# The interaction of materials and biology: simulations of peptides, surfaces, and biomaterials

Lewis J. Martin

A thesis submitted in fulfilment of the requirements for the degree of

*Doctor of Philosophy*

Faculty of Science

University of Sydney

March 2018

# Abstract

Biomaterials were originally designed to augment or replace damaged tissue in the body, but now encompass a wider range of applications including drug delivery, cancer vaccines, electronic sensor devices, and non-fouling coatings for ship hulls. At the heart of all of these applications is the interface between synthetic materials and biology. Modern techniques for studying this interface are limited to the macro and micro scales. With the advent of high performance computing clusters, molecular simulation is now capable of simulating the interface at the nano-scale.

This thesis demonstrates how simulation adds important insights to the understanding of biomaterials. It begins with a comprehensive outline of the theoretical aspects of simulating the interface between water and solid surfaces. After this, small surface-bound biological molecules are modelled to explain experiments showing that they can capture cells on the surface. Finally, a new and practical, scalable technique for controlling biological molecules at the surface is developed. This work advances the field of biomaterials by explaining important processes that occur at the interface of biology and technology.

# Acknowledgements

No one is an island, and nor is any thesis. I owe my gratitude to many people, some of whom I would especially like to mention here. In no particular order, thankyou:

To my parents Pam and Errol, whose support made it possible in the first place and whose faith (and proofreading!) got this thesis over the line.

To my brother, who always answers complicated questions.

To Bek, whose encouragement made me believe I could do it. You make me want to be the best.

To my supervisor Marcela, who has an amazing ability to understand basically anything. My writing owes a lot to your teaching.

To my co-supervisor Serdar, whose expertise led me around many pitfalls.

To my co-supervisor Behnam, who dedicated so many hours in the ~~trenches~~ lab with me.

To Giselle Yeo, who taught me to run experiments.

To Ben Corry, who gave me a start in science.

# Statement of originality

This is to certify that, to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any other degree or other purposes.

Lewis J. Martin

Signature:

29/3/18

# Authorship attribution statement

## Chapters published as papers:

Chapter 3 of this thesis has been published in the journal *Interface Focus*. I performed the research and drafted the article.

Chapter 6 of this thesis has been published in the journal *Nature Communications*. I designed the peptide and the dual buffer system, created the figures, and performed the following experiments: simulations, peptide immobilization on the radical functionalized surface, enzyme-linked immunosorbent assay, and electrokinetic analysis. Dr. Behnam Akhavan performed the radical functionalized surface design and preparation and X-ray photoelectron spectroscopy. Time of flight secondary ion mass spectrometry was peformed by Dr. John Denman at the Australian Microscopy & Microanalysis Research Facility. I contributed equally with Prof. Marcela Bilek and Dr. Behnam Akhavan to analyse the results and write the manuscript.

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Lewis J. Martin

Signature:

29/3/18

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Prof. Marcela M. M. Bilek

Signature:

29/3/18

# List of abbreviations and acronyms

| | |
|---|---|
| $\varepsilon$ | Lennard-Jones parameter defining the depth of the potential energy minimum |
| $\theta$ | Water contact angle |
| $\sigma$ | Lennard-Jones parameter defining the distance where potential energy is zero |
| $\varphi$ | Backbone dihedral phi angle |
| $\psi$ | Backbone dihedral psi angle |
| .pdb | Protein database file |
| .psf | Protein structure file |
| Å | Ångström |
| AMBER | Assisted Model Building with Energy Refinement software |
| AMOEBA | Atomic Multipole Optimized Energetics for Biomolecular Applications software |
| CHARMM36 | Chemistry at Harvard Macromolecular Mechanics force field |
| CGenFF | CHARMM General Force Field |
| CPU | Central Processing Unit |
| dPCA | Dihedral angle Principal Component Analysis |
| ECM | Extra-Cellular Matrix |
| ELISA | Enzyme-Linked Immunosorbent Assay |
| EPR | Electron Paramagnetic Resonance spectroscopy |
| FF | Force Field |
| FMOC | Fluorenylmethyloxycarbonyl chloride |
| fs | Femtosecond |
| GROMOS | Groningen Molecular Simulation force field |
| GPU | Graphics Processing Unit |
| IDP | Intrinsically Disordered Peptide |
| $k_B$ | Boltzmann constant |
| LJ | Lennard-Jones |
| MD | Molecular Dynamics |
| NAMD | Not (just) Another Molecular Dynamics software |
| NMR | Nuclear Magnetic Resonance |

| | |
|---|---|
| ns | Nanosecond |
| OPLS | Optimized Potentials for Liquid Simulations force field |
| PCA | Principal Component Analysis |
| PC | Principal Component |
| PMF | Potential of Mean Force |
| PTFE | Polytetrafluoroethylene |
| ps | Picosecond |
| RFPP | Radical-Functionalized Plasma Polymer |
| RMSD | Root-Mean-Square Deviation |
| SAM | Self-Assembled Monolayer |
| TCL | Tool Command Language |
| TIP3P | Three-site Transferable Intermolecular Potential |
| ToF-SIMS | Time-of-flight Secondary Ion Mass Spectrometry |
| VMD | Visual Molecular Dynamics |
| XPS | X-ray Photoelectron Spectroscopy |

**List of figures**

Table of Contents

# Chapter 1. Introduction

An introduction to biomaterials and molecular simulations, explaining their uses and the motivation for this thesis. This chapter serves as a broad introduction to the field that outlines the fundamental problems encountered when attempting to simulate interactions of biomolecules with solid surfaces. All of the topics mentioned here, along with the specific knowledge gaps and how they are addressed, are covered in detailed studies in Chapters 3 to 6.

# 1.1 Introduction

Biomaterials are materials designed to interact with biological tissue. The modern form of biomaterials arose in the mid 20th century with the observation by an ophthalmologist that plastic shards, splintered from the cockpits of wartime aeroplanes, had persisted in the eyes of pilots without causing inflammation[1]. Since then, improvements in materials science and a deeper understanding of the biology of interfaces has led to implants and prostheses for a range of hard and soft tissues[2]. Modern biomaterials cross the boundary between implant and host, extending into the surrounding tissue to direct local cell growth and improve healing after surgery and injury[3]. Going further and inverting the original concept of an implanted biomaterial, decellularized organs can now be used to grow part or whole organs – in this case, the biomaterial is being implanted by biological tissue[4].

The science of biomaterials now influences many other fields, going far beyond tissue engineering. Implantable technologies are being used to induce immune responses as a cancer vaccine[5] or to overcome common biological barriers to drug delivery[6]. In chemical, fuel and food production, surface-immobilized enzymes show promise as effective, high throughput alternatives to the existing methods[7]. In biological research, scaled down organs made from inorganic material and biomolecules can improve research outputs using so-called labs-on-a-chip[8]. The original goal of biomaterial interfaces was to create a surface so inert it resists interaction with nearby cells. This led, for example, to the creation of anti-fouling coatings on ship hulls that resist microorganism growth and reduce friction and fuel consumption[9]. The new goal, however, is to create surfaces that actively integrate in the surrounding tissue.

There are difficult problems to solve to fully realise these applications. For instance, while titanium has emerged as an effective replacement for bone due to its bulk physical properties, there is an associated cost and many implants currently in use have poor biocompatibility[10]. Antifouling coating solutions, particularly those with biocidal properties, can leech into the ocean and lead to toxic environmental damage[11]. For covalent attachment, controlling the orientation of the attached biomolecule requires additional chemical steps[12]. For medical biomaterials, the ultimate obstacle is the response of the host

– after implanting into the body, biomaterials often cause a foreign body response that jeopardises the biomaterial's function. This occurs when biomolecules, such as proteins, come into contact with the biomaterial, are damaged in the process and induce the body's natural defence mechanism: inflammation[13].

This consideration is the broad problem in biomaterial science – how can we create biomaterials that perfectly integrate with biology? There have been two approaches to answering this question. The first uses treatment of the biomaterial surface to try to match its chemical, electrical, or physical properties to the intended biological target[14]. This is enough for applications in non-fouling materials or sometimes cell culturing where there is no immune response. The second approach addresses biointegration, which is the process of integrating with the surrounding tissue in an organism and requires a more nuanced approach. Extracellular adhesion proteins are the natural biomolecules that create linkages between cells, and reproducing the chemical signals that they provide to cells on the surface of the biomaterial can aid biointegration. These signals can be provided by whole proteins adsorbed onto the surface, but this is transitory since the attachment strength is weak[15]. A new, alternative, approach has been developed by Professor M. Bilek and others at the University of Sydney. Treatment of the surface by high energy, ionized gas creates reactive chemical species at the interface that strongly bind whole proteins, or synthetic derivatives called peptides[16].

The biomaterial interface with cell-adhesive signals is a complex environment that, due to its small size, is difficult to probe. Biomaterials interact with biology on multiple scales – the macro-scale for bulk physical properties, the micro-scale for cellular adhesion, and the nanoscale for chemical recognition. Yet, while the larger scales are amenable to spectroscopy, the small time and length behaviour at the nanoscale interface goes beyond most available probe techniques[17]. Here, molecular simulations are an effective alternative because they model the behaviour of proteins in atomic detail[18].

A chemical system such as a biomolecule surrounded by liquid water (and perhaps a material surface) can be modelled by computers using algorithms based on classical physics. In molecular dynamics simulations, each atom is represented as a single particle with x-, y-, and z-coordinates within a nanoscale volume of space that comprises the model system. At every stage of the simulation the

forces on each atom are calculated and their positions in space in response to these forces are recalculated accordingly. The process of calculating forces and updating coordinates continues iteratively, frame by frame, to reproduce the motion of the protein. Since the pressure and temperature are maintained at experimental conditions, and the algorithms relate the force to time, the motions closely resemble the true behaviour of the biomolecule in water at small timescales.

For biomaterial science, the most advanced molecular dynamics simulation packages available today have one shortcoming – their treatment of the interface can be inaccurate. When atoms are represented as single particles, the atomic charges across a molecule are set as fixed values. This modelling approximation reduces computational calculation time, allowing for long simulations with better statistical accuracy, and works well for biomolecule-in-water systems[19].

In real molecular systems, however, charge across a molecule constantly varies in response to the electrical environment. In particular, at solid-liquid interfaces, the sharp change in the electrical environment means this approximation is not applicable. For biomaterials, where the interface is a crucial part of proper function, using such simulations requires extra care.

This thesis investigates biomaterials by examining the interaction between synthetic material surfaces and biological molecules. Advances are made by: devising a way to simulate weakly interacting surface interfaces without the need for including polarizability; using the simulations of peptide-surface interactions to explain intriguing experimental results in a biointerfacial system; and developing a new paradigm for controlled peptide immobilisation using both simulations and experimental techniques. To achieve these outcomes, this thesis employs multiple techniques: simulations to characterize the interactions in atomic detail, and experimental techniques to understand the macroscale effects.

Chapter 2 provides an overview of simulation methods as applied to molecules in solution, as well as details of the specific simulations methods that are used in later chapters. This chapter is an introduction for readers who are not familiar with molecular modelling simulations. Chapter 3 is a literature review and has in-depth treatment of these techniques, paying particular attention to how they relate to simulating interactions between surfaces and biomolecules. The chapter, which is included in a

published manuscript format, reviews the field to uncover the obstacles that need to be overcome, and recommends a strategy for parameterizing surface models. Chapter 4 discusses the specific problems with biomaterial integration and implements the strategy developed in Chapter 3 to develop a computational model of a surface that is then used to simulate the surface interactions of biomolecules. These simulations offer a viable explanation for cell adhesion trends previously observed in experiment. This addresses half of a biomolecule functionalised biomaterial's function – the biomolecule in contact with a surface. In Chapter 5, the interaction of these biomolecules with their receptor on the cell is rigorously discussed and explored using simulation, completing the link between the surface and the cell. Combining the information from all of these chapters, Chapter 6 uses simulation and experiment to report the development of a new method to control biomolecules at the interface, potentially paving the way to effective biomaterials in future. Finally, Chapter 7 concludes the thesis by putting the discoveries in context.

Overall, this thesis scales from mostly theoretical to mostly experimental. At each stage, the focus is on the interface between synthetic materials and biology and, in particular, how nanoscale interactions affect biomaterial function.

## 1.2 References


1 Apple, D. J. & Sims, J. Harold Ridley and the invention of the intraocular lens. *Survey of ophthalmology* **40**, 279-292 (1996).

2 Scholz, M.-S. *et al.* The use of composite materials in modern orthopaedic medicine and prosthetic devices: A review. *Composites Science and Technology* **71**, 1791-1803 (2011).

3 Hubbell, J. A. Biomaterials in tissue engineering. *Nature Biotechnology* **13**, 565-576 (1995).

4 Crapo, P. M., Gilbert, T. W. & Badylak, S. F. An overview of tissue and whole organ decellularization processes. *Biomaterials* **32**, 3233-3243 (2011).

5 Gu, L. & Mooney, D. J. Biomaterials and emerging anticancer therapeutics: engineering the microenvironment. *Nature Reviews Cancer* **16**, 56 (2016).

6 Blanco, E., Shen, H. & Ferrari, M. Principles of nanoparticle design for overcoming biological barriers to drug delivery. *Nature biotechnology* **33**, 941 (2015).

7 Rodrigues, R. C., Ortiz, C., Berenguer-Murcia, Á., Torres, R. & Fernández-Lafuente, R. Modifying enzyme activity and selectivity by immobilization. *Chemical Society Reviews* **42**, 6290-6307 (2013).

8 Craighead, H. Future lab-on-a-chip technologies for interrogating individual molecules. *Nature* **442**, 387 (2006).

9 Yebra, D. M., Kiil, S. & Dam-Johansen, K. Antifouling technology—past, present and future steps towards efficient and environmentally friendly antifouling coatings. *Progress in organic coatings* **50**, 75-104 (2004).

10 Nasab, M. B., Hassan, M. R. & Sahari, B. B. Metallic biomaterials of knee and hip-A review. *Trends Biomater. Artif. Organs* **24**, 69-82 (2010).

11 Thomas, K. & Brooks, S. The environmental fate and effects of antifouling paint biocides. *Biofouling* **26**, 73-88 (2010).

12 Hoarau, M., Badieyan, S. & Marsh, E. N. G. Immobilized enzymes: understanding enzyme–surface interactions at the molecular level. *Organic & biomolecular chemistry* **15**, 9539-9551 (2017).

13 Ratner, B. D. & Bryant, S. J. Biomaterials: where we have been and where we are going. *Annu. Rev. Biomed. Eng.* **6**, 41-75 (2004).

14 Elbert, D. L. & Hubbell, J. A. Surface treatments of polymers for biocompatibility. *Annual Review of Materials Science* **26**, 365-394 (1996).

15 Wilson, C. J., Clegg, R. E., Leavesley, D. I. & Pearcy, M. J. Mediation of biomaterial–cell interactions by adsorbed proteins: a review. *Tissue engineering* **11**, 1-18 (2005).

16 Bilek, M. M. *et al.* Free radical functionalization of surfaces to prevent adverse responses to biomedical devices. *Proceedings of the National Academy of Sciences* **108**, 14405-14410 (2011).

17 Buehler, M. J. & Genin, G. M. Integrated multiscale biomaterials experiment and modelling: a perspective. *Interface focus* **6**, 20150098 (2016).

18 Dror, R. O., Dirks, R. M., Grossman, J., Xu, H. & Shaw, D. E. Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics* **41**, 429-452 (2012).

19 Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC biology* **9**, 71 (2011).

# Chapter 2. General methods

This thesis makes extensive use of molecular dynamics (MD) simulations. This chapter is a general introduction to these simulations, as well as detailing the methods that are common to all the results chapters. In depth treatment of molecular dynamics simulations, and in particular how they relate to simulating biointerfaces, is given in the literature review in Chapter 3. The current chapter also provides some methodological discussion that is important for the following chapters but, in order to present them in context, the specifics and discussion of those methods will be given in the respective chapters.

## 2.1 Molecular dynamics simulation

As explained in the introduction, MD simulations model atoms to generate physically relevant states that often are not measurable by experiment[1]. Each simulation consists of a cell of atoms. Atoms are composed of subatomic particles – protons and neutrons in the nucleus, and electrons orbiting this nucleus. Modelling these explicitly leads to an enormous increase in the computational demand of any simulation, due to the complexity of solving the quantum chemical equations governing their movement[2]. To reduce the computational demand, most biological macromolecule simulations utilise the Born-Oppenheimer approximation, in which the motion of electrons (which are approximately 2000 fold lighter in mass than the particles in the nucleus) is assumed to be so fast compared to the motion of the nucleus that the entire atom can be modelled as a single unit[3]. While there are techniques for modelling electron movement through time, labelled *ab initio* molecular dynamics[4], the computational expense typically limits them to timescales and system sizes below those required for most macromolecules.

At each step in a molecular dynamics simulation, the potential energy of every atom is calculated using an algorithm based on interatomic interactions. This is then used to determine the force on each atom. Finally, integrating Newton's equations of motion leads to an estimation of new coordinates for each particle after a small interval of simulated time. The timestep most commonly used in biomolecular modelling is 1-2 femtoseconds, which is based on the fastest motion of the underlying degrees of freedom – with the Born-Oppenheimer approximation, the fastest movement is bond vibration of hydrogen atoms (~10 fs)[5]. Fixing the fast degrees of movement can allow for longer timesteps, but this then leads to discretization error that arises from taking finite approximations to the equations of motion[6]. The coordinates of the atoms are then updated and the process is repeated. This time-step forms a single unit of a MD simulation, and is repeated as many times as necessary to generate physically relevant information about the atomic system, such as the configurational space of a protein, equilibrium properties like thermodynamic quantities, or dynamic behaviour[7]. There are several nuances, relevant to this thesis, that are worth mentioning. Other well established aspects of molecular simulations are covered in recommended textbooks[8,9].

Bounding the atomic system by hard barriers would introduce edge-effects, so simulations take advantage of periodic boundary conditions. Using these, a simulation system consists of a geometric shape such as a cube. When the coordinates of a particle move outside of the cube, it re-appears on the opposite face. The effect of these conditions is that a simulation ultimately represents an infinitely extended system consisting of repeated blocks of atoms. The forces between atoms are applied across the periodic boundary to seamlessly integrate each block. Periodic boundary conditions are a useful approximation to remove edge effects but, like all approximations, their use is not without risk. Recent modelling of haemoglobin suggests that the hydrophobic effect (the tendency for water to exclude non-polar molecules) can extend to large distances, requiring box sizes far greater than what is typically used in protein simulations[10]. This likely affects most simulations of protein conformations, which involve large changes in the surface area accessible to solvent.

Correct treatment of the simulation box also extends to controlling the temperature and pressure. While typically the number of atoms in a simulation will not change through time, the temperature and pressure can vary wildly without appropriate controls. In the canonical ensemble (aka NVT), the number of particles (N), the volume (V), and the temperature (T) are set in advance[11]. The temperature is maintained by coupling to a fictional heat bath (a 'thermostat') that periodically changes the velocities of the particles in accordance with the defined temperature. Some so-called velocity rescaling algorithms do not maintain a correct relationship between the temperature of the system and the average energies. This has been rigorously demonstrated recently for the Berendsen thermostat, suggesting it should be avoided[12]. Another ensemble, NPT, maintains constant pressure by varying the size of the simulation box. Such simulations use a barostat, which analogously to thermostats, maintains constant pressure[13]

There are several components to the potential energy function governing atomic forces. Simulations in this thesis use the potential energy function as defined by the NAMD software[14]. Intramolecular energy terms account for the vibration and rotation of atomic bonds within a molecule. Intermolecular energy terms approximate the interaction between electron clouds in non-bonded atoms. These interactions can form a significant portion of the calculations of a molecular dynamics simulation, because the interaction between every pair of atoms needs to be accounted for[15]. When atoms are approximated

as single particles, the electron distribution is treated as a fixed partial charge (covered in detail in Chapter 3). Charged particles interact with each other following Coulombs law, which relates the force between two particles to the magnitude of their charge. Another intermolecular force is the Lennard-Jones (LJ) potential that approximates van der Waals interactions, which is the attractive force felt by molecules as a result of induced dipoles from the movement of their electron clouds. The LJ potential relates the potential energy of two atoms to the distance between them. Since Chapter 4 discusses this thoroughly, the LJ potential is reproduced in Figure 1**.** The parameters of interest are $\varepsilon$ and $\sigma$, which are the depth of the minimum and the distance at which the potential energy is zero, respectively. In this thesis, the potential energy function is calculated using the CHARMM27 force field values, which are parameterized to model biomolecules[16].

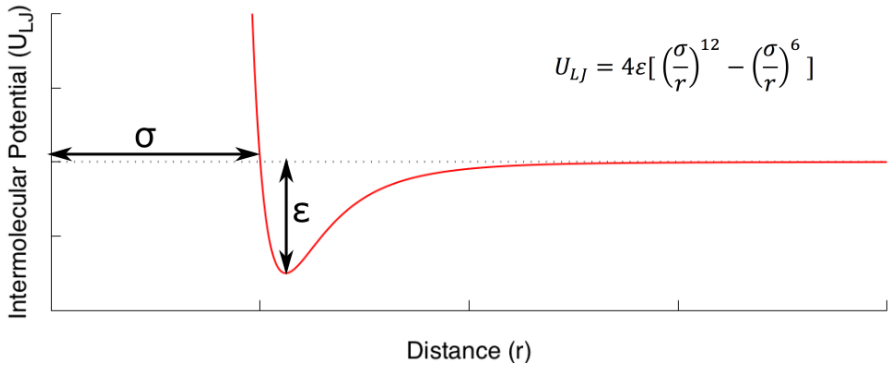



**Figure 1.** The Lennard-Jones potential approximating van der Waals interactions. Two atoms experience a repulsive force at close distances (less than $\sigma$). The strength of the energetic minimum, when atoms are attracted, is defined by $\varepsilon$.

In comparison to intramolecular forces, which are for nearest neighbour bonded atoms only, the intermolecular forces are time-consuming calculations in MD. For a system size of $N$ particles, each step requires calculating $N \times N$ pairwise interactions. To reduce the computational load, both electrostatic and LJ interactions are usually cut off after a certain distance (usually around 1 – 1.5 nm)[15]. The correct treatment of intermolecular cut-offs is beyond the scope of this chapter, but one consequence is that the cut-off length determines the minimum size of a simulation cell. System sizes below two times the cut-off distance means atoms can interact with periodic images of themselves. To prevent this, most system sizes are at least the length of the longest molecule, with an additional buffer of the cut-off length in each of the X, Y, and Z axes.

These approximations reduce the time taken to calculate a single time-step. With these, MD simulations are well suited to simulations of biomolecules like peptides and proteins. Such systems, with a biomolecule surrounded by water and ions, span a range of about 30,000 to 300,000 atoms with timescales of hundreds or thousands of nanoseconds attainable. The scale of simulations is heavily dependent on the computer resources available, with many biomolecular systems requiring highly parallelized computer facilities. The resources at these facilities can reduce the time to perform a simulation by a factor of several hundred.

## 2.2 Advanced Simulation Techniques

The simulation setting described above is known as equilibrium dynamics. This term refers to thermodynamic equilibrium, where no extrinsic bias is applied that could alter the system. These simulations can generate physically relevant states of biomolecules, but often the most biologically interesting states take a long time to arise without external input. This is because unbiased simulations are free to wander through irrelevant phase-space, without discovering new, physically relevant, states. To remedy this, an applied bias can direct the simulation toward a desired state. Subsequently removing the bias, and calculating the energy change between the starting and end states, gives access to an important quantity called free energy.

Free energy is a thermodynamic quantity that represents the proportion of time spent in a particular state. Calculating free energy values by simulation is a common way to verify that they are physically accurate, since it can be compared to experimental values. It is also used to compare the preference of a biomolecule for different biologically relevant states. In theory, an equilibrium simulation could generate accurate free energy values by gathering enough positional data, but the timescales required are usually not practical, so biases are preferred. Such biases have been called ‘molecular tweezers’, since they direct the molecules in a simulation toward or away from a given state[17]. One such method is based on probability distributions. By applying a bias to the potential energy function that controls the position of molecules in a simulation intermediate states, that are generally inaccessible due to being high energy, can be explored in detail. Performing several simulations with slight shifts in the position of the bias leads to overlapping descriptions of a desired reaction pathway[18]. Subsequently removing

the bias can yield the underlying changes in free energy[19]. Another variant is methods based on so-called alchemical transformations. These are analogous to positional biases, but instead the chemical identity of the molecules in the simulation are altered. In this case the intermediate states are nonphysical, but these methods can still yield a free energy difference between the end states. These approaches are covered in detail in [20].

One common quantity measured by free energy values is the strength of binding. Here, the bound and the unbound states are the two endpoints being compared. The free energy change in the intermediate regions can also be calculated, generating a potential of mean force (PMF). Strictly defined as the potential that gives the average force (calculated as the negative of the PMF gradient) over all possible configurations of the system between the two states, it can be understood intuitively as measuring the strength of attraction to the bound state.

This thesis uses two methods of calculating PMFs. The first is umbrella sampling. Here, a restraining potential is applied to an atom to influence its position. The position of the restraint is shifted in separate simulations, forcing the system to explore all configurations between the bound and unbound states. The attraction to the bound state causes deviations from the restrained positions, which are related to the PMF[18]. The second method of calculating PMFs uses the Jarzynski equality[21]. In this method a moving restraint is applied to an atom, pulling it from the bound to the unbound state. The resistance of the atom against the restraint reflects the energetic change. By repeating this process several times using different starting coordinates for the water, the friction of water can be removed to yield a PMF. Both methods are capable of calculating PMFs, but they require different levels of computer resources. In general, umbrella sampling is more accurate while the Jaryznski equality is well-suited to quickly calculate relative, not exact, binding strengths[22].

## 2.3 Method details

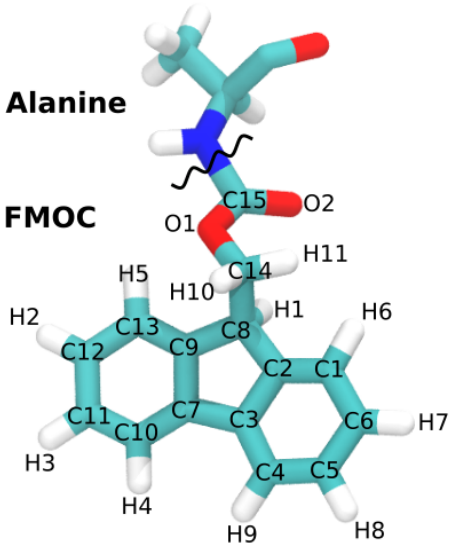
The simulations in this thesis use some tools and parameters that are common to all chapters. To prevent repetition across chapters, the details of those methods are presented here.

### 2.3.1 FMOC parameterization

Chapters 4 and 5 present MD simulations of a peptide with an attached FMOC group. FMOC is used as a chemical blocking group during peptide synthesis, but can also function as a hydrophobic moiety to influence peptide behaviour[23]. There are no simulation parameters for FMOC included in the CHARMM27 force field, so new ones were determined. Parameterization used the CGenFF server to determine partial charges, LJ parameters, and bond parameters of the FMOC group[24]. These parameters were then added as a patch to the alanine residue in the CHARMM27 force field. The bonds between the alanine backbone and the FMOC group were parameterized by analogy to similar bonds in the force field.

Automatic parameterization of small molecules can be successful but the results require scrutiny to ensure they are chemically reasonable. To guide researchers using the CGenFF, the results are associated with penalty scores indicating the chance of poor parameter fitting. In the case of FMOC, several of the partial charges were associated with high penalty scores. The aromatic carbons in the fluorene group were assigned partial charges of up to -0.3 e. This contrasts with the typical aromatic partial charges in the CHARMM forcefield for aromatic carbons of -0.115 e (with the associated hydrogens being 0.115 e). For a hydrophobic group such as FMOC, the interaction with peptide moieties and the water phase is crucial to generating accurate modelling results. The partial charges as originally assigned by CGenFF would likely lead to a more hydrophilic FMOC group than chemical intuition would suggest, so the aromatic carbons were changed back to the standard -0.115 e (Figure 2). Subsequently, an extensive parameterization protocol was published for FMOC moieties for use with CHARMM forcefields[25]. These parameters were validated by free energy calculations and the formation of supramolecular assemblies, which are exquisitely sensitive to the partial charges. The close agreement to the partial charges used here, being within ~95% accuracy for the aromatic carbons, and ~85% accuracy for the aromatic group hydrogens, justifies this chemical intuition approach. The other moiety likely to have an effect on the behaviour of the FMOC group is the two oxygens. These charges are 55% and 72% of the equivalent charges in the published parameters, which is suitable for conformational analysis but may lead to errors when performing free energy calculations (which are not performed here).

a




b

| Atom Name | Partial Charge | Atom Name | Partial Charge |
|---|---|---|---|
| C1 | -0.115 | C15 | 0.848 |
| C2 | 0 | O1 | -0.515 |
| C3 | 0 | O2 | -0.615 |
| C4 | -0.115 | H1 | 0.09 |
| C5 | -0.115 | H2 | 0.115 |
| C6 | -0.115 | H3 | 0.115 |
| C7 | 0 | H4 | 0.115 |
| C8 | -0.058 | H5 | 0.115 |
| C9 | 0 | H6 | 0.115 |
| C10 | -0.115 | H7 | 0.115 |
| C11 | -0.115 | H8 | 0.115 |
| C12 | -0.115 | H9 | 0.115 |
| C13 | -0.115 | H10 | 0.09 |
| C14 | 0.07 | H11 | 0.09 |

**Figure 2.** FMOC partial charges. a. The FMOC cofactor, showing the atom names and point of attachment to alanine. b. The partial charges for the atoms in the FMOC cofactor.

### 2.3.2 Peptide generation

Chapters 4, 5 and 6 include simulations of peptides. These peptides were generated as .pdb files using the Avogadro peptide builder using a linear conformation. .psf files were then generated using the psfgen software available in the VMD software[26]. Simulations were begun by setting the atomic velocities of all atoms to create a system at a temperature of 298 K. Their high flexibility justified this approach, since they quickly equilibrated to an unfolded state, and had no enduring conformations that could be influenced by the starting state. Each chapter describes the specifics of the simulation systems used therein.

### 2.3.3 Simulation parameters

In each of the results chapters (4, 5, and 6), the molecular dynamics simulations used NAMD 2.11[14]. Periodic boundary conditions were applied, with constant pressure of 1 atm maintained using a Langevin Piston and constant temperature of 298 K maintained using Langevin dynamics to simulate in the NPT ensemble. The CHARMM27 force field was used for protein and water[16], with ions using parameters from Joung and Cheatham[27]. The aqueous phase was represented by TIP3P water[28] and was generated using the Solvate plugin of VMD. Rigid bonds were turned on, fixing the bond distance between hydrogen atoms and their parent heavy atoms[29]. This allowed for a time-step of 2 fs, doubling

the sampling rate compared to simulations without fixed hydrogens. System coordinates were stored every 5 ps to facilitate analysis. The non-bonded cut-off was set to 1.2 nm.

## 2.4 References


1 Dror, R. O., Dirks, R. M., Grossman, J., Xu, H. & Shaw, D. E. Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics* **41**, 429-452 (2012).

2 Hofer, T. S. From macromolecules to electrons—grand challenges in theoretical and computational chemistry. *Frontiers in chemistry* **1**, 6 (2013).

3 Bechstedt, F. in *Many-Body Approach to Electronic Excitations* 3-11 (Springer, 2015).

4 Marx, D. & Hutter, J. Ab initio molecular dynamics. *Parallel computing* **309**, 327 (2009).

5 Feenstra, K. A., Hess, B. & Berendsen, H. J. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry* **20**, 786-798 (1999).

6 Leimkuhler, B. J., Reich, S. & Skeel, R. D. in *Mathematical Approaches to biomolecular structure and dynamics* 161-185 (Springer, 1996).

7 Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural and Molecular Biology* **9**, 646 (2002).

8 Frenkel, D. & Smit, B. *Understanding molecular simulation: from algorithms to applications*. Vol. 1 (Elsevier, 2001).

9 Allen, M. P. & Tildesley, D. J. *Computer simulation of liquids*. (Oxford university press, 2017).

10 El Hage, K., Hédin, F., Gupta, P. K., Meuwly, M. & Karplus, M. Valid molecular dynamics simulations of human hemoglobin require a surprisingly large box size. *eLife* **7**, e35560 (2018).

11 Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics* **72**, 2384-2393 (1980).

12 Braun, E., Moosavi, M. S. & Smit, B. Anomalous effects of velocity rescaling algorithms: the flying ice cube effect revisited. *arXiv preprint arXiv:1805.02295* (2018).

13 Chow, K.-H. & Ferguson, D. M. Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. *Computer physics communications* **91**, 283-289 (1995).

14 Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **26**, 1781-1802 (2005).

15 Sagui, C. & Darden, T. A. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual review of biophysics and biomolecular structure* **28**, 155-179 (1999).

16 Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ1 and χ2 dihedral angles. *Journal of chemical theory and computation* **8**, 3257-3273 (2012).

17 Chipot, C. Frontiers in free-energy calculations of biological systems. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**, 71-89 (2014).

18 Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry* **13**, 1011-1021 (1992).

19 Grossfield, A. *WHAM: the weighted histogram analysis method*, <http://membrane.urmc.rochester.edu/content/wham> (2016).

20 Chipot, C. & Pohorille, A. *Free energy calculations*. (Springer, 2007).

21 Jarzynski, C. Nonequilibrium equality for free energy differences. *Physical Review Letters* **78**, 2690 (1997).

22 Baştuğ, T. & Kuyucak, S. Application of Jarzynski's equality in simple versus complex systems. *Chemical Physics Letters* **436**, 383-387 (2007).

23 Jayawarna, V. *et al.* Introducing chemical functionality in Fmoc-peptide gels for cell culture. *Acta biomaterialia* **5**, 934-943 (2009).

24 Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* **31**, 671-690 (2010).

25 Sasselli, I. R., Ulijn, R. & Tuttle, T. CHARMM force field parameterization protocol for self-assembling peptide amphiphiles: the Fmoc moiety. *Physical Chemistry Chemical Physics* **18**, 4659-4667 (2016).

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

27 Joung, I. S. & Cheatham III, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The journal of physical chemistry B* **112**, 9020-9041 (2008).

28 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926-935 (1983).

29 Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry* **13**, 952-962 (1992).

# Chapter 3. Simulating surfaces and biomolecules

Molecular dynamics simulations can generate physically relevant states of biomolecule systems in atomic detail. There is a need for more detailed information of biomaterial systems and molecular dynamics simulations are an ideal solution, but this requires introducing a surface into the simulation. When using the fixed-charge approximation polarization effects are not modelled, so care must be taken in the model formulation. Some researchers have modified the force field to include polarization, although this requires validation and adds to the computational load. This chapter reviews molecular dynamics simulations of surfaces and discusses the process of model parameterization, suggesting that the fixed-charge approximation is appropriate in certain cases. It lays the groundwork for the results chapters that follow.

This chapter has been published in *Interface Focus*.

INTERFACE FOCUS

rsfs.royalsocietypublishing.org

Review

**Cite this article:** Martin L, Bilek MM, Weiss AS, Kuyucak S. 2016 Force fields for simulating the interaction of surfaces with biological molecules. *Interface Focus* **6**: 20150045. http://dx.doi.org/10.1098/rsfs.2015.0045

One contribution of 19 to a theme issue 'Integrated multiscale biomaterials experiment and modelling: towards function and pathology'.

**Subject Areas:**
biomaterials, biotechnology, computational biology

**Keywords:**
biointerface, molecular dynamics, parametrization, surface, peptide

**Author for correspondence:**
Lewis Martin
e-mail: lewis@physics.usyd.edu.au

# Force fields for simulating the interaction of surfaces with biological molecules

Lewis Martin[1], Marcela M. Bilek[1], Anthony S. Weiss[2,3,4] and Serdar Kuyucak[1]

[1]Department of Applied Physics, [2]Charles Perkins Centre, [3]Department of Molecular Bioscience, and [4]Bosch Institute, University of Sydney, Sydney, New South Wales, Australia

The interaction of biomolecules with solid interfaces is of fundamental importance to several emerging biotechnologies such as medical implants, anti-fouling coatings and novel diagnostic devices. Many of these technologies rely on the binding of peptides to a solid surface, but a full understanding of the mechanism of binding, as well as the effect on the conformation of adsorbed peptides, is beyond the resolution of current experimental techniques. Nanoscale simulations using molecular mechanics offer potential insights into these processes. However, most models at this scale have been developed for aqueous peptide and protein simulation, and there are no proven models for describing biointerfaces. In this review, we detail the current research towards developing a non-polarizable molecular model for peptide–surface interactions, with a particular focus on fitting the model parameters as well as validation by choice of appropriate experimental data.

## 1. Introduction

The physical interface between biological tissue and abiotic surfaces has long been exploited in nature, and it is a major determinant of the performance of a wide range of modern technologies. For example, in marine environments biofouling of ships can become a significant economic burden, and there is a demand for a cheap and non-toxic anti-fouling coating [1]. Implantable medical devices are often associated with poor tissue integration due to bio-incompatibility, leading to scarring, sub-optimal performance and in some cases rejection. The idea that these adverse medical outcomes could be remedied with a biofunctionalized interface [2] is driving a large sector of modern biomaterials research. Biomolecule–surface interactions are also being explored on gold nanoparticles for cancer diagnostics [3] and on polymers to construct protein screening microarrays [4]. Yet, there is still debate about exactly what properties of a surface make it biocompatible or not—for example, the hydrophilicity [5] and surface topology [6] have been shown to strongly influence the behaviour of proteins and cells on surfaces. Our incomplete understanding of the roles, both individual and in combination, of chemical and physical structures of surfaces makes the rational design of optimum biological interfaces challenging. A better understanding of how biology interacts with surfaces is needed to facilitate progress in the field.

For mammalian cells, adhesion occurs through receptor proteins that are normally used to bind to specific external motifs such as protein constituents of the body's extracellular matrix (figure 1). Hence the question of how proteins react to different surfaces is central to the problem. Proteins tend to bind weakly to hydrophilic surfaces, but can maintain their native conformation, whereas they typically unfold and adhere strongly to hydrophobic surfaces [8]. This is because energy minimization of the surface protein system is favoured by contact of the hydrophobic interior segments of the protein to the hydrophobic surface. Unfolding or denaturation of proteins at surfaces needs to be avoided as unfolding proteins signal the presence of a foreign body to the host tissues and initiate harmful immune reactions and foreign body responses [9]. Molecules from the extracellular matrix can be incorporated into the surface to mimic mammalian tissues [10] and provide appropriate biological signals to cells at the interface. Although immobilization of extracellular matrix proteins has been shown to be beneficial [11] for promoting healthy cell integration attention is

THE ROYAL SOCIETY PUBLISHING

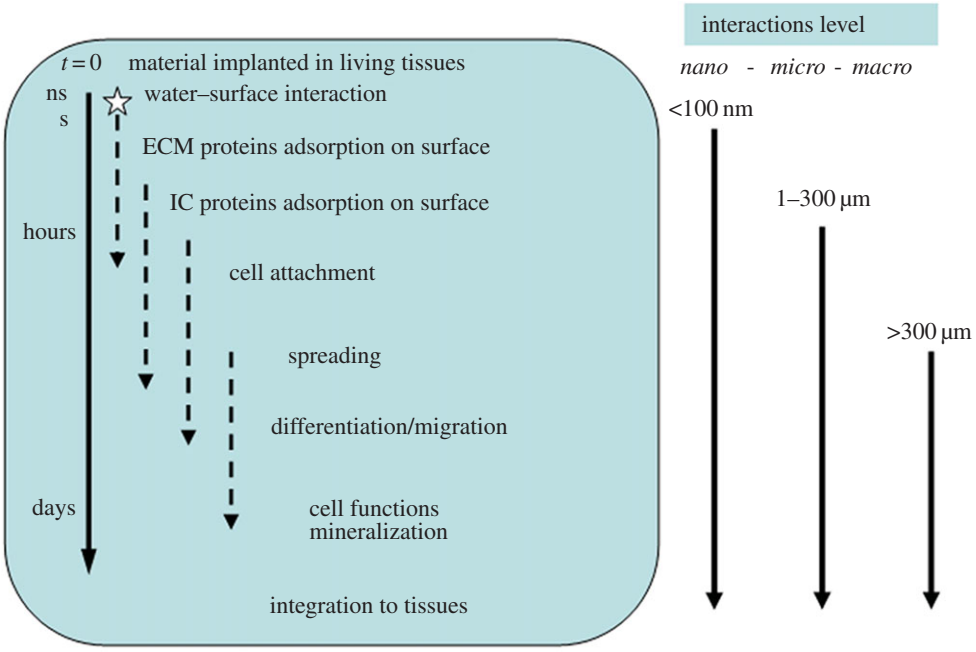



**Figure 1.** A schematic of the multiscale processes controlling the interaction of biological tissue with biomaterials. Adapted with permission from [7], copyright © 2007 Springer. (Online version in colour.)

shifting to peptides and protein segments. Biomimetic peptides are preferable to whole proteins in manufacturing environments because they can be cheaply synthesized, require less demanding purification and are more robust during processing and handling than entire protein molecules that need to maintain fragile conformations. The rapidly growing body of knowledge in biochemistry that ascribes functions performed by protein molecules to particular segments and peptides is opening up the possibility of driving biological responses through surface immobilized fragments/peptides that target single recognition sites [12] with fewer non-specific effects.

Progress in this field will require a deeper understanding of the adsorption of peptides to surfaces. Different-sized silica nanoparticles, for instance, can attract peptides with only 20% sequence similarity [13] and the mechanism for different specificity is unclear. Experimental methods operating at their current limits of resolutions are largely inadequate in this sphere. Single molecule force microscopy, one of the highest resolution experimental probes available, can measure the interactions of whole peptides, but cannot interrogate the roles of individual amino acids [14] in the process. While there have been recent advances in experimental studies of surface-adsorbed protein structure [15] and conformation [16] using ensemble averaging techniques such as nuclear magnetic resonance (NMR) and circular dichroism, nanoscale modelling is proving to be an invaluable complementary tool as it is able to achieve a resolution higher than any experimental method currently available.

Recognizing the absence of an 'ultimate microscope' with the capability to measure the position and identity of individual atoms at femtosecond speed, atomic-scale resolution is only available by using molecular modelling simulations. However, making the simulations computationally tractable typically requires some assumptions and some restrictions in the number of degrees of freedom of the system being modelled. Theoretically, the most detailed model describes the quantum mechanical (QM) nature of the valence electrons that take part in forming covalent bonds within molecules and surfaces. *Ab initio* approaches, such as density functional theory, approximate solutions to the Schrödinger equation and can be used to determine electron densities within molecules [17]. The computational effort of QM modelling methods limits their application in practice to systems of up to hundreds of atoms. These approaches do not scale well computationally so the simulation of a solvated peptide interacting with a surface (tens of thousands of atoms) is not yet feasible at this level of accuracy. By treating electron density implicitly as a force field (FF) between atoms, a significantly less computationally intensive molecular model can be formulated using Newton's equations of motion to govern atomic interactions [18]. The degrees of freedom can be further reduced by using coarse-grained models [19] (where groups of atoms are treated as a single particle) or with continuum models [20] (where water is represented by a continuous dielectric medium rather than individual molecules). These methods reduce the computational load and increase the system size scale and timescale that can be explored by compromising on resolution.

While coarse-grain and continuum methods are suitable for modelling the behaviour of bulk materials, accurate calculation of molecular interaction energies from a single peptide molecule requires an all-atom representation, meaning all atoms are accounted for individually. All-atom simulations are based on molecular mechanics, which is a method of generating physically realistic conformations of a system made up of atoms bound into molecules. These conformations are generated by using a potential energy function, known as an FF, which can be used to calculate the force on each atom at small time intervals in order to generate a time evolution of conformations. Within any FF, there are two considerations to gauge the accuracy of a simulation: sampling and parametrization. Sampling refers to how closely the conformations represent a Boltzmann-weighted ensemble average, and thus measures the confidence in their thermodynamic accuracy, and for protein–surface interactions this has been reviewed thoroughly by Latour [14]. Parametrization is the process of

assigning accurate values for atomic properties such as charge or bond length to the atoms within a simulation.

Several FFs already exist that have rigorously determined parameter sets for aqueous protein simulations—the CHARMM [21], AMBER [22], GROMOS [23] and OPLS [24] FFs are among the best known, and thus well tested, and they have been shown to fairly accurately reproduce aqueous protein dynamics. In this review, we focus mainly on the CHARMM FF, because it has been used in many of the efforts to develop new FFs for simulating proteins and solid surfaces. In molecular dynamics simulations, there is a time evolution of a molecular system whereby integration of the potential energy function determines the force on each atom at small time intervals (femtoseconds) and allows for iterative updates of their positions. The biophysical FFs mentioned above all use a similar potential energy function for these calculations. Because of its prevalence in parametrizing interfacial systems, the potential energy function as it is expressed in the CHARMM FF is an appropriate choice to demonstrate this method. The potential energy function in the CHARMM FF is

$$U_{\text{total}} = \sum_{\text{bonds}} K_b(b-b_0)^2 + \sum_{\text{angles}} K_\theta(\theta-\theta_0)^2 + \sum_{\text{UB}} K_{\text{UB}}(S-S_0)^2 + \sum_{\text{dihedral}} K_\chi(1+\cos(n\chi-\delta)) + \sum_{\text{impropers}} K_{\text{impr}}(\varphi-\varphi_0)^2 + \sum_{\text{LJ}_{i\neq j}} \varepsilon_{ij}\left[\left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{6}\right] + \sum_{\text{coulomb}} \frac{q_i q_j}{\varepsilon_i r_{ij}},$$

where the adjustable intramolecular (bonded) parameters are $b$ (bond length), $\theta$ (bond angle), $S$ (Urey–Bradley), $\chi$ (bond rotation), $\varphi$ (improper term for planar ring moieties). For intermolecular (non-bonded) interactions, van der Waals interactions are modelled with the Lennard-Jones (LJ) potential [25] with parameters $\epsilon$ for well depth, $R_{\min}$ for the point of minimum energy, while coulombic attraction is calculated using partial charge $q$. The subscript '0' refers to the equilibrium value. In this paradigm, atoms of an element in a given chemical environment are classified as an 'atom type' that shares values of these parameters characteristic to that environment. A change in environment will necessitate either new parameters, or that the original parameter set incorporates the average of both environments. Most protein FFs today have fixed values for the partial charge and are thus referred to as 'additive' FFs, as the total electrostatic potential energy is the sum of all the two-body potentials. Non-additive FFs (also known as polarizable FFs), in which the molecular charge distribution can change in response to the local environment, are also being developed [26]. While additive FFs are computationally less expensive they perform best in one particular dielectric medium, which will be illustrated later using water behaviour at phase interfaces. Another consideration for the success of any biophysical FF is the rigorous parametrization of atom types and molecules for the potential energy function, which ensures that it is both *balanced* and properly *validated*.

A *balanced* FF can accurately reproduce molecular behaviour at the atomic scale, while also being capable of reproducing macroscale thermodynamic quantities. This is as important for simulating small-molecule drugs binding to proteins as it is for peptides binding to surfaces, because a binding event is an emergent property that is the sum of not only the individual atomic interactions (an enthalpic quality) but also the entropic change in bulk water (if the net change favours binding, this is known as the hydrophobic effect). Balance is ingrained into an FF at the parametrization stage by carefully choosing a broad range of experimental measures to use as target data for fitting the parameters. In the CHARMM FF, this means satisfying structural data like QM torsional potential energy scans as well as bulk properties like the enthalpy of vaporization. The non-trivial nature of the parameter solution means this is a laborious process with a lot of scope for parameter correlation—where errors in a parameter set cancel, allowing the erroneous parameters to nevertheless fit the target data. A multiscale set of target data can alleviate this issue.

When a parameter set for a molecule is determined, its utility must be evaluated by reproducing some experimental measure that is outside the training set—this is known as *validation*. Ideally, validation should integrate many degrees of freedom to demonstrate their aggregate behaviour—for the protein FFs, folding into native conformations relies upon all of the bond vibrations and rotations, electronic charges and implicitly treated dispersion (van der Waals) forces. The validation of the CHARMM lipid parameter set followed a similar process [27]. For solid surfaces with limited dynamic behaviour, the validation must rely mostly on the non-bonded parameters—combining the effect of both the enthalpic potential and the entropic effect of surface-adsorbed water structure.

For interfacial systems, parametrizing an FF to accurately represent these contributions is difficult because of the extra attraction afforded by polarization on the surface due to charged regions of the peptide [28]. Polarization effects have to be handled implicitly in additive FFs—as an example, the dipole moment of water is 1.85 D in the gas phase [29] and it increases to 3 D in liquid water due to polarization. The TIP3P water model [30] used in many biophysical FFs has a dipole moment of 2.35 D, which represents a middle ground that accommodates both pure aqueous water as well as solute behaviour. To our knowledge, no measurements of water dipole moment at solid surfaces are available, but one would expect the dipole moment to decrease near proteins or surfaces with lower dielectric constants. Herein lies the difficulty of simulating interfacial systems with a fixed-charge FF. The fixed partial charges used in the current FFs are optimized for protein interactions in bulk aqueous environments and incorporate the effects of polarization through a mean-field approximation. Such an implicit treatment of polarization seems to work well in bulk, but polarization effects are dependent on the environment, and this could lead to problems in inhomogeneous systems as amply demonstrated in the narrow gramicidin channel [31]. In the light of this, there has been some debate about the ability of existing FFs to model an interfacial system [32], but there are promising signs that with rigorous parametrization it is possible, and that it can be assessed using proper validation. In this review, we will cover some of the efforts at parametrizing a fixed-charge FF for interfacial peptide–surface interactions.

## 2. Interfacial force fields and surface models

There have been several attempts at FF development for the purpose of peptide–surface interactions. We will cover three materials, gold, silica and polymers, because they are important for a range of biotechnological applications and they illustrate the primary concerns of FF development

such as parameter balance and validation, but there is ongoing work on FFs for other materials like metals [33,34], metal oxides [35], minerals [36,37] and graphene [38,39].

For gold surfaces, there is a strong polarization component to binding, which increases with peptide charge [28]. This is a consequence of mobile electrons that contribute to 'image-charge' effects [40]—an oppositely charged pseudo-molecule in the metal that is induced within the metal surface. As a consequence, many FFs include some degree of polarization in the surface. The first such FF for use with proteins, GolP [41], used a rotating dipolar rod as a method of reproducing image charge effects, with an additional potential to account for the chemisorption of sulfur. With polarization accounted for, the other terms (the LJ parameters) were tuned to reproduce both QM calculations [41] and experimental adsorption energies of model alkane molecules determined by helium atom reflectivity [42]. Validation of the FF with further alkane molecules produced reasonable agreement, and while analysis of the free energy of adsorption correlated with experiment, the resolution of the experimental measures was not fine enough to determine a quantitative comparison—more approaches to validation with adsorption free energy are discussed later. The next advance in gold–protein interactions was the GolP-CHARMM FF [43] that was integrated with the existing biophysical CHARMM FF.

The GolP-CHARMM FF reparametrized GolP from first principles and extended its use to both Au(111) and Au(100) surfaces, while retaining the capability for polarizability all within the existing CHARMM FF. While not completely consistent, requiring several new atom types to be added to the existing parameters for proteins, it successfully showed quantitative agreement with amino acid adsorption energies (but not free energies, which include entropic effects). The parametrization effort was based mostly on *ab initio* results for the (111) surface, and entirely on *ab initio* data for the extension to the (100) surface, owing to the lack of experimental data. In this case, where there are no bonded parameters (the gold atoms were fixed) and the charge parameters are handled by the rotating rod model, the main consideration is the LJ well depth, which was fitted to QM interaction energies of alkanes approaching a gold surface in the gas phase. Ordinarily, interaction energies from gas-phase QM approximations are not applicable in condensed-phase fixed-charge FFs, because the electron density varies so much from the gas phase to the condensed phase (and condensed-phase QM simulations are computationally prohibitive). The polarizable model used here accounted for this effect. One concern is that the LJ cutoff was changed during simulation, as it is agreed that parametrization of existing FFs incorporates, and relies upon, the native cutoff [27]. In this respect, it will be interesting to see whether this FF can be rigorously validated in protein adsorption free energy simulations. It is implied that the fitting of enthalpic adsorption energies with model molecules (including water) will translate to an accurate thermodynamic treatment of the interface, and this is best tested with free energy simulations. Recent results measuring the free energy of adsorption of the gold-binding peptide AuBP1 on three different gold facets compare favourably with experimental data and also help explain facet selectivity and the role of water in binding [44]. Fitting the FF to the bulk material properties of the surface, by removing the fixed atom restraint on gold atoms, will also allow for validation using interfacial energies.

Another thoroughly parametrized interfacial FF is also integrated with biomolecular FFs and is also fixed-charge. The INTERFACE FF [45] of Heinz *et al*. [46] is a collection of parameter sets for inorganic materials, having started with the CHARMM-METAL FF and subsequently being extended to several clay minerals, silica, sulfates and phosphates. Parametrization of these materials followed a nuanced approach that took into account broad sources of data that are applicable to each material to create a thermodynamically consistent FF. As an example, the CHARMM-METAL FF aimed to accurately simulate face-centred cubic metals. Prior to this FF, existing biomolecular FFs were parametrized using vaporization energies of model alkane compounds as target data. However, the temperature-dependence of the LJ approximation means it is not applicable to parametrizing metals at their boiling points an order of magnitude above the temperature of interest [46], so vaporization energy was no longer applicable. Heinz *et al.* thus used experimental material properties like density and surface energy as target data to fit the LJ atomic radii and well depth, respectively (in fixed-charge FFs for metals, the partial charge can be set to zero). This meant incorporating the attractive effect of polarization into the LJ well depth, which was significantly higher than other atom types from the biomolecular FFs.

Highly polarizable surfaces expose the main question about fixed-charge FFs in interfacial systems. The net polarization in peptide adsorption is the polarization of the surface *by* the peptide, minus the loss of its polarization by displaced water [28]. Fixed-charge FFs, with averaged parameters to implicitly incorporate polarization effects, may accurately determine thermodynamic quantities because in simulation these values are calculated using the interaction of a single molecule, whereas in experiment millions or billions of molecules are measured. Small inaccuracies resulting from the averaging of parameters can lead to large inaccuracies in the absolute value of the quantity being measured. To our knowledge, there has not been a thermodynamic investigation of peptide adsorption using the CHARMM-METAL FF, so this could be a valuable next step in validating its place within the CHARMM FF.

On insulating surfaces, there is less of an attractive force from polarization as there are no image charge effects. We will now consider one such surface, silica, and its parametrization [47]. For silica, a crystal, there is a useful set of target data from X-ray diffraction. For the INTERFACE silica parameters, electron densities from X-ray diffraction of alpha-quartz were integrated within the silicon and oxygen atomic radii to determine atomic partial charges [48]. On the surface, hydroxyl and siloxide groups were given partial charges rationalized by their similar pH to already parametrized functional groups. By then fitting the bonded and LJ parameters to X-ray and Raman spectra target data, the FF was fitted to bulk material properties (in this case, heat of immersion). Validation included computed water contact angles and adsorption isotherms. This procedure is an ideal example of balanced parametrization of a surface—there is detailed scrutiny of surface morphology at the atomic scale, as well as on bulk properties. A companion paper attempted to validate the FF using peptide adsorption [49]. This exercise demonstrated qualitative agreement of binding strength to silica surfaces with different surface groups, and allowed for an atomistic understanding of the trends controlling binding. However, the comparison of experimental adsorption data with the simulated percentage of time the

peptides spent on surfaces (which share a logarithmic, not linear, relationship) was an uncommon measure not suitable for quantitative comparisons. There are enhanced sampling methods available to extend the effective sampling beyond 5 ns, allowing for the calculation of free energies [50]. Given that a single set of partial charge parameters were used across several different FFs, each with different water molecules, there is the possibility of parameters not being suitable for the water model and associated protein parameters on the atomic scale, despite good agreement at the bulk level. The extent of agreement with measured free energies of adsorption can be used to assess accuracy at the atomic scale. This approach to validation is covered below.

The most thorough work in validating FFs for interfacial systems comes from the Latour laboratory working with peptides on self-assembled monolayers (SAMs). While metal and crystal surfaces have a well-defined structure, for most organic surfaces such as polymers there is the added complication of a high level of heterogeneity in the surface composition. Their bulk properties are best considered with coarse-grain approaches [19], inherently unsuitable for determining the effect of specific functional group interactions. An added difficulty in simulating these systems is the presence of contaminants, and also the heterogeneity introduced by surface treatments, such as by plasma [51]. Given that there is no atomic structure for these materials, researchers have turned to SAMs of alkanes that can be more easily controlled in experiment, and more explicitly defined in simulation. These can be a proxy for polymer surfaces because the surface-tethered alkanes can be functionalized with specific functional groups that are presented on the surface and mimic those groups that would be found on the surface of a polymer. This has been used by Wei and Latour as a powerful source of both experimental and simulated data to facilitate a comparison for parametrization and validation purposes [52].

Latour is the first, and to our knowledge only, investigator to consider the free energy of peptide adsorption as a property for the validation of an interfacial FF [53]. For these studies, Latour used a simulated model of SAMs using parameters for similar moieties in the CHARMM FF. First, Sun *et al.* tried calculating the adsorption free energy of peptides using implicit water, but it became clear from this work that explicit water is necessary for accurately calculating the free energy of peptide adsorption [54]. The extra computational load of explicit water was addressed using a novel enhanced sampling method [55,56]. Since that work, computer power has increased and the need for enhanced sampling has diminished—longer periods of simulation can now adequately sample slower degrees of freedom [14]. Experimentally measured adsorption data were used as a validation dataset for the parameters transferred from the CHARMM FF and showed that the existing CHARMM parameters were not completely transferable to an SAM surface [32]. Specifically, there was good agreement for the calculated free energy of adsorption for the test peptide on SAM surfaces with hydrophilic end groups, but poor agreement for the same system with hydrophobic end groups.

The failure of the existing CHARMM parameters to simulate an SAM surface led to the development of a new FF, the Dual-FF, that scaled the interactions between the solid and liquid phases such that the adsorption free energies from experiment were reproduced in the simulation [32]. In particular, the effective LJ well depth of the water and the SAM carbons was altered, as well as the effective charge on regions of the peptide. These changes had no effect on the interactions within each phase separately (and so preserved the peptide solution-phase behaviour), but scaled the potential energy calculated *between* phases to compensate for errors in peptide behaviour over hydrophobic SAMs. This method successfully reproduced the free energy of adsorption on hydrophobic SAMs without affecting the same measure over hydrophilic SAMs. The Dual-FF approach has also been applied to silica [57] and high-density polyethylene [58] to reproduce experimental adsorption data. While this solution achieves the goal of reproducing the desired experimental data, it is a practical solution, which neglects the possibility of underlying errors in the parametrization of the surface. Additionally, by using peptide adsorption free energies as the target data for parametrizing the FF, the reproduction of peptide adsorption free energies becomes a foregone conclusion and the experimental data available for validation is reduced. Another approach, detailed below, is to first parametrize the surface based on external datasets and then to test its behaviour in the target system.

The CHARMM FF has adopted the principle of transferability, whereby parameters for atoms in a class of functional groups can be transferred to similar groups without further optimization. The experimental and simulation work by Latour *et al.* shows that for parametrizing interfacial systems, the transferability of existing CHARMM parameters is compromised. In agreement with this, it has been determined by the CHARMM developers that monolayers cannot be simulated self-consistently by the latest lipid bilayer CHARMM parameter set [27]. This might not preclude simulating peptide–surface interactions consistently with the CHARMM FF, but it may require additional atom typing and parametrization. Given that the system in question is an alkane monolayer, it is instructive to look at the parametrization of a similar system, the lipid bilayer.

Following the parametrization philosophy adopted in CHARMM [59], the initial step to parametrizing a new molecule is to develop internal degrees of freedom to reproduce lowest energy molecular configurations, usually determined by QM modelling. Following this, a first guess of the partial charges is made and then LJ parameters are determined (figure 2). This is an iterative process, which can be reiterated from any stage according to how well the model reproduces the experimental target data (figure 2). Some simulation data have shown monolayer structure to be sensitive to the LJ terms [60], so this may provide an avenue to address parameter correlation and guide the parametrization process. As asserted by Zhu *et al.* [59], the task can be approached by degrees—having validated the non-bonded terms, the torsional terms can be refit to reproduce accurate dihedral behaviour and ensure the correct material dynamics are preserved.

A thorough parametrization process such as this can potentially produce a surface model that is balanced—reproducing both monolayer material properties as well as bulk thermodynamic quantities. Then, validation using the experimental peptide adsorption data can assess the rigour of the parameters.

As demonstrated on the gold and silica surfaces, there is no hard and fast rule for experimental target data, because the nuances of each material will be different. In the following section, we will describe some of the experimental tools that can be used to characterize surfaces for simulation,

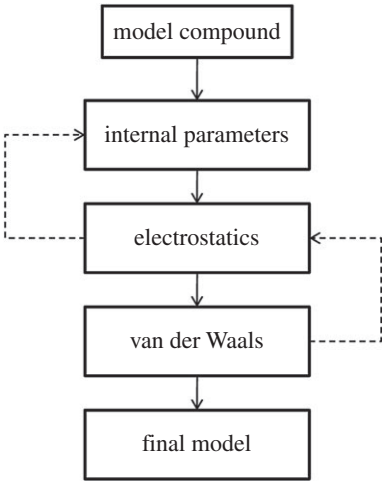



**Figure 2.** The parametrization process used in the development of the CHARMM FFs. The process is reiterated until satisfactory reproduction of target data is achieved. Adapted with permission from [59], copyright © 2011 John Wiley and Sons.

particularly those data that can be reproduced by simulation so that the validity of the model can be tested.

## 3. Experimental target data and validation data

All surfaces will require some measure of experimental data to be used in parametrization. For some materials, explicit topological data can be determined. These surfaces have either a crystal or lattice structure that can be probed by X-ray diffraction. Others, such as monolayers or lipid bilayers, do not have a single atomic-level description and must instead be understood by dynamic means, based on a statistical distribution function. Even still, contamination of the experimental surface can frustrate parametrization and validation, so a careful understanding of the experimental conditions is required. This section will demonstrate how these considerations have been addressed by the choice of experimental target data, as well as covering some simulation techniques for validation.

The LJ terms and electrostatic terms, thought of as non-bonded terms, are not typically expected to influence the structure of bonded molecules. However, in most FFs, non-bonded interactions are ignored only between atoms that are bonded by up to three bonds. This means the non-bonded parameters can still affect the torsional behaviour of a dynamic molecule, and thereby the structure of mobile phases such as mono- or bilayers. For the CHARMM lipid FF, this meant that parameter correlation in the non-bonded terms was an issue that affected lipid structure. Ordinarily, in biomolecular FFs, after partial charges were determined, the LJ parameters were developed by fitting to experimental heats of vaporization and relative free energies of solvation [61,62]. However, the solutions to these fits were non-trivial, with scope for parameter correlation, such that errors in one parameter could be compensated by another. The solution was to use QM modelling of model molecules interacting with helium and neon, which carry no partial charge to interfere with the calculations, as a means to determine the *relative* van der Waals minimum interaction distances. This reduced the number of possible combinations of parameters, informing the choice of the correct set.

Dynamic materials like mono- and bilayers are usually described using the surface area per lipid, thickness of the bilayer or acyl chain order parameters. Despite the differing timescales of simulation and experiment (nanoseconds for simulation compared to microseconds for $^{1}$H NMR [63]), these measures have been used successfully for the parametrization of lipid bilayers [29]. A specific consideration that has emerged from this process is the importance of using the particle mesh Ewald, a method of calculating long-range electrostatics, rather than a cutoff, because the surface area of lipid is sensitive to the balance of forces acting on the headgroup and alkane tails [64]. Whether or not this is true for alkanethiol SAMs such as those commonly used in peptide binding experiments remains to be determined.

The type and quantity of the functional groups appearing on the surface of a material will significantly affect the ability of peptides to bind. For example, by varying the chemistry of silica surfaces with pH changes, adsorbant peptides have less than 5% sequence similarity [47]. As there are 20 natural amino acids, this equates to random sequence homology, and the reasons for this binding pattern are unexplained at the atomic level. There are a wide range of methods to determine the identity and the area density of these surface chemical groups, and they can be used to inform the development of the surface model. To some extent the atomic composition of the surface will be known *a priori*, but X-ray photoelectron spectroscopy (XPS) can be used in this context to ascertain the presence of contaminants that might affect peptide interactions at the surface [47,65,66]. On silica, specific surface area in combination with thermogravimetric analysis, in which the loss of mass during heating is measured, were used in tandem to determine the area density of surface siloxide groups [47]. Particularly where pH effects are concerned, potentiometric titration and zeta potential are useful for measuring the degree of ionization on the surface.

Attenuated total reflection infrared spectroscopy and XPS typically penetrate further into the substrate than is desired for surface characterization, where only the first few atomic layers are of interest [67]. These techniques thus combine information about the chemistry of the bulk material with that of the surface, but peptide interactions are dominated by the chemical groups immediately on the surface. For this reason, in cases where there are charged groups on the surface, atomic force microscopy has been used to determine the surface potential and through this, the area density of surface charges [68]. Similarly, for SAMs where the composition is known *a priori*, but can also be confirmed with XPS, surface plasmon resonance (SPR) can measure the pKa of the surface functional groups and help in modelling the correct area density of charges under different pH conditions [69].

The presence of specific functional groups on a surface is necessary for any all-atom model of an interface with peptides in solution to function, but for a balanced representation, the surface must also reproduce bulk material properties. For solid surfaces such as metals or polymers, where there is very little dynamic behaviour, it is likely that the main determinants of peptide binding will be the structure of water on the surface. This is because in bulk aqueous solution, water will form a dynamic network of freely moving water molecules, with a large number of stabilizing hydrogen bonds [70]. This is an enthalpically and entropically favourable situation. However on hydrophobic surfaces, this network is disturbed and water molecules become oriented, losing orientational freedom

and hydrogen bond contacts. This is associated with a reduced water density on the surface, called de-wetting. In this context, the adsorption of a peptide can be regarded as minimizing the surface contact with water and so returning its favourable entropy and reducing interfacial tension [70,71]. This can be a significant contributor to binding [72], and so must be addressed by any surface model used for peptide binding.

Current methods of experimentally determining the bulk properties at the solid–water interface use some combination of surface tension, interfacial tension, water contact angle, zeta potential or heat of immersion. The water contact angle, an implicit measure of the polarity of the surface, is a technique readily available in most laboratories and has a large number of published data for most surfaces of interest. In simulation, the water contact angle can also be determined at the nanoscale to compare with experimental measurements, with some added considerations. While experimental contact angles represent the average of trillions of molecular interactions, nanoscale droplets can be somewhat influenced by the specific composition of a handful of polar and non-polar groups on the surface; however this effect can be quantified and removed by consideration of the solvent-accessible surface area of those groups [73]. The process differs from macroscale measurements—while on some simulated surfaces it is enough to simply fit a circle to the two-dimensional density profile of a nanodroplet of water [47,74] and measure the angle at the interface, for other surfaces the line tension can affect the size of the droplet [75]. This relationship between the energy at the perimeter of the droplet and the size of the circumference of the droplet at the surface can lead to size dependence of the calculated water contact angles, and so the level of dependence on this value should be determined [76] before calculation, as in [77].

Macroscopic properties like heat of vaporization or the free energy of hydration are also common measures. Clearly, heat of vaporization is not appropriate for a solid-phase surface. With lipids, the free energy of solvation has been determined in simulation using a free energy method that involves transforming the target molecule in and out of a water box [78], but it is unclear how this approach would be implemented for surfaces, and there are no published examples using the biomolecular FFs mentioned here. The enthalpy of aqueous immersion, however, is an experimental quantity that can also be calculated in simulations. Determined using calorimetry, this quantity can be calculated using three separate simulations of the surface–water interface, the surface alone, and the water alone, as demonstrated for metal [46], silica [47] and most recently hydroxyapatite [79].

Using a thorough surface characterization such as through the methods outlined here, highly heterogeneous or amorphous surfaces can be modelled without exact atomic coordinates of the surface. If the functional groups on the surface are present in the correct area density, and the surface is parametrized so as to reproduce bulk thermodynamic quantities, then it is likely that it can also quantitatively measure the adsorption free energy of peptides and thus be used as a predictive tool for understanding peptide–surface interactions. To our knowledge, there are no published examples of this process being followed rigorously for heterogeneous surfaces—in which the exact atomic coordinates of the surface are used for free energy validation and prediction. However, a number of studies have used organic models for surfaces and explored the trends resulting from altering the atomic composition of the surface. Dai *et al.* [74] developed a model for polypropylene with parameters borrowed from the CHARMM FF, and compared the effect of common hydrophilization procedures on water contact angle in experiment. The model reflected the trends from experiment, but the parameters were not optimized for polypropylene so absolute values could not be compared. Most recently, the dual-FF philosophy for CHARMM originally developed for SAMs was applied to high-density polyethylene [58]. Surface characterization included atomic composition and water contact angle, but these were not used for optimization of the surface parameters. These models represent promising approaches, yet the lack of parametrization in each case precludes any assessment of their potential for predictive simulations.

Surface models that are justified in their atomic composition and bulk thermodynamic properties must then be validated in order to be quantitative for peptide interactions. The quantity of most interest is the free energy of adsorption, because it accounts for both the enthalpic interaction with the surface as well as the entropic change in the system. Experimentally, the free energy of the interaction is more difficult to determine for peptides on surfaces than it is for interactions between functional proteins. In experimentally measuring the binding affinity of small molecules to proteins, there is often some functional effect (such as current–voltage curves for ion channels) that will be altered by some critical concentration of the small molecule, and this concentration can be used to determine the binding free energy. On surfaces, however, the binding free energy of a peptide is more difficult to determine. On electrically conductive surfaces, SPR is a common technique to observe and measure peptide adsorption [65,80,81]. For surfaces with more than a nanometre-thick non-conductive layer such as organic polymers, Thyparambil *et al.* have measured peptide binding by determining a correlation between desorption via atomic force microscopy and the binding free energy from SPR experiments. Atomic force microscopy is a powerful method of measuring adsorption free energy. Because a pulling force is applied, it does not rely on adsorption and subsequent desorption of the peptide, which can limit SPR measurements to only those peptides with reversible binding.

Importantly, there are several simulation analogues that can be used to compare with experiment for validating model surface parameters. Umbrella sampling is one method, in which peptides are restricted at increasing distances from the material surface. At each distance, the force of attraction between the peptide and the surface can be calculated from the movement away from the restrictive potential towards the surface. These positional data can then be deconvoluted into a continuous free energy profile, and this is used to determine an adsorption free energy, as in [82], to compare with experimental adsorption free energies. There is another popular approach that may have advantages over umbrella sampling in terms of simplicity and parallelization. In this approach, steered molecular dynamics is used to pull the bound peptide away from the surface while measuring resistance. This is a non-equilibrium process, which used to be considered inappropriate for determining equilibrium quantities like free energy. However, using a statistical mechanical relation discovered relatively recently [83], the free energy of adsorption can be calculated. This has already been used for surface adsorption free energies [84]. While this technique is suitable for determining relative binding free energies, and is attractive for its ease of use, it offers no sampling benefits

over umbrella sampling when calculating absolute values for adsorption free energy [85].

Another enhanced simulation technique that has recently been applied to interfacial peptide surface adsorption is metadynamics. Here, one or two degrees of freedom are biased via the potential energy function, forcing the peptide to escape from low free energy wells while also recreating the free energy landscape as it is explored. As the computational effort grows dramatically with each degree of freedom that is biased, generally only one or two are chosen—such as the peptide distance or angle from a surface. To ensure full sampling while only biasing a couple of degrees of freedom, metadynamics has been combined with replica exchange [86] to measure the free energy of adsorption for peptides at interfaces. This approach was demonstrated on the alanine dipeptide on gold [87] and has been used to calculate the free energy of adsorption for oligopeptides on SAM surfaces [88], silica [89] and on gold [44]. These studies also demonstrate how a re-weighting scheme can be applied to undo the metadynamics bias and analyse additional degrees of freedom in an equilibrium Boltzmann distribution to assess their free energy landscape.

## 4. Polarizable force fields for interfacial systems

While this review focuses on additive FFs, it is important to consider it in the context of the ongoing development of polarizable FFs. Ultimately, if a rigorously parametrized surface fails to reproduce peptide adsorption free energies using an additive FF, this is most likely due to the lack of polarizability, assuming proper sampling of the peptide conformational entropy at the surface. On this basis, a polarizable FF will be needed. The GolP-CHARMM FF [43] mentioned previously incorporates polarization in a gold surface using a rotating rod model, retaining the fixed-charge description in the aqueous phase. Walsh and co-workers have also recently applied this method to silver [34] and graphite [39] surfaces, where they take advantage of the computational efficiency afforded by the fixed-charge treatment of the aqueous phase to calculate amino acid binding free energies.

There are also two fully polarizable FFs of interest for simulating biological interfaces. The AMOEBA FF introduces polarization with multipole–multipole interactions and has recently been extended for use with proteins [90]. While the most recent update to the FF has not been applied to interfaces, previously an FF based on AMOEBA was extended for use with peptides on a graphene surface [91]. The CHARMM Drude-2013 FF is similar to the CHARMM additive FF, with the addition of a charged particle within non-hydrogen atoms that moves in response to the electric field, also known as the charge-on-spring [26]. The use of these FFs in protein simulations has been reviewed recently [92], but current parametrization efforts focus mostly on biomolecules in aqueous environments and there is significant work yet to be done towards simulating an interfacial system.

## 5. Conclusion

Understanding peptide–surface interactions has enormous potential in biotechnological applications, but their characterization at the atomic level is beyond the resolution of current experimental techniques. For this reason, molecular mechanics is used to model this interaction in atomic detail. However, most FFs currently in use are not tailored for the interaction of proteins with surfaces and do not account for polarization effects. New FFs for the simulation of peptide–surface interactions developed with a rigorous parametrization procedure taking into account these effects are required. In order for the model to accurately reproduce binding behaviour the parametrization must be balanced, reproducing both atomic-scale interactions and bulk thermodynamic properties. This will allow it to account for both the enthalpic and entropic contributions to peptide adsorption. Additionally, it must be followed by validation with experimental quantities that were not used during parametrization. Significant ground has already been covered in understanding the process of peptide adsorption, but strict parametrization and validation promise to advance this field even further in future.

Authors' contributions. L.M. performed research and drafted the article; S.K. provided technical oversight and editing of the text; M.B. and A.W. provided intellectual input and edited the text for meaning and readability.

Competing interests. We declare we have no competing interests.

Funding. We acknowledge funding from the Australian Research Council and National Health & Medical Research Council. A.S.W. received additional funding from the National Institutes of Health (EB014283) and the Wellcome Trust (103328).

## References

1. Muller W, Wang X, Proksch P, Perry C, Osinga R, Garderes J, Schroder H. 2012 Principles of biofouling protection in marine sponges: a model for the design of novel biomimetic and bio-inspired coatings in the marine environment? *Mar. Biotechnol.* **15**, 375–398. (doi:10.1007/s10126-013-9497-0)
2. Hirano Y, Mooney D. 2004 Peptide and protein presenting materials for tissue engineering. *Adv. Mater.* **16**, 17–25. (doi:10.1002/adma.200300383)
3. Alkilany A, Lohse S, Murphy C. 2013 The gold standard: gold nanoparticle libraries to understand the nano-bio interface. *Acc. Chem. Res.* **46**, 650–661. (doi:10.1021/ar300015b)
4. Bilek M. 2014 Biofunctionalization of surfaces by energetic ion implantation: review of progress on applications in implantable biomedical devices and antibody microarrays. *Appl. Surf. Sci.* **310**, 3–10. (doi:10.1016/j.apsusc.2014.03.097)
5. Schilp S, Kueller A, Rosenhahn A, Grunze M, Pettitt M, Callow M, Callow J. 2007 Settlement and adhesion of algal cells to hexa(ethylene-glycol)-containing self-assembled monolayers with systematically changed wetting properties. *Biointerphases* **2**, 143–150. (doi:10.1116/1.2806729)
6. Lord M, Foss M, Besenbacher F. 2010 Influence of nanoscale surface topography on protein adsorption and cellular response. *Nano Today* **5**, 66–78. (doi:10.1016/j.colsurfb.2014.09.040)
7. Roach P, Eglin D, Rohde K. 2007 Modern biomaterials: a review—bulk properties and implications of surface modification. *J. Mater. Sci. Mater. Med.* **18**, 1263–1277. (doi:10.1007/s10856-006-0064-3)
8. Grinnell F, Feld M. 1981 Fibronectin adsorption on hydrophilic and hydrophobic surfaces detected by antibody binding and analyzed during cell adhesion in serum-containing medium. *J. Biol. Chem.* **257**, 4888–4893. (doi:10.1073/pnas.0601617103)
9. Anderson J, Rodriguez A, Chang D. 2008 Foreign body reaction to biomaterials. *Semin. Immunol.* **20**, 83–85. (doi:10.1016/j.smim.2007.12.008)

10. Shin H, Jo S, Mikos A. 2003 Biomimetic materials for tissue engineering. *Biomaterials* **24**, 4353–4364. (doi:10.1002/adfm.201103083)
11. Bax D, McKenzie D, Weiss A, Bilek M. 2010 The linker-free covalent attachment of collagen to plasma immersion ion implantation treated polytetrafluoroethylene and subsequent cell-binding activity. *Biomaterials* **31**, 2526–2534. (doi:10.1016/j.biomaterials.2009.12.009)
12. Kokkoli E, Mardilovich A, Wedekind A, Rexeisen E, Garg A, Craig J. 2006 Self-assembly and applications of biomimetic and bioactive peptide-amphiphiles. *Soft Matter* **2**, 1015–1024. (doi:10.1039/B608929A)
13. Patwardhan S. 2011 Biomimetic and bioinspired silica: recent developments and applications. *Chem. Commun.* **47**, 7567–7582. (doi:10.1039/c0cc05648k)
14. Latour R. 2014 Perspectives on the simulation of protein–surface interactions using empirical force field methods. *Colloids Surf. B Interfaces* **124**, 25–37. (doi:10.1016/j.colsurfb.2014.06.050)
15. Mirau P, Naik R, Gehring P. 2011 Structure of peptides on metal oxide surfaces probed by NMR. *J. Am. Chem. Soc.* **133**, 18 243–18 248. (doi:10.1021/ja205454t)
16. Fears K, Petrovykh D, Photiadis S, Clarke T. 2013 Circular dichroism analysis of cyclic $\beta$-helical peptides adsorbed on planar fused quartz. *Langmuir* **29**, 10 095–10 101. (doi:10.1021/la401544c)
17. Hofer T. 2013 From macromolecules to electrons—grand challenges in theoretical and computational chemistry. *Front. Chem.* **1**, 6. (doi:10.3389/fchem.2013.00006)
18. Van Gunsteren W, Berendsen H. 1990 Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.* **29**, 992–1023. (doi:10.1002/anie.199009921)
19. Marrink S, Tieleman D. 2013 Perspective on the Martini model. *Chem. Soc. Rev.* **42**, 6801–6822. (doi:10.1039/C3CS60093A)
20. Dzubiella J, Swanson J, McCammon J. 2006 Coupling hydrophobicity, dispersion, and electrostatics in continuum solvent models. *Phys. Rev. Lett.* **96**, 087802. (doi:10.1103/PhysRevLett.96.087802)
21. MacKerell Jr AD *et al.* 1998 All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616. (doi:10.1021/jp973084f)
22. Cornell W *et al.* 1995 A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197. (doi:10.1021/ja00124a002)
23. Oostenbrink C, Villa A, Mark A, van Gunsteren W. 1998 A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656–1676. (doi:10.1002/jcc.20090)
24. Jorgensen W, Tirado-Rives J. 1988 The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666. (doi:10.1021/ja00214a001)
25. Jones E. 1924 On the determination of molecular fields. II. From the equation of state of a gas. *Proc. R. Soc. Lond. A* **106**, 463–477. (doi:10.1098/rspa.1924.0082)
26. Huang J, Lopes P, Roux B, MacKerell Jr AD. 2014 Recent advances in polarizable force fields for macromolecules: microsecond simulations of proteins using the classical Drude oscillator model. *J. Phys. Chem. Lett.* **5**, 3144–3150. (doi:10.1021/jz501315h)
27. Pastor R, MacKerell Jr AD. 2011 Development of the CHARMM force field for lipids. *J. Phys. Chem. Lett.* **2**, 1526–1532. (doi:10.1021/jz200167q)
28. Heinz H, Jha KC, Luettmer-Strathmann J, Farmer L, Naik RR. 2011 Polarization at metal–biomolecular interfaces in solution. *J. R. Soc. Interface* **8**, 220–232. (doi:10.1098/rsif.2010.0318)
29. Gubskaya A, Kusalik P. 2002 The total molecular dipole moment for liquid water. *J. Chem. Phys.* **117**, 5290–5302. (doi:10.1063/1.1501122)
30. Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M. 1983 Comparison of simple potential functions for simulating water. *J. Chem. Phys.* **79**, 926–935. (doi:10.1063/1.445869)
31. Baştuğ T, Kuyucak S. 2006 Energetics of ion permeation, rejection, binding, and block in gramicidin A from free energy simulations. *Biophys. J.* **90**, 3941–3950. (doi:10.1529/biophysj.105.074633)
32. Biswas P, Vellore N, Yancey J, Koccukkal T, Collier G, Brooks B, Stuart S, Latour R. 2012 Simulation of multiphase systems utilizing independent force fields to control intraphase and interphase behavior. *J. Comput. Chem.* **33**, 1458–1466. (doi:10.1002/jcc.22979)
33. Heinz H, Farmer B, Pandey R, Slocik J, Patnaik S, Pachter R, Naik R. 2009 Nature of molecular interactions of peptides with gold, palladium, and Pd-Au bimetal surfaces in aqueous solution. *J. Am. Chem. Soc.* **131**, 9704–9714. (doi:10.1021/ja900531f)
34. Hughes Z, Wright L, Walsh T. 2013 Biomolecular adsorption at aqueous silver interfaces: first-principles calculations, polarizable force-field simulations, and comparisons with gold. *Langmuir* **29**, 13 217–13 229. (doi:10.1021/la402839q)
35. Vallee A, Humblot V, Pradier C. 2010 Peptide interactions with metal and oxide surfaces. *Acc. Chem. Res.* **43**, 1297–1306. (doi:10.1021/ar100017n)
36. Heinz H, Koerner H, Anderson K, Vaia R, Farmer B. 2005 Force field for mica-type silicates and dynamics of octadecylammonium chains grated to montmorillonite. *Chem. Mater.* **17**, 5658–5669. (doi:10.1021/cm0509328)
37. Heinz H, Vaia R, Farmer B. 2006 Interaction energy and surface reconstruction between sheets of layered silicates. *J. Chem. Phys.* **124**, 224713. (doi:10.1063/1.2202330)
38. Walsh T. 2008 Modelling the nanoscale patterning of nucleic acid base pairs deposited on graphite. *Mol. Phys.* **106**, 1613–1619. (doi:10.1080/00268970802225533)
39. Hughes Z, Tomásio S, Walsh T. 2014 Efficient simulations of the aqueous bio-interface of graphitic nanostructures with a polarisable model. *Nanoscale* **6**, 5438–5448. (doi:10.1039/c4nr00468j)
40. Lang N, Kohn W. 1973 Theory of metal surfaces: induced surface charge and image potential. *Phys. Rev. B* **7**, 3541–3550. (doi:10.1103/PhysRevB.7.3541)
41. Iori F, di Felice R, Molinari E, Corni S. 2009 GolP: an atomistic force-field to describe the interaction of proteins with Au(111) surfaces in water. *J. Comput. Chem.* **30**, 1465–1476. (doi:10.1002/jcc.21165)
42. Wetterer S, Lavrich D, Cummings T, Bernasek S, Scoles G. 1998 Energetics and kinetics of the physisorption of hydrocarbons on Au(111). *J. Phys. Chem. B* **102**, 9266–9275. (doi:10.1021/jp982338+)
43. Wright L, Rodger P, Corni S, Walsh T. 2013 GolP-CHARMM: first-principles based force fields for the interaction of proteins with Au(111) and Au(100). *J. Chem. Theory Comput.* **9**, 1616–1630. (doi:10.1021/ct301018m)
44. Wright L, Palafox-Hernandez J, Rodger P, Corni S, Walsh T. 2015 Facet selectivity in gold binding peptides: exploiting interfacial water structure. *Chem. Sci.* **6**, 5204–5214. (doi:10.1039/C5SC00399G)
45. Heinz H, Lin T, Misha R, Emami F. 2013 Thermodynamically consistent force fields for the assembly of inorganic, organic, and biological nanostructures: the INTERFACE force field. *Langmuir* **29**, 1754–1765. (doi:10.1021/la3038846)
46. Heinz H, Vaia R, Farmer B, Naik R. 2008 Accurate simulation of surfaces and interfaces of face-centered cubic metals using 12–6 and 9–6 Lennard-Jones potentials. *J. Phys. Chem. C* **112**, 17 281–17 290. (doi:10.1021/jp801931d)
47. Emami F, Puddu V, Berry R, Varshney V, Patwardhan S, Perry C, Heinz H. 2014 Force field and a surface model database for silica to simulate interfacial properties in atomic resolution. *Chem. Mater.* **26**, 2647–2658. (doi:10.1021/cm500365c)
48. Heinz H, Suter U. 2004 Atomic charges for classical simulations of polar systems. *J. Phys. Chem. B* **108**, 18 341–18 352. (doi:10.1021/jp048142t)
49. Emami F, Puddu V, Berry R, Varshney V, Patwardhan S, Perry C, Heinz H. 2014 Prediction of specific biomolecule adsorption on silica surfaces as a function of pH and particle size. *Chem. Mater.* **26**, 5725–5734. (doi:10.1021/cm5026987)
50. Chipot C, Pohorille A. 2007 *Free energy calculations: theory and applications in chemistry and biology*. Berlin, Germany: Springer.
51. Wakelin E, Kondyurin A, Wise S, McKenzie D, Davies M, Bilek M. 2015 Bio-activation of polyether ketone using plasma immersion ion implantation: a kinetic model. *Plasma Process. Polym.* **12**, 180–193. (doi:10.1002/ppap.201400149)

52. Collier G, Vellore N, Yancey J, Stuart S, Latour R. 2012 Comparison between empirical protein force fields for the simulation of the adsorption behavior of structured LK peptides on functionalized surfaces. *Biointerphases* **7**, 24. (doi:10.1007/s13758-012-0024-z)
53. Raut V, Agashe M, Stuart S, Latour R. 2005 Molecular dynamics simulations of peptide-surface interactions. *Langmuir* **21**, 1629–1639. (doi:10.1021/la047807f)
54. Sun Y, Latour R. 2006 Comparison of implicit solvation models for the simulation of protein-surface interactions. *J. Comput. Chem.* **27**, 1908–1922. (doi:10.1039/b714141f)
55. Li X, Latour R, Stuart S. 2009 TIGER2: an improved algorithm for temperature intervals with global exchange of replicas. *J. Chem. Phys.* **130**, 174106. (doi:10.1063/1.3129342)
56. Li X, Latour R. 2011 The TIGER2 empirical accelerated sampling method: parameter sensitivity and extension to a complex molecular system. *J. Comput. Chem.* **32**, 1091–1100. (doi:10.1002/jcc.21689)
57. Snyder J, Abramyan T, Yancey J, Thyparambil A, Wei Y, Stuart S, Latour R. 2012 Development of a tuned interfacial force field parameter set for the simulation of protein adsorption to silica glass. *Biointerphases* **7**, 56. (doi:10.1007/s13758-012-0056-4)
58. Abramyan T, Snyder J, Yancey J, Thyparambil A, Wei Y, Stuart S, Latour R. 2015 Parameterization of an interfacial force field for accurate representation of peptide adsorption free energy on high-density polyethylene. *Biointerphases* **10**, 021002. (doi:10.1116/1.4916361)
59. Zhu X, Lopes P, MacKerell Jr AD. 2012 Recent developments and applications of the CHARMM force fields. *WIREs Comput. Mol. Sci.* **2**, 167–185. (doi:10.1002/wcms.74)
60. Venable R, Chen L, Pastor R. 2009 Comparison of the extended isotropic periodic sum and particle mesh Ewald methods for simulations of lipid bilayers and monolayers. *J. Phys. Chem. B* **113**, 5855–5862. (doi:10.1021/jp900843x)
61. Yin D, MacKerell Jr AD. 1998 Combined ab initio/empirical approach for optimization of Lennard–Jones parameters. *J. Comput. Chem.* **19**, 334–348. (doi:10.1002/(SICI)1096-987X(199802)19:3<334::AID-JCC7>3.0.CO;2-U)
62. Chen I, Yin D, MacKerell Jr AD. 2002 Combined ab initio/empirical approach for optimization of Lennard–Jones parameters for polar-neutral compounds. *J. Comput. Chem.* **23**, 199–213. (doi:10.1002/jcc.1166)
63. Vermeer L, de Groot B, Reat V, Milon A, Czaplicki J. 2007 Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular dynamics simulations and comparison with $^2$H NMR experiments. *Eur. Biophys. J.* **36**, 919–931. (doi:10.1007/s00249-007-0192-9)
64. Anézo C, de Vries A, Höltje H, Tieleman D, Marrink S. 2003 Methodological issues in lipid bilayer simulations. *J. Phys. Chem. B* **107**, 9424–9433. (doi:10.1021/jp0348981)
65. Wei Y, Latour R. 2009 Benchmark experimental data set and assessment of adsorption free energy for peptide–surface interactions. *Langmuir* **25**, 5637–5646. (doi:10.1021/la8042186)
66. Bax D, Kondyurin A, Waterhouse A, McKenzie D, Weiss A, Bilek M. 2014 Surface plasma modification and tropoelastin coating of a polyurethane co-polymer for enhanced cell attachment and reduced thrombogenicity. *Biomaterials* **35**, 6797–6809. (doi:10.1016/j.biomaterials.2014.04.082)
67. Briggs D. 1998 *Surface analysis of polymers by XPS and static SIMS*. Cambridge, UK: Cambridge University Press.
68. Tran C, Kondyurin A, Chrzanowski W, Bilek M, McKenzie D. 2013 Influence of pH on yeast immobilization on polystyrene surfaces modified by energetic ion bombardment. *Colloids Surf. B Biointerfaces* **104**, 145–152. (doi:10.1016/j.colsurfb.2012.11.029)
69. Fears K, Creager S, Latour R. 2008 Determination of the surface pK of carboxylic- and amine-terminated alkanethiols using surface plasmon resonance spectroscopy. *Langmuir* **24**, 837–843. (doi:10.1021/la701760s)
70. Berne B, Weeks J, Zhou R. 2009 Dewetting and hydrophobic interaction in physical and biological systems. *Annu. Rev. Phys. Chem.* **60**, 85–103. (doi:10.1146/annurev.physchem.58.032806.104607)
71. Young T, Hua L, Huang X, Abel R, Friesner R, Berne B. 2010 Dewetting transitions in protein cavities. *Proteins* **78**, 1856–1869. (doi:10.1002/prot.22699)
72. Setny P, Baron R, McCammon J. 2010 How can hydrophobic association be enthalpy driven? *J. Chem. Theory Comput.* **6**, 2866–2871. (doi:10.1021/ct1003077)
73. Wang J, Bratko D, Luzar A. 2010 Probing surface tension additivity on chemically heterogeneous surfaces by a molecular approach. *Proc. Natl Acad. Sci. USA* **108**, 6374–6379. (doi:10.1073/pnas.1014970108)
74. Dai J, Ling J, Huang X, Wan L, Xu Z. 2001 Molecular simulation on the interactions of water with polypropylene surfaces. *J. Phys. Chem. C* **115**, 10 702–10 708. (doi:10.1021/jp201040g)
75. Santiso E, Herdes C, Muller E. 2013 On the calculation of solid–fluid contact angles from molecular dynamics. *Entropy* **15**, 3734–3745. (doi:10.3390/e15093734)
76. Werder T, Walther J, Jaffe R, Halicioglu T, Koumoutsakos P. 2003 On the water–carbon interaction for use in molecular dynamics simulations of graphite and carbon nanotubes. *J. Phys. Chem. B* **107**, 1345–1352. (doi:10.1021/jp0268112)
77. Cruz-Chu E, Aksimentiev A, Schulten K. 2006 Water-silica force field for simulating nanodevices. *J. Phys. Chem. B* **110**, 21 497–21 508. (doi:10.1021/jp063896o)
78. Klauda J, Venable R, Freites J, O'Connor J, Tobias D, Mondragon-Ramirez C, Vorobyov I, MacKerell Jr AD, Pastor R. 2010 Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843. (doi:10.1021/jp101759q)
79. Wang K, Leng Y, Lu X, Ren F. 2014 Molecular dynamics simulation of protein effects on interfacial energy between HA surfaces and solutions. *Mater. Lett.* **123**, 191–194. (doi:10.1016/j.matlet.2014.02.089)
80. Tamerler C, Duman M, Oren E, Gungormus M, Xiong X, Kacar T, Parviz B, Sarikaya M. 2006 Materials specificity and directed assembly of a gold-binding peptide. *Small* **2**, 1372–1378. (doi:10.1021/bm8009895)
81. Singh N, Husson S. 2006 Adsorption thermodynamics of short-chain peptides on charged and uncharged nanothin polymer films. *Langmuir* **22**, 8443–8451. (doi:10.1021/la0533765)
82. Vellore N, Yancey J, Collier G, Latour R. 2010 Assessment of the transferability of a protein force field for the simulation of peptide-surface interactions. *Langmuir* **26**, 7396–7404. (doi:10.1021/la904415d)
83. Jarzynski C. 1997 Nonequilbrium equality for free energy differences. *Phys. Rev. Lett.* **78**, 2690–2693. (doi:10.1103/PhysRevLett.78.2690)
84. Mijajlovic M, Penna M, Biggs M. 2013 Free energy of adsorption for a peptide at a liquid interface via nonequilibrium dynamics. *Langmuir* **29**, 2919–2926. (doi:10.1021/la3047966)
85. Baştuğ T, Chen P, Patra S, Kuyucak S. 2008 Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. *J. Chem. Phys.* **128**, 155104. (doi:10.1063/1.2904461)
86. Sugita Y, Okamoto Y. 1999 Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151. (doi:10.1016/S0009-2614(99)01123-9)
87. Bellucci L, Corni S. 2014 Interaction with a gold surface reshapes the free energy landscape of alanine dipeptide. *J. Phys. Chem. C* **118**, 11 357–11 364. (doi:10.1021/jp502494k)
88. Deighan M, Pfaendtner J. 2013 Exhaustively sampling peptide adsorption with metadynamics. *Langmuir* **29**, 7999–8009. (doi:10.1021/la4010664)
89. Meißner R, Schneider J, Schiffels P, Ciacchi L. 2014 Computational prediction of circular dichroism spectra and quantification of helicity loss upon peptide adsorption on silica. *Langmuir* **30**, 3487–3494. (doi:10.1021/la500285m)
90. Shi Y, Xia Z, Zhang J, Best R, Wu C, Ponder J, Ren P. 2013 Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory Comput.* **9**, 4046–4063. (doi:10.1021/ct4003702)
91. Akdim B, Pachter R, Kim S, Naik R, Walsh T, Trohalaki S, Hong G, Kuang Z, Farmer B. 2013 Electronic properties of a graphene device with peptide adsorption: insight from simulation. *ACS Appl. Mater. Interfaces* **5**, 7470–7477. (doi:10.1021/am401731c)
92. Lopes P, Guvench O, MacKerell Jr AD. 2015 Current status of protein force fields for molecular dynamics simulations. In *Molecular modelling of proteins* (ed. A Kukol). Methods in Molecular Biology, vol. 1215, pp. 47–71. New York, NY: Springer.

# Chapter 4. Surface simulations of bioactive peptides

Molecular dynamics (MD) simulations of a biomaterial interface require special care when the fixed charge approximation is used. Nevertheless, they might still be used to develop a nanoscale understanding of the interface. In this chapter, MD simulations of hydrophobic surfaces are justified and a model for polytetrafluoroethylene (PTFE) is parameterized using water contact angle as the training data. This model surface is then used to investigate the behaviour of four peptide variants at the interface. These peptides have a bioactive epitope that, in some cases, is able to bind cell surface receptor proteins. Several types of MD simulations are performed to first explore the behaviour of the peptides on the surface and then to develop likely molecular explanations for why some peptide variants are cell active when on the PTFE surface while others are not. While the PTFE model is not rigorously validated for its physical properties here, it is intended only to represent the hydrophobic aspect of a surface. The good agreement of peptide surface adhesion strength with experimental cell activities justifies this approach.

## 4.1 Introduction

Biomaterials have become a promising avenue of research for a wide range of applications. They have come to be defined as synthetic or natural substances used in technology that interact with living matter[1]. This broad definition now encompasses a wide range of uses such as implants and prostheses, diagnostics, genetic and tissue engineering, drug delivery, and imaging[2]. In an industry worth billions of dollars, the most mature category of biomaterials is used in constructing implants and prostheses[3]. But these devices are imperfect and remain under active development as when implants contact living tissue, they influence the physiology of the surrounding area, sometimes causing an adverse response[4]. To remedy this, one of the main areas of focus in the development of more advanced implants is refining the physico-chemical makeup of the material surface to facilitate integration and reduce possible adverse effects[5].

At the microscale, the interaction between a biomaterial and the surrounding microenvironment can affect the immune response. In healthy human tissue, the interstices between cells are filled by soluble proteins and the extracellular matrix (ECM), a mixture of water and fibrous protein chains. The presence of a foreign object in this space disturbs the tissue and causes an immune response. This begins when proteins from solution adsorb on foreign objects, misfolding in the process, aggregating and exchanging with previously adsorbed proteins and hence presenting normally-hidden interior structures into the ECM. This dynamic process is known as the Vroman effect[6,7]. These denatured protein structures are recognised by immune cells, leading to encapsulation of the implant by fibrotic tissue (i.e. scarring)[8]. Misfolding of adsorbed proteins occurs because of the mismatch in the surface energies of the implant and the protein in its native conformation. Hydrophobic surfaces in particular cause structural reorganization of proteins in order to reduce the Gibbs free energy of the entire system by increasing contact between the surface and the hydrophobic interior of the protein.

The Vroman effect and the misfolding process can be mitigated by chemical alteration of the surface. The most rudimentary kind of treatment involves increasing the hydrophilicity of the surface by acid[9] or irradiation[10]. This improves cell adhesion but, for very hydrophilic surfaces, may also repel biomolecules by strongly adhering a layer of water to the surface[11]. More advanced treatments use covalent attachment to tether drugs or biomolecules to the surface[12]. These interact with cells to either kill pathogens[13] or bind directly to extracellular matrix elements or cell membrane proteins, which in turn integrate the biomaterial into the surrounding tissue to bypass the fibrotic response[14]. As well as being more robust, an advantage of covalent linkage is that the biomolecule becomes tethered to the surface at a single point. In contrast, noncovalent adsorption requires a larger surface area between the biomolecule and the surface to sustain attachment, which can enclose the active site or distort the protein's shape. To facilitate attachment and further enhance accessibility of the active site, reactive polymeric spacers or linkers may be used, but these require additional wet chemical steps[15]. Linker-free attachment is a scalable solution, often achieved via plasma treatment to create reactive radicals, and has been shown to facilitate attachment of biomolecules that encourage cellular adhesion[16].

Cellular adhesion is the result of cooperation between several intracellular and extracellular proteins and, for this reason, many of the proteins used to functionalize surfaces come from the ECM. The adhesion complex is termed a 'focal adhesion' and is comprised of intracellular cytoskeletal proteins, transmembrane proteins, and ECM proteins. Together, these serve as a structural and signalling conduit for the cell[17]. For biomaterials the integrin class of surface receptors, being a large family of proteins that bind a range of adhesion protein ligands, forms an important receptor class[18]. The ligands for this family include soluble adhesion proteins such as fibronectin and vitronectin[4], or ECM components like collagen and elastin[19]. All of these extracellular components have been used in some way to generate synthetic biomaterials[20].

In recent years, synthetic peptides have been used in place of whole proteins. The adhesion proteins that bind to integrin receptors contain active regions that are relatively small compared to the protein[21]. Due to their small size, these regions can be generated synthetically as peptides. So-called biomimetic peptides can provide adhesive signals to cell-surface receptors without loss of affinity[19]. These biomimetic peptides hold great potential for incorporation into biomaterials due to their clean solid-

phase synthesis and relative robustness under the post-synthesis sterilization processes applied to implantable biomedical devices. The prototypical biomimetic peptide is based on the RGD sequence (using the single letter amino acid code) derived from proteins in the ECM[22], and this has been incorporated into many biomaterials since its discovery[23]. To generate biomaterials that induce a particular response, or are specific to a certain tissue type, researchers are investigating ECM proteins for other biomimetic sequences.

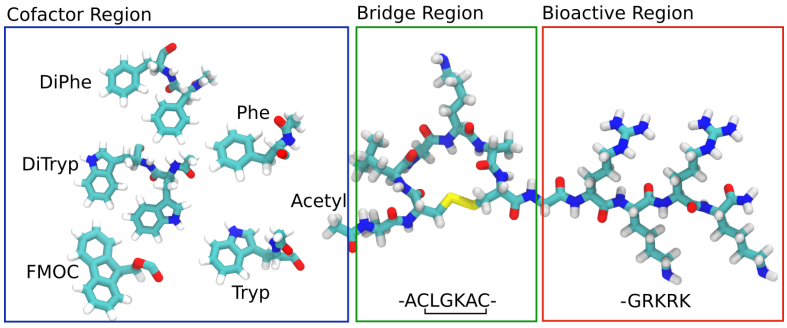
One such ECM protein is elastin - a highly elastic, fibrous protein that is found in most tissues of the body. It confers elasticity, allowing tissues to return to their prior shape after stretching or contracting, and, for this reason, is crucial for the normal functioning of virtually all tissues and organs. The soluble precursor of elastin is tropoelastin, which is constructed according to 34 exons of the genetic code[24]. These exons are the DNA sequences within a gene that code for the amino acid sequence that makes tropoelastin. A large and growing body of literature is developing the understanding of cellular interactions with the various sections of the protein. The focus of that research is on three sections of tropoelastin[25]: the N-terminal domain, which accounts for the molecule's elasticity; the hinge region, contributing flexibility; and the C-terminal domain, which is important for cellular interactivity.

As well as its functional role in maintaining tissue flexibility, tropoelastin, and in particular the C-terminus, has an important cell signalling role. The C-terminal domain of tropoelastin (named domain 36 because in humans exons 34 and 35 have been deleted[26]) is unique in three ways: its amino acid sequence, GGACLGKACGRKRK, is highly charged; it is highly conserved across species; and it contains the only two cysteine amino acids in the entire protein[27]. These cysteines are connected by a disulfide bond, the removal of which disrupts the ability of the molecule to form elastin fibres[28]. Disulfide bonds typically support structural integrity of the surrounding region[29], suggesting an important role for the highly charged C-terminal $-R_9KRK_{12}$ sequence (the two leading glycines have been removed in these simulations to match the experimental data – see Figure 1). Indeed, opposing this charge by introducing negative charges in the region prevents cell binding of tropoelastin to multiple cell types[30]. On the cell surface, the tropoelastin C-terminus interaction with cells has been localised to either of two protein receptors – glycosaminoglycans[31] or integrin $\alpha_v\beta_3$[32], depending on cell type. The C-terminus's

potent cell interactions provide motivation for this research in which we seek to study the potential of a peptide generated from domain 36 to reproduce the cellular interactions of tropoelastin.

The capacity of the C-terminal domain of tropoelastin to support cellular adhesion suggests that it can be used as a biomimetic peptide. This possibility has been explored using synthetically prepared domain 36 to adhere cells to tissue culture plates[30,33]. Given the importance of the charged $–R_9KRK_{12}$ region, it is expected that these charges are presented into solution to interact with cell-surface receptors. Recently, further interrogation of cellular adhesion to surfaces coated with the domain 36 peptide has revealed a complicating factor. Unpublished results of our research group show that when adsorbed onto PTFE, which is commonly used in biomaterial design, N-acetylated domain 36 is incapable of adhering cells without the addition of a hydrophobic N-terminal moiety. A number of N-terminal modifications (a.k.a. cofactors), based on either natural amino acids or the synthetic N-terminal blocking group FMOC, are listed in Figure 1 along with their capacity to bind cell surface receptors as per experimental data provided by Dr. Giselle Yeo (see Appendix A). These residues were chosen because they are easily ordered from most peptide synthesis companies and resulted in good cellular adhesion without further modifications.

a




b

| Variant | Cell adhesion on PTFE |
|---|---|
| Acetyl-$A_1$CLGKACGRKRK$_{12}$ | Inactive |
| Phe-$A_1$CLGKACGRKRK$_{12}$ | Inactive |
| Tryp-$A_1$CLGKACGRKRK$_{12}$ | Inactive |
| DiPhe-$A_1$CLGKACGRKRK$_{12}$ | Active |
| DiTryp-$A_1$CLGKACGRKRK$_{12}$ | Active |
| FMOC-$A_1$CLGKACGRKRK$_{12}$ | Active |

**Figure 1.** The domain 36 variants used in this chapter. a. The peptides have three structural sections—the cofactor, bridge, and bioactive regions. The cofactor can be: acetyled alanine (Acetyl), one phenylalanine (Phe),

one tryptophan (Tryp), two phenylalanines (DiPhe), two tryptophans (DiTryp) or the aromatic FMOC moiety. The bridge region consists of six residues connected by a disulfide bridge. The bioactive region is where the peptides bind to the tropoelastin receptor. It is connected to the bridge by a flexible glycine and is highly positively charged. **b.** The cell adhesion activities of the variants. The acetyl, Phe, and Tryp variants are inactive, while the DiPhe, DiTryp, and FMOC variants are active. Appendix A describes the experimental data (from Dr. Giselle Yeo) demonstrating the cell adhesion capacity.

The acetyl variant is incapable of adhering cells to a PTFE surface - some quality of the peptide has been lost by excising it from tropoelastin. The sequence close to domain 36 in tropoelastin is quite hydrophobic[34], so perhaps hydrophobic moieties contribute to binding. The Tryp- and Phe- variants have the hydrophobic amino acids tryptophan and phenylalanine as cofactors, respectively, yet do not bind cells. Only larger hydrophobic cofactors such as DiPhe-, DiTryp-, or the aromatic FMOC- moiety adhere cells. This suggest that, for surfaces, large hydrophobic cofactors like FMOC, DiTryp, and DiPhe, interact with either the surface or $-R_9KRK_{12}$ to recapitulate the effect of the native tropoelastin molecule. Doing this restores the ability to bind cells. Since it is difficult to determine the basis behind these effects using experimental techniques alone, atomistic simulations of the behaviour of the peptide variants is a useful next step.

As shown in Figure 1, the presence of a hydrophobic N-terminal modification imparts domain 36 with the ability to adhere cells to a surface. These data are for domain 36 peptides adsorbed onto a polytetrafluorethylene (PTFE) surface. Simulating the peptide variants on a similar surface might show how their behaviour is influenced by the modifications but, as outlined in the previous chapter, there are difficulties in simulating interfacial systems using nonpolarizable force fields (FF). This is because, lacking electronic polarization, these FFs also lack an enthalpic contribution to surface adhesion caused by charge polarization of molecules at or near the interface, and this may lead to nonphysical behaviour. Yet there are some indications that, for hydrophobic surfaces, nonpolarizable FFs are suitable. For peptides, which require extensive sampling to explore the conformational landscape, this is desirable because nonpolarizable FFs have lower computational demand than polarizable FFs and thus enable greater sampling.

To assess nonpolarizable FFs in modelling interfacial systems, the binding of ions at aqueous interfaces has been used as a test case. Traditionally ions have been ordered in a series, known as the Hofmeister series, by their effect on protein solubility[35]. This order has since been recognised to match their relative preference to bind at aqueous interfaces, and this property can be readily calculated with MD simulations[36]. Comparison of the calculated ion-interface binding using multiple FFs has shown that for nonpolarizable FFs, errors appear in the order of the negatively charged ions[37]. Positively charged ions, which are less polarizable, are modelled more accurately. This error has been attributed to the lack of polarization, and used as justification for the rejection of fixed charge force fields for interfacial modelling[38]. However, decomposition of the energetic contributions to surface binding reveals that the Hofmeister series actually arises from a combination of many factors, and that ion- or water-polarization alone cannot explain ion-surface binding[39]. For peptides, which are more variable in terms of shape and interactions with water, other factors such as the shape of the cavity required to solvate each amino acid would be expected to contribute even more strongly than for ions.

Peptides have several energetic contributions to surface adsorption: the dispersion force; the hydrophobic effect; entropic forces; and steric effects from the protein backbone. At the interface between water and a hydrophobic surface, which has a low dielectric constant, these are all explicitly modelled using nonpolarizable force fields[40]. Given this, polarization may be only a minor determinant of surface binding, where the combined effects of the above explicitly defined forces are the majority. For certain surfaces, estimates of the energetic contribution of polarization to surface binding are as low as 10%[41]. Indeed, on flat hydrophobic surfaces, a comparison of experimental and simulated peptide adsorption shows good agreement[42]. On such surfaces, where all energetic contributions except for polarization are accounted for, modelling by nonpolarizable FFs can still yield worthwhile results.

Before simulating interfacial systems, the model system must be correctly parameterized. This requires determining accurate intermolecular interaction parameters and surface topography. Both of these are important particularly for metals: these surfaces have high dielectric constants, leading to strong interaction with molecules in the aqueous phase. This can create attractive image charges – induced effective charges residing within the metal in response to real charges in the aqueous phase[41] – and

lead to polar binding sites created by specific crystal facets[43]. These are highly sensitive to the parameterization of the metal surface.

For hydrophobic surfaces, which have low dielectric constants and negligible polarizability, parameterization is simpler. Dispersion forces arising from the surface are small, so there are no binding sites associated with the crystal facet, and image charges (which are repulsive for surfaces with low dielectric constant) arise directly from the oriented dipoles of water molecules at the interface, and not by polarization of the surface itself[40]. This means the main determinant of molecular binding to hydrophobic surfaces is the density and orientation of water at the interface[44,45] and the hydrophobic effect, both of which are included in nonpolarizable FFs. Nonpolarizable simulations of adsorption to and diffusion across hydrophobic surfaces has been tested and shows qualitative and quantitative agreement to experiment[42,46]. While there are demonstrated issues for reproducing ion-surface binding with nonpolarizable FFs, peptide adsorption is more complex, depending upon many other factors. Given that nonpolarizable FFs have shown agreement with experiment, it is reasonable to consider surface interactions of biomimetic peptides on hydrophobic surfaces without polarization.

In this chapter, we present MD simulations of the interaction between a hydrophobic surface and biomimetic peptides derived from tropoelastin to provide understanding of the *in vitro* cell adhesion trends. We apply nanoscale water contact angle measurement to parameterize a hydrophobic surface model that represents PTFE, and use this to generate 2 μs simulations of each of the four domain 36 variants in the surface-bound state. Analysis of the peptide conformation and distance from the surface demonstrates how the hydrophobic cofactor affects the surface-bound state. These results prompt detailed study of the molecular determinants and strengths of atomic interactions at the surface, which is used to offer an explanation of the experimental data. This approach is an ideal complement to experimental methods, since it illustrates the role of cofactor molecules in surface adhesion and rationalizes this in molecular detail.

## 4.2 Methods

### 4.2.1 Development of a surface model

Developing a surface model has two practical aspects: creation of an *in silico* representation of the physical surface; and parameterizing the forces of attraction or repulsion between it and the surrounding atoms such as water, ions, and peptide. The model used here ('surface model') is a flat surface made of a hydrogen-terminated diamond slab. While the model does not have any three-dimensional surface topography, the effects of this are more prominent for strongly-interacting surfaces such as metals and less so for hydrophobic surfaces. The movement of individual chains of PTFE ('PTFE surface') is expected to be slow compared to the diffusion of a peptide. Since there are no specific surface interactions, and the surface is practically stationary, the flat surface model is used. This approximates well the average of the PTFE surface where the processing-induced roughening would be on a length scale much larger than that of the simulations, presenting a locally flat surface. In addition there are numerous examples of successful conclusions drawn from flat idealised surfaces[47-49] The diamond slab was generated by replicating and translating the coordinates of carbon atoms of the unit cell of a diamond crystal to make the desired size. Hydrogen atoms were added to the terminating carbons. These processes were automated by code written in house using Python and 'bash' programming languages. The slabs were 1.6 nm high in order to be larger than the non-bonded cut-off used in simulations, ensuring that protein and water molecules at the surface did not interact with water on the other side of the slab in the adjacent periodic cell.

The parameters governing the interaction between atoms within the diamond slab and between the diamond slab and the aqueous phase must account for bonded and non-bonded interactions. All of these parameters must be explicitly defined or the simulation code will fail. Since there are no existing parameters for diamond in the CHARMM27 force field, these parameters were added to the force field manually. The bonded parameters for carbon-carbon and carbon-hydrogen bonds were created by analogy with existing parameters for CT2-type carbons and H-type hydrogens from the CHARMM27 force field. The non-bonded parameters are partial charge $q$ and the LJ parameters $\sigma$ and $\varepsilon$. For highly hydrophobic surfaces the partial charge of the surface atoms are set to zero, since these surfaces have low or zero polar surface energy. While this is in contrast to hydrophobic molecules in some force fields, these are likely introduced to reproduce their complex conformational behaviour that is not applicable

here, and it is in line with previous research using zero partial charge for solid hydrophobic surfaces[46,47]. The position of the LJ potential y-intercept, $\sigma$, was the same as for CT2- and H-type atoms. The depth of the LJ potential minimum, $\varepsilon$, was altered to change the non-polar surface energy of the model.

### 4.2.2 Surface model molecular docking of domain 36 variants

Molecular docking of the domain 36 peptides was performed in order to quickly find the lowest energy conformation, reducing the equilibration time of the MD simulations. Docking calculations use empirically fitted scoring functions to find conformations of small molecules, such as drugs or peptides, in energetic minima[50]. They contrast with molecular dynamics simulations in that they have no explicit treatment of solvent molecules and do not evolve the conformations through time. The original intent of molecular docking was to generate realistic conformations, rather than calculating binding affinities, and the failure of docking to predict binding affinities continues to be noted[51]. In contrast, generating accurate conformations was solved early on by many docking algorithms. For the domain 36 peptides, where the intention of molecular simulation on the surface model is exploratory data analysis, docking is an ideal solution for quickly generating realistic conformations on the surface.

Docking was performed using the AutoDock Vina program[52]. This requires .pdbqt files for receptor and ligand, which were generated using AutoDock 4[53]. For each peptide variant, the docked poses could be aligned by minimizing the RMSD of the backbone alpha-carbon atoms. Poorly clustered docking poses are indicative of multiple energetic minima of similar depth, and due to the difficulty predicting binding affinity with high accuracy, choosing the best surface-bound conformation is non-trivial. As the structure of the top 20 aligned poses all appeared to be of the same conformation by aligning and visually inspecting them, the affinity problem was judged to be a minor concern in this case. As a result of this, the single top-scoring pose was chosen as the initial peptide coordinates for MD.

### 4.2.3 Simulation system

Equilibrium simulations of the domain 36 variants were performed in order to explore their behaviour while adsorbed on the model surface. Equilibrium simulations used a model system based on the PTFE model surface. The surface was hydrated with ≅ 4000 explicit TIP3P water molecules using the VMD *solvate* plugin[54]. The hydrated surface was used to generate systems for the four variants - the docked

peptide structure was added and water molecules within 2 Å were removed. These systems were neutralized, and the salt concentration set to 150 mM, with $Na^+$ and $Cl^-$ ions using the VMD *ionize* plugin. The water was equilibrated by restraining the peptide atoms and simulating for 2 ns, after which the peptide restraints were released. A harmonic restraining potential with a force constant of 0.1 kcal/mol/$Å^2$ was applied to a one-atom thick layer of carbons in the middle of the surface to prevent the system drifting. The systems for umbrella sampling of arginine and lysine, and pulling simulations of the cofactors, were generated in the same way except for the number of water molecules. The umbrella sampling systems used ≅ 900 water molecules, while the pulling simulation systems used ≅ 4000 water molecules. These allowed for at least one cut-off distance between the peptide or amino acid in the fully solvated state and the bottom of the PTFE model of the next periodic image.

### 4.2.4 Water Contact Angle

The water contact angle on the model surface was measured to parameterize the LJ $\varepsilon$ parameter. For PTFE, which has zero polar surface energy, the water contact angle is determined by the apolar surface energy. This is determined by van der Waals forces that are approximated by the LJ potential[55]. It follows that altering the depth $\varepsilon$ of the LJ potential will change the contact angle. This is a convenient method to parameterize the surface because water contact angles are readily determined by experiment or from published values.

Measuring the macroscopic contact angle on the nanoscale requires modelling nano-droplets of water. This is viable, but the small size of the droplets compared to their curvature introduces an additional force called line tension[56]. Accounting for this requires simulating multiple droplet sizes, which represents added computational cost. Line tension can be removed by taking advantage of the periodic boundary conditions, creating a semi-cylindrical droplet with no curvature that extends into infinity along one axis[57]. The water contact angle was determined using a semi-cylindrical nano-droplet of approximately 3000 water molecules on the model surface. This droplet size is larger than the commonly used size of 2000 water molecules from several years ago[58], reflecting greater computational capabilites, but as explained above semi-cylindrical droplets have no size dependence. The nano-droplets began as cubes on the surface, but within 1 ns formed a droplet shape. These droplet and surface systems were simulated using equilibrium dynamics to collect data.

Analysis of the simulations was performed using custom-made Python scripts that used the MDAnalysis Python library for accessing molecular dynamics trajectories. The trajectory was split into blocks for block averaging (see Appendix B). Looping through every frame of each block, first the coordinates of the water droplets were transformed to account for drift of the centre of mass of the droplet. Then, the coordinates of the water molecules were binned into 0.1 nm bins in the X and Y dimensions (bisecting the droplet). The density of water occupancy of each bin was calculated, creating a 2-dimensional histogram of the droplet shape. The edges of the droplet were determined by the density (see Supplementary Figure 1), creating a 2-dimensional semi-circle. This was fit to a circle using the method of least squares[59]. The water contact angle was then calculated using:

$$\theta = 2\times \tan^{-1}\frac{h}{d}$$

where $\theta$ = contact angle, $h$ = droplet height, and $d$ = droplet width at the surface.

### 4.2.5 Cofactor binding strength

Having determined some behavioural differences between the variants on the surface using equilibrium simulations, the surface adhesion of the cofactors was measured in more detail. In particular, the relative adhesion strength on the surface was measured. Because only the relative adhesion strength was required, a combination of pulling simulations and the Jarzynski equality was used, as opposed to umbrella sampling (which would provide more precise results at the cost of greater simulation time). The relative adhesion strength of the cofactors was measured using steered MD and the Jarzynski equality. The cofactors were simulated for 500 ps, saving coordinates every 50 ps. These were the starting coordinates for 10 simulations using steered MD. A moving harmonic restraint of strength 20 kcal/mol/Å$^2$ was applied to the centre of mass of the cofactor atoms. The restraint was moved away from the surface at a speed of 0.25 nm/ns and the force applied to the centre of mass was written every 0.02 ps. This combination of pulling strength and speed ensured that the movement of the peptide closely followed the position of the restraint point, generating a time series of the force that could be used to calculate the work. Using Python scripting the area under this time series was integrated to give the work, and the free energy change was calculated using the Jarzynski equality[60]:

$$e^{-\Delta G/k_B T} = \langle e^{-W/k_B T} \rangle$$

where $\Delta G$ is the free energy change, $k_B$ is the Boltzmann constant, $T$ is the temperature, and $W$ is the work.

### 4.2.6 Umbrella Simulations

In contrast to pulling simulations and using the Jarzynski equality to determine relative potentials of mean force (PMF) of the cofactors, detailed PMFs of the binding of two charged residues were required. These simulations used umbrella sampling, which provides precise measurement of PMFs at the cost of greater simulation and analysis time. These were more feasible for the two residues analysed here, since the simulation system is smaller. PMFs of arginine and lysine on the model surface were calculated using umbrella sampling. Harmonic restraining potentials were applied to constraint the Z-coordinate of the centre of mass of the amino acids. There were 30 restraints centred at 0.05 nm intervals moving away from the surface, generating a PMF of 0 – 1.5 nm from the surface. Positional data from the simulations were recorded every 0.8 ps, and deconvoluted to yield a PMF using a software implementation[61] of the weighted histogram analysis method[62]. This method generates PMFs by deconvoluting the positional data of adjacent simulations to determine the relative preference for the amino acid across the reaction coordinate (in this case, the distance from the surface). Obtaining accurate results thus requires good overlap of the explored region of the PMF between adjacent simulations. The 0.05 nm separation between biases in adjacent simulations here led to considerable overlap as well as providing convenient generation of input files. The systems were simulated for 5 ns. The first nanosecond of data was discarded as equilibration and after 5 ns adding more data did not alter the PMF, which indicates convergence.

## 4.3 Results and Discussion

### 4.3.1 Surface model parameterization

In order to simulate domain 36 variants on a hydrophobic surface with MD, a model of a PTFE surface was developed. This was achieved by parameterizing one of the Lennard-Jones parameters of the

carbons in a flat, hydrogen-terminated diamond slab. This parameter, $\varepsilon$, approximates the effect of van der Waals forces in nonpolarizable MD simulations[63] and, for hydrophobic surfaces, also determines the surface energy. Parameterization was achieved by calibrating $\varepsilon$ to reproduce the surface energy of PTFE.

A cross-section of the simulated system is shown in Figure 1A. The surface energies of the diamond slabs at varying $\varepsilon$ were determined by measuring the contact angle of a nanodroplet of water consisting of explicit water molecules interacting with each other and the model surface. Using periodic boundary conditions, the water extends to create an infinitely long cylindrical water nanodroplet, having a linear contact boundary between the surface and the droplet. Spherical droplets have circular contact boundaries with the surface and, due to the high ratio of curvature to the droplet volume, have significant forces distorting the droplet[57]. The simulation system used here avoids the additional simulations needed to account for these forces. The cylindrical nanodroplets were stable, allowing for the determination of the macroscopic contact angle.

At the nanoscale, thermal motion of water molecules resulted in perturbations of the nanodroplet that deviate from a circular shape. By averaging across several hundred picoseconds, as well as correcting for lateral movement, the fluctuations converge to a circular arc. Plotting the averaged number density of water molecules results in a density map that can be fitted to a circle, from which the contact angle was measured (Figure 1b). The perimeter of the droplet was chosen as the point where the density was half that at the centre of the nanodroplet: This definition encompasses the majority of the water molecules used in the simulation without increasing the size of the nanodroplet beyond normal water density (Supplementary Figure 1).

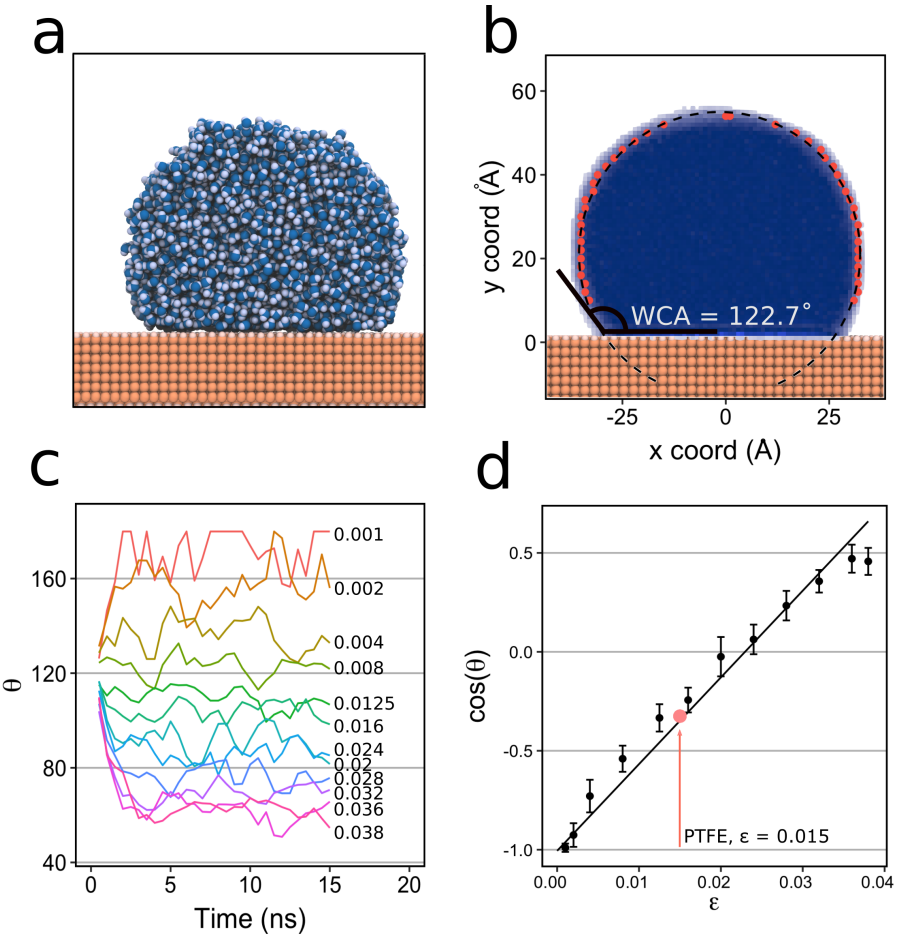



**Figure 2:** The water contact angle from droplet simulations can be used to calibrate $\varepsilon$ for the diamond surface. **a)** Cross-section of the simulation system - the water molecules (blue, white) form a cylindrical nanodroplet on the diamond surface (orange). **b)** Density map averaged over 1 ns of simulation. The edge of the droplet (red points) is used to fit a circle (black, dashed line), from which the contact angle is calculated (as indicated on the left side of the circle). **c)** Simulated nanodroplet contact angles for varying $\varepsilon$. All droplets were equilibrated by 2 nanoseconds and the average of the remaining simulation was used to determine the contact angle. **d)** The $\varepsilon$ and the cosine of the contact angle show a linear relationship up to $\varepsilon \cong 0.036$. This can be used to parameterize the diamond surface to represent materials with low polar surface energy such as PTFE.

Because the water contact angle fluctuates due to thermal motion, 30 nanoseconds of nanodroplet simulation was used to obtain better statistics. The time series of the nanodroplets on surfaces with varying $\varepsilon$ is shown in Figure 1c. All simulations were started from the same coordinates and the first two nanoseconds were discarded since the system is clearly moving towards equilibrium during this time. The simulated contact angles after two nanoseconds fluctuated around a consensus value, so

only these remaining data were used to calculate the contact angle. The time series clearly shows that increasing the $\varepsilon$ of the surface reduces the contact angle, indicating that the surface has higher surface energy and is more hydrophilic. The data points presented in Figure 1c are the contact angle calculated as the average over blocks of 500 picoseconds. Block averaging (see Appendix B) of the trajectories determined that increasing the size of the blocks does not change the standard deviation. These data points are therefore uncorrelated, and can be averaged and used to generate error bars.

Fitting the cosine of the contact angle to $\varepsilon$ yields a linear relationship (Figure 2d). According to Young's law, the cosine of the contact angle is linearly dependent on the interaction energy between the water and the surface[64]. In these simulations, where the only interaction between the liquid and solid phases is from the Lennard Jones potential, it follows that the cosine of the contact angle is linearly dependent on $\varepsilon$, the maximum depth of the LJ potential energy function. A full theoretical description has been provided along with droplet simulations previously but, in that work, the fit erroneously led to contact angles less than 180° at $\varepsilon > 0$[47]. Removing line tension by using a cylindrical droplet, as shown here, avoids this error. This fit confirms that the nanodroplet-surface system models the effects of van der Waals forces in macroscale surface-droplet systems and so provides a useful way to parameterize model flat hydrophobic surfaces. Using these data, $\varepsilon$ can be set to match the water contact angle of many hydrophobic polymers. To model PTFE, the water contact angle of the polymer can simply be measured or read from a table. Then, by finding the cosine of this angle and using the relationship determined in Figure 2d, the correct $\varepsilon$ value to recreate that surface is determined.

It should be noted that there is a practical upper limit to $\varepsilon$. Increasing $\varepsilon$ beyond a certain point does not accurately reflect most hydrophilic surfaces – in reality the surface energy will increasingly have some polar component. This is demonstrated for the contact angle at $\varepsilon > 0.036$, which deviates from the trend. For surfaces more hydrophilic than this, some polar moieties should be included but these will require further parameterization of the charge distribution. For PTFE, whose surface energy is 100% nonpolar, this poses no problem and so this approach is suitable to parameterize a surface model of PTFE, whose contact angle is around 109°[65] and so has $\varepsilon = 0.015$.

### 4.3.2 Surface-peptide docking

Three of the hydrophobic N-terminal modifications cause domain 36 peptides to become cell-adherent (FMOC, DiPhenylalanine, and DiTryptophan). This is likely due to their effect on either the conformation, by unlocking active binding structures[66], or through changing the orientation to present the bioactive region away from the surface and towards the cell. The conformations of the variants can be determined using molecular dynamics simulations on the PTFE surface model but, due to energetic attraction to the surface restricting movement, equilibrating the peptides is slow. Molecular docking with AutoDock Vina, which uses flexible fitting to find low energy conformations[52], is initially used to speed equilibration.

Figure 3 shows the top-scoring docking conformations of six domain 36 variants on the PTFE surface model. The variants can be grouped in pairs by the direction the cofactor bends with respect to the bioactive region. The Acetyl- and FMOC- variants have the cofactor bending to the right when viewed from above. The Phe and Tryp variants show an approximately even split between left and right, while the DiPhe and DiTryp variants have the cofactor bending to the left.

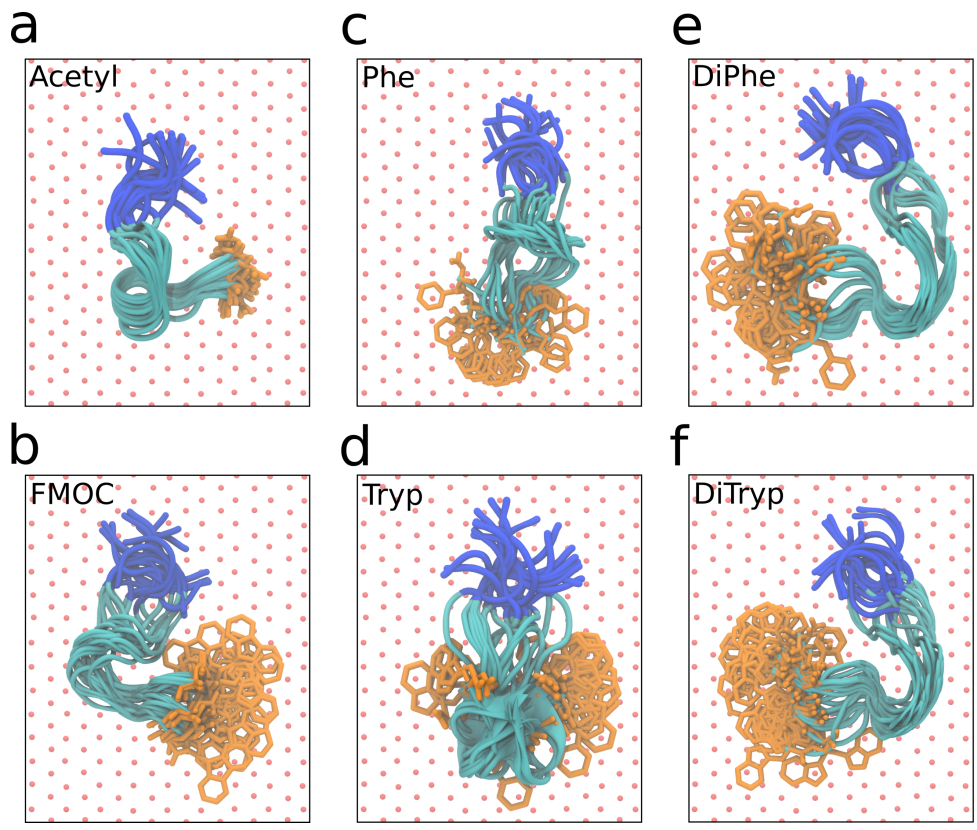



**Figure 3:** Molecular docking calculations show the effect of the cofactor on domain 36 variant conformation. Birds-eye view of alignments (**a-f**) of the top 20 conformations of each variant showing the N-terminal cofactor (orange, stick representation), peptide backbone (cyan, ribbon representation) and bioactive region (blue, ribbon representation) on the top-most layer of the PTFE model (red, ball representation). The docked conformations are aligned in the x-y plane to facilitate comparison. There are three types of binding defined by the direction of bending of the cofactor: The Acetyl and FMOC (**a, b**) variants bend to the right, the Phe and Tryp (**c, d**) variants bend in both directions, while the DiPhe and DiTryp (**e, f**) variants bend left.

These surface binding modes bring to light the role of the cofactor size. The Acetyl variant can be viewed as the ‘native’ surface-bound structure, having the smallest cofactor that is not expected to interact strongly with the rest of the peptide or the surface. The Phe and Tryp variants have intermediate sized cofactors, which adsorb more strongly onto the surface and shift the orientation of the cofactor to the left via torsion of the peptide backbone. The largest cofactors, DiPhe and DiTryp, have the largest size, and fully shift the cofactor orientation to the left, likely because of stronger interaction with the surface. This does not occur for the FMOC- variant which, while similar in size, has high flexibility due to the non-peptidic connection to the peptide, allowing the native conformation to exist. According to these results, the cofactor plays a role in the conformation of the peptide as a

whole, and this could include the bioactive region. However, the docking results do not show a common conformation for all of the cell-adhering variants. So, to further probe the structure, molecular dynamics simulation is required.

### 4.3.3 Surface-peptide equilibrium molecular dynamics

2 μs equilibrium MD simulations of four of the domain 36 variants (Acetyl, FMOC, DiPhe, DiTryp), starting from their docked conformation, were used to explore the conformational landscape in more detail. The Phe and Tryp cofactor variants were not analysed using equilibrium MD simulations because they did not show cell adhesion. They are, however, included in the more detailed follow up simulations. While the Acetyl variant also lacked cell adhesion, it provides a useful starting point to compare conformational changes in equilibrium simulations, since it is closest to the native tropoelastin sequence and represents the surface-bound structure without the influence of any large cofactor. Due to the sampling restrictions imposed by interactions with the surface, these simulations are unlikely to completely sample the adsorbed conformational landscape. While there are techniques that enhance sampling by escaping from energy basins[67,68], these are aimed at generating complete Boltzmann-weighted population ensembles and are associated with significant computational and analytical cost. The purpose of these equilibrium simulations is exploratory data analysis to guide the more targeted simulations that follow.

An example of the simulation systems used is shown in Figure 4a. Compared to docking, MD simulations are more accurate because they explicitly include water molecules, use FFs designed to reproduce physical interactions, and evolve the peptide conformation through time. All four peptides stayed adsorbed to the surface for the entirety of the simulations. They freely translated across the surface as has been observed previously on flat hydrophobic surfaces[42,46]. This indicates that no binding sites are created by the crystal facet surface, as occurs for metals[43]. Additionally, the peptides had similar translational diffusion coefficients, so this is unlikely to be the mechanism that causes cell-adhesion.

Aligned snapshots from the MD simulations are shown in Figure 4b, showing that the conformations seen in docking are largely preserved. The main movements are rotations of the bioactive region and

flexing about the bridge region that changes the position of the cofactors with respect to the surface and the bioactive region. These movements indicate areas where differences between the peptides that affect their cell adhesive activity might arise.

The secondary structure of the variants may be a discriminating factor that separates the cell-adhering variants from the Acetyl variant. For a comparison of this, Figure 4c shows the time series of the peptide secondary structure elements observed in each simulation. At first, there is no obvious differentiating factor that separates the variants. Each variant has a turn (beta turn, 3-10 helix, or alpha helix) across the bridge region, likely caused by the torsion from the disulfide bond. However, this turn element is shifted by 1-2 residues towards the N-terminal for the cell-adhering variants compared to the Acetyl variant. Since it only occurs in these variants, it is likely due to the tight adsorption of the hydrophobic cofactors to the surface.

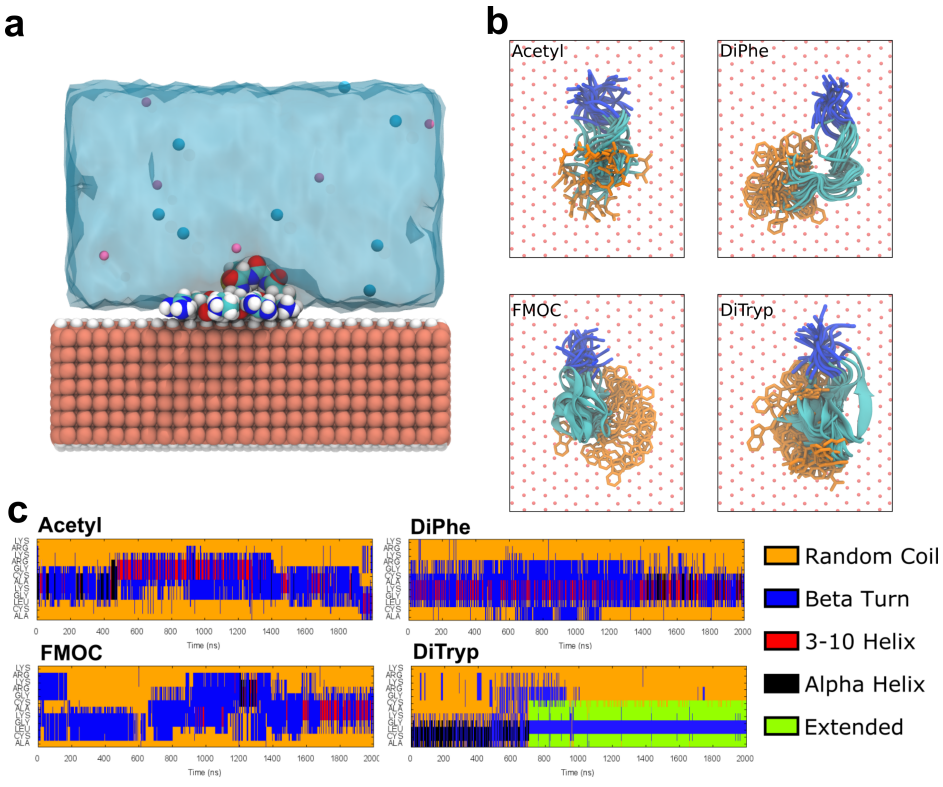



**Figure 4:** Equilibrium simulations show no discernible effect of the cofactor region on the secondary structure of the bioactive region, hence more high-resolution analysis is required. **a)** Cross-section of the simulation system used for equilibrium MD of the domain 36 peptide variants on a PTFE model. The initial coordinates for the peptide (space-filling model, blue, white, red and yellow) were taken from docking simulations. **b)** Aligned snapshots, shown as ribbons and stick models for clarity, taken from every 200 ns of the equilibrium simulations.

The snapshots show that the docking conformations are largely preserved, and that for the hydrophobic variants, the cofactor (orange) can bend around the bridge (cyan) to contact the bioactive region (blue). **c)** Time series of secondary structure elements for the four peptide variants across the equilibrium simulations show no discriminatory elements in the bioactive region.

The charged bioactive region, made up of $–R_9KRK_{12}$, is known to be the active epitope in domain 36[32]. Changes in this region may affect how it interacts with the cell. The positive charges of the amino acids in this region are thought to be the chemical determinant of binding[28] and so their position with respect to each other, and to the surface, is a possible point of difference between the variants. Density plots of the peptide backbone angles of each residue, also known as Ramachandran plots, describe the bond angles in the peptide backbone between each amino acid and thus the structure of the backbone. These plots provide 'at a glance' visualizations of the entirety of a simulation, allowing a fast comparison of the conformational landscape that was explored. Analysis of the Ramachandran plots of the charged amino acids ($-R_9KRK_{12}$) in the bioactive region does not show any difference that separates the FMOC, DiTryp, and DiPhe peptides from the Acetyl peptide (Supplementary Figure 2).

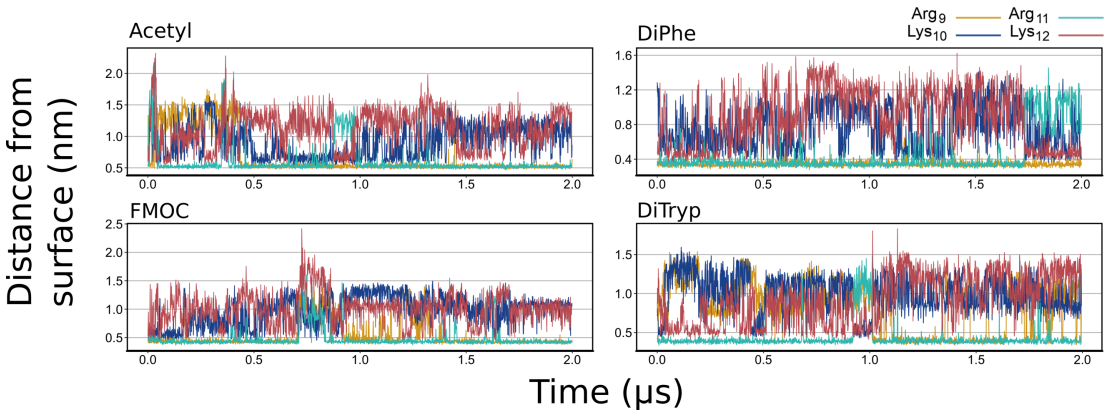



**Figure 5.** Time series showing the surface separation distance of the binding residues. In each case, both arginines (orange, cyan) are bound to the surface, solvated by the first interfacial water layer. The lysines (blue, red) are solvated for a majority of the simulation. The movement of the cofactor therefore has no effect on the surface-adhesion behaviour of the bioactive region.

The position of the positive charges in the bioactive region relative to the surface may affect how accessible this region is to the cell. Figure 5 shows time series of the surface separation distance of

these charges for the four variants. In each case, the arginine charges are closely adsorbed to the surface. In contrast, the lysine charges are separated from the surface by a further ~4 Å. This is potentially a result of the differences in both the topology and the flexibility of these amino acids. Arginine, which has a lower charge density due to separation of the charge across two amine groups, also presents a flatter and wider profile than the pointed lysine. This facilitates good enthalpic interactions with water molecules at the surface. The lysine, which is more flexible, allows for the carbon stem to lay flat along the surface, a consequence of the hydrophobic effect, while the charge sits further away. Yet again, however, there are no apparent discriminating factors between the cell-adhering and non-cell-adhering variants from the equilibrium simulations alone.

Since the conformation of the bioactive region does not show any differences that could explain the cell binding properties of the variants, it may be that the orientation of the peptide on the surface, which facilitates contact between the bioactive region and the cell, is changed by the presence of the cofactors. To better understand how the cofactors affect the peptide orientation, the surface adsorption behaviour of both the cofactor region and the bioactive region needs to be analysed.

### 4.3.4 Surface binding of positively-charged amino acids

Docking and equilibrium molecular dynamics did not demonstrate any factors that could discriminate the cell-adhering from the non-cell-adhering variant. Assuming a reasonable characterisation of the conformational behaviour at the surface they do, however, show interesting differences in the surface adsorption of different parts of the peptide. In particular, there was tight adsorption to the surface by the hydrophobic cofactors (but not the Acetyl) and some differences in the preference for the surface between the charged Arginine and Lysine residues. These differences imply a more detailed study of the surface adsorption behaviour is needed. This is because an alternative mechanism of cell adhesion to conformational differences involves the presentation of the bioactive region to the solution. By preferentially adsorbing through the cofactors, the active variants could be altering the presentation of the bioactive region, facilitating better contact with cell-surface receptors. This could be a function of the adsorption strength of the bioactive region.

Figure 5 shows some differences in the adsorption of the charges in the bioactive region. To further explore this, umbrella sampling of these residues on the PTFE model surface was performed. Umbrella sampling is a MD technique using many replicate simulations with a change in position of a bias to force sampling across a reaction coordinate and so determine a potential of mean force (PMF), which describes the energetic interaction of the atoms along the path of movement. It is an ideal technique to measure the atomic determinants controlling adsorption because it offers fine spatial resolution and accuracy. Arginine and lysine have the same charge (+1e) but different molecular volumes and shapes, so they are a suitable model to determine how these factors influence surface binding. This analysis is especially useful here since it also controls how the cofactors adsorb.

The PMFs of the alpha-carbon of arginine and lysine moving away from the surface is shown in Figure 6a. Arginine has a slightly larger free energy of adsorption and this can be rationalized by observation of the lowest-energy binding structures shown in Figures 6b and 6c. For arginine, whose positive charge is spread across two amine groups, the charge lays flat along the surface, while the lysine's single amine group is lifted away from the surface. This is due to the layers of water at the interface – Figures 6d and 6e show how the charge and mass density of water near the surface is separated into two layers. This has been observed previously[47,69] and is one determinant of peptide binding on hydrophobic surfaces[45]. Arginine, whose charge is spread across a planar moiety, lays flat in the first water layer, getting favourable enthalpic contacts with the high density negative charge in this region. Lysine, in comparison, achieves better solvation between the first and second water layers due to its more rounded charge density, which points both up- and downwards to the first and second layers of negative charge created by the interfacial water. Surface adsorption to PTFE is generally considered to be caused by exclusion from the aqueous phase. The conformations observed here demonstrate that shape and polarity also influence binding by interacting with the surface water layers. These effects are not limited to charged residues, so they may similarly influence the way the cofactors bind.

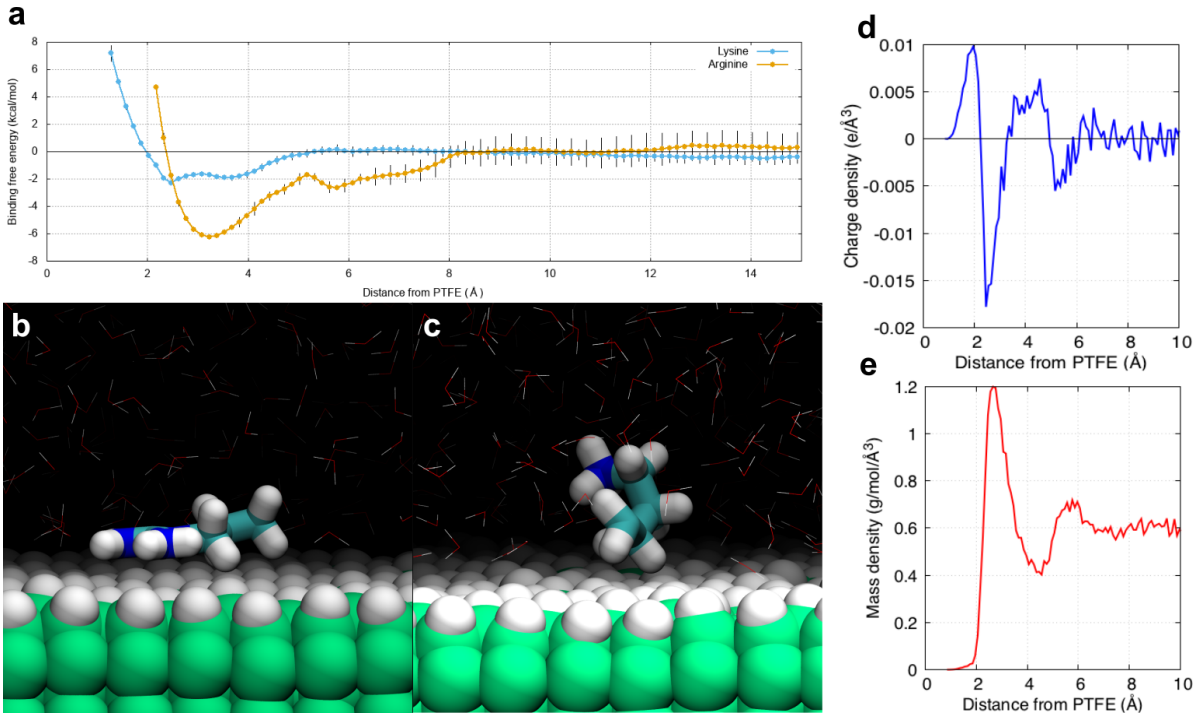



**Figure 6.** Potentials of mean force (PMF) for the positively-charged arginine and lysine amino acids demonstrates how the interfacial water layers change surface binding. **a)** PMFs of the alpha-carbons of arginine and lysine amino acids moving away from the PTFE surface model. **b)** Lowest energy surface-bound conformation of arginine. **c)** Lowest energy surface-bound conformation of lysine. **d)** Charge density of water at the interface showing dipole moments created by oriented water molecules. **e**) Mass density of water at the interface showing alternating layers of high- and low-density water.

### 4.3.5 Peptide orientation on the surface

The equilibrium MD simulations were further analysed to probe the adhesion of the cofactor and bioactive regions on the surface. Figure 7 shows the surface separation distance of the centres of mass of the cofactors and the bioactive regions for the four domain 36 variants. For the charged bioactive regions, in each case there is prolonged association with the surface. It is capable of desorbing from the surface, and this can be seen particularly for the FMOC variant from ~700-900 ns. However, since this also occurs for the Acetyl variant, and is not observed for either the DiPhe or DiTryp variants, this occurrence is not likely caused by any interaction with the cofactor and so is not, by itself, facilitating cell adhesion. The ability of the bioactive region to desorb spontaneously does suggest that it can easily unbind from the surface, explaining how it binds to the tropoelastin cell surface receptor.

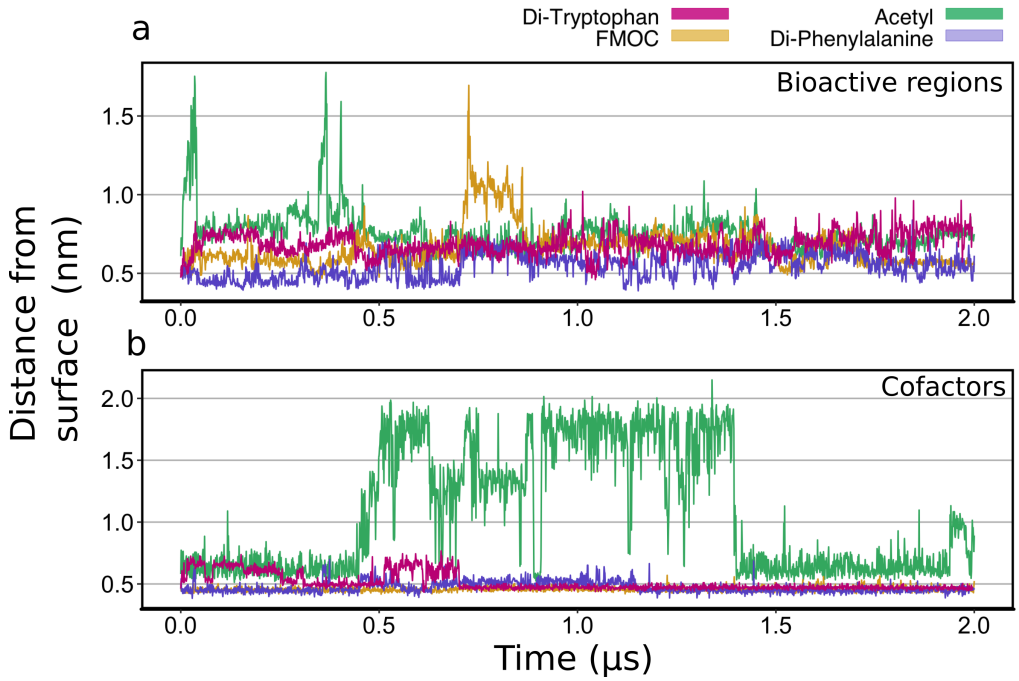



**Figure 7.** Time-series showing the distance from the surface of a) the binding or b) the cofactor. The bioactive region is mostly surface-bound in each case, but is capable of unbinding 1-4 residues for periods of up to 100 ns. Unlike the hydrophobic cofactors, the acetyl cofactor is only weakly adsorbed onto the surface and is capable of unbinding completely for extended periods.

On the other end of the peptide, the movement of the cofactor itself may affect how the peptide interacts with cells. The larger surface area of the three hydrophobic cofactors results in strong surface adsorption for the entirety of the simulations. In contrast, the non-cell adhering Acetyl variant demonstrates complete desorption of the cofactor region, with the peptide being anchored to the surface by only the bioactive region. This is the only discriminatory factor observed so far that separates the hydrophobic cofactor variants from the Acetyl variant.

### 4.3.6 Surface adhesion strength of the cofactors

Since neither the conformation of the peptides nor the adhesion of the bioactive region are discriminatory for cell adhesion, the interaction between the cofactor itself and the surface must be facilitating cellular adhesion. To explore this, the relative strength of adhesion of the cofactors was determined. Pulling simulations, in combination with the Jarzynski equality, were used to determine the adsorption energy of the cofactors. This method, while not as accurate as umbrella sampling, can

more efficiently calculate the relative adhesion strength[70]. Figure 8a shows the simulation system used, in which the cofactor, starting from the adsorbed conformation, was pulled from the PTFE model surface. This generated the PMFs shown in Figure 8b. The order of the PMFs supports the justification of the docked binding poses from Figure 3 – the Phe and Tryp cofactors have reduced surface adsorption strength compared to the DiPhe and DiTryp variants, and so are less able to introduce torsion into the backbone.

The surface adsorption on the PTFE model is a function of both the shape, which contributes to the hydrophobic effect by displacing water molecules back into the aqueous phase, as well as the presence of polar groups, which interact favourably with dipoles aligned parallel to the surface. This was demonstrated by the preferred surface binding conformations of arginine and lysine residues seen in umbrella sampling, and can be observed again in the surface adsorption trends of the cofactors. Figure 8c shows a comparison of the molecular volume and the number of polar chemical groups (defined as oxygen or nitrogen-containing moieties) in the side-chains of the cofactors. The molecular volume contributes to surface adsorption by expelling water from the surface back into the aqueous phase[71], while the number of polar moieties indicates the capacity for favourable enthalpic interactions with the high density water layer at the surface.

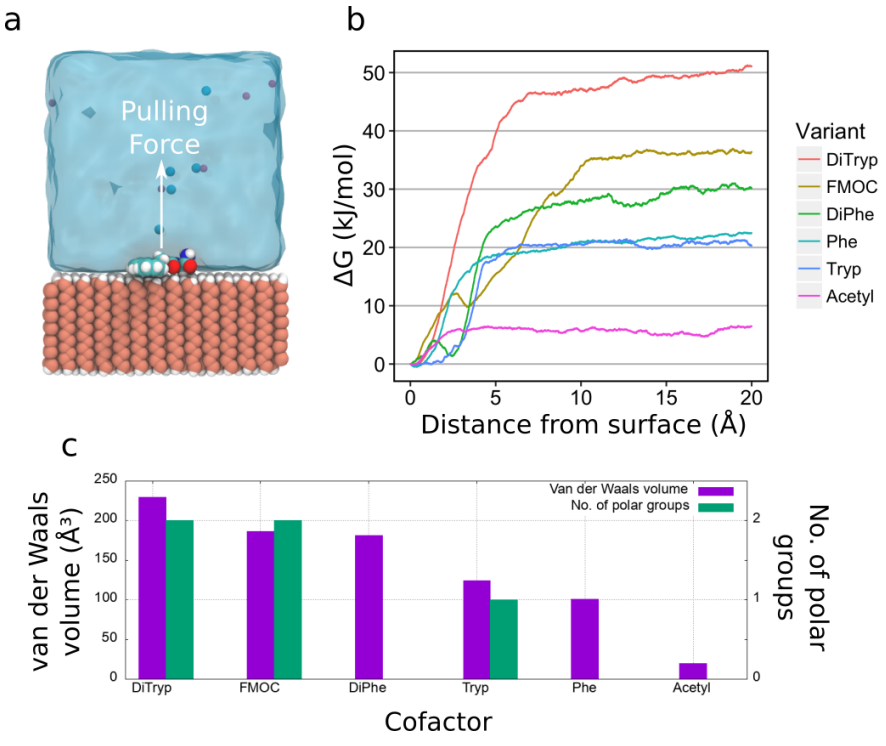


**Figure 8.** Pulling simulations of the cofactors reveal the order of surface adhesion strengths of these regions. **a)** Simulation system used for calculating potentials of mean force (PMF) for surface binding. The cofactor (space-filling model, blue, white and red) initially is equilibrated at the PTFE model surface, and is pulled away through a water box (transparent, blue). The PMF is estimated using the Jarzynski equality. **b)** PMF of surface adhesion for the peptide cofactors. The cofactors from the cell-binding domain 36 variants (FMOC, DiTryp, DiPhe) show the highest surface adhesion. Single-amino acid cofactors have intermediate adhesion strength, while the Acetyl cofactor shows minimal adhesion. **c)** Comparison of the molecular volumes and the number of polar moieties of each cofactor. Higher molecular volume contributes to surface adsorption by expelling water back into the aqueous phase, while some polar character also interacts favourably with the in-plane dipole moments of the surface water layer. These trends justify the surface adsorption strengths calculated by pulling simulations.

The order of surface adsorption strengths correlates well with the combination of molecular volumes and the number of polar moieties. The DiTryp cofactor has the highest surface adsorption, which occurs because of the combination of higher molecular volume in addition to the polar amine group on each amino acid – these make contact with water molecules in the first water layer, reducing the enthalpic cost of water molecules leaving the aqueous phase. The FMOC cofactor also has two polar moieties but smaller molecular volume and, consequently, has a smaller surface adsorption strength. This trend continues down to the Acetyl cofactor with no polar moieties, the smallest molecular volume, and the lowest surface adsorption strength.

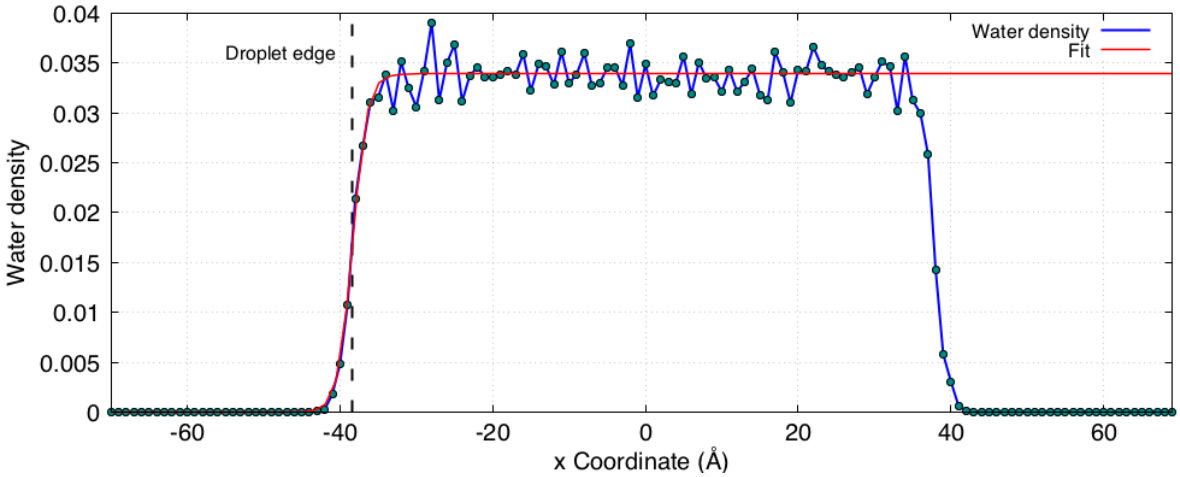
The trends in surface adsorption strength provide a feasible mechanism to explain the cell binding behaviour on PTFE observed in experiment. The bioactive region, which is capable of desorbing from the surface, can attract a region of negative charge on the cell surface receptor due to its high concentration of positive charges. Once this region is bound, cellular adhesion relies on the surface adsorption strength of the cofactors to stay anchored to the surface. The higher adsorption strength of the DiTryp, FMOC and DiPhe cofactors can support anchoring to the surface, but evidently once the adsorption strength of the cofactor region drops below some critical point, the peptide can no longer anchor the cell to the surface. This occurs for the Phe, Tryp and Acetyl variants.

## 4.4 Conclusion

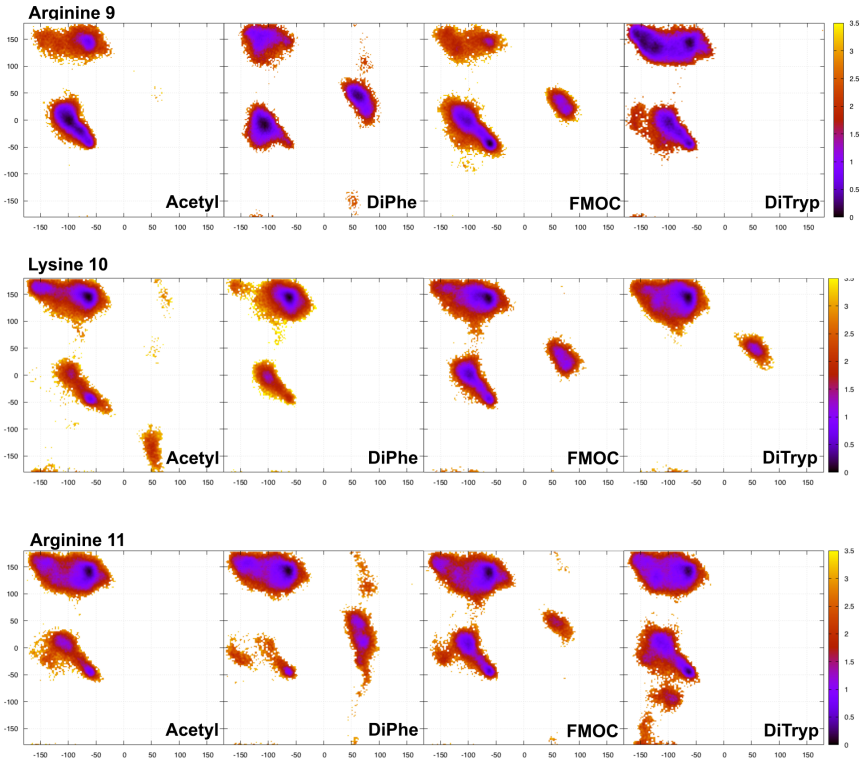
The ability to use fixed charge MD simulations of a hydrophobic surface to rationalise cell adhesion trends was far from a foregone conclusion. There are demonstrated difficulties in simulating even simple ion-surface adsorption, and the existence of specialised forcefields to simulate interfacial systems (covered in Chapter 3) suggests some researchers believe that nonpolarizable FFs can not be used in this case. There have been nonpolarizable simulations of an ideal case - diamond surfaces - that show good agreement with experiment, but this is not as complex as the biological system investigated here.

The calibration of a surface model for simulation showed that the nano-droplet method obeyed the physical relationship expected between surface energy and contact angle that is observed on the macroscale. This relationship was used to generate a model PTFE surface capable of simulating surface interactions of the domain 36 variants. Docking and equilibrium simulations of the cell active variants showed no common conformation that might explain their interaction with cells. The cofactors themselves, however, showed differences in their adsorption strength that correlate to the experimental cell adhesion data and thus offer an explanation for them. The active variants had much stronger surface adhesion, which suggests they can anchor the peptide to the surface when the cell is bound to the bioactive region. The strong adsorption of the hydrophobic cofactors follows from their larger volumes, but also the presence of polar moieties that point into a high density layer of water. Despite doubts about simulating interfacial systems using the fixed charge approximation, the PTFE model is able to explain the experimental cell adhesion results as well as offer insight into the molecular mechanism of surface adsorption. This suggests that the fixed charge approximation is appropriate to simulate inert, hydrophobic surfaces.

## 4.5 Supplementary figures




**Supplementary Figure 1.** Cross section of a water nanodroplet illustrating the cutoff used to determine the droplet's outer edge. Near the base, the droplet is approximately 80 Å wide. The density of water molecules (blue curve), calculated as the average of 500ps of simulation, rises from the vapor phase to maximal density in a sigmoid pattern. Fitting one half of the density to a sigmoid curve yields the fit in red, whose inflection point is less than 0.1 Å from the actual cut-off used Figure 1**.** This demonstrates how the half-maximal density is an appropriate choice for the edge of the nanodroplet, since it encompasses most of the water volume without over-extending the edge.



**Supplementary Figure 2**. Ramachandran plots of the bioactive region do not show any structural motifs common only to the binding variants. The plots are represented as density heat-maps showing the regions of highest sampling. Lysine 12 is omitted since it has no psi angle.

## 4.6 References


1 Ratner, B. D., Hoffman, A. S., Schoen, F. J. & Lemons, J. E. *Biomaterials science: an introduction to materials in medicine*. (Academic press, 2004).

2 Williams, D. F. On the nature of biomaterials. *Biomaterials* **30**, 5897-5909 (2009).

3 Ratner, B. D. & Bryant, S. J. Biomaterials: where we have been and where we are going. *Annu. Rev. Biomed. Eng.* **6**, 41-75 (2004).

4 Wilson, C. J., Clegg, R. E., Leavesley, D. I. & Pearcy, M. J. Mediation of biomaterial–cell interactions by adsorbed proteins: a review. *Tissue engineering* **11**, 1-18 (2005).

5 Roach, P., Eglin, D., Rohde, K. & Perry, C. C. Modern biomaterials: a review—bulk properties and implications of surface modifications. *Journal of Materials Science: Materials in Medicine* **18**, 1263-1277 (2007).

6 Hirsh, S. L. *et al.* The Vroman effect: competitive protein exchange with dynamic multilayer protein aggregates. *Colloids and Surfaces B: Biointerfaces* **103**, 395-404 (2013).

7 Hirsh, S. *et al.* A comparison of covalent immobilization and physical adsorption of a cellulase enzyme mixture. *Langmuir* **26**, 14380-14388 (2010).

8 Anderson, J. M., Rodriguez, A. & Chang, D. T. in *Seminars in immunology.* 86-100 (Elsevier).

9 Curtis, A., Forrester, J., McInnes, C. & Lawrie, F. Adhesion of cells to polystyrene surfaces. *The Journal of cell biology* **97**, 1500-1506 (1983).

10 Lee, J. H., Kim, H. G., Khang, G. S., Lee, H. B. & Jhon, M. S. Characterization of wettability gradient surfaces prepared by corona discharge treatment. *Journal of colloid and interface science* **151**, 563-570 (1992).

11 Bacakova, L., Filova, E., Parizek, M., Ruml, T. & Svorcik, V. Modulation of cell adhesion, proliferation and differentiation on materials designed for body implants. *Biotechnology advances* **29**, 739-767 (2011).

12 Tomizaki, K. y., Usui, K. & Mihara, H. Protein-Detecting Microarrays: Current Accomplishments and Requirements. *ChemBioChem* **6**, 782-799 (2005).

13 Schierholz, J. & Beuth, J. Implant infections: a haven for opportunistic bacteria. *Journal of Hospital Infection* **49**, 87-93 (2001).

14 Pesen, D. & Haviland, D. B. Modulation of cell adhesion complexes by surface protein patterns. *ACS applied materials & interfaces* **1**, 543-548 (2009).

15 Bilek, M. M. & McKenzie, D. R. Plasma modified surfaces for covalent immobilization of functional biomolecules in the absence of chemical linkers: towards better biosensors and a new generation of medical implants. *Biophysical reviews* **2**, 55-65 (2010).

16 Yin, Y. *et al.* Covalent immobilisation of tropoelastin on a plasma deposited interface for enhancement of endothelialisation on metal surfaces. *Biomaterials* **30**, 1675-1681 (2009).

17 Romer, L. H., Birukov, K. G. & Garcia, J. G. Focal adhesions: paradigm for a signaling nexus. *Circulation research* **98**, 606-616 (2006).

18 Hynes, R. O. Integrins: versatility, modulation, and signaling in cell adhesion. *Cell* **69**, 11-25 (1992).

19 Hubbell, J. A. Biomaterials in tissue engineering. *Nature Biotechnology* **13**, 565-576 (1995).

20 Lutolf, M. & Hubbell, J. Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering. *Nature biotechnology* **23**, 47-55 (2005).

21 Yamada, K. M. Adhesive recognition sequences. *J Biol Chem* **266**, 12809-12812 (1991).

22 Ruoslahti, E. RGD and other recognition sequences for integrins. *Annual review of cell and developmental biology* **12**, 697-715 (1996).

23 Hersel, U., Dahmen, C. & Kessler, H. RGD modified polymers: biomaterials for stimulated cell adhesion and beyond. *Biomaterials* **24**, 4385-4415 (2003).

24 Mithieux, S. M. & Weiss, A. S. Elastin. *Advances in protein chemistry* **70**, 437-461 (2005).

25 Wise, S. G. *et al.* Tropoelastin: a versatile, bioactive assembly module. *Acta biomaterialia* **10**, 1532-1541 (2014).

26 Szabó, Z. *et al.* Sequential loss of two neighboring exons of the tropoelastin gene during primate evolution. *Journal of molecular evolution* **49**, 664-671 (1999).

27 Tamburro, A. M., Bochicchio, B. & Pepe, A. Dissection of human tropoelastin: exon-by-exon chemical synthesis and related conformational studies. *Biochemistry* **42**, 13347-13362 (2003).

28 Brown, P. L., Mecham, L., Tisdale, C. & Mecham, R. P. The cysteine residues in the carboxy terminal domain of tropoelastin form an intrachain disulfide bond that stabilizes a loop structure and positively charged pocket. *Biochemical and biophysical research communications* **186**, 549-555 (1992).

29 Sevier, C. S. & Kaiser, C. A. Formation and transfer of disulphide bonds in living cells. *Nature reviews Molecular cell biology* **3**, 836-847 (2002).

30 Akhtar, K. *et al.* Oxidative modifications of the C-terminal domain of tropoelastin prevent cell binding. *Journal of Biological Chemistry* **286**, 13574-13582 (2011).

31 Broekelmann, T. J. *et al.* Tropoelastin interacts with cell-surface glycosaminoglycans via its COOH-terminal domain. *Journal of Biological Chemistry* **280**, 40939-40947 (2005).

32 Bax, D. V., Rodgers, U. R., Bilek, M. M. & Weiss, A. S. Cell adhesion to tropoelastin is mediated via the C-terminal GRKRK motif and integrin αVβ3. *Journal of biological chemistry* **284**, 28616-28623 (2009).

33 Rodgers, U. & Weiss, A. S. Integrin αVβ3 binds a unique non-RGD site near the C-terminus of human tropoelastin. *Biochimie* **86**, 173-178 (2004).

34 Vrhovski, B. & Weiss, A. S. Biochemistry of tropoelastin. *The FEBS Journal* **258**, 1-18 (1998).

35 Hofmeister, F. Zur lehre von der wirkung der salze. *Naunyn-Schmiedeberg's Archives of Pharmacology* **25**, 1-30 (1888).

36 Dang, L. X. Computational study of ion binding to the liquid interface of water. *The Journal of Physical Chemistry B* **106**, 10388-10394 (2002).

37 Wick, C. D. & Dang, L. X. Recent advances in understanding transfer ions across aqueous interfaces. *Chemical Physics Letters* **458**, 1-5 (2008).

38 Wick, C. D., Kuo, I.-F. W., Mundy, C. J. & Dang, L. X. The effect of polarizability for understanding the molecular structure of aqueous interfaces. *Journal of chemical theory and computation* **3**, 2002-2010 (2007).

39 Caleman, C., Hub, J. S., van Maaren, P. J. & van der Spoel, D. Atomistic simulation of ion solvation in water explains surface preference of halides. *Proceedings of the National Academy of Sciences* **108**, 6838-6842 (2011).

40 Koelsch, P. *et al.* Specific ion effects in physicochemical and biological systems: simulations, theory and experiments. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **303**, 110-136 (2007).

41 Heinz, H., Jha, K. C., Luettmer-Strathmann, J., Farmer, B. L. & Naik, R. R. Polarization at metal–biomolecular interfaces in solution. *Journal of The Royal Society Interface*, rsif20100318 (2010).

42 Horinek, D. *et al.* Peptide adsorption on a hydrophobic surface results from an interplay of solvation, surface, and intrapeptide forces. *Proceedings of the National Academy of Sciences* **105**, 2842-2847 (2008).

43 Heinz, H. & Ramezani-Dakhel, H. Simulations of inorganic–bioorganic interfaces to discover new materials: insights, comparisons to experiment, challenges, and opportunities. *Chemical Society Reviews* **45**, 412-448 (2016).

44 Horinek, D. & Netz, R. R. Specific ion adsorption at hydrophobic solid surfaces. *Physical Review Letters* **99**, 226104 (2007).

45 Penna, M., Mijajlovic, M., Tamerler, C. & Biggs, M. J. Molecular-level understanding of the adsorption mechanism of a graphite-binding peptide at the water/graphite interface. *Soft matter* **11**, 5192-5203 (2015).

46 Serr, A., Horinek, D. & Netz, R. R. Polypeptide friction and adhesion on hydrophobic and hydrophilic surfaces: a molecular dynamics case study. *Journal of the American Chemical Society* **130**, 12408-12413 (2008).

47 Sendner, C., Horinek, D., Bocquet, L. & Netz, R. R. Interfacial water at hydrophobic and hydrophilic surfaces: Slip, viscosity, and diffusion. *Langmuir* **25**, 10768-10781 (2009).

48 Dai, Z.-W., Ling, J., Huang, X.-J., Wan, L.-S. & Xu, Z.-K. Molecular simulation on the interactions of water with polypropylene surfaces. *The Journal of Physical Chemistry C* **115**, 10702-10708 (2011).

49 Wang, J., Bratko, D. & Luzar, A. Probing surface tension additivity on chemically heterogeneous surfaces by a molecular approach. *Proceedings of the National Academy of Sciences* **108**, 6374-6379 (2011).

50 Gohlke, H. & Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie International Edition* **41**, 2644-2676 (2002).

51 Pantsar, T. & Poso, A. Binding Affinity via Docking: Fact and Fiction. *Molecules* **23**, 1899 (2018).

52 Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **31**, 455-461 (2010).

53 Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry* **30**, 2785-2791 (2009).

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

55 van Oss, C. J. Use of the combined Lifshitz–van der Waals and Lewis acid–base approaches in determining the apolar and polar contributions to surface and interfacial tensions and free energies. *Journal of Adhesion Science and Technology* **16**, 669-677 (2002).

56 Werder, T., Walther, J., Jaffe, R., Halicioglu, T. & Koumoutsakos, P. On the water− carbon interaction for use in molecular dynamics simulations of graphite and carbon nanotubes. *The Journal of Physical Chemistry B* **107**, 1345-1352 (2003).

57 Rafiee, J. *et al.* Wetting transparency of graphene. *Nature materials* **11**, 217 (2012).

58 Wu, Y. & Aluru, N. Graphitic carbon–water nonbonded interaction parameters. *The Journal of Physical Chemistry B* **117**, 8802-8813 (2013).

59 Riano, L. *Least Square Circle*, <https://gist.github.com/lorenzoriano/6799568 - file-least_square_circle-py> (2016).

60 Jarzynski, C. Nonequilibrium equality for free energy differences. *Physical Review Letters* **78**, 2690 (1997).

61 Grossfield, A. *WHAM: the weighted histogram analysis method*, <http://membrane.urmc.rochester.edu/content/wham> (2016).

62 Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry* **13**, 1011-1021 (1992).

63 Halgren, T. A. The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *Journal of the American Chemical Society* **114**, 7827-7843 (1992).

64 Rowlinson, J. S. & Widom, B. *Molecular theory of capillarity*. (Courier Corporation, 2013).

65 Fox, H. & Zisman, W. The spreading of liquids on low energy surfaces. I. polytetrafluoroethylene. *Journal of Colloid Science* **5**, 514-531 (1950).

66 Du, X. *et al.* Insights into protein–ligand interactions: Mechanisms, models, and methods. *International journal of molecular sciences* **17**, 144 (2016).

67 Li, X., Latour, R. A. & Stuart, S. J. TIGER2: an improved algorithm for temperature intervals with global exchange of replicas. *The Journal of chemical physics* **130**, 174106 (2009).

68 Wright, L. B. & Walsh, T. R. Efficient conformational sampling of peptides adsorbed onto inorganic surfaces: insights from a quartz binding peptide. *Physical Chemistry Chemical Physics* **15**, 4715-4726 (2013).

69 Jensen, T. R. *et al.* Water in contact with extended hydrophobic surfaces: Direct evidence of weak dewetting. *Physical review letters* **90**, 086101 (2003).

70 Baştuğ, T. & Kuyucak, S. Application of Jarzynski's equality in simple versus complex systems. *Chemical Physics Letters* **436**, 383-387 (2007).

71 Jamadagni, S. N., Godawat, R. & Garde, S. How surface wettability affects the binding, folding, and dynamics of hydrophobic polymers at interfaces. *Langmuir* **25**, 13092-13099 (2009).

# Chapter 5. Simulations of bioactive peptides in solution

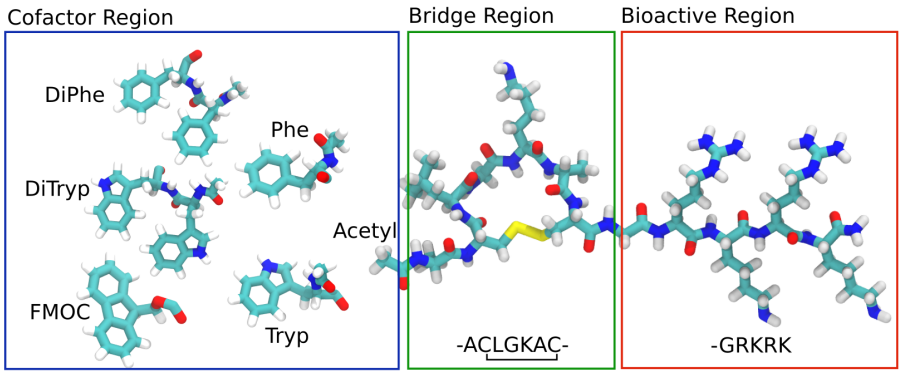
The acetylated domain 36 peptide cannot adhere cells to PTFE surfaces due to the weak adsorption strength of the acetyl group to the surface. The addition of hydrophobic N-terminal modifications, termed cofactors, allows for stronger adsorption at the peptide-PTFE interface. To show that the peptides bind specifically to the tropoelastin receptor, they have previously been experimentally tested for the ability to inhibit cell adhesion on surfaces coated with tropoelastin. In these experiments some variants of the peptide, in solution, bound to and blocked the available tropoelastin receptor binding sites. One of the peptide variants showed different binding activity in solution than on the surface. This chapter interrogates the four domain 36 variants in solution, using simulation techniques to offer an explanation of the experimental cellular adhesion data.

## 5.1 Introduction

Domain 36 of tropoelastin can bind to its receptor on the cell surface when attached to the tropoelastin molecule and also, as discussed in chapter 2, from a PTFE surface when strongly adsorbed. The binding site of domain 36 on the cell surface is not well characterized. The molecular structure of the site is unknown and hence the mechanism of binding is unclear. Determining the molecular details of binding will inform the design of surfaces that are better able to adhere cells and will shed light on the binding of endogenous tropoelastin.

To further characterise the domain 36 peptides, Dr Giselle Yeo from the Charles Perkins Centre at the University of Sydney has tested the capacity of the four domain 36 variants introduced in Chapter 4 to bind to the tropoelastin receptor directly from solution using inhibition assays (see Appendix A). The variants were added to a solution of cells on a surface coated with a monolayer of tropoelastin molecules and then tested for their ability to inhibit adhesion by binding to, and thus blocking, the available receptor sites. Unlike the same molecules on the PTFE surface, not all of the hydrophobic variants could bind to the cell (Figure 1). While the DiTryp and FMOC variants were capable of binding, both the Acetylated and Diphenylalanine variants did not inhibit adhesion and are thus incapable of binding to the cell surface receptor when in solution.

a




b

| Variant | Cell adhesion on PTFE | Cell binding in solution |
|---|---|---|
| Acetyl-$A_1$CLGKACGRKRK$_{12}$ | Inactive | Inactive |
| DiPhe-$A_1$CLGKACGRKRK$_{12}$ | Active | Inactive |
| DiTryp-$A_1$CLGKACGRKRK$_{12}$ | Active | Active |
| FMOC-$A_1$CLGKACGRKRK$_{12}$ | Active | Active |

**Figure 1.** The domain 36 variants used in this chapter. **a)** The peptides have three structural sections—the cofactor, bridge, and bioactive regions. The cofactor can be: acetyled alanine (Acetyl), two phenylalanines (DiPhe), two tryptophans (DiTryp) or the aromatic FMOC moiety. The bridge region consists of six residues connected by a disulfide bridge. The bioactive region is the region where the peptides are likely to bind to the tropoelastin receptor. It is connected to the bridge by a flexible glycine and is highly positively charged. **b)** The cell activities of the four variants, measuring the ability to bind the tropoelastin cell surface receptor. The acetyl variant is always inactive, while the FMOC and DiTryp variants are always active. The DiPhe variant is active on a PTFE surface, but not in solution.

These results suggest that: 1) the acetylated domain 36 peptide loses something essential to binding, possibly a hydrophobic support, by being separated from the tropoelastin molecule; 2) two of the hydrophobic cofactors (FMOC, DiTryp) are able to recapitulate this function; and 3) the Diphenylalanine cofactor, while it is also hydrophobic, is incapable of providing this function. Can we more closely examine the mechanism of binding without knowing the makeup of the binding site? Insight into the effect of the cofactor on the peptide structure and its dynamics will shed light on this question, and molecular simulations may be able to generate the high-resolution information required.

Atomic-resolution structural studies of biomolecules generally proceed using Nuclear magnetic resonance (NMR) spectroscopy or X-ray crystallography, but there are some molecules which do not

yield a coherent structure using either of these techniques. Tropoelastin is known to be highly flexible and mostly disordered[1], hampering structural studies that use these techniques. Because of this, there is no available crystal structure of tropoelastin. The domain 36 peptide has been studied by NMR and simulation, indicating that it has some beta turn content in the bridge region but is mostly unfolded[2]. Given its small size and high flexibility, domain 36 is likely a member of a class of molecules known as intrinsically disordered peptides (IDPs).

IDPs have two defining features: 1) They are typically shorter than structured proteins, so the total intramolecular force that can support folding is weak compared to their thermal energy. Longer polypeptides have more opportunity for enthalpic contacts between residues and a greater entropic benefit to the expulsion of water from an internal globular phase. These processes result in the folded state existing in a free energy well. As they are removed (often simply by reducing the size of the chain), the barriers between free energy wells are reduced, ultimately becoming crossable by thermal energy alone. This makes folded conformations transient. Energetically, they can be described by a flat free energy landscape with low barriers between structurally discrete populations (Figure 2)[3]. 2) Many conformations that IDPs do visit lack secondary structure content. That is, they have a low percentage of well-ordered structural motifs like alpha helices and beta turns[4], and are mostly 'random coil' - a term from polymer science that refers to randomly distributed orientations between adjacent monomers or, in this case, amino acids. For peptides, the free energy landscape of rotation around the backbone does have some minima that weight the distribution towards certain structures, but these are not recognised as folded due to the lack of secondary structure content. IDPs such as the domain 36 variants are thus structurally heterogeneous - they have no native structure and exist as an ensemble of transient populations that are mostly random coil. Because of this, they are best understood as a statistical distribution of structures rather than a single conformation.

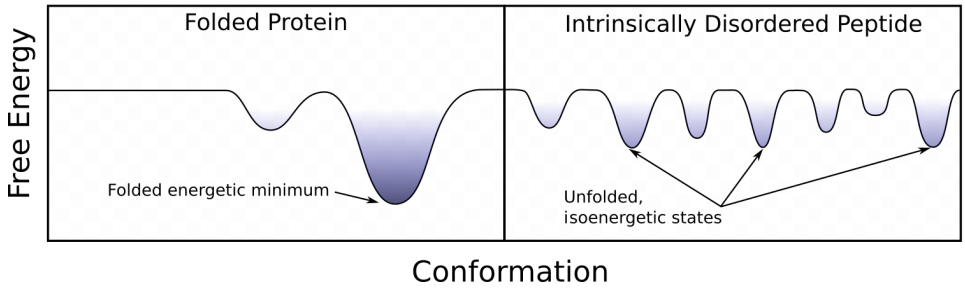


**Figure 2.** Schematic diagram of the free energy landscapes of a folded protein and an intrinsically disordered peptide (IDP). Energy wells indicate a preferred conformation – folded proteins usually have a single, global minimum, while IDPs have many minima of the same or similar energy that are physiologically relevant.

IDPs are disordered in order to access alternative binding modes to folded proteins[5]. According to the 'lock-and-key' binding mechanism, natively folded protein ligands bind to their substrates by fitting into a pocket of complementary shape[6]. This constricts their motion, which thermodynamically is described as a loss of translational and rotational entropy. The loss in entropy is balanced by an increase in enthalpic energy via ionic bonds, van der Waals interactions, or the hydrophobic effect (favourable entropy and enthalpy in the surrounding water molecules). For IDPs, whose structural disorder gives them comparatively high entropy, the entropic cost of binding is higher. They compensate for this with increased enthalpic benefit of binding via hydrogen bonding to the exposed backbone and/or by a stepwise mechanism of binding, using charged hotspot residues that minimize entropic cost while gradually increasing enthalpic benefit[7]. These mechanisms are illustrative of two further kinds of binding modes – 'conformational selection' and 'induced fit'.

The high flexibility of IDPs gives rise to a large number of discrete structures, leading to these two types of binding. In conformational selection, the ensemble of populations explored by an IDP in the solution phase includes many structures that are not suitable for binding, although some of the time, the peptide visits a subpopulation that is compatible with the binding site. If the ensemble of populations is weighted enough towards this structure then, with enough interaction events between the IDP and its receptor, binding occurs. This is essentially the lock-and-key mechanism but relies on the interaction occurring while the IDP is in the correct subpopulation.

In the other type of binding of IDPs, induced fit, the peptide never visits the binding conformation on its own. Initially, charge interactions attract it towards the binding pocket and, facilitated by high flexibility, it fits into the pocket to achieve shape complementarity. For purely induced fit binding, the peptide can visit the bound conformation only by interaction with the receptor. In practice, most IDP binding is likely to be some combination of the two methods, i.e. a subpopulation of the ensemble is more likely to bind since it has lower entropy and reduced entropic cost to binding. This allows it to

flexibly fit into the binding site. The challenge in optimizing binding is first to understand how much each of these contribute to the interaction.

Often, since NMR and X-ray crystallography are inapplicable for IDPs in the solution-phase, atomistic simulations are required to determine the mechanism of binding. Where there is a crystal structure of the *bound* IDP/receptor dimer, simulations can essentially proceed in reverse, moving from the bound state to the solution phase, and the degree of structural change is revealed. This is beneficial computationally, since it avoids sampling large regions of inactive conformations in the aqueous phase unnecessarily. If the bound structure also exists as a part of the solution-phase ensemble, then conformational selection is the likely binding mechanism. Modifications that stabilize this structure are expected to increase binding. If the bound structure is not part of the solution-phase structural landscape, then induced fit is more likely. In this case, binding is enhanced by modifications that slow the dynamics of the peptide, reducing the entropic cost of fitting the binding site.

Determining the binding mechanism using simulation is considered difficult or impossible without experimental data such as a crystal structure[8]. In the case of domain 36, where there is no knowledge of the binding structure and all structures are equally likely to be physiologically relevant, another external source of screening data is required. Cell adhesion data like those shown in Figure 1 can be used for this purpose. These data indicate that chemical modifications change either the dynamics or the structural ensemble of the domain 36 peptide. Comparing the structural ensembles of the variants in binding and inactive situations reveals the extent of conformational selection, an approach that has been taken with beta amyloid peptides[9].

Comparing conformational ensembles first requires a Boltzmann-weighted ensemble of structures which are then clustered into discrete populations. Finding an effective clustering method is complex for IDPs since they, unlike native proteins that are easily clustered by tertiary and secondary structure, have low secondary structure content. Folded proteins can be classified by collective variables such as the number of hydrogen bonds or the radius of gyration, but how can structures of solvated, random coil IDPs be differentiated from one another? The peptide backbone dihedral angles can describe different structures, yet even the shortest peptides have tens or hundreds of backbone

angles. In these circumstances, it is necessary to reduce the number of dimensions of movement to a manageable number so they can be plotted graphically and clustered.

Dimensionality reduction is the procedure of collapsing many degrees of freedom, or dimensions, to a low number that allows for analysis. Movement across the low dimensional space represents changes in multiple parameters from the high-dimensional space. A handful of dimensions in the reduced space represents the total variation in the data. One example of dimensionality reduction applied to binding mechanisms of IDPs is the 'sketch-map', a 2-dimensional projection of points that are clustered together to faithfully represent cluster patterns from high-dimensional space[10].

Another method of dimensionality reduction is principal component analysis (PCA). Theoretically, this procedure applies a linear transformation to simulation trajectory data to distil all of the observed variation into several modes that each describe multiple underlying parameters. A PCA classifies all of the types of structural changes observed in the data and orders them according to magnitude. This technique allows high-dimensional structure changes consisting of many elements to be described using only a handful of dimensions. PCA can be applied to Cartesian coordinates of atoms in a simulation, although any translational and rotational diffusion must first be removed to prevent these movements being interpreted as structural change. For peptides, which have nontrivial alignments, internal coordinates such as backbone dihedral angles are more suitable[11]. These internal coordinates summarize whole-peptide movement and have successfully been used in a dihedral PCA (dPCA) to separate IDP conformations into low-dimensional space[12,13].

Sketch-maps and PCA are both capable of low-dimensional representations of movement in high-dimensional space, but they are mathematical descriptions and so, without interpretation, lack the intuitive description given by collective variables. Additionally, while sketch-mapping is capable of unravelling nonlinear data topologies, PCA is not[10]. Only a PCA can be interpreted in terms of the underlying structural change. For example, movement along one principal component from a dPCA is equivalent to a linear combination of movements along the underlying dihedrals, but this is not true for sketch-maps. This means dPCAs can be interpreted in terms of the structural changes of collective

variables. For this reason, dPCA is the logical first choice for a rigorous understanding of peptide structural properties.

The tropoelastin-derived domain 36 peptide is likely an IDP and, as such, might be best studied by dPCA. Experimental analysis of this peptide by circular dichroism (which probes secondary structure), NMR and modelling have all indicated that it has very little secondary structure, no prolonged conformations, and that there are several isoenergetic states[2]. Additionally, it relies heavily on charge interactions to bind its receptor[14]. These are the hallmarks of intrinsic disorder, and so the modification by cofactors likely either changes the conformation (implying conformational selection) or the dynamics (implying induced fit) of the peptide. This can be determined by molecular dynamics simulation. In comparison to Chapter 4, where the peptides were constrained by being adsorbed to the surface, here equilibrium simulations can generate a more complete picture of the conformational landscape. This is useful for determining conformational selection or induced fit binding mechanisms. Using dPCA, both the conformational landscape and kinetic properties of the four peptides are compared. This is achieved by brute force equilibrium simulations, which preserve the residence time of each conformation. Additionally, this is achieved without the benefit of a known binding conformation or even a known binding partner. Rigorous statistical interrogation of the conformational ensembles is combined with more understandable collective variable analysis to give a comprehensible picture of the mechanism of binding of domain 36 of tropoelastin.

## 5.2 Methods

### 5.2.1 Equilibrium Simulation

To determine the differences in conformation and dynamics of the four peptides, they were modelled using molecular dynamics equilibrium simulations. The simulations were carried out in water boxes that extended in each direction by the length of the peptide plus two lengths of the non-bonded interaction cut-off distance (1.2 nm). This ensured that even at full extension and alignment along an axis, the peptide could at no stage interact with a periodic copy of itself. The peptides were simulated for 3 μs each, totalling 12 μs of data.

### 5.2.2 Analysis

The coordinates of all atoms in the simulations were saved every 20 picoseconds, which resulted in 600,000 trajectory frames. All of the analyses were performed by looping through these frames and passing the data to various scripts. Secondary structure analysis of the peptide backbones used the STRIDE program[15] based on the original STRIDE secondary structure assignment algorithm, which classifies structures according to specific combinations of backbone angles[4]. For the dihedral angle analysis, backbone angles from the trajectories were output by scripting in VMD[16], which uses the TCL programming language. Calculating the mean and variance of the dihedral angles used Python scripting of circular statistics, which accounts for the errors introduced by periodicity in angular measurements (see Appendix C).

The backbone dihedral angles were also analysed by dPCA in order to reduce the number of dimensions and simplify the analysis of the equilibrium simulations. The backbone dihedral angles from each frame of the trajectory were first transformed into a dataset containing the *cos* and *sin* of each $\varphi$ and $\psi$ angle. This ensures that values around -180° or 180° do not jump across the boundary and vary by a full 360°. In the past, dPCA has been performed on single peptides only. By concatenating the dihedral data from each peptide, here all of the four peptides can be directly compared within a single dPCA analysis. The full dataset consisted of an $N \times M$ matrix where $N = 44$, corresponding to the *sin* and *cos* of each of the 22 backbone dihedral angles for $A_1$ – $K_{12}$, and $M = 600{,}000$, which is the total number of trajectory frames used. The number of backbone angles for the twelve amino acids of domain 36 is 22 since the first and last amino acids have only one peptidic bond, meaning they have only one peptide dihedral. The dihedrals from the DiPhe and DiTryp cofactors were not included since they had no equivalent in the Acetyl and FMOC variants.

The dPCA was performed using the scikit-learn library for Python, which is a set of tools for machine learning and data analysis[17]. There are three outputs of this process. The first is a list of the eigenvalues in descending order. These are referred to as the 'Principal Components' (PC) in the results. The magnitude of these represents the fraction of total variation in the dataset that is caused by that eigenvalue, or PC. Because of this, the sum of all eigenvalues is one. The second output is the set of eigenvectors – one for each of the eigenvalues. The eigenvectors are sets of numbers associated with

the variables (the dihedral angles) that explain the extent of change at each variable within the PC. As an example, two peptide structures can be described by their value along the PC1 dimension. When they have different values along PC1 then their dihedral angles are also different, which is controlled by the first eigenvector. The third output is a matrix with the same dimensions as the input data. Each row of this matrix is made up of the values of PCs 1 to 44 that completely describe the underlying structure of that trajectory frame.

These data have been simplified for representation. The eigenvectors have length 44, but refer to only 22 variables. To reduce the size of the eigenvectors, the absolute value of the elements corresponding to the *sin* and *cos* of each dihedral are summed, giving a single value for each variable (i.e. 22 in total).

### 5.2.3 Sequence Alignment

The sequences of domain 36 from a range of species were downloaded from the UniProt online server[18]. These sequences were aligned using the T-Coffee online server[19].

## 5.3 Results and Discussion

### 5.3.1 Secondary structure of peptide variants

The ability of the domain 36 variants to bind to cells depends on the type of N-terminal cofactor present. To determine how the cofactor affects the structure and dynamics of the peptides, which in turn control the degree of conformational selection or induced fit, the conformation and flexibility of the peptides is measured. Extended solution-phase equilibrium simulations of the four peptides were performed to determine the conformational and kinetic differences between the active and inactive variants. Because there is no structure of domain 36 in the bound state, the structures of the peptide variants cannot be easily scored based on their similarity to this state. Additionally, there is no structure of the domain 36 receptor, meaning docking cannot be performed. So, to determine the molecular determinant of binding, increasingly stringent analysis is applied to the simulation trajectories to find the conformational or dynamic properties that are unique to the cell active variants.

The secondary structure of the variants was calculated from the equilibrium simulations to analyse the conformation. This analysis classifies the peptide backbone phi and psi angles into common structural motifs. All four variants show the hallmarks of intrinsic disorder, for example fast dynamics compared to peptide folding[20], and few secondary structure elements with a high percentage of random coil structure (Figure 3a). The main secondary structural element observed is a beta turn occurring within the bridge region. This is caused by torsion introduced by the disulfide bond that joins the two cysteine residues containing the bridge region. A second beta turn occurs in the active region of all variants. This region, being highly charged and responsible for binding the receptor, is a possible source of unique conformations that lead to binding. However, since this turn is equally probable in all of the variants, it cannot be used to differentiate them. There is little other secondary structure in the active region, so the structure here is mostly defined as 'random coil'. These results agree with experimental measures of domain 36 structure, which found no alpha helix content, some beta turn content, with the peptide mostly unfolded[2].

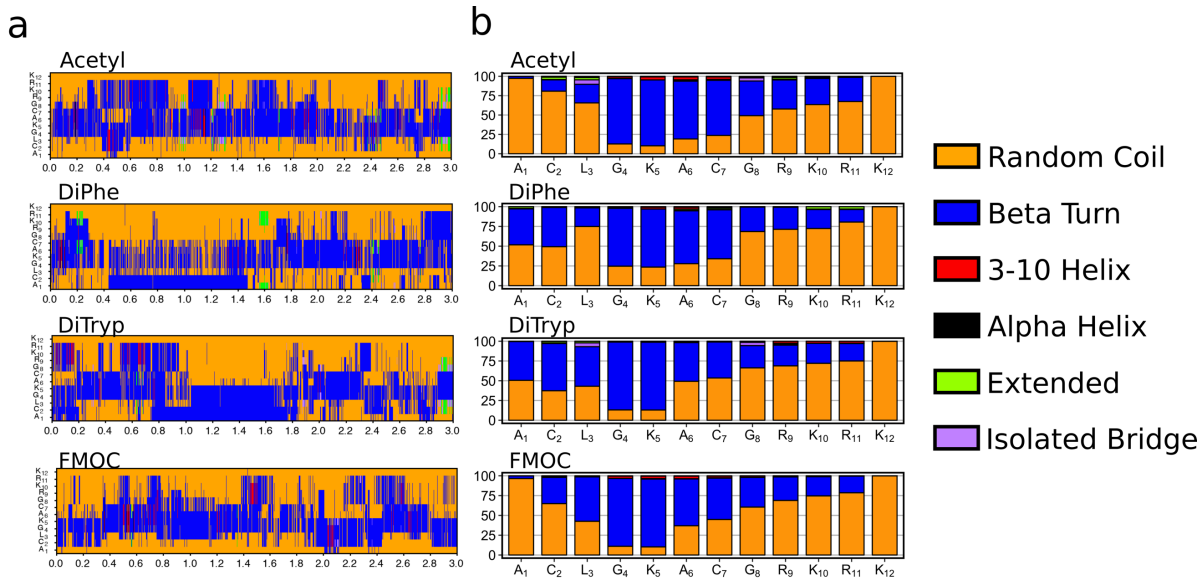



**Figure 3.** Secondary structure of the four peptide variants shows mostly beta turn and random coil. **a)** Time series of the peptide trajectories coloured by secondary structure elements. **b**) Proportion of the trajectories spent in each secondary structure type. The bridge region is mostly beta turn, while the rest of the peptide is random coil. The active peptides (FMOC and DiTryp) show slightly higher beta turn content at residue $L_3$.

The beta turn in the bridge region is shifted towards the N-terminus in the three variants with hydrophobic cofactors (DiPhe, DiTryp, and FMOC). It occurs across residues $G_4LAC_7$ for the Acetyl variant, but through $A_1CLG_4$ for the DiPhe and DiTryp variants, and also $C_2LGK_5$ for the FMOC and DiTryp variants. The shift in the beta-turn might be attributed to the interaction between the hydrophobic cofactor and the rest of the peptide, causing further torsion in the bridge region. This behaviour is analogous to protein folding in globular proteins caused by the hydrophobic effect – the hydrophobic regions have unfavourable enthalpic and entropic interactions with water and so are more likely to associate with other hydrophobic regions because doing so reduces contact with the aqueous phase. Because of this similarity to globular proteins, this behaviour will be called 'hydrophobic folding', although the peptide is still intrinsically disordered and does not undergo complete folding like a globular protein.

The beta-turn through $C_2LGK_5$ occurs mostly in the FMOC and DiTryp variants, which are the two bioactive variants, and since this also involves a charged residue ($K_5$) it could potentially alter the charge distribution across these variants to a position more suitable for cell binding. This turn leads to a slightly higher percentage of beta-turn content for the $L_3$ residue than in the inactive variants, and so becomes, this far in the analysis, the only defining feature of the bioactive variants (Figure 3b).

There are limits to secondary structure analysis, particularly for IDPs. Since IDPs are unfolded, these peptides lack secondary structure. Furthermore, this type of analysis groups the whole range of peptide backbone rotation, which varies over 360°, into 6 kinds of structure. A more fine-grained approach may refine the differences hidden within the 'random coil' and 'beta turn' classifications.

### 5.3.2 Dihedral angle analysis of the peptide variants

The mean and the variance, which is measured here as the *R* parameter (see Appendix C), of the backbone angles can be used to analyse both the structure and dynamics of the peptides. The means of the backbone angles represent the average structure over the course of the trajectories, which would be altered by changes in the population ensemble. Similarly, the *R* parameter represents the average flexibility of the peptides, which is an indirect measure of entropy. These values also show more subtle differences between the variants than the secondary structure time series do (Figure 4) because they

are more fine-grained. Near the N-terminus of the peptide, hydrophobic folding can be seen in the change of the mean and *R* of the $C_2$-phi value, which is shifted just for the three hydrophobic cofactor variants. For the Acetyl variant, which has little torsion introduced by the N-terminal alanine residue, this angle samples the most energetically favourable states almost exclusively. When a hydrophobic cofactor is present, hydrophobic folding gives rise to a second population characterised by a change in this angle. This change aligns with the increased beta turn at this residue for the hydrophobic cofactor variants (Figure 3b).

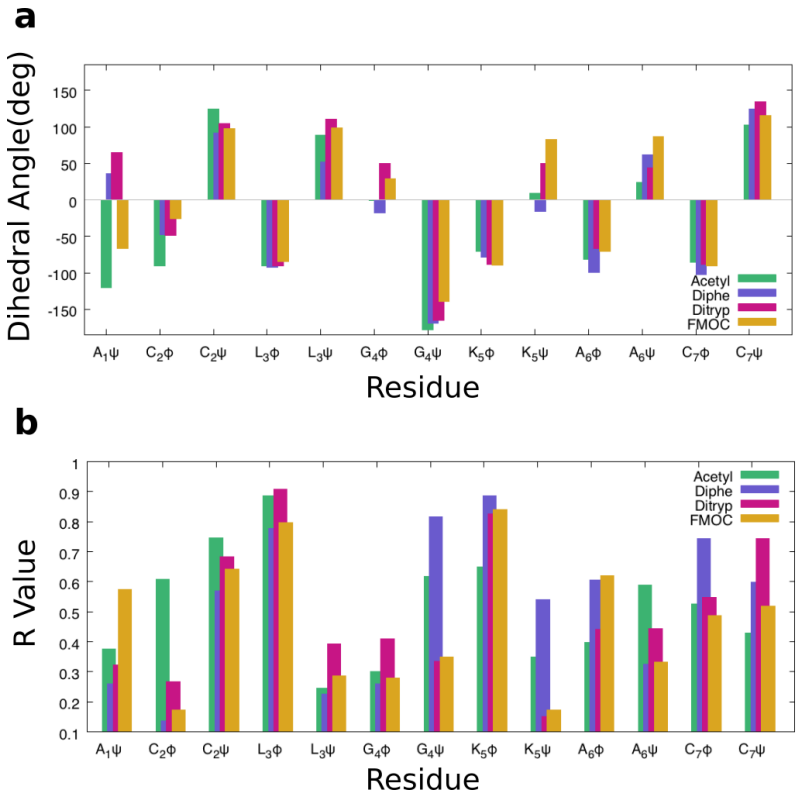



**Figure 4.** Means and variances of the phi and psi angles of the peptide backbone from residue A1 to C7, showing more detail than secondary structure analysis. **a)** Means of the phi and psi backbone angles. **b)** Variance of the phi and psi backbone angles, reported as *R* (see Appendix C).

In the bridge region there are also indications of a separate population that exists only for the bioactive variants (i.e. FMOC and DiTryp). The reduced variance of the $G_4$-psi angle and the variance and mean of the $K_5$-psi angle are unique to the FMOC and DiTryp variants. Given that the beta turn in the bridge region, which was identified by secondary structure analysis (Figure 3), is also unique to these peptide variants, this region could be the origin of structural change that differentiates the cell binding variants

from the non-binding variants. In the active region, the backbone dihedral angles show little difference between the variants (Supplementary Figure 1), so the average structure of this region is not discriminatory for the binding variants. There are, however, some differences in this region in *R*, which fluctuates but with no preference for any of the cofactors. Since it does not group the cell binding and inactive variants separately, it is not discriminatory.

Analysis of the mean and *R* of the backbone dihedral angles gives a higher resolution picture of the structure of the peptide than the secondary structure analysis in Figure 3. However, while the secondary structure time series showed fine detail in time, these two values are averages, i.e. a single value representing the entire simulation. As a result, this analysis can miss transitory, and perhaps physically relevant, structures. These transitory structures are described fully by linear combinations of all of the backbone dihedral angles at once. To resolve these structures properly, a more fine-grained, method that offers high resolution both in space and in time is required.

### 5.3.3 dPCA: Conformational selection

The ideal measure for analysing all of the conformations must represent all dihedral angles from all structures in the ensemble, and preferably also be in a low dimensional space to facilitate analysis. A principal component analysis of dihedral angles using dPCA achieves all of these goals – every single structure from the ensemble can be described by several principal components, each of which in turn describes several associated dihedrals. This analysis also separates movement between important conformations from the ubiquitous fluctuations associated with temperature. The first several principal components describe the largest, and so usually the slowest and most biologically important, structural changes[21], while the rest reflect thermal vibrations. A dPCA of the concatenated trajectories of the four peptide variants allows direct comparisons between all of the structures visited by each variant within a single set of principal components. This helps to determine unique populations that might contribute to conformational selection binding. This analysis also allows simple comparison of the frequency of those movements (i.e. the peptide dynamics) to determine the importance of induced fit.

dPCA of the peptides confirms that they are intrinsically disordered – the eigenvalues, which measure the percentage of variation contributed by each principal component, are flat with no single principal

component (PC) contributing more than 13% (Supplementary Figure 2). This indicates there is no single, dominant structural change that describes the peptides as might be the case for movement in globular proteins[22]. Many types of movement occur with more or less equal contribution to the total structure, so the peptides are structurally heterogeneous. This also means that to describe a significant portion of the structural variation requires a large number of principal components – possibly defeating the purpose of dimensionality reduction by dPCA in the first place. However, given that the aim of generating these data was not an exhaustive description of the peptide structural phase space – it was already known from NMR measurements to be heterogeneous with likely only a few biologically relevant structures – most PCs can be discarded if they a) do not have a separate population of both the cell binding variants, DiTryp and FMOC, or b) describe biologically irrelevant movements such as ubiquitous thermal vibrations.

Discarding some PCs using these criteria leaves only the first three, accordingly with the three largest eigenvalues. The type of structural movement described by each PC is shown in Figure 5a, alongside histograms showing the distribution of each PC that each variant visits (Figure 5b), as well as bar graphs of the individual dihedrals whose movement contributes to that PC (i.e. the eigenvectors, Figure 5c).

PC 1 describes movement of the hydrophobic cofactor outwards or inwards with respect to the peptide, causing changes mainly in the $A_1C_2$ dihedral angles. This PC is consistent with the hydrophobic folding behaviour observed in both the secondary structural (Figure 3) and dihedral angle analyses (Figure 4). As noted previously, this behaviour does not occur for the Acetyl variant, so it is likely caused by the hydrophobic cofactor. This can be seen clearly in the histogram of PC 1, where PC 1 > 0.0 corresponds to hydrophobic folding and is populated almost exclusively by the hydrophobic DiPhe, DiTryp, and FMOC variants.

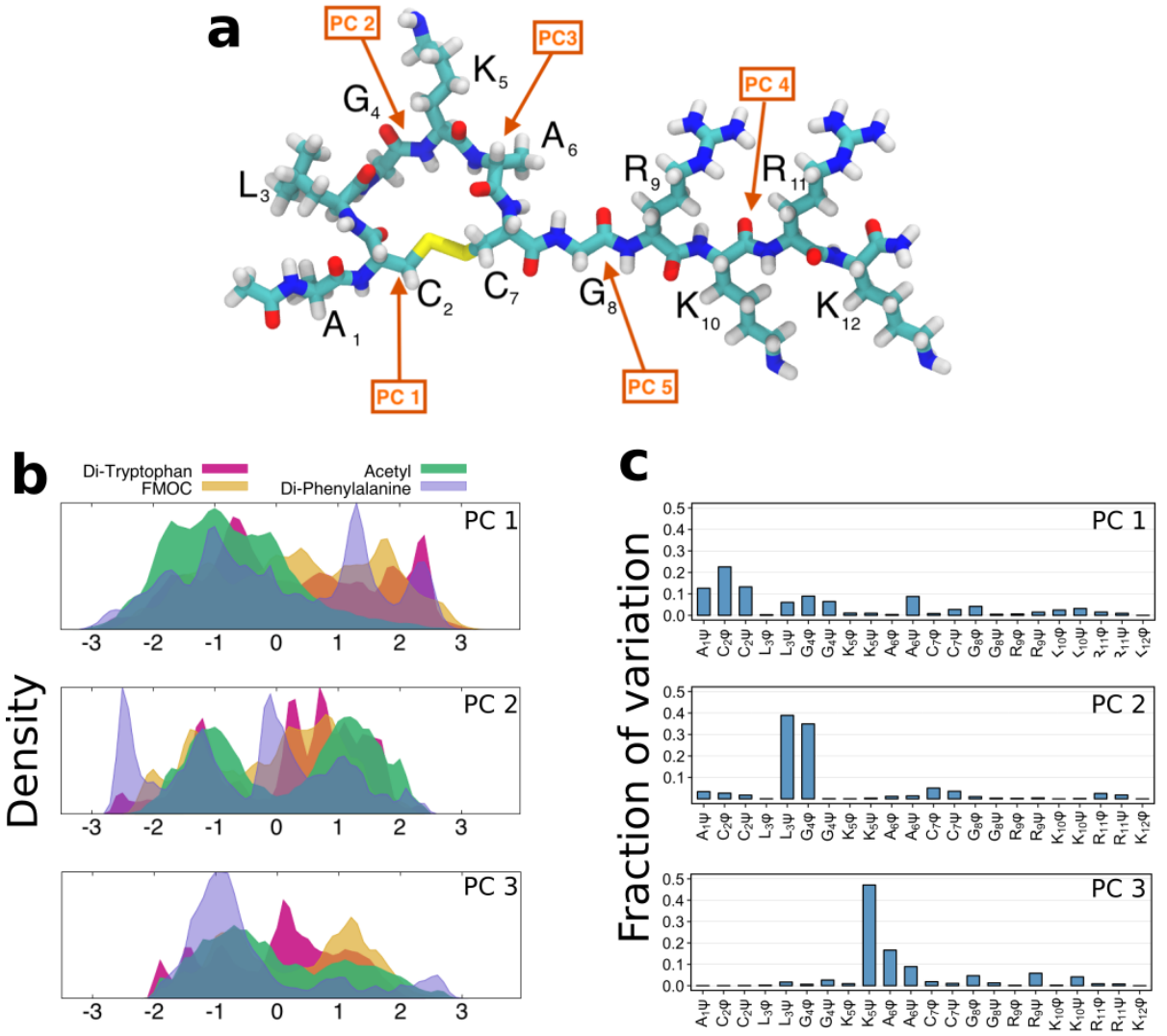



**Figure 5.** The first three principal components describe the relevant structural changes in the peptide variants. **a)** Structure of the acetyl domain 36 variant showing the point in the backbone where movement occurs for each principal component. **b)** Histograms of the first three PCs showing differences in the proportions of structures visited by each variant. **c)** The eigenvectors associated with the first three PCs showing the contribution of each backbone dihedral angle.

PCs 2 and 3 both describe movement within the bridge region and so can be considered together, although the structural movements they describe can occur independently. These PCs describe rotation at two points within the bridge region. Since this region forms a loop that is constrained by the disulfide bond, rotation at these points manifests as torsion of the bridge, shifting it from a roughly circular shape. PC 2 represents a twisting motion at the $L_3G_4$ peptide bond, whereas PC 3 represents twisting at $K_5A_6$. All four variants fully explore the range of these PCs (Figure 5b), but the DiTryp and FMOC variants

have higher representation at 0 < PC 2 < 1 and 0 < PC 3 < 2. Notably, the movements associated with these PCs frames residue $K_5$, which carries a charge. It follows that different distributions across these PCs can rearrange the position of this charge relative to the other charges in the active region.

A PCA orders the principal components by the size of variance observed.

To combine the PCs together, all of the structures across the four trajectories can be filtered based on rules that remove structures also visited by the inactive variants. The only remaining conformations are those that are visited almost exclusively by the cell-binding FMOC and DiTryp variants. The following rubric is designed to select for DiTryp and FMOC variant structures that are less likely to occur for the DiPhe and Acetyl variants using the discriminatory PCs identified above:

1) PC 1 > 0
2) 0 < PC 2 < 1
3) 0 < PC 3 < 2

The number of trajectory frames captured by this filtration is indicative of how often, across the whole ensemble, these structural motifs combine. Figure 6 shows that the DiTryp and FMOC variants spend a greater amount of time in this structure compared to the DiPhe and Acetyl variants. Thus, hydrophobic folding (PC 1) and torsion of the bridge region (in two places, PCs 2 and 3) occur simultaneously for the DiTryp and FMOC variants, but not for the DiPhe and Acetyl variants.

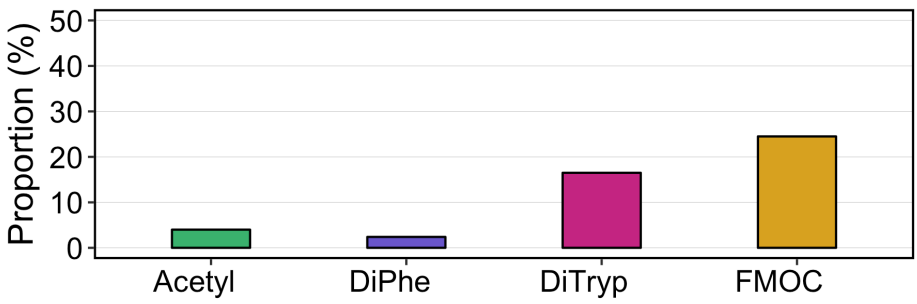



**Figure 6.** Proportion of the trajectory that each variant spends in the putative binding conformation.

This unique population suggests a possible mechanism of binding to the tropoelastin receptor. The PCs defining this population control movement of the cofactor and the bridge region. PC 1 controls hydrophobic folding towards the bioactive region. PCs 2 and 3 are mainly associated with movement of the dihedral angles either side of residues $G_4$, which is highly flexible, and $K_5$, which is charged. When they occur together, these movements shift the position of $K_5$ and lead to a unique conformation of this residue with respect to the other four charges in domain 36. If this behaviour is essential to form the binding conformation, then it implies that residues $G_4$ and $K_5$ are just as crucial as $R_9KRK_{12}$ to the function of domain 36. One way to test this is by aligning domain 36 sequences from different species, which is shown in Figure 7.


**Figure 7.** Sequence alignment of part of domain 36 from several species. Some regions are highly conserved across species, suggesting a vital function. The cysteines (yellow) are disulfide bonded and so likely have a structural role. The glycines and charged residues (green) could manoeuvre the charges into a binding conformation.

Mutations that change the sequence of a protein are common. When they occur in a region that is crucial to the function of the protein then there is an evolutionary cost and the mutation is unlikely to persist. The conservation of some parts of domain 36 across a range of species therefore suggests that these areas have a functional role in tropoelastin. The bioactive region $G_8RKRK_{12}$ is highly conserved due to its role in binding the tropoelastin receptor via charge interactions[23]. The cysteines ($C_2$ and $C_7$) are also conserved, and since they are the only disulfide bond in tropoelastin they likely have some structural function. The conservation of residues $G_4K_5$ in the bridge region suggests they too have a function. In light of the dPCA results, it is likely that the high flexibility of $G_4$ facilitates movement of

charged $K_5$, changing the overall arrangement of charges. If domain 36 peptides bind by conformational selection, then this mechanism can explain the cell activity trends of the four variants.

### 5.3.4 dPCA: Induced fit

In the other class of IDP binding, induced fit, peptides with slower dynamics bind a receptor because they have a lower entropic barrier. The dynamics of the domain 36 peptides can be indirectly measured by way of the mean displacement of each PC. This measure corresponds to the amount of movement through PC-space for each of the variants, allowing a comparison of the frequency and magnitude of structural change.

An example of this comparison is shown in Figure 8a-c. The time series of PC 1 (Figure 8a), which describes hydrophobic folding, can be plotted as the mean displacement in PC-space (Figure 8b). The gradients of the mean displacement correspond to the frequency of movement at the cofactor region, which is an indirect measure of the dynamics (Figure 8c). For the three hydrophobic cofactor variants, hydrophobic folding of the cofactor is stabilized by nonpolar interactions between the bridge and the rest of the peptide, as well as favourable hydrophilic interactions in the surrounding water (i.e. the hydrophobic effect). Accordingly, the gradient of the mean displacement for these three variants is significantly lower than the Acetyl variant, whose cofactor is not stabilized and, as a result, is more mobile. Since the Acetyl variant has the most fluctuation in this region, and the DiPhe variant has the slowest, the dynamics change in this region alone cannot be responsible for cell binding of the DiTryp and FMOC variants.

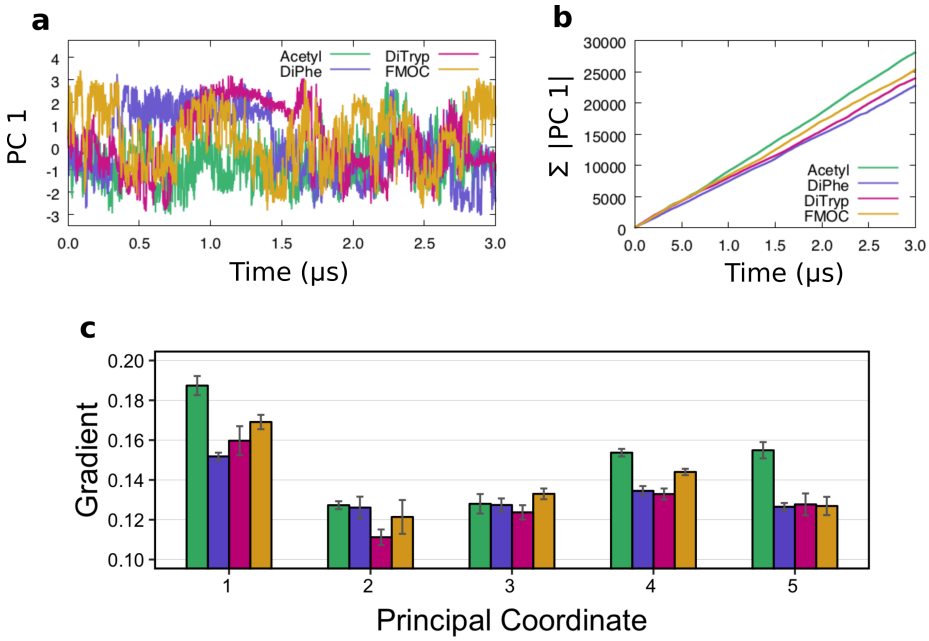



**Figure 8.** Comparison of the movement through PC-space shows no difference between the cell binding and inactive variants. **a**) Time series of PC 1. **b**) Cumulative sum of the displacement of PC 1. **c)** Mean displacement per simulation step of PCs 1-5 show the frequency of change of the top five structural movements in the domain 36 variants. There are differences between the four variants, but they do not explain the cell binding trends. Error bars = s.d. calculated by blocking average (see Appendix B).

The analysis is the same for PC 2 to PC 5: While there are differences in peptide dynamics, they don't separate thee cell active and inactive variants. As a result the peptide dynamics can't be used to rationalize the cell activity trends, and induced fit is not the mechanism of domain 36 peptide binding to the tropoelastin receptor.

The method of simulation used here, equilibrium dynamics, is an unbiased method that preserves the time constant of conformational change. This allows for analysis of peptide dynamics or, indirectly, entropy. Other techniques are well-established for analysing peptide conformation, in particular replica exchange molecular dynamics[24]. This technique, which uses simultaneous computation of replica simulations at higher temperature to overcome energetic barriers between conformations, effectively reduces the magnitude of the energetic barrier between structures. This helps to generate a Boltzmann-weighted population ensemble of structures more quickly. Without such non-equilibrium biasing techniques, limited exploration of the structural landscape becomes a significant weakness for simulations of proteins and folded peptides. For IDPs, though, the ensemble is mainly defined not by

the depth of the landscape, but by the breadth of available structures. Most structures are separated by barriers smaller than kT at room temperature (see Figure 1), which gives rise to the large range of structures observed for domain 36. In this case, then, the additional computational requirements of replica exchange MD are not justified, since they have no benefit over the equivalent amount of simulation in the equilibrium state.

In addition, while generating a complete conformational landscape (which can be done by equilibrium or nonequilibrium methods) is required to gauge the contribution of conformational selection, it is not enough for determining the contribution of induced fit. Equilibrium simulations allow straightforward analysis of kinetic rates, which in turn facilitates a comparison of entropy to determine the contribution of induced fit. The time series of PC 1, which describes the largest magnitude structural movement, shows that change can occur on the scale of ~100 ns (Figure 8a). This is fast enough that unbiased, equilibrium simulation is justified, although reaching a fully converged population ensemble may require more time than 3 μs. The use of equilibrium simulations to measure peptide conformation and dynamics in this way has precedent[25], and while kinetic rate determination from biased simulations is possible, its difficulty has been noted[26].

One way to gauge the degree of sampling of the equilibrium simulations is the autocorrelation coefficient of the principal components[27]. This measures the degree of correlation of each principal component over time. Complete sampling is achieved when the trend of the autocorrelation coefficient has decayed towards zero. The autocorrelation coefficients for the domain 36 peptides display different levels of sampling (Supplementary Figure 3). PC1, which describes the movement of the hydrophobic cofactor, shows slow variations on the scale of 1 μs, indicating more sampling would be required to get a complete description of its movement. PCs 2 and 3, describing the movement of the charged residue $K_5$, indicate better sampling with PC2 quickly decaying to zero and PC3 trending around zero. The autocorrelation coefficients of PC2 and PC3 demonstrate a peculiarity of PCAs of peptide simulations – unlike for proteins, the largest variance movement is not necessarily the slowest. PC2, which is larger in magnitude, decays to zero much more quickly than PC3. This problem has prompted the use of time-lagged independent component analysis, which automatically

chooses independent components based on the correlation behaviour[28]. As new tools are released this method is becoming more accessible[29].

### 5.3.5 Specific interactions

The analysis of conformations and dynamics of the solution-phase domain 36 peptides leads to the following questions: How do all three hydrophobic cofactor variants exhibit hydrophobic folding, yet only the DiTryp and FMOC variants exhibit this in combination with movement of residue $K_5$? What is the origin of the difference in structure that yields the binding population in these two variants? Connecting the mathematical description by PCA back to the relevant interactions allows for a comprehensible description of the molecular mechanisms causing this difference.

Guided by the dPCA, determination of the relevant collective variables can be achieved by focusing on the interaction between the cofactor and the bioactive region. There is no formal procedure for determining specific interactions from dPCA – this analysis can be performed only by close analysis of the trajectories while using the dPCA results as a guide. The cofactor does not interact directly with the bridge region due to steric hindrance caused by their close proximity along the peptide backbone, so interaction with the bioactive region must be the origin of the changed distribution for PCs 2 and 3. Inspection of the trajectories reveals a subtle difference in the interaction between the cofactors and the bioactive region: the DiPhe cofactor interacts mainly with the aliphatic sidechain atoms of the active region amino acids ($G_9RKRK_{12}$), while the FMOC and DiTryp cofactors interact with both the aliphatic sidechain atoms and the charged atoms, forming pi-cation bonds. Figure 9 shows the proportion of pi-cation bonds for each variant over the entire trajectory and demonstrates that the DiTryp and FMOC variants approximately 50% more likely to form these interactions.

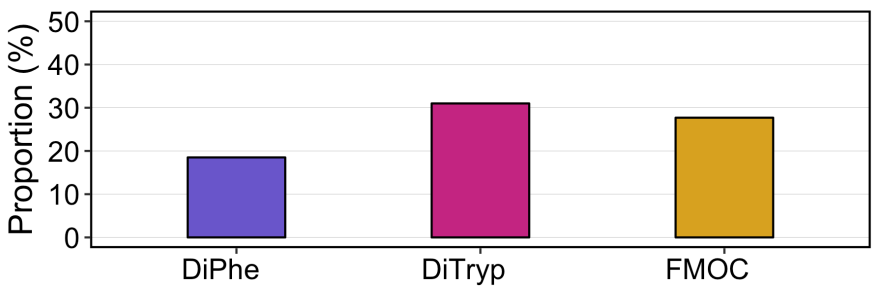


**Figure 9.** Proportion of total trajectory showing a pi-cation bond between the cofactor and the bioactive region of the domain 36 variants. The variants that bind to the tropoelastin receptor (DiTryp, FMOC) are approximately 50% more likely to pi-cation bond. The Acetyl variant is not shown because it has no aromatic groups to participate in this type of bond.

This difference can be attributed to both the charge and shape of the cofactors. A pi-cation bond has its origin in electrostatic interactions. Due to the higher electronegativity of carbon than hydrogen, aromatic rings have negative charge density at their centre. This attracts positively charged groups (cations) like lysine and arginine, with affinity comparable to binding a water molecule[30]. For this reason tryptophan, which is more polar than other aromatic amino acids, forms pi-cation bonds more readily in proteins[31]. This is demonstrated here by the higher number of such bonds for the DiTryp variant than the DiPhe variant. The CHARMM force fields have no explicit interaction term to facilitate pi-cation bonds, however the partial charges already do account for the polarity of the aromatic rings. When electron movement is taken into account (i.e. polarization), these bonds are likely to become even stronger. Thus the effect seen here is likely somewhat understated and would be enhanced with a polarizable force field.

The pi-cation bonding preference of the FMOC cofactor, whose polarity approximates a phenylalanine sidechain, cannot be explained in the same way. Here, the broad surface area of the FMOC aromatic moiety, more than twice as large as a phenylalanine, may be facilitating pi-cation bonding by maintaining contact with charged groups even when thermal motion threatens to separate them. The FMOC parameters used in this study, which were generated by the CHARMM CgenFF[32] along with analogy to existing aromatic groups, have not been thoroughly tested here. They are, however, close to previously determined and validated parameters using either the CHARMM[33] or a closely related forcefield[34].

Like a PCA, pi-cation interactions require the movement of multiple underlying variables, but unlike PCA they are intuitively understandable and lead to biological insight. Here, they shed light on how domain 36 binds to cells from native tropoelastin. The dPCA of the domain 36 variants shows that the unique binding population is characterised by 1) hydrophobic folding, 2) movement of residue $K_5$, and 3) pi-

cation bonding between the cofactor and the active region. These three movements, which do not coincide in the inactive variants, manoeuvre the 5 charges on the peptide into the binding conformation.

In the past it has been theorised that the four charges on domain 36, in native tropoelastin, are bent into a pocket shape that fits into a negatively charged region on the receptor[35], implying conformational selection. Others hypothesized that it is simply the clustering of charges so close together that attracts the ligand towards the receptor[14]. According to the hotspot theory of IDP binding[7], this would suggest an induced fit mechanism whereby charge interactions overcome the peptide entropy. The results presented here show that it is the relative position of the five charges that facilitates cell binding. It follows that in native tropoelastin the hydrophobic backing provided by the adjacent domain 33, as well as structural stabilization provided by the disulfide bond in the bridge region, also act to hold the charges in the correct position to bind by conformational selection.

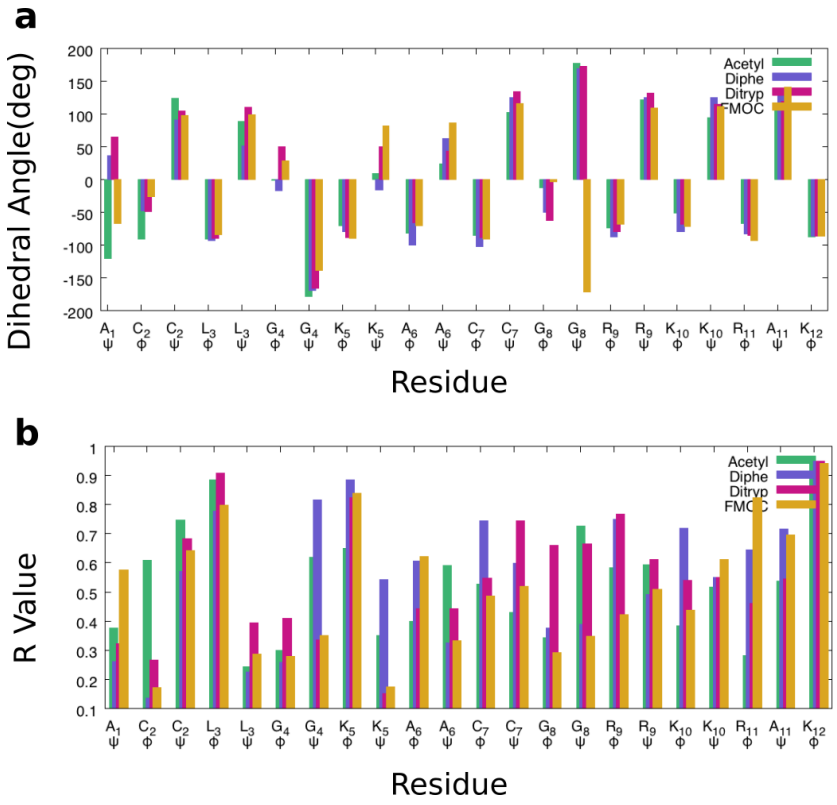
## 5.4 Conclusion

Typically, the mechanism of action of peptide ligands is determined in simulation by starting from the bound state and moving to the free, aqueous state. Computationally this is favourable since it avoids sampling large regions of irrelevant conformational space in the aqueous phase. Without a crystal structure of the bound peptide-target dimer, this process is considered extremely difficult or impossible[8]. For the domain 36 peptide, the lack of even a known molecular target (i.e. the tropoelastin receptor) escalated the difficulty further since the bound state could not be directly verified.

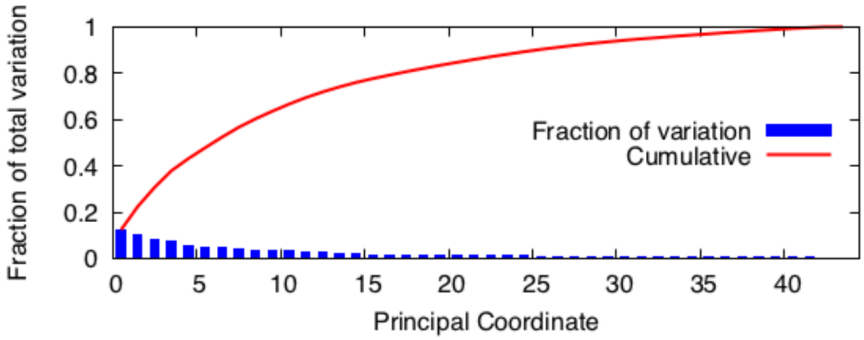
Despite this, the dynamics and structural landscape of known binding and non-binding variants was used, in a process of elimination, to assess the degree of binding caused by induced fit and conformational selection. This relied on a method of clustering the conformations. Methods such as sketch-map can separate conformations, but in this chapter we have used dPCA. While both methods can be used to tease apart the large number of conformations associated with IDPs, only dPCA is capable of measuring the peptide dynamics. This is because it allows structural change to be plotted as a time series, while sketch-map does not. This procedure was used to show that there is a unique conformation of the domain 36 peptide that occurs only for the binding variants and that the dynamics were not discriminatory between the variants.

The domain 36 peptide binds by conformational selection, and this likely occurs for the native tropoelastin protein as well. The stark difference in cell binding activity between cofactors that are chemically similar indicates the difficulty in rationally designing peptide ligands, but combining computational and experimental approaches can overcome this.

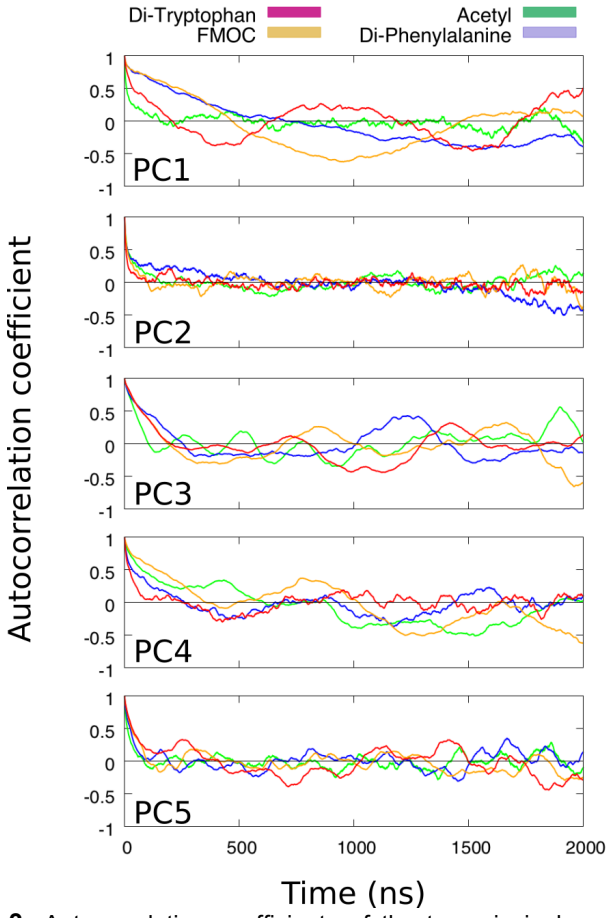
## 5.5 Supplementary figures




**Supplementary Figure 1.** Backbone dihedral angle statistics for domain 36 **a)** Mean backbone dihedral angles for the domain 36 peptide. **b)** *R* values, which measure variance, of the backbone dihedral angles. Values close to 1 indicate small variance, while values close to zero indicate high variance.




**Supplementary Figure 2**. The eigenvalues of the 44 principal components (PCs) of the dPCA. The PCs are ordered by the fraction of the total variation in the data that they explain. The first PC explains less than 15% of the variation. This is typical for intrinsically disordered peptides, which have many different structures.




**Supplementary Figure 3.** Autocorrelation coefficients of the top principal components of the domain 36 peptides. Decay towards zero indicates complete sampling of the explored phase space. These autocorrelation coefficients show varying degrees of decay towards zero, indicating the level of sampling of each type of movement of the peptide.

## 5.4 References


1 Muiznieks, L. D. & Weiss, A. S. Flexibility in the solution structure of human tropoelastin. *Biochemistry* **46**, 8196-8205 (2007).

2 Floquet, N. *et al.* Structure and modeling studies of the carboxy-terminus region of human tropoelastin. *Matrix biology* **24**, 271-282 (2005).

3 Fisher, C. K. & Stultz, C. M. Constructing ensembles for intrinsically disordered proteins. *Current opinion in structural biology* **21**, 426-431 (2011).

4 Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics* **23**, 566-579 (1995).

5 Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proceedings of the National Academy of Sciences* **97**, 8868-8873 (2000).

6 Du, X. *et al.* Insights into protein–ligand interactions: Mechanisms, models, and methods. *International journal of molecular sciences* **17**, 144 (2016).

7 London, N., Movshovitz-Attias, D. & Schueler-Furman, O. The structural basis of peptide-protein binding strategies. *Structure* **18**, 188-199 (2010).

8 Kurcinski, M., Kolinski, A. & Kmiecik, S. Mechanism of folding and binding of an intrinsically disordered protein as revealed by ab initio simulations. *Journal of Chemical Theory and Computation* **10**, 2224-2231 (2014).

9 Bellucci, L. *et al.* The interaction with gold suppresses fiber-like conformations of the amyloid β (16–22) peptide. *Nanoscale* **8**, 8737-8748 (2016).

10 Ceriotti, M., Tribello, G. A. & Parrinello, M. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proceedings of the National Academy of Sciences* **108**, 13023-13028 (2011).

11 Mu, Y., Nguyen, P. H. & Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins: Structure, Function, and Bioinformatics* **58**, 45-52 (2005).

12 Jain, A., Hegger, R. & Stock, G. Hidden complexity of protein free-energy landscapes revealed by principal component analysis by parts. *The Journal of Physical Chemistry Letters* **1**, 2769-2773 (2010).

13 Espinoza-Fonseca, L. M., Ilizaliturri-Flores, I. & Correa-Basurto, J. Backbone conformational preferences of an intrinsically disordered protein in solution. *Molecular BioSystems* **8**, 1798-1805 (2012).

14 Akhtar, K. *et al.* Oxidative modifications of the C-terminal domain of tropoelastin prevent cell binding. *Journal of Biological Chemistry* **286**, 13574-13582 (2011).

15 *STRIDE*, <http://webclu.bio.wzw.tum.de/stride/>

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

17 Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of machine learning research* **12**, 2825-2830 (2011).

18 Consortium, U. UniProt: the universal protein knowledgebase. *Nucleic acids research* **45**, D158-D169 (2016).

19 Notredame, C., Higgins, D. G. & Heringa, J. T-coffee: a novel method for fast and accurate multiple sequence alignment1. *Journal of molecular biology* **302**, 205-217 (2000).

20 Kubelka, J., Hofrichter, J. & Eaton, W. A. The protein folding 'speed limit'. *Current opinion in structural biology* **14**, 76-88 (2004).

21 Hayward, S. & Go, N. Collective variable description of native protein dynamics. *Annual review of physical chemistry* **46**, 223-250 (1995).

22 Bakan, A. & Bahar, I. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proceedings of the National Academy of Sciences* **106**, 14349-14354 (2009).

23 Bax, D. V., Rodgers, U. R., Bilek, M. M. & Weiss, A. S. Cell adhesion to tropoelastin is mediated via the C-terminal GRKRK motif and integrin αVβ3. *Journal of biological chemistry* **284**, 28616-28623 (2009).

24 Ostermeir, K. & Zacharias, M. Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1834**, 847-853 (2013).

25 Lindorff-Larsen, K., Trbovic, N., Maragakis, P., Piana, S. & Shaw, D. E. Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *Journal of the American Chemical Society* **134**, 3787-3791 (2012).

26 Leahy, C. T., Kells, A., Hummer, G., Buchete, N.-V. & Rosta, E. Peptide dimerization-dissociation rates from replica exchange molecular dynamics. *The Journal of Chemical Physics* **147**, 152725 (2017).

27 Ernst, M., Sittel, F. & Stock, G. Contact-and distance-based principal component analysis of protein dynamics. *The Journal of Chemical Physics* **143**, 12B640_641 (2015).

28 Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G. & Noé, F. Identification of slow molecular order parameters for Markov model construction. *The Journal of chemical physics* **139**, 07B604_601 (2013).

29 Harrigan, M. P. *et al.* MSMBuilder: statistical models for biomolecular dynamics. *Biophysical journal* **112**, 10-15 (2017).

30 Dougherty, D. A. Cation-π interactions involving aromatic amino acids. *The Journal of nutrition* **137**, 1504S-1508S (2007).

31 Gallivan, J. P. & Dougherty, D. A. Cation-π interactions in structural biology. *Proceedings of the National Academy of Sciences* **96**, 9459-9464 (1999).

32 Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* **31**, 671-690 (2010).

33 Sasselli, I. R., Ulijn, R. & Tuttle, T. CHARMM force field parameterization protocol for self-assembling peptide amphiphiles: the Fmoc moiety. *Physical Chemistry Chemical Physics* **18**, 4659-4667 (2016).

34 Mu, X., Eckes, K. M., Nguyen, M. M., Suggs, L. J. & Ren, P. Experimental and computational studies reveal an alternative supramolecular structure for Fmoc-dipeptide self-assembly. *Biomacromolecules* **13**, 3562-3571 (2012).

35 Brown, P. L., Mecham, L., Tisdale, C. & Mecham, R. P. The cysteine residues in the carboxy terminal domain of tropoelastin form an intrachain disulfide bond that stabilizes a loop structure and positively charged pocket. *Biochemical and biophysical research communications* **186**, 549-555 (1992).

# Chapter 6. Controlling peptides at the interface

Molecular dynamics simulations of peptides have proven useful to characterise the interactions that occur at synthetic surfaces. The simulations show how molecule shape and charge can change the function of bioactive peptides. How can these interactions be exploited to create more effective biomaterials? This chapter demonstrates a new technique to control the surface orientation and concentration of a peptide that becomes covalently bound to a polymer coating. The design of the peptide used in this chapter incorporates the knowledge developed in the previous chapters. Because the underlying interactions are the same, in future this technique can be generalised to other bioactive ligands.

This chapter has been published in *Nature Communications*.

## 6.1. Methods

Because the format of *Nature Communications* does not permit exposition on methods, this section introduces the methods used in analysing the functionalization with biomolecules presented in the article.

### 6.1.1 Peptide design

The main body of Chapter 6 is a published manuscript that contributes to the field of biomaterials by developing a new technique. This technique offers control over how small molecules, such as peptides, covalently bind to a putative biomaterial surface. The development of this technique required a test molecule that fulfils multiple purposes at once. Not all of these would be required in the envisioned end use case, but they are essential for building an evidence base that compellingly supports the proposed model of the technique presented in the paper.

The peptide:

- should be mostly linear, so that it has easily defined orientations (‘upward facing’ and ‘downward facing’, with respect to the surface).
- must have a dipole moment that interacts with electric fields, and whose direction is parallel to the linear structure of the peptide. Ideally the dipole moment should remain parallel to the long axis of the peptide this across a wide pH range.
- should be small enough such that any dipole moment induced by an electric field is less than the permanent dipole moment. The peptide must also be small enough so that the torque from the interaction of the dipole with an electric field is strong enough to rotate it to align it with the electric field.
- should have a net charge, so as to feel a net force due to the electric fields. Ideally this also should persist across a wide pH range.
- must have some bioactivity to be consistent with the intended application of biomaterial design, and which is only active in the ‘upward facing’, orientation.
- must have some method of detection by biological assays or spectroscopy.

Using simulations described below, many variants of the peptide were tested. All but one failed at least one of the criteria. For example peptides that are small enough to be orientable by electric fields also have a tendency towards conformational disorder, which has been described in Chapter 5. As a result, maintaining linearity is unlikely. A dipole requires positive and negative charges separated by a distance. According to:

$$\vec{\mu} = \sum_i q_i \vec{r_i},$$

where $\vec{\mu}$ is the dipole moment vector, $q_i$ is the magnitude of the $i^{th}$ charge, and $\vec{r_i}$ is the vector representing the position of the $i^{th}$ charge, the magnitude of the dipole is increased by separating the charges by a greater distance. A strong dipole is beneficial for controlling orientation, because it reduces the electric field strength necessary to overcome thermal vibration (kT). However, due to the flexibility of most short peptides, separated charges can form salt bridges, potentially lessening the dipole magnitude. Using folded peptides could maintain the charge separation, but then the molecule may fail the smallness and linearity criteria. On top of this, most bioactive epitopes are charged because of "hotspot residues"[1], described in Chapter 5, and so potentially interfere with dipoles elsewhere.

Detection by assays or spectroscopy requires the presence of a chromophore or some other tag. Due to the restriction in available space, two dimensional monolayers of peptides have such low numbers of molecules compared to the three dimensional solution phase that chromophores are expected to produce very low signal, and most tags need to be presented into solution, in the same place as the bioactive epitope would be. Also, tags such as a six-histidine tag are sensitive to pH changes, potentially confounding the results.

The peptide used in these experiments combines the function of each region in order to satisfy all criteria. The sequence – Ac-FFMMMAAAAAAAAAADDDDDK-$NH_2$ – has a charged 'FLAG' tag on one end (-DDDDK)[2]. This has an overall charge as well as an inherent dipole and binds antibodies in enzyme-linked immunosorbent assays (ELISA, described below), which allows sensitive detection of peptide on the surface as well as illustrating a form of bioactivity. The methionines ('M') introduce sulphur atoms. Since the polymerized biomaterial surface did not include sulphur, this also allowed detection of their presence by X-ray Photoelectron Spectroscopy (XPS). The methionines are

hydrophobic, and in combination with the phenylalanines ('F') this gave the peptide a hydrophobic base. When separated by a chain of alanines ('A'), which have small sidechains and so are chemically inert[3], the charged FLAG tag and hydrophobic regions are pushed apart due to their different affinities for water, and this enforces linearity. Finally, the direction of the dipole is congruous with the overall net negative charge. That is, a positively charged surface would attract the peptide in greater concentrations as well as orienting them such that the bioactive FLAG region faces "up" into solution where it can interact with biological molecules.

### 6.1.2 Dual buffer system

The technique developed in this chapter uses electric fields to control the approach of peptides to the surface. One way to induce electric fields on a surface is by altering the pH of the solution used to deliver the peptides. This changes the number of hydrogens present on ionisable groups at the surface, and thus the surface charge. The pH of a solution is commonly controlled using buffers – conjugate acid and base pairs that maintain the number of hydrogens in solution by preserving the hydrogen concentration (i.e. pH). When the number of hydrogens is decreased, as by the addition of an acid, the acid part of the buffer turns to its conjugate base by donating hydrogens back in solution[4].

The peptide immobilization was performed at two pHs, which requires two buffers with different pKa (the pH at which half of the buffer population is acid and half is conjugate base). Because the biomaterial surface used here is reactive, it is undesirable to have additional chemical species that potentially react with it. To control for this, these experiments use a dual buffer system that is capable of maintaining different pH levels while keeping the ionic strength and the identity of all chemical species constant. One buffer is active at a lower pH, while the other is entirely depleted, and vice versa. This has the advantage that any specific chemical reactions between the surface and the solution would thus be the same in all experimental conditions. The dual buffer system was purpose built for these experiments in order to maximally alter the surface charge without changing the charge on the peptide. It derives from the concept of the 'universal buffer', first described in French language[5] but also reported in English language textbooks since[6].

### 6.1.3 Peptide immobilization

The surfaces used in this experiment are plasma polymerized coatings developed and produced by co-author Dr. Behnam Akhavan[7]. These surfaces contain reactive free radicals that are capable of covalently binding biological molecules (described in the main text). Due to the sensitivity of the detection techniques, the peptide immobilization procedure required a careful approach to ensure even immobilisation across samples and avoid contamination. As described in the manuscript, the surface was cut into 1cm x 1cm samples. These fit into the standard 24-well plate that is commonly used in microbiology experiments (such as ELISA). To accommodate these, the wells in the device used to apply the electric field (provided by Dr. Edgar Wakelin) were the same dimensions as a 24-well plate. This necessitated several rounds of washing and sonication of the device with ethanol and distilled water to ensure proper cleaning between immobilization runs.

### 6.1.4 Peptide Detection

This chapter uses several experimental techniques to detect the presence of peptides on the surface. Since any single technique is unable to completely characterise the surface, a combination of the techniques was required to build a cogent picture of the surface composition and the concentration and orientation of the bound peptides.

X-ray photoelectron spectroscopy (XPS) is an electron spectroscopy technique using the photoelectric effect to determine the atomic number and concentration of atoms in the surface. An XPS experiment proceeds by placing the sample in a vacuum and irradiating it using photons in the X-ray range. Depending on the energy some of these photons will impinge on atoms in the sample leading to emission of photoelectrons, which can be detected[8]. Electron detection is used to determine the properties of the emitting atoms - the energy at which the electrons are emitted is related to their binding energy within the atom of origin, and the number of electrons emitted with a particular energy is related to the concentration of atoms with electrons in shells with that binding energy in the sample. As all atoms have electron shells with distinctive binding energies, XPS spectra reveal the atomic percentage elemental concentrations (for all elements except hydrogen). Practically, this technique has lower spatial resolution than other surface measurements like scanning probe

microscopy, so is more suited to detecting the presence of biomolecules rather than the spatial arrangement[9]. In addition the sampling depth is up to 10 nm, meaning that both biomolecule and surface atoms are likely to be detected (the photons can penetrate further, but the photoelectrons emitted from greater depths are scatted by collisions with the sample so do not arrive at the detector)[8]. For peptides, which are usually up to 1-3 nm, this means the signal arising from the biomolecule is small compared to the signal arising from the surface. In addition to this, the surface used in this chapter has a similar atomic composition to the peptide, requiring the addition of sulphur-containing methionines to the peptide sequence to enable detection. Since methionine is a hydrophobic residue there are limits on the number that can be added before sacrificing solubility of the peptide so, as discussed in the main text, detection of the peptide was at the lower limits of the capability of XPS.

Time of flight secondary ion mass spectrometry (ToF-SIMS) is a sputtering technique that, in contrast to photons for XPS, uses high energy ions as the impinging particles. ToF-SIMS analysis uses bombardment by a primary ion beam to perturb a small region in the sample surface by the deposition of significantly high energy. The primary ions have higher kinetic energy than the bond energies in the sample, so the perturbed region undergoes bond fragmentations. This leads to the production of so-called secondary ions, whose mass/charge ratio is then measured[10]. The mass/charge ratios of common fragments produced by proteins are well characterised, so the identity of the amino acid sequence of proteins or peptides on the surface can be determined. The sampling depth of ToF-SIMS is far smaller than XPS, being on the order of 1-2 nm. As a result the sputtered fragments sample a single region of the biomolecule so, if there is a preferred orientation (as opposed to random orientations) as well as unequal distribution of amino acid types within the biomolecule, this preferred orientation can then be inferred[8]. Like XPS, ToF-SIMS is performed in vacuum, meaning protein structures can change relative to the solvated state. For a linear peptide such as used here this likely leads to some level of collapse onto the surface. Even if the rest of the peptide is pushed towards the surface under vacuum, ultimately only one end is covalently bound, and this end will be closest to the surface. Thus, on average, the most accessible parts of the peptide (those measured by ToF-SIMS) will correspond to the orientation in the aqueous phase.

In addition to the vacuum-phase atomic detection techniques, ELISA was used to measure the presence of the FLAG epitope. ELISA is a technique that uses antibodies for the detection of biomolecules. Antibodies, also known as immunoglobulin, are proteins produced by cells of the adaptive immune system. This system is capable of creating antibodies that bind to proteins on exogenous biological material, such as bacteria and viruses, flagging them for attack by the immune system and so developing 'learned' immunity to pathogens. The antibodies consist of a common structural body with highly variable tips, which are capable of binding a large array of biological epitopes[11].

In ELISA, special antibodies are created in live animals by injecting the necessary epitope (the FLAG tag, in this case), creating an immune response and up-regulation of the production of the antibody. The antibodies can be extracted, transported and used to bind the epitope on which they were 'raised' in assays. After linkage to the epitope by the so-called primary antibody, ELISAs require a second stage of detection by enzymatic digestion of a fluorophore[12]. The change in colour of the fluorophore indicates the presence of the secondary antibody, which implies presence of the primary antibody[13]. In the assay used in this chapter a secondary antibody, with the enzyme covalently attached, is used to detect the primary one. To increase detection, this chapter uses polyclonal secondary antibodies. These can bind to multiple points on the primary antibody and so amplify the signal detection. This allows for high sensitivity of small amounts of peptide. It does, however, mean that each detected peptide is associated with multiple, bulky antibodies. On a two dimensional surface, as compared to three dimensional volumes in solution, space is at a premium and we found, as discussed in the main text, that this quickly led to saturation of the assay even at low peptide surface concentration because of molecular crowding of the secondary antibodies.

### 6.1.5 Electro-kinetic analysis

One method of controlling the electric field near a surface is to alter the surface charge by modulating the pH of the solution phase using buffer chemistry, as described above. To determine how much the surface charge was being altered by the solution, electro-kinetic characterisation of the surface across a wide range of pH was performed. This term refers to the magnitude of force applied to a flowing liquid of variable pH across a surface. In this technique, the charge created on the surface by the pH

of the solution leads to an electric field extending into solution, which interacts with ions in the solution to constrict their movement and influence the liquid flow. The models describing the behaviour of the solution and ions near the surface are highly complex, and are best approached by the textbook by Hunter[14].

Briefly, the electric field is strongest at the interface between the charged surface and the solution. The change in the electric field or potential from the surface as one moves into solution is dependent on the positions taken up by proximate charged particles such as ions. Initially, the model of Helmholtz posited that a monolayer of immobile ions of opposite charge that neutralized the charge of the surface – a double layer of electrical charge – was attracted from solution[15]. Due to having an invariable distance between the separated charges (being adjacent, they would be separated by just the width of an atom) this model predicted an invariable capacitance, which is dependent on separation distance. This prediction is easily disproven by charging a metal surface in solution to different voltages and observing variable capacitance. The Gouy-Chapman model followed, which accounted for the thermal motion of the ions in solution[16,17]. This can be conceptualised as lifting the neutralizing ion layer from the surface due to the atomic kinetic energy, and spreading the layer into solution across several nanometers. This model combines Poisson's equation, which relates charge to electrical potential, with a Boltzmann distribution of ions in solution. The combination of these predicts a decay of electrical potential moving away from the charged surface caused by changing concentrations of ions in solution, which are attracted to the charged surface. Due to the stronger than exponential growth in salt concentration approaching the surface, this predicts impossibly high salt concentrations at highly charged surfaces. Accounting for this is the contribution of Stern, which adds a Helmholtz-like layer of immobile ions that reduces the effective potential of the surface to levels associated with a realistic concentration of ions[18].

In this chapter, electrical double layer theory as described here has two relevant consequences. The first is that even small voltages are screened by the neutralizing ions over a distance of just several nanometers. For the voltages used here, which are achievable with common household batteries, this leads to an electric field strength approaching billions of volts per meter, amplifying the force exerted on the charged, polar peptide. The second consequence is that the application of a flowing current

across the surface leads to shearing of the ions away from the surface due to collisions with water molecules. The resulting charge separation resists the flow of water. By measuring the resistance to the applied flow the electrical potential, known as ζ potential, can be calculated. This is best described by Kirby and Hasselbrink[19]. The ζ potential measures the electric potential at the outer Helmholtz plane, sometimes called a Stern layer, where ions are mobile. While this is not the same as the electric potential at the interface itself, it is a useful proxy for determining how the surface is charged at different pH levels. This measurement thus provided information on how the surface charge and electric field at the surface depended on pH, allowing for the control of peptide movement at the interface simply by changing the pH of the aqueous phase.

## 6.2 References


1 London, N., Movshovitz-Attias, D. & Schueler-Furman, O. The structural basis of peptide-protein binding strategies. *Structure* **18**, 188-199 (2010).

2 Einhauer, A. & Jungbauer, A. The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. *Journal of biochemical and biophysical methods* **49**, 455-465 (2001).

3 Morrison, K. L. & Weiss, G. A. Combinatorial alanine-scanning. *Current opinion in chemical biology* **5**, 302-307 (2001).

4 Beynon, R. & Easterby, J. *Buffer solutions*. (Taylor & Francis, 2004).

5 Fernández, C. M. & Martin, V. C. Preparation d'un tampon universel de force ionique 0, 3 M. *Talanta* **24**, 747-748 (1977).

6 Perrin, D. *Buffers for pH and metal ion control*. (Springer Science & Business Media, 2012).

7 Akhavan, B., Wise, S. G. & Bilek, M. M. Substrate-regulated growth of plasma-polymerized films on carbide-forming metals. *Langmuir* **32**, 10835-10843 (2016).

8 Ratner, B. D., Hoffman, A. S., Schoen, F. J. & Lemons, J. E. *Biomaterials science: an introduction to materials in medicine*. (Academic press, 2004).

9 Castner, D. G. & Ratner, B. D. Biomedical surface science: Foundations to frontiers. *Surface Science* **500**, 28-60 (2002).

10 Leonard, D. & Mathieu, H. Characterisation of biomaterials using ToF-SIMS. *Fresenius' journal of analytical chemistry* **365**, 3-11 (1999).

11 Alberts, B. *et al.* Molecular Biology of the Cell (Garland, New York, 2002). *Google Scholar*.

12 AVrtAMEAS, S. Coupling of enzymes to proteins with glutaraldehyde. *Immunochemistry* **6**, 43-52 (1969).

13 Miles, L. & Hales, C. Labelled antibodies and immunological assay systems. *Nature* **219**, 186 (1968).

14 Hunter, R. J. *Zeta potential in colloid science: principles and applications*. Vol. 2 (Academic press, 2013).

15 Gongadze, E., Petersen, S., Beck, U. & van Rienen, U. in *COMSOL Conference.* 14-16.

16 Gouy, M. Sur la constitution de la charge électrique à la surface d'un électrolyte. *J. Phys. Theor. Appl.* **9**, 457-468 (1910).

17 Chapman, D. L. LI. A contribution to the theory of electrocapillarity. *The London, Edinburgh, and Dublin philosophical magazine and journal of science* **25**, 475-481 (1913).

18 Stern, O. Zur theorie der elektrolytischen doppelschicht. *Zeitschrift für Elektrochemie und angewandte physikalische Chemie* **30**, 508-516 (1924).

19 Kirby, B. J. & Hasselbrink Jr, E. F. Zeta potential of microfluidic substrates: 1. Theory, experimental techniques, and effects on separations. *Electrophoresis* **25**, 187-202 (2004).

nature COMMUNICATIONS

104

ARTICLE

DOI: 10.1038/s41467-017-02545-6 OPEN

# Electric fields control the orientation of peptides irreversibly immobilized on radical-functionalized surfaces

Lewis J. Martin [1], Behnam Akhavan [1,2] & Marcela M.M. Bilek [1,2,3,4]

Surface functionalization of an implantable device with bioactive molecules can overcome adverse biological responses by promoting specific local tissue integration. Bioactive peptides have advantages over larger protein molecules due to their robustness and sterilizability. Their relatively small size presents opportunities to control the peptide orientation on approach to a surface to achieve favourable presentation of bioactive motifs. Here we demonstrate control of the orientation of surface-bound peptides by tuning electric fields at the surface during immobilization. Guided by computational simulations, a peptide with a linear conformation in solution is designed. Electric fields are used to control the peptide approach towards a radical-functionalized surface. Spontaneous, irreversible immobilization is achieved when the peptide makes contact with the surface. Our findings show that control of both peptide orientation and surface concentration is achieved simply by varying the solution pH or by applying an electric field as delivered by a small battery.

[1] School of Physics, University of Sydney, Sydney, NSW 2006, Australia. [2] School of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, Sydney, NSW 2006, Australia. [3] Charles Perkins Centre, University of Sydney, Sydney, NSW 2006, Australia. [4] University of Sydney Nano Institute, University of Sydney, Sydney, NSW 2006, Australia. Lewis J. Martin and Behnam Akhavan contributed equally to this work. Correspondence and requests for materials should be addressed to B.A. (email: behnam.akhavan@sydney.edu.au) or to M.M.M.B. (email: marcela.bilek@sydney.edu.au)

Modern medicine increasingly relies on implantable biomedical devices[1]. The function(s) of these devices are often limited due to unsuccessful integration with host tissue; and in extreme cases this necessitates replacement of the device by revision surgery[2]. Bio-devices such as pacemakers and bone implants can cause unfavourable reactions in the surrounding host tissue, e.g. foreign body responses that lead to encapsulation by fibrotic tissue or the formation of bacterial biofilms resulting in untreatable infections[3]. These responses cause significant pain and suffering, as well as considerable economic burden for national health-care systems[4]. Functional coatings can mitigate such problems by masking the implanted devices and promoting successful integration with the body. Biologically functionalized surfaces, in particular, have the potential to direct optimal host responses by providing biological cues through molecules immobilized at the host-device interface[5].

Bio-functionalized devices are surface-engineered to present bioactive molecules. These molecules influence the biology of nearby cells by providing signals via specific interactions with cell surface receptor proteins[6]. Covalent bio-functionalization is necessary to avoid adsorption-induced denaturing[7] and/or loss of the biomolecules through protein exchange such as occurs in the Vroman effect[8,9]. Established methods for covalent attachment, however, involve cumbersome multistep wet chemistry, often using reagents that may present hurdles for regulatory approval[10]. Recently, a nonspecific, chemical linker-free approach to achieve covalent attachment of bioactive molecules directly from buffered solution has been demonstrated[11]. Covalent immobilization is achieved through reactions with radicals embedded under the surface by energetic ion bombardment[12]. Radical-functionalized surfaces can be created on both non-polymeric[13,14] and carbon-rich polymeric surfaces[15,16]. The buried radicals are capable of diffusing to the surface[17], where they react with biomolecules. The ion-treated surfaces are typically hydrophilic due to reactions with atmospheric oxygen[18]; thus, the immobilized proteins do not suffer denaturation through physical interactions with the surface. Due to the non-specific nature of the reaction, immobilized orientation could be controlled by orienting the biomolecules on approach.

Protein-functionalization of surfaces has been applied to improve the biocompatibility of medical devices[6]. Practical applications are, however, impeded by loss of function in post-packaging sterilization, denaturation-induced thrombosis[19] and the possibility of pathogen transfer from proteins produced in micro-organisms[20]. Surface functionalization with bioactive peptides may offer a solution. The smaller size of peptides compared to proteins makes them more resilient to sterilization. Furthermore, peptides can be generated synthetically as opposed to in microorganisms. Mimetic peptides are derived from the active amino acid sequences of proteins and therefore can provide the key functionality of the protein[21]. The most commonly used peptide mimics are derived from extracellular matrix proteins such as fibronectin or collagen, since they support native cell adhesion and hence integration of an implanted surface into tissue[22,23]. The RGD peptide, for example, has often been used as a proxy for fibronectin[24,25], and there are a number of other peptide mimics of growth factors[26,27].

Surface-attached biomolecules can interact with the microenvironment and direct cell behaviour only if bound in an orientation that allows access to the active site[28,29] and in a conformation in which the active site is structurally intact[30,31]. The density of surface attachment also plays a role in determining the effectiveness of the immobilized biomolecules especially with reference to promoting adhesion and spreading of cells[32]. The cell adhesive activity of the RGD peptide, in particular, is known to be sensitive to surface concentration[33], and molecular crowding of immobilized peptides has been shown to affect presentation of the active site[34,35].

Control of the surface orientation of peptides by changing their chemistry and achieving site-specific attachment has been studied before. For example, site-specific phosphorylation of serine residues in a silicon-based peptide[36] or the insertion of a cysteine residue[37]. However, controlling the orientation of peptides with these traditional chemical approaches is typically cumbersome and tedious for large-scale manufacturing. Electric fields might present an opportunity to orient peptides at biomaterial interfaces. Charged interfaces repel and attract like-charged and oppositely charged regions of a biomolecule, respectively, resulting in orientation of molecules that are asymmetrically charged.

The electric field at the surface can be manipulated by changing the solution pH or by the application of an external electric field. Electric field-induced bidirectional alignment along the long axis of large proteins has been achieved with DC fields[38] and in dielectrophoresis with AC electric fields[39]. A unidirectional orientation can only be achieved via a permanent dipole moment and hence requires small molecules (<~100 kDa), for which the induced dipole moment does not dominate[40]. Furthermore, as the dielectrophoretic force scales with particle size, increasing the arrival rate and surface immobilized density of peptides would require unfeasibly high field strengths to overcome thermal motion[41].

Here we demonstrate the use of electrostatic manipulation to control the orientation of peptides through electric field interactions with permanent dipoles. Guided by computational simulations, we strategically design a peptide incorporating a FLAG-tag functional epitope used in generating fusion proteins. A radical-functionalized plasma polymer (RFPP) surface, with a high concentration of reactive radicals, is used to covalently immobilize this peptide upon arrival at the surface. We demonstrate the control of both concentration and orientation of the immobilized peptide by varying the solution pH initially and then extend the achievable range by employing externally applied electric fields. Our findings shed light on mechanisms of biomolecule immobilization that are extremely important for the design of synthetic peptides and the production of advanced bio-functionalized materials.

## Results

**Radical-functionalized surface**. We tested the ability of electric fields to control the orientation of peptides as they approach and bind to a radical-functionalized surface (Fig. 1). To do this, we used an RFPP coating. The coating was fabricated on titanium substrates using a unique plasma polymerization configuration, where the substrate is negatively biased in a pulsed manner during the deposition. Pulsed biasing of the substrate results in enhanced bombardment of the growing film by accelerated ions, thus allowing the generation of a high concentration of radicals within the structure of the coating. The EPR spectrum of the RFPP coating confirms that the surface is permeated by radicals (Fig. 2a). The EPR spectrum shows a single resonance peak, centred at 3513G (*g*-value of 2.003), ascribed to unpaired electrons associated with radical-containing compounds. The radicals are capable of forming covalent bonds with biomolecules, as previously demonstrated on similar plasma polymerized structures[42,43]. Plasma polymerization parameters were optimized to produce mechanically and chemically robust RFPP coatings capable of covalently binding peptide molecules directly from solution at a wide range of pH values.

Surface chemistry and charge are the most important characteristics influencing the interaction of surfaces with peptides[44,45]. The XPS survey spectrum of RFPP coating shows

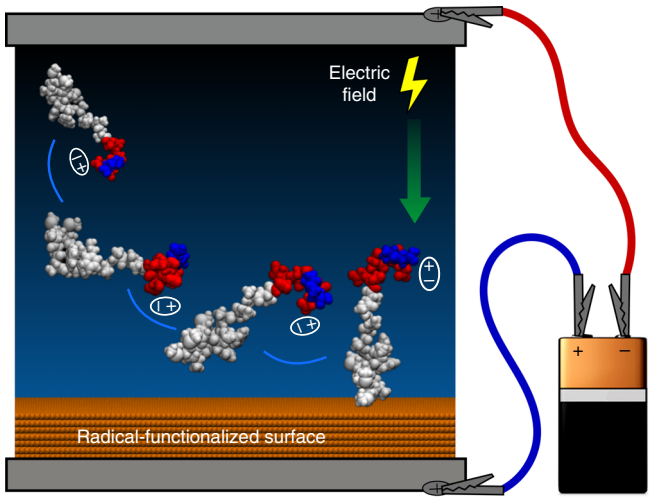



**Fig. 1** Control of peptide orientation by electric field. Charge separation on one end of the peptide creates a dipole moment (indicated by ellipses) that aligns with the electric field and rotates the entire molecule. Once contact is established with the radical-functionalized surface, covalent linkage anchors the peptide in this orientation

that the surface is composed of carbon, nitrogen and oxygen with atomic concentrations of 71.3, 19.6 and 9.1%, respectively (Fig. 2b). No titanium signal is detected, indicative of complete coverage of the underlying titanium surface by the RFPP layer with a thickness of 32.6 nm as measured by spectroscopic ellipsometry. The detected oxygen is primarily originating from post-deposition oxidation, an inevitable process that occurs upon the exposure of samples to atmospheric oxygen[46,47].

The high resolution C 1 s XPS spectrum was fitted with components associated with C–C/C–H at binding energy (BE) ≅ 284.6 eV, C–O/C–N at BE ≅ 286.5 eV, C=O/N–C=O at BE ≅ 287.5 eV and COOH at BE ≅ 289 eV (Fig. 2d)[48,49]. The four components indicate the broad range of chemical environments within the RFPP, in particular C=O and COOH moieties that become charged in aqueous environments. The presence of nitrogen and oxygen on the surface suggests that both basic and acidic groups are present, leading to the formation of charged moieties such as $NH_3^+$ and $COO^-$. The presence of these compounds allows modulation of the surface charge by altering the immobilization solution pH. The changes of surface zeta potential as a function of pH indicate that the isoelectric point (where the surface has no charge) is at pH = 4.5 (Fig. 2e). The surface becomes progressively more positive below this point and progressively more negative above it. This behaviour allows us to determine the effect of surface charge on interfacial interactions with a charged peptide that has a constant surface charge over this pH range.

**Peptide design and simulation**. The FLAG peptide was designed in such a way as to allow the evaluation of the influence of electric fields on the orientation and concentration of the immobilized peptide. The sequence (Ac-FFMMMAAAAAAAAAADDDDDK-$NH_2$) demonstrates several features that facilitate this evaluation: Methionine (M) residues introduce sulfur so that surface concentration can be measured by XPS. These residues are proximate to phenylalanine (F) residues that form a hydrophobic region (FFMMM) on one side of the peptide. On the other side of the peptide, there is a sequence based on the FLAG epitope[50] (DDDDK) capable of binding antibody molecules. This segment allows for the sensitive ELISA assay to assess how the surface immobilized peptide is oriented. This sequence also introduces both a net negative charge of −4 e and an electric dipole created by the aspartic acid (D) and lysine (K) residues. Crucially, due to the close proximity of the charges in the sequence, they cannot be separated and the dipole destroyed by structural changes in the peptide. Additionally, the peptide is unlikely to fold due to the different polarities of the hydrophobic and hydrophilic ends. The separation of these ends is supported by a bridge of alanines (A), which can exhibit a disordered structure. This arrangement allows the peptide to be linear, with definitive orientations with respect to the surface, one of which provides access to the charged epitope.

To predict the solution structure of the peptide, we performed an 800 ns equilibrium simulation in water. The secondary structure[51] shows a clear majority population consisting of a beta-turn through residues 3–7 connected by a beta-bridge between residues 2 and 10, while other residues are in random coil conformation (Fig. 3a). While there are minor populations, defined mainly by a beta-turn in one or two of residues 8–21, these do not correspond to folding of the peptide and do not change the accessibility of the FLAG epitope. This is demonstrated by the time series of the radius of gyration ($R_g$) (Fig. 3b). It is clear that after 40–50 ns, the peptide is equilibrated from its fully extended starting structure. The initial reduction in $R_g$ from full extension corresponds to folding into a beta-turn within the hydrophobic FFMMM- region. After this point, fluctuations to minor populations that include beta-turns do not significantly affect the peptide extension.

The charged region of the peptide creates a dipole moment that may affect the orientation of the peptide in an electric field. The vector angle between the dipole moment and the peptide principal, or 'long', axis fluctuates around 12.5 degrees, indicating close alignment between the dipole and the peptide direction of extension (Fig. 3c). This ensures that in the presence of an electric field, when alignment of the dipole is achieved the whole peptide will become parallel to the electric field lines.

The residues in the charged region, i.e. lysine (K) and aspartic acid (D), hold different charge states depending on the pH of the solution. Changing the pH alters the peptide net charge, which also alters the interaction with charged surfaces that controls the arrival rate and surface concentration (Fig. 3d). Changing the pH can thus remove or alter the strength of the dipole, affecting the peptide orientation at the surface. When the charged region is maximally charged, the peptide carries a net charge of −4 e. This situation occurs within the pH range of 5–10, where all residues are at least 90% charged. The peptide immobilization was therefore carried out at two pH values of 5.4 and 9.8, striking a balance between low and high RFPP surface charge, respectively, and optimal charging of the peptide.

A representative peptide structure chosen from the simulation trajectory demonstrates the peptide features (Fig. 3e). The beta-turns across the hydrophobic residues (F, M) lead to a folded section, likely formed in order to reduce water contact. The less-hydrophobic alanine (A), as per our design, forms a linear bridge between this region and the charged, polar FLAG epitope with an electric dipole aligned along the peptide principal axis. According to these simulation results, the designed sequence successfully constructs a linear, unstructured peptide that exposes the FLAG epitope to solution.

**Orientation and concentration control by pH**. XPS and ELISA can be used to detect the presence of surface protein or peptide. Here, we use XPS to measure the concentration of surface-bound peptide via the sulfur atomic concentration from methionine. The absence of sulfur signals in the XPS data of the RFPP coating before immobilization (<0.15 at. %, the XPS signal sensitivity) allows the use of this measurement to quantify the surface-bound

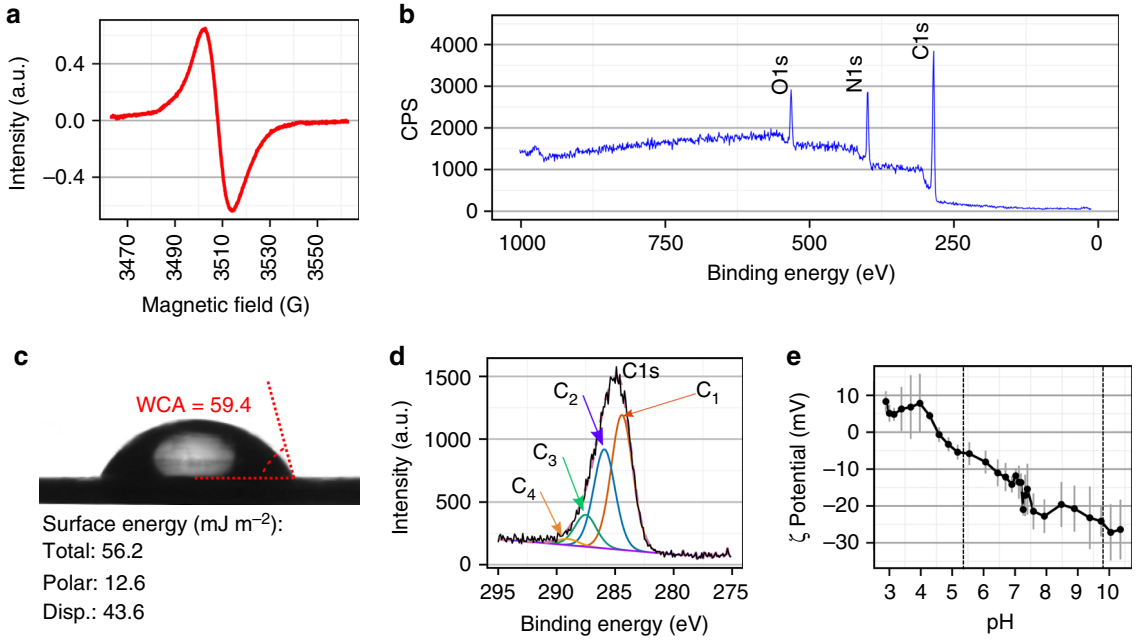



**Fig. 2** The surface has radicals and a range of charge states. **a** Electron paramagnetic resonance spectrum of the radical-functionalized plasma polymer (RFPP) showing a broad and symmetrical peak, indicative of unpaired electrons associated with radicals. **b** X-ray photoelectron spectroscopy (XPS) survey spectrum of the RFPP film showing a surface elemental composition of carbon (71.3 at.%), nitrogen (19.6 at.%) and oxygen (9.1 at.%). No sulfur or titanium is detected on the surface. **c** Water contact angle (WCA) of the RFPP surface. The total, polar and dispersive (Disp.) surface energies are given. **d** XPS C 1 s high-resolution spectra of the RFPP fitted with four components, $C_1$: C−C/C−H, $C_2$: C−O/C−N, $C_3$: C=O/N−C=O and $C_4$: COOH. **e** The changes of zeta potential as a function of pH show that the surface is negatively charged above pH 4.5, with the negative charge saturating at pH ≅ 7.5. The dashed lines indicate the pH values used in the immobilizing solutions. Error bars are s.d. (**e**)

peptide. We also use ELISA to indicate the presence of the functional FLAG epitope. For this experiment, the peptide was immobilized at pH 5.4, the lower of the two pH values used, because the surface charge is nearly neutral at this value.

The XPS and ELISA detection of surface-bound peptide increase with increasing concentration of peptide in the immobilization solution (Fig. 4a, b). The sulfur atomic concentration is undetectable below 5 μg mL$^{-1}$ (Fig. 4a), but the increase after this point indicates an increasing density of surface immobilized peptide. The ELISA assay indicates that the FLAG epitope is accessible by antibody in solution, and that this assay is concentration-dependent up to 20 μg mL$^{-1}$ (Fig. 4b). At concentrations higher than this, the signal plateaus and saturates (see Supplementary Fig. 1), most likely because steric hindrance, due to the large size (relative to the peptides) of the antibodies, prevents further antibody binding. This proof-of-concept experiment illustrates that the density of the immobilized peptide is measurable over a wide range of surface immobilized concentrations by a combination of XPS and ELISA. While the sensitivity of XPS is limited to immobilizing concentrations above 5 μg mL$^{-1}$, the ELISA assay is sensitive to peptide concentrations at the lowest levels tested, but the signal saturates at higher concentrations. The combined use of XPS and ELISA can, therefore, measure the presence of peptide on the surface at both high and low surface densities.

The ELISA assay should also be capable of differentiating the orientations of bound peptide. Here we assign $FLAG^{UP}$ to peptide bound through the hydrophobic end, which presents the FLAG epitope into solution, and $FLAG_{DOWN}$ to peptide bound via the FLAG epitope itself. It has been previously reported that surface-immobilized peptides can lie horizontally as opposed to normal to the surface[34,52,53], potentially concealing the FLAG epitope even in the $FLAG^{UP}$ orientation. This situation is, however, unlikely to happen for our study since the epitope is charged and has lower energy when fully hydrated as compared to that when contacting the RFPP surface. In addition, the presence of the alanine spacer between the hydrophobic end and the FLAG epitope should increase antibody binding in the $FLAG^{UP}$ orientation by facilitating more contact in solution, a phenomenon known as the spacer effect[25,34,54].

The solution pH was changed during the immobilization to explore the effect of surface charge on peptide attachment. Increases in pH increase the negative surface charge and demonstrate the effect of the associated electric field. Informed by the zeta potential measurements (Fig. 2e), we chose immobilizing solution pH values of 5.4 and 9.8 to compare peptide binding while remaining within the constraints of our buffer system. For this experiment, the peptide concentration during immobilization was chosen to allow detection by XPS, and hence was in the saturated region of concentrations for ELISA.

The XPS analysis of peptide immobilized at the pH 9.8 condition yielded a zero sulfur signal, while ~ 0.7% sulfur was detected for the sample prepared at pH = 5.4 (Fig. 4c). The peptide maintains a constant charge of −4 e across both pH conditions (Fig. 3d). It is, therefore, the surface charge only that controls the arrival rate and binding of the peptide to the RFPP film. The RFPP surface is several-fold more negatively charged at pH 9.8 compared to pH 5.4, and hence the surface gives rise to an electric field that repels negative charges. The RFPP surface at such conditions represents a potential barrier that results in a slower peptide arrival rate to the surface and consequently a lower degree of covalent binding over the incubation time. Even though the sulfur level is not detectable by XPS at pH 9.8, there may be peptide bound on the surface. For the pH = 5.4 condition, where sulfur signals are detected, the surface is nearly neutral and so does not have a steep potential barrier for peptide binding. This surface-binding behaviour is consistent with the trends observed for adhesion of proteins on charged surfaces[55–58].

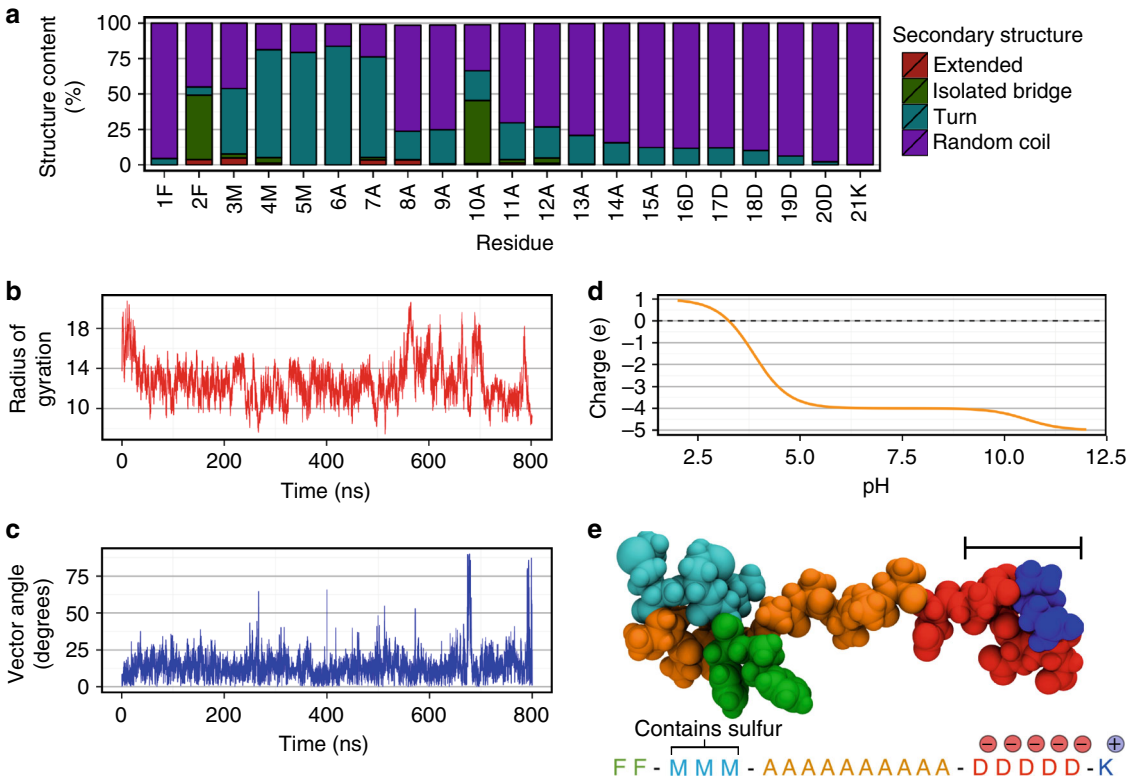



**Fig. 3** The simulated peptide is linear and mostly unstructured. **a** Stacked bar chart labelled with amino acid one-letter codes showing the per-residue structure content that is mostly random coil. **b** The time series of $R_g$ shows that after 40 ns of equilibration, the peptide maintains a single length. **c** Time series of the vector angle between the first principal axis and the dipole moment created by the charged residues. This shows that the dipole is closely aligned to the long axis of the peptide. **d** Net charge of the peptide across a range of pH values. All residues in the charged region are charged between approximately pH 5–10, leading to a net charge of −4 e. **e** Space-filling model of a representative structure of the FLAG-tag peptide, showing the turn across residues F2–A7, and the extended random coil structure of the polar, charged residues (D16–K21). Scale bar: 1 nm

The detection of sulfur even after Tween 20 or sodium dodecyl sulfate (SDS) washing indicates that the peptide is attached covalently to the surface. The slight loss of signal is likely due to the removal of additional adsorbed peptide. SDS and Tween 20 are detergents that disrupt physical interactions between adsorbed solutes and surfaces while leaving covalent bonds intact.

The ELISA results for the RFPP surface functionalized with peptide show an increase in signal over the control sample for both pH values of 5.4 and 9.8, indicating the presence of peptide on the surface (Fig. 4d). The high concentration used here is saturated for the pH 5.4 condition, and since the pH 9.8 condition shows the same absorbance signal, it is also likely to be saturated. Despite the utility of ELISA for detecting low concentrations of surface peptide, the size differential between the peptide and the antibody molecules means that once bound, antibody molecules block access to other peptide molecules underneath the antibody footprint. This indicates the assay saturates at concentrations far below a peptide monolayer. Such behaviour is an intrinsic limitation of two-dimensional ELISA, particularly for peptides, which have a much smaller footprint than antibody proteins and so saturate at low coverage densities.

Although XPS and ELISA indicated the surface concentration of peptide, they could provide only limited information about its orientation. The orientation of peptides has been deduced before using highly surface-sensitive techniques such as sum frequency generation spectroscopy[59] and time of flight secondary ion mass spectroscopy (ToF-SIMS)[60]. In this study, we used ToF-SIMS as a highly surface-sensitive technique with a sampling depth of 1–2 nm[61]. This technique enables an assessment of peptide orientation by revealing amino acids dominant in the top-most region of the peptide layer. ToF-SIMS is performed in vacuum; therefore, the conformation of the peptide may be different compared to that in the aqueous environment. Nevertheless, the covalent linkage will minimize the sputtering of amino acids from the tethered end. To assess the orientation, we used the average positive SIMS counts associated with methionine and phenylalanine, normalized to the total 42 protein-associated secondary ion fragments[62] (Fig. 4e). The other amino acid residues from the peptide (alanine, lysine and aspartic acid) are excluded, because they cannot be unambiguously discriminated from either fragments originating from the RFPP layer or from each other (See Supplementary Fig. 2). Both residues show greater counts for the peptide immobilized at pH 9.8 compared to that immobilized at pH 5.4. These two amino acids make up the hydrophobic end of the peptide (Fig. 3e), indicating that there is a greater number of molecules in the $FLAG_{DOWN}$ orientation for pH 9.8 than for pH 5.4. The difference in counts can be due to either a change in overall density of the peptide on the surface or in the preferred orientation of the immobilized peptides. XPS data, however, showed that the pH 5.4 condition yields a substantially higher overall peptide density. It can, therefore, be concluded that there is a strong preference for $FLAG_{DOWN}$ orientation for peptide immobilized at pH 9.8 relative to that immobilized at pH 5.4.

XPS, ELISA and ToF-SIMS results together show the effect of immobilization pH on both the surface concentration and the orientation of bound peptide. At pH 5.4, where the surface charge is close to neutral, the repulsion between the RFPP and the negatively charged peptide is negligible, leading to higher

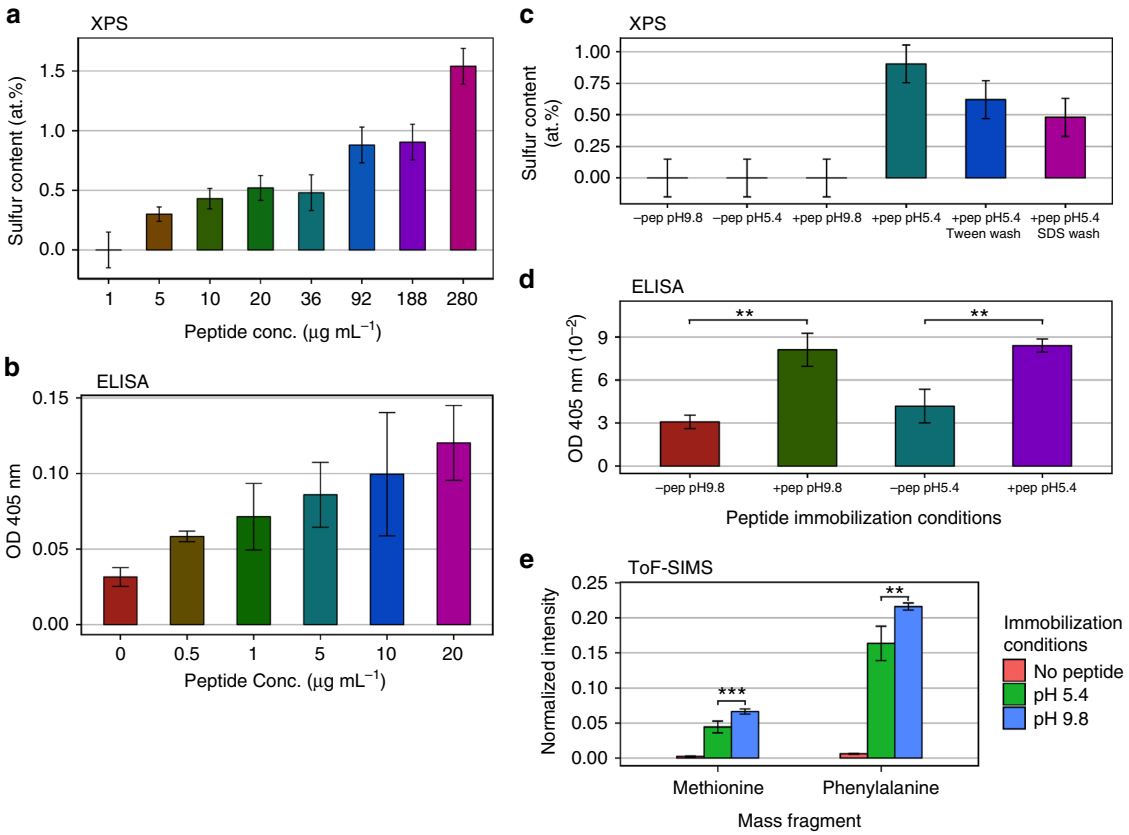



**Fig. 4** pH controls the peptide concentration and orientation. **a** Sulfur content, as measured by X-ray photoelectron spectroscopy (XPS), indicates increasing surface peptide with solution concentration but is undetectable for concentrations below 5 μg mL$^{-1}$. **b** Peptide titration shows that enzyme-linked immunosorbent assay (ELISA) is discriminatory for peptide concentrations down to at least 0.5 μg mL$^{-1}$ but saturates at approximately 20 μg mL$^{-1}$ due to the large footprint of the primary and secondary antibodies compared to the peptide. Hence, ELISA is not capable of discerning immobilized peptide surface density for high densities while XPS is not capable of detecting low peptide surface concentrations. **c** XPS sulfur atomic concentration of peptide-coated surfaces compared to uncoated controls for immobilization pH values of 5.4 and 9.8. The sulfur content shows surface peptide at pH 5.4, but no detectable peptide at pH 9.8. Presence of peptide after Tween 20 or sodium dodecyl sulfate (SDS) washing indicates covalent attachment. **d** The absorbance measured with ELISA for peptide-coated samples for immobilization pH values of 5.4 and 9.8 compared with uncoated controls. ELISA shows that peptide is present at both pH values. **e** Time of flight secondary ion mass spectrometry (ToF-SIMS) normalized mass fragments of the hydrophobic residues indicate higher number of $FLAG_{DOWN}$ peptide for pH 9.8, suggesting that a greater proportion of peptide for pH = 5.4 is in the $FLAG^{UP}$ orientation. Error bars are s.d. and *P*-values are from the Student's two-tailed *t*-test: $^{**}P < 0.01$, $^{***}P < 0.001$ (**b**, **d** and **e**). XPS error bars are calculated from the background noise (**a**, **c**)

functionalization density. Because of the negligible charge interaction with the surface, the orientation is most likely random. Given that the dispersive component of the RFPP surface energy is higher than the polar component, there may be a preference for $FLAG^{UP}$ orientation due to hydrophobic interactions between the surface and the hydrophobic end of the peptide. In contrast, at pH = 9.8 the RFPP is more negatively charged, repelling the peptide and reducing its density on the surface by reducing the arrival rate. Under such conditions, the peptide is also preferentially oriented on the surface in the $FLAG_{DOWN}$ configuration. This orientation is favoured, because it aligns the dipole in the FLAG epitope created by the C-terminal aspartate and lysine residues with the electric field at the surface. These findings demonstrate that pH can be utilized to control surface orientation and concentration of immobilized peptide. The mechanism for this control is the appearance of an electric field normal to the surface when the surface is charged. This effect would increase as the surface charge increases, because the electric field is proportional to the surface charge density. For extra control in situations where the pH is constrained, the application of an external electric field could be an efficient alternative.

**Orientation and concentration control by external E-field.** Surface charge created by changes in solution pH is often limited by the types of chemical groups present on the surface. In situations where linker chemistry is used to immobilize biomolecules, pH ranges are often restricted to those that facilitate the required chemistry. Here we show that an externally applied electric field will overcome these limitations, enabling greater control of orientation and concentration. Specifically, the RFPP surface in this experiment is not able to achieve a significant positive charge using solution pH and hence is unable to attract the negatively charged peptide. We applied external electric fields in both directions during the peptide immobilization using a custom-made well-plate and a simple power supply. We denote the electric field direction using $E^{up}$ and $E_{down}$, corresponding to the positive terminal connecting to the lower or upper plate, respectively. The negative terminal was connected to the opposite plate in each case. The $E^{up}$ field is expected to attract additional peptide molecules to the surface leading to higher surface density. We used initially a low peptide concentration (1 μg mL$^{-1}$) in the immobilizing solution to demonstrate the enhancing effect of electric fields on surface density, while a high peptide concentration (180 μg mL$^{-1}$) was used to increase the signal-to-noise

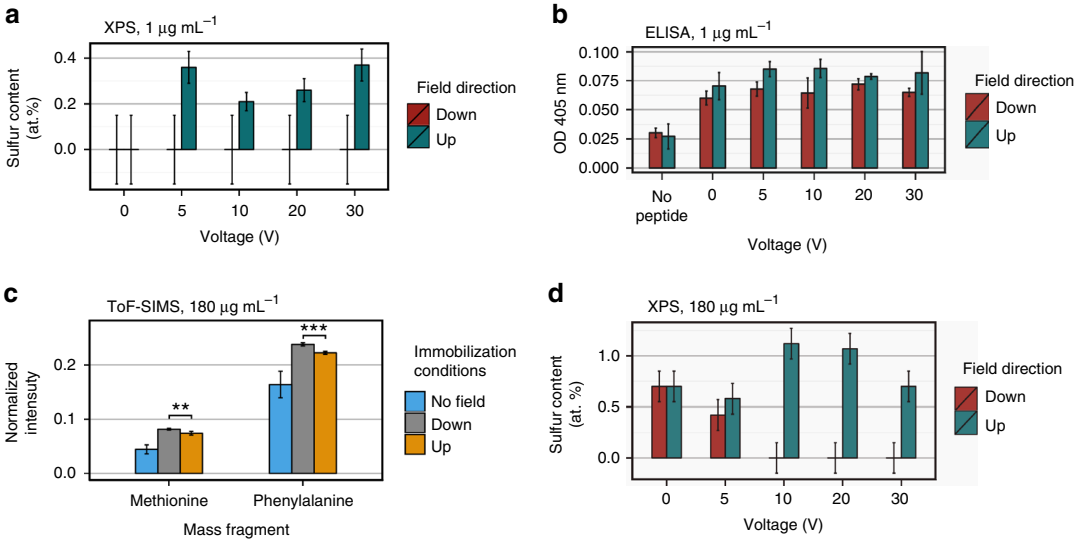



**Fig. 5** External field controls peptide density and orientation. **a** The sulfur atomic concentration of the peptide-functionalized surface from X-ray photoelectron spectroscopy (XPS) as a function of applied voltage (peptide solution concentration = 1 μg mL$^{-1}$). An increase in peptide arrival rate is observed for the $E^{up}$ field, while no peptide is detectable for the $E_{down}$ field. **b** The absorbance measured with enzyme-linked immunosorbent assay (ELISA) as a function of applied voltage showing that peptide is present after immobilization in both $E^{up}$ and $E_{down}$ field directions (peptide solution concentration = 1 μg mL$^{-1}$). **c** Time of flight secondary ion mass spectrometry (ToF-SIMS) normalized mass fragments of the hydrophobic peptide residues indicate approximately equal amounts of $FLAG_{DOWN}$ peptide (peptide solution concentration = 180 μg mL$^{-1}$). **d** The sulfur atomic concentration of the peptide-functionalized surface from XPS as a function of applied voltage (peptide solution concentration = 180 μg mL$^{-1}$). Surface peptide is increased by $E^{up}$, and decreased by $E_{down}$ relative to the case when no field is applied. Given the higher concentration for the $E^{up}$ field but equal amount of hydrophobic residues (indicating $FLAG_{DOWN}$ peptide) in ToF-SIMS, there is a higher proportion of $FLAG^{UP}$ peptide for this field direction compared to the $E_{down}$ direction. Error bars are s.d. and *P*-values are from the Student's two-tailed *t*-test: $^{**}P < 0.01$, $^{***}P < 0.001$ (**b**, **c**). XPS error bars are calculated from the background noise (**a**, **d**)

ratio for the evaluation of orientation by XPS and ToF-SIMS. In all of the following experiments, the pH of the immobilizing solution was kept at 5.4 to reduce the influence of charged surface moieties.

We used XPS, ELISA and ToF-SIMS to characterize surfaces prepared with an electric field applied either in the $E^{up}$ or $E_{down}$ directions (Fig. 5). It is observed that the low peptide concentration in the immobilizing solution leads to a surface concentration below the detection limit of XPS for the $E_{down}$ field, while for $E^{up}$ the peptide is detectable through XPS sulfur signals reaching to 0.2–0.35 at% (Fig. 5a). This level of functionalization is comparable to that achieved using solution concentrations 5- to 10-fold greater in the absence of an electric field (see Fig. 4a). Such improvement of peptide surface density suggests a great potential to use an external electric field during immobilization to increase the yield of bound peptide, which is particularly important for surface bio-functionalization using expensive molecules.

There is an ELISA signal above the no-peptide control at each voltage and field direction applied, indicating that peptide is present in all cases, even for $E_{down}$ where peptide was not detected by XPS (Fig. 5b). Unlike the XPS signal, the ELISA absorbance does not significantly vary across the range of applied voltages, indicating that the signal is saturated. The difference between the saturated values for the $E^{up}$ and $E_{down}$ conditions is statistically significant ($P < 0.001$). This possibly arises due to suboptimal packing of antibody on the lower density surface ($E_{down}$) caused by greater gaps without peptide on the surface where antibodies cannot bind. On the $E^{up}$ surface, the higher density of peptide facilitates closer packing of the larger antibodies and therefore a slightly higher saturated signal.

To investigate differences in orientation, we employed XPS and ToF-SIMS. As discussed before, only the phenylalanine and methionine amino acid residues from the hydrophobic region of the peptide are analysed to infer orientation by ToF-SIMS data. Both electric field directions show higher concentrations of methionine and phenylalanine on the topmost layer compared to the sample prepared in the absence of electric fields (Fig. 5c). This is indicative of greater concentration of $FLAG_{down}$ oriented peptide in both cases. The two field directions show the same relative amounts, with a slightly higher signal for each residue for $E_{down}$ than $E^{up}$.

ToF-SIMS results can be better understood by taking into account the XPS results for samples prepared with and without the application of external electric fields (Fig. 5d). The sulfur atomic concentration is increased for the $E^{up}$ field and decreased for the $E_{down}$ field, indicating that the peptide concentration for the $E_{down}$ field reduces as the electrical potential is increased. Similar to the pH 9.8 condition, the $E_{down}$ field creates a potential barrier for the negatively charged peptide as it approaches the surface and hence repels the peptide molecules. In contrast, when the field is in the upward direction ($E^{up}$), a potential well for negatively charged peptide is created at the surface, attracting greater numbers of peptide molecules to the surface. This situation did not arise in the pH comparison experiment since we were limited to a negative surface charge in order to have a fully charged peptide. The application of the $E^{up}$ field thus demonstrates the extra control possible over functionalization. This is particularly useful given the importance of optimizing the surface concentration of bioactive peptides[24,26,35,63,64].

Using the information about the density of surface peptide in each case, as well as the concentration of $FLAG_{DOWN}$-oriented peptide molecules given by ToF-SIMS, we can now infer the concentration of $FLAG^{UP}$-oriented peptide. According to XPS results, a higher density of surface peptide is achieved by applying the $E^{up}$ field compared to the $E^{down}$ configuration. ToF-SIMS

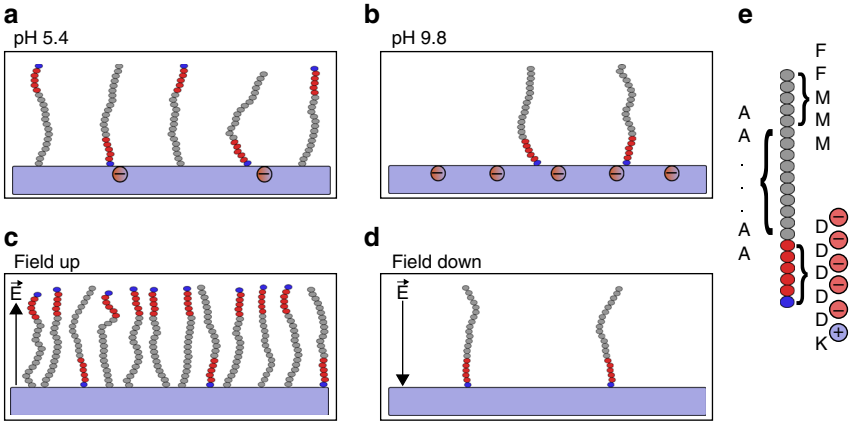



**Fig. 6** Illustration of peptide orientation and concentration. The immobilization conditions were **a** pH 5.4, **b** pH 9.8, and with applied electric fields **c** $E^{up}$ and **d** $E_{down}$ both at pH 5.4. The radical-functionalized plasma polymer surface is indicated by purple rectangles, with charges on both the surfaces and the peptide indicated by circled + or − signs. **e** The enlarged peptide indicates the position of amino acids (lysine (K) blue, aspartic acid (D) red, and all others shown in grey)

data, however, suggested that the concentration of $FLAG_{DOWN}$ molecules are approximately equal in both cases. It can, therefore, be concluded that the majority of peptide molecules immobilized using the $E^{up}$ field are in a $FLAG^{UP}$ orientation. In contrast, for the $E_{down}$ field there is mostly $FLAG_{DOWN}$ peptide. This distinction between the surfaces balances out such that the total presentation of hydrophobic residues is approximately equal, as shown by ToF-SIMS. We have therefore demonstrated that stark differences in orientation and concentration of peptide molecules are achieved through the application of external electric fields during the immobilization process. Importantly, unlike other recent studies where orientation is used to optimize peptide activity[28,31,34,52,53,65], such differences are achieved without adding any extra amino acids for covalent linkage or changing the immobilizing solution.

## Discussion

Here we discuss our findings for the variations of concentration and orientation as a function of electric field changes induced by pH or potential differences applied across the immobilizing solution. For controlling the peptide functionalization by modulating the immobilizing solution pH, higher surface densities were achieved when the surface was less charged. Since the surface charge is small, and so has a negligible electrostatic effect on the peptide, the arrival rate and subsequent covalent linkage are mediated by the hydrophobic effect and van der Waals interactions only. In this case, the peptide binds in a random orientation (Fig. 6a). When the surface charge is more negative at higher pH values (Fig. 6b), the total peptide density is lower due to the potential barrier created by the increased negative surface charge. However, when peptide molecules do bind to the surface, they are oriented $FLAG_{DOWN}$ in order to align their dipole moment with the electric field created by the charged surface. An externally applied electric field is also capable of influencing the peptide-surface interactions. For the $E_{down}$ field (Fig. 6d), the negative bias on the surface repels peptide molecules, reducing the density of peptides immobilized on the surface and simultaneously orienting them $FLAG_{DOWN}$ (as for immobilization at pH 9.8). For the $E^{up}$ field (Fig. 6c), which was not achievable by pH modulation, the surface is positively biased and thus more peptide molecules are attracted, leading to a high immobilized surface density. Additionally, the field orients peptides in the $FLAG^{UP}$ orientation. For surface functionalization, this represents the best-case scenario where high concentration of bound peptide as well as an increased proportion of the correct orientation are achieved.

We believe that our findings on the mechanisms of peptide immobilization controlled by electric fields have important implications for the design of synthetic biomolecules and biofunctionalization of advanced implantable materials. In particular, guidelines for the design of synthetic peptides to optimize their presentation and density when immobilized can be inferred. For a particular functional peptide sequence and radical-functionalized surface, the amino acid charge sequence and surface potential as a function of pH must be determined so that an optimal pH for immobilization can be identified. If there is no pH at which the peptide net charge and dipole moment are favourable, then additional amino acids can be appended to the peptide end that is to be bound at the surface so as to adjust both the dipole moment and the charge. Alternatively, electric fields as shown in this work can be applied to achieve the optimal orientation and density. Our findings suggest that tuning the immobilization solution pH and/or the application of electric fields during immobilization have the potential to improve biofunctionalization methods, offering better outcomes for implantable devices used in modern biomedicine.

## Methods

**Materials**. Titanium substrates (thickness = 0.07 mm) were obtained from Firmetal, China, and were ultrasonicated in acetone and ethanol, rinsed with Milli-Q water, and dried using a nitrogen gas stream prior to RFPP film deposition. High purity argon, acetylene and nitrogen gases were supplied by BOC, Australia. Buffer reagents, Tween 20 and SDS were purchased from Sigma Aldrich. Goat primary polyclonal anti-DDDDK antibody and rabbit anti-goat secondary antibody were obtained from Abcam, Australia. The designed peptide (Ac-FFMMMAAAAAAAAAADDDDDK-$NH_2$) was purchased from Auspep at 95% purity.

**Plasma polymerization**. The deposition of a thin RFPP layer on titanium substrates was performed using a custom-made plasma polymerization system, described elsewhere in detail[46,66]. A precursor gas mixture of argon, acetylene and nitrogen was injected into the chamber at flow rates of 15, 5 and 10 standard cubic centimetres per minute (sccm), respectively. The system working pressure was kept constant at 110 mTorr, while the base pressure was below $5\times10^{-2}$ mTorr. Plasma polymerization was performed for 2 min at an RF input power of 50 W provided by an ENI radio frequency power generator (13.56 MHz). Voltage pulses of −500 V (pulse duration = 20 µs, frequency = 3 kHz) were applied to the substrate holder using an RUP-6 pulse generator (GBS Elektronik GmbH). Titanium samples were cleaned by argon plasma prior to plasma polymerization as the final cleaning step. To achieve a stable surface chemistry, the PP-coated samples were stored in the laboratory environment for 7 days prior to incubation with the peptide.

**X-ray photoelectron spectroscopy (XPS)**. Surface chemistry of PP-coated surfaces before and after peptide immobilization was analysed using a SPECS FlexMod spectrometer. The instrument was equipped with a monochromatic Al $K_\alpha$ ($h\vartheta$ =

1486.7 eV) radiation source operating at 200 W. A PHOIBOS 150 hemispherical analyser and an MCD9 electron detector were used for spectroscopy. The electron take-off angle was 90° referenced to the sample surface, and the base pressure was always maintained below $5\times10^{-8}$ mbar. Survey spectra were recorded in an energy range of 0–1000 eV with a resolution of 0.5 eV (pass energy = 30 eV). High resolution C 1s, S 2p, N 1s, and O 1s spectra were collected at a resolution of 0.1 eV and a pass energy of 20 eV. The high-resolution spectra were used for surface chemical composition calculations. Processing and atomic concentration calculations were performed using CasaXPS software (version 2.3.1).

**Contact angle measurements**. The contact angles of the RFPP surface were measured using a Krüss DSA Mk2 goniometer. Eight measurements were taken for both water and diiodomethane using the sessile drop method by applying 1 µL drops. The surface energy was calculated using the Owens–Wendt model[67], which relates the contact angle and polar and dispersive components of the liquids to the surface energy.

**Electron paramagnetic resonance (EPR) spectroscopy**. EPR spectroscopy was conducted using a Bruker EMXplus Xband to evaluate the radical functionalization of the RFPP coating. RFPP-coated polystyrene films (7 cm×7 cm) were rolled into Wilmad Borosilicate glass NMR tubes, and the spectrometer was calibrated using a weak pitch sample. Spectra were recorded with a central magnetic field of 3510 G, modulation amplitude of 3 G, microwave frequency of 9.8 GHz and power of 25 mW. The field modulation frequency was $1\times10^{5}$ and the sampling time was 85 ms. Ten scans were averaged per sample.

**Electro kinetic analysis**. The zeta potential of the RFPP surface was measured as a function of pH using an Anton Paar SurPASS Electrokinetic Analyzer with an automatic titration unit to adjust the pH between 3 and 10. Two 10×20 mm samples of RFPP were fixed to an adjustable cell sample holder with a gap of approximately 100 µm between the samples. The measurement solution of $1\times10^{-3}$ M KCl was adjusted to an initial pH of 10 using appropriate volumes of 1 M NaOH. At each pH value, this solution was pumped through the sample holder, in both directions, with a pressure ramp to measure the streaming current. Zeta potential was then calculated using the Helmholtz-Smoluchowski equation. Four measurements were taken at each pH, before adjustment by aliquots of HCl by the automatic titration unit. The reported zeta potential values are averages at each pH value.

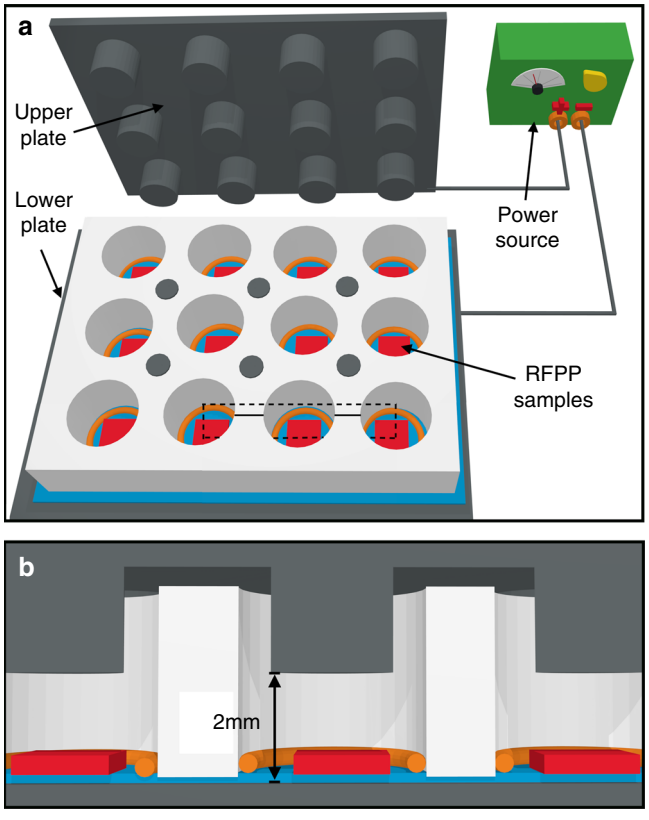



**Fig. 7** The sample holder for applying external electric fields. **a** The radical-functionalized plasma polymer-coated samples (red) lie in wells in a polyether ether ketone sample holder with 12 wells (white). Rubber O-rings (orange) seal each well, which are insulated by a 0.1 mm polytetrafluoroethylene sheet (blue). The upper plate fits on top of the sample holder, and electrical potential is applied across the wells using a variable power source. Dotted lines indicate the cross-section in (**b**). **b** Cross-section of three wells. The extensions from the top plate lower into each well, leaving a 2-mm gap between the bottom and top plates. The sample is raised 0.1 mm above this level by the insulating material

**Peptide simulation**. Two molecular dynamics equilibrium simulations were run to assess the secondary-structural behaviour of the FLAG peptide, each using different starting structures. The peptide starting coordinates were generated using Avogadro[68] software as either fully linear or alpha-helix. These structures were solvated with TIP3P water using the VMD Solvate plugin[69], with box sizes 14 Angstrom larger than the peptide's longest axis in the *x*, *y* and *z* directions using periodic boundary conditions. This ensured the peptides did not interact with their periodic images in an adjacent box. Each system had 150 mM NaCl added. The systems were minimized for 4000 steps and the water equilibrated for 1 ns by restraining the alpha-carbon atoms.

The alpha helical peptide equilibrium simulation was run for 300 ns. The alpha helix conformation was unstable and by this point had unfolded. This simulation was stopped since the peptide was longer and could interact with periodic copies of itself. The linear peptide simulation was run for 800 ns. Both simulations used a constant pressure of 1 atm and temperature of 298 K maintained using a Langevin Piston and Langevin thermostat respectively. The CHARMM27 force field[70] was used for the protein parameters, and the simulations, using a 2 fs timestep, were run using NAMD2.91.

**Peptide immobilization**. The buffer used for immobilization was designed to have equal ionic strength and buffer capacity at a low and a high pH value. A mixed buffer system was therefore used, with two buffers with different pKa values, allowing similar buffering capacity at two pH values. The buffer system consisted of an acetate/acetic acid buffer (pKa 4.8) and an ammonia/ammonium buffer (pKa 9.2) in the same solution. The pH was increased by 0.6 from the pKa to ensure that the peptide would fully dissolve at the lower pH value while maintaining its maximum charge. At pH = 5.4 the acetate/acetic acid buffer predominates, whereas at pH = 9.8 the ammonia/ammonium buffer predominates, while the buffering capacity is equal in both cases. The buffers were prepared by adjusting the pH of an ammonium acetate solution using either concentrated HCl or NaOH and adding NaCl to equalize the ionic strength.

The FLAG peptide was dissolved in 10 mM buffer at either pH 5.4 or pH 9.8. The incubation on the plasma polymer surface was performed by pipetting 500 µL of peptide solution onto a 1 cm×1 cm sample in a 24-well plate, when no external electric field was being applied. For samples with applied external electric field, the incubation was performed using a custom-made apparatus, schematically shown in Fig. 7. This apparatus consisted of wells of the same dimensions as the 24-well plate, but with electrodes above and below the sample. Each sample was incubated for 20 h at room temperature with or without an electric field applied. A polytetrafluoroethylene foil was placed over the bottom electrode, which ensured that no current was drawn. After incubation, the peptide solution was removed, and the sample was washed three times using 1 mL de-ionized water. To validate the covalent attachment of the peptide, the detergent-washed samples were placed in a 5% Tween 20 or 5% SDS solution heated to 70 °C for 1 h, then washed three times using 1 mL de-ionized water.

**Time of flight secondary ion mass spectrometry (ToF-SIMS)**. To evaluate the orientation of the immobilized peptide on the RFPP-coated surfaces, ToF-SIMS measurements were performed using a Physical Electronics Inc. PHI TRIFT V nanoTOF instrument. A pulsed primary $^{79}Au^{+}$ ion beam (30 keV) was used for the ionization of species. Dual charge neutralization was achieved using a PHI system by applying a combination of low energy argon ions (10 eV) and electrons (25 eV). The base pressure was $5\times10^{-4}$ Pa or less during the measurements. The spectra were collected in bunched instrument settings to optimize mass resolution. Positive SIMS data were acquired over areas of 100 µm×100 µm for an acquisition time of 60 s. For each sample, six spectra were recorded at different locations to assess repeatability. Species associated with protein secondary ion fragments[62] were selected, and the counts were normalized to the total intensity of all selected peaks. Acquired data were processed and integrated using WincadenceN software (Physical Electronics Inc. Chanhassen, MN, USA). Error bars for ToF-SIMS represent the 95% confidence interval calculated from the six recorded spectra.

**Enzyme-linked immunosorbent assay (ELISA)**. ELISA was performed to detect bound peptide on the RFPP surface. The peptide-functionalized samples were blocked with 3% BSA solution. Peptide functionalization was detected using the goat primary polyclonal anti-DDDDK antibody ab1257, followed by a rabbit anti-goat secondary antibody ab6741. The samples were added to ABTS solution and absorbance was measured at 405 nm after 1 h incubation. Statistical significance was calculated using single factor analysis of variance.

**Data availability**. The authors declare that all data supporting the findings of this study are available within the article and its supplementary information files or from the corresponding authors upon request.

Received: 4 August 2017 Accepted: 7 December 2017
Published online: 24 January 2018

## References

1. Langer, R. & Tirrell, D. A. Designing materials for biology and medicine. *Nature* **428**, 487–492 (2004).
2. Cobelli, N., Scharf, B., Crisi, G. M., Hardin, J. & Santambrogio, L. Mediators of the inflammatory response to joint replacement devices. *Nat. Rev. Rheumatol.* **7**, 600–608 (2011).
3. Anderson, J. M., Rodriguez, A. & Chang, D. T. Foreign body reaction to biomaterials. *Semin. Immunol.* **20**, 86–100 (2008).
4. Dias, J., Kay, P., Porter, M. & Briggs, T. *Restoring Your Mobility*. Available at http://www.boa.ac.uk/wp-content/uploads/2014/01/BOA-Practise-Strategy.pdf (British Orthopaedic Association, 2012).
5. Wise, S. G. et al. Plasma ion activated expanded polytetrafluoroethylene vascular grafts with a covalently immobilized recombinant human tropoelastin coating reducing neointimal hyperplasia. *ACS Biomater. Sci. Eng.* **2**, 1286–1297 (2016).
6. Place, E. S., Evans, N. D. & Stevens, M. M. Complexity in biomaterials for tissue engineering. *Nat. Mater.* **8**, 457–470 (2009).
7. Wei, Q. et al. Protein interactions with polymer coatings and biomaterials. *Angew. Chem. Int. Ed.* **53**, 8004–8031 (2014).
8. Vroman, L. Effect of adsorbed proteins on the wettability of hydrophilic and hydrophobic solids. *Nature* **196**, 476–477 (1962).
9. Hirsh, S. L. et al. The Vroman effect: competitive protein exchange with dynamic multilayer protein aggregates. *Colloids Surf. B. Biointerfaces* **103**, 395–404 (2013).
10. Mao, A. S. & Mooney, D. J. Regenerative medicine: current therapies and future directions. *Proc. Natl. Acad. Sci. USA* **112**, 14452–14459 (2015).
11. Bilek, M. M. M. Biofunctionalization of surfaces by energetic ion implantation: review of progress on applications in implantable biomedical devices and antibody microarrays. *Appl. Surf. Sci.* **310**, 3–10 (2014).
12. Bilek, M. M. M. et al. Free radical functionalization of surfaces to prevent adverse responses to biomedical devices. *Proc. Natl. Acad. Sci. USA* **108**, 14405–14410 (2011).
13. Waterhouse, A. et al. In vivo biocompatibility of a plasma-activated, coronary stent coating. *Biomaterials* **33**, 7984–7992 (2012).
14. Waterhouse, A. et al. The immobilization of recombinant human tropoelastin on metals using a plasma-activated coating to improve the biocompatibility of coronary stents. *Biomaterials* **31**, 8332–8340 (2010).
15. Bax, D. V., McKenzie, D. R., Weiss, A. S. & Bilek, M. M. The linker-free covalent attachment of collagen to plasma immersion ion implantation treated polytetrafluoroethylene and subsequent cell-binding activity. *Biomaterials* **31**, 2526–2534 (2010).
16. Bax, D. V. et al. Binding of the cell adhesive protein tropoelastin to PTFE through plasma immersion ion implantation treatment. *Biomaterials* **32**, 5100–5111 (2011).
17. Wakelin, E. A., Davies, M. J., Bilek, M. M. M. & McKenzie, D. R. Temperature activated diffusion of radicals through ion implanted polymers. *ACS Appl. Mater. Interfaces* **7**, 26340–26345 (2015).
18. Kondyurin, A., Naseri, P., Fisher, K., McKenzie, D. R. & Bilek, M. M. M. Mechanisms for surface energy changes observed in plasma immersion ion implanted polyethylene: the roles of free radicals and oxygen-containing groups. *Polym. Degrad. Stab.* **94**, 638–646 (2009).
19. R. W. Colman. *Hemostasis and Thrombosis: Basic Principles and Clinical Practice* (Lippincott Williams & Wilkins, Philadelphia, PA, 2006).
20. Collier, J. H. & Segura, T. Evolving the use of peptides as components of biomaterials. *Biomaterials* **32**, 4198–4204 (2011).
21. Vagner, J., Qu, H. & Hruby, V. J. Peptidomimetics, a synthetic tool of drug discovery. *Curr. Opin. Chem. Biol.* **12**, 292–296 (2008).
22. Lutolf, M. & Hubbell, J. Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering. *Nat. Biotechnol.* **23**, 47–55 (2005).
23. Huang, A. H. & Niklason, L. E. Engineering of arteries in vitro. *Cell. Mol. Life. Sci.* **71**, 2103–2118 (2014).
24. Frith, J. E., Mills, R. J. & Cooper-White, J. J. Lateral spacing of adhesion peptides influences human mesenchymal stem cell behaviour. *J. Cell. Sci.* **125**, 317–327 (2012).
25. Kong, L. et al. RGD peptide-modified dendrimer-entrapped gold nanoparticles enable highly efficient and specific gene delivery to stem cells. *ACS Appl. Mater. Interfaces* **7**, 4833–4843 (2015).
26. Deng, Y. et al. Long-term self-renewal of human pluripotent stem cells on peptide-decorated poly (OEGMA-co-HEMA) brushes under fully defined conditions. *Acta Biomater.* **9**, 8840–8850 (2013).
27. Wang, P.-Y., Thissen, H. & Kingshott, P. Modulation of human multipotent and pluripotent stem cells using surface nanotopographies and surface-immobilised bioactive signals: a review. *Acta Biomater.* **45**, 31–59 (2016).
28. Shen, L., Ulrich, N. W., Mello, C. M. & Chen, Z. Determination of conformation and orientation of immobilized peptides and proteins at buried interfaces. *Chem. Phys. Lett.* **619**, 247–255 (2015).
29. Ding, B., Jasensky, J,. Li, Y,. & ChenZ. Engineering and characterization of peptides and proteins at surfaces and interfaces: a case study in surface-sensitive vibrational spectroscopy. *Acc. Chem. Res.* **49**, 1149–1157 (2016).
30. Stevens, M. M. & George, J. H. Exploring and engineering the cell surface interface. *Science* **310**, 1135–1138 (2005).
31. Ogorzalek, T. L. et al. Molecular-level insights into orientation-dependent changes in the thermal stability of enzymes covalently immobilized on surfaces. *Langmuir* **31**, 6145–6153 (2015).
32. Pesen, D. & Haviland, D. B. Modulation of cell adhesion complexes by surface protein patterns. *ACS Appl. Mater. Interfaces* **1**, 543–548 (2009).
33. Chollet, C. et al. The effect of RGD density on osteoblast and endothelial cell behavior on RGD-grafted polyethylene terephthalate surfaces. *Biomaterials* **30**, 711–720 (2009).
34. Han, X. et al. Different interfacial behaviors of peptides chemically immobilized on surfaces with different linker lengths and via different termini. *J. Phys. Chem. B* **118**, 2904–2912 (2014).
35. White, S. J. et al. The influence of two-dimensional organization on peptide conformation. *Angew. Chem.* **127**, 988–992 (2015).
36. Price, J. V. et al. 'On silico' peptide microarrays for high-resolution mapping of antibody epitopes and diverse protein-protein interactions. *Nat. Med.* **18**, 1434–1440 (2012).
37. Madl, C. M., Mehta, M., Duda, G. N., Heilshorn, S. C. & Mooney, D. J. Presentation of BMP-2 mimicking peptides in 3D hydrogels directs cell fate commitment in osteoblasts and mesenchymal stem cells. *Biomacromolecules* **15**, 445–455 (2014).
38. Emaminejad, S. et al. Tunable control of antibody immobilization using electric field. *Proc. Natl. Acad. Sci. USA* **112**, 1995–1999 (2015).
39. Lapizco-Encinas, B. H. & Rito-Palomares, M. Dielectrophoresis for the manipulation of nanobioparticles. *Electrophoresis* **28**, 4521–4538 (2007).
40. Takashima, S. Dielectric dispersion of protein solutions in viscous solvent. *J. Polym. Sci. Part Polym. Chem.* **56**, 257–265 (1962).
41. Washizu, M., Suzuki, S., Kurosawa, O., Nishizaka, T. & Shinohara, T. Molecular dielectrophoresis of biopolymers. *IEEE Trans. Ind. Appl.* **30**, 835–843 (1994).
42. Yin, Y. et al. Comparison on protein adsorption properties of diamond-like carbon and nitrogen-containing plasma polymer surfaces. *Thin Solid Films* **520**, 3021–3025 (2012).
43. Yin, Y. et al. Covalent immobilisation of tropoelastin on a plasma deposited interface for enhancement of endothelialisation on metal surfaces. *Biomaterials* **30**, 1675–1681 (2009).
44. Chen, S., Li, L., Zhao, C. & Zheng, J. Surface hydration: principles and applications toward low-fouling/nonfouling biomaterials. *Polymer (Guildford)* **51**, 5283–5293 (2010).
45. Li, L., Chen, S. & Jiang, S. Protein adsorption on alkanethiolate self-assembled monolayers: nanoscale surface structural and chemical effects. *Langmuir* **19**, 2974–2982 (2003).
46. Akhavan, B., Wise, S. G. & Bilek, M. M. Substrate-regulated growth of plasma-polymerized films on carbide-forming metals. *Langmuir* **32**, 10835–10843 (2016).
47. Akhavan, B., Jarvis, K. & Majewski, P. Evolution of hydrophobicity in plasma polymerised 1, 7-octadiene films. *Plasma Process. Polym.* **10**, 1018–1029 (2013).
48. Jarvis, K. L. & Majewski, P. Plasma polymerized allylamine coated quartz particles for humic acid removal. *J. Colloid Interface Sci.* **380**, 150–158 (2012).
49. Nisol, B. & Reniers, F. Challenges in the characterization of plasma polymers using XPS. *J. Electron Spectrosc. Relat. Phenom.* **200**, 311–331 (2015).
50. Einhauer, A. & Jungbauer, A. The FLAG$^{TM}$ peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods* **49**, 455–465 (2001).
51. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Bioinform.* **23**, 566–579 (1995).
52. Li, Y. et al. Effects of peptide immobilization sites on the structure and activity of surface-tethered antimicrobial peptides. *J. Phys. Chem. C* **119**, 7146–7155 (2015).
53. Wang, Z., Han, X., He, N., Chen, Z. & Brooks, C. L. III Molecular structures of C-and N-terminus cysteine modified cecropin P1 chemically immobilized onto maleimide-terminated self-assembled monolayers investigated by molecular dynamics simulation. *J. Phys. Chem. B* **118**, 5670–5680 (2014).
54. Nowinski, A. K., Sun, F., White, A. D., Keefe, A. J. & Jiang, S. Sequence, structure, and function of peptide self-assembled monolayers. *J. Am. Chem. Soc.* **134**, 6000–6005 (2012).

55. Höök, F., Rodahl, M., Kasemo, B. & Brzezinski, P. Structural changes in hemoglobin during adsorption to solid surfaces: effects of pH, ionic strength, and ligand binding. *Proc. Natl. Acad. Sci. USA* **95**, 12271–12276 (1998).
56. Demanèche, S., Chapel, J.-P., Monrozier, L. J. & Quiquampoix, H. Dissimilar pH-dependent adsorption features of bovine serum albumin and α-chymotrypsin on mica probed by AFM. *Colloids Surf. B. Biointerfaces* **70**, 226–231 (2009).
57. Bremer, M. G., Duval, J., Norde, W. & Lyklema, J. Electrostatic interactions between immunoglobulin (IgG) molecules and a charged sorbent. *Colloids Surf. Physicochem. Eng. Asp.* **250**, 29–42 (2004).
58. Rabe, M., Verdes, D. & Seeger, S. Understanding protein adsorption phenomena at solid surfaces. *Adv. Colloid Interface Sci.* **162**, 87–106 (2011).
59. Weidner, T., Breen, N. F., Li, K., Drobny, G. P. & Castner, D. G. Sum frequency generation and solid-state NMR study of the structure, orientation, and dynamics of polystyrene-adsorbed peptides. *Proc. Natl. Acad. Sci. USA* **107**, 13288–13293 (2010).
60. Chen, R. et al. Characterization of chemoselective surface attachment of the cationic peptide melimine and its effects on antimicrobial activity. *Acta Biomater.* **8**, 4371–4379 (2012).
61. Vickerman, J. C. & Briggs, D. *ToF-SIMS: Surface Analysis by Mass Spectrometry* 168 (IM Publications and SurfaceSpectra Limited, Chichester, UK, 2001).
62. Henry, M. & Bertrand, P. Surface composition of insulin and albumin adsorbed on polymer substrates as revealed by multivariate analysis of ToF-SIMS data. *Surf. Interface Anal.* **41**, 105–113 (2009).
63. Chen, L.-Y. et al. Effect of the surface density of nanosegments immobilized on culture dishes on ex vivo expansion of hematopoietic stem and progenitor cells from umbilical cord blood. *Acta Biomater.* **8**, 1749–1758 (2012).
64. Kuo, Y.-C. & Chung, C.-Y. TATVHL peptide-grafted alginate/poly (γ-glutamic acid) scaffolds with inverted colloidal crystal topology for neuronal differentiation of iPS cells. *Biomaterials* **33**, 8955–8966 (2012).
65. North, S. H., So, C., Fears, K. & Taitt, C. R. in *SPIE Sensing Technology +Applications* 911213 (eds Southern, Š. O., Mentzer, M. A., Rodriguez-Chavez, I. & Wotring, V. E.) (International Society for Optics and Photonics, Washington, 2014).
66. Akhavan, B. & Bilek, M. Controlled deposition of plasma activated coatings on zirconium substrates. In *SPIE Micro+Nano Materials, Devices, and Applications* 96685V (International Society for Optics and Photonics, Washington, 2015).
67. Owens, D. K. & Wendt, R. Estimation of the surface free energy of polymers. *J. Appl. Polym. Sci.* **13**, 1741–1747 (1969).
68. Hanwell, M. D. et al. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Chemin.* **4**, 1 (2012).
69. Humphrey, W., Dalke, A. & Schulten, K. VMD—visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
70. MacKerell A. D., Feig, M. & Brooks, C. L. 2004 Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations *J. Comput. Chem.* **25**, 1400–1415 (2004).

## Acknowledgements

We acknowledge the Australian Research Council for funding this research and the National Computational Infrastructure (NCI), supported by the Australian Government, for the computational resources used for our peptide simulations. L.J.M. acknowledges scholarship top-up funding from the CRC for Cell Therapy Manufacturing. Many thanks to Edgar Wakelin for design of the sample holder. We thank Dr. Clare Hawkins of the Heart Research Institute for access to and assistance with EPR and Professor Anthony Weiss and Dr. Giselle Yeo for access to and assistance with ELISA. Thanks also to Shao Dai and Ahmad Jabbarzadeh of the School of Aerospace, Mechanical and Mechatronic Engineering for access to and assistance with zeta-potential measurements. We acknowledge the facilities of the Australian Microscopy & Microanalysis Research Facility. We are thankful to Dr. John Denman for undertaking ToF-SIMS measurements and to Professors Richard Payne and David McKenzie and Associate Professor Boris Kuhlmey for providing helpful comments on the manuscript.

## Author contributions

L.J.M. and B.A. contributed equally to this work. M.M.M.B. proposed the concept. L.J.M. and B.A. performed the experiments. L.J.M., B.A. and M.M.M.B. analysed the results and wrote the manuscript.

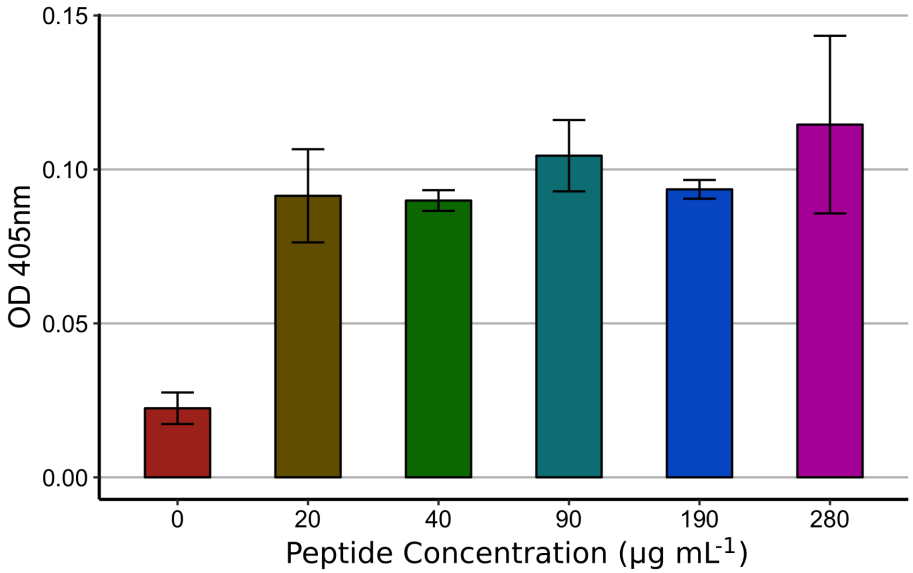
## Additional information

**Supplementary Information** accompanies this paper at https://doi.org/10.1038/s41467-017-02545-6.

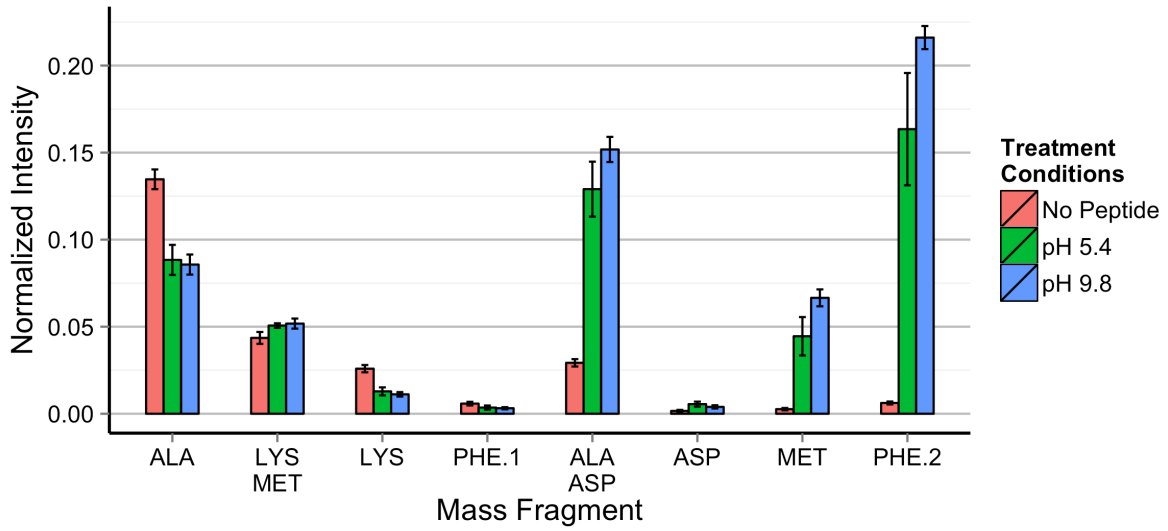
**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.




**Supplementary Figure 1:** Absorbance saturates at high peptide concentration. Enzyme-linked immunosorbent assay does not show concentration dependence above 20 mg $mL^{-1}$, indicating the signal is already saturated. Concentrations above 20 μg $mL^{-1}$ are rounded to the nearest 10 μg $mL^{-1}$. Error bars are s.d.




**Supplementary Figure 2:** Mass spectroscopy of the surface peptide. Time of flight secondary ion mass spectroscopy mass fragments of all residues from the FLAG peptide. The first four fragments (ALA, LYS/MET, LYS, and PHE.1) are discarded because they are present in the No Peptide case. This is because they arise from the radical-functionalized plasma polymer surface itself, confusing the analysis of the pH 5.4 and pH 9.8 cases. The ALA/ASP and ASP fragments are also discarded because the differences in intensity cannot be unambiguously attributed to either ALA or ASP, since they have an unknown contribution to the ALA/ASP fragment. The remaining fragments, MET and PHE.2, are used to analyse peptide orientation. Error bars are s.d.

# Chapter 7. Conclusion

This thesis has used modelling to understand the molecular interactions between peptides and surfaces in atomic detail. Understanding the surface interactions led to the design of a peptide that was used to develop a new method of controlling the orientation and concentration of peptides on surfaces. This chapter connects the conclusions of the previous chapters and proposes a model for creating bioactive peptides that can be manipulated by electric fields.

## 7.1 Conclusion

Knowledge of how synthetic materials and biology interact deepens our understanding of biomaterials. While experimental techniques are capable of analysing this interaction at the macro- and micro-scales, they are mostly incapable of resolving time and length on the atomic scale. This thesis uses molecular modelling to fill this knowledge gap, and the results make it clear that molecular forces on the nanoscale control biomaterial function. Further, the work advances our understanding of the forces on this scale, leading to the possibility of manipulating these forces to create more effective biomaterials.

Chapter 3 reviewed molecular modelling of interfaces, laying the groundwork for all of the following chapters. The problem facing the field stems from the fixed-charge approximation used in most molecular dynamics force fields. By fixing the charge on each atom, molecular dynamics codes take less time to calculate each step and are able to achieve longer simulations than polarizable force fields. The corollary is that when moving from the aqueous phase to the interface, the atomic charge doesn't change to reflect the new electrostatic environment. Responses to this problem normally require modifications of the force field. The evidence presented in the review suggests that, in some cases, careful parameterization can lead to a physically accurate interfacial model without force field modifications. In doing so it discusses some of the crucial elements of surface interactions like adhesion strength, the Lennard-Jones potential, entropy and enthalpy of water, and electric fields.

Chapter 4 built on the reasoning in the review. Surfaces with low surface energy, like PTFE, can be represented by fixed charge molecular modelling because most of the surface interactions are explicitly modelled. Unlike on strongly interacting surfaces, polarization and the topography of the surface is a minor contributor to binding. Parameterization by simulating a water droplet and calculating the contact angle led to a PTFE model with physically accurate surface energy. This model was used to rationalize cell binding trends observed in experiment. One outcome of this was an explanation of how the shape and charge of chemical groups interact with the surface and adjacent water layers to affect the adhesion strength.

Extending our understanding of the bioactive peptides, the focus of Chapter 5 is their behaviour in solution. Despite lacking molecular detail about the receptor, the determinant of binding was localised to the movement of a charge in the middle of the peptide. Extended equilibrium simulations identified the contributions of conformational selection and induced fit. Dimensionality reduction using dihedral principal component analysis showed how this can be done. The unique role of movement at two residues in the bridge region suggests a reason for their high conservation across species. The specific interaction underlying this behaviour may be due to a pi-cation bond. Compared with phenylalanine residues, this occurs more often on tryptophan because of slightly higher polarity and on FMOC due to the large flat shape. These pi-cation interactions are directly analogous to adhesion on the PTFE surface – in each case, the choice between solvating in water or binding a solute is determined by the interplay between the shape and charge of the chemical groups.

The simulations of surfaces and bioactive peptides leads to a better understanding of the interactions at the interface. A low surface energy material like PTFE introduces a weakly interacting boundary into the aqueous phase. This causes water layers with alternating charge and density to arise. Solutes like peptides, that are expelled from the aqueous phase due to the hydrophobic effect, are pushed onto the surface. The tendency to remain on the surface (i.e. the adhesion strength) is highly dependent on their shape, which dictates the volume of water that gets pushed back into the aqueous phase, and their charge, which determines the strength of interaction with the electric dipoles in the water layers.

Integrating all of these effects, Chapter 6 reports the development of a new method to manipulate peptides at the interface. The peptide used in this chapter is rationally designed to combine several useful features that exploit all of the nanoscale interactions identified so far. Its sequence has sulphur atoms that can be measured by spectroscopic techniques, as well as a charged bioactive site that binds antibody proteins to aid in detection. The separation of hydrophobic and charged segments exploits solvation effects to ensure the peptide maintains a linear conformation. Here it is important to note that pi-cation bonds between the two segments are unlikely – most of the charges are negative (i.e. anionic), and the hydrophobic residues are phenylalanines, which are less likely to participate in pi-cation bonds. The linearity and arrangement of charges leads to an electric dipole in line with the long axis of the peptide. As demonstrated in Chapter 6 this leads to a useful effect: specific orientation of the peptide

along electric field lines. This is distinct from alignment, as created by induced dipoles, which occur in two opposing orientations.

Knowledge of these molecule-scale interactions has a striking consequence - the ability to manipulate nanoscale properties by changing macroscopic conditions. Shifting the zeta potential by the simple mechanism of buffer pH changes the electric field strength and alters the orientation and strength of attraction to the surface. Where the surface or biomolecule chemistry prevent tuning by pH, the field can be applied externally by inexpensive equipment. The low strength of the required electric field suggests that this mechanism involves an electric double layer, which increases the field strength at the interface[1].

The approach taken in this thesis represents a potential workflow for biomaterial design. Hydrophobic surfaces are parameterizable using the water contact angle approach in Chapter 2. If knowledge of the number density of charged surface sites is available, then this method can easily be extended to parameterizing these to simulate hydrophilic surfaces. For bioactive peptides, the simulation and analysis techniques demonstrated in Chapter 4 and 5 show an efficient and targeted approach to understanding experimental data. The surface models can also be used for full characterization of surface-peptide behaviour using advanced sampling techniques if desired. The approach in Chapter 6 shows how simulation can guide experimental design of bioactive peptides.

These new levers of biomolecule control are specific in that they influence the peptide without affecting the chemistry of the surface. On the other hand, they have broad significance since they can be applied on a wide range of surfaces and to any nanoscale particle with a permanent dipole stronger than its induced dipole. Guided by the knowledge of nanoscale interactions and anticipating the use of this technique in future, a model for biomaterial ligand design can now be proposed.

### 7.1.1 A model for biomaterial ligand design

A ligand can be built by starting from an active epitope. For the purposes of cellular adhesion this can be an epitope identified in the extracellular matrix, but the procedure can be generalised to non-biological polymers or linkers. These epitopes are often charged in order to exploit hot-spot interactions

with a protein receptor[2]. Charged epitopes sometimes have a dipole (the canonical fibronectin epitope RGD is a common example[3]), but if not then extra charges added to the sequence can create one. The sequence is then extended with a linear linker region, resulting in a ligand with a principal axis and specific orientations. The addition of a hydrophobic segment may aid in linearising the ligand. Molecular dynamics simulations are ideal to test the linear nature and strength of the dipole to aid in this process. Electric fields may even be applied in molecular dynamics simulations and this could be used to determine the field strength required for orientation of a peptide[4].

A possible improvement on this procedure is the use of an alpha-helical peptide. These peptides are highly ordered, folding to create linear particles with an inherent dipole due to the orientation of the backbone atoms[5]. Short alpha helices are already used in biomaterial design[6]. A biologically inert peptide in a stable alpha helix could be added as a base to any bioactive epitope in a modular fashion to create a bioactive ligand with controllable orientation. Non-peptidic alpha helical mimics are promising candidates for this base since they are stable and proteolytically inert and therefore resistant to degradation in a physiological context[7].

In combination with the radical functionalized surface used in Chapter 6, this model has significant benefits over existing approaches. As described in that chapter, existing immobilization techniques require the addition of specific residues that may interfere with the peptide function. While the radical functionalized plasma polymer does away with the need for wet chemical steps for covalent linkage and allows nonspecific covalent linkage without specific residues, it leads to uncontrolled orientations. The combination of the radical functionalized plasma polymer, which can be coated onto numerous underlying materials, with a modular alpha-helical peptide attached to any specific bioactive epitope represents a flexible approach to biomaterial development that is potentially applicable to any surface and any desired biological use case.

### 7.1.2 Future research

There have been several advances in the field of molecular simulation during the production of this thesis. These advances have the potential to improve the statistics and capabilities of simulations of interfaces. Several molecular dynamics codes have outsourced time-consuming calculations to

graphical processing units (GPUs). GPU cards are low cost yet highly parallelized, allowing them to perform simulations at speeds equivalent to hundreds of central processing units (CPUs).

In addition, some codes have implemented software techniques enabling faster equilibrium molecular dynamics. Hydrogen mass repartitioning increases the mass on hydrogens while decreasing the parent atom's mass by an equivalent amount to conserve total mass. This slows the fastest motion in simulations and allows for longer time-steps with minimal impact on protein structure[8]. Alternatively, a multiple time stepping technique using a novel integrator, where some forces are calculated less frequently, shows fast performance that is well-suited to biomolecule simulation[9].

These advances have made, in extreme cases, millisecond-long simulations of whole proteins feasible[10]. For peptide systems, which are smaller and faster to model, the implication is clear. The fixed-charge approximation that is used to reduce the cost of each time-step was implemented to increase simulation output. Advances in hardware and simulation techniques now allow for long time-scale simulations even with the added cost of explicit polarization. These will become particularly useful to explore peptide conformational landscapes on strongly interacting surfaces like metals, where polarization effects dominate and conformational sampling is slow.

### 7.1.3 The interface of materials and biology

The focus of this thesis has been to advance our understanding of biomaterial function by characterising the nanoscale interactions of peptides with synthetic surfaces. The development of a simulation model for hydrophobic surfaces led to an atomic-detail picture of exon 36 peptide adhesion. This model was able to explain how shape and charge affect the adhesion of different chemical groups. The principles extend to solution, governing how exon 36 binds to its receptor on the cell. Finally, this detailed understanding of the interactions at the interface led to a new technique to manipulate charged peptides at the interface. The technique is broadly applicable to the production of many biomaterials. The understanding of surface interactions gained by molecular simulation will thus improve the development and function of new biomaterials.

## 7.2 References


1 Carnie, S. L. & Torrie, G. M. The statistical mechanics of the electrical double layer. *Advances in Chemical Physics, Volume 56*, 141-253 (2007).

2 London, N., Movshovitz-Attias, D. & Schueler-Furman, O. The structural basis of peptide-protein binding strategies. *Structure* **18**, 188-199 (2010).

3 Wilson, C. J., Clegg, R. E., Leavesley, D. I. & Pearcy, M. J. Mediation of biomaterial–cell interactions by adsorbed proteins: a review. *Tissue engineering* **11**, 1-18 (2005).

4 Tarek, M. Membrane electroporation: a molecular dynamics simulation. *Biophysical journal* **88**, 4045-4053 (2005).

5 Wada, A. The alpha-helix as an electric macro-dipole. *Advances in biophysics*, 1-63 (1976).

6 Banwell, E. F. *et al.* Rational design and application of responsive α-helical peptide hydrogels. *Nature materials* **8**, 596 (2009).

7 Henchey, L. K., Jochim, A. L. & Arora, P. S. Contemporary strategies for the stabilization of peptides in the α-helical conformation. *Current opinion in chemical biology* **12**, 692-697 (2008).

8 Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *Journal of chemical theory and computation* **11**, 1864-1874 (2015).

9 Leimkuhler, B. & Matthews, C. Efficient molecular dynamics using geodesic integration and solvent–solute splitting. *Proc. R. Soc. A* **472**, 20160138 (2016).

10 Feinberg, E. N. *et al.* Machine Learning Harnesses Molecular Dynamics to Discover New μ Opioid Chemotypes. *arXiv preprint arXiv:1803.04479* (2018).

# Appendix A

## Cell activity of domain 36 peptide variants

Exon 36, on the C-terminus of tropoelastin, is known to have a role in cellular adhesion via either glucosaminoglycans[1] or integrin $\alpha_V b\beta_3$[2]. This suggests that peptides based on exon 36 could be biomimetic surface molecules that help to integrate biomaterials into organic tissue. Research performed by Dr. Giselle Yeo of the Charles Perkins Centre at the University of Sydney has led to a profile describing the binding activity of several peptides derived from the exon 36 sequence. The research used cellular adhesion assays to determine peptide binding to the tropoelastin receptor. It is included here as experimental context to the simulations results presented in Chapters 4 and 5.

The sequences of the peptide variants are based on the sequence of human tropoelastin exon 36 (ACLGKACGRKRK using the one-letter amino acid code) and are labelled here 'Pep36Am'. This is actually a shortened version of exon 36, since it includes only the elements that are crucial for cellular adhesion such as the charged binding region (RKRK) and a disulfide bond. The Pep36Am sequence is prepended with various cofactors that effect how the variants bind to the tropoelastin receptor. These cofactors are described fully in the main text.

The acetylated peptide variant (AcPep36Am), which most closely resembles the native exon 36 sequence, is incapable of adhering cells to a PTFE surface even when applied to the surface in high concentrations (Figure 1A). The FMOC variant (FmocPep36Am), in contrast, does facilitate cellular attachment. In order to probe the effect of the hydrophobic FMOC cofactor, several other variants using hydrophobic amino acids were tested. Single tryptophan (TrpPep36Am) and single phenylalanine (PhePep36Am) residues did not recapitulate the effect of the FMOC group. The larger double phenylalanine (PhePhePep36Am) group had similar adhesive properties to FMOC. These results suggest that the size of the hydrophobic cofactor facilitates cellular adhesion. To test this, a double tryptophan cofactor was tested (Figure 1B). This variant showed similar cellular adhesion as the FMOC and DiPhe variants. The mechanism of this pattern of cellular attachment to PTFE is explained in Chapter 4.

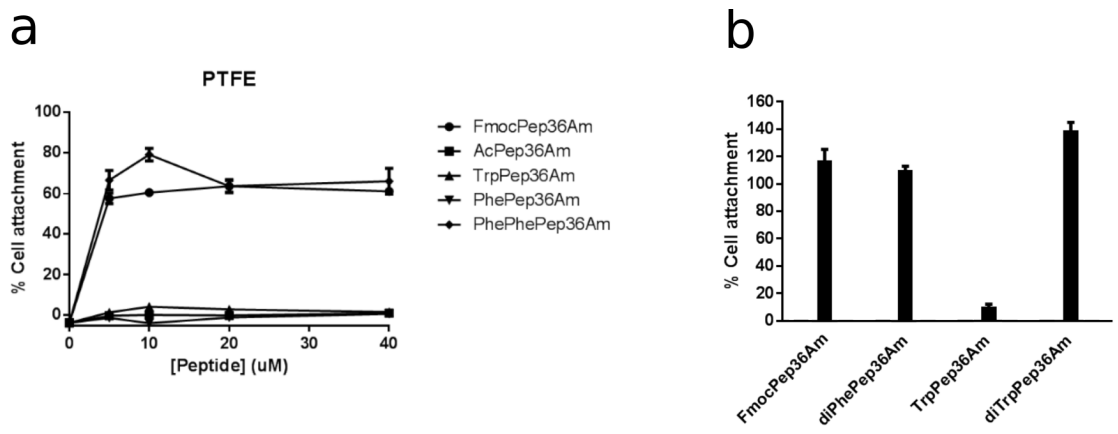



**Figure 1.** Cellular attachment to PTFE, mediated by exon 36 variants. **a)** When adsorbed onto PTFE, domain 36 peptides with acetylated and single phenylalanine cofactors do not cause cell adhesion to the surface. In contrast, the FMOC and double phenylalanine cofactors are able adhere to cells. **b)** The double tryptophan cofactor can cause cellular adhesion, but the single tryptophan cofactor can not.

To fully characterise the variants, they were also tested for the ability to bind the tropoelastin receptor directly from solution. In this case, the surface is coated in tropoelastin proteins. When the peptide variants bind to the receptor on the cell, they block the binding site and prevent interaction with tropoelastin. This inhibits adhesion.

Testing the hydrophobic variants from Figure 1A showed a new pattern of binding. From the aqueous phase, only the FMOC variant was capable of inhibiting adhesion (Figure 2A). The ability of the DiPhe variant to bind the tropoelastin receptor is abolished when in solution. The DiTryp variant, like FMOC, retained the ability to bind the receptor (Figure 2B).

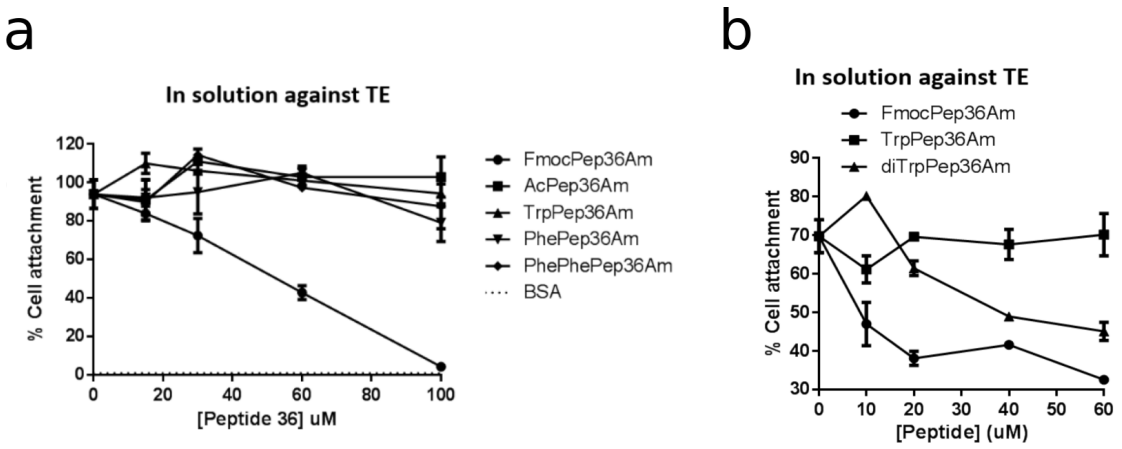



**Figure 2.** Some exon 36 variants can inhibit cellular attachment to tropoelastin by binding to the tropoelastin-specific receptor and blocking the available sites, inhibiting adhesion on tropoelastin-coated surfaces. **a)** In solution the FMOC peptide is able to bind the receptor but the double phenylalanine variant is not. **b)** The double tryptophan cofactor is also able to bind the tropoelastin-specific receptor.

These results suggest that the behaviour of the exon 36 variants is different in the PTFE and aqueous phases. Developing a complete understanding of the mechanism of action of these peptides requires detail on the molecular scale. This motivated the molecular dynamics simulations presented in Chapters 4 and 5.

## References

1 Broekelmann, T. J. *et al.* Tropoelastin interacts with cell-surface glycosaminoglycans via its COOH-terminal domain. *Journal of Biological Chemistry* **280**, 40939-40947 (2005).

2 Bax, D. V., Rodgers, U. R., Bilek, M. M. & Weiss, A. S. Cell adhesion to tropoelastin is mediated via the C-terminal GRKRK motif and integrin αVβ3. *Journal of biological chemistry* **284**, 28616-28623 (2009).

# Appendix B

## Block averaging

Block averaging is a procedure to determine the confidence interval around the mean for highly correlated datasets.

Given a time series of independent measurements, the mean of some property is trivial to calculate and is given by:

$$\langle x \rangle = \frac{x(1) + x(2) + \cdots + x(N)}{N}$$

Where $\langle x \rangle$ is the population mean, $x(1,2 \ldots N)$ are the sample independent measurements, and *N* is the number of measurements.

If $\sigma_x$ is the standard deviation of $x$, applying the central limit theorem gives us the standard deviation of the sample:

$$\sigma_{\langle x \rangle} = \frac{\sigma_x}{\sqrt{N}}$$

This relation assumes that the data points are independent, but that is not always the case. For many time series, the data are recorded far more frequently than the underlying property varies. This means that the value of adjacent data points are likely to be similar simply by virtue of their closeness in time – i.e. the data are highly correlated. The number of independent data points is fewer than the number of measured data points.

The number of independent data points in a correlated sample can be found by “blocking” the data into a few, averaged measurements where the blocks are large enough such that the means of each block are uncorrelated. Block averaging is a procedure to determine the correct size of the blocks so that they are independent.

Figure 1 shows the time series of a dihedral angle from a simulation lasting 1 ns with data recorded every 0.5 ps, giving 2000 data points. The mean is 38.0° and applying the central limit theorem gives

$\sigma_{\langle x \rangle} = \frac{38.0}{\sqrt{2000}}$, which is 0.6°. This value is artificially low. By averaging adjacent data points into blocks, the sample has fewer total data points. With increasing block size (i.e. at first blocking the whole data set using 2 points at a time, then 3, and 4 and so on) the data points become further apart in time and are therefore less correlated.

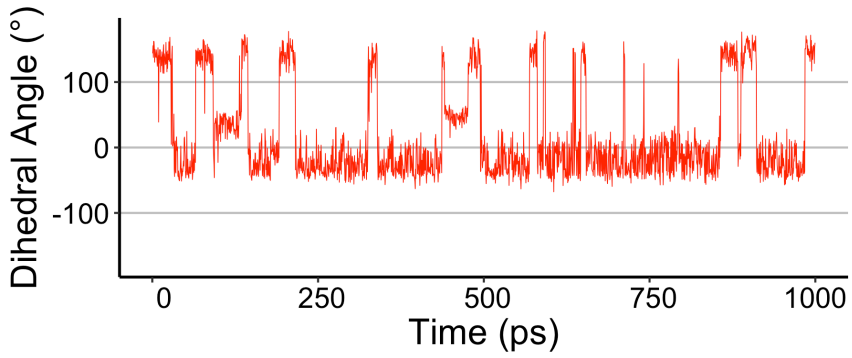



**Figure 1.** Time series of a peptide backbone dihedral angle from a simulation lasting 1 ns. Because the data are recorded so frequently, adjacent data points are highly correlated. This leads to an artificially low standard deviation.

Figure 2 shows the progression of $\sigma_{\langle x \rangle}$ as the block size *B* is increased. The sample standard deviation increases up to $B \cong 100$, after which it fluctuates around a consensus value. The correct block size to achieve independent data is thus $B \cong 100$. To determine the standard deviation exactly, the block averaging data can be fit to find an asymptote. Although there is no accepted curve to fit the block averaging data, good results can be found by fitting to $y = a \times \tan^{-1}(bx)$, where the asymptote is given by $\sigma_{\langle x \rangle} = \frac{a\pi}{2}$.

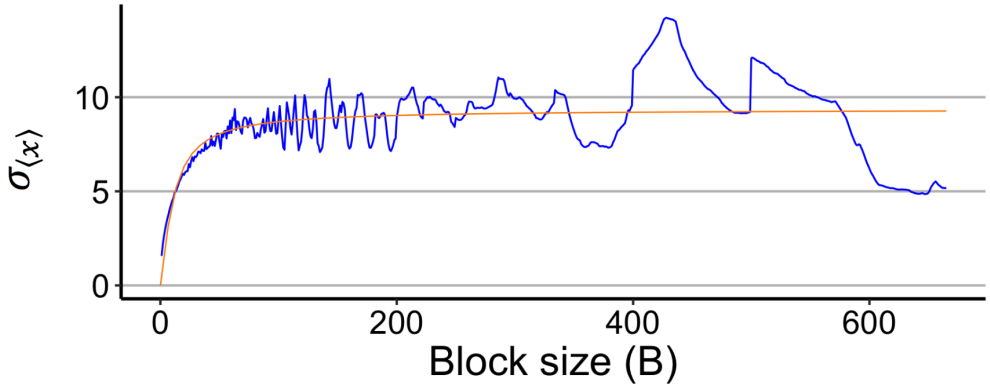



**Figure 2.** The population standard deviation of the dihedral angle time series with different block sizes. Without blocking the data (B = 1), the calculated standard deviation is too low. With increasing block size the standard deviation rises to the true value, indicating uncorrelated data points.

# Appendix C

## Circular Statistics

When working with angular data, the mean and variance cannot always be calculated in the usual way. As an example, consider two angles: 30° and 330°. The arithmetic mean of these angles is 180° but, when seen graphically, the mean clearly should be 0° (Fig. 1). The mean and variance should therefore take into account the periodicity of the units. This is useful for simulations where protein backbone dihedral angles are being measured.

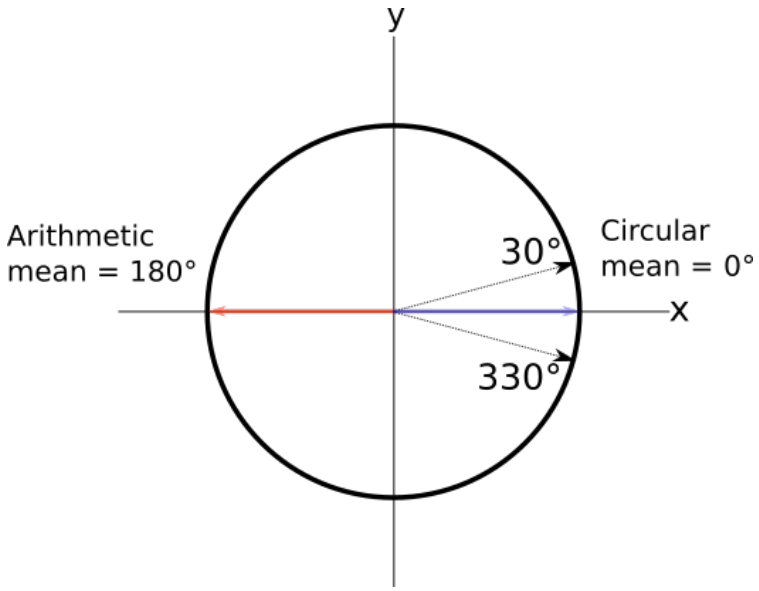



**Figure 1.** The mean of two angles (30° and 330°) calculated using arithmetic mean, red, or circular mean, blue.

## Circular mean.

The circular mean, $\bar{\theta}$, is found by first determining the average of the *sin* and, separately, *cos* of the angles, i.e.

$$\bar{s} = \frac{1}{N}\sum_{n=1}^{N} \sin\theta_n$$

and

$$\bar{c} = \frac{1}{N}\sum_{n=1}^{N} \cos\theta_n$$

Then, depending on these values, the average is[1]:

$$\bar{\theta} = \begin{cases} \arctan\left(\frac{\bar{s}}{\bar{c}}\right), & \bar{s} > 0, \bar{c} > 0 \\ \arctan\left(\frac{\bar{s}}{\bar{c}}\right) + 180°, & \bar{c} < 0 \\ \arctan\left(\frac{\bar{s}}{\bar{c}}\right) + 360°, & \bar{s} < 0, \bar{c} > 0 \end{cases}$$

## Circular variance.

The variance of angular data used here is not directly analogous to the variance of linear quantities. Instead an *R* parameter, which varies from 0 to 1, is used[2]. This quantity indicates the degree of clustering of the angles, where tightly clustered angles (low variance) have *R* values close to 1 and widely distributed angles (high variance) have *R* values close to 0. The variance is calculated as follows:

$$\frac{R}{N}^2 = (\sum_{n=1}^{N} \cos\theta_n)^2 + (\sum_{n=1}^{N} \sin\theta_n)^2$$

## References


1 Allen, F. H. & Johnson, O. Automated conformational analysis from crystallographic data. 4. Statistical descriptors for a distribution of torsion angles. *Acta Crystallographica Section B: Structural Science* **47**, 62-67 (1991).

2 MacArthur, M. W. & Thornton, J. M. Conformational analysis of protein structures derived from NMR data. *Proteins: Structure, Function, and Bioinformatics* **17**, 232-251 (1993).