soc.*, 143:6244, 2012.

B. R. Peelle, E. M. Krauland, K. D. Wittrup, and A. M. Belcher. Design criteria for engineering inorganic material specific peptides. *Langmuir* , 21:6929, 2005.

S. Paul and A. Chandra. Hydrogen bond dynamics at vapour-water and metal-water interfaces. *Chem. Phys. Lett.*, 386:218, 2004

H.-C. Peng, S. Xie, J. Park, X. Xia, and Y. Xia. Quantitative analysis of the convergence density of Br^- ions on Pd$\{100\}$ facets and its role in controlling the shape of Pd nanocrystals. *J. Am. Chem. Soc.*, 135:3780, 2013.

J. P. Perdew, K. Burke and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865, 1996.

A. Phan, T. A. Ho, D. Cole, and A. Striolo. Molecular structure and dynamics in thin water films at metal oxide surfaces: Magnesium, aluminum and silicon oxide surfaces. *J. Phys. Chem. C*, 116:15982, 2012.

S. Piana, K. Lindorff-Larsen, and D. E.Shaw. How robust are protein folding simulations with respect to force field parameterisation? *Biophysical J.*, 100:47, 2011.

V. Puddu and C. C. Perry. Peptide adsorption on silica nanoparticles: Evidence of hydrophobic interactions. *ACS Nano*, 6:6356–6363, 2012.

D. Quigley and P. M. Rodger. Metadynamics simulations of ice nucleation and growth. *J. Chem. Phys.*, 128:154518, 2008

P. Ren and J. W. Ponder. Polarizable atomic mulitpole water model for molecular mechanics simulation. *J. Phys. Chem. B*, 107:5933, 2003.

S. W. Rick. Replica exchage with dynamical scaling. *J. Chem. Phys.*, 126:054102, 2007.

A. Rimola, B. Civalleri, and P. Ugliengo. Neutral vs zwitterionic glycine forms at the water/silica interface:. *Langmuir* , 24:14027, 2008.

A. Rimola, M. Sodupe, and P. Ugliengo. Affinity scale for the interaction of amino acids with silica surfaces. *J. Phys. Chem. C* , 113:5741, 2009.

A. Rimola, M Aschi, R. Orlando, and P. Ugliengo. Does adsorption at hydroxyapatite surfaces induce peptide folding? Insights from large-scale B3LYP calculations. *J. Am. Chem. Soc.*, 134:10889, 2012.

G. Roman-Perez and J. M. Soler. Efficient implementation of a van der Waals density functional: Application to double-wall carbon nanotubes. *Phys. Rev. Lett.*, 103: 096102, 2009.

L. Romaner, D. Nobok, P. Puschnig, E. Zojer, and C. Ambrosch-Draxl. Theoretical study of PTCDA adsorbed on the coinage metal surfaces Ag(111), Au(111), Cu(111). *New J. Phys.*, 11:053010, 2009.

M. Rosa, S. Corni, and R. di Felice. A density functional theory study of cytosine on Au(111). *J. Phys. Chem. C*, 116:21366, 2012.

D. E. Rosenfeld and C. A. Schmuttenmaer Dynamics of the water hydrogen bond network at ionic, nonionic and hydrophobic interfaces in nanopores and reverse micelles *J. Phys. Chem. B*, 115:1021, 2011

L. Ruan, H. Ramezani-Dakhel, C.-Y. Chiu, Y. Li, H. Heinz, and Y. Huang. Tailoring molecular specificity toward a crystal face: A lesson from biorecognition toward Pt{111}. *Nano Lett.*, 13:840, 2013.

K.-I. Sano and K. Shiba. A hexapeptide motif that electrostatically binds to the surface of titanium. *J. Am. Chem. Soc.* , 125(47):14234, 2003.

K-I Sano, H Sasaki, and K Shiba. Specificity and biomineralisation activities of Ti-binding peptide-1 (TBP-1). *Langmuir* , 21:3090, 2005.

J. Schneider and L. Colombi Ciacchi. Specific material recognition by small peptides mediated by the interfacial solvent structure. *J. Am. Chem. Soc.*, 134:2407, 2012.

S. Schnur and A. Gross. Properties of metal-water interfaces studied from first principles. *New J. Phys*, 11:125003, 2009.

A. P. Schoen, D. T. Schoen, K. N. L. Huggins, M. A. Arunagirinathan, and S. C. Heilshorn. Template engineering through epitope recognition: a modular biomimetic strategy for inorganic nanomaterial synthesis. *J. Am. Chem. Soc.* , 133:18202, 2011.

P. Schravendijk, L. M. Ghiringhelli, L. delle Site, and N. F. A. van der Vegt. Interaction of hydrated amino acids with metal surfaces: A multiscale modelling description. *J. Phys. Chem. C*, 111:2631, 2007.

K.-P. Schroder, J. Sauer, M. Leslie, C. R. A. Catlow, and J. M. Thomas. Bridging hydroxyl groups in zeolite catalysts: A computer simulation of their structure, vibrational properties and acidity in protonated faujasites (H-Y zeolites). *Chem. Phys. Lett.*, 188(4):320, 1992.

T. Schwabe and S. Grimme. Double-hybrid density functionals with long-range dispersion corrections: Higher accuracy and extended applicability. *Phys. Chem. Chem. Phys.*, 9(26):3397, 2007.

U. O. S. Seker and H. V. Demir. Material binding peptides for nanotechnology. *Molecules*, 16:1426, 2011.

M. Sethi and M. R. Knecht Experimental studies on the interactions between Au nanoparticles and amino acids: Bio-based formation of branched linear chains. *ACS Appl. Mater. Inter.*, 1:1270, 2009.

M. Sethi and M. R. Knecht Understanding the mechanism of amino acid Au nanoparticle chain formation. *Langmuir*, 26:9860, 2010.

A. A. Shemetov, I. Nabiev, and A. Sukhanova. Molecular interaction of proteins and peptides with nanoparticles. *ACS Nano*, 64585, 2012.

H. Shi and C. Stampfl. Shape and surface structure of gold nanoparticles under oxidising conditions. *Phys. Rev. B*, 77:94127, 2008.

D. Shuai, D. C. McCalman, J. K. Choe, J. R. Shapley, W. F. Schneider, and C. J. Werth. Structure sensitivity study of waterbourne contaminant hydrogenation using shape- and size-controlled Pd nanoparticles. *ACS Catal.*, 3:453, 2013.

A. A. Skelton, T. N. Liang, and T. R. Walsh. Interplay of sequence, conformation, and binding at the peptide-titania interface as mediated by water. *ACS Appl. Mater. Interfaces*, 1:1482, 2009.

A. A. Skelton, P. Fenter, J. D. Kubicki, D. J. Wesolowski, and P. T. Cummings. Simulations of the quartz (10 11)/water interface: A comparison of classical forcefields, ab initio molecular dynamics, and X-ray reflectivity experiments. *J. Phys. Chem. C* , 1152076, 2011a.

A. A. Skelton, D. J. Wesolowski, and P. T. Cummings. Investigating the quartz (10 10)/water interface using classical and ab initio molecular dynamics. *Langmuir* , 27:8700, 2011b.

P. Skoluda. The influence of carbonyl aliphatic compounds on extension of the reconstructed Au(100) surface. *Electrochem. Commun.*, 6:785, 2004.

P. Skoluda. Inhibition of potential-induced surface reconstruction on Au(100) electrode by the products of tyramine electro-oxidation. *Electrochimica Acta*, 56:8625, 2011.

J. M. Slocik, R. R. Naik, M. O. Stone, and D. W. Wright. Viral templates for gold nanoparticle synthesis. *J. Mater. Chem.*, 15:749, 2005a.

J. M. Slocik, M. O. Stone, and R. R. Naik. Synthesis of gold nanoparticles using multifunctional peptides. *Small*, 1:1048, 2005b.

J. A. Snyder, T. Abramyan, J. A. Yancey, A. A. Thyparambil, Y. Wei, S. J. Stuart, and R. A. Latour. Development of a tuned interfacial force field parameter set for the simulation of protein adsorption to silica glass. *Biointerphases*, 7:56, 2012.

C. R. So, J. L. Kulp, E. E. Oren, H. Zareie, C. Tamerler, J. S. Evans, and M. Sarikaya. Molecular recognition and supramolecular self-assembly of a genetically engineered gold binding peptide on Au(111). *ACS Nano* , 3:1525, 2009.

E. Soto-Cantu, S. Turksen-Selcuk, J. H. Qiu, Z. Zhou, P. S. Russo, and M. C. Henk. Silica-polypeptide composite particles: Controlling shell growth. *Langmuir* , 26 15604, 2010.

Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protien folding. *Chem. Phys. Lett.*, 314:141, 1999.

Y. Sugita, A. Kitao, and Y. Okamoto. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.*, 113:6042, 2000.

M. Sulpizi, M.-P. Gaigeot and M. Sprik. The silica-water interface: How the silanols determine the surface acidity and modulate the water properties. *J. Chem. Theory Comput.*, 8:1037, 2012.

M. Sumper and E. Brunner. Silica biomineralisation in diatoms: The model organism thalassiosira pseudonana. *Chem. Bio. Chem.*, 9:1187, 2008.

K. Sun, M. Kohyama, S. Tanaka, and S. Takeda. A theoretical study of co adsorption on gold by Huckel theory and density functional theory calculations. *J. Comput. Chem.*, 32:3276, 2011.

A. Surrey, D. Pohl, L. Schultz, and B. Rellinghaus. Quantitative measurement of the surface self-diffusion on Au nanoparticles by aberratin-corrected transmission electron microscopy. *Nano Lett.*, 12:6071, 2012.

M. B. Swindells, M. W. MacArthur, and J. M. Thornton. Intrinsic ϕ,ψ propensities of amino acids, derived from the coil regions of known structures. *Nature Structural Biology*, 2(7):596, 1995.

D. Syomin and B. E. Koel. IRAS studies of the orientation of acetone molecules in monolayer and multilayer films on Au(111) surfaces. *Surf. Sci.*, 498:53, 2002a.

D. Syomin and B. E. Koel. Probing the reactivity of C-6-hydrocarbons on Au surfaces: Cyclohexane, cyclohexyl and cyclohexene on Au(111). *Surf. Sci.*, 498: 61, 2002b.

D. Syomin, J. Kim, B. E. Koel, and G. B. Ellison. Identification fo adsorbed phenyl (C_6H_5) groups on metal surfaces: Electron-induced dissocaiation of benzene on Au(111). *J. Phys. Chem. B*, 105:8387, 2001.

C. Tamerler and M. Sarikaya. Genetically designed peptide-based molecular materials. *ACS Nano* , 3:1606, 2009.

C. Tamerler, T. Kacar, D. Sahin, H. Fong, and M. Sarikaya. Genetically engineered polypeptides for inorganics: A utility in biological materials science and engineering. *Mat. Sci. Eng. C-Bio S*, 27:558, 2007.

C. Tamerler, D. Khatayevich, M. Gungormus, T. Kacar, E. E. Oren, M. Hnilova, and M. Sarikaya. Molecular biomimetics: GEPI-based biological routes to technology. *Biopolymers* , 94:78, 2010.

Z. Tang, J. P. Palafox-Hernandez, W. Law, Z. E. Hughes, M. T. Swihart, P. N. Prasad, M. R. Knecht, T. R. Walsh. Biomolecular recognition principles for bionanocombinatorics: and integrated approach to elucidate enthalpic and entropic factors. *ACS Nano*, 7:9632, 2013.

T. Terakawa, T. Kameda, and S. Takada. On easy implementation of a varient of the replica exchange with solute tempering in Gromacs. *J. Comput. Chem.*, 32: 1228, 2011.

T. Thonhauser, V. R. Cooper, S. Li, A. Puzder, P. Hyldgaard, and D. C. Langreth. Van der Waals density functional: Self-consistent potential and the nature of the van der Waals bond. *Phys. Rev. B*, 76:125112, 2007.

A. A. Thyparambil, Y. Wei, and R. A. Latour. Determination of peptide-surface adsorption free energy for material surfaces not conducive to SPR or QCM using AFM. *Langmuir*, 28:5687, 2012.

G. Tiana. Estimation of microscopic averages from metadynamics. *Eur. Phys. J. B*, 63:235, 2008.

A. Tilocca and A. N. Cormack. The initial stages of bioglass dissolution: a Car-Parrinello molecular-dynamics study of the glass-water interface. *Proc. R. Soc. London, Ser. A*, 467:2102, 2011.

A. Tkatchenko and M. Scheffler. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.*, 102: 073005, 2009.

K. Toyoda, Y. Nakano, I. Hamada, K. Lee, S. Yanagisawa, and Y. Morikawa. First principles study of benzene on noble metal surfaces: Adsorption states and vacuum level shifts. *Surf. Sci.*, 603:2912, 2009.

D. Trzesniak, A.-P. E. Kunz, and W. F. vanGunsteren. A comparison of methods to compute the potential of mean force. *Chem. Phys. Chem.*, 8:162, 2007.

V. N. Uversky and A. K. Dunker. Understanding protein non-folding. *Biochim. Biophys. Acta*, 18041231, 2010.

D. Vanderbilt Soft Self-Consistent Pseudopotentials in a Generalized Eigenvalue Formalism. *Phys. Rev. B*, 41:7892, 1990.

A. C. T. vanDuin, S. Dasgupta, F. Lorant, and W. A. Goddard(III). A reactive force field for hydrocarbons. *J. Phys. Chem. A*, 105:9396, 2001.

G. Vecchio, A. Galeone, V. Brunetti, G. Maiorano, L. Rizzello, S. Sabella, R. Cingolani, and P. P. Pompa. Mutagenic effects of gold nanoparticles induce aberrant phenotypes in drosophila melanogaster. *Nanomed. Nanotech. Biol. Med.*, 8:1, 2012a.

G. Vecchio, A. Galeone, V. Brunetti, G. Maiorano, L. Rizzello, S. Sabella, R. Cingolani, and P. P. Pompa. Concerntration-dependent, size-indepedent toxicity of citrate capped AuNPs in drosophila melangaster. *PLoS one*, 7:e29980, 2012b.

N. A. Vellore, J. A. Yancey, G. Collier, and R. A. Latour. Assessment of the transferability of a protein forcefield for the simulation of peptide-surface interactions. *Langmuir* , 267396, 2010.

S. Venkatachalam, P. Kaghazchi, L. A. Kibler, D. M. Kolb, and T. Jacob. First priniciples studies of the potential-induced lifting of the Au(100) surface reconstruction. *Chem. Phys. Lett.*, 455:47, 2008.

A. V. Verde, J. M. Acres, and J. K. Maranas. Investigating the specificty of peptide adsorption on gold using molecular dynamics simulations. *Biomacromolecules*, 10:2118, 2009.

A. V. Verde, P. J. Beltramo, and J. K. Maranas. Adsorption of homopolypeptides on gold investigated using atomistic molecular dynamics. *Langmuir* , 27:5918, 2011.

N. N. Vlasova and L. P. Golovkova. Adsorption of amino acids on the surface of highly dispersed silica. *Colloid J.*, 66(6):657, 2004.

D. Walczyk, F. B. Bombelli, M. P. Monopoli, I. Lynch, and K. A. Dawson. What the cell sees in bionanoscience. *J. Am. Chem. Soc.* , 132:5761, 2010.

F. Wang, S. J. Stuart, and R. A. Latour. Calculation of adsorption free energy for solute-surface interactions biased replica-exchange molecular dynamics. *Biointerphases*, 39, 2008.

L. Wang, R. A. Friesner, and B. J. Berne. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B*, 115:9431, 2011.

X.-Q. Wang. Phases of the Au(100) surface reconstruction. *Phys. Rev. Lett.*, 67: 3547, 1991.

Z. W. Wang and R. E. Palmer. Mass spectrometry and dynamics of gold adatoms observed on the surface of size-selected Au nanoclusters. *Nano Lett.*, 12:91, 2012.

Y. Wei and R. A. Latour. Correlation between desorption force measured by AFM and adsorption free energy measured by SPR for peptide-surface interactions. *Langmuir*, 26(24):18852, 2010.

J. Wellendorff, A. Kelkkanen, J. J. Mortensen, B. I. Lundqvist, and T. Bligaard. RPBE-vdW description of benzene adsorption on Au(111). *Top. Catal.*, 53:378, 2010.

S. M. Wetterer, D. J. Lavrich, T. Cummings, S. L. Bernasek, and G. Scholes. Energies and kinetics of the physisorption of hydrocarbons on Au(111). *J. Phys. Chem. B*, 102:9266, 1998.

R. L. Willet, K. W. Baldwin, K. W. West, and L. N. Pfeiffer. Differential adhension of amino acids to inorganic surfaces. *Prot. Natl. Acad. Sci.*, 1027817, 2005.

K. D. Wittrup. Protein engineering by cell-surface display. *Curr. Opin. Biotech.*, 12395, 2001.

L. B. Wright and T. R. Walsh. Facet selectivity of binding on quartz surfaces: free energy calculations of amino-acid analogue adsorption. *J. Phys. Chem. C*, 116: 2933, 2012a.

L. B. Wright and T. R. Walsh. First-principles molecular dynamics simulations of NH_4^+ and CH_3COO^- adsorption at the aqueous quartz interface. *J. Chem. Phys.*, 137:224702, 2012b.

L. B. Wright and T. R. Walsh. Efficient conformational sampling of peptides adsorbed onto inorganic surface: Insights from a quartz binding peptide. *Phys. Chem. Chem. Phys.*, 15:4715, 2013.

L. B. Wright, C. L. Freeman, and T. R. Walsh. Benzene adsorption at the aqueous (011) α-quartz interface: is surface flexiblity important? *Mol. Sim.*, 39:1093, 2013a.

L. B. Wright, P. M. Rodger, S. Corni, and T. R. Walsh. GolP-CHARMM: First principles based force-fields for the interaction of proteins with Au(111) and Au(100). *J. Chem. Theory Comput.*, 9:1616, 2013b.

L. B. Wright, P. M. Rodger, T. R. Walsh, and S. Corni. First-principles based force fields for the interaction of protiens with Au(100)(5x1): An extension of GolP-CHARMM. *J. Phys. Chem. C*, 117:24292, 2013c.

L. Xia and Z. Li. Poly-L-lysine mediated biomimetic silica synthesis: Effects of mixing sequences and counterion concentrations. *Langmuir* , 27(3):1116, 2011.

W. Xie, J. Pu, A. D. MacKerellJr, and J. Gao. Development of a polarizable intermolecular potential function (PIPF) for liquid amides and alkanes. *J. Chem. Theory and Comput.*, 3:1878, 2007.

R. L. York, G. J. Holinga, and G. A. Somorjai. An investigation of the influence of chain length on the interfacial ordering of L-lysine and L-proline and their homopeptides at hydrophobic and hydrophilic interfaces studied by SFG and QCM. *Langmuir*, 25:9369, 2009.

J. Yu, M. L. Becker, and G. A. Carri. A molecular dynamics simulation of the stablity-limited growth mechanism of peptide-mediated gold nanoparticle synthesis. *Small*, 62242, 2010.

J. Yu, L. Becker, and G. A. Carri. The influence of amino acid sequence and functionality on the binding process of peptides onto gold surfaces. *Langmuir*, 28: 1408, 2012.

J. H. Zar. Significance testing of spearman rank correlation coefficient. *J. Am. Stat. Assoc.*, 67:578, 1972.

D. M. Zehner, B. R. Appleton, T. S. Noggle, J. W. Miller, J. H. Barret, L. H. Jenkins, and E. Schow III. Characterization of reordered (001) Au surfaces by positive-ion-channeling spectroscopy, LEEDS and AES. *J. Vac. Sci. Technol.*, 12: 454, 1975.

Y. L. Zhao, S. Köppen, and T. Frauenheim. An SCC-DFTB/MD study of the adsorption of zwitterionic glycine on a geminal hydroxylated silica surface in an explicit water environment. *J. Phys. Chem. C*, 115:9615, 2011.

A. R. Zimmerman, K. W. Goyne, J. Chorover, S. Komarneni, and S. L. Brantley. Mineral mesopore effects on nitrogenous organic matter adsorption. *Org. Geochem.*, 35:355, 2004.