Potential and Free Energy Surfaces of Adsorbed Peptides

James Andrew Ross-Naylor

School of Chemical Engineering and Advanced Materials

Submitted January 2020 for the degree of Doctor of Philosophy

Table of Contents

Citation Listing

Chapter 4

ROSS-NAYLOR, J. A., MIJAJLOVIC, M., ZHANG, H. & BIGGS, M. J. 2017. Characterizing the Switching Transitions of an Adsorbed Peptide by Mapping the Potential Energy Surface. *J. Phys. Chem. B,* 121**,** 11455-11464.

Chapter 5

ROSS-NAYLOR, J. A., MIJAJLOVIC, M. & BIGGS, M. J. 2020. Energy Landscape Mapping and Replica Exchange Molecular Dynamics of an Adsorbed Peptide. *J. Phys. Chem. B,* 124**,** 2527-2538.

Chapter 6

ROSS-NAYLOR, J. A., MIJAJLOVIC, M. & BIGGS, M. J. 2020. Energy Landscapes of a Pair of Adsorbed Peptides. *J. Phys. Chem. B,* 124**,** 2401-2409.

Statements of Authorship for all citations are included prior to the beginning of the relevant chapter.

Abstract

Peptide adsorption on solid surfaces is a common process that occurs in nanotechnology and biology, with applications in the formation of nanomaterials, biosensing and drug delivery, amongst many others. Peptide adsorption involves complex processes that are difficult to characterise experimentally. Computational approaches such as molecular dynamics (MD) are often employed to better understand biomolecular systems. However, the computationally demanding nature of such systems combined with the long characteristic timescales of peptide adsorption means MD is not well suited to its study with current computing capacities.

An alternative computational approach to characterising the behaviour of atoms and molecules is mapping the potential energy surface (PES) – the molecular energy as a function of the positions of all atoms – by determining its local energy minima and saddle points, which represent stable configurations and transition states that lie between them. These minima and saddle points may be located using optimisation algorithms. Harmonic approximations yield information about transition rates between minima via saddle points as well as the free energy surface (FES). This methodology – which is referred to as ‘energy landscape mapping’ (ELM) hereafter – is able to characterise fast and slow processes equally, only being limited by the size and complexity of the system studied, and the applicability of the potential energy models used. In the past, it has largely been applied to atomic and molecular clusters, and to biomolecules. It has never been applied to adsorption of peptides or any other biomolecule.

In three journal papers included in this thesis, this approach is for the first time applied to adsorbed peptides. Firstly, ELM was applied to polyalanine adsorbed on surfaces of varying interactions strengths. Results obtained were comparable results to those obtained in a prior study of the same system using an evolutionary algorithm. In the second paper, ELM was applied to met-enkephalin at a gas/graphite interface, and compared with a molecular simulation technique designed for accelerating the simulation of slow processes, replica exchange molecular dynamics (REMD). In the final paper, ELM was applied to two met-enkephalin molecules at a gas/graphite interface, introducing an additional level of complexity and a step towards practical application, given real peptide adsorption processes often occur *en masse*. In all of these studies, information about transitions between conformations, energy barriers, rates, and the nature of the overall PES and FES, all of which were previously unknown for the systems studied, was obtained by ELM.

The work conducted here has demonstrated the applicability of ELM to peptide/surface systems. Future work may consist of applying ELM to other similar processes of practical importance, developing and validating potential energy models suitable for modelling interfacial systems, including the effect of solvents, and continual development of the methodology to accelerate calculations.

Thesis Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

James Andrew Ross-Naylor, 31 January 2020

Acknowledgements

This project has been a long and arduous labour of love, and I owe much to many people, without whom this work would not have been possible.

Most prominently, I would like to acknowledge the work and assistance of my supervisory panel. Mark Biggs, who first recommended this project to me some seven-and-a-half years ago, whose widespread expertise and knowledge was invaluable from start to finish, and who spent long hours editing my papers when there was much else he could have been doing.

Milan Mijajlovic, whose past work set the foundation for my research, and who as a co-supervisor provided both technical know-how and advice, and a friendly, can-do attitude when I needed it.

Hu Zhang, who took on the role of principal supervisor when both Mark and Milan moved overseas, and handled a difficult situation admirably.

I would also like to thank my parents, extended family and friends, for lending a listening ear, offering a positive outlook and occasionally providing an interesting new perspective on my work from an outsider's point of view. I love you all: you helped me than you can know.

Finally, I would like to thank God, the Creator, Designer and Ruler of this beautiful world. It's my privilege to study the works of Your hand, and I thank You for staying by my side when things got tough. This is all for You.

We did it!

1 Introduction

1.1 Contextual Background

Peptides are small-to-medium-sized biomolecules that are found across nature. Chemically, peptides, as well as proteins, are defined as chains of amino acids that are linked together by peptide bonds, as illustrated in the example in Figure 1-1. Peptides are smaller than proteins, although the distinction between them is not precisely defined, but in general, a molecule may be considered a peptide if it comprises fewer than around 50 to 100 amino acid residues (Bodanszky, 1988, Craik et al., 2013).

Regardless of precise definitions, both peptides and proteins are of widespread interest and relevance. The three-dimensional structure of proteins is vital to their biological function, and has resulted in a vast field of study of their folding

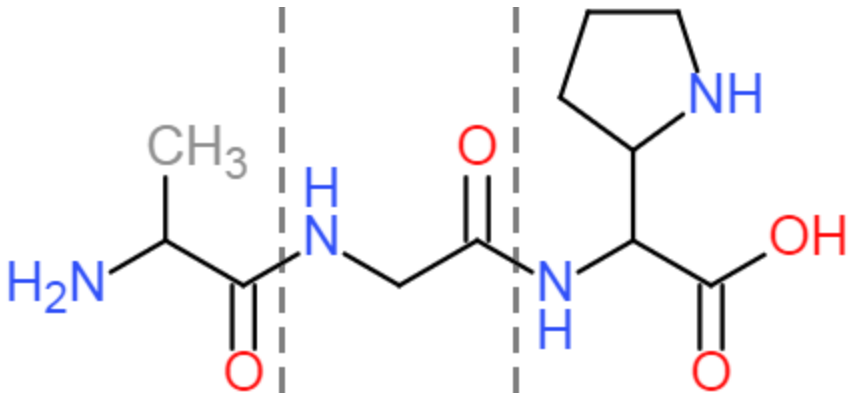

Figure 1-1. Structure of an example tripeptide. The dashed lines separate the three amino acid residues comprising the peptide, which are (in order from left to right) alanine, glycine and proline.

(Dobson, 2003, Onuchic and Wolynes, 2004, Dill and MacCallum, 2012), and the creation of the Protein Data Bank, which contains all known protein structural data (Berman et al., 2003). Peptides, meanwhile, are also of wide interest. They are frequently used as components or facilitators in the synthesis of nanomaterials (Slocik and Naik, 2017, Walsh and Knecht, 2017) and biomaterials (Sunna et al., 2017). Additionally, peptide-based therapeutics are a source of ongoing interest and development (Craik et al., 2013, Fosgerau and Hoffmann, 2015), and antimicrobial peptides provide a promising avenue toward combatting growing bacterial immunity to antibiotics (Pasupuleti et al., 2012, Hoyos-Nogués et al., 2018). Peptides may also be used in catalysis (Zozulia et al., 2018) and the development of synthetic vaccines (Skwarczynski and Toth, 2016).

The adsorption of biomolecules to solid surfaces is a well-documented phenomenon (Nakanishi et al., 2001, Gray, 2004, Rabe et al., 2011, Marruecos et al., 2018). A particular example of this occurrence in nature is the tendency for proteins and other biomolecules to adsorb to foreign objects within the human body such as implants and nanoparticles (Docter et al., 2015, Trindade et al., 2016). On the other hand, the adsorption properties of peptides can be exploited in a number of ways. Peptides designed or found naturally to specifically bind to particular materials are now routinely used in fabrication of nanoparticles (Briggs and Knecht, 2012) and protective coatings for implants (Pagel and Beck-Sickinger, 2017, Mas-Moruno, 2018). Their surface adsorption properties have also been exploited for biosensing (Demir et al., 2016, Liang et al., 2016, Chen

and Nugen, 2019), and the delivery of peptide-based therapeutics (Kumar et al., 2012, Huotari et al., 2013, He et al., 2018).

To facilitate further development in these fields, it is desirable to gain a deeper understanding of how peptide adsorption occurs at the molecular level, and the behaviour and changes in conformation of adsorbed peptides. Such an understanding would aid in the design of peptides with optimised properties and functionality (Costa et al., 2015, Schwaminger et al., 2018). As such, a vast array of reported studies of peptide adsorption exists, incorporating both theory and experiment. Experimental methods such as microscopy and spectroscopy can elucidate the amount and compositions of peptides adsorbed on a surface, and their three-dimensional structure (Costa et al., 2015, Schwaminger et al., 2018). A step forward is provided by computational methods, which can optimise the preferred conformation of adsorbed peptides and simulate their behaviour at a surface for timescales up to the microsecond range (Latour, 2008, Heinz and Ramezani-Dakhel, 2016, Ozboyaci et al., 2016a, Ramakrishnan et al., 2017).

1.2 Research Problem and Objective

One of the primary challenges to computational studies of peptide adsorption today is the limitations on what molecular processes can realistically be characterised by molecular simulation methods. Changes in conformation of peptides already adsorbed on a surface can occur over timescales far longer than the current state of the art in simulations, and as such, simulations frequently show a peptide taking on and remaining in a given state whilst not visiting others

that are of physical importance (Latour, 2008, Ozboyaci et al., 2016a). Increases in the system's complexity with the goal of better modelling real processes, such as the simulation of multiple peptides adsorbing in bulk or the inclusion of solvent effects, tends to exacerbate these problems. Enhanced sampling methods have been applied in an attempt to address such issues, but these have limitations that will be discussed in the following chapter.

A computational approach that provides an alternative to simulation is to consider the energy landscapes of the system from a mathematical perspective. The term 'energy landscape' is a visual manner of describing the potential energy surface (PES) and free energy surfaces (FES), wherein the potential or free energy is represented as a multidimensional hypersurface (Wales, 2003). Minima on the landscape, for which any small change in configuration will result in a higher energy, correspond to locally stable configurations, while first-order saddle points, which maximise the energy in one direction while minimising it in all others, correspond to transition states. A representative example of a surface containing defined minima and saddle points is illustrated in Figure 1-2. Steepest descent paths connect saddle points to the minima they lie between, and their energies relative to one another determine energy barriers and transition rates. Locating minima and saddle points using mathematical algorithms can thus elucidate and quantify molecular transitions, regardless of the rate at which they occur in real time. The difference between simulation and finding minima and saddle points may be likened to the difference between walking on a landscape (where distance and steep climbs may forbid access to some areas) and viewing it from above in

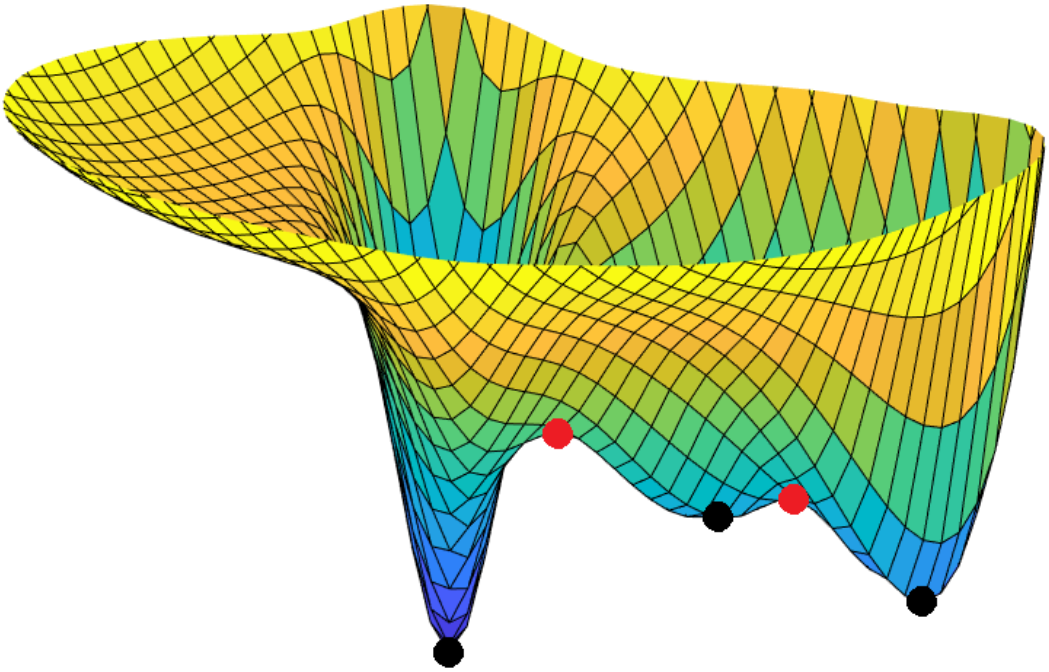

Figure 1-2. Three dimensional example of a hypersurface (Müller and Brown, 1979). Minima and saddle points are highlighted in black and red, respectively.

order to map distinguishing features such as valleys (minima) and mountain passes (saddle points). Consequently, this work adopts the term 'energy landscape mapping' (ELM) for this approach.

ELM is typically applied to glass-forming liquids, atomic and molecular clusters, and biomolecules (Wales, 2003); a recent review of its applications to biomolecules was provided by Joseph et al. (2017). During the course of this project, the first ELM study of an amino acid adsorbed on a surface was published (Sacchi et al., 2017). However, to the best of the author's knowledge, peptide adsorption has not previously been studied using ELM before.

In brief, the objective of this project is to apply ELM to a variety of peptide adsorption systems, and to demonstrate its value in complementing molecular

simulation and yielding additional information regarding conformational changes that cannot otherwise be elucidated.

1.3 Thesis Outline and Scope

In chapter 2, a survey of literature relating to the project is undertaken. Firstly, an overview of existing experimental and computational techniques that are used for studying peptide adsorption is given, and the benefits and limitations of each is assessed. Secondly, methods used in ELM are outlined, including algorithms to find minima and saddle points, and means of assessing and exploiting the resulting networks of stationary points. Finally, models for the calculation of the potential energy of peptide-surface systems are assessed, along with the challenges associated with such models. Chapter 3 presents greater detail about the ELM methods utilised in the work conducted for this project, including algorithms and schemes for finding minima and saddle points, and the calculation of rate constants and free energies.

The body of this thesis by publication is presented in three journal papers, which are included as chapters 4, 5 and 6. The first of these was published by the *Journal of Physical Chemistry B* in 2017, while the latter two were published by the same journal in 2020. Chapter 4 reports the first ever application of ELM to an adsorbed peptide. The system studied was polyalanine adsorbed at solid surfaces of varying interaction strengths, which had previously been studied by Mijajlovic and Biggs (2007) using an evolutionary algorithm. The present work yielded comparable results to theirs, and expanded on them by providing detailed

transition mechanisms and estimated rate constants between competing adsorbed conformations, thus validating the applicability of ELM to adsorbed peptides.

In chapter 5, the conformations of a different peptide, met-enkephalin, at a graphite surface is investigated using two different techniques: ELM and replica exchange molecular dynamics, an established methodology for enhanced sampling in molecular simulation. The results obtained by both techniques were in good agreement. The paper assesses the information and insights provided by both methodologies, and the challenges to each.

In chapter 6, ELM is applied to study the conformations of two met-enkephalin molecules adsorbed on a graphite surface, and the transitions between them. While the previous two papers had established the applicability and benefit of ELM in studying adsorbed peptides compared to other methodologies, here ELM was used to study a system for which detailed computational studies using other techniques are sparse: the interactions of multiple adsorbed peptides with one another. This is a phenomenon well worth studying, since in practical applications most peptide adsorption processes occur in the non-dilute regime. As such, the results of this study are of greater relevance from scientific and technological perspectives, and represent a step towards the study of bulk adsorption.

Chapter 7 concludes the thesis with a discussion of the implication of the results obtained, and recommendations for future work. An appendix is included, containing the Supporting Information for Publication associated with chapter 4. Finally, a list of references for the entire thesis is given.

Before proceeding to the Literature Review, the limitations on the scope of this project are worth stating. As previously mentioned, the primary goal of the project was to apply ELM to adsorbed peptides and to demonstrate its applicability and value in terms of results. As such, although the next chapter outlines several competing methodologies for finding minima and saddle points and for modelling potential energies, the studies reported here do not compare such methodologies and models or attempt to determine which is optimal. Additionally, although solvent is known to play a significant role in adsorption processes (Jena and Hore, 2010, Penna et al., 2014, Xu et al., 2018), there are presently challenges in modelling such effects accurately, as outlined in the next chapter. As such, the studies reported in this thesis focus on the gas/solid interface, with no solvation effects included.

2 Literature Review

2.1 Introduction

In this chapter, existing literature relevant to the use of energy landscape mapping (ELM) in studying adsorbed peptides is assessed. As outlined in the previous chapter, the motivation behind this work is the fact that ELM has not yet been utilised for this purpose, and the gap in knowledge that exists as a result. This literature review, therefore, assesses multiple separate, although in some aspects related, bodies of work. Firstly, an overview is given of existing techniques that are commonly used to study peptide adsorption. Secondly, the development of ELM methods and the current state of the art is outlined. This includes algorithms to find minima and saddle points, as well as methods to visualise, interpret and exploit the resulting information about the energy landscapes. Finally, attention is given to the potential energy models, or 'force fields', available to model peptide-interface systems. These force fields are normally used for molecular simulations – the most widespread means of studying the phenomenon computationally – and are intrinsically necessary for ELM. Relevant force fields are presented and the challenges to accurately modelling such systems are assessed.

2.2 Existing Techniques for Studying Peptide Adsorption

Studies of adsorbed or adsorbing peptides are widespread in literature, owing to their ubiquitous applications as outlined in chapter 1. Information regarding the binding mechanisms and affinities of peptides towards surfaces, and their adsorbed conformations, may be elucidated either by theory or experiment, or a

combination thereof. Recent reviews of peptide adsorption studies that focus on experiment with some reference to computational studies and techniques have been published by Costa et al. (2015, 2016) and Schwaminger et al. (2018). Meanwhile, Heinz and Ramezani-Dakhel (2016), Ozboyaci et al. (2016a) and Ramakrishnan et al. (2017) have provided broad reviews of peptide adsorption studies carried out by computational methods, particularly molecular simulations. This section focuses on the most prevalent techniques used to study peptide adsorption, the information and insights that can be gleaned from each, and the challenges and limitations involved in applying them effectively to peptide-surface systems.

2.2.1 Experimental Methods

2.2.1.1 Trial and Error

The simplest and most intuitive means of investigating a peptide's adsorption mechanism experimentally is by testing the binding affinities of related molecules, such as individual amino acids, their sidechain analogues and sub-sequences from within the peptide, and making inferences from these results. Vogel and Altstein (1977) used this approach to study the adsorption of met-enkephalin and leu-enkephalin on polystyrene, determining that in each case, the C-terminal dipeptide sequence (Phe-Met and Phe-Leu, respectively) was responsible for the adsorption. In a more recent study using a similar approach, Zhang et al. (2011) quantified the binding strengths of several amino acids on graphene oxide, and used these results to design peptides for adsorption to the surface. Such studies are constrained only by the ability to obtain or synthesise the required amino acids and peptides, but

the information they yield is correspondingly limited. While they give an indication of which residues of a peptide adsorb to a surface, other techniques are generally required to elucidate the nature and mechanics of this adsorption (Zhang et al., 2011).

2.2.1.2 Microscopy

The three-dimensional structure of peptides adsorbed on a surface may be observed by scanning probe microscopy, a class of microscopes that physically probe systems with atomic-level precision (Salapaka and Salapaka, 2008) and are thus capable of directly imaging the surface (Costa et al., 2015). The scanning tunnelling microscope (STM) (Binnig and Rohrer, 1983, Binnig and Rohrer, 2000) and atomic force microscope (AFM) (Binnig et al., 1986, Eaton and West, 2010) were both invented by Binnig and co-workers in the 1980s, and today are the most popular microscopic methods for studying adsorbed biomolecules (Costa et al., 2015, Schwaminger et al., 2018). Of the two, STM offers superior resolution, although at the risk of interfering with or damaging the system being probed (Costa et al., 2015).

Images obtained through STM and AFM display the corrugations on the surface, therefore elucidating how adsorbed peptides ore oriented with respect to the surface and one another. For example, Sek et al. (2005) used STM to observe a polyalanine monolayer formed on Au [111] and estimate the intramolecular distance. More recently, also using STM, Méthivier et al. (2017) showed that Gly-Pro dipeptides self-organise on a Au [110] surface in linear patterns, depending on whether they are in the uncharged or zwitterionic form, as shown in Figure 2-1.

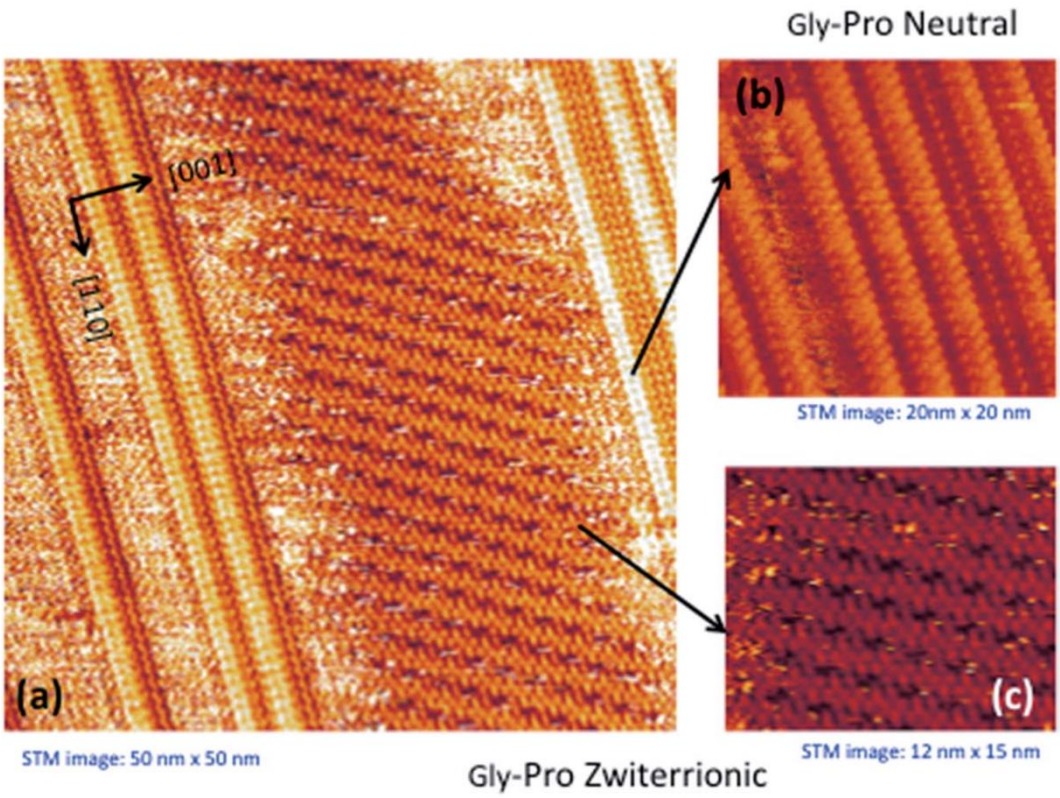

Figure 2-1. STM images of uncharged and zwitterionic Gly-Pro dipeptides self-assembling at a Au [110] surface (Méthivier et al., 2017).

Meanwhile, Nakayama et al. (2015) used AFM to study the assembly of a peptide, Fmoc-EVVKV, on a mica surface, noting the thickness and organisation of the fibres formed as a function of the methanol/water ratio used as a solvent. These studies illustrate the manner in which microscopy represents a step towards understanding the adsorption mechanism and adsorbed conformations, even if not yielding direct information about either.

2.2.1.3 Spectroscopy

The final major class of experimental techniques utilised in peptide adsorption studies are spectroscopic methods. These include infrared reflection absorption

spectroscopy (IRRAS), X-ray photoelectron spectroscopy (XPS), circular dichroism (CD) and others. Each differ in their theoretical background and applications. IRRAS is a form of vibrational spectroscopy: a molecule exposed to infrared light will absorb radiation at frequencies characteristic to its functional groups. Changes in the spectra when adsorbed compared to in the gas phase can indicate adsorption sites and interactions between adsorbed molecules (Hoffmann, 1983). XPS, meanwhile, is a form of photoemission spectroscopy, and can yield information regarding the chemical composition and electronic state of a system. As such, IRRAS and XPS are often complementary and may be used together to gain a more complete understanding of an adsorbate's final state (Costa et al., 2015). The aforementioned study of Méthivier et al. (2017) is one example of this. CD is a methodology that utilises the responses of chiral molecules to polarised light, and is thus very useful for many biomolecules. Of particular use is the fact that CD spectra differ according to a molecule's conformation (Woody, 1995). As such, when CD is applicable, it can be used to detect changes in structure (Slocik et al., 2019) or directly identify secondary structures (Slocik et al., 2011, Liu et al., 2016).

2.2.2 Molecular Dynamics

Molecular dynamics (MD) is a method of computationally simulating the movements of atoms in a molecular system (Frenkel and Smit, 2002). Beginning at a specified initial condition, the new positions of all atoms are calculated at regular intervals and coordinates are regularly saved, creating an ordered series of three-dimensional configurations known as a trajectory. MD trajectories

essentially simulate the atoms' movement over the timescale covered by the simulation, and are thus highly useful in studying biomolecular processes (Karplus and McCammon, 2002).

Young (2001) and Frenkel and Smit (2002) provide overviews of the fundamentals of MD in their works. Leimkuhler and Matthews (2015) describe the mathematical formulations involved in considerable detail, although the particular implementations vary according to the software package used (Plimpton, 1995, Phillips et al., 2005, Brooks et al., 2009, Salomon-Ferrer et al., 2013, Abraham et al., 2015). Here, a brief explanation of the procedure for MD is given. Initially, all atoms are allocated random velocities. Then, based on the positions of the atoms and the potential energy model in use, the forces on each atom (partial derivatives of energy with respect to distance) are calculated. A numerical method such as the Verlet (Verlet, 1967) or velocity Verlet (Swope et al., 1982) algorithm is used to integrate Newton's equations of motion and calculate positions and velocities for the next time step. The force calculations and integrations are then repeated until a desired amount of simulated time has been reached.

Standard MD sees widespread use in studies of peptide adsorption. It is particularly commonly used as a complement to experimental studies of solid-binding peptides, determining the adsorbed conformation of the peptide (Kim et al., 2011, Ramezani-Dakhel et al., 2015, Gladytz et al., 2016, Liu et al., 2016). Figure 2-2(a) illustrates the contrasting adsorption structures of a graphene-binding peptide (GBP) and carbon nanotube-binding peptide (CBP) obtained by

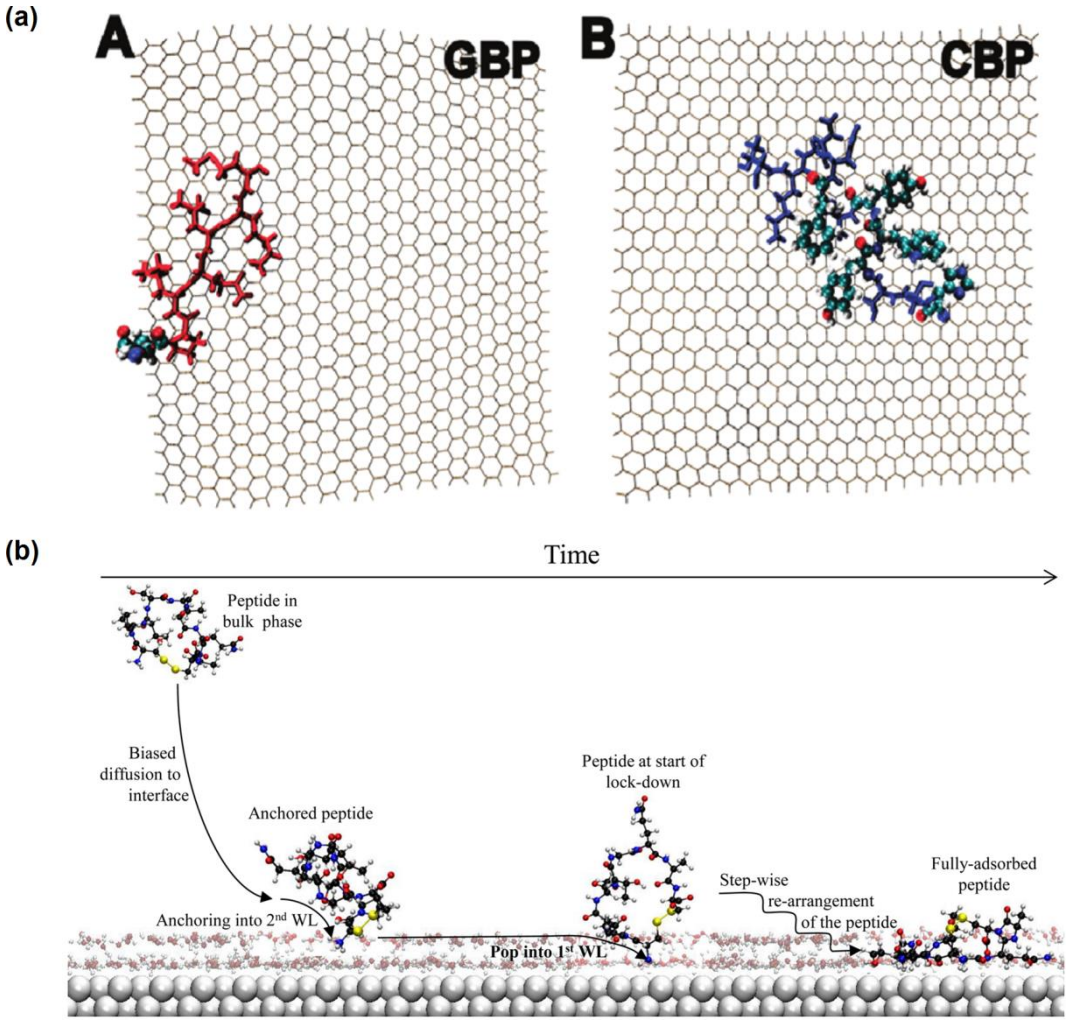

Figure 2-2. (a) Adsorbed conformations of two separate peptides on graphene surfaces (Kim et al., 2011), and (b) proposed mechanism of adsorption of a peptide at a water/solid interface based on a MD study (Penna et al., 2014).

Kim et al. (2011). MD is also used to simulate the adsorption process and thus propose the adsorption mechanism (Penna et al., 2014, Ozboyaci et al., 2016b, Luo et al., 2019), to assess the effect of variables such as pH, surface structure and surface charge (Braun et al., 2002, Emami et al., 2014, Ramezani-Dakhel et al., 2015, Zhang and Sun, 2018), and to determine the conformations of multiple peptides adsorbing in bulk (Mu et al., 2014, Ramezani-Dakhel et al., 2015, Liu et

al., 2016, Zhang and Sun, 2018). Recently, a generalised mechanism of adsorption for peptides at a water/solid interface was proposed based on MD simulations (Penna et al., 2014). This is illustrated in Figure 2-2(b).

Critical to the success of MD simulations is the choice of the potential energy model used for energy and force calculations (Ramakrishnan et al., 2017, Walsh, 2017). Most often, potentials that simulate interactions between individual atoms, known as 'force fields', are used. There are many such force fields in existence, and the optimal choice tends to depend on the application; consideration is given to the force fields used for simulating peptide-surface interactions in Section 2.4. *Ab initio* MD (Kresse and Hafner, 1993) presents an alternative means of performing MD using quantum mechanical techniques, which are discussed in more detail in Section 2.2.4.1. However, calculations using these techniques are far more computationally expensive, limiting *ab initio* MD simulations to far smaller systems and shorter timescales (Heinz and Ramezani-Dakhel, 2016, Ozboyaci et al., 2016a).

2.2.3 Molecular Dynamics with Enhanced Sampling

Sampling of configurational space has been identified as a major roadblock to the effective application of MD to peptide adsorption (Latour, 2008, Ozboyaci et al., 2016a, Walsh, 2017). For molecular systems as geometrically complex as an adsorbed or adsorbing peptide, the energy landscapes tend to be correspondingly complex, frequently including multiple low-energy states with high energy barriers between them (Latour, 2008). Individual MD simulations at reasonable temperatures such as 298 K, therefore, frequently become trapped in one of the

low-energy states and fail to identify the others (Latour, 2008, Ozboyaci et al., 2016a). A related problem is the amount of simulation time required to characterise some important adsorption and rearrangement processes, which can occur on longer timescales than classical all-atom MD is currently capable of simulating (Ozboyaci et al., 2016a). As a result of these sampling issues with peptide adsorption and other similarly complex phenomena, a number of enhanced sampling methods for conducting MD have been proposed and applied as outlined in the following sub-sections.

2.2.3.1 Replica Exchange Molecular Dynamics

Replica exchange molecular dynamics (REMD) was first applied by Sugita and Okamoto (1999) as an analogue of the Monte Carlo concept of parallel tempering (Swendsen and Wang, 1986). REMD consists of multiple 'replicas' of the same MD simulation with a given characteristic varying for each replica. The simulations are run in parallel, and at regular intervals exchanges are attempted between replicas with neighbouring characteristics (Sugita et al., 2019). The most common formulation is the original REMD (Sugita and Okamoto, 1999), where replicas are simulated at different temperatures, and exchanges are accepted or rejected based on the Metropolis criterion (Metropolis et al., 1953):

$$P_{ij} = min\left(1, e^{-\left(\frac{1}{k_B T_j} - \frac{1}{k_B T_i}\right)(U_i - U_j)}\right) \qquad (2\text{-}1)$$

where P_{ij} is the probability of replicas i and j with adjacent temperatures T_i and T_j and potential energies U_i and U_j exchanging temperatures. A schematic of a temperature REMD simulation is illustrated in Figure 2-3. REMD has also been

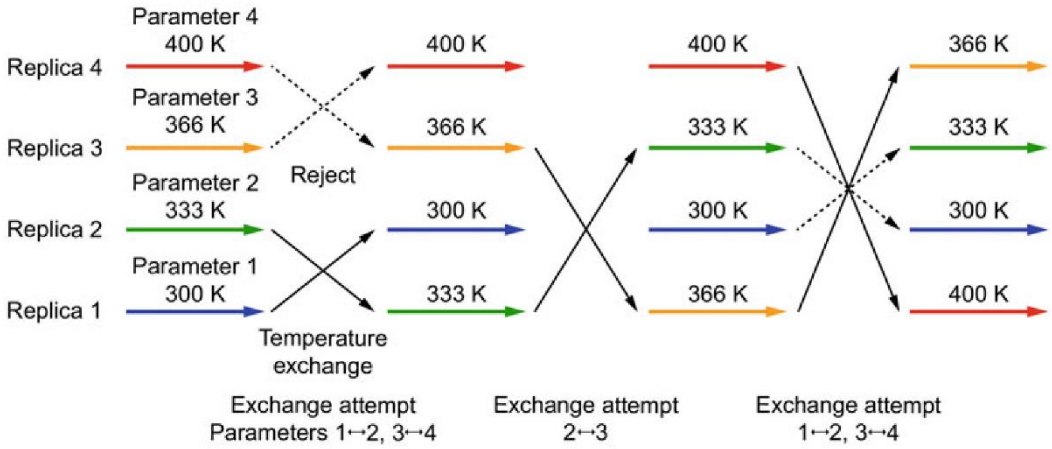

Figure 2-3. Schematic of a temperature-based REMD simulation (Sugita et al., 2019).

applied using bias potentials (Fukunishi et al., 2002). For REMD simulations in which the temperature is the varied parameter, the intent is that the higher temperatures will permit the system to reach areas of configurational space that are not usually accessible otherwise, and it will remain in those areas upon cooling to the minimum temperature (which is generally the replica of interest), resulting in their adequate sampling.

The primary drawback of REMD is the computational effort required in order to achieve meaningful results. For large systems with many degrees of freedom, particularly simulations that include explicit solvent, the number of replicas required to cover a significant temperature range while maintaining a realistic acceptance probability becomes unreasonable (Liu et al., 2005). This motivated the development of replica exchange with solute tempering MD (REST-MD), which holds the solvent temperatures of all replicas constant while only varying and exchanging the solute temperatures (Liu et al., 2005). REST-MD has been shown to achieve similar degrees of sampling to REMD at a fraction of the cost

(Wright and Walsh, 2013). This development has permitted a multitude of replica-exchange simulations of peptides at a water/solid interface in recent years, with the improved sampling typically enabling the elucidation of multiple stable configurations (Wright and Walsh, 2013, Hughes and Walsh, 2015, Bellaiche and Best, 2018).

Irrespective of the REMD method used, the required lengths of REMD simulations to ensure thorough mixing of replicas are considerable and frequently not met by REMD studies reported in literature (Abraham and Gready, 2008). Additionally, a number of commonly used methods for assessing the quality and convergence of MD simulations cannot be applied to REMD, due to the discontinuities in the trajectory at the temperature of interest (Grossfield and Zuckerman, 2009, Zhang et al., 2010). For the same reason, REMD simulations generally cannot yield information about transitions between low-energy conformations; they can only identify the conformations themselves. Finally, it has been observed that temperature-based REMD fails to sufficiently sample peptides that adsorb strongly to surfaces (Latour, 2008, Ozboyaci et al., 2016a). For this reason, it has been recommended that temperature REMD be utilised in combination with bias potential methods such as umbrella sampling (Latour, 2008) and metadynamics (Ozboyaci et al., 2016a).

2.2.3.2 Bias Potential Methods

The use of bias potentials is the primary alternative to REMD for enhancing MD sampling (Latour, 2008). Umbrella sampling (Torrie and Valleau, 1977) involves the identification of a key collective variable, X, and the addition of a bias term

$U_b(X)$ to the existing potential energy function (Kästner, 2011). Often, this takes the form of a harmonic potential constraining X to a target value, X_t:

$$U_b(X) = k(X - X_t)^2 \qquad (2\text{-}2)$$

for some constant k regulating the strength of the bias. Parallel simulations, or 'windows', may be carried out, varying k, X_t, or both (Kästner, 2011). If these values are exchanged between parallel simulations, what results is the bias potential REMD formulation referred to earlier (Fukunishi et al., 2002); these simulations are therefore sometimes referred to as replica exchange umbrella sampling (Sugita et al., 2019). The flexibility in the choice of collective variable makes umbrella sampling useful for a variety of purposes; it has been used to calculate the free energy of adsorption (Xu et al., 2018), and to investigate the impact of surface curvature on adsorption processes (Yin et al., 2019).

Metadynamics was presented by Laio and Parrinello (2002) as a means of constructing a bias potential in terms of one or more selected collective variables without any prior knowledge of the nature of the energy landscapes as a function of those variables. Based on the configurations the system has taken thus far in the simulation, the bias potential is constructed to push the system out of well-sampled regions of the collective variable space (Laio and Parrinello, 2002, Valsson et al., 2016). As such, if the collective variables constitute an adequate representation of the energy landscape, metadynamics can provide effective sampling of low-energy minima as well as transition states and regions (Pfaendtner, 2019). It is frequently combined with REMD or REST-MD to achieve improved sampling (Hughes and Walsh, 2015, Wright et al., 2015,

Hughes et al., 2017, Sprenger et al., 2018), and has even been applied to generate an entire FES of alanine dipeptide adsorbed on gold (Bellucci and Corni, 2014). The effectiveness of umbrella sampling, metadynamics and other methodologies that utilise collective variables, however, is contingent on the choice of variables. If the wrong ones are chosen, or if the energy landscape cannot easily be reduced to a computationally feasible number of variables, then the results obtained may be non-convergent or misleading (Pfaendtner, 2019).

2.2.4 Other Computational Methods

2.2.4.1 Density Functional Theory

Quantum mechanical (QM) techniques model atomic and molecular systems as a set of positively charged nuclei and negatively charged electrons, as opposed to all-atom force fields, which represent atoms as complete entities. Fundamentally, QM methods attempt to solve the Schrödinger equation given the positions of the nuclei and the spins of the electrons (Koch and Holthausen, 2001, Saura et al., 2019). Of the various approaches available for doing this (Schermann, 2008, Saura et al., 2019), the one most often applied to peptide-surface systems is density functional theory (DFT), which represents a middle ground between the accuracy of more rigorous QM techniques and the computational accessibility of all-atom molecular simulations (Ozboyaci et al., 2016a). DFT is generally considered more accurate than all-atom force fields, particularly in modelling atomic interactions (Latour, 2008, Heinz and Ramezani-Dakhel, 2016, Ramakrishnan et al., 2017). Its primary drawback is that many common approximations of DFT, which are necessary for computational efficiency, do not

explicitly take into account van der Waals dispersion forces; these must be catered for with a correction term that is generally system-dependent (Schermann, 2008, Ozboyaci et al., 2016a).

DFT is based on the theorem of Hohenberg and Kohn (1964), which states that the electron density function of a molecule, if known, is sufficient to calculate its energy. Kohn and Sham (1965) developed a practical application of this theorem for calculating the energy, which includes the exchange-correlation term E_{xc}, a function of the electron density. A number of functionals have been proposed to model this term, and the optimal choice of functional is system-dependent (Sousa et al., 2007). As well as optimised molecular structures, a variety of physical properties can be determined through DFT (Koch and Holthausen, 2001). DFT calculations can be applied to known individual structures, such as those obtained through all-atom MD (Gladytz et al., 2016, Hayamizu et al., 2016). However, with improvements in computational resources, the use of DFT calculations in performing *ab initio* MD simulations is increasing (Ozboyaci et al., 2016a). Recently, the adsorption of a peptide with as many as 12 residues was studied using DFT-MD (Poudel et al., 2017); however, the simulation lasted only 2.5 picoseconds and its utility was limited to the calculation of properties for a given state that had been optimised and locally relaxed prior to simulation.

2.2.4.2 Monte Carlo Simulations

Monte Carlo (MC) simulations offer a form of all-atom simulation alternative to MD. The general form of the MC algorithm is simple: given an initial set of atom positions, a perturbation is made, such as the displacement of an atom or molecule

(Young, 2001). The energies of the initial and final configurations are compared, and, similarly to the replica exchange method, the Metropolis criterion is applied to accept or reject the new configuration (Metropolis et al., 1953):

$$P_{ij} = min\left(1, e^{-\frac{1}{k_B T}(U_i - U_j)}\right) \quad (2\text{-}3)$$

This differs from equation 2-1 in that the temperature of interest, *T*, remains constant. Iteration of this algorithm in the infinite limit leads to the sampled configurations approaching a Boltzmann distribution, which allows for the accurate calculation of thermally averaged properties at *T* by averaging over all configurations (Metropolis et al., 1953, Frenkel and Smit, 2002). A number of enhanced sampling methods used in MD may also be applied to MC, such as umbrella sampling (Torrie and Valleau, 1977) and replica exchange (Swendsen and Wang, 1986). MC is typically less computationally intensive than MD (Young, 2001), but in studies of peptide adsorption processes, MD is typically preferred due to its generation of smooth and physically realistic trajectories, which MC does not generate (Schermann, 2008, Ramakrishnan et al., 2017). In recent times, MC peptide adsorption studies have largely been conducted using coarse-grained models (Carnal et al., 2016, Qiu et al., 2017, Li et al., 2019), which in conjunction with the computational speed of MC allow for very large systems or wide areas of configurational space to be explored, at the cost of accuracy.

2.3 Energy Landscape Mapping

As outlined in chapter 1, ‘energy landscapes’ collectively refers to the potential energy surface (PES) and free energy surfaces (FES) of an atomic or molecular

system. In practice, the FES is often calculated by averaging free energies from a MD simulation over a set of user-defined collective variables (Gruebele, 2002, Gfeller et al., 2007). This approach suffers from the inherent sampling challenges of simulation, and typically requires high-level enhanced sampling methods to be successful (Awasthi and Nair, 2019). The reduction of configurational space to collective variables, meanwhile, can lead to important minima being omitted, particularly when suitable collective variables for a system are not obvious (Caflisch, 2006). The PES, on the other hand, describes the system's potential energy as a function of the coordinates of all mobile atoms, a purpose to which all-atom force fields are well suited for molecules (Wales, 2003). In such cases, energies and gradients may be calculated for any set of coordinates, allowing the use of the minimisation and saddle point finding methods outlined in this section.

The general approach of studying a molecular system by identifying stationary points on the PES does not have a consistently and uniquely applied name in literature. Although the algorithmic location of such points is frequently referred to as 'geometry optimisation' (Baker, 1987, Wales, 2003, Schlegel, 2011), the systematic characterisation of an energy landscape according to these points represents a step beyond optimisation algorithms (Berry, 1993). The full set of stationary points is uniquely determined by the PES, and constitutes a 'map' of the surface, with any selected pair of minima linked to each other by one or more sequences of intervening saddle points and minima. This is the rationale behind the use of the term 'energy landscape mapping' in this work. This constitutes an

adoption of the terminology of Berry (1993), who first referred to the approach as ‘mapping the landscape’ in an early review.

Complete characterisation of the minima and saddle points on a PES is often not realistic, since the number of stationary points scales exponentially with the system size (Wales, 2003). Nonetheless, systematic stationary point searches allow for the construction of extensive networks of connected minima and saddle points. Such networks permit highly multidimensional energy landscapes to be represented on paper in a relatively simple and accessible manner, as discussed in Section 2.3.3, and may yield a variety of qualitative and quantitative conclusions (Wales et al., 1998, Noé and Fischer, 2008). In particular, sampling of stationary point paths between minima can yield the kinetics and dynamics of transitions, with a computational efficiency that, unlike time-based simulations, is independent of the energy barriers (Wales, 2002). This makes ELM particularly suited to studying ‘rare events’ that occur on timescales longer than simulation methods such as MD can reliably or feasibly model.

Before detailing the various aspects of ELM, it is worth summarising the early historical development of the approach. The systematic characterisation of a PES by its minima was pioneered by Hoare and Pal (1971). At this time, the development of saddle point finding methods was in its infancy (Crippen and Scheraga, 1971, McIver and Komornicki, 1972), and the systematic application of such methods had to await algorithmic improvements (Berry, 1993). The first studies to attempt to characterise entire PES by minima and saddle points without *a priori* knowledge were carried out in the late 1980s (Berry et al., 1988, Wales,

1990), and the ensuing three decades have seen further developments in algorithms, methods and analytical techniques, largely by Wales and co-workers (Wales, 2003, Wales and Bogdan, 2006, Wales, 2018). In this section, the state of the art in algorithms for constructing networks of stationary points and means of deriving relevant results and conclusions from them is outlined.

2.3.1 Finding Potential Energy Minima

2.3.1.1 Minimisation Algorithms

Minimisation techniques, of course, are ubiquitous across many areas of science and mathematics, and many such techniques have been applied to locally minimise the potential energy of molecules (Schlegel, 2011). Historically (Stillinger and Weber, 1982, Berry, 1993), steepest descent minimisation has been applied, in which a point is iteratively moved in the direction of its negative gradient (Saad et al., 2010). However, this has been superseded by faster and more efficient algorithms (Berry, 1993). The method of conjugate gradients (Hestenes and Stiefel, 1952) represents a speed-up on the steepest descent method without requiring matrix storage (Nocedal and Wright, 2006). When following conjugate gradients, one of a number of formulations for the step size may be used (Saad et al., 2010). The Polak-Ribiére update (Polak, 1971) is generally considered most efficient, despite requiring storage of an additional vector (Nocedal and Wright, 2006), and occasionally giving non-convergent results (Saad et al., 2010).

Newton's method (Nocedal, 1996, Saad et al., 2010) achieves faster convergence and improved stability by making use of the Hessian matrix of partial second

derivatives. Essentially, the surface at the point of interest is represented as a multidimensional quadratic, and the stationary point of this quadratic is chosen as the next estimate. For Newton's method to seek out a minimum, as opposed to a saddle point or maximum, the Hessian must be positive definite, *i.e.* all of its eigenvalues must be positive (Schlegel, 2011). Strategies for achieving this include initial use of the conjugate gradient method before switching to Newton's (Nocedal, 1996), or shifting the Hessian's eigenvalues using a Lagrange multiplier (Schlegel, 2011). The latter approach may be generalised to the use of different Lagrange multipliers for each eigenvalue (Wales, 1990), allowing the approach to seek out first-order saddle points as well as minima. This approach is known as 'eigenvector following', and is discussed in detail in Section 2.3.2.1.

The primary obstacle to the use of Newton-based methods in problems including many variables is the computationally intensive requirement to compute, diagonalise and store Hessians (Nocedal and Wright, 2006, Saad et al., 2010). This is an extremely relevant problem for systems involving peptides, which may contain hundreds of degrees of freedom. This challenge motivated the development of quasi-Newton methods (Dennis and Moré, 1977), which approximate the Hessian by iterative updates, rather than calculating it outright. In particular, the limited memory version of the Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS) (Dennis and Moré, 1977, Nocedal, 1980, Liu and Nocedal, 1989) does not even explicitly calculate any Hessian approximations, only requiring vector calculations and the storage of vectors for a limited number of steps. As a result, the storage requirements of the algorithm scale linearly with the

number of variables (Schlegel, 2011). Although other limited memory quasi-Newton methods are theoretically possible (Nocedal and Wright, 2006), the L-BFGS algorithm seems to see the most use due to its combination of established formulation, Newton-like efficiency and low computational cost. Since its first known application to ELM nearly two decades ago (Kumeda et al., 2001), it has remained the state of the art; recent examples of its use include Joseph et al. (2019) and Wales et al. (2019).

2.3.1.2 Sampling of Minima

The choice of minimisation algorithm only partially addresses the goal of finding minima on the PES: a means of systematically applying the selected algorithm to find many different minima is also required for effective mapping of the surface (Berry, 1993). Perhaps the most intuitive means of doing so is the 'grid search' approach: choose one or more parameters that together span a wide region of the system's configurational space (for example, the orientation and dihedral angles of a peptide), choose reasonable and reasonably spaced values for those parameters (for example, varying an angle between 0° and 360° at 30° increments), and perform minimization starting from configurations with each possible combination of values (Young, 2001, Schermann, 2008). This is effective if configurational space can reasonably be approximated by a small number of parameters, but as the number of parameters increases, the cost rapidly becomes prohibitive (Schermann, 2008). An alternative approach that has been applied is to perform a MD simulation and to apply energy minimisation to configurations obtained by this means (Stillinger and Weber, 1982, Berry, 1993). This suffers

from the same limitations as conventional MD, in that the system may not be able to overcome high energy barriers and effectively sample configurational space. This may be mitigated by the use of enhanced sampling methods, as outlined in Section 2.2.3.

Currently, the most effective means of sampling large numbers of relevant potential energy minima from across configurational space is basin hopping (BH). This technique was first proposed by Li and Scheraga (1987), and a general implementation was introduced by Wales and Doye (1997). Both groups developed BH for the purpose of global optimisation – that is, locating the lowest potential energy minimum on the surface – but the methodology is also effective in systematically producing large numbers of well-sampled minima, and has frequently been used for that purpose (Strodel and Wales, 2008, Oakley and Johnston, 2012, Somani and Wales, 2013).

BH consists of a MC simulation, as described in Section 2.2.4.2, where the energy is minimised at each step, and the minimised energies are used to determine whether to accept or reject the step. In effect, it constitutes a MC simulation on a transformed PES, in which each point is mapped to a nearby energy minimum. This creates flat regions, or ‘basins’, corresponding to areas of the surface that map to the same minimum, as depicted in Figure 2-4. The primary advantage of BH is the lack of energy barriers, which effectively addresses the sampling problems inherent in other simulation methods (Wales and Doye, 1997). Once again, techniques to further enhance sampling may be applied to BH. BH with parallel tempering (replica exchange) was introduced by Strodel et al. (2010),

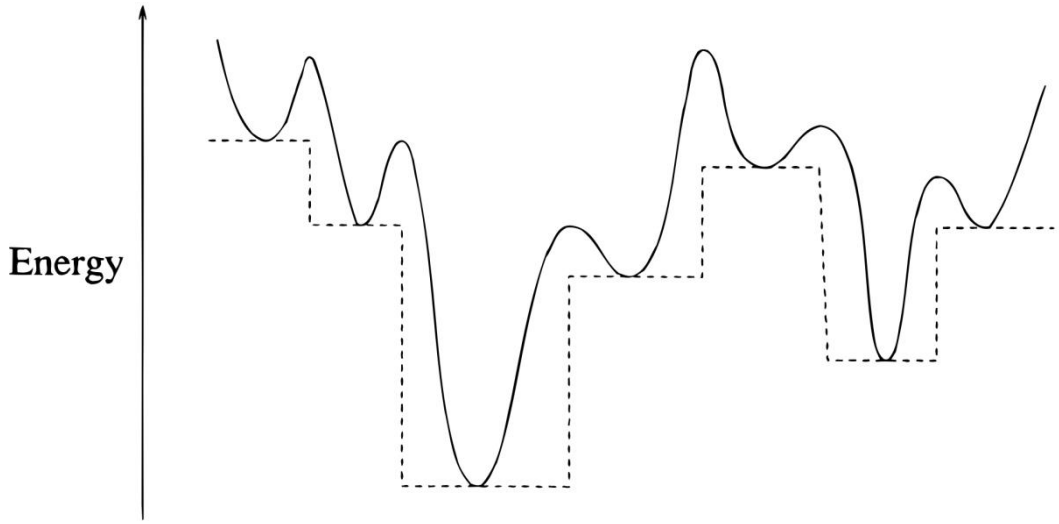

Figure 2-4. Two-dimensional schematic of a PES (solid line) and transformed PES (dotted line) for BH (Wales and Doye, 1997).

while the 'basin filling' approach of Kushima et al. (2009) is essentially BH with metadynamics.

2.3.2 Finding Potential Energy Saddle Points

Since the definition of a transition state as a first-order saddle point was first deduced half a century ago (Murrell and Laidler, 1968), many methods and algorithms for locating them have been proposed. Reviews that detail several algorithms include those of Bell and Crighton (1984), Czerminski and Elber (1990), Henkelman et al. (2002), Olsen et al. (2004), Alhat et al. (2008), Schlegel (2011), and Tao et al. (2012). In comparison to minimisation algorithms, the performance of saddle point finding algorithms relative to one another is not well established. This can partially be attributed to the sheer number of independently conceived algorithms, many of which are not often used outside of the developers' research group, and partially to a lack of high-quality comparative studies. Among

the studies dedicated to comparing different algorithms in detail (Henkelman et al., 2002, Olsen et al., 2004, Koslover and Wales, 2007, Sheppard et al., 2008, Klimeš et al., 2010), none are comprehensive, and most focus on only a handful of techniques that are of interest to the authors, and/or dated implementations of algorithms that have since undergone improvements.

Given these limitations, it is not possible to reliably conclude which of the many algorithms available is best. Instead, this section will review two of the main classes of algorithms that are most commonly used and have undergone the most development. Many others have been omitted due to their general disuse or lack of validation outside their immediate developing groups. In addition to the algorithms outlined in this section, the α-branch and bound method of Westerberg and Floudas (1999) is worth mentioning as it theoretically guarantees the location of all stationary points; however, it is only realistically applicable to very small systems.

2.3.2.1 Eigenvector Following

Eigenvector following methods have alternately been referred to as ‘surface walking’ (Simons et al., 1983, O'Neal et al., 1984), ‘uphill walks’ (Bell and Crighton, 1984, Bofill et al., 2019), ‘mode following’ (Baker, 1986, Czerminski and Elber, 1990) and other names. They are methods that, given a starting point, attempt to move towards a saddle point by taking steps along the eigenvectors of the Hessian matrix. In doing so, they exploit the concept that a first-order saddle point is a stationary point where the Hessian matrix has exactly one negative eigenvalue, and the corresponding eigenvector is the ‘lowest curvature

eigenvector' (LCE), *i.e.* the direction in which the energy is maximised (Murrell and Laidler, 1968). An illustration of an eigenvector following method moving stepwise to a saddle point is given in Figure 2-5.

The earliest example of an eigenvector following method is probably that of Crippen and Scheraga (1971), whose aim was to move from one minimum to another via a saddle point. They did so by an iterative process of computing the Hessian and its eigenvectors, taking steps 'uphill' (in the direction of increasing energy) along the LCE, and minimising the energy in all orthogonal directions, until the saddle point had been passed. Cerjan and Miller (1981) formulated this approach as an algorithm based on Newton's method for finding stationary points, as described in Section 2.3.1.1. If the eigenvectors of the Hessian are represented

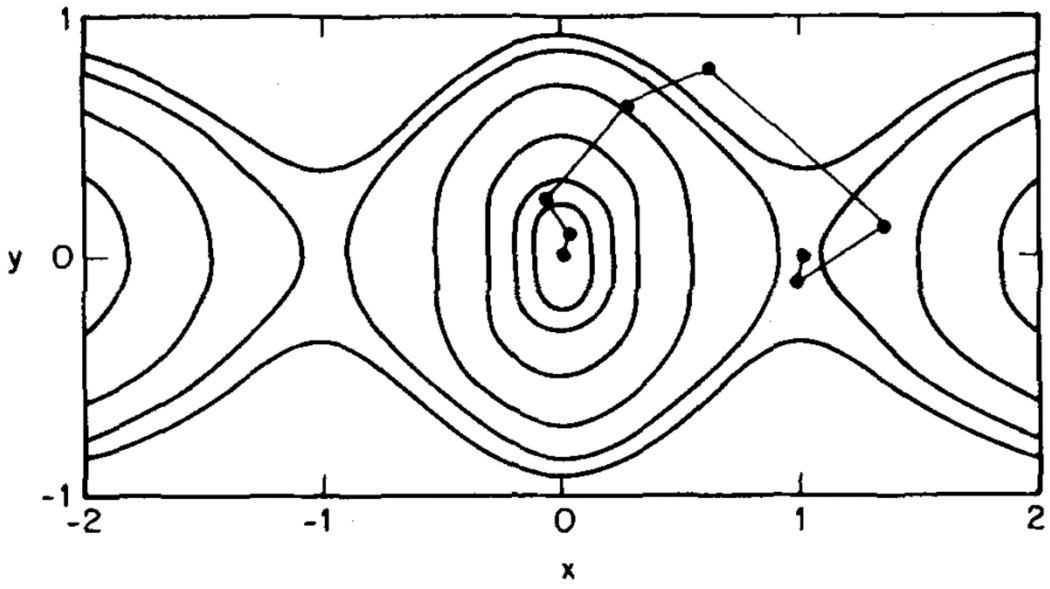

Figure 2-5. Contour plot illustrating an early eigenvector following method (Cerjan and Miller, 1981) starting from a minimum (centre) and moving to a saddle point (right).

as $\boldsymbol{v_1}$ to $\boldsymbol{v_N}$, and its corresponding eigenvalues e_1 to e_N, then the step, $\Delta\boldsymbol{x}$, taken by Newton's method is:

$$\Delta\boldsymbol{x} = \sum_{i=1}^{N} -\frac{g_i}{e_i}\boldsymbol{v_i} \qquad (2\text{-}4)$$

where g_i is the component of the gradient in the direction of $\boldsymbol{v_i}$. This guarantees that the step will be taken uphill for negative eigenvalues, and downhill for positive eigenvalues. To control which order of stationary point is sought, Cerjan and Miller introduced a Lagrange multiplier, λ, which alters the step to:

$$\Delta\boldsymbol{x} = \sum_{i=1}^{N} \frac{g_i}{\lambda - e_i}\boldsymbol{v_i} \qquad (2\text{-}5)$$

The choice of λ thus determines how many eigenvectors are followed uphill and downhill. Cerjan and Miller proposed a formulation for λ that guaranteed that exactly one eigenvector, the LCE, was followed uphill. Most subsequent eigenvector following developments have been based on alternative choices for λ (Simons et al., 1983, Banerjee et al., 1985, Nichols et al., 1990, Wales, 1990, Helgaker, 1991, Culot et al., 1992, Tsai and Jordan, 1993, Wales, 1994, Mauro et al., 2005), and/or implementing methods to limit step sizes, since large steps compromise the accuracy of the quadratic approximation (Simons et al., 1983, Baker, 1986, Nichols et al., 1990, Wales and Walsh, 1996, Khait and Puzanov, 1997, Besalú and Bofill, 1998).

As with Newton's method, eigenvector following achieves excellent convergence, but at the cost of calculating, storing and performing computations with Hessian

matrices, which becomes prohibitive for large systems (Henkelman et al., 2002). Several (Simons et al., 1983, Banerjee et al., 1985, Baker, 1986, Bofill, 1994) have addressed this by approximating the Hessian using quasi-Newton updates. Munro and Wales (1999) proposed two 'hybrid' eigenvector following schemes, the first in which the LCE is computed in a way that avoids diagonalising the Hessian, and the second not requiring the Hessian at all. In either case, after the uphill step along the LCE, the energy is subsequently minimised in all other directions, in this sense returning to the original formulation of Crippen and Scheraga (1971). Along similar lines, Bergeler et al. (2015) applied another means of estimating the LCE without diagonalising the Hessian, known as 'mode-tracking'.

A popular variant of eigenvector following that does not require Hessian calculation is the dimer method. While conceptually similar to prior eigenvector following methods, it differs significantly in implementation. In the original formulation of the dimer method introduced by Henkelman and Jónsson (1999), two points on the surface are chosen, each with a small displacement from their midpoint. The dimer is rotated about its midpoint to minimise the combined energy, which is equivalent to orienting it in the direction of the LCE, using a combination of the Newton and conjugate gradient methods. It is then translated by means of applying a modified gradient, $\boldsymbol{g}_{dimer}$:

$$\boldsymbol{g}_{dimer} = \boldsymbol{g} - 2\boldsymbol{g}_{par} \qquad (2\text{-}6)$$

where $\boldsymbol{g}_{par}$ is the component of the gradient, $\boldsymbol{g}$, parallel to the orientation of the dimer. Translating the dimer in the direction of $-\boldsymbol{g}_{dimer}$ amounts to moving it uphill

along the direction of orientation and downhill in all other directions, hence seeking out the saddle point. Subsequent developments of the dimer method have focused on reducing the number of gradient calls required (Olsen et al., 2004, Heyden et al., 2005, Kästner and Sherwood, 2008, Plasencia Gutiérrez et al., 2016), interfacing with more efficient minimisation techniques (Kästner and Sherwood, 2008, Shang and Liu, 2010, Gould et al., 2016), and the use of different techniques to calculate the LCE (Nikodem et al., 2012, Plasencia Gutiérrez et al., 2016). A variant, the shrinking dimer method, reduces the length of the dimer as it approaches the saddle point, guaranteeing convergence (Zhang and Du, 2012, Zhang et al., 2016).

It has been noted that eigenvector following methods can perform poorly when a suboptimal starting point is used, such as a minimum (Czerminski and Elber, 1990). Occasionally, in such cases, they may not converge to a saddle point (Jensen, 1995). As a result, eigenvector following is most effectively used to refine approximate saddle point estimates, such as those obtained from chain-of-states methods (Czerminski and Elber, 1990, Henkelman and Jónsson, 2000, Trygubenko and Wales, 2004, Nikodem et al., 2012, Zimmerman, 2013).

2.3.2.2 Chain-of-States Methods

In contrast with eigenvector following, which works by iteratively refining a single point, chain-of-states methods attempt to characterise the transition path, or 'minimum energy path' (MEP) between them (Koslover and Wales, 2007, Sheppard et al., 2008), representing them as a series of linked points (a 'chain of states') from one minimum to the other: an example is shown in Figure 2-6. Since

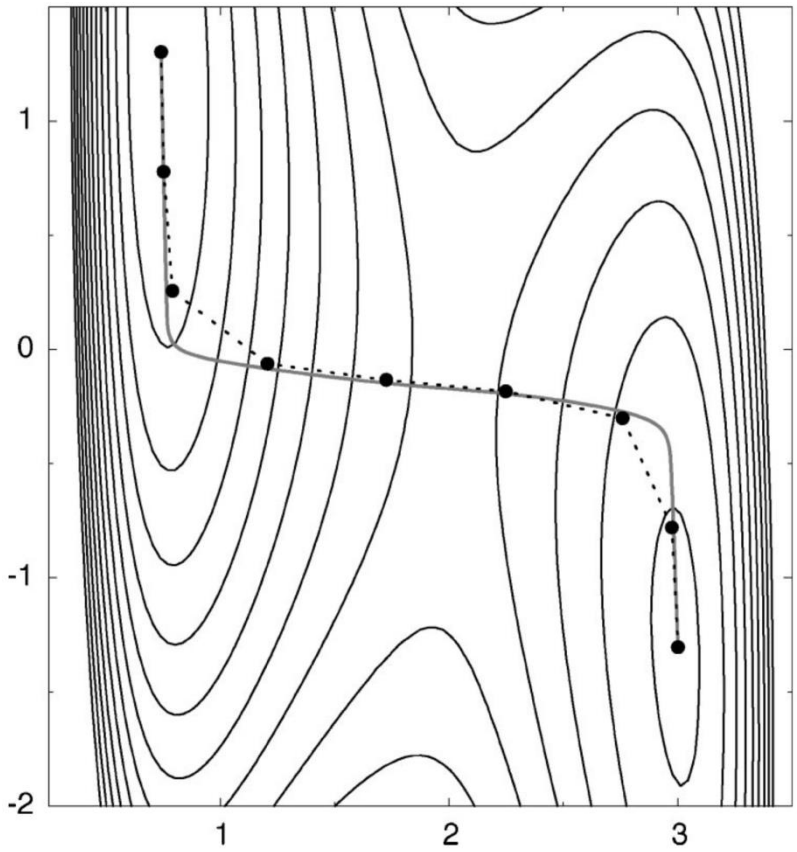

Figure 2-6. Contour plot illustrating a converged chain of states (dotted line) and the true MEP (solid line) (Henkelman and Jónsson, 2000).

the maxima along the MEP are saddle points, this makes chain-of-states methods well suited to searching for saddle points connecting a selected pair of minima. Early such methods include the synchronous transit method (Halgren and Lipscomb, 1977), the Dewar-Healy-Stewart method (Dewar et al., 1984), line integral minimisation (Elber and Karplus, 1987), and conjugate peak refinement (Fischer and Karplus, 1992). While none of these are commonly used today, aside from conjugate peak refinement, which is implemented in the CHARMM program (Brooks et al., 2009), they have inspired many contemporary chain-of-states methods.

The nudged elastic band (NEB) method, along with its variants, is probably the chain-of-states method in most widespread use today. It is implemented in both the CHARMM and AMBER packages (Brooks et al., 2009, Salomon-Ferrer et al., 2013), is included in most reviews and comparative studies (Henkelman et al., 2002, Koslover and Wales, 2007, Sheppard et al., 2008, Klimeš et al., 2010, Schlegel, 2011, Tao et al., 2012), and has undergone continual development and refinement across multiple research groups over the last twenty years. The original formulation was by Jónsson et al. (1998) and is based on the concept of optimising an 'elastic band' that stretches between the endpoint minima, with spring forces acting on each point (or 'image') along the path. The artificial gradient, $\boldsymbol{g}_{i,NEB}$, applied to the *i*th image, combines the true gradient and the spring force:

$$\boldsymbol{g}_{i,NEB} = \boldsymbol{g}_{i,perp} - [k_{i+1}(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i) - k_i(\boldsymbol{x}_i - \boldsymbol{x}_{i-1})]_{par} \qquad (2\text{-}7)$$

where $\boldsymbol{x}_i$ and $\boldsymbol{g}_i$ are respectively the coordinates and gradient of image i, $\boldsymbol{x}_0$ and $\boldsymbol{x}_{M+1}$ are the coordinates of the minima (M being the number of images), and k_1 to k_{M+1} are constants that regulate the strength of the spring forces. Shifting each image in the direction of $-\boldsymbol{g}_{i,NEB}$ thus minimises the energy of each image while keeping the elastic band relatively smooth and the images fairly equidistant. 'Perp' and 'par' indicate only the components perpendicular and parallel to the tangent of the elastic band are kept, respectively, which is necessary to keep the images from sliding down towards minima.

Subsequent development of the NEB algorithm has improved the calculation of the tangent (Henkelman and Jónsson, 2000) and the initial estimate for the path, for where linear interpolation is suboptimal (Smidstrup et al., 2014). A derivative, the climbing image NEB (CI-NEB) method (Henkelman et al., 2000), was designed to specifically optimise the elastic band near a saddle point in an attempt to locate it precisely, and further strategies for doing so were implemented by Kolsbjerg et al. (2016). Some (Crehuet et al., 2005, Lindgren et al., 2019) have found that optimising only one, or a subset, of the images at a time improves computational efficiency. Another area of improvement has been the algorithm used to optimise the elastic band. Jónsson et al. (1998) used a modified version of the velocity Verlet algorithm used in MD (Verlet, 1967, Swope et al., 1982), and alternatives have been proposed by Chu et al. (2003) and Bohner et al. (2013). Optimisation using the L-BFGS method has been shown to provide optimal efficiency (Sheppard et al., 2008), but Trygubenko and Wales (2004) noted instability in the L-BFGS method's performance with the NEB method, and proposed a 'doubly nudged' variant (DNEB) that resolves this, at the cost of never achieving full convergence. This may be resolved by switching to the regular NEB method when the DNEB path approaches convergence (Sheppard et al., 2008).

Other chain-of-states methods primarily fall into the class of 'string methods', and differ fundamentally from the NEB method in that they characterise the MEP as a continuous curve connecting the endpoint minima, rather than straight lines between images. In practice, string methods still use a series of discrete images,

but the interpolated path connecting them is presumed to be smooth and not necessarily linear. The original string method (SM) proposed by E et al. (2002) iteratively shifts all images downhill with respect to their gradient perpendicular to the path tangent (essentially equivalent to the NEB method without the spring force), and at regular intervals the images are redistributed so they are equally spaced along the path. Path interpolation and calculating tangents may be addressed using cubic splines: third-order polynomials that connect each adjacent pair of images and link smoothly with one another (Peters et al., 2004). Later, E et al. (2007) simplified their method to shift images using their entire gradient, removing the need for a tangent estimate: the issue of images collapsing towards minima as a result is resolved by the redistribution step. In the same work, they also included an energy weighting to the redistribution, reducing the distance between images close to saddle points, and demonstrated that the same climbing image strategy included in the CI-NEB method may be implemented in the SM (CI-SM).

Multiple variants of the SM have since been proposed. The growing string method (GSM) developed by Peters et al. (2004) has the same form as the SM, but begins with only the endpoints and iteratively creates new images when the perpendicular gradient of the adjoining image drops below a defined tolerance, thus ‘growing’ the string from the endpoints to the centre; this largely addresses cases where linearly interpolated images correspond to implausible configurations. This may also, or alternatively, be addressed by using non-linear interpolation methods, such as linear synchronous transit (Halgren and Lipscomb, 1977, Behn et al.,

2011b). Using a similar philosophy, Behn et al. (2011a) developed a freezing string method (FSM) that generates a pair of points at a time and converges them completely before generating the next two, again moving from the endpoints to the centre. The searching string method (SSM) (Chaffey-Millar et al., 2012) iteratively places additional images on the string in the area closest to the saddle point. Burger and Yang (2006) proposed a quadratic string method (QSM) that uses local quadratic approximations to shift the images, including a quasi-Newton updating scheme to estimate the Hessian; other formulations have used the conjugate gradient method (Goodrow et al., 2008) or the L-BFGS method (Galván and Field, 2008, Sheppard et al., 2008). Finally, variants of the SM (Ren and Vanden-Eijnden, 2013) and GSM (Zimmerman, 2015) have been developed for the purpose of finding saddle points near one minimum whilst leaving the other undetermined: one end remains fixed at the minimum while the other searches for a saddle point.

Comparison of the NEB and string methods have concluded that they have similar accuracy and efficiency (Koslover and Wales, 2007, Sheppard et al., 2008), and hybrids incorporating elements of both methods have been used (Galván and Field, 2008, Aleksandrov and Field, 2012). In general, as previously stated, the usefulness of chain-of-states methods derives from the fact that they explicitly search for saddle points between a specified pair of minima. Their weakness is that they do not necessarily find saddle points, even when fully converged to the MEP, since the set of images is discrete and the saddle point will generally lie between two of them. Alternative approaches include using eigenvector following

to refine saddle points using maximum-energy images along the MEP as an initial estimate, as outlined in the previous section, or to use the CI-NEB and CI-SM methods, which do precisely find the saddle point. The latter has the additional computational cost of refining other parts of the path that are not of interest. On the other hand, it has been observed (Goodrow et al., 2010) that saddle point estimates from chain-of-states methods are sometimes too inaccurate for eigenvector following to work reliably. Interpolation methods between images have been used to improve these estimates (Goodrow et al., 2010).

2.3.3 Visualising Stationary Point Networks

Databases of minima and saddle points found *via* the described algorithms, with each saddle point linked to the pair of adjacent minima, form a network that discretises the highly multidimensional PES. Various means of representing such networks have been proposed. In cases where the PES contains only a few stationary points, a potential energy diagram can be drawn such as in Figure 2-7(a), in which all minima, saddle points, their connectivity and their energies are described. However, such diagrams become impractical for large networks. Kunz and Berry (1995) pioneered a manner of representing and analysing a PES using sets or 'basins' of monotonic sequences: that is, all the connected series of minima and saddle points beginning at one minimum for which every subsequent minimum increases in energy. This is depicted in Figure 2-7(b). Comparing such basins can give an indication of how quickly and readily a system seeks its minimum-energy configuration (Ball et al., 1996). However, such analyses do not

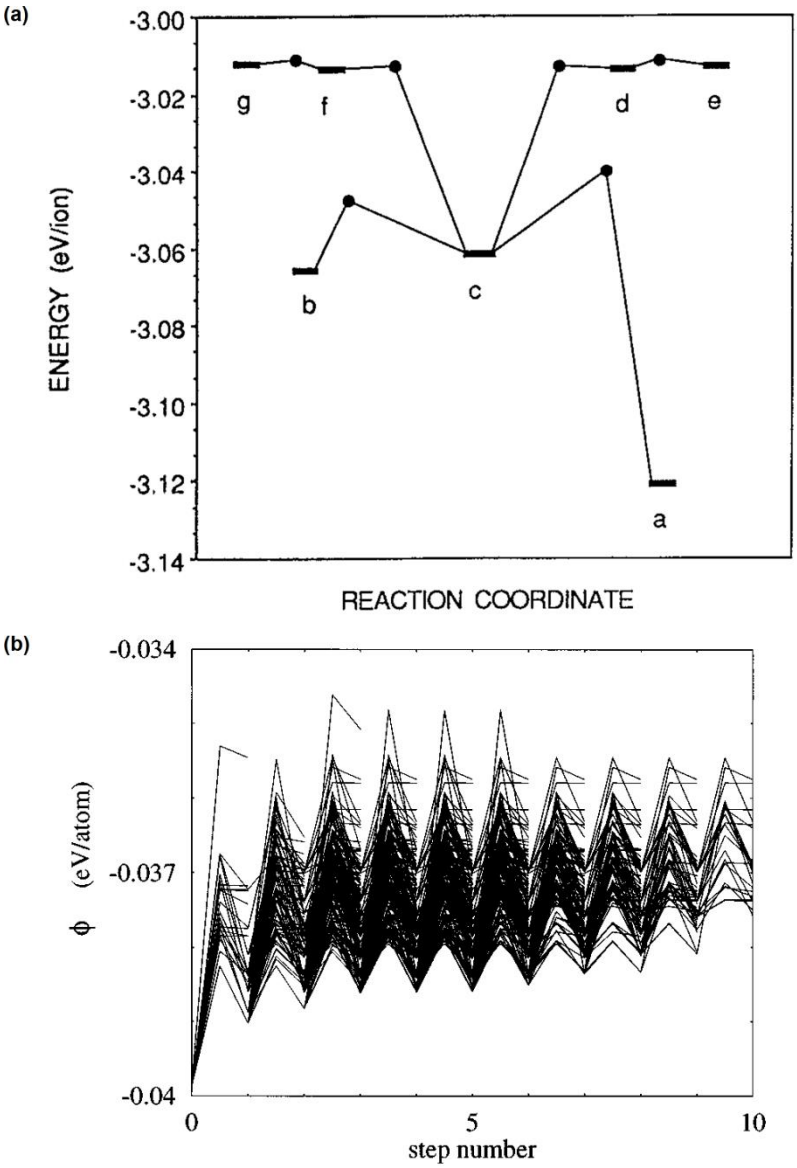

Figure 2-7. Two alternative forms of representing the PES using (a) a potential energy diagram (Rose and Berry, 1992) and (b) monotonic sequences (Kunz and Berry, 1995).

take into account the fact that there may be competing low-energy minima with separate basins (Wales, 2003).

Becker and Karplus (1997) proposed the disconnectivity graph, which has since become a popular means of visually representing energy landscapes (Wales, 2003,

Caflisch, 2006, Schermann, 2008). It was designed for illustrating PES, but is equally applicable to FES (Krivov and Karplus, 2002). An early version of the disconnectivity graph was used by Czerminski and Elber (1989). Under Becker and Karplus' formulation, the disconnectivity graph is a connected tree as illustrated in Figure 2-8(a), in which each terminus corresponds to a minimum. The *y*-axis normally corresponds to energy (Wales, 2003), but other formulations include temperature (Becker and Karplus, 1997) and mean transition time

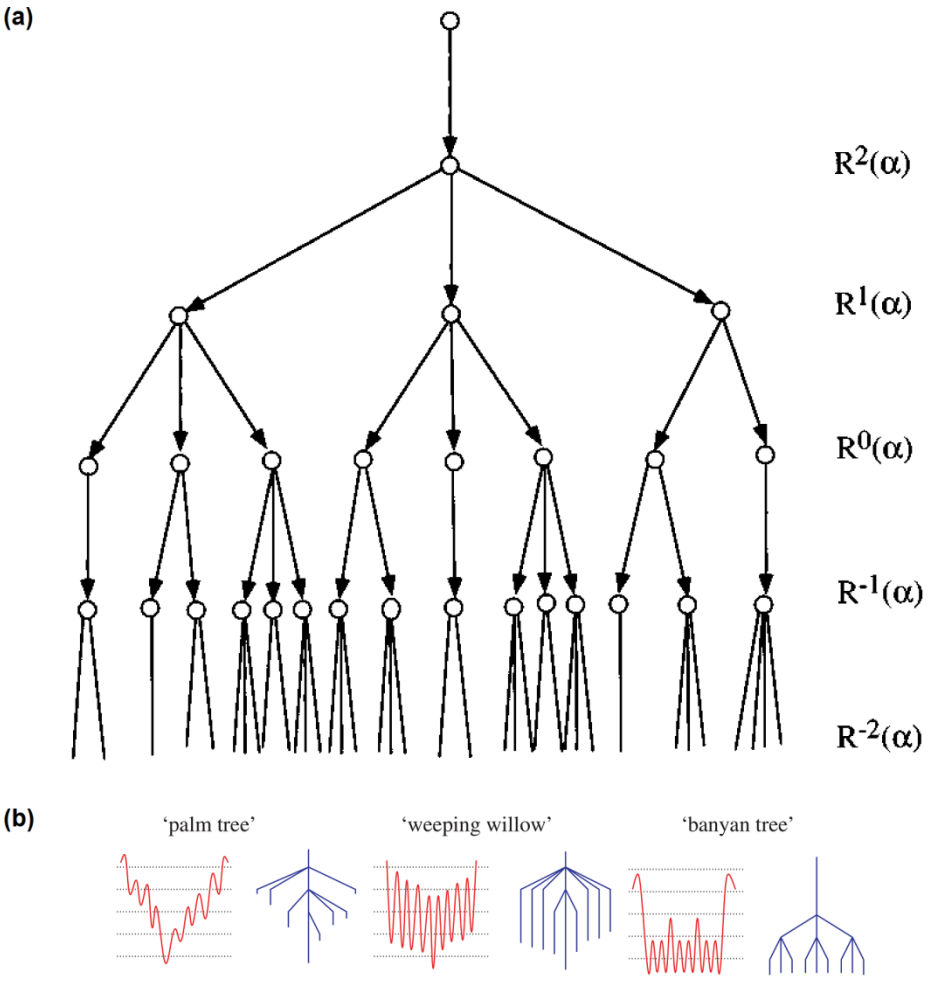

Figure 2-8. (a) Schematic of a disconnectivity graph (Becker and Karplus, 1997), and (b) some archetypal disconnectivity graph structures (Wales, 2006).

(Lempesis et al., 2013). At intervals along the *y*-axis, minima separated by an energy, temperature or transition time lower than the specified level are merged. Optionally, different parts of the energy landscape may be further illustrated by using user-specified order parameters as metrics. This may take the form of varying the thickness of lines (Komatsuzaki et al., 2005), their colour (Rylance et al., 2006), the width of the graph (Wales and Bogdan, 2006), or by utilising the width and/or depth axes of the graph as a scale (Smeeton et al., 2014).

Disconnectivity graphs are useful for their ability to yield important information about the energy landscapes and properties of a system through simple observation, which may then be supplemented with calculations (Wales, 2018). In particular, a number of archetypal shapes and structures are common amongst disconnectivity graphs, some of which are illustrated in Figure 2-8(b). In this figure, the 'palm tree' structure represents a landscape with shallow minima that drive the system to its global minimum. On the other hand, the 'weeping willow' and, especially, the 'banyan tree' pattern represent a rough landscape with many low- energy minima separated by high barriers (Wales et al., 1998, Wales, 2003). It is also possible for landscapes to have more than one of these structures, in a parallel or hierarchical manner (Wales, 2003). In such cases, one would expect a system to undergo different transitions on widely varying timescales, depending on the different energy barriers involved (Doye and Wales, 1999).

Stationary point networks may also be studied and illustrated using complex network analysis (Wales, 2003, Caflisch, 2006). In network diagrams such as illustrated in Figure 2-9, each minimum is represented by a node, and connections

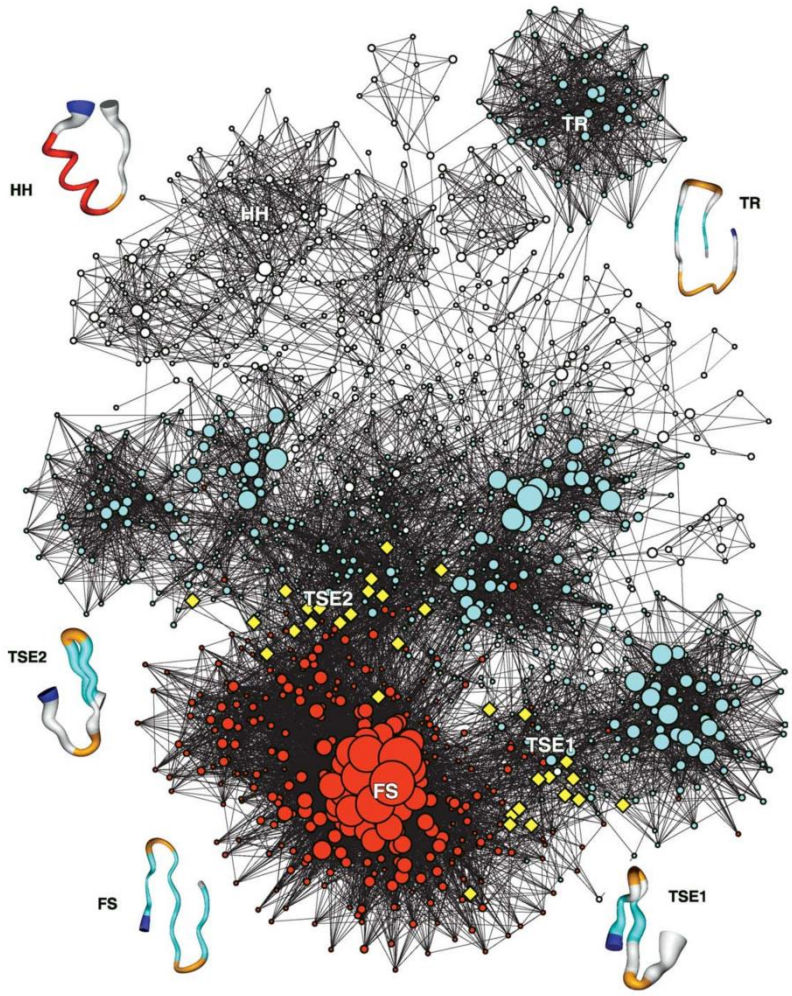

Figure 2-9. Network diagram of the conformations of a protein, with significant and well-connected minima highlighted (Rao and Caflisch, 2004).

to other nodes indicate that they are directly connected by a saddle point. Principles of complex network analysis (Newman, 2003) may then be applied; in particular, to assess the connectivity between minima (Wales, 2003, Caflisch, 2006). The primary disadvantage of network diagrams is that they cannot easily represent the energy barrier associated with a saddle point: they only indicate whether or not one exists. As such, their primary use in computational chemistry

is for assessing transitions observed in simulation (Bowman et al., 2010), not quantitative stationary points obtained through ELM.

2.3.4 Exploiting Stationary Point Networks

2.3.4.1 Discrete Path Sampling

Discrete path sampling (DPS) (Wales, 2002) was proposed as a means to study rare event transitions between a desired pair of configurations with particular relevance; for example, the native and unfolded states of a protein. Its primary goal is to calculate the rate constant between the two configurations. DPS is the conceptual analogue of transition path sampling (Dellago et al., 2002), a molecular simulation method that attempts to characterise a transition by choosing and exploiting a suitable order parameter. Unlike transition path sampling, though, DPS does not require an order parameter, instead characterising transition paths by 'discrete paths': sequences of minima and saddle points that connect the endpoints.

DPS works by defining a subset of minima as group A and another subset as group B, representing the endpoint configurations. These may be individual minima or groups of minima that all represent variants of each configuration. The rate constants are thus calculated as a summation of every discrete path between each A minimum and each B minimum, using the individual rate constants that may be calculated for any pair of minima connected by a saddle point, and weighted by the equilibrium probability that the configuration will correspond to each minimum. Intuitively, this is more likely to be true for low-energy minima,

but in reality, connectivity and energy barriers play a role. Extensive derivations for this formulation of the rate constant are provided by Wales (2002, 2003) and summarised in chapter 3.

The reliability of DPS depends on whether all relevant minima and saddle points have been found. While every stationary point makes some contribution to the rate constant calculation, those that lie on discrete paths between the states make the greatest impact. Consequently, a necessary step in DPS is to carefully probe configurational space for saddle points that will create such paths, or improve on existing ones. This typically involves both single-ended searches that try to find saddle points around known minima, typically using eigenvector following, and double-ended searches that try to find saddle points connecting a pair of known minima, typically using a chain-of-states method to obtain an initial guess that is refined by eigenvector following (Wales, 2002). In the latter case, given the sheer number of pairs of minima that could be selected, a strategy is needed for determining which minima to connect. These strategies include pairing a minimum on the current ‘best path’ (the discrete path that contributes most to the rate constant) with another minimum closest in configurational space (Wales, 2002), pairing minima on the best path that are close in configurational space but separated by intervening minima (Carr and Wales, 2005), pairing minima on the best path on either side of the largest energy barrier (Strodel et al., 2007), and pairing a minimum on the best path with another minimum separated by the largest energy barrier (Strodel et al., 2007). Strategies also exist for creating a discrete path if one does not already exist (Evans and Wales, 2004, Carr et al.,

2005, Wales and Carr, 2012), and for determining the best path (Evans and Wales, 2004, Carr and Wales, 2008, Sharpe and Wales, 2019).

Whilst DPS works by using a weighted ensemble of discrete paths and therefore does not technically predict one specific transition, transitions may nonetheless be studied by considering one of the discrete paths, most commonly the best path, as a representative. Visualising the three-dimensional configurations that correspond to each minimum and saddle point along the past yields an approximation of the transition mechanism, and plotting the potential energy as the path progresses makes clear where the highest energy barriers occur. An example is illustrated in Figure 2-10. In this manner, DPS may be used to fully characterise the mechanisms of rare events that are difficult or impossible to study using other means such as MD simulation.

2.3.4.2 Free Energy Surfaces

The true behaviour of molecular systems does not depend solely on their potential energy, but also the effects of entropy. This is particularly true of proteins and other biomolecules, for which entropy plays a significant role in determining their conformations (Gruebele, 2002). Consequently, the FES of such systems can yield more valid conclusions than the PES. The difficulty lies in computing the FES, since the temperature-dependent effects of entropy are not trivially calculated, and computation from molecular simulations face challenges as discussed previously.

Means of deriving the FES from potential energy stationary point networks were independently proposed by Krivov and Karplus (2002) and by Evans and Wales

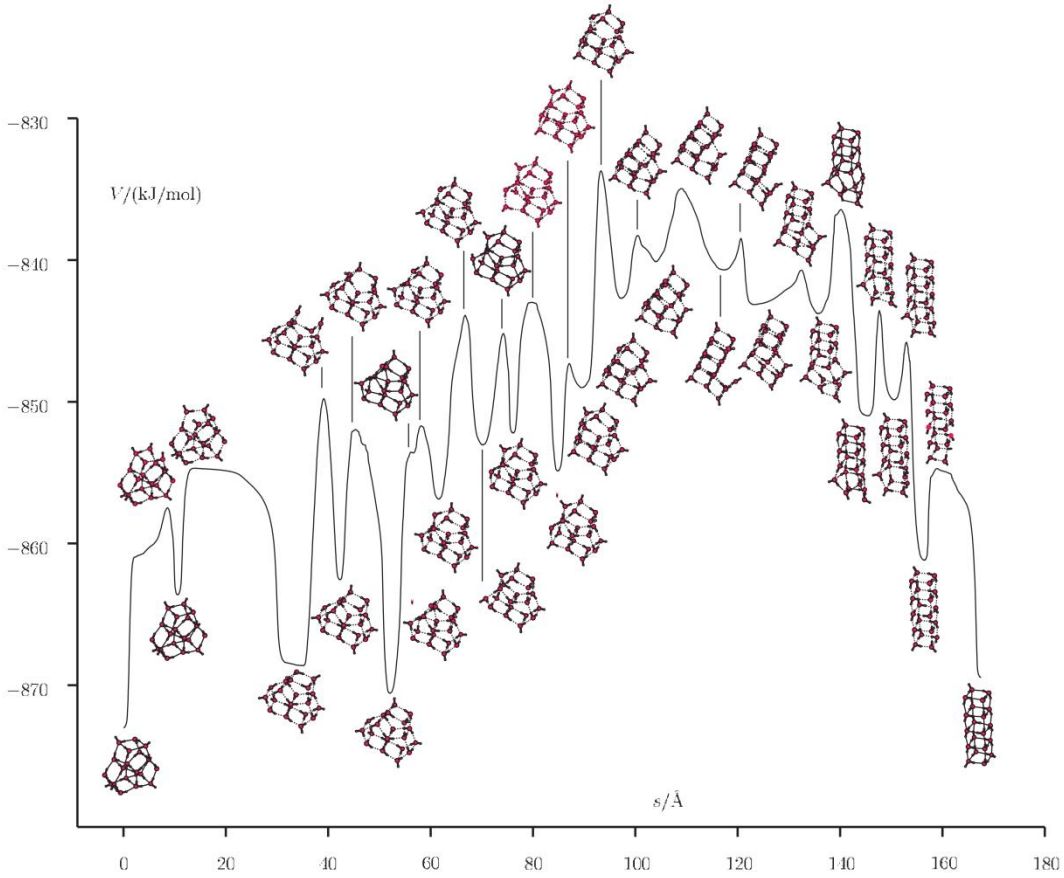

Figure 2-10. Plot of potential energy against path distance for a cluster of water molecules, with inset structures showing the transition mechanism (Wales, 2004).

(2003). The two methods are essentially equivalent. Given the potential energy difference between a minimum and adjoining saddle point, and using a harmonic approximation for the local density of states at each minimum, free energy differences between the minimum and saddle point may be estimated. Iterating this process over the entire network allows relative free energies to be calculated for every stationary point. This may then be used to represent the free energy on a disconnectivity graph, in much the same way as potential energy. Evans and Wales used the lowest energy saddle point between each pair of minima to determine free energy differences, while Krivov and Karplus compared this

approach with including contributions from different paths, and a free energy disconnectivity graph derived from molecular simulation; all were qualitatively similar. Entire FES based on collective variables have also been constructed from stationary point networks, and show good agreement with simulation (Strodel and Wales, 2008).

Free energy-based transition networks allow a variety of observables to be calculated or estimated. They directly determine equilibrium probabilities and transition times and rates (Noé and Fischer, 2008). In practice, these are difficult to compute for large networks, and methods have been proposed to overcome this (Trygubenko and Wales, 2006, Wales, 2009). Additionally, the ability to evaluate the FES at, theoretically, any temperature allows for important temperature-dependent properties to be characterised, such as the temperature at which conformational transitions occur (Wales and Bogdan, 2006), and heat capacity curves (Wales, 2017).

2.4 Potential Energy Models

Like the molecular simulation methods described in Section 2.2, ELM for peptide-surface interactions necessarily requires potential energy models for calculating the intramolecular potential of the peptide, the interaction potential between the peptide and the surface, and, if present, the effect of solvent. Currently, all-atom force fields are the most widely used models for adsorbing peptides (Latour, 2008, Martin et al., 2016), and this section will focus on those models. Coarse-grained force fields exist that further reduce the number of

degrees of freedom, such as the Martini (Monticelli et al., 2008, de Jong et al., 2012) and Gō (Hills and Brooks, 2009) models, but these are most useful for macromolecules, such as proteins, where probing timescales relevant to key events is very difficult indeed with molecular methods.

2.4.1 Modelling Intramolecular Forces

Classical all-atom force fields represent the intramolecular potential energy as a sum of terms. A typical form is the following (Best, 2019):

$$U = \sum_{\text{bonds},i} \frac{1}{2}k_{b,i}(r_i - r_{0,i})^2 + \sum_{\text{angles},i} \frac{1}{2}k_{a,i}(\theta_i - \theta_{0,i})^2 + \sum_{\text{dihedrals},i} \sum_{n} \frac{1}{2}k_{d,i,n}\cos(n\varphi_i - \varphi_{0,i}) + \sum_{\text{atoms},i,j} 4\varepsilon_{i,j}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{\text{atoms},i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2\text{-}8)$$

The five terms represent bond lengths, bond angles, dihedral angles, van der Waals forces and repulsion, and electrostatic interactions, respectively. r_i, θ_i, φ_i, and r_{ij} are the bond length, bond angle, dihedral angle and distance between atoms, and are calculated directly from the atomic coordinates. Other parameters are specific to the force field, and are determined by the type of atom or atoms involved in the bond, angle or dihedral angle being considered. For example, an aromatic carbon and an amide hydrogen are given different parameters. These parameters are usually fitted to experimental data, quantum mechanical calculations for sample molecules or both, under the assumption that the parameters are transferable to atoms of the same type in different molecules (Best,

2019). The main classical force fields are AMBER (Weiner and Kollman, 1981, Cornell et al., 1995, Wang et al., 2000, Maier et al., 2015), CHARMM (Brooks et al., 1983, MacKerell et al., 1998, Best et al., 2012), OPLS (Jorgensen and Tirado-Rives, 1988, Jorgensen et al., 1996, Kaminski et al., 2001, Harder et al., 2015), and GROMOS (Oostenbrink et al., 2004, Schmid et al., 2011). All of these force fields have undergone continual development and refinement over time, and all are applicable to peptides and other biomolecules. Other force fields exist in which the energy contribution terms are more complex and include cross-terms with one another, such as CVFF (Dauber-Osguthorpe et al., 1988), CFF (Maple et al., 1998) and COMPASS (Sun, 1998). Compared to the classical force fields, these are more computationally expensive, but provide a greater degree of accuracy for smaller systems (Latour, 2008).

2.4.2 Modelling Surface Interaction

The presence of a surface introduces two challenges to effectively modelling the system. The first is the question of how the surface should be modelled. Force fields for simulation of inorganics typically take a very different functional form compared to the organic force fields already discussed, since the main goal is to reproduce the structure of the surface, and not necessarily characterise interactions with it (Ozboyaci et al., 2016a). Often, literature parameters are used to fix the inorganic lattice and model the interactions between atoms, groups or molecules and the surface; for example, using a Lennard-Jones potential (Mahaffy et al., 1997, Penna et al., 2015, Luo et al., 2019). To reduce the computational effort required to compute interactions with each atom in the lattice, it is possible to

approximate the surface as a smooth, infinite continuum, such as in the model proposed by Steele (1973), which has been noted to produce results consistent with experiment (Steele, 1978, Qian et al., 1997).

The second challenge posed by surfaces is the suitability of general all-atom force fields to simulating peptide-surface interactions. With the increasing number of biomolecule-surface studies in the recent past, it has become apparent that classical force fields, which are parameterised for simulating peptides in the vacuum or aqueous phase, model interfacial systems very poorly (Martin et al., 2016, Walsh, 2017). Consequently, there has been considerable work in recent years towards developing specialised force fields for surface interactions with molecules. Most notably, Heinz and co-workers have developed the INTERFACE force field (Heinz et al., 2013), an all-atom force field with parameters optimised for surface interactions, and including surface models and parameters for a variety of metals and minerals based on prior and ongoing work. Similarly, Walsh and co-workers have developed a number of force fields optimising interfacial systems with metals. The first such force field was GolP-CHARMM for adsorption processes on gold (Wright et al., 2013), based on earlier work by Iori et al. (2009), and they subsequently have developed similar force fields for silver (Hughes et al., 2013) and palladium (Hughes and Walsh, 2018). Other notable force fields designed for interfacial systems include ProMetCS (Kokh et al., 2010) and the dual force field approach of Biswas et al. (2012).

2.4.3 Modelling Solvation Effects

Since solvent molecules are known to actively influence biomolecular adsorption (Jena and Hore, 2010, Penna et al., 2014, Xu et al., 2018), the reliable modelling of these effects is desired in computational studies of adsorption processes. This is typically achieved through one of two approaches: either explicit, in which water molecules are included in the system and interact according to a specialised force field, or implicit, in which terms in the intramolecular potential are added or modified to simulate the solvent's effects. Popular force fields for explicitly modelling water include SPC (Berendsen et al., 1981, Hermans et al., 1984), SPC/E (Berendsen et al., 1987), SPC/L (Glättli et al., 2002), TIP3P (Jorgensen et al., 1983), TIP4P (Jorgensen et al., 1983), TIP4P-Ew (Horn et al., 2004), and TIP4P/2005 (Abascal and Vega, 2005). Implicit solvation, meanwhile, can be achieved by a variety of means, such as by altering the dielectric constant used to model electrostatic interactions in force fields (Latour, 2008), using the Poisson–Boltzmann equation or its less costly approximation, the generalized Born model (Brooks et al., 2009), or by using a specialised model such as EEF1, which was designed for use with the CHARMM force field (Lazaridis and Karplus, 1999).

Similarly to intramolecular and surface models, the applicability of solvation models to interfacial systems is critical. Studies have shown that explicit solvent models can, if they are chosen carefully to suit the intramolecular force field, characterise adsorption processes acceptably (Walsh, 2017). However, the use of explicit solvent in ELM is not computationally feasible. Water molecules in explicit solvent systems typically account for the vast majority of the degrees of

freedom (Latour, 2008), and it follows that most of the minima and saddle points of an explicitly solvated system would correspond to rearrangements of water molecules. Implicit solvent, meanwhile, has historically been noted to model peptide and protein adsorption poorly (Sun and Latour, 2006, Sun et al., 2007). More recently, however, it was demonstrated that implicit solvation can yield accurate results when studying hydrophobic surfaces, where water is typically not present between the adsorbing molecule and the surface (Cohavi et al., 2010). Hydrophilic surfaces require specialised parameterisation and additional terms to model the solvent-surface interactions and the distorting effect of the peptide (Ozboyaci et al., 2016a). The ProMetCS force field of Kokh et al. (2010), developed for peptide/Au(111) interactions, provides one example of this. While the development of such force fields making use of implicit solvation methodologies is still in its infancy, it represents the most likely path towards accurately modelling adsorption processes without requiring an unreasonable computational load.

3 Methods

In the preceding chapter, literature pertaining to the theory and application of energy landscape mapping (ELM) was reviewed, including overviews of a number of techniques and methodologies for locating minima and saddle points, systematically creating networks and deriving transition paths, rate constants and free energy networks. In this chapter, those methods that were used in the conduction of this work are outlined in greater detail. Each of the journal papers included in this thesis contains its own methodology section, in which the methods particular to that work are stated. The purpose of this chapter is to give the reader the theoretical understanding necessary to interpret these methodology sections without requiring a detailed reading of the citations provided.

3.1 The L-BFGS Minimisation Algorithm

The limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm seeks to identify the local minimum of an unconstrained, smooth function, f, of n variables, where both the function itself and the gradient, $\boldsymbol{g}$, are known. The vector $\boldsymbol{g}$, which is of size n, contains the derivatives of f with respect to each variable. The L-BFGS algorithm was developed by Nocedal (1980) for the purpose of solving problems with a large number of variables, for which storage requirements become prohibitive for other quasi-Newton methods. It was later presented with minor modifications by Liu and Nocedal (1989). The latter version of the algorithm was used throughout the work presented here, with one revision that will be noted. In all cases, the L-BFGS method was used to minimise potential energy as a function of Cartesian coordinates.

The algorithm proceeds as follows. Let $\boldsymbol{x_0}$ be the starting point of the algorithm, m the number of prior iterations stored by the algorithm, and H_0 an initial estimate for the inverse of the Hessian matrix. In the present work, $m = 4$ and $H_0 = 0.1$ Å^{2}mol/kcal multiplied by the identity matrix, I. For the kth step, beginning at $k = 0$, the following vectors are calculated:

$$\boldsymbol{d_k} = -H_k \boldsymbol{g}(\boldsymbol{x_k}) \quad (3\text{-}1)$$

$$\boldsymbol{s_k} = \alpha_k \boldsymbol{d_k} \quad (3\text{-}2)$$

$$\boldsymbol{x_{k+1}} = \boldsymbol{x_k} + \boldsymbol{s_k} \quad (3\text{-}3)$$

$$\boldsymbol{y_k} = \boldsymbol{g}(\boldsymbol{x_{k+1}}) - \boldsymbol{g}(\boldsymbol{x_k}) \quad (3\text{-}4)$$

In the original algorithm, a line search is used to compute α_k. Here, the formulation of Trygubenko and Wales (2004) was used, in which the line search is removed and, instead, α_k is chosen so that the step size does not exceed 0.4 Å.

Then, denoting as k' the minimum of k and $m - 1$, the updated inverse Hessian is computed:

$$\begin{aligned} & H_{k+1} \\ & = \left({V_k}^{\mathrm{T}} \cdots {V_{k-k'}}^{\mathrm{T}}\right)(\gamma_k I)(V_{k-k'} \cdots V_k) \\ & + \rho_{k-k'}\left({V_k}^{\mathrm{T}} \cdots {V_{k-k'+1}}^{\mathrm{T}}\right)\boldsymbol{s_{k-k'}}\boldsymbol{s_{k-k'}}^{\mathbf{T}}(V_{k-k'+1} \cdots V_k) \\ & + \rho_{k-k'+1}\left({V_k}^{\mathrm{T}} \cdots {V_{k-k'+2}}^{\mathrm{T}}\right)\boldsymbol{s_{k-k'+1}}\boldsymbol{s_{k-k'+1}}^{\mathbf{T}}(V_{k-k'+2} \cdots V_k) + \cdots \\ & + \rho_k \boldsymbol{s_k}\boldsymbol{s_k}^{\mathbf{T}} \end{aligned} \quad (3\text{-}5)$$

where:

$$V_k = I - \rho_k \boldsymbol{y_k}\boldsymbol{s_k}^{\mathbf{T}} \quad (3\text{-}6)$$

$$\rho_k = \frac{1}{\boldsymbol{y_k}^{\mathrm{T}} \boldsymbol{s_k}} \qquad (3\text{-}7)$$

$$\gamma_k = \frac{\boldsymbol{y_k}^{\mathrm{T}} \boldsymbol{s_k}}{\boldsymbol{y_k}^{\mathrm{T}} \boldsymbol{y_k}} \qquad (3\text{-}8)$$

The algorithm then proceeds to the (k + 1)th step. The most recent m iterations of $\boldsymbol{y_k}$ and $\boldsymbol{s_k}$ are stored, as these are necessary to calculate the approximate inverse Hessian matrix. All other information from prior steps is discarded. The algorithm proceeds until the root mean square (RMS) gradient at $\boldsymbol{x_k}$ is lower than a threshold value, at which point f is considered to be at a local minimum. In the work conducted here, a threshold RMS gradient of 10^{-7} kcal/molÅ was used, except where otherwise stated.

3.2 Sampling of Minima: Basin Hopping with Simulated Annealing

Basin hopping, as formulated by Wales and Doye (1997), was implemented in conjunction with a simulated annealing procedure (SA-BH) to ensure the configuration gradually converges to a low-energy region of the potential energy surface. While basin hopping was designed for the purpose of global optimisation (that is, finding the lowest energy on the potential energy surface), here it is intended to search for low-lying minima across configurational space, providing a basis for systematically mapping the entire potential energy surface, as described in Section 3.5. To this end, many SA-BH procedures were typically run in parallel, with the aim of locating minima in as many low-lying areas of the potential energy surface as possible.

Basin hopping requires an initial configuration. In the studies conducted here, configurations obtained from past work (Mijajlovic and Biggs, 2007, Mijajlovic et al., 2011) were used. The starting structure's potential energy is first locally minimised using the L-BFGS method, which is used for all energy minimisations. At this stage, it is not necessary to locate the minimum precisely, and hence a less stringent RMS gradient convergence criterion is typically used; here, a criterion of 10^{-3} kcal/molÅ was applied.

Then, for each step, all backbone dihedral angles are perturbed randomly up to a maximum value, and the resulting configuration is locally minimised, also using the weaker RMS gradient criterion. Checks are performed to ensure that the chirality of all residues and the isomeric states of all peptide bonds have been preserved, and that the peptide remains entirely in the gas phase, above the surface. If all checks are passed, then the Metropolis criterion is applied to determine the probability, P_{ij}, that the current step is accepted, based on the potential energy of the previous configuration, U_i, that of the new configuration, U_j, and the current Boltzmann temperature of the simulation, k_BT:

$$P_{ij} = min\left(1, e^{-\frac{1}{k_BT}(U_i - U_j)}\right) \qquad (3\text{-}9)$$

If the step is accepted, then the new configuration is used for the following step. If it is rejected, then the current configuration is retained. It may be noted that equation 3-9 is identical to the formulation of the Metropolis criterion for standard Monte Carlo techniques, as expressed in equation 2-3. The difference between the methods is the local minimisation after each basin hopping step.

The number of steps is specified prior to the simulation. In the studies presented here, the number of steps chosen depended on the size of the system and the computational resources available, but typically, a number in the order of 10^5 was considered reasonable. The maximum perturbation of backbone dihedral angles was initially set to ±360°. This is double the maximum perturbation of ±180° required for complete randomisation of secondary structure, which ensured that the structure was randomised for at least several hundred steps at the beginning of the simulation until the maximum perturbation was decreased to below ±180°. After every 50 steps, if the ratio of accepted steps over the preceding 50 steps was greater than 0.5, the maximum perturbation was increased by 5%, or decreased by 5% otherwise. Simulated annealing is involved through initially setting the Boltzmann temperature to k_BT = 5.0 kcal/mol, and decrementing it by a small percentage after each step. This percentage was chosen according to the number of steps, with the aim of achieving a Boltzmann temperature of approximately 0.1 kcal/mol by the final step of the simulation.

When the simulation has run to completion, all of the accepted configurations are subjected to energy minimisation, this time using the stringent RMS gradient criterion of 10^{-7} kcal/molÅ. The converged local minima obtained by this means from multiple SA-BH procedures are amalgamated into a single database of minima, from which systematic searches for saddle points are conducted.

3.3 Finding Approximate Saddle Points: The DNEB Method

The doubly nudged elastic band (DNEB) method of Trygubenko and Wales (2004) was used for double-ended saddle point searches, where the saddle point(s) directly connecting a designated pair of minima are desired. The DNEB method approximates the minimum energy path (MEP) through a series of points on the potential energy surface, and in general one or more of those points will be used as a starting point for an exact saddle point search, the procedure for which is described in Section 3.4. The DNEB method is designed for stability when combined with powerful minimisation algorithms, such as the L-BFGS algorithm, with which some alternative methodologies are unstable (Trygubenko and Wales, 2004).

Initially, a series of M points or 'images' are placed on the potential energy surface, with coordinates denoted by $\boldsymbol{x}_1$ through $\boldsymbol{x}_M$, by linear interpolation between the endpoint minima. In the work conducted here, M = 10 was chosen. The coordinates of the minima are denoted $\boldsymbol{x}_0$ and $\boldsymbol{x}_{M+1}$. The gradients of all images, $\boldsymbol{g}_1$ through $\boldsymbol{g}_M$, are calculated. A spring force is also applied to each image, pulling it towards the two adjacent images with the following gradient:

$$\boldsymbol{g}_{i,spr} = -k[(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i) - (\boldsymbol{x}_i - \boldsymbol{x}_{i-1})] \qquad (3\text{-}10)$$

In this equation, k is a user-defined spring constant. In the work conducted here, k = 1 kcal/molÅ^2 was used.

The DNEB method works by minimising with respect to a combination of both the true and spring gradients, which are split into components parallel and

perpendicular to a vector $\boldsymbol{\tau_i}$, which is a unit vector tangent to the elastic band at image *i*. The perpendicular component of the true gradient is retained, as is the parallel component and a portion of the perpendicular component of the spring gradient. The DNEB gradient for each image is given by:

$$\boldsymbol{g_{i,DNEB}} = \boldsymbol{g_i} - (\boldsymbol{g_i} \cdot \boldsymbol{\tau_i})\boldsymbol{\tau_i} + \boldsymbol{g_{i,spr}} - \{[\boldsymbol{g_{i,spr}} - (\boldsymbol{g_{i,spr}} \cdot \boldsymbol{\tau_i})\boldsymbol{\tau_i}] \cdot \hat{\boldsymbol{g}}_i\}\hat{\boldsymbol{g}}_i \quad (3\text{-}11)$$

where $\hat{\boldsymbol{g}}_i$ is the unit vector parallel to the perpendicular component of the real gradient:

$$\hat{\boldsymbol{g}}_i = \frac{\boldsymbol{g_i} - (\boldsymbol{g_i} \cdot \boldsymbol{\tau_i})\boldsymbol{\tau_i}}{|\boldsymbol{g_i} - (\boldsymbol{g_i} \cdot \boldsymbol{\tau_i})\boldsymbol{\tau_i}|} \quad (3\text{-}12)$$

Meanwhile, $\boldsymbol{\tau_i}$ is approximated by the direction to the adjacent image with higher potential energy, *U*, if there is exactly one such image:

$$\boldsymbol{\tau_i} = \begin{cases} \boldsymbol{\tau_i}^+ = \dfrac{\boldsymbol{x_{i+1}} - \boldsymbol{x_i}}{|\boldsymbol{x_{i+1}} - \boldsymbol{x_i}|} & \text{if } U_{i+1} > U_i > U_{i-1} \\ \boldsymbol{\tau_i}^- = \dfrac{\boldsymbol{x_i} - \boldsymbol{x_{i-1}}}{|\boldsymbol{x_i} - \boldsymbol{x_{i-1}}|} & \text{if } U_{i+1} < U_i < U_{i-1} \end{cases} \quad (3\text{-}13)$$

If both adjacent images have higher potential energy or both have lower, a linear combination of the two directions is used, as proposed by Henkelman and Jónsson (2000):

$$\boldsymbol{\tau_i} = \begin{cases} \Delta U_{max}\boldsymbol{\tau_i}^+ + \Delta U_{min}\boldsymbol{\tau_i}^- & \text{if } U_{i+1} > U_{i-1} \\ \Delta U_{max}\boldsymbol{\tau_i}^- + \Delta U_{min}\boldsymbol{\tau_i}^+ & \text{if } U_{i+1} < U_{i-1} \end{cases} \quad (3\text{-}14)$$

where:

$$\Delta U_{max} = \max(|U_{i+1} - U_i|, |U_i - U_{i-1}|) \quad (3\text{-}15)$$

$$\Delta U_{min} = \min(|U_{i+1} - U_i|, |U_i - U_{i-1}|) \quad (3\text{-}16)$$

The L-BFGS algorithm is used to minimise with respect to the DNEB gradient, $\boldsymbol{g_{DNEB}}$, which consists of the gradient on each image, $\boldsymbol{g_{i,DNEB}}$, as formulated in equation 3-11. It is thus a vector of length nM, where n is the number of degrees of freedom on the potential energy surface. The objective function corresponding to $\boldsymbol{g_{DNEB}}$ is unknown, but is not required, since the L-BFGS algorithm only requires the gradient. Since precise convergence of the MEP is not necessary for an adequate estimate for the saddle point(s), minimisation of the elastic band using the L-BFGS algorithm generally requires a weak criterion. In the work conducted here, the procedure concluded when the RMS gradient was less than 10^{-2} kcal/molÅ, or after 300 iterations of the L-BFGS algorithm, whichever occurred first. At this point, all images with potential energy higher than both of the adjacent images are used as starting points for exact saddle point searches.

3.4 Finding Exact Saddle Points: Hybrid Eigenvector Following

Exact saddle point searches in this work utilised the hybrid eigenvector following (HEF) algorithm proposed by Munro and Wales (1999). True eigenvector following diagonalises the Hessian matrix to calculate all eigenvalues and eigenvectors, and takes steps along each eigenvector. However, when searching for a first-order saddle point, only the lowest eigenvalue, e_{min}, and the corresponding eigenvector, $\boldsymbol{v_{min}}$, need be calculated, as this is the direction in which a step is taken to increase the potential energy; all other eigenvectors are followed in the direction of decreasing energy, and thus may effectively be replaced by minimising the energy in the space perpendicular to $\boldsymbol{v_{min}}$.

e_{min} and $\boldsymbol{v_{min}}$ may be calculated by a variety of means that do not involve the computationally intensive task of diagonalising the Hessian. Two methods proposed by Munro and Wales (1999) involve iteratively pre-multiplying a vector by the Hessian, and minimising the Rayleigh-Ritz ratio; the latter does not require the Hessian at all. In the version of the algorithm utilised here, e_{min} and $\boldsymbol{v_{min}}$ are calculated using the DSYEVR routine in the LAPACK package of linear algebra subroutines (Anderson et al., 1999), which is based upon the algorithm of Dhillon (1997).

When e_{min} and $\boldsymbol{v_{min}}$ are known, the size of the step, h, along $\boldsymbol{v_{min}}$ is calculated. If e_{min} is positive, indicating a saddle point is not close, a large step is used, for which h = 0.5 Å was chosen here. Otherwise, the step is calculated by:

$$h = \min\left(h_{max}, \frac{2\boldsymbol{g}\cdot\boldsymbol{v_{min}}}{|e_{min}|\left[1+\sqrt{1+4\left(\frac{\boldsymbol{g}\cdot\boldsymbol{v_{min}}}{e_{min}}\right)^2}\right]} \right) \qquad (3\text{-}17)$$

where $\boldsymbol{g}$ is the gradient and $\boldsymbol{v_{min}}$ is normalised to unit vector size. The maximum step size, h_{max}, was initially set here to 0.2 Å. At each step, h_{max} is multiplied by 1.1 if a trust radius, r, is less than 2, and divided by 1.1 if r exceeds 2, while always constraining the value between 0.01 Å and 0.5 Å, using the formulation of Wales and Walsh (1996):

$$r = \left| \frac{[\boldsymbol{g}\cdot\boldsymbol{v_{min}}]_{current} - [\boldsymbol{g}\cdot\boldsymbol{v_{min}}]_{previous}}{h_{previous} e_{min,current}} - 1 \right| \qquad (3\text{-}18)$$

After the step h is taken, energy minimisation using the L-BFGS algorithm is conducted in the space perpendicular to $\boldsymbol{v_{min}}$, using the adjusted gradient:

$$\boldsymbol{g}_{perp} = \boldsymbol{g} - (\boldsymbol{g} \cdot \boldsymbol{v}_{min})\boldsymbol{v}_{min} \quad (3\text{-}19)$$

The above steps are iterated until *h* and the RMS gradient are both less than specified thresholds: here, those thresholds were 0.02 Å and 10^{-7} kcal/molÅ, respectively. At this point, the current coordinates are deemed to correspond to a saddle point. Additionally, in this work, a maximum of 50,000 iterations were permitted; if, at this point, the algorithm had not converged to a saddle point, the search was concluded unsuccessfully.

3.5 Systematic Construction of Stationary Point Networks

The construction of stationary point networks requires the storing and updating of databases of known minima and saddle points. These databases store the coordinates of the points, along with other necessary information. The potential energy of each point is retained in the database, for use in constructing disconnectivity graphs, and calculating rate constants and free energies as outlined in Section 3.6. Other parameters necessary for rates and free energies are also retained, such as vibrational frequencies calculated from the positive Hessian eigenvalues. In addition to this information, for saddle points, the identities of the adjoining minima are also stored. Since each saddle point is a transition state on the minimum-energy path between two minima (Murrell and Laidler, 1968), steepest descent paths starting from the saddle point in either direction along $\boldsymbol{v}_{min}$ will lead to these adjoining minima. These are approximated by perturbing the saddle point by a small step, 0.01 Å in this work, in both the positive and negative directions of $\boldsymbol{v}_{min}$, and in each case then applying L-BFGS minimisation to locate a minimum. The L-BFGS algorithm will not generally follow the steepest descent

path, but will normally find the same minimum at a considerably lower computational cost than steepest descent methods (Wales, 2006). It is possible that minima found in this manner will be previously unknown; any such minima are added to the existing database.

In order to effectively build the network, methods are required to determine appropriate starting points for saddle point searches using the DNEB or HEF algorithms. A number of such methods were used throughout the studies conducted here. The NEWCONNECTIONS scheme involves searching for saddle points in the region of configurational space near known minima. A minimum is selected and all of its Cartesian coordinates are randomly perturbed by up to a small user-specified threshold, for which 0.01 Å was used here. This is used as a starting point for a saddle point search using HEF. This may be repeated a specified number of times for every minimum in the database. In the work conducted here, for the purpose of computational feasibility, a threshold potential energy (dependent on the system and the size of the database) was chosen, and saddle point searches were only attempted for minima with potential energy below this threshold.

Other methods build the network by attempting to determine the optimal transition path (as represented by a connected series of minima and saddle points) between two pre-defined minima. These may be structures that are of interest to the study, for which it is desirable to calculate the transition path between them. The endpoints may also be minima of low potential energy that are distant in configurational space, in which case probing the region between them may yield

important information about the energy landscape. This applies particularly if the minima currently are separated by a high potential energy barrier, or if no connecting path of minima and saddle points has yet been found.

If the latter is the case, then the DIJINIT scheme is applicable (Carr et al., 2005). In this scheme, the entire set of known minima is represented as a graph in which each pair of minima, *i* and *j*, is connected by edges of the following weights:

$$w(i,j) = \begin{cases} 0 & \text{if minima } i \text{ and } j \text{ are directly linked by a saddle point} \\ \infty & \text{if } n_{max} \text{ attempts have been made to link minima } i \text{ and } j \\ d(i,j)^2 & \text{otherwise} \end{cases} \quad (3\text{-}20)$$

where *d(i,j)* is the Euclidean distance between the minima, minimised with respect to orientation and permutational isomers. The algorithm of Dijkstra (1959) is used to determine the shortest path between the endpoint minima, based on the edge weights defined above. For any edges along that path that are non-zero in weight, the DNEB method, followed by HEF on approximate saddle points thus found, is used to attempt to connect each such pair of minima. Since this procedure is deterministic for a given pair of minima, any further attempt to connect the same minima is redundant, and hence $n_{max} = 1$ is chosen. When all of these procedures have been completed and new saddle points and minima added to the database, the edge widths are recalculated and the procedure is repeated. The scheme concludes when Dijkstra's algorithm locates a path of zero total length, indicating a connected path of stationary points between the endpoint minima.

When such a connected path exists, a variety of schemes may be used to search for further saddle points in the region of configurational space between the endpoint minima. In the SHORTCUT scheme (Carr and Wales, 2005), the 'best path', which is the sequence of stationary points making the largest contribution to the steady state rate constant between the minima as outlined in Section 3.6.2, is computed. The scheme then selects pairs of minima on the best path for a connection attempt using DNEB/HEF. In this manner, SHORTCUT attempts to make improvements to the best path by finding 'shortcuts': saddle points linking minima that previously were separated on the path by multiple intervening minima. Since in most cases in the work conducted here, the best path consisted of not more than approximately 30 minima, the SHORTCUT scheme was generally allowed to run to completion, *i.e.* attempting to connect every pair of minima on the path.

The DIJPAIR and UNTRAP schemes (Strodel et al., 2007) both attempt to lower energy barriers that are artificially high due to a high saddle point previously being found, and thus construct new viable paths alternative to the putative best path. DIJPAIR calculates the best path between one of the endpoint minima and every other minimum that is connected to it by a series of stationary points. Pairs of minima on these paths that are separated by the highest potential energy barrier are selected for connection using DNEB/HEF. UNTRAP, meanwhile, directly calculates the potential energy barriers between an endpoint minimum and each other connected minimum, and attempts to connect those pairs for which the barrier is highest. Both of these schemes were found to be less efficient than

SHORTCUT, and could not feasibly be run to completion for the systems studied here due to the large number of connected stationary points. Nonetheless, they were effective in correcting a bad initial guess for the best path.

3.6 Analysing Stationary Point Networks

3.6.1 Rate Constants

The calculation of rate constants and transition paths in the work conducted here followed the discrete path sampling (DPS) procedure of Wales (2002), using transition state theory calculations for which full derivations are provided elsewhere (Wales, 2003). In summary, a harmonic approximation to the local density of states at any given stationary point yields, in the canonical ensemble which is used here, the following equation for the rate constant at which the system leaves a minimum *i* through a saddle point *t*:

$$k_{ti} = \frac{o_i}{o_t} \frac{{\nu_i}^n}{{\nu_t}^{n-1}} e^{-\frac{1}{k_B T}(U_t - U_i)} \qquad (3\text{-}21)$$

In this equation, o_i and o_t are the point group orders of *i* and *t*, respectively; for adsorbed peptides, where each structure corresponds to a unique minimum, these are reduced to unity and may be ignored. The quantities ν_i and ν_t are the geometric mean vibrational frequencies, U_i and U_t the potential energies, *n* the number of degrees of freedom, and $k_B T$ the Boltzmann temperature at which the rates are calculated. The two-state rate constant, k_{ji}, between minima *i* and *j* is the sum of the rate constants, k_{ti}, for all saddle points, *t*, that directly connect *i* and *j*.

DPS comprises characterising two endpoint configurations by separate groups of minima, group A and group B. Provided these two groups are connected by at least one sequence of connected minima and saddle points, the overall steady state rate constants, k_{AB} for transitions from group B to group A and k_{BA} for the reverse, may be estimated as summations over all paths between any minima *a* in A and *b* in B:

$$k_{AB} = \frac{1}{{P_B}^{eq}} \sum_{a \leftarrow b} \frac{k_{ai_1} k_{i_1 i_2} \cdots k_{i_N b} {P_b}^{eq}}{\sum_{\alpha_1} k_{\alpha_1 i_1} \sum_{\alpha_2} k_{\alpha_2 i_2} \cdots \sum_{\alpha_N} k_{\alpha_N i_N}} \qquad (3\text{-}22)$$

$$k_{BA} = \frac{1}{{P_A}^{eq}} \sum_{b \leftarrow a} \frac{k_{bi_1} k_{i_1 i_2} \cdots k_{i_N a} {P_a}^{eq}}{\sum_{\alpha_1} k_{\alpha_1 i_1} \sum_{\alpha_2} k_{\alpha_2 i_2} \cdots \sum_{\alpha_N} k_{\alpha_N i_N}} \qquad (3\text{-}23)$$

In these equations, i_1 through i_N are intervening minima on a path between *a* and *b*, and α_1 through α_N are counters over all minima directly connected *via* a saddle point to i_1 through i_N, respectively. P_A^{eq} and P_B^{eq} are the probabilities that the system is in group A and B, respectively, while P_a^{eq} and P_b^{eq} are the corresponding probabilities for the individual minima, *a* and *b*. These occupational probabilities may be calculated directly from the two-state rate constants using a master equation approach (Wales, 2003).

In practice, calculating the rate constant in this manner becomes difficult for large systems with a large number of possible paths, and can lead to numerical problems. The graph transformation (GT) method (Trygubenko and Wales, 2006, Wales, 2009) provides a means of overcoming this. Firstly, equations 3-22 and 3-23 are rewritten in terms of branching probabilities (the probability, P_{ji}, that the system's next move from minimum *i* to another minimum will be to minimum *j*)

and mean waiting times (the expected time, τ_i, before the system moves out of minimum i):

$$k_{AB} = \frac{1}{{P_B}^{eq}} \sum_{a \leftarrow b} \frac{P_{ai_1} P_{i_1 i_2} \cdots P_{i_N b} {P_b}^{eq}}{\tau_b} \quad (3\text{-}24)$$

$$k_{BA} = \frac{1}{{P_A}^{eq}} \sum_{b \leftarrow a} \frac{P_{bi_1} P_{i_1 i_2} \cdots P_{i_N a} {P_a}^{eq}}{\tau_a} \quad (3\text{-}25)$$

$$P_{ji} = \frac{k_{ji}}{\sum_\gamma k_{\gamma i}} \quad (3\text{-}26)$$

$$\tau_i = \frac{1}{\sum_\gamma k_{\gamma i}} \quad (3\text{-}27)$$

GT comprises iteratively removing intervening minima, and renormalising the branching probabilities and waiting times. If minimum α is removed, then the following renormalisations occur for all minima β connected to α, and all minima γ connected to either α or β, excluding α itself:

$${P_{\gamma\beta}}' = P_{\gamma\beta} + \frac{P_{\gamma\alpha} P_{\alpha\beta}}{1 - P_{\alpha\alpha}} \quad (3\text{-}28)$$

$${\tau_\beta}' = \tau_\beta + \frac{P_{\alpha\beta} \tau_\alpha}{1 - P_{\alpha\alpha}} \quad (3\text{-}29)$$

When all minima except those in groups A and B have been removed, the rate constants are calculated by:

$$k_{AB} = \frac{1}{{P_B}^{eq}} \sum_{b \in B} \frac{{P_b}^{eq}}{\tau_b} \sum_{a \in A} {P_{ab}}' \quad (3\text{-}30)$$

$$k_{BA} = \frac{1}{{P_A}^{eq}} \sum_{a \in A} \frac{{P_a}^{eq}}{\tau_a} \sum_{b \in B} {P_{ba}}' \quad (3\text{-}31)$$

Here, P_{ab}' and P_{ba}' are the final branching probabilities, which will generally have been renormalised multiple times. If the original waiting times are used (τ_a and τ_b as calculated using equation 3-27), then the steady state rate constants are retained, while using the final waiting times, τ_a' and τ_b' (calculated using equation 3-29), yield non-steady state rate constants that consider only the states corresponding to groups A and B.

3.6.2 Best Paths

In the work conducted here, particular emphasis was given to the 'best path'; that is, the sequence of connected stationary points that make the largest contribution to the steady state rate constant. The best path corresponds to the most likely transition pathway between two structures, and is thus of qualitative and quantitative interest. In the original formulation of Wales (2002), best paths were evaluated by considering every possible path and calculating the contribution of each to the overall rate constant, with the use of an adjacency matrix to include 're-crossings', *i.e.* instances of the system moving back and forth along the path before eventually reaching the endpoint.

For large networks, however, the number of distinct discrete paths makes this approach unfeasible. An alternative approach is to use the shortest path algorithm of Dijkstra (1959) with the weights proposed by Evans and Wales (2004) for any two minima, *i* and *j*, connected by a saddle point:

$$w(i,j) = \begin{cases} \ln \dfrac{{P_b}^{eq}}{{P_B}^{eq}} k_{ij} & \text{if } i \text{ is in A and } j \text{ is in B} \\ \ln k_{ij} & \text{if } i \text{ is in A and } j \text{ is in neither A nor B} \\ \ln \dfrac{k_{ij}}{\sum_\gamma k_{\gamma i}} & \text{if } i \text{ and } j \text{ are in neither A nor B} \\ \ln \dfrac{{P_b}^{eq}}{{P_B}^{eq}} \dfrac{k_{ij}}{\sum_\gamma k_{\gamma i}} & \text{if } i \text{ is in neither A nor B and } j \text{ is in B} \end{cases} \quad (3\text{-}32)$$

Summing the weights of a discrete path from a minimum in B to a minimum in A will yield the natural logarithm of that path's contribution to k_{AB}, and hence the path with the largest weight will correspond to the path with the greatest contribution. The reverse formulation yields the path with the greatest contribution to k_{BA}, which is not necessarily the same, but in practice the paths were usually found to be identical.

In the work conducted here, this algorithm was used for formulating putative best paths for use in the SHORTCUT and DIJPAIR schemes. However, it was noted that, when calculating the rate constant for the individual path using the GT approach outlined above, the algorithm sometimes did not find the path with the highest rate constant. For this reason, when determining the best path for the analysis and presentation of results, the KSHORTESTPATHS algorithm (Jiménez and Marzal, 1999) was used instead to determine a list of 10,000 paths with the largest weight according to the above formulation. Rate constants were calculated for each of these paths and that with the highest rate constant was considered the best path.

3.6.3 Free Energies

Free energy transition networks are calculated according to the formulation of Evans and Wales (2003). In brief, it involves a similar derivation to the calculation of rate constants, using the same harmonic approximation of the local density of states at each stationary point. Given the rate for leaving a minimum i through a saddle point t, k_{ti} as defined above, the free energy difference between that minimum and saddle point is calculated as follows:

$$G_i - G_t = k_B T \ln \frac{h k_{ti}}{k_B T} \qquad (3\text{-}33)$$

G_i and G_t are the free energies of the minimum and saddle point, respectively, and h is Planck's constant. Using this equation, beginning from a reference point (typically the global potential energy minimum, which is set to a free energy of zero), the free energies of all other stationary points are recursively calculated. Typically, a variety of temperatures were chosen for free energy calculations, in order to elucidate the temperature-dependence of the free energy surface.

When recalculating free energies, potential energy minima are grouped together if the two-state rate constants connecting them are higher than a threshold rate, for which 10^9 s^{-1} was used here. Such a rate constant indicates that the configurations corresponding to those potential energy minima readily interconvert on a nanosecond time-scale at the temperature of interest. The free energies of merged groups of potential minima and saddle points connecting them are calculated using a summation of the natural logarithms of their partition functions (Evans

and Wales, 2003), thus preserving the contribution of the local entropy of each minimum and the configurational entropy of the multiple minima.

Statement of Authorship

Title of Paper	Characterizing the Switching Transitions of an Adsorbed Peptide by Mapping the Potential Energy Surface.
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Ross-Naylor, J.A., Mijajlovic, M., Zhang, H. & Biggs, M.J. 2017. J. Phys. Chem. B, 121, 11455-11464.

Principal Author

Name of Principal Author (Candidate)	James Ross-Naylor
Contribution to the Paper	Performed computational work, interpretations and analysis of results. Wrote manuscript.
Overall percentage (%)	90
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 22 Jan 2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Milan Mijajlovic
Contribution to the Paper	Supervised conducting of work. Provided research suggestions and technical support. Assisted in interpreting results and editing manuscript.
Signature	Date 24/1/2020

Name of Co-Author	Hu Zhang
Contribution to the Paper	Supervised conducting of work.
Signature	Date Jan 22, 2020

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Mark Biggs
Contribution to the Paper	Supervised conducting of work. Provided research suggestions. Assisted in interpreting results. Evaluated and edited manuscript.
Signature	Date 29/1/2020

4 Characterizing the Switching Transitions of an Adsorbed Peptide by Mapping the Potential Energy Surface

4.1 Introduction

The interaction of peptides and proteins with solid surfaces has attracted significant attention over many years (Nakanishi et al., 2001, Gray, 2004, Rabe et al., 2011). Central to much of this prior work has been the application of peptide and protein adsorption in nanotechnology (Nel et al., 2009, Mahmoudi et al., 2011, Seker and Demir, 2011). For example, nanoparticles have been implicated in the denaturing of proteins in the human body, leading to toxicity (Laera et al., 2011). More positively, peptides may be used to form nanoparticle assemblies (Coppage et al., 2011), or to functionalize nanomaterials for applications such as cancer diagnosis and treatment (Cho et al., 2011, Kumar et al., 2012, Soum et al., 2015), biosensing (Zhang et al., 2011, Mannoor et al., 2012, Qu et al., 2013) and drug delivery (Huotari et al., 2013). Peptide adsorption is also of relevance in biology and medicine. For example, it is implicated in the fouling and degradation of medical implants (Meyers and Grinstaff, 2011), and may play a role in the fibrillation of proteins associated with degenerative brain diseases such as Alzhemier's (Mu et al., 2014, Bellucci et al., 2016).

Given the widespread applications of peptide and protein adsorption, it is desirable to understand the molecular processes involved in this phenomenon. All-atom molecular dynamics (MD) has been widely used to study preferred structures of adsorbed peptides (Latour, 2008, Ozboyaci et al., 2016a), and studies

in recent years have succeeded in estimating the free energy of adsorption (Biswas et al., 2012, Mijajlovic et al., 2013, Bellucci and Corni, 2014), and proposing mechanisms for the peptide adsorption process (Penna et al., 2014, Penna et al., 2015, Ozboyaci et al., 2016b). However, all-atom MD is only useful for adsorption events lasting up to the order of a microsecond, and as such, slower but no less important processes, such as the change in secondary structure of an already adsorbed peptide, are as yet less well understood (Rabe et al., 2011, Ozboyaci et al., 2016a). Up until the present time, these processes have only been investigated using spectroscopic experimental methods and coarse-grained simulations, which offer less definitive understanding than all-atom models (Rabe et al., 2011, Ozboyaci et al., 2016a).

Discrete path sampling (DPS) (Wales, 2002) is a promising all-atom methodology for investigating transitions between two conformations of a system that are too slow to be probed by all-atom MD. In short, this approach involves finding stationary points on the potential energy surface (PES) – that is, the minima (stable configurations) and saddle points (maximum-energy transition states) – between two conformations of interest. Transition pathways are then approximated as sequences of stationary points between the desired endpoints, and rate constants may be estimated using occupational probabilities derived from harmonic densities of states (Wales, 2002, Trygubenko and Wales, 2006, Wales, 2009). This methodology has been applied to study the rearrangements of clusters (Calvo et al., 2012, Farrell et al., 2013) and biomolecules (Prentiss et al., 2010, Carr et al., 2015, Neelamraju et al., 2015) in studies dating up to the recent past.

However, to the best of our knowledge, it has not been used to investigate changes in conformation of adsorbed peptides.

In a previous study (Mijajlovic and Biggs, 2007) in which molecular modeling was used to study the adsorption of polyalanine at gas-solid interfaces, some of the authors here showed that the polyalanine can take on one of the three helical structures illustrated in Figure 4-1 depending on the strength of the solid-peptide

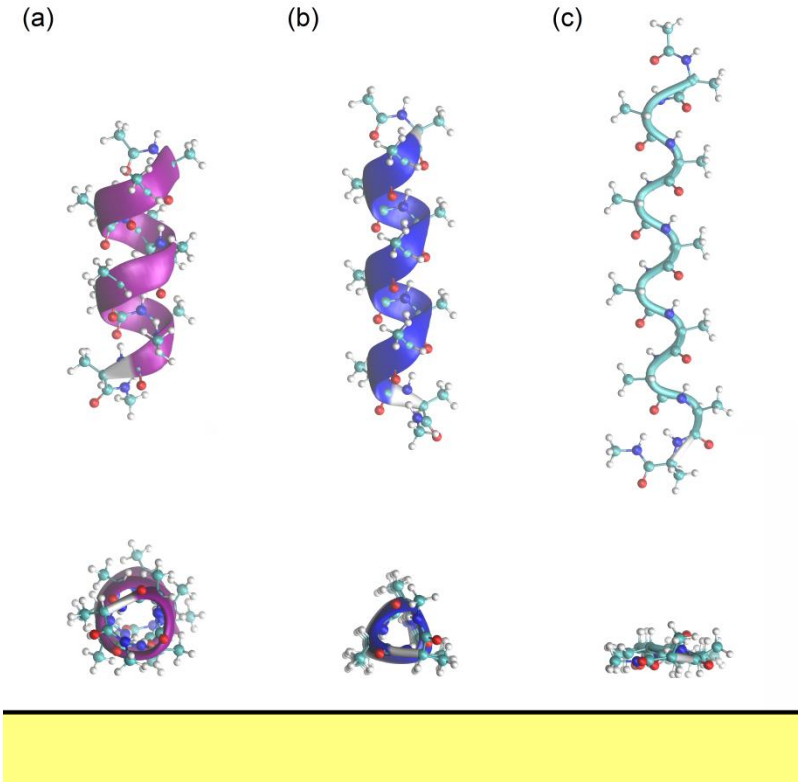

Figure 4-1. (a) Top and side views of the three helical forms taken by 10-alanine depending upon the strength of the interaction between the peptide and the solid surface (Mijajlovic and Biggs, 2007): (a) α-helix, which is found when the solid interacts weakly with the peptide (and also when in the gas phase); (b) 3_{10}-helix, which is found when the solid interacts more strongly with the peptide; and (c) 2_7-helix, which is found when the interaction between the solid and peptide is high. The structures were generated using Visual Molecular Dynamics (Humphrey et al., 1996).

interaction, and that switching occurs between these structures at specific solid-peptide interaction strengths. In the study reported here, DPS was used to elucidate the transitions between these conformations within the vicinity of the switching points. The nature of the PES of the system was also explored as a function of the solid-peptide interaction strength. In this report, we first outline the molecular model and methods used along with the details of the study undertaken. The switching energies, transition paths and associated rate constants are then presented and discussed. The report ends with a summary of key conclusions and an outline of future work.

4.2 Methodology

4.2.1 Model

The model is essentially composed of a molecular model of polyalanine above a solid surface. The details of this model were set to replicate those of Mijajlovic and Biggs (2007). The polyalanine molecules were capped by an acetyl group (CH_3CO) at the N-terminus and an amino-methyl group ($NHCH_3$) at the C-terminus. Intramolecular interactions were modeled using the Amber94 force field (Cornell et al., 1995).

The interaction between the peptide and solid surface was modeled by the Steele potential (Steele, 1973):

$$E_{Au} = 2\pi\rho \sum_{j} \sum_{l=0}^{L-1} \varepsilon_{sj} {\sigma_{sj}}^2 \left[\frac{2}{5} \left(\frac{\sigma_{sj}}{z_j + l\Delta} \right)^{10} - \left(\frac{\sigma_{sj}}{z_j + l\Delta} \right)^{4} \right] \qquad (4\text{-}1)$$

where $l = 0, \cdots, L-1$ is a counter over the layers of solid atoms in the surface up to the maximum, L, ρ the density of atoms in each of the layers, Δ the distance between the layers, z_j the perpendicular distance of peptide atom-j from the surface, and ε_{sj} and σ_{sj} the associated Lennard-Jones energy and length parameters, respectively. The solid surface parameters for Au [111], which are summarized in Table 4-1, were derived from the literature (Mahaffy et al., 1997, Lide, 2009). Specifically, the Lorentz-Berthelot rules were used to calculate the surface Lennard-Jones parameters, ε_s and σ_s, using the literature parameters for CH_2 interacting with Au (Mahaffy et al., 1997), and these same rules were then applied with the Amber94 force field parameters to calculate ε_{sj} and σ_{sj}. A range of surface-peptide interaction strengths, E_s, were investigated through a simple expedient of multiplying the potential energy arising from the Steele model by a factor, E_s/E_{Au}, which we call the surface energy ratio.

4.2.2 Methods

The study is split into two parts. The first, which essentially replicates the previous work reported in Mijajlovic and Biggs (2007) using a different method to identify the minimum energy polyalanine conformations, involved determining

Table 4-1. Steele model parameters for Au [111].

parameter	value
ρ	0.13886 atoms/Å^2
ε_s	0.0905 kcal/mol
σ_s	3.359 Å
Δ	2.3545 Å
L	2

which of the three helix forms is the most favored for a range of surface energies. This then underpinned identification of the surface energies where switching between two helix forms occurs.

The minimum energy polyalanine conformations and associated energy were identified for each of the helix forms as follows. Using a canonical form of the helix in question – either α-, 3_{10}- or 2_7-helix – the limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm (Liu and Nocedal, 1989) was used to find the minimum energy conformation starting from 1296 initial conformations that were generated by systematically varying the two angles defined in Figure 4-2 by increments of 10° between the possible limits (where this process resulted in a peptide atom being within 1Å of the surface, the peptide was

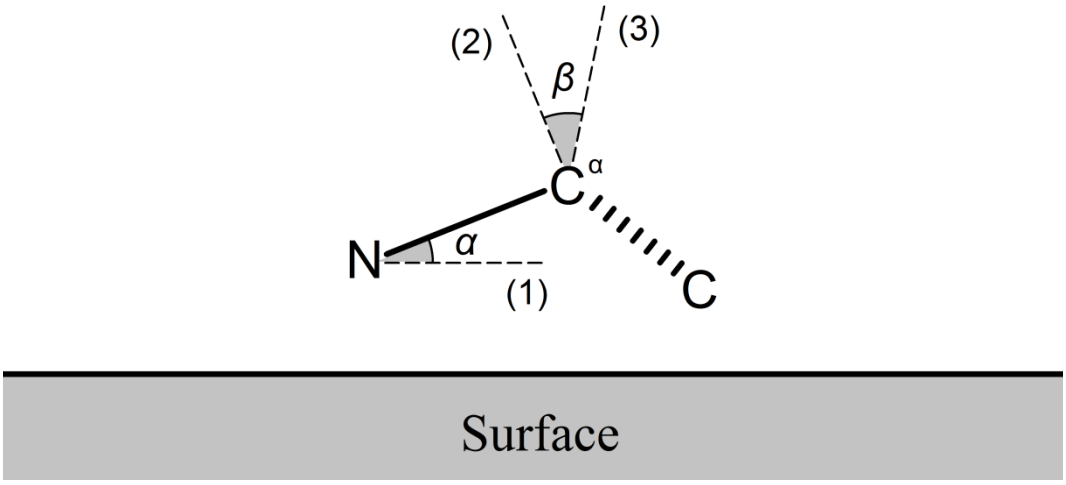

Figure 4-2. The angles, α and β, that were varied to generate initial adsorbed polyalanine configurations. The atoms shown belong to the residue nearest to the N-terminus. The viewing plane contains the N–C^{α} bond and is perpendicular to the surface. The labeled lines are (1) parallel to the surface in the viewing plane, (2) perpendicular to the N–C^{α} bond in the viewing plane, and (3) normal to both bonds.

shifted normal to the surface until this minimum distance was satisfied). The canonical forms of the α-, 3_{10}- and 2_7-helixes were generated by setting all the backbone dihedral angles to (ψ = –60°, ϕ = –45°), (–55°, –15°) and (–75°, 70°), respectively. The most favored helix form for a given surface energy is that identified as having the lowest potential energy for the surface energy. The switching point between two helix forms is that surface energy where the energies of two different forms are identical.

The second part of the study reported here was concerned with identifying the switching pathway and associated energy barrier for the two switches (*i.e.* $\alpha \rightarrow 3_{10}$ and $3_{10} \rightarrow 2_7$). As illustrated in the flow diagram shown in Figure A-1 in the Supporting Information, this involved the application of DPS to extensively enumerate the stationary points (*i.e.* minima and saddle points) on the PES of the systems in the vicinity of the switching points and the pathways between them. The details of this process are outlined in the remainder of this section below.

The iterative enumeration of stationary points and pathways between them for a PES started by repeatedly using basin-hopping (Wales and Doye, 1997) combined with simulated annealing (SA-BH) to expand the database of minima identified through the first part of this study. SA-BH involves randomly perturbing all backbone dihedral angles up to a maximum followed by relaxing the resulting structure using the LBFGS algorithm (Liu and Nocedal, 1989). The Metropolis criterion is then applied to determine whether to accept or reject the step, and this procedure is repeated for a specified number of steps. The SA component of the algorithm sees the temperature used in application of the Metropolis criterion

gradually reduce over the course of the simulation from an initial high value. The maximum step change in the dihedral angles was initially set at 360° and then adjusted regularly to maintain an acceptance ratio of 0.5.

Following considerable expansion of the minima database using SA-BH, a variety of selection schemes were used to choose pairs of minima to be used as starting points for a saddle point search. Details of the saddle point searches are provided in the following paragraph. Firstly, the DIJINIT (Carr et al., 2005) scheme was used to iteratively construct a putative pathway of connected stationary points between the minima associated with the two helix forms that define the switch being considered. Once this putative pathway was established, further selection schemes were applied iteratively to find additional stationary points that may be used to construct more energetically accessible pathways between the endpoint minima. These schemes included SHORTCUT (Carr and Wales, 2005), DIJPAIR (Strodel et al., 2007) and UNTRAP (Strodel et al., 2007), which all attempt to shorten the putative minimum energy pathway and decrease the energy barrier, by finding unknown stationary points near the pathway. The NEWCONNECTIONS (Wales Group Home Page) scheme was also employed, which iteratively selects each known minimum below the energy barrier of the current minimum energy pathway and applies a small random perturbation. An additional scheme, which we call CONNUSERMIN, was also used in this study to connect a user-specified list of minima with their closest neighbors in Euclidean space; this was used to cross-check the putative pathways between the same helix forms at different

surface energies, ensuring structures along the transition path at one surface energy ratio were adequately sampled at the others.

The identification of saddle points between pairs of minima first involved use of the doubly nudged elastic band (DNEB) method (Trygubenko and Wales, 2004) to identify saddle point candidates between the two chosen minima. Each candidate was then in turn refined using eigenvector following (Munro and Wales, 1999). Following the precise location of a saddle point, the LBFGS algorithm was once again used to identify the two minima adjacent to the saddle point; if any of these minima were previously unknown, they were added to the database of stationary points. When using the NEWCONNECTIONS scheme, the DNEB method was skipped entirely and eigenvector following was applied directly to the perturbed minimum.

After applying any given scheme, the KSHORTESTPATHS algorithm (Carr and Wales, 2008) was used to determine the lowest energy pathway and calculate the rate constant (Wales, 2009). The refining of the putative switching pathway continued until the change in the rate constant of the minimum energy pathway between the two forms of helix of interest did not substantially change with the number of discovered minima. As illustrated in Figure A-2 in Supporting Information, the rate of change of the rate constant with the number of discovered minima approached zero for both switches, suggesting that the PES were well-mapped near the helixes of interest and the limiting rate constants can be associated with the switching pathway.

The various algorithms mentioned above were accessed using the software of Wales and co-workers, which is freely available on their website (Wales Group Home Page) . A summary of these algorithms and schemes is given in the Supporting Information. Their usage is this study required interfacing the software with the CHARMM program (Brooks et al., 2009), which was used to evaluate the potential energy for a given peptide conformation. The CHARMM program was modified to include the Steele potential.

4.2.3 Study Details

As *per* the authors' previous study (Mijajlovic and Biggs, 2007), polyalanine molecules of length ranging from 6 to 12 residues were considered for surface energies ranging from $E_s/E_{Au} = 0.0$ (*i.e.* gas-phase polyalanine) to $E_s/E_{Au} = 5.5$ at increments of 0.1. The structures obtained from minimization were classified as α-, 3_{10}- or 2_7-helixes or otherwise, under the criterion that at least two thirds of the residues must have both backbone dihedral angles within ±20° of the values associated with the canonical values given above.

The α→3_{10} and 3_{10}→2_7 transitions were each studied for 10-alanine at three surface energy ratios near to the switching point. The multiple energies were studied to investigate whether there were any significant changes to the transition pathways either side of the switching point.

Each SA-BH simulation was run for a total of 50,000 steps, with an initial temperature of $k_BT = 5.0$ kcal/mol, which was decremented by 8×10^{-3}% after each step. The maximum step size was adjusted up or down by 5% every 50 steps

if the acceptance ratio for the last 50 steps was greater than or less than 0.5, respectively.

The DNEB method was applied with 10 images and a maximum of 300 iterations. The LBFGS algorithm, which was used in both the DNEB and eigenvector following methods, was implemented with a memory of the last 4 iterations, a maximum step size of 0.4 Å, initial guesses for the diagonals of the Hessian matrix of 0.1 $Å^2$mol/kcal, and a convergence criterion of the RMS gradient not exceeding 10^{-7} kcal/molÅ.

Perturbations to minima for the NEWCONNECTIONS scheme were randomized to a maximum of 0.01Å for each Cartesian coordinate. Paths and rate constants calculated using the KSHORTESTPATHS algorithm were evaluated at a temperature of $k_BT = 0.3$ kcal/mol.

4.3 Results and Discussion

4.3.1 Switching Behavior of Polyalanine Molecules

Figure 4-3 shows for 10-alanine the variation of the potential energy of the minimum-energy α-, 3_{10}- and 2_7- helixes with the surface energy ratio; the variation is similar for the other polyalanine molecules investigated here. The helix structure with the lowest energy for a given surface energy ratio is the most favored at that surface energy. Thus, in line with our previous study (Mijajlovic and Biggs, 2007), the α-helix is the most favored in the absence of the surface (*i.e.* in the gas phase) and when the solid surface interacts weakly with the peptide, the

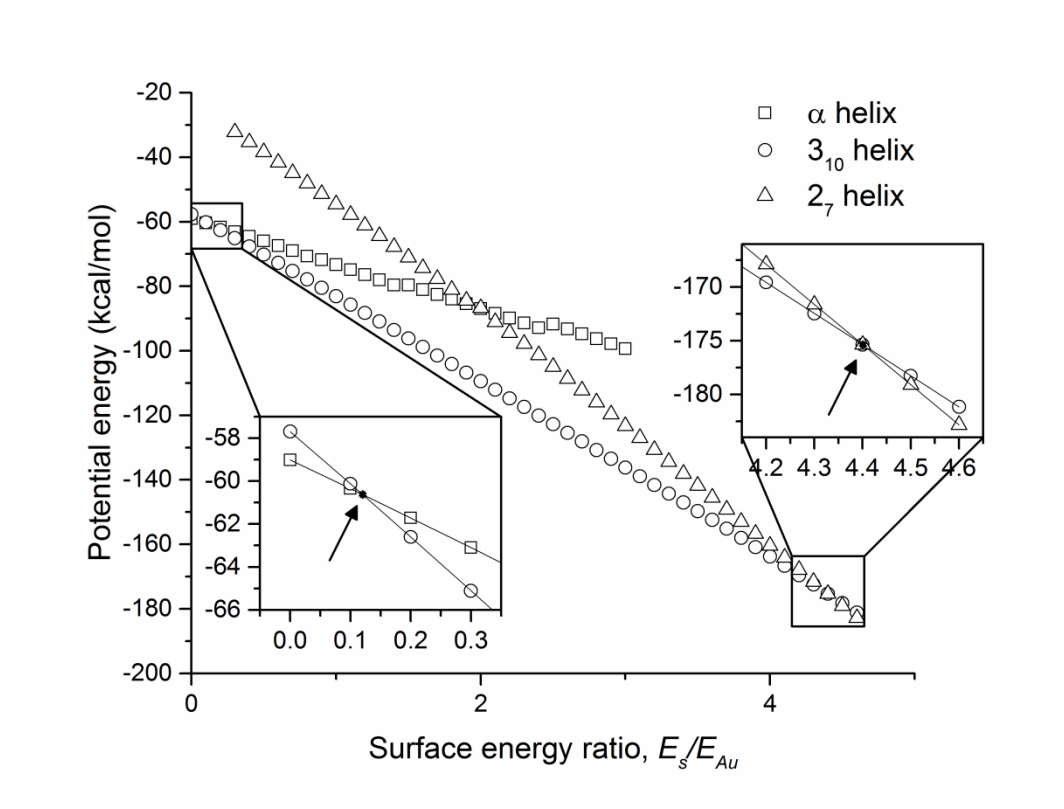

Figure 4-3. Variation of the potential energy of the α-, 3_{10}- and 2_7-helix structures of 10-alanine with the surface energy ratio. Insets show the behavior near the α→3_{10} and 3_{10}→2_7 switching points. Lines shown in the insets are used to locate the switching points, which are at the intersections of the lines indicated by the arrows.

2_7-helix is preferred at the highest surface energies investigated here, and the 3_{10}-helix is favored for energies between these two extremes. There are two switching points: where the α- and 3_{10}-helixes are equally favored at a low surface energy ratio, and where the 3_{10}- and 2_7-helixes are equally favored at a high surface energy ratio. For 10-alanine, the α→3_{10} switching point is $E_s/E_{Au} = 0.121$, whilst that of the 3_{10}→2_7 switch is $E_s/E_{Au} = 4.400$.

Figure 4-4 shows the switching point surface energies as a function of the size of the polyalanine molecule. The variation seen in this figure is similar to that obtained by Mijajlovic and Biggs (2007) (Figure 9(a) in their work). However, there are two clear differences. The first is the preference for the 3_{10}-helix in the gas phase and lower surface energies when the number of residues drops below 10. This preference in the gas phase for short polyalanines is not inconsistent with earlier results of others (Schafer et al., 1993, Topol et al., 2001). It is also in line with the work of Park and Goddard (2000), who showed that polyalanine α-helixes are stabilized by dipole-dipole interaction energy that increases with the

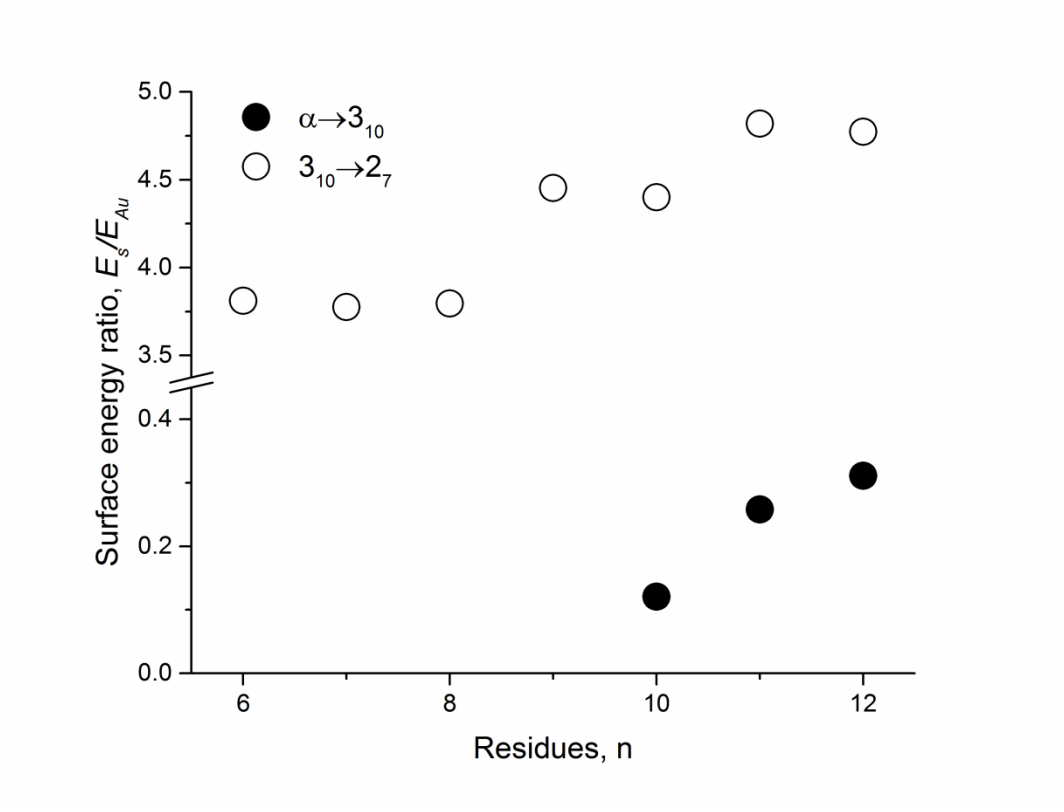

Figure 4-4. Variation of the switching surface energy ratio with the polyalanine length for the $\alpha \rightarrow 3_{10}$ and $3_{10} \rightarrow 2_7$ switches.

number of residues. This result is corroborated by other studies that show α-helical global minima for long polyalanines (Zhang and Hermans, 1994, Okamoto and Hansmann, 1995, Mortenson et al., 2002, Rossi et al., 2010). More recently, studies of polyalanine and polyalanine derivatives adsorbing on solid surfaces have indicated the formation of α-helixes through simulation (*via* a coarse-grained Monte Carlo study) (Mu, 2011), and both α-helixes (Sek et al., 2005, Hu et al., 2010) and 3_{10}-helixes (Hu et al., 2010) through experiment (*via* synthesis of self-assembled monolayers), although it must be noted that none of these studies precisely replicate the system studied here. The second difference between the results obtained here and those of Mijajlovic and Biggs (2007) is the absolute values of the switching energies: they are significantly higher and lower for $3_{10} \rightarrow 2_7$ and $\alpha \rightarrow 3_{10}$ switches, respectively.

The differences observed between the results obtained in the study reported here and the study of Mijajlovic and Biggs (2007) are easily explained by one fundamental difference in the methodologies used. While the force field and surface models are identical, Mijajlovic and Biggs constrained the bond lengths and angles during the energy minimization, varying only the dihedral angles (*via* an evolutionary algorithm). In the study reported here, on the other hand, bond lengths and angles were free to change as all atoms were permitted to move freely when minimizing the structures. The additional localized flexibility means the same three structures are preferred as the surface energy increases but the energies at which they are preferred are shifted relative to when the flexibility is absent.

4.3.2 Potential Energy Surfaces for 10-Alanine

Based on the switching points determined for 10-alanine, the PES for this molecule were investigated in detail along with the switching transitions for surface energy ratios of $E_s/E_{Au} = 0.0$, 0.1 and 0.2, which bracket the $\alpha \rightarrow 3_{10}$ transition, and $E_s/E_{Au} = 4.2, 4.4$ and 4.6, which straddle the $3_{10} \rightarrow 2_7$ transition. Figure 4-5 shows the disconnectivity graphs (Becker and Karplus, 1997) for $E_s/E_{Au} = 0.0$, 0.1 and 4.4; corresponding graphs for the other surface energy ratios investigated are shown in Figure A-3 in Supporting Information. It is immediately obvious from these graphs that the presence of the solid surface greatly increases the complexity of the PES of the system, a phenomenon that has previously been noted as an obstacle to studies of peptide-surface interactions (Wright and Walsh, 2013). It is also clear that the degree of complexity broadly increases with the strength of the peptide-surface interaction. This is reflected in Table 4-2, which shows the number of minima and saddle points identified through the work reported here. Whilst the number of points identified for each surface energy ratio are considerable, the fact that the solid surface 'roughens' the PES means comprehensive enumeration of all stationary points is unrealistic (Wright and Walsh, 2013). However, as outlined in Section 4.2.2, given the variety of methods used to survey the PES and the convergence of the rate constant of the preferred switching pathway with the number of discovered minima, we are confident that the pathways between the three helix forms of interest here have been comprehensively sampled.

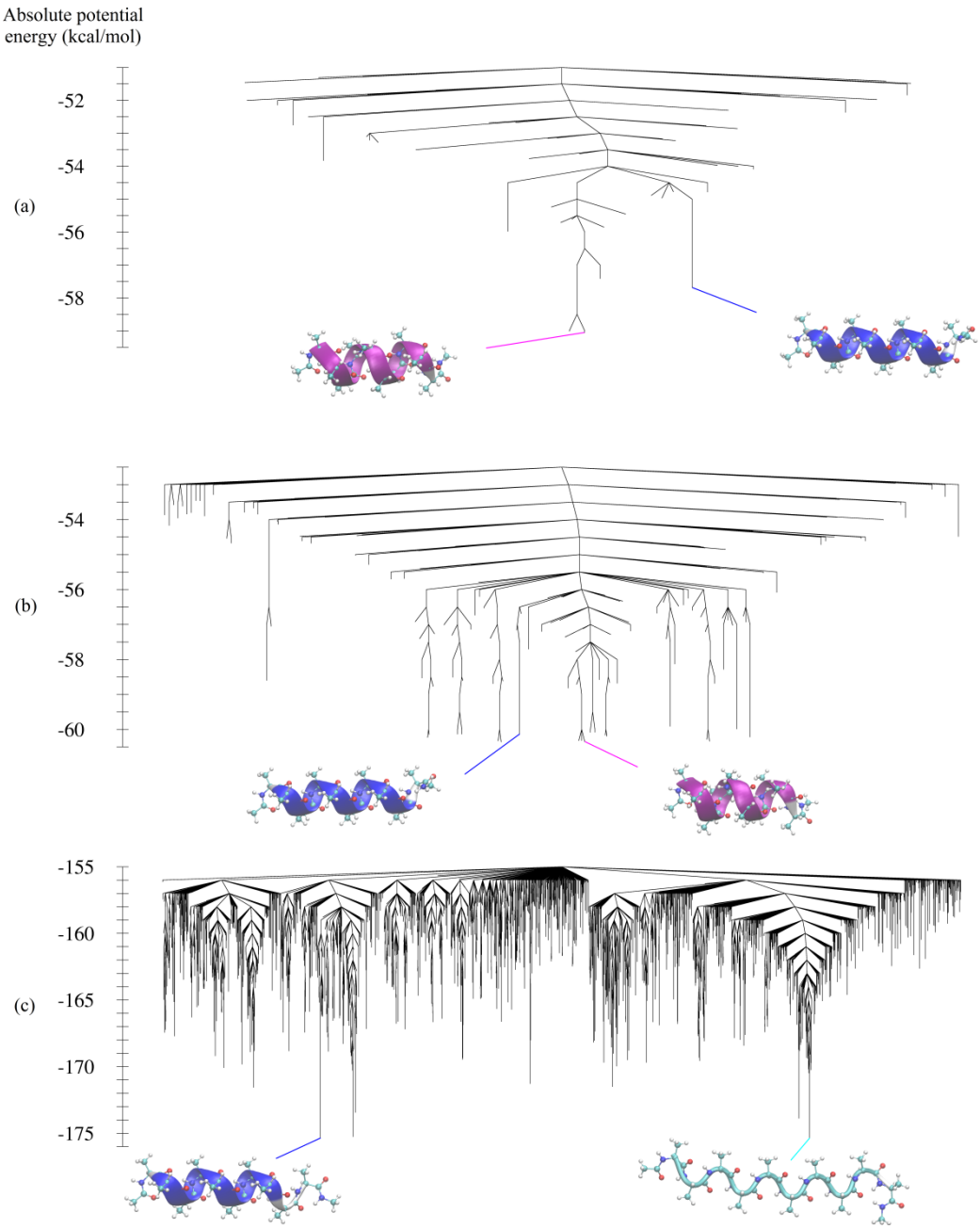

Figure 4-5. Disconnectivity graphs showing the lowest minima found for 10-alanine at surface energy ratios of: (a) $E_s/E_{Au} = 0.0$, (b) $E_s/E_{Au} = 0.1$ and (c) $E_s/E_{Au} = 4.4$. The minima corresponding to the minimum-energy α- (magenta), 3_{10}- (blue) and 2_7-helices (cyan) are highlighted.

Table 4-2. Number of stationary points discovered for the PES of 10-alanine when in the gas phase and adsorbed on surfaces of various surface energies considered in detail here.

surface energy ratio, E_s/E_{Au}	**number of**	
	minima	**saddles**
0.0	6364	8288
0.1	6448	9064
0.2	7194	10116
4.2	28361	38331
4.4	24791	32825
4.6	25681	33357

The termini in the disconnectivity graphs correspond to minima whilst the nodes represent the energy barrier between two minima or groups of minima (Becker and Karplus, 1997). The appearance of a graph can then be used to draw inferences about the underlying PES (Wales et al., 1998). The gas-phase disconnectivity graph shown in Figure 4-5(a) exhibits a classic 'funnel' structure with a number of shallow minima guiding the system down to the global minimum α-helix. The 3_{10}-helix is one of those 'guiding' minima, separated by a small energy barrier. Figure 4-5(b), which displays the graph at a low surface energy, $E_s/E_{Au} = 0.1$, where the α- and 3_{10}-helixes are favored, is similarly shaped to the gas-phase graph but with a number of minima of similar energies, corresponding to these helixes at different orientations to the surface. The lower part of this graph resembles the 'banyan tree' structure proposed by Wales and co-workers (Wales et al., 1998), and indicates that at any time the molecule may take any of these structures with similar probabilities.

Figure 4-5(c) shows the disconnectivity graph at a high surface energy, $E_s/E_{Au} = 4.4$, at which the 3_{10}- and 2_7-helixes have similar energies. This exhibits a multi-funnel structure with the most prominent funnel leading down to the 2_7-helix. Another funnel contains the 3_{10}-helix along with another competing minimum, which represents a stable $3_{10}/2_7$ hybrid we will return to below. There are several further funnels that indicate other forms 10-alanine may possibly take, albeit with higher potential energy. The relative wideness of the 2_7-funnel indicates that the 2_7-helix may be more favored when entropy is taken into consideration.

Comparing Figure 4-5(b-c) with the corresponding graphs in Figure A-3 indicates the PES do not change dramatically as the switching points are crossed. This suggests that the switching mechanism will be similar in either direction (*i.e.* from $3_{10} \rightarrow 2_7$ and from $2_7 \rightarrow 3_{10}$, for example) and, thus, are not ratchet-like. This absence of a ratchet-like switching may well arise from the symmetry (Cheng et al., 2015) of polyalanine. This possibility could potentially be revealed by applying the analysis undertaken here to helix-forming asymmetric peptides.

4.3.3 Switching Transitions for 10-Alanine

The transition path identified for the change between the α-helix and 3_{10}-helix at $E_s/E_{Au} = 0.1$ is illustrated in Figure 4-6, with inset figures showing major intermediates and transition states. The switch from the marginally more favored α-helix structure to the 3_{10}-helix was found to occur in three distinct stages. Firstly, the α-helix rolls about its longitudinal axis to assume a slightly less stable position on the surface, but one from which the 3_{10}-helix is more readily accessible (stage I): the 3rd and 7th residues are nearest to the surface, compared to

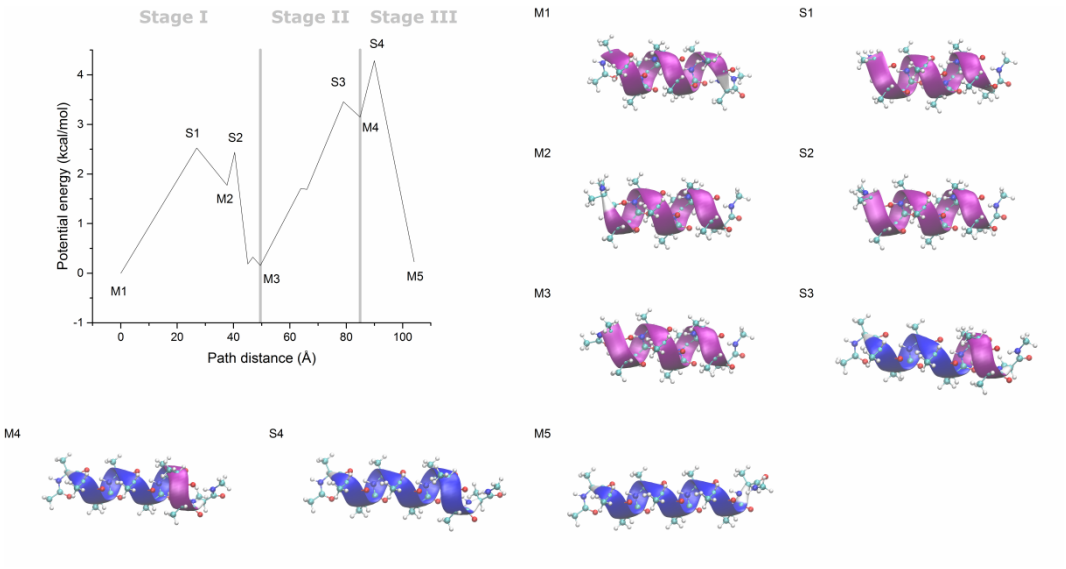

Figure 4-6. Variation of the potential energy along the transition pathway between the α-helix (magenta) and 3_{10}-helix (blue) at $E_s/E_{Au} = 0.1$, and images of selected structures along the path, viewed from above the surface. The reference potential energy is that of the minimum-energy α-helix adsorbed on the surface. The path distance is the minimized Euclidean distance between neighboring stationary points. Lines between stationary points are provided as a guide to the eye only.

the starting structure where the 4th and 8th residues are nearest. In the next step, the two turns closest to the N-terminus switch from an α to a 3_{10} configuration, forming a marginally stable intermediate between the two helixes (stage II). Finally, the remaining α-helical residues, which are adjacent to the C-terminus, switch to yield a full 3_{10}-helix (stage III); it is this stage that presents the highest potential energy barrier in this switching process.

The paths for the transitions identified at $E_s/E_{Au} = 0.0$ (gas-phase 10-alanine) and $E_s/E_{Au} = 0.2$, which are shown in Supporting Information (Figures A-4 and A-5), are similar in nature, although stage I is omitted in the gas-phase transition

for obvious reasons. These figures reveal that the path length increases with the surface energy ratio, in line with the increase in the complexity of the PES (see above).

The transition path identified for the change between the 3_{10}-helix and 2_7-helix at $E_s/E_{Au} = 4.4$ is shown in Figure 4-7, again with inset figures showing intermediate structures. This path is more complex than that for the $\alpha \rightarrow 3_{10}$ switch at lower surface energy, and involves a number of small changes in conformation

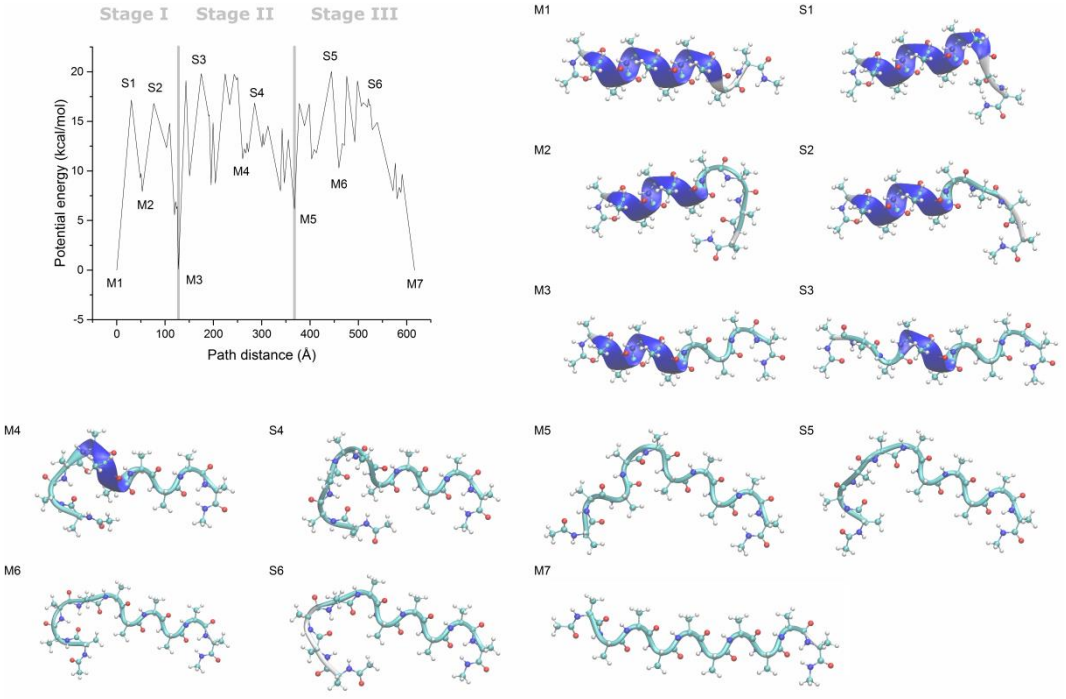

Figure 4-7. Variation of the potential energy along the transition pathway between the 3_{10}-helix (blue) and 2_7-helix (cyan) at $E_s/E_{Au} = 4.4$, and images of selected structures along the path, viewed from above the surface. The reference potential energy is that of the minimum-energy 3_{10}-helix adsorbed on the surface. The path distance is the minimized Euclidean distance between neighboring stationary points. Lines between stationary points are provided as a guide to the eye only.

with, in many cases, high potential energy barriers in between. However, as with the α→3_{10} switch, a three-stage mechanism is apparent, punctuated by two notable intermediates between the helixes. Firstly, the 3_{10} structure closest to the C-terminus becomes two 2_7-helix turns, forming a $3_{10}/2_7$ hybrid of comparable stability to the pure helixes (stage I). Secondly, the remainder of the 3_{10}-helix, near the N-terminus, also switches to 2_7, but coiling in the middle of the structure produces a moderately stable 'broken' 2_7-helix (stage II). Finally, the segment near the N-terminus flips upside down, producing a full 2_7-helix (stage III). The highest energy barrier occurs at the final stage, although the energy barriers of all three stages are comparable.

Corresponding figures for the transitions identified for $E_s/E_{Au} = 4.2$ and 4.6 are provided in Supporting Information (Figures A-6 and A-7). These show the same three-stage process identified here. As with the α→3_{10} switching transition, the path length appears to increase with the surface energy ratio. Also, for $E_s/E_{Au} = 4.6$, the highest energy barrier occurs in stage II, possibly due to the improved stability of 2_7-helical structures at the greater surface energy ratio.

It is evident from the results outlined above that the switch from the 3_{10}-helix to the 2_7-helix is significantly more impeded than the α→3_{10} switch; the transition path is approximately six times as long and the energy barrier about five times as great. The rate constants, which were evaluated using transition state theory incorporating local harmonic densities of states (Wales, 2002), quantify in a more direct way the differences in the timescales of each switch. The rate constants for α→3_{10} switching both in the gas phase and at $E_s/E_{Au} = 0.1$, which are shown in

Figure 4-8 as a function of temperature, are greater than 10^8 s^{-1} for temperatures above 200K. This corresponds to a mean transition time of up to a few nanoseconds, which is well within the capabilities of MD simulation (Ozboyaci et al., 2016a). It is worth noting that an early MD study of gas-phase 10-alanine at 300 K yielded a switch from 3_{10} to α in a few picoseconds (Zhang and Hermans, 1994), which corresponds well with our results here. It is also notable that the rates are slower in the presence of the surface by about an order of magnitude, despite the surface favoring the formation of 3_{10}-helixes; this is possibly another consequence of the increasing complexity of the PES.

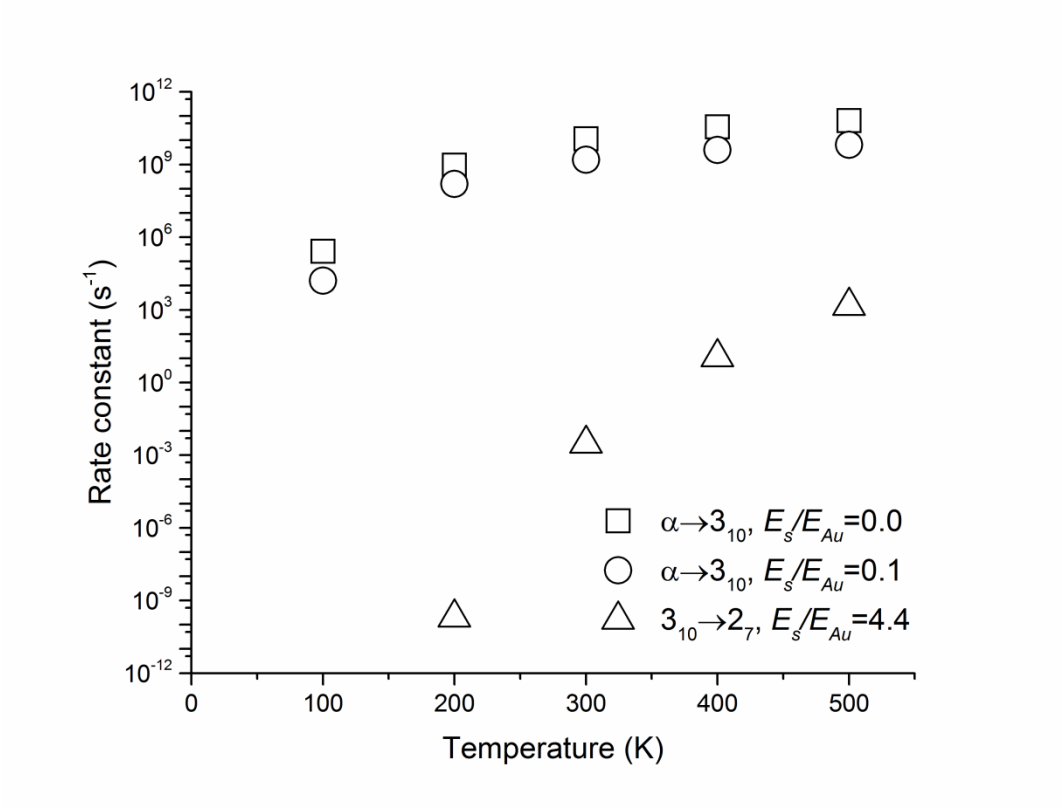

Figure 4-8. Variation of the rate constants with temperature for the switching transitions at $E_s/E_{Au} = 0.0$, 0.1 and 4.4.

The rate constants for $3_{10}\rightarrow2_7$ switching at $E_s/E_{Au} = 4.4$ are far lower than for $\alpha\rightarrow3_{10}$ switching. Even at the maximum temperature sampled of 500 K, the rate constant of approximately 10^3 s^{-1} corresponds to a transition time of about a millisecond, which conventional MD simulations cannot currently reach. It should be noted that only the rate constant for the most preferred path from the higher-energy minimum to the lower-energy minimum was considered; in reality, the transition is reversible and there are other transition paths and competing structures that would complicate the process and potentially lengthen the necessary simulation timescales.

4.3.4 Approximate Analysis of Computational Effort

Given DPS is generally adopted when MD simulation is unable to probe the necessary timescales within a reasonable level of computational resource (Wales, 2002), it is worthwhile discussing the effort required to generate the results here. For a system of the size considered here, a full saddle point search incorporating the DNEB and eigenvector following methods took up to a few minutes on a single core of an AMD Opteron 6212 CPU. Thus, in order to construct a database of the size of those found in this study (*c.f.* Table 4-2), at least a few thousand core hours are necessary, and this is further increased by the finding of redundant minima and saddle points, and the extra time taken to read and manage a large database. A conservative estimate of the computational time required is, therefore, around 10,000 core hours. This compares unfavorably with MD for the $\alpha\rightarrow3_{10}$ switch, where MD simulations of the order of 5 core hours on these Opteron CPUs would be sufficient. In the case of the $3_{10}\rightarrow2_7$ switch, on the other hand, the

MD simulation time would increase to the order of 10 million core hours even at 500K, making it entirely impractical; DPS is the only viable option. Whilst the computational time for DPS compares unfavorably for simpler systems such as we consider here at lower surface energies, it is worth noting that this form of analysis is embarrassingly parallelizable (many saddle point searches can be run essentially independently) and it can yield insights that are difficult to gain from MD alone as demonstrated by the present study.

A related point of discussion is how DPS scales with the size of the adsorbed peptide. It is known that the number of stationary points on a potential energy surface generally has an exponential relationship with the size of the system (Wales, 2003). This study also demonstrates the effect of the surface in increasing the complexity of the potential energy surface. This suggests that while larger biomolecules have been studied using DPS in the past (Prentiss et al., 2010), studying the adsorption of such molecules may present greater challenges. Further study will, however, be required before this is fully understood.

4.4 Conclusions

Three known helical structures formed by polyalanine at a gas/solid interface were subjected to energy minimization and compared. In agreement with past findings (Mijajlovic and Biggs, 2007), a switching phenomenon was observed from α-helix to 3_{10}-helix to 2_7-helix, in order of increasing strength of the surface-peptide interaction. A discrete path sampling procedure was used to investigate this switching process for 10-alanine, and general stage-by-stage mechanisms for the

$\alpha \rightarrow 3_{10}$ and $3_{10} \rightarrow 2_7$ switches were identified. The former switch is a relatively simple and energetically accessible process that involves the molecule 'rolling' on the surface and then transforming piecewise from an α-helix to a 3_{10}-helix. The $3_{10} \rightarrow 2_7$ switch is more complex, involving a number of small changes and intermediates, with high energy barriers and low rate constants. As a consequence, while the $\alpha \rightarrow 3_{10}$ switch occurs rapidly enough that it may easily be observed using molecular dynamics simulation, the $3_{10} \rightarrow 2_7$ switch will normally be beyond current MD simulation capabilities. The information gleaned about the $3_{10} \rightarrow 2_7$ transition demonstrates the possibilities inherent in this form of study, which could well be applied to other cases of biomolecule adsorption that occur over long periods of time.

This study largely focused on the PES; entropic effects were only included when calculating rate constants, and these only took into consideration the most favored path in one direction. The disconnectivity graphs of the PES obtained here suggest that entropy may have some influence at least quantitatively – we are, therefore, working on including entropic contributions more fully (Krivov and Karplus, 2002, Evans and Wales, 2003), which we will report on in the future. We will also investigate the variation of complexity and computational effort as a function of the peptide size and other peptide characteristics. Finally, as it is anticipated that a solvent will affect the switching behavior explored here (Mijajlovic and Biggs, 2007), incorporation of solvent in DPS analysis is of interest. Unfortunately, this is a major challenge at present (Wright and Walsh, 2013), and will need to await further methodological developments.

Statement of Authorship

Title of Paper	Energy Landscape Mapping and Replica Exchange Molecular Dynamics of an Adsorbed Peptide
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Ross-Naylor, J.A., Mijajlovic, M. & Biggs, M.J. 2020. Submitted to J. Phys. Chem. B.

Principal Author

Name of Principal Author (Candidate)	James Ross-Naylor
Contribution to the Paper	Performed computational work, interpretations and analysis of results. Wrote manuscript.
Overall percentage (%)	90
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 22 Jan 2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate to include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Milan Mijajlovic
Contribution to the Paper	Supervised conducting of work. Provided research suggestions and technical support. Assisted in interpreting results and editing manuscript.
Signature	Date 24/1/2020

Name of Co-Author	Mark Biggs
Contribution to the Paper	Supervised conducting of work. Provided research suggestions. Assisted in interpreting results. Evaluated and edited manuscript.
Signature	Date 29/1/2020

Please cut and paste additional co-author panels here as required.

5 Energy Landscape Mapping and Replica Exchange Molecular Dynamics of an Adsorbed Peptide

5.1 Introduction

Peptides are often used in the functionalization of solid surfaces due to their high degree of programmability (Shiba, 2017). Applications of material-binding peptides include the formation and functionalization of nanomaterials (Care et al., 2015, Slocik and Naik, 2017, Walsh and Knecht, 2017), protective or bioactive coatings for medical implants and other biomaterials (Pagel and Beck-Sickinger, 2017, Mas-Moruno, 2018), biosensing (Demir et al., 2016, Liang et al., 2016), cancer diagnosis (Soum et al., 2015) and drug delivery (Szweda et al., 2016). Peptide adsorption can also be used as a model for that of larger biomolecules such as proteins, which is implicated in desired and otherwise bodily reactions (Docter et al., 2015, Trindade et al., 2016). For instance, the effect of a surface on the fibrillation of the amyloid beta protein, which causes Alzheimer's disease, has been studied using polypeptide segments of this protein (Bellucci et al., 2016, Liu et al., 2016, Soltani and Gholami, 2017).

Computational studies have led to an improved understanding of peptide-surface interactions in recent years (Heinz and Ramezani-Dakhel, 2016, Ozboyaci et al., 2016a, Ramakrishnan et al., 2017). Modeling interactions that occur over timescales longer than the microsecond range remains a challenge, however, since all-atom simulation methods such as molecular dynamics (MD) are not currently capable of probing these scales sufficiently (Ozboyaci et al., 2016a). This

limitation poses a particular barrier to assessing the conformations of many material-binding peptides, which possess complex, flexible natures that cannot be reliably sampled in so short a time (Walsh, 2017). Coarse-grained models that significantly reduce the number of degrees of freedom in the molecule improve computational speed to the point of enabling adequate probing of timescales into the second range (Heinz and Ramezani-Dakhel, 2016). However, this occurs at the cost of the model's accuracy, and the development of reliable coarse-grained models remains a work in progress (Ramakrishnan et al., 2017).

An alternative approach to MD simulation is the use of mathematical algorithms to locate stationary points on the underlying potential energy surface (PES) of the system, the local minima (stable configurations) and saddle points (maximum-energy transition states) (Berry, 1993, Wales, 2003, Wales and Bogdan, 2006, Wales, 2018). These algorithms may be applied iteratively to create large connected networks of stationary points that constitute a map of the PES, in contrast with MD and other molecular simulation methods that explore the free energy surface (FES) stochastically; we therefore refer to the former approach as 'energy landscape mapping' (ELM). Stationary point networks constructed using ELM may be used in a variety of ways including, for example, constructing the FES using harmonic approximations (Evans and Wales, 2003), understanding the nature of the PES and FES through disconnectivity graphs (Becker and Karplus, 1997, Wales et al., 1998) and transition mechanisms between stable configurations *via* discrete path sampling (DPS) (Wales, 2002).

Recently, the authors here applied ELM to an adsorbed peptide for the first time ever, successfully reproducing a previously reported conformational switching phenomenon and identifying the transitions (Ross-Naylor et al., 2017). The study reported here further explores the potential of ELM for adsorbed peptides by comparing results obtained from it with those produced by replica exchange molecular dynamics (REMD), an MD technique designed to ameliorate the timescale limitations of standard MD (Sugita and Okamoto, 1999). The system considered was met-enkephalin adsorbed at the gas/graphite interface, which has been previously studied by some of the authors (Mijajlovic et al., 2011). The primary conformations and energy landscapes obtained by the two approaches were compared, along with the heat capacity curves computed by each method. Transition paths and rate constants between prominent conformations were also evaluated using ELM. The models and methodologies used and the study details are summarized in the next section, followed by a presentation and discussion of results and, finally, a concluding summary.

5.2 Methodology

5.2.1 Model

The model consists of a single met-enkephalin molecule above a flat, infinite graphite surface. Met-enkephalin, a pentapeptide with the amino acid sequence Tyr-Gly-Gly-Phe-Met, was represented in simple uncharged form, with NH_2 and COOH groups at the N- and C-termini, respectively. This form was observed to yield a broader range of conformations and a more complex PES than the

zwitterionic form, which was also considered as part of the study (results not shown). Intramolecular interactions were modeled using the CHARMM36m force field, a variation of the CHARMM36 force field with improved modeling of intrinsically disordered peptides and proteins (Huang et al., 2017). The CHARMM force field used has broken symmetry, which has been noted to cause problems in this form of studies, due to inconsistent energies arising for permutational isomers of the same structure (Małolepsza et al., 2010). However, since none of the residues or termini in uncharged met-enkephalin are asymmetric in the CHARMM36m force field, the results of the present study are unaffected by this.

Graphite was represented as two graphene layers, since it has been previously observed that layers beyond the first two have negligible interaction with adsorbing molecules (Braun et al., 2002). In ELM calculations, peptide-surface interactions were modeled by the Steele potential (Steele, 1973):

$$E_s = 2\pi\rho \sum_j \sum_{l=0}^{L-1} \varepsilon_{sj} {\sigma_{sj}}^2 \left[\frac{2}{5} \left(\frac{\sigma_{sj}}{z_j + l\Delta} \right)^{10} - \left(\frac{\sigma_{sj}}{z_j + l\Delta} \right)^4 \right] \qquad (5\text{-}1)$$

where $l = 0, \ldots, L - 1$ is a counter over the layers of solid atoms in the surface up to the maximum, L, ρ the density of atoms in each of the layers, Δ the distance between the layers, z_j the distance from the surface of peptide atom j, and ε_{sj} and σ_{sj} the Lennard-Jones energy and length parameters, respectively. The Lennard-Jones parameters were calculated from the corresponding parameters for the surface, ε_s and σ_s, and the atom, ε_j and σ_j, using the Lorentz-Berthelot rules. The surface parameters are summarized in Table 5-1.

Table 5-1. Steele model parameters for graphite.

parameter	value	reference
L	2	[a]
ε_s	0.05564 kcal/mol	[b]
σ_s	3.40 Å	[b]
ρ	0.3807 atoms/Å^2	[c]
Δ	3.3555 Å	[c]

[a] (Braun et al., 2002)

[b] (Steele, 1973)

[c] Derived from literature bond length and unit cell height data (Trucano and Chen, 1975).

In the case of the REMD simulations, graphite was represented using an atomistic model, with the atomic structure and force field parameters set to replicate those used in the Steele model for ELM. A periodic box with dimensions 51.723 × 51.192 × 50.000 Å^3 was used to achieve the same in-plane extent as in ELM, with the graphite sheets positioned at the bottom of the box. The non-bonded interactions in the REMD simulations were truncated using a smooth cutoff function between 10 Å and 12 Å in order to maintain a reasonable computational effort whilst minimizing the differences with the ELM energies.

5.2.2 Methods

5.2.2.1 Energy Landscape Mapping

ELM was used to construct an extensive database of connected stationary points on the PES. Basin-hopping with simulated annealing (SA-BH) was first used to identify a large number of local potential energy (PE) minima. BH involves conducting a Monte Carlo simulation on a transformed PES, where every point is mapped to the locally minimized energy, thus allowing the system to bypass

energy barriers (Wales and Doye, 1997). A random step is taken from a starting structure, followed by local minimization, and the Metropolis criterion is then applied to accept or reject the step. This procedure is repeated for a specified number of steps. Combining this methodology with SA, where the temperature used in application of the Metropolis criterion is decreased over the course of the simulation, allows the algorithm to gradually seek out lower-lying minima. In this study, the steps were taken by randomly perturbing the backbone dihedral angles of the met-enkephalin molecules up to a maximum step size, and minimization was carried out using the limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm (Liu and Nocedal, 1989). A noted weakness of SA is that, as the temperature decreases, the system may become trapped in a region of the PES that does not contain the global minimum (Wales and Doye, 1997). This was mitigated by running in parallel many independent SA-BH simulations starting from different initial structures, and using a large starting step size so as to encourage broad sampling of the PES. Parallel tempering, in which the temperatures of parallel simulations are periodically exchanged, provides an alternative approach to addressing this issue, and has been benchmarked in conjunction with BH (Strodel et al., 2010, Joseph and Wales, 2018). However, in this study, it was observed that parallel SA-BH simulations were sufficient to effectively sample large numbers of PE minima, without requiring temperature exchanges.

Following the completion of all SA-BH simulations, the stationary point database was expanded using single-ended transition state searches. In such a search, a

small perturbation is applied to a known local minimum, eigenvector following (Munro and Wales, 1999) is used to locate a nearby saddle point, and the two minima adjoining this saddle point are located using the LBFGS algorithm. All newly found stationary points and their connectivity are recorded in the database. At regular intervals, the database was depicted using disconnectivity analysis (Becker and Karplus, 1997), and all stationary points not connected to the global minimum within a reasonable PE threshold were removed.

When further single-ended searches resulted in no qualitative change in the disconnectivity graph, DPS (Wales, 2002) techniques were then employed to find further saddle points separating low-energy minima that had been overlooked by the single-ended searches. DPS was applied to pairs of low-energy minima separated by high barriers. In the DPS procedure used here, an approach based on Dijkstra's shortest-path algorithm (Dijkstra, 1959) was used to determine the fastest transition path (Evans and Wales, 2004), and double-ended transition state searches were applied between pairs of minima on this path, selected in order of increasing Euclidean separation (Carr and Wales, 2005). A double-ended transition state search involves the use of the doubly nudged elastic band (DNEB) method (Trygubenko and Wales, 2004) to identify saddle point candidates between the two endpoints, eigenvector following to precisely locate these saddle points, and the LBFGS algorithm to identify the connecting minima. DPS was iteratively applied until connection attempts had been attempted for all pairs of minima within 2.5 kcal/mol of the global minimum that were separated by energy barriers of more than 7.5 kcal/mol.

As a final step in ELM, the PE stationary point database was used to derive the FES of the system at 300 K and 350 K. For a chosen temperature, rate constants between PE minima connected by a saddle point were calculated using occupational probabilities derived from harmonic densities of states (Stillinger and Weber, 1984, Wales, 2003). PE minima connected by rate constants above 10^9 s^{-1} were combined to form new FE minima, and harmonic densities of states were then used to re-calculate energies for these minima and associated transition states (Evans and Wales, 2003). Additionally, optimal transition pathways between selected low PE minima were determined using the KSHORTESTPATHS algorithm (Carr and Wales, 2008), again using harmonic densities of states to calculate rate constants. Constant volume heat capacity curves also were computed based on the same harmonic approximation, by considering the contribution of each PE minimum in the database (Wales, 2017). Finally, overall phenomenological rate constants, summing over an ensemble of transition pathways, were calculated for comparison with the single-pathway rate constants using graph transformation (Trygubenko and Wales, 2006).

ELM was implemented using the software of Wales and co-workers, which is freely available on their website (Wales Group Home Page) . This software was interfaced with the CHARMM program for PE calculations (Brooks et al., 2009), which was modified to include the Steele potential.

5.2.2.2 Replica Exchange Molecular Dynamics

REMD requires identification of the temperature range over which the replicas are spread. The lower bound was set at the temperature of interest, 300 K. The upper

bound was selected by identifying the temperature at which the molecule was able to flip easily between the two observed predominant adsorbed orientations to the surface, which are as illustrated in Figure 5-1(a). The orientation of the molecule to the surface is best described by the backbone normal angle, which is defined as the angle to the graphite surface of the backbone normal vector, Figure 5-1(b). For a strongly adsorbed molecule, a backbone parallel to the surface is anticipated, corresponding to a backbone normal angle approaching either 90° or –90°, depending on the orientation as displayed in Figure 5-1(a).

The REMD upper bound temperature was determined by undertaking a series of 100 ns conventional MD simulations at temperatures ranging from 300 K to 1500 K at 200 K intervals. Snapshots of the molecule were taken at 1 ps intervals and, from those snapshots with a center of mass within 8 Å of the graphite surface, all those conformations with backbone normal angles above 60° and below –60° were classified into separate groups reflecting the orientations illustrated in Figure 5-1(a). The temperature at which the fractions of these orientations converged was chosen as the upper bound temperature. These conventional simulations were also used to determine the remaining REMD study details, which are given and justified in the following section.

During the REMD simulation, snapshots of each replica were taken directly prior to each attempted temperature exchange. After completion, the simulation was analyzed by applying a clustering algorithm to all snapshot structures at the lowest temperature, 300 K, adsorbed on the surface: the starting structure and 32 further snapshots had a center of mass more than 8 Å from the surface and were, thus,

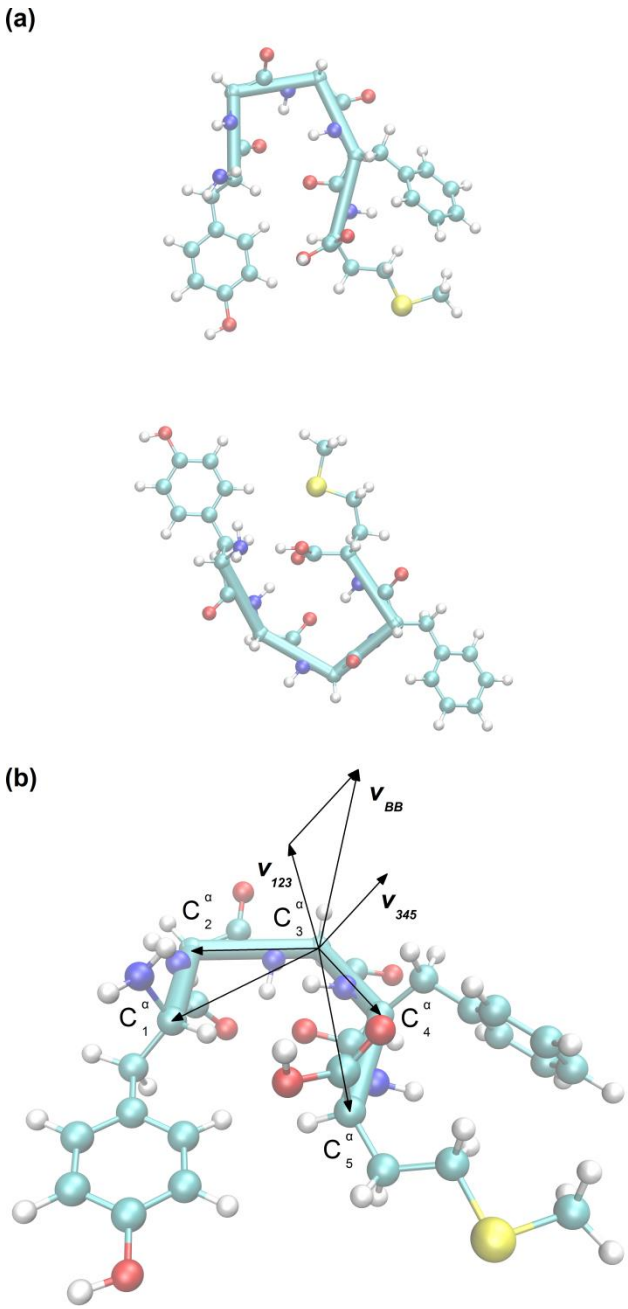

Figure 5-1. The orientation of met-enkephalin to the graphite surface: (a) the two predominant adsorbed orientations viewed from directly above the surface, which is not shown; and (b) the definition of the backbone normal vector, $\boldsymbol{v}_{BB}$, that is used to define the orientation of the molecule to the surface, which is the sum of $\boldsymbol{v}_{123}$ and $\boldsymbol{v}_{345}$ where the former is the vector cross product of the vectors from C^{α}_{3} to C^{α}_{2} and from C^{α}_{3} to C^{α}_{1}, and the latter the vector cross product of the vectors from C^{α}_{3} to C^{α}_{5} and from C^{α}_{3} to C^{α}_{4}.

discarded. Clustering involves grouping similar structures together according to an appropriate metric, often the root mean square deviation (RMSD), and thus provides a means of determining the main conformations found by the simulation and their prominence. While many algorithms and approaches exist to classify snapshots from MD simulations into clusters (Shao et al., 2007), the most useful and reliable algorithms require the calculation of the metric between every pair of snapshots. However, this was not computationally feasible for the number of snapshots produced in this study. Instead, a new clustering algorithm based on the methodology of Lyman and Zuckermann (2006) was used. This new algorithm is outlined here.

Starting from the full list of snapshots, the algorithm iteratively removes randomly selected snapshots from the list, along with all other snapshots within a specified cut-off distance. Here, this distance was set at 1 Å RMSD (calculated using only the C^{α} atoms), as this was found to generate an appropriate number and size of clusters for meaningful analysis, a finding consistent with our earlier work (Mijajlovic et al., 2013). Under Lyman and Zuckermann's approach, once all snapshots have been removed from the list, the set of randomly selected snapshots are termed 'reference structures', and each snapshot is allocated to a cluster depending on which reference structure it is nearest to. In the algorithm employed in this study, the LBFGS algorithm is used to determine the PE minimum associated with each snapshot, and the PE minima associated with the randomly selected snapshots are used as reference structures. The resulting clusters are then refined by iteratively applying two steps. Firstly, any clusters with reference

structures separated by less than a threshold of 0.5 Å RMSD are merged, with the reference structure of the smaller cluster being discarded; this threshold was identified based on detailed inspection of structures as a function of deviation from reference structures. Secondly, the associated minima of all snapshots in each cluster are tallied, and the most populated minimum becomes the new reference structure for that cluster. These steps are alternated until their successive application results in no change to the set of reference structures. Finally, the distribution of backbone normal angles for the snapshots in each cluster is examined, and if this distribution is multimodal (indicating a conformation that adsorbs on the surface in different ways), appropriate threshold angles are chosen to split the distribution into separate sub-clusters. Following this, the clusters are deemed final.

The self-consistency of the REMD simulation was analyzed by measuring the convergence of the cluster probability distribution (that is, the fraction of snapshots allocated to each cluster) for the 300 K replica. Two means were used: (a) self-referential convergence analysis (SRCA); and (b) constancy of cluster entropy (CCE). SRCA involves splitting the simulation into halves and comparing their cluster probability distributions; their similarity is an indication of the degree of self-consistency of the simulation (Lyman and Zuckerman, 2006). In CCE, the cluster entropy is calculated as an order parameter of the cluster probability distribution as a function of the elapsed time in the simulation:

$$S(t) = -\sum_{i=1}^{N} p_i(t) \log p_i(t) \qquad (5\text{-}2)$$

where $p_i(t)$ is the probability that a snapshot taken between the beginning of the simulation and time t belongs to cluster i, and N is the total number of clusters. A convergence in this cluster entropy over time also implies a self-consistent simulation (Sawle and Ghosh, 2016).

Constant volume heat capacities were also calculated from the REMD simulation, to provide an additional metric for the comparison of REMD with ELM. In the canonical ensemble, the heat capacity is given by:

$$C_V = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2} \qquad (5\text{-}3)$$

where $\langle E \rangle$ and $\langle E^2 \rangle$ are the average values of the energy and squared energy, respectively; k_B is Boltzmann's constant and T is the temperature. Average energies and squared energies are calculated at each of the temperatures used in the REMD simulation, by discarding the first 50 snapshots taken at each temperature and averaging over the rest.

MD and REMD simulations were carried out using Nanoscale Molecular Dynamics (NAMD) (Phillips et al., 2005). Analysis of simulation results, including the calculation of RMSD values and backbone normal angles, and the generation of molecular structure images used in this publication, was performed using Visual Molecular Dynamics (VMD) (Humphrey et al., 1996).

5.2.3 Study Details

5.2.3.1 Energy Landscape Mapping

A total of 100 SA-BH simulations were conducted, each of 250,000 steps. A temperature of k_BT = 5.0 kcal/mol was used for the first step, and this was decremented by 1.6×10^{-3} % after each step. The maximum step size for each dihedral angle was initially set to 360°, simulating complete randomization of the secondary structure. This was adjusted every 50 steps within the range of (0, 360°], being increased by 5% if the Metropolis acceptance ratio for the preceding 50 steps was greater than 0.5, and decreased by 5% otherwise. Starting points for single-ended transition state searches were produced by applying random deviations of up to 0.01 Å in each Cartesian coordinate from the local PE minima. The DNEB method was applied with 10 images and a maximum of 300 iterations. All applications of the LBFGS algorithm used a memory of the last 4 iterations, a maximum step size of 0.4 Å, initial guesses for the diagonals of the Hessian matrix of 0.1 $Å^2$mol/kcal, and a convergence criterion of the RMS gradient not exceeding 10^{-7} kcal/molÅ.

5.2.3.2 Replica Exchange Molecular Dynamics

All MD simulations used a time step of 1 fs. An upper bound REMD temperature of 1100 K was chosen according to the methodology outlined in Section 5.2.2.2. In order to determine the number of replicas and allocation of temperatures, the potential energy distributions from the conventional MD simulations were approximated as Gaussian distributions. The details were then chosen according to

the recommendations of Rathore *et al.* (2005), with a target exchange probability of 0.2 for all adjacent pairs of replicas. Using this approach, nine replicas were chosen for use in the REMD simulation, at temperatures of 300 K, 354 K, 418 K, 492 K, 579 K, 680 K, 799 K, 938 K and 1100 K. The exchange period and number of exchange attempts were chosen according to the criteria that the exchange period should exceed the autocorrelation time and the number of attempts should exceed the transit number by at least two orders of magnitude (Abraham and Gready, 2008). The conventional MD simulations produced autocorrelation times ranging from approximately 1 ps to 2.5 ps, and the number of replicas and target exchange probability in this study correspond to a transit number of approximately 125 (Abraham and Gready, 2008). Accordingly, an exchange period of 5 ps and 40,000 exchange attempts were chosen, yielding a total simulation length of 200 ns.

5.3 Results and Discussion

5.3.1 Performance of REMD Simulation

Figure 5-2 shows the exchange probability achieved in the REMD simulation by each adjacent pair of replicas. A steady increase in the exchange probability with temperature is evident, ranging from approximately 0.2 at 300 K to 0.25 at the highest. These exchange rates are within a range that corresponds to ideal mixing of the replicas (Rathore et al., 2005), although the variation does indicate that the temperatures chosen for the replicas are not distributed in the most computationally efficient way.

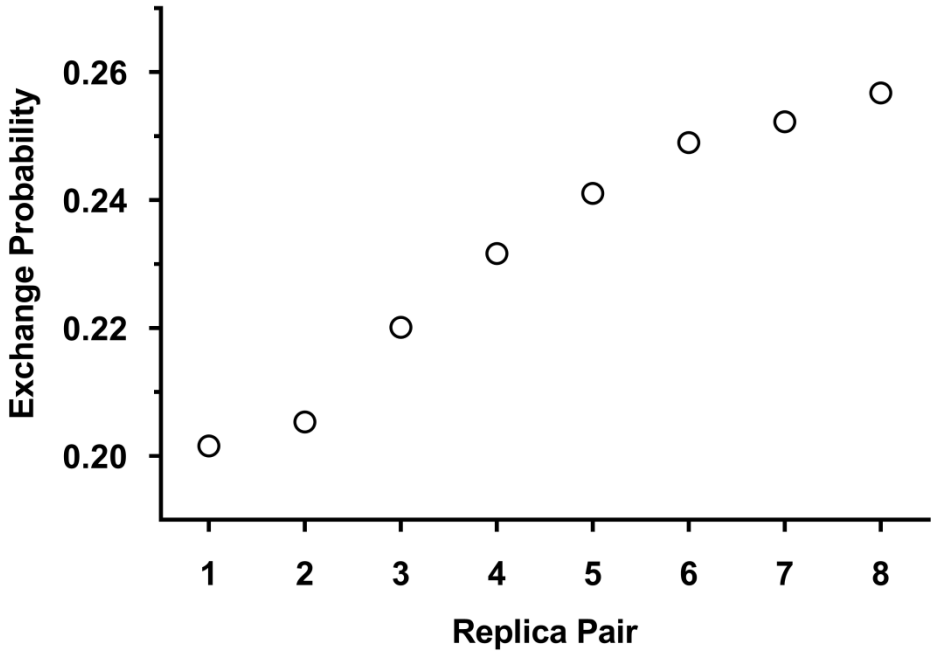

Figure 5-2. The REMD exchange probability for each pair of replicas where pairs are ordered by increasing temperature: (1) 300 K and 354 K, (2) 354 K and 418 K, and so forth.

Figure 5-3(a) and (b) display the cluster probability distributions of the 300 K replica for the first and second 100 ns halves of the REMD simulation. These are consistent in the sense that the probabilities for each of the first eight clusters, which account for approximately 79% of all structures, are well within an order of magnitude. Meanwhile, Figure 5-3(c) shows that the cluster entropy for the 300 K replica stabilizes very quickly at around 2.5, remaining within 5% of the final value after approximately 113 ns. These results are all indicative of a self-consistent simulation. It is worth noting that self-consistency does not guarantee convergence, as there may be areas of the FES entirely unexplored by the simulation (Sawle and Ghosh, 2016). However, given the steps taken to ensure

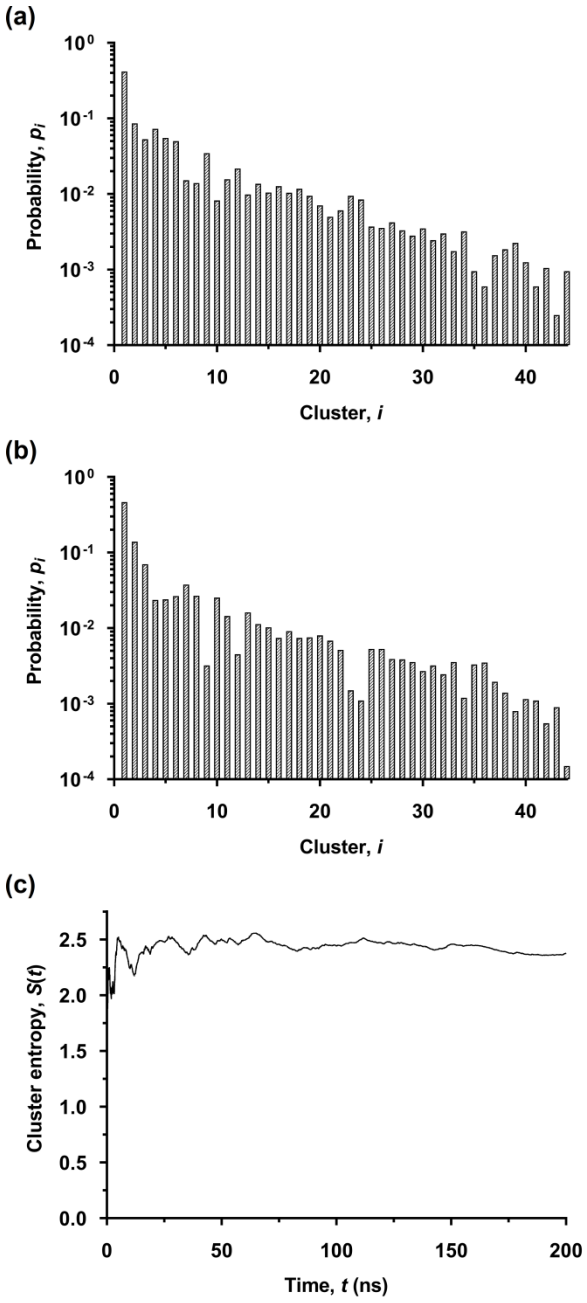

Figure 5-3. Results illustrating the self-consistency of the REMD simulation: (a) Cluster probability distribution of the 300 K replica in the first 100 ns of the REMD simulation; (b) cluster probability distribution of the 300 K replica in the second 100 ns of the simulation; and (c) variation of the cluster entropy of the 300 K replica with time.

sufficient mixing of the replicas, we are confident that this possibility has been minimized as far as reasonably practicable.

5.3.2 REMD Cluster Analysis

Figure 5-4(a) shows for each snapshot in the REMD simulation the backbone normal angle and the RMSD from the PE minimum reference structure of cluster A. It also gives an indication of the many varied structures the adsorbed molecule may take. As shown in Figure 5-4(b), which displays the cluster probability distribution for the entire simulation, a total of 44 clusters were identified from the REMD simulation using the procedure outlined above. The eight clusters that account for at least 2% of the total number of snapshots, which are denoted A to H in order of decreasing probability, are shown in color in Figure 5-4(a); these collectively account for nearly 79% of all snapshots. This demonstrates the success of the clustering algorithm in that these eight clusters are distinct with very little overlap.

Figure 5-5 displays the reference structure for each cluster, illustrating the conformation corresponding to each. The most common three conformations (A to C) are essentially flat to the surface, which is reflected by their backbone normal angles clustering towards ±90° as seen in Figure 5-4(a). The others possess more three-dimensional secondary structures, with various sites binding to the surface. In either case, the secondary structures vary appreciably, generally permitting intramolecular hydrogen bonding and interactions between various functional groups. Common motifs include interactions between the Phe and Met sidechains (structures A and B), the uncharged N- and C-termini (A and C), the glycine

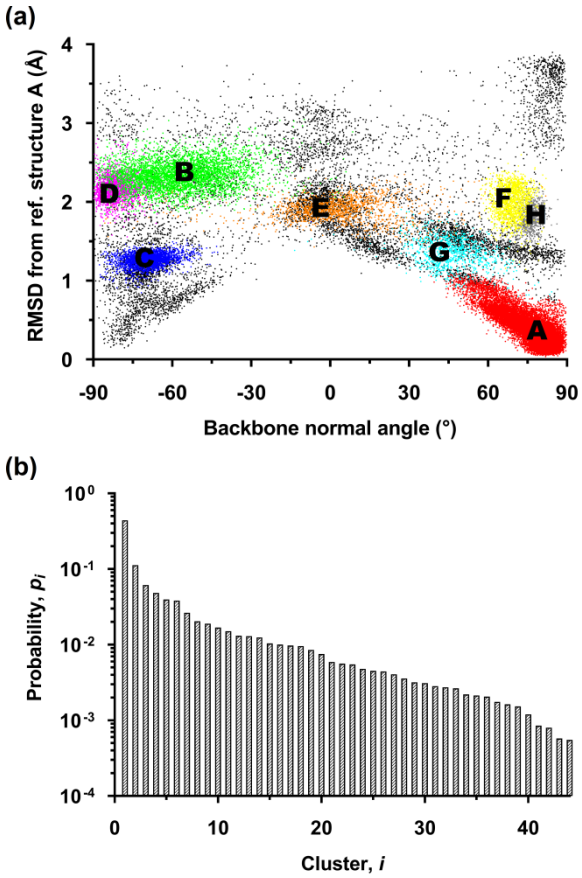

Figure 5-4. Results of the REMD clustering algorithm: (a) all adsorbed snapshots taken during the REMD simulation, plotted according to their backbone normal angle and RMSD from the reference structure of cluster A, with snapshots belonging to clusters A through H colored accordingly; and (b) cluster probability distribution, with clusters ordered from most populated to least.

strand and the Tyr sidechain (D and E) and the C-terminus and the hydroxyl group of the Tyr residue (F, G and H). Structures A and C, representing two of the three most populated clusters which account for more than half of all snapshots, are essentially equivalent to conformations identified in the authors' previous study of this system, which used both REMD and an evolutionary algorithm (Mijajlovic et al., 2011), indicating good agreement with that prior work. Structure E also is of

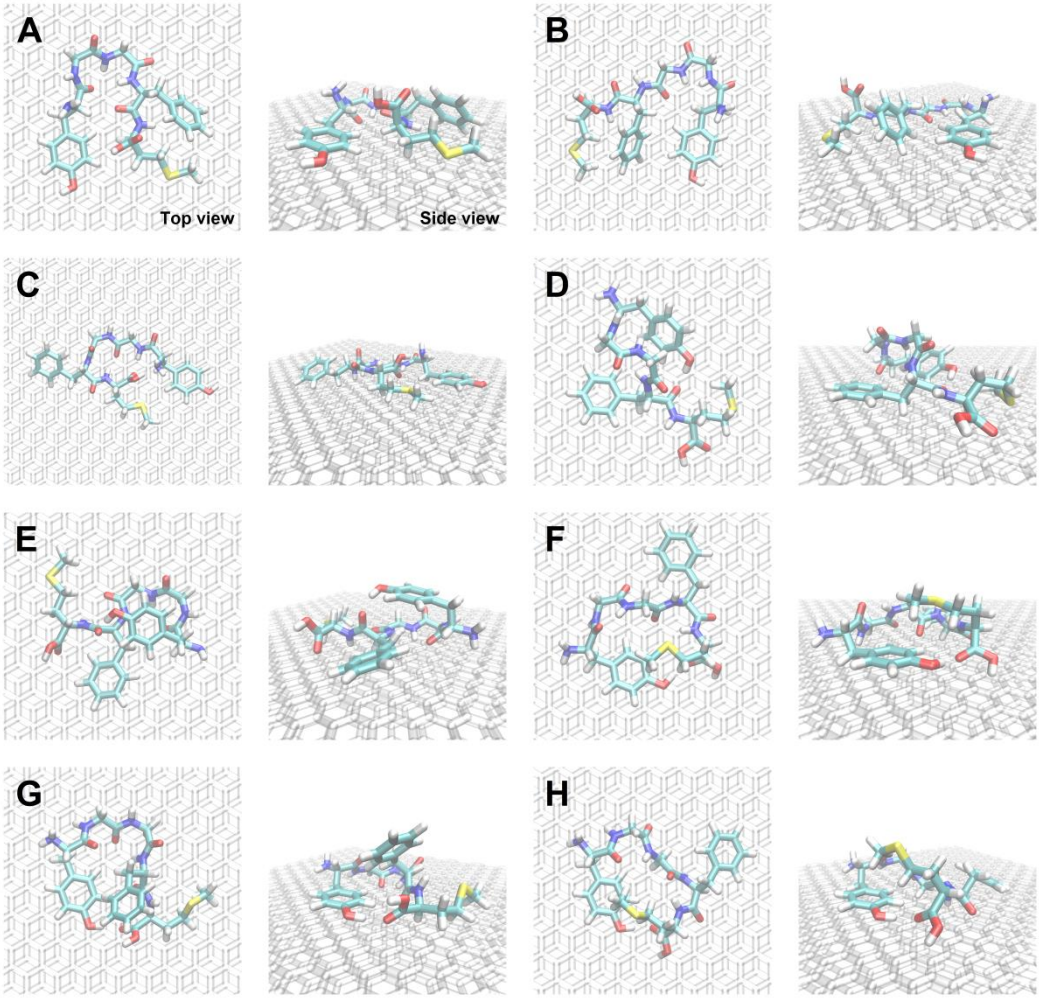

Figure 5-5. Reference structures for clusters A through H.

note for its twisted, S-shaped backbone. This phenomenon would result in the two component vectors of the backbone normal vector (as displayed in Figure 5-1) pointing in opposite directions with respect to the surface, accounting for the near-zero backbone normal angle observed in Figure 5-4(a).

5.3.3 ELM and Comparison with REMD

A total of 706,461 minima and 558,053 saddle points on the PES were located using ELM. As has been previously noted, the PES of peptide/interface systems

contain so many stationary points that its exhaustive mapping is impractical (Wright and Walsh, 2013, Ross-Naylor et al., 2017). However, since only the lowest PE minima and the stationary points connecting them are likely to make a significant contribution to the molecule's dynamics (Wales, 2003), the evaluation of these points provides an indicator of how extensively ELM has probed the PES. Figure 5-6, which shows the variation of the number of PE minima within 2.5 kcal/mol of the global minimum over the course of the mapping, indicates that nearly all of these low-lying minima were identified within the first 100,000 minima discovered by ELM. The iterative DPS procedure then ensured the paths between all these low-lying minima were well sampled.

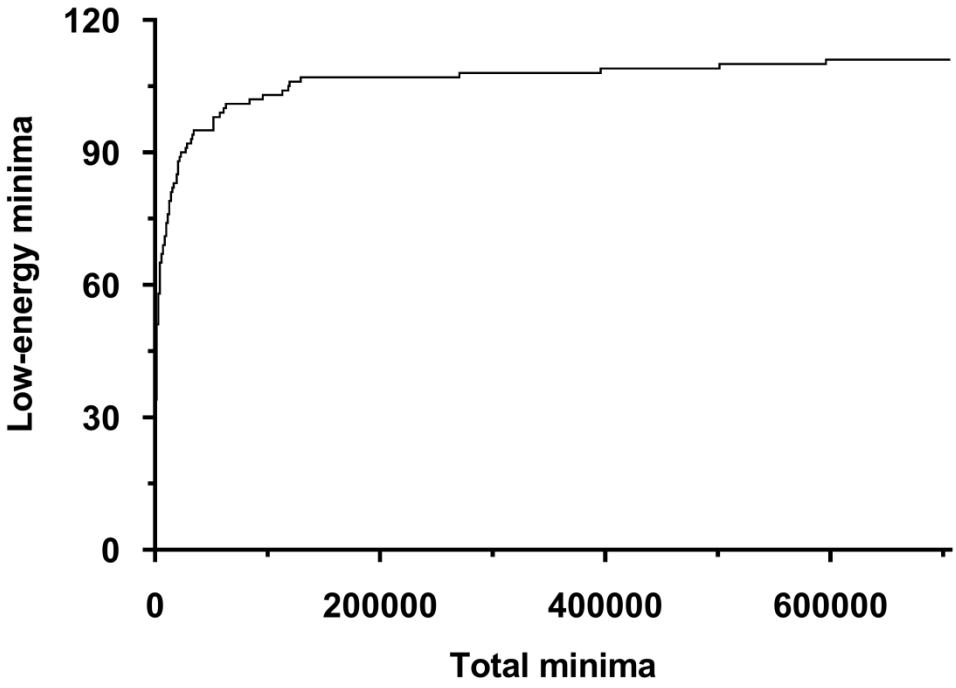

Figure 5-6. Variation of the number of low-lying minima (i.e. within 2.5 kcal/mol of the global minimum) with the total number of minima found through the ELM simulation.

A further indication of ELM's comprehensiveness can be gathered from comparison with the REMD simulation, in particular by tallying the number of REMD snapshots corresponding to PE minima that were not found using ELM, or for which no connection was found to the global minimum. Since no significantly low-lying minima were connected to the global minimum by a PE barrier greater than 20 kcal/mol, this was used as a cut-off for analysis. The total number of snapshots, 39,968, is significantly lower than the number of minima found by ELM, and as a result, most of the minima are inevitably not associated with any snapshots. On the other hand, snapshots that correspond to a minimum not discovered or not connected to the global minimum by ELM indicate that the PES was not sufficiently mapped in that region of configurational space.

Table 5-2 shows that of the PE minima associated with each REMD snapshot, 4.5% were not located at all by ELM, while a further 6.0% were not connected to the global minimum, confirming that the ELM procedure was not entirely exhaustive. Most of the largest REMD-identified clusters were almost completely characterized, however, the only prominent exception being cluster E, for which approximately 16% of associated minima were not connected to the global minimum below the threshold, and a further 7% were not discovered at all. The smaller clusters collectively denoted as 'the rest' were generally less well sampled, which is to be expected given they correspond to higher-energy regions of the PES that were not prioritized when conducting ELM.

Table 5-2. Proportion of REMD-discovered snapshots for which the associated PE minimum was discovered or otherwise by ELM. Only connections via a maximum-energy saddle point below a PE threshold of 20 kcal/mol are considered.

REMD-identified cluster	**REMD snapshots**	**number of REMD-discovered PE minima discovered by ELM and connected to the global minimum**		**number of REMD-discovered PE minima discovered by ELM but not connected to global minimum**		**number of REMD-discovered PE minima not discovered by ELM**	
		number	**fraction**	**number**	**fraction**	**number**	**fraction**
A	17595	17189	0.977	247	0.014	159	0.009
B	4496	4302	0.957	106	0.024	88	0.020
C	2454	2423	0.987	19	0.008	12	0.005
D	1928	1722	0.893	92	0.048	114	0.059
E	1581	1223	0.774	247	0.156	111	0.070
F	1528	1457	0.954	29	0.019	42	0.027
G	1055	977	0.926	36	0.034	42	0.040
H	813	737	0.907	42	0.052	34	0.042
the rest	8518	5750	0.675	1563	0.183	1205	0.141
total	39968	35780	0.895	2381	0.060	1807	0.045

Figure 5-7 displays the lowest PE stationary points connected to the global minimum and their connectivity using a disconnectivity graph. A disconnectivity graph pictorializes an energy landscape by representing individual energy minima as termini, which are joined at nodes corresponding to the maximum-energy saddle point on the optimal path connecting them (Becker and Karplus, 1997). The PE disconnectivity graph displayed here is broad and contains a large number of small ‘funnels’ and individual minima that are separated by high energy barriers. This phenomenon, known as frustration, occurs commonly when a system has a number of dissimilar low-energy structures, and also when geometric barriers hinder a molecule from transitioning between two related structures

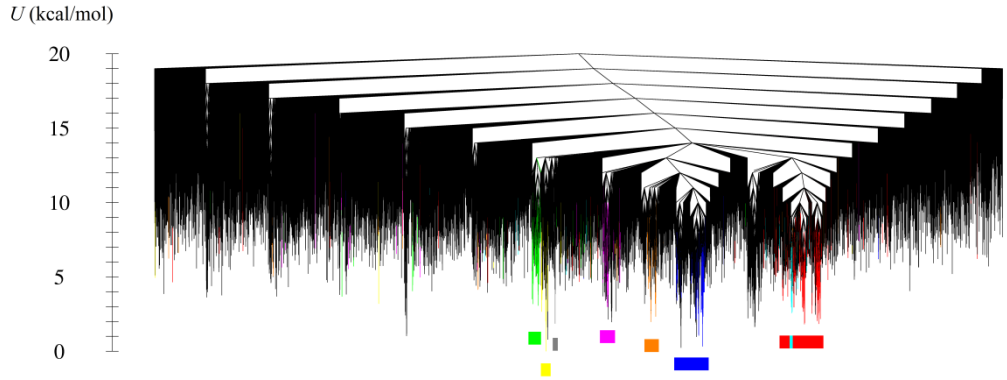

Figure 5-7. PE disconnectivity graph. All minima connected to the global minimum by a maximum-energy saddle point below U = 20 kcal/mol are shown. Colored minima are the associated minima of REMD snapshots from the most prominent clusters: A (red), B (green), C (blue), D (pink), E (orange), F (yellow), G (light blue) and H (grey). Colored bars indicate regions of the disconnectivity graph dominated by a particular cluster.

(Wales, 2003). Both of these are certainly possible in the system considered here: the REMD simulation results suggest met-enkephalin adsorbed on graphite may take on any of several diverse conformations as displayed in Figure 5-5, while the presence of the graphite surface may provide a significant geometric obstacle to some conformational transitions.

The colored minima in Figure 5-7 correspond to snapshots found in the REMD simulation. These provide a useful basis for comparing the two methodologies, and a number of conclusions may be drawn. ELM was successful in characterizing most of the clusters obtained using REMD, with minima corresponding to the same cluster typically near each other on the graph. Whilst the global PE minimum sits within cluster F, the small number of minima located within the funnel corresponding to this cluster means entropic effects may be

small, making it potentially less favored than other structures. Conversely, clusters A and C also contain PE minima close to but greater than the global PE minimum but occupy significant portions of the disconnectivity graph corresponding to many minima and, thus, presumably more favored entropically. These observations are reflected in the REMD results, which show conformations A and C are significantly more favored at 300 K than F. It is worth noting that the sections of the graph corresponding to each cluster are relatively narrow compared to the breadth of the graph, indicating that most minima do not correspond closely to the eight most prominent clusters identified with REMD, instead representing other configurations that REMD failed to characterize.

There are discrepancies between the results obtained from the two techniques, however. The funnel containing cluster B, while well-defined, is both narrow and has relatively high PE, implying that this conformation should be relatively rare. However, REMD indicates that it is the second most preferred conformation. According to Table 5-2, ELM was highly successful in locating the minima corresponding to cluster B, and as such, the reason for this cluster's underrepresentation in the ELM results compared to the REMD counterparts is not immediately obvious. A possibility, given its wide range of backbone normal angles as displayed in Figure 5-4(a), is that it covers a large area of configurational space that has relatively few minima on the PES (i.e. a large flattish hyper-plane), and is thus favored by entropy in a manner that is not reflected in disconnectivity analysis.

Cluster G presents another point of inconsistency between the results obtained through ELM and REMD, in that it appears to exist on the disconnectivity graph as a subset of cluster A identified through the REMD analysis. The low energy barrier between the two clusters suggests that they may interconvert as readily as slightly different variants of the A conformation. This contrasts with Figure 5-4(a), which quite clearly distinguishes the two clusters according to their order parameters. It has been observed previously that order parameters can fail to accurately represent the relationships between structures in kinetic transition networks (Wales and Head-Gordon, 2012, Wales, 2015). As such, this result is not surprising, and illustrates the utility of ELM and disconnectivity analysis in yielding a richer, more insightful description of a system's conformations and the relationships between them compared to analysis based on order parameters only.

Figure 5-8 shows the FE disconnectivity graphs of the system at 300 K and 350 K. The disconnectivity graph at 300 K indicates that the FES at this temperature retains the essential characteristics of the PES, although various funnels are condensed into individual minima by the technique of amalgamating groups of PE minima. There exist low FE minima corresponding to clusters A, C, D, F, G and H, while the PE minima corresponding to clusters B and E are assimilated within clusters A and D, respectively. At this point, the global FE minimum belongs to cluster A, in accordance with the results of the REMD simulation, while the other five minima (i.e. C, D, F, G and H) are all within about $2k_BT$ of the global minimum. It is known (Latour, 2008) that the ratio of the sampling probabilities

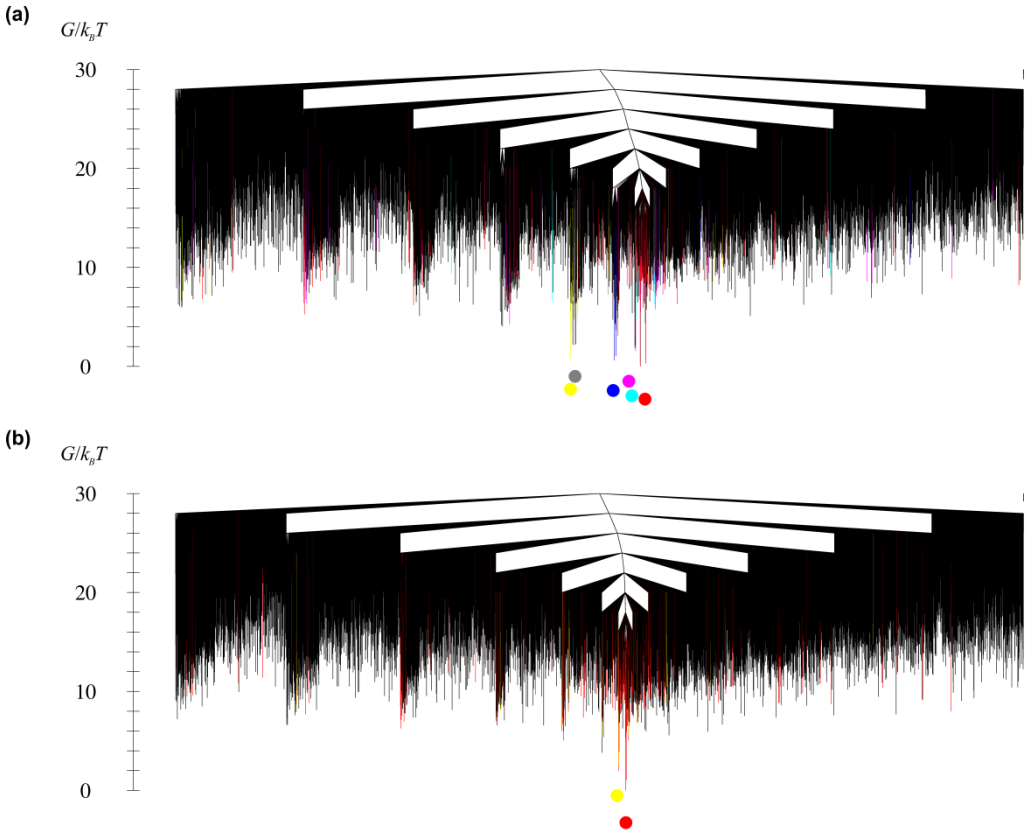

Figure 5-8. FE disconnectivity graphs at: (a) 300 K and (b) 350 K. All minima connected to the global minimum by a maximum-energy saddle point below $G = 30k_BT$ are shown. Colored minima are the associated minima of REMD snapshots from the most prominent clusters: A (red), B (green), C (blue), D (pink), E (orange), F (yellow), G (light blue) and H (grey). Colored dots indicate the lowest minimum on the disconnectivity graph corresponding to each cluster.

of two states *i* and *j*, p_i/p_j, is related to the difference in FE between the corresponding minima, ΔG_{ij}, according to the equation:

$$\frac{p_i}{p_j} = e^{-\frac{\Delta G_{ij}}{k_B T}} \qquad (5\text{-}4)$$

Based on this equation, the FE differences at 300 K obtained through ELM and the probabilities at the same temperature yielded by REMD as displayed in Figure

5-4(b) are in good agreement. There also exist other prominent FE minima at slightly higher energies, which correspond to clusters with lower probabilities. The assimilation of clusters B and E is surprising, given their graphical separation from the parent clusters A and D in Figure 5-4(a), and the high energy barriers evident on the PE disconnectivity graph. This, firstly, presents further examples of relationships between structures that are not evident from REMD, and secondly, illustrates the importance of FE analysis in ELM, which may yield information overlooked by considering the PE only. Meanwhile, it is notable that the precise ordering of these minima obtained through ELM differs from REMD. Inaccuracies due to the harmonic approximations used in calculating free energies, or an undersampling of PE stationary points, present possible explanations for this. Another potential source of error is the surface corrugation effect of the atomistic graphite model used for REMD, which is overlooked by the smooth approximation utilized in ELM. However, this effect has been demonstrated to be extremely small except at very low temperatures (Jiang et al., 1993), and is therefore unlikely to account for any observable discrepancies in results.

At 350 K, the PE minima for all clusters other than F are connected by rate constants above the threshold of 10^9 s^{-1} utilized in forming FE minima. This causes them all to merge to form a well-defined global FE minimum corresponding to all of these conformations, while F retains a distinct FE minimum. As a whole, this FE graph shows that most of the major conformations of adsorbed met-enkephalin interconvert relatively easily even at low temperatures. This implies that 1100 K was an excessive choice as an upper

bound temperature for REMD. A lower choice would still have provided sufficient sampling of conformational space, while lessening the computational effort due to the fewer replicas required.

Further comparison between the two methodologies may be made by comparing the constant volume heat capacity curves obtained from each, which are displayed in Figure 5-9. Since the heat capacities calculated from REMD are directly computed from the canonical ensemble, they provide an effective means of assessing the accuracy of those calculated from ELM. As Figure 5-9 shows, ELM overestimates the heat capacity by about 10%. It correctly predicts a peak at approximately 400 K, but significantly underestimates its strength. ELM also

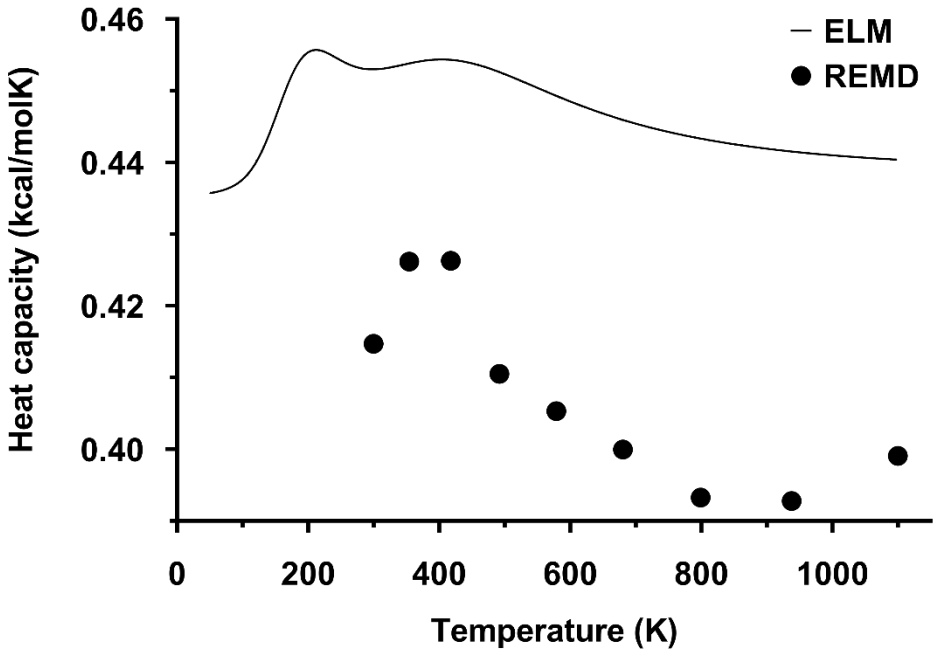

Figure 5-9. Constant volume heat capacities, as a function of temperature, calculated from the stationary point database obtained from ELM, and separately from the REMD simulation.

predicts a second heat capacity peak at about 200 K, which is below the minimum temperature used for REMD simulation here. It has been suggested that calculations of heat capacity features from ELM by using harmonic approximations are most reliable at low temperatures, where the contributions of low PE minima are dominant, and the assumption of harmonicity is most accurate (Wales, 2017). At higher temperatures, anharmonic densities of states and the undersampling of higher PE minima become prominent factors, and these likely account for the differences in results here. It is notable that the maximum REMD temperature, 1100 K, corresponds to an increase in heat capacity that is likely due to the peptide's detachment from the surface; this is not taken into consideration by ELM, since all PE minima correspond to adsorbed structures. Aside from this, the similarity of the profiles of the curves shown in Figure 5-9 suggests that, even in the absence of comprehensive enumeration of PE minima, calculation of the heat capacity using harmonic approximations may yield reasonable qualitative conclusions regarding the heat capacity features of the landscapes of adsorbed peptides, but they should not be relied on for the production of accurate quantitative data.

Figure 5-10 displays a transition pathway connecting three of the REMD-determined reference structures that were most prominent on the disconnectivity graphs: those corresponding to clusters A, C and F. The pathway consists of the best path between F and A, and the best path between A and C. The best path between F and C (not shown) contains a significantly larger number of stationary points and lower rate constant, and hence is unlikely to contribute as significantly

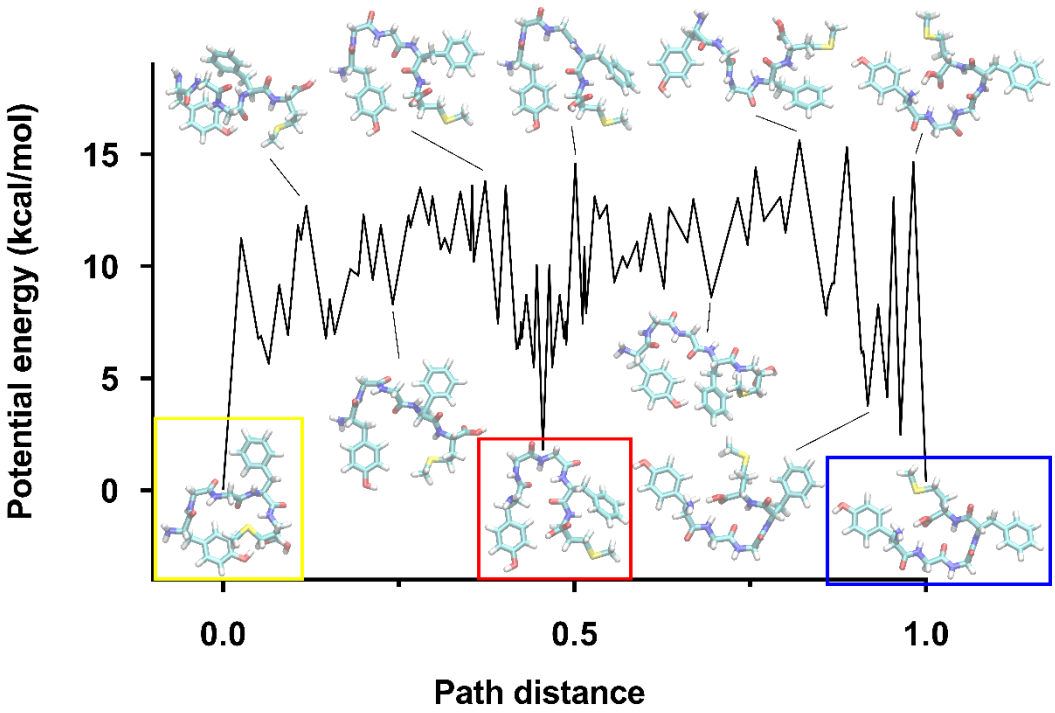

Figure 5-10. Variation of the PE along the transition pathway between the reference structures for cluster F (left, yellow), A (center, red) and C (right, blue). Selected structures along the pathway are shown as insets. The reference PE is that of structure F, the global PE minimum. The path distance is the minimized Euclidean distance between neighboring stationary points, scaled to the total length of the pathway. Lines between stationary points are provided as a guide to the eye only.

to the system's dynamics as the shown path. As displayed in the inset structures on Figure 5-10, the transition represents a rearrangement of the adsorbed peptide from a compact, folded structure, F, through an intermediate, A, to a flat, extended structure, C. Such folding and unfolding mechanisms are characteristic of proteins, and it has been noted that solid surfaces play a role in protein unfolding (Marruecos et al., 2018). These results thus demonstrate a significant advantage of

ELM as applied to adsorbed peptides, in its ability to characterize and provide insight into such processes.

Rate constants for the transition path as shown in Figure 5-10 at 300 K, 350 K and 400 K are displayed in Figure 5-11(a). These are split into the forward and backward transitions between F and A, and between A and C. It is notable that the A→F rate constants in all cases are 1 to 2 orders of magnitude greater than all others, suggesting that the F conformation should be most favored, at odds with

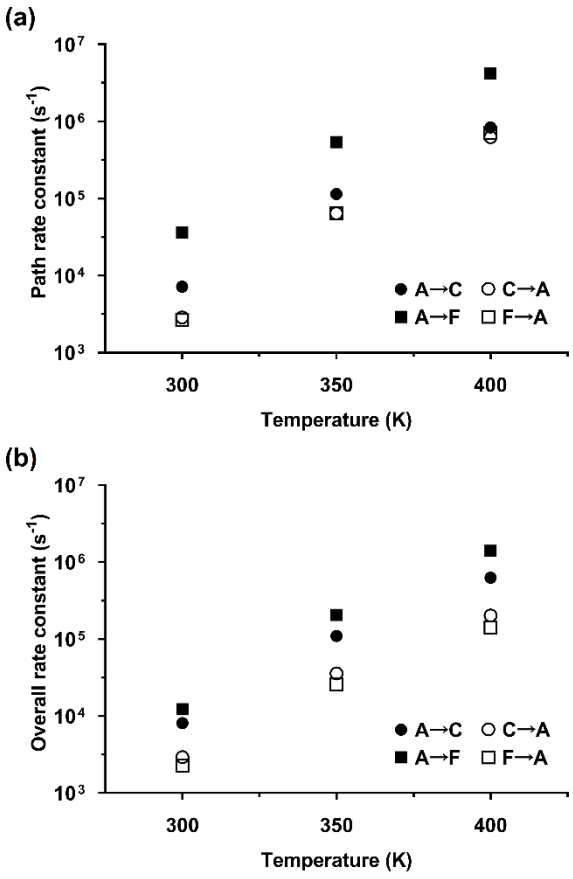

Figure 5-11. (a) Variation of the rate constants with temperature for the transition paths between structures F, A and C as shown in Figure 5-10, and (b) variation of the overall phenomenological rate constants between F, A and C with temperature.

the results for both REMD and free energy analysis of the ELM-derived stationary point database. However, the rate constants shown in Figure 5-11(a) only account for a single pathway, and in reality, an ensemble of transition paths would contribute to the kinetics of the system. The phenomenological rate constants displayed in Figure 5-11(b) take this transition path ensemble into account: here, it is notable that the A→C rate constants are now competitive with those from A→F, although the F conformation is still preferred. This apparent preference may be due to the funnel corresponding to F containing a very well-defined reference structure with no competing low minima in its vicinity, as displayed in Figure 5-7, while A and C contain numerous such competing minima. As a result, the probabilities of the system corresponding precisely to the reference structures for A and C are artificially reduced. It is also notable that most of the phenomenological rate constants are lower than the pathway rate constants, by up to an order of magnitude. This indicates that the complexity of the pathway ensemble generally has the effect of slowing down the system's transitions relative to the rates expected based on the most favored path alone.

5.4 Conclusions

Two techniques for exploring energy landscapes characterized by high energy barriers and inherently long timescales were applied to study the adsorbed conformations of a flexible peptide, met-enkephalin, at a gas/graphite interface. A number of significant conformations were identified using REMD, and the energy barriers and natures of the energy landscapes, which included significant frustration at low temperatures, were further elucidated using ELM. Both

methodologies were in substantial agreement on several key conformations, and their energetic favorability relative to one another. It can be concluded that ELM offers greater insight into the conformational space and the kinetics and dynamics of transition processes compared to REMD, as well as information about energy barriers between conformations and the effects of entropy.

The challenge remains of ensuring ELM is performed to the highest level of completion possible, as its efficacy depends significantly on whether all important minima and saddle points are found. The procedure followed in this study involving single-ended and double-ended transition state searches represents a step towards achieving this, but further work is needed towards more rigorous methods of quantifying the degree to which the PES is mapped. The difficulties in extensively evaluating stationary points pose a particular obstacle to the calculation of properties that depend upon mapping high PE regions of the PES, such as the heat capacity at moderate to high temperatures. PE minimum sampling schemes such as basin sampling (Wales, 2013), however, provide a promising means of addressing this (Wales, 2017).

Although REMD and ELM were undertaken separately in this study, it would be possible to address the limitations and challenges of each method by using both techniques in tandem, provided the size of the system and the computational resources available allow for this possibility. Using REMD and the presently used clustering approach, one may identify PE minima corresponding to the most significant conformations. DPS along with single-ended searches may then be used to investigate the configurational space between these minima, and build up

an essentially complete picture of the energy landscape that includes all of the conformations found using REMD. This may provide a more efficient and reliable means of conducting ELM than attempting to build a stationary point database from scratch, and is worthy of future investigation.

Given many applications of peptide and protein adsorption occur in aqueous environments, it is desirable to incorporate the use of solvent in studies of this phenomenon. Unfortunately, implicit solvation models have proven largely inadequate in modeling peptide adsorption at liquid/solid interfaces (Latour, 2008). The requirement of explicit solvent to reliably model the system greatly increases the number of degrees of freedom, making both ELM and REMD more challenging (Frenkel and Smit, 2002, Fukunishi et al., 2002, Wales, 2003). All-atom force fields designed for use in biointerfacial systems may provide a way forward for the use of implicit solvation in studies such as this, but their development and validation is ongoing (Walsh, 2017). Finally, the feasibility of using ELM to characterize larger and more complex systems of biomolecule adsorption is worth investigating. In particular, since the aggregation or assembly of peptides on surfaces is relevant to a number of its applications (Care et al., 2015, Pagel and Beck-Sickinger, 2017, Slocik and Naik, 2017, Walsh and Knecht, 2017, Mas-Moruno, 2018), the adsorption of multiple peptides at a surface is a topic of interest, and is the subject of ongoing work.

Statement of Authorship

Title of Paper	Energy Landscapes of a Pair of Adsorbed Peptides
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Ross-Naylor, J.A., Mijajlovic, M., & Biggs, M.J. 2020. Submitted to J. Phys. Chem. B.

Principal Author

Name of Principal Author (Candidate)	James Ross-Naylor
Contribution to the Paper	Performed computational work, interpretations and analysis of results. Wrote manuscript.
Overall percentage (%)	90
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date: 22 Jan 2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Milan Mijajlovic
Contribution to the Paper	Supervised conducting of work. Provided research suggestions.
Signature	Date: 24/1/2020

Name of Co-Author	Mark Biggs
Contribution to the Paper	Supervised conducting of work. Provided research suggestions. Assisted in interpreting results. Evaluated and edited manuscript.
Signature	Date: 29/1/2020

Please cut and paste additional co-author panels here as required.

6 Energy Landscapes of a Pair of Adsorbed Peptides

6.1 Introduction

Applications of peptide adsorption are widespread, including the formation, functionalization and impact of nanoparticles (Slocik and Naik, 2017, Walsh and Knecht, 2017, Karim et al., 2018, Kojima et al., 2018, Boge et al., 2019, Quan et al., 2019), biosensing (Demir et al., 2016, Liang et al., 2016, Chen and Nugen, 2019), and biomedicine (Povimonsky and Rapaport, 2017, Yang et al., 2017, He et al., 2018, Huq et al., 2018, Mas-Moruno, 2018, Ortiz-Hernandez et al., 2018). Biomolecular adsorption is also of relevance to processes that occur in nature, and polypeptide segments of biomolecules are frequently used to study these processes, which include bodily reactions to foreign entities (Docter et al., 2015, Trindade et al., 2016), cell adhesion (Inoue et al., 2018), and the fibrillation of the amyloid-beta protein that is thought to have a role in Alzheimer's disease (Bellucci et al., 2016, Liu et al., 2016, Bellaiche and Best, 2018). This wide prevalence of peptide adsorption means it has attracted much study, both computational and experimental (Rabe et al., 2011, Heinz and Ramezani-Dakhel, 2016, Ozboyaci et al., 2016a, Ramakrishnan et al., 2017). Computational studies of peptide adsorption have yielded a significant degree of insight into the conformations formed by adsorbed peptides, and the key characteristics of adsorption processes (Latour, 2008, Rabe et al., 2011, Heinz and Ramezani-Dakhel, 2016, Ozboyaci et al., 2016a, Ramakrishnan et al., 2017). However, many such studies simulate only the adsorption of a single molecule, and adsorption studies that incorporate multiple peptides are typically limited to determining the

adsorbed conformations on a short timescale (Liu et al., 2016, Zhang and Sun, 2018). Since, in reality, peptide adsorption processes often involve multiple molecules adsorbing together, it is desirable to gain a deeper understanding of such systems.

Since processes involving adsorbed peptides typically occur over long time periods (Ross-Naylor et al., 2017, Walsh, 2017), required computational effort is a limiting factor in computational simulation of such processes (Rabe et al., 2011, Ozboyaci et al., 2016a). As the computational effort typically scales in a nonlinear way with the number of degrees of freedom (Frenkel and Smit, 2002), this factor becomes even more of an issue as the number of peptides increases. Coarse-graining models offer one way of addressing this issue (Heinz and Ramezani-Dakhel, 2016), but much development and validation work still remains to be done before such models can be used to model peptide adsorption with confidence (Ramakrishnan et al., 2017).

Energy landscape mapping (ELM) provides an alternative approach for computational study of atomic and molecular processes that is suited to probing the long timescales and rare events that typically characterize peptide adsorption (Berry, 1993, Wales, 2003, Wales and Bogdan, 2006, Wales, 2018). Although the exact details of its implementation vary, ELM may be generally defined as the use of mathematical algorithms to locate minima and saddle points on the potential energy surface (PES). A minimum corresponds to a stable conformation, while a first-order saddle point with steepest-descent paths leading to two minima represents a transition state between these minima. Constructing large-scale

connected databases of these minima and saddle points can yield information about the underlying potential energy and free energy surfaces, and detailed path information and rates can be obtained through discrete path sampling (DPS) (Wales, 2002). In recent work, the authors have demonstrated the applicability of ELM to a single adsorbed peptide (Ross-Naylor et al., 2017, Ross-Naylor et al., 2020). Here we demonstrate its application to two met-enkephalin peptides at a gas/graphite interface. The models, methodologies and study details are outlined in the following section. Results are then presented and discussed, followed by the key conclusions and an outline of future areas of study.

6.2 Methodology

6.2.1 Model

In a previous study (Ross-Naylor et al., 2020), the authors investigated a single met-enkephalin molecule above a flat, infinite graphite surface. This model is extended here to two met-enkephalin molecules (Tyr-Gly-Gly-Phe-Met), modeled using the CHARMM36m force field (Huang et al., 2017). However, unlike the previous study in which the uncharged form was investigated due to its greater flexibility and thus interest, here the molecules were represented in the zwitterionic form (i.e. NH_3^+ and COO^- at the N- and C-termini, respectively), since it was presumed that the interactions of the charged termini would result in a more definite coupling of the peptides. A similar study on the uncharged form will be considered in future work. The CHARMM force field was not symmetrized, which may lead to a rougher PES/FES compared to the more correct symmetrized

force field and some invalid stationary points corresponding to permutational isomers of the zwitterionic C-terminal residue (Małolepsza et al., 2010); this was considered in the analyses of the PES and FES undertaken here.

Graphite was represented as two graphene layers, since it has been previously observed that layers beyond the first two have negligible interaction with adsorbing molecules (Braun et al., 2002). The Steele potential was used to model interactions between the peptide atoms and the surface. This potential is given by the following equation (Steele, 1973):

$$E_s = 2\pi\rho \sum_j \sum_{l=0}^{L-1} \varepsilon_{sj}\sigma_{sj}{}^2 \left[\frac{2}{5}\left(\frac{\sigma_{sj}}{z_j+l\Delta}\right)^{10} - \left(\frac{\sigma_{sj}}{z_j+l\Delta}\right)^{4}\right] \qquad (6\text{-}1)$$

where $l = 0, \ldots, L-1$ is a counter over the layers of solid atoms in the surface up to the maximum, L, ρ the density of atoms in each of the layers, Δ the distance between the layers, z_j the distance from the surface of peptide atom j, and ε_{sj} and σ_{sj} the Lennard-Jones energy and length parameters, respectively. The Lennard-Jones parameters were calculated from the corresponding parameters for the surface, ε_s and σ_s, and the atom, ε_j and σ_j, using the Lorentz-Berthelot rules. The surface parameters are summarized in Table 6-1.

6.2.2 Methods

The ELM procedure applied here is similar to that used in the authors' previous study (Ross-Naylor et al., 2020). Firstly, a basin-hopping procedure with simulated annealing (SA-BH) was used to build an extensive database of local potential energy (PE) minima. Basin-hopping consists of a Monte Carlo

Table 6-1. Steele model parameters for graphite.

parameter	value	reference
L	2	[a]
ε_s	0.05564 kcal/mol	[b]
σ_s	3.40 Å	[b]
ρ	0.3807 atoms/Å^2	[c]
Δ	3.3555 Å	[c]

[a] (Braun et al., 2002)

[b] (Steele, 1973)

[c] Derived from literature bond length and unit cell height data (Trucano and Chen, 1975).

simulation on a transformed PES, where every point is mapped to the locally minimized PE. This has the effect of eliminating the difficulties posed by PE barriers that are excessively high compared to the neighboring minima (Wales and Doye, 1997). Starting from a given structure, a random perturbation is applied followed by local PE minimization, and the Metropolis criterion is then applied to accept or reject the step. This procedure is repeated for a specified number of steps. The simulated annealing portion of the methodology involves gradually reducing the temperature used in application of the Metropolis criterion, allowing the algorithm to seek out lower-lying minima. In this study, the steps were taken by randomly perturbing the backbone dihedral angles of the met-enkephalin molecules up to a maximum step size, and minimization was carried out using the limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm (Liu and Nocedal, 1989). A large number of SA-BH simulations were run independently in parallel, in order to quickly locate a large number of local minima and allow for the possibility that some simulations may converge to a

region of the PES that does not contain the global minimum, which is a noted disadvantage of simulated annealing (Wales and Doye, 1997). Each simulation was initialized with a random structure and high starting step size for perturbations, ensuring effective randomization of the results.

Given the database of local minima obtained by SA-BH simulations, transition state searches were then applied to find saddle points connecting these minima, as well as additional minima that had not initially been found using SA-BH. Single-ended transition state searches were first used to probe the neighborhood of all known minima. In every such search, starting from a selected minimum, a small perturbation is applied, eigenvector following (Munro and Wales, 1999) is used to locate a nearby saddle point, and the two minima adjoining this saddle point are located using the LBFGS algorithm. All newly found stationary points and their connectivity are recorded in the database.

When each known minimum had been used to seed a single-ended search a specified number of times, with additional searches conducted using low-lying minima within 5 kcal/mol of the global minimum, DPS (Wales, 2002) techniques were then employed to continue searching for overlooked saddle points. Disconnectivity analysis (Becker and Karplus, 1997) was used to identify minima within 5 kcal/mol of the global minimum that were separated by barriers of more than 10 kcal/mol, or disconnected from the global minimum entirely. DPS was then applied to all minima thus identified. For minima disconnected from the global minimum, an approach based on Dijkstra's shortest-path algorithm (Dijkstra, 1959) was used to search for connecting saddle points in order to

establish a connected pathway of saddle points and minima (Carr et al., 2005). For pairs of low-lying minima connected to the global minimum but separated by a high PE barrier, a separate approach also based on Dijkstra's algorithm was used to determine the fastest transition path between the two endpoints (Evans and Wales, 2004), and double-ended transition state searches were applied between pairs of minima on this path, selected in order of increasing Euclidean separation (Carr and Wales, 2005). A double-ended transition state search involves the use of the doubly nudged elastic band (DNEB) method (Trygubenko and Wales, 2004) to identify saddle point candidates between the two endpoints, eigenvector following to precisely locate these saddle points, and the LBFGS algorithm to identify the connecting minima.

When attempts had been made to connect all low-lying minima in this way, significant well-defined structures were identified using disconnectivity analysis, and DPS was applied to refine the minimum energy pathway (MEP) between each pair of such structures. The MEP was calculated at each stage using the KSHORTESTPATHS algorithm (Carr and Wales, 2008), with rate constants estimated using harmonic densities of states (Stillinger and Weber, 1984, Wales, 2003). New saddle points and minima were found using the DPS procedure as previously stated, and also by applying the UNTRAP algorithm (Strodel et al., 2007), which attempts to connect the endpoint structures with other minima near in configurational space but separated by high barriers. When both of these algorithms were unable to improve the MEP any further, the stationary point database was considered to be complete.

To complement the PES computed directly from the database of PE stationary points, these stationary points were also used to derive the free energy (FE) stationary points, and thereby the free energy surface (FES), at specific temperatures. For each temperature, rate constants between PE minima connected by a saddle point were calculated using occupational probabilities derived from harmonic densities of states (Wales, 2002). PE minima connected by rate constants above 10^9 s^{-1} were combined to form new FE minima, and harmonic densities of states were then used to re-calculate energies for these minima and associated transition states (Evans and Wales, 2003). This was done at 300 K and 350 K in order to illustrate how the nature of the FES varies with temperature. The harmonic approximation for densities of states was also used to calculate the constant volume heat capacity curve, as a summation over every PE minimum (Wales, 2017).

This procedure was implemented using the software of Wales and co-workers, which is freely available on their website (Wales Group Home Page) . This software was interfaced with the CHARMM program for PE calculations (Brooks et al., 2009), which was modified to include the Steele potential. Molecular structure images used in this publication were generated using Visual Molecular Dynamics (Humphrey et al., 1996).

6.2.3 Study Details

A total of 100 SA-BH simulations were conducted, each of 200,000 steps. A temperature of k_BT = 5.0 kcal/mol was used for the first step, and this was decremented by 2×10^{-3} % after each step. The maximum step size for each

dihedral angle was initially set to 360°, simulating complete randomization of the secondary structure. This was adjusted within the range of (0, 360°] every 50 steps, being increased by 5% if the Metropolis acceptance ratio for the preceding 50 steps was greater than 0.5, and decreased by 5% otherwise. Starting points for single-ended transition state searches were produced by applying random deviations of up to 0.01 Å in each Cartesian coordinate from the local PE minima. The DNEB method was applied with 10 images and a maximum of 300 iterations. All applications of the LBFGS algorithm used a memory of the last 4 iterations, a maximum step size of 0.4 Å, initial guesses for the diagonals of the Hessian matrix of 0.1 $Å^2$mol/kcal, and a convergence criterion of the RMS gradient not exceeding 10^{-7} kcal/molÅ. 50 single-ended transition state searches were attempted from each local minimum within 5 kcal/mol of the global minimum, and 5 searches were attempted from each local minimum above this threshold.

6.3 Results and Discussion

6.3.1 ELM Results and Validity

ELM was successful in locating a total of 1,223,837 minima and 1,116,410 saddle points. These numbers compare favorably with studies of comparable size and complexity (Ross-Naylor et al., 2017, Joseph and Wales, 2018, Neelamraju et al., 2018, Röder and Wales, 2018, Wales et al., 2019, Ross-Naylor et al., 2020), which typically report total numbers of stationary points in the order of 10^4 to 10^6. This database of stationary points will not be exhaustive owing to the complexity of the system: complete mapping of even a single adsorbed molecule has been

noted as impractical (Wright and Walsh, 2013), and the presence of a second molecule with additional degrees of freedom further exacerbates this, since the number of stationary points scales exponentially with the number of degrees of freedom (Wales, 2003). However, based on our previous study where we obtained good agreement between ELM and replica exchange molecular dynamics simulation for a single adsorbed peptide (Ross-Naylor et al., 2020), we believe the application of ELM here has identified the lowest PE minimum and probed that part of the PE surface connected to it that is likely to have the greatest impact on the dynamics of the system (Wales, 2003). In past work, a similar implementation achieved broad agreement with a replica exchange molecular dynamics simulation on the key conformations present, validating its ability to sample configurational space sufficiently (Ross-Naylor et al., 2020). The latter is in particular supported by Figure 6-1, which shows that the number of minima discovered within 5 kcal/mol of the global minimum plateaus over the course of the PES exploration, notwithstanding a sharp spike after about 1,160,000 minima that corresponds to the beginning of the DPS phase of the exploration. Of 8,674 minima below the 5 kcal/mol threshold, 8,279 were connected to the global minimum by a PE barrier of less than 20 kcal/mol, indicating considerable success in identifying paths between low-lying minima.

6.3.2 Structures and Energy Landscapes

Disconnectivity analysis on the database of stationary points resulted in the identification of three key conformations, each corresponding to a PE minimum at or near the global minimum PE and lower than any other minimum in its

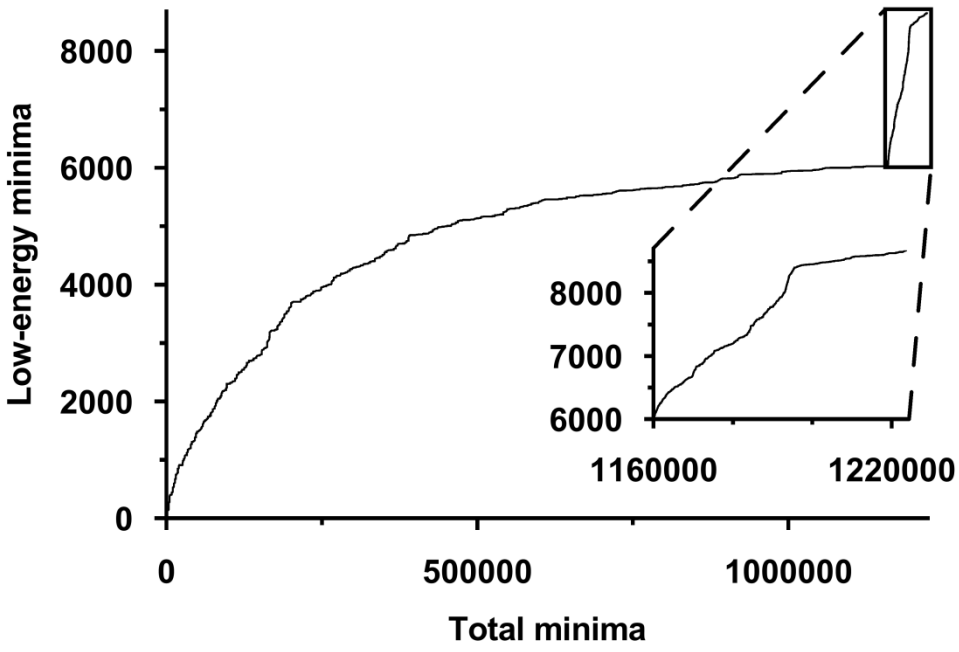

Figure 6-1. Variation of the number of minima found within 5 kcal/mol of the global minimum over the course of the exploration of the potential energy surface, which is defined in terms of the total number of minima found. Inset shows the additional minima found through the DPS analysis.

neighborhood on the PES. One of these structures is the global minimum, denoted as A, while the other two are denoted as B and C. Figure 6-2 displays these structures as viewed from the top (above the surface) and from the side. By inspection, it is clear that they are not permutational isomers and hence unrelated to the lack of symmetrization of the force field. All three structures share a common motif of the two met-enkephalin molecules lying anti-parallel on the surface, with the backbones interacting with one another and the Tyr_1, Phe_4 and Met_5 sidechains extending out in opposite directions. A and B are almost identical, with the backbones binding strongly through hydrogen bonds between the N-terminal NH_3^+ group in one molecule and the Gly_3 carboxyl group in the

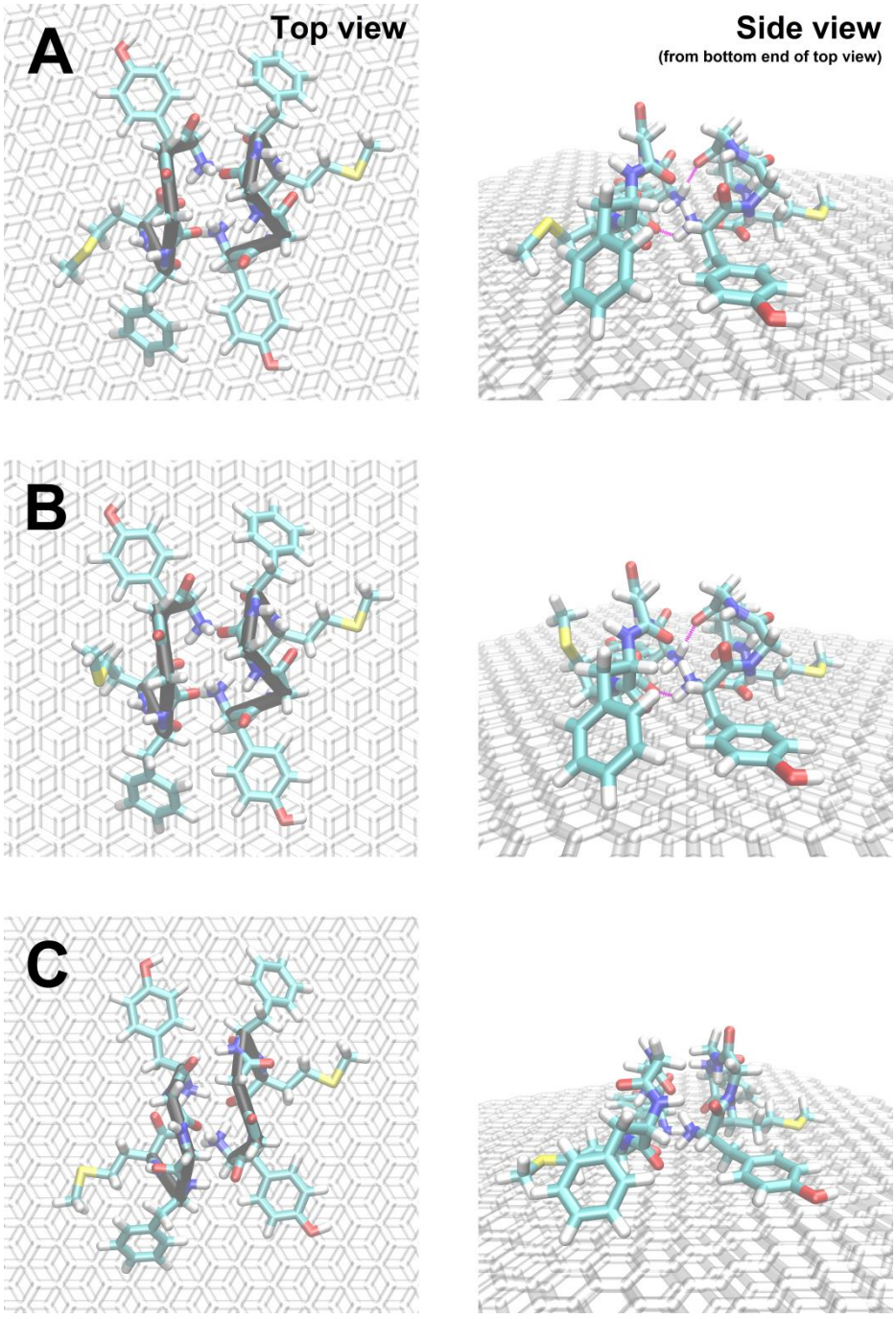

Figure 6-2. Global PE minimum structure (A) and the two other key local PE minimum structures (B and C) discovered for two zwitterionic met-enkephalin molecules adsorbed at the gas/graphite interface. The peptide backbones are represented on the leftmost images by black tubes connecting the C_α atoms of each residue. Pink dashed lines on the rightmost images denote hydrogen bonds referred to in the text.

other. The only discernible difference between A and B occurs in one of the Met_5 sidechains (left of the image), which adheres to the surface in A but peels upwards in B, interacting with the peptide backbone. C has a more distinct structure, with both molecules having arched backbones and the charged N- and C-termini all interacting with one another close to the surface. A random sampling of other stationary points discovered by ELM verified that structures lacking the anti-parallel motif observed here were found, but every such structure possessed a PE more than 20 kcal/mol higher than the global minimum. It can be concluded that all conformations reasonably contributing to the dynamics of the system possess this anti-parallel motif.

Figure 6-3 shows these three key structures, along with all the lowest connected stationary points, in a disconnectivity graph. A disconnectivity graph represents an underlying energy landscape as a connected tree, with termini corresponding to energy minima, and nodes corresponding to the energy barrier between any pair of minima (Becker and Karplus, 1997). Compared to the corresponding PES obtained in our previous study of a single met-enkephalin molecule with uncharged termini adsorbed at the gas/graphite interface (Ross-Naylor et al., 2020), the ‘palm tree’ motifs leading down to A, B and C suggests a PES with more defined structures; this is likely to be due to the presence of the charged termini in this study. The palm tree motif represents a funnel in the PES that energetically drives the conformation towards the minimum PE structure, and multiple such funnels indicate competing structures (Röder et al., 2019). In this case, the regions of the PES containing structures A and B are separated by a PE

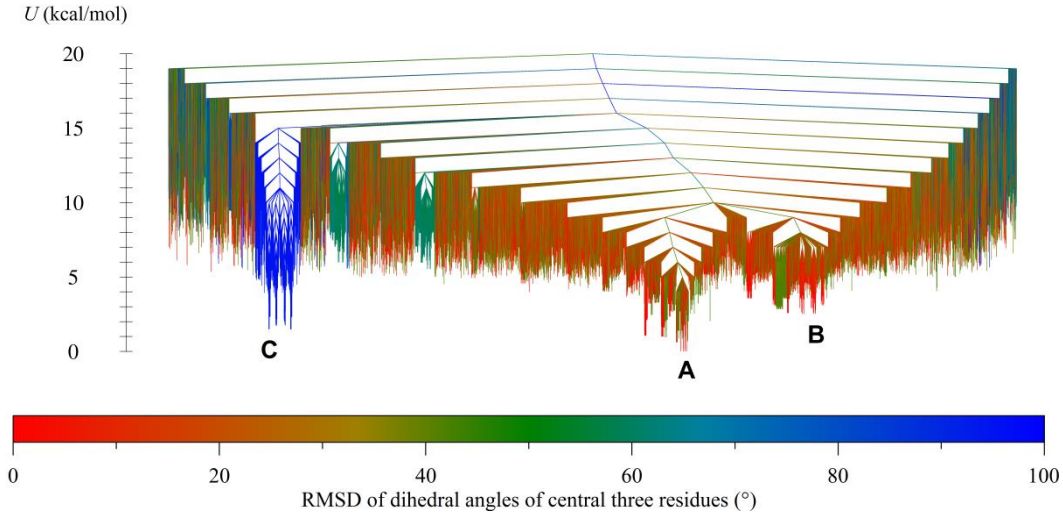

Figure 6-3. PE disconnectivity graph for two zwitterionic met-enkephalin molecules adsorbed at the gas/graphite interface. All minima connected to the global minimum by a maximum PE saddle point below U = 20 kcal/mol are shown. The labels A, B and C correspond to the structures shown in Figure 6-2. Minima are colored according to the RMSD of the dihedral angles of the central three resides from the global minimum, using the displayed scale.

barrier of approximately 10 kcal/mol, while the more distinct structure C is separated by a barrier of over 15 kcal/mol. These PE barriers are very sizeable given the qualitative similarity of the conformations. As elucidated by the transition paths presented and discussed in the next section, it is likely that these high barriers are caused by the geometric complexity of the system, which poses a significant energetic challenge to even minor transitions. Each of the funnels associated with structures A, B and C possesses four low-PE minima corresponding to the same structure; these are a consequence of the broken symmetry of the CHARMM force field. The lowest PE minimum corresponding to each structure was chosen for transition analysis.

The color scale on Figure 6-3 provides further insight into the structure of the PES. Termini and connecting nodes are colored according to the root mean standard deviation of the six central backbone dihedral angles (φ_2, ψ_2, φ_3, ψ_3, φ_4 and ψ_4) of both molecules, relative to the global minimum. This was chosen as an order parameter in light of past studies, which have shown the central three residues of met-enkephalin contribute most significantly to its conformation (Smiatek and Heuer, 2011, Banerjee and Cukier, 2014). This order parameter failed to distinguish between structures A and B, which is unsurprising given the only qualitative difference between the two structures occurs in the Met_5 sidechain. In fact, the variations in color from red to green in both funnels indicate that modest deviations in the peptide backbones prove more accessible than movement of the sidechain, which seems to play a significant role in stabilizing the conformation. The parameter is more successful in distinguishing structure C, and other individual minima and minor funnels that are separated from the global minimum by high PE barriers. Other order parameters, such as the RMSD of atom positions and the number of peptide-peptide contacts, also failed to separate A from B and in some cases could not distinguish between any of the three (results not shown). This demonstrates that in systems where peptides and biomolecules are adsorbing in contact with one another, properties and functional groups that would not normally be considered when assessing the configurational space of a single molecule may become highly relevant, such as the Met_5 sidechain in this case.

The disconnectivity graphs in Figure 6-4 display the FES at temperatures of 300 K and 350 K, and complement the PES by showing the temperature dependence of the system. At both temperatures, structure A remains the global minimum, while B and C remain the only significant competing minima. It is notable that

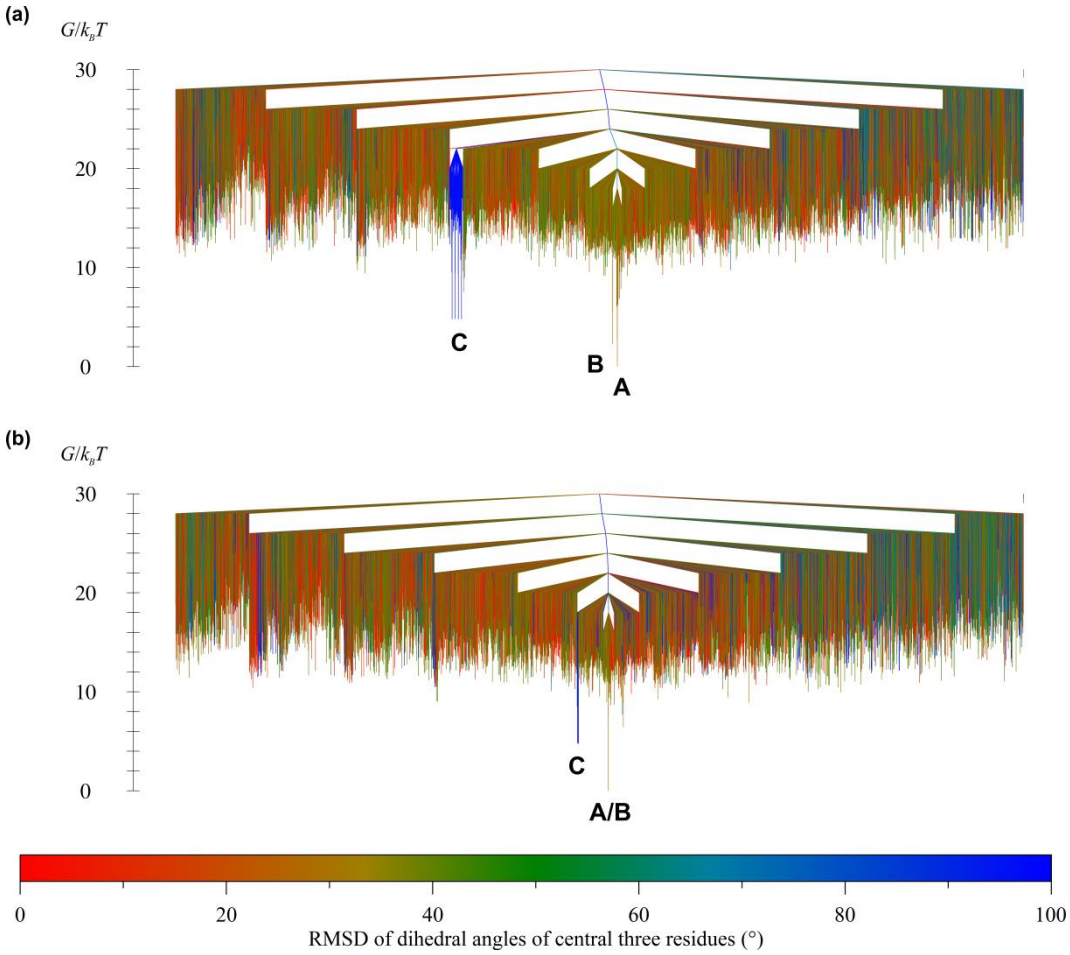

Figure 6-4. FE disconnectivity graphs for two zwitterionic met-enkephalin molecules adsorbed at the gas/graphite interface at temperatures at: (a) 300 K; and (b) 350 K. All minima connected to the global minimum by a maximum PE saddle point below $G = 30k_BT$ are shown. The labels A, B and C correspond to the structures shown in Figure 6-2. Minima are colored according to the RMSD of the dihedral angles of the central three resides from the global minimum, using the displayed scale.

although the PE of structure C is lower than that of structure B, the reverse is true of their FE at 300 K. This is intuitively due to the effect of entropy: as Figure 6-3 indicates, the region of configurational space corresponding to B contains more stationary points than C and is likely larger, permitting a greater degree of movement and perturbation within the molecules. The same effect is evident at 350 K, at which point the amalgamation of minima connected by rate constants above 10^9 s^{-1} causes structures A and B to merge. The FE of structure C relative to the global minimum holds steady at approximately $4k_BT$, while the FE barrier decreases relative to the temperature, indicating, unsurprisingly, that the transition between structures A/B and C becomes easier as the temperature is increased.

Figure 6-5 displays the heat capacity of the system, calculated as a function of temperature. This may be compared with the heat capacity obtained for a single uncharged met-enkephalin molecule at a gas/graphite interface (Ross-Naylor et al., 2020). Here, the heat capacity curve has one solitary peak, and its height, indicated by the scale of the *y*-axis, is quite small, indicating only a minor variance with temperature. This is consistent with the highly structured energy landscape of the two adsorbed peptides, as shown in Figure 6-3 and Figure 6-4, which contrasts with the rougher landscape of the single molecule (Ross-Naylor et al., 2020). A point of inflection is observable on the heat capacity curve, at approximately 200 K. This could indicate that the curve comprises two overlapping peaks: a minor peak at approximately 150 K and a major one at about 300 K. These peaks, intuitively, are likely to correspond to the low-energy transition between A and B and the high-energy transition between A/B and C,

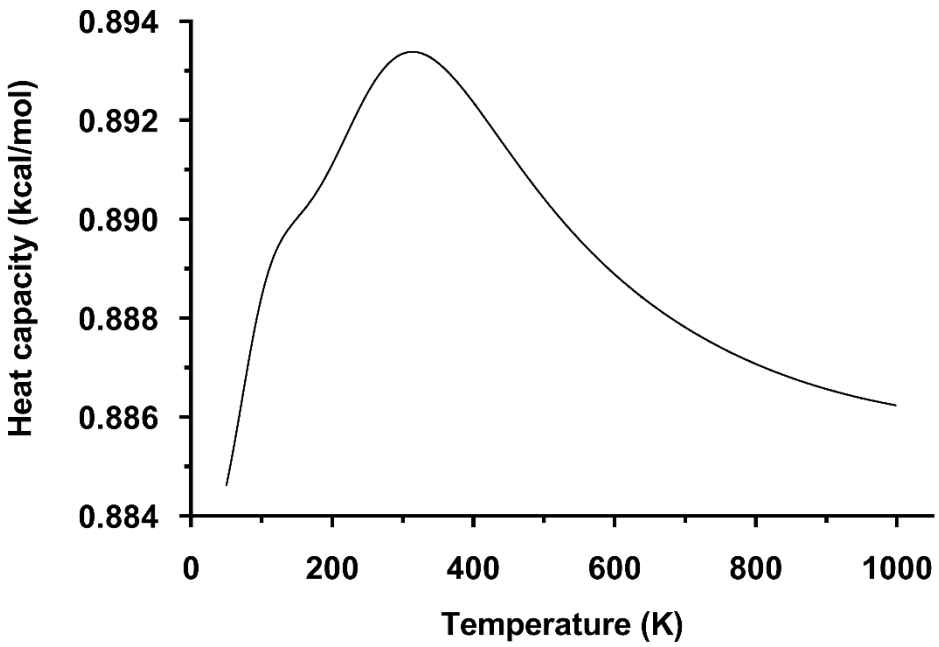

Figure 6-5. Constant volume heat capacity, as a function of temperature, calculated from the stationary point database obtained from ELM.

respectively. It is worth noting that the harmonic approximation used in the calculation of heat capacities in this work may lead to inaccuracies, particularly at higher temperatures (Wales, 2017). However, comparison with simulation methods for computing the heat capacities of adsorbed peptides showed that the harmonic approximation was sufficient to replicate the qualitative features of the heat capacity curve, despite quantitative errors (Ross-Naylor et al., 2020).

6.3.3 Transition Pathways

A more complete understanding of the key structures formed by the system is gained by considering the energy profile of the MEP between each pair of structures, and the conformations formed along this pathway. Figure 6-6

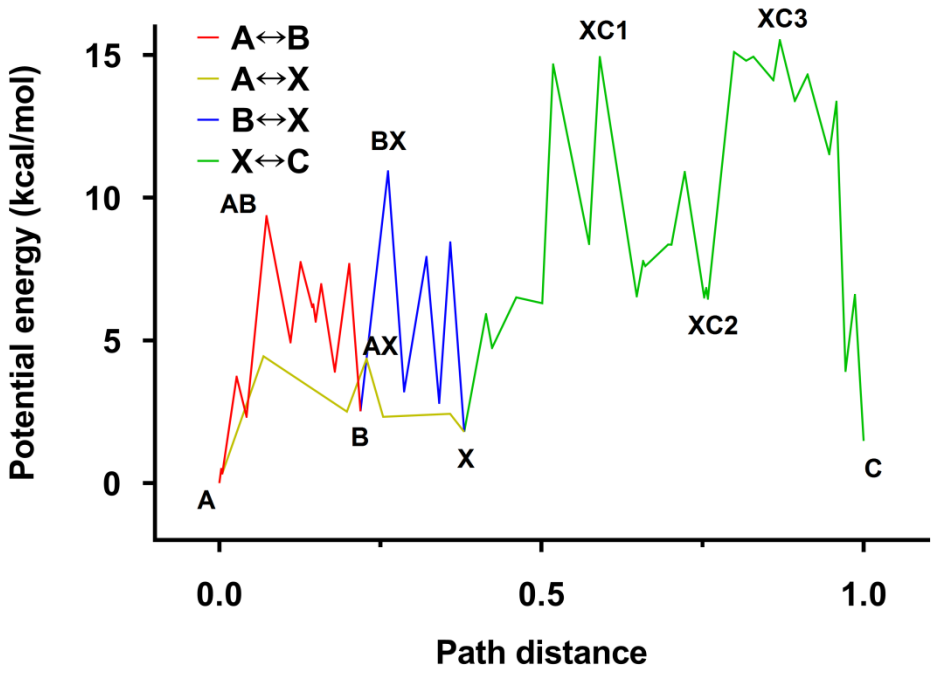

Figure 6-6. Variation of the PE along the MEPs between structures A, B and C shown in Figure 6-2, including intermediate structures labeled on this graph and displayed in Figure 6-7. The reference PE is that of the global minimum structure, A. The path distance is the minimized Euclidean distance between neighboring stationary points, scaled to the path between A and C. Lines between stationary points are provided as a guide to the eye only.

illustrates the paths between all three pairs of structures (A↔B, A↔C and B↔C). It should be noted that the A↔B path is far shorter than the other two, which is to be expected given the structures' similarity. Also, the A↔C and B↔C paths share much in common, with both A and B transitioning to a meta-stable intermediate, denoted X, from which the full transition to C proceeds.

Major structures and intermediates highlighted in Figure 6-6 are illustrated in Figure 6-7, in the form of a flow chart showing the transitions. In general, these transitions consist of a series of structural changes that impact the peptides'

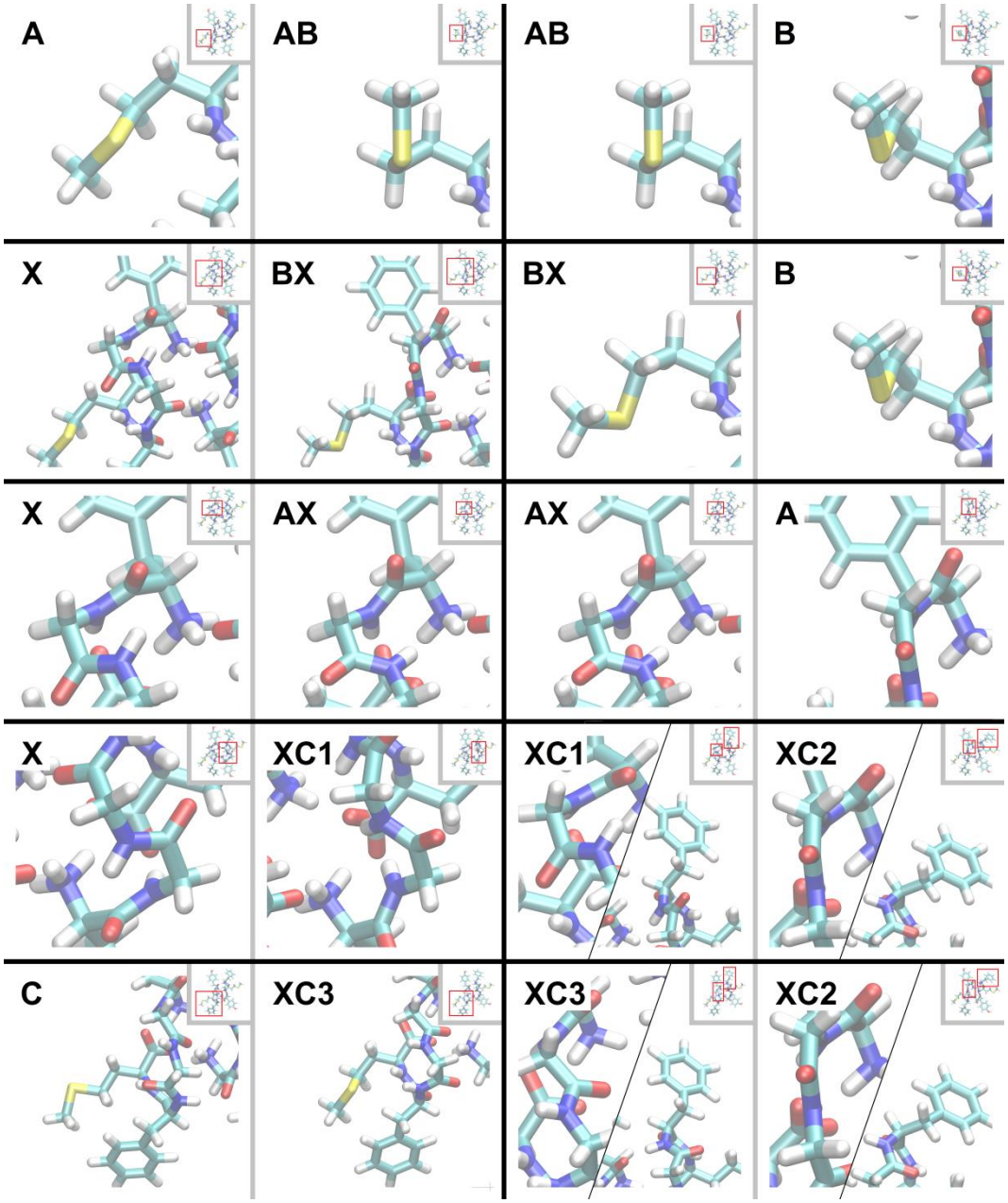

Figure 6-7. Drawings of transitions between structures A, B and C shown in Figure 6-2, and intermediates highlighted in Figure 6-6. The main images are zoomed in on the transitioning functional groups. Full structures are shown in insets.

interaction sites with one another and with the surface to a significant degree. This is presumably responsible for the high energy barriers present at a number of

points along the transition paths, and hence the multi-funnel nature of the energy landscapes as illustrated in Figure 6-3 and Figure 6-4. The primary motif in the A↔B transition, and indeed that from B to X as well, is the rotation of the left hand (as viewed) Met_5 sidechain, between lying flat on the surface and peeling away to interact with the peptide backbone. Only in the final stages of the transition from B to X is there significant movement elsewhere, with the Gly_2 residue on the left-hand molecule twisting parallel to the surface. The transition from A to X, meanwhile, involves no change to the sidechain, and a PE barrier approximately half that of the A↔B and B↔X transitions. This further supports the hypothesis of the Met_5 sidechains stabilizing the system.

The transition path between X and C takes place in two general stages, each with approximately equal PE barriers. Starting from X, firstly, both molecules' backbones straighten and lengthen via the Gly_2 residues rotating to arch over the surface, causing the Tyr_1 and Phe_4 sidechains to peel away from each other. Meanwhile, the Gly_3 residue of the right-hand molecule rotates 180 degrees, breaking the hydrogen bond with the N-terminus of the left-hand molecule. The combined effect instead brings this N-terminus into closer contact with the C-termini of both molecules, resulting in a moderately stable intermediate, XC2. Following this, both the Gly_2 and Gly_3 residues of the left-hand molecules rotate similarly, breaking the other Gly_3–N-terminus hydrogen bond and instead facilitating interactions between all four termini. At the final step, the left-hand Met_5 sidechain rotates to lie flat on the surface, forming the final C structure.

The energy profiles displayed in Figure 6-6 give a general comparison of the energetic favorability of different paths and stages, but quantifying the timescales of the transitions requires the calculation of rate constants. Figure 6-8 displays rate constants for each of the reversible transitions between A, B and C as a function of temperature, taking entropic effects into consideration. As the energy barrier would suggest, A and B interconvert relatively readily at 300 K; at this temperature, the favored transition from B to A possesses a rate constant of approximately 10^7 s^{-1}, corresponding to a mean transition time of 100 ns. Transitions to and from C are far slower, requiring times in the millisecond range at 300 K, which is likely to be out of reach of most standard MD simulations, and

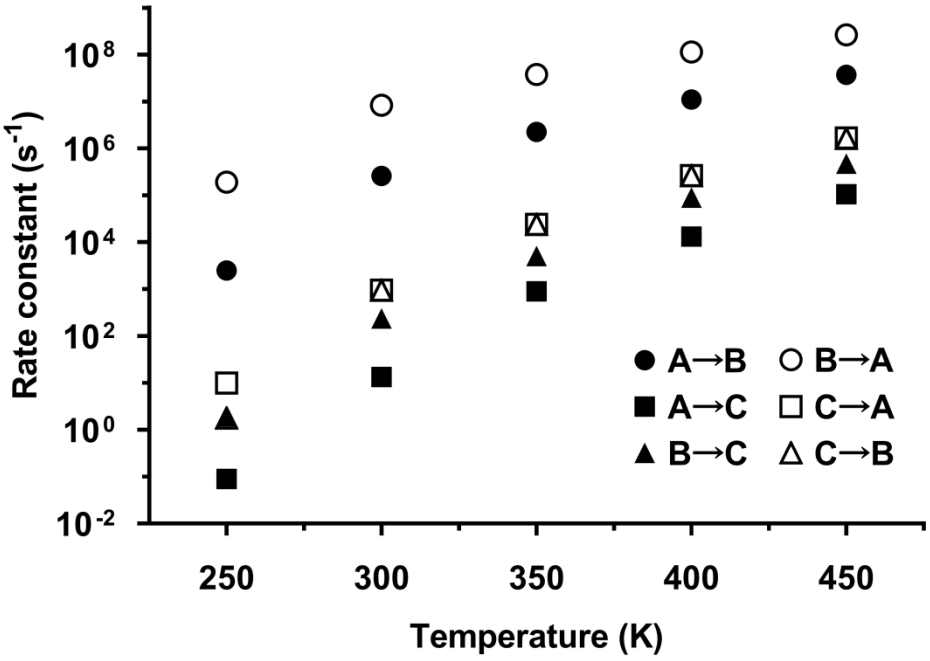

Figure 6-8. Variation of the rate constants with temperature for the transition paths between structures A, B and C as shown in Figure 6-6.

in the microseconds at 450 K. The narrowing gap between the forward and backward reactions for A↔B and A↔C with increasing temperature indicate that alternate structures to the global minimum A may be found in greater quantities at higher temperature. The B↔C transition does not follow this trend: the rate constants are approximately equal at 250 K, and the transition from C to B is favored thereafter; this likely arises due to structure B's entropic favorability overcoming its marginally higher PE with increasing temperature. In practice, it can be concluded that, due to the prevalence of A, the A↔B and A↔C transitions would dominate at all the surveyed temperatures. It should be noted that these rate constants consider only the single most favored path and ignore other less significant paths that would nonetheless contribute to the dynamics of the system.

6.4 Conclusions

The conformations and energy landscapes of two zwitterionic met-enkephalin molecules at a gas/graphite interface were investigated using ELM. It was found that the system favors structures where the two peptides lie antiparallel on the surface with the backbones in close proximity and interacting through hydrogen bonds. Major conformations are distinguished by the orientation of the Met_5 sidechains to the surface, and the functional groups to which the N-termini hydrogen-bond. Although these constitute only minor structural changes, the geometry of the system, involving two molecules in close proximity and the presence of a rigid surface, makes transitions between them relatively difficult and complex, occurring up to the millisecond range at 300 K in some cases. Both the specific conformations taken by two adsorbed met-enkephalin molecules, and the

ease of transitions between them, differed appreciably from a prior study of a single adsorbed met-enkephalin molecule at the gas/graphite interface (Ross-Naylor et al., 2020). This has significant implications for studies of peptide adsorption: it is evident that results obtained from studies of a single adsorbed molecule will be of limited applicability to adsorption in systems that are not dilute.

The present work characterized a system similar to those benchmarked in prior studies (Mijajlovic et al., 2011, Ross-Naylor et al., 2020). Future applications of ELM to multiple adsorbed peptides may consider peptides with aggregation properties, such as beta-sheet and helix formers. In general, however, further work is needed to validate and improve the applicability of ELM to biomolecule-surface interactions. Since the computational feasibility of ELM depends significantly on the number of degrees of freedom (Wales, 2003), modeling macromolecules such as proteins and simulating aqueous environments using explicit solvent presents difficulties. Efforts are ongoing to produce all-atom force fields compatible with liquid/solid interfacial systems (Walsh, 2017) and reliable coarse-grained models for proteins (Ramakrishnan et al., 2017). Finally, while the present work investigating the adsorption of two molecules on a surface represents a step towards modeling bulk adsorption, greater insights would be achieved by considering a larger number of molecules, and the development of a generalized strategy to this end would be desirable.

7 Conclusion

Energy landscape mapping was applied to a variety of peptide adsorption systems. In chapter 4, the conformational switches of polyalanine at a solid surface were studied as a function of the strength of the surface interaction: the first application of ELM to the adsorption of a biomolecule. The key outcomes achieved were comparable to the results obtained in prior work on the same system (Mijajlovic and Biggs, 2007). The work also elucidated in detail the transition between the 3_{10}- and 2_7-helix conformations of polyalanine, which occurs too slowly to be elucidated by widely used molecular dynamics methods.

Chapter 5 detailed the study of met-enkephalin at a gas/graphite interface using both ELM and replica exchange molecular dynamics. Considerable agreement was observed between the results of the two methodologies, including the major adsorbed conformations of the peptide, and their relative residence probabilities. ELM also further elucidated the energy barriers between these conformations, revealing a significant degree of frustration in the energy landscapes, and effectively describing relationships between conformations that could not be determined analytically or through the use of order parameters.

In chapter 6, adsorption of two met-enkephalin molecules with charged termini at a gas/graphite interface were studied. The major conformations were identified, along with a common motif of the peptides lying antiparallel on the surface. Transition mechanisms and rates between these conformations were elucidated, and it was noted that distinct adsorbed structures could not be distinguished by

standard order parameters, which serves as a cautionary note for future studies of multiple adsorbed peptides, which are of far wider practical interest compared to the now common single-molecule studies.

Although the peptides studied, polyalanine and met-enkephalin, were largely chosen as model peptides due to connections with past work and their ubiquity in literature, both have practical applications and relevance that is worth discussing here. Trinucleotide expansion mutation in humans can cause the generation of long polyalanine stretches in proteins, which are thought to be responsible for several diseases (Albrecht and Mundlos, 2005, Polling et al., 2015). The helix-forming tendency of polyalanine has also been exploited in the creation of biosensors (Akbulut et al., 2014) and antimicrobial peptides (Cardoso et al., 2016). Met-enkephalin, meanwhile, is a human neuropeptide with opioid function (Hughes et al., 1975); in biological applications it is often known as opioid growth factor (Rogosnitzky et al., 2013). Met-enkephalin has been noted as a promising avenue for the treatment of multiple cancers (Rogosnitzky et al., 2013, McLaughlin and Zagon, 2014, Zagon and McLaughlin, 2014, Wang et al., 2017), and delivery *via* nanoparticles could provide a way forward for such treatment in practice (Szweda et al., 2016). An understanding of the adsorption behaviour of met-enkephalin may, thus, be of assistance in determining more effective means of delivery.

As outlined in the Introduction, peptide adsorption is of widespread relevance in the fields of biology, nanotechnology and biomedicine, and numerous experimental and simulation studies of peptide/surface assemblies have been

conducted. The work here introduces a new means of studying such systems that is particularly suited to investigating the behaviour of adsorbed peptides over longer timescales that are of biological relevance, including in the non-dilute phase, and thus would provide complementary information about these assemblies. The studies conducted here operate on relatively simple peptides and utilise idealised assumptions, such as the absence of solvent and a continuous surface model. As such, the application of these methods to biomolecules of greater relevance, incorporating the effect of solvent and using more realistic surface models, constitutes a logical next step. General studies of peptide/surface interactions may also elucidate common motifs of adsorption properties in peptides, enabling the design of surface-functional peptides, which is a current topic of interest (Costa et al., 2015, Schwaminger et al., 2018). Finally, studies of peptide adsorption may be used as a model for, or extended to, protein adsorption processes, such as the formation of the protein corona on nanoparticles (Docter et al., 2015).

Further work on ELM and other associated methodologies is necessary if it is to be used to study peptide adsorption more widely. The greatest challenge at present is the inability for ELM to realistically incorporate explicit solvent models, which are the most accurate means of incorporating the effect of solvents under any circumstances, but particularly so for peptide adsorption in the most general case (Walsh, 2017). Implicit solvent approaches are computationally more feasible in the ELM context and do exist for peptide/protein adsorption. They are, however, far less well developed, particularly for non-hydrophobic surfaces. Further work

on investigating their implementation for ELM of peptide adsorption is, thus, needed. Force fields optimised for peptide/solid interactions have advanced significantly over the last decade (Martin et al., 2016), but further development and validation of these force fields for a variety of processes would be desirable. Meanwhile, a challenge encountered in the studies conducted here was the computational viability of mapping the landscape sufficiently so as to yield meaningful and reliable results. The presence of the surface introduced a high degree of complexity to the potential energy surface, and it is presumed that studies of larger molecules, or larger numbers of molecules necessary for the study of true bulk adsorption, would result in further scale-up. Methodologies for efficiently sampling areas of the potential energy surface that are most critical to the system's behaviour and properties are discussed and implemented in the present studies, but further development of such methodologies is desirable. Another possible source of improved computational efficiency lies in the algorithms for locating minima and, in particular, saddle points. As noted in the Literature Review, these algorithms are rarely directly compared or systematically optimised, and a more in-depth study of these methods may yield long-term benefits in efficiency.

Appendix A: Supporting Information for Chapter 4

A.1 Discrete Path Sampling Procedure

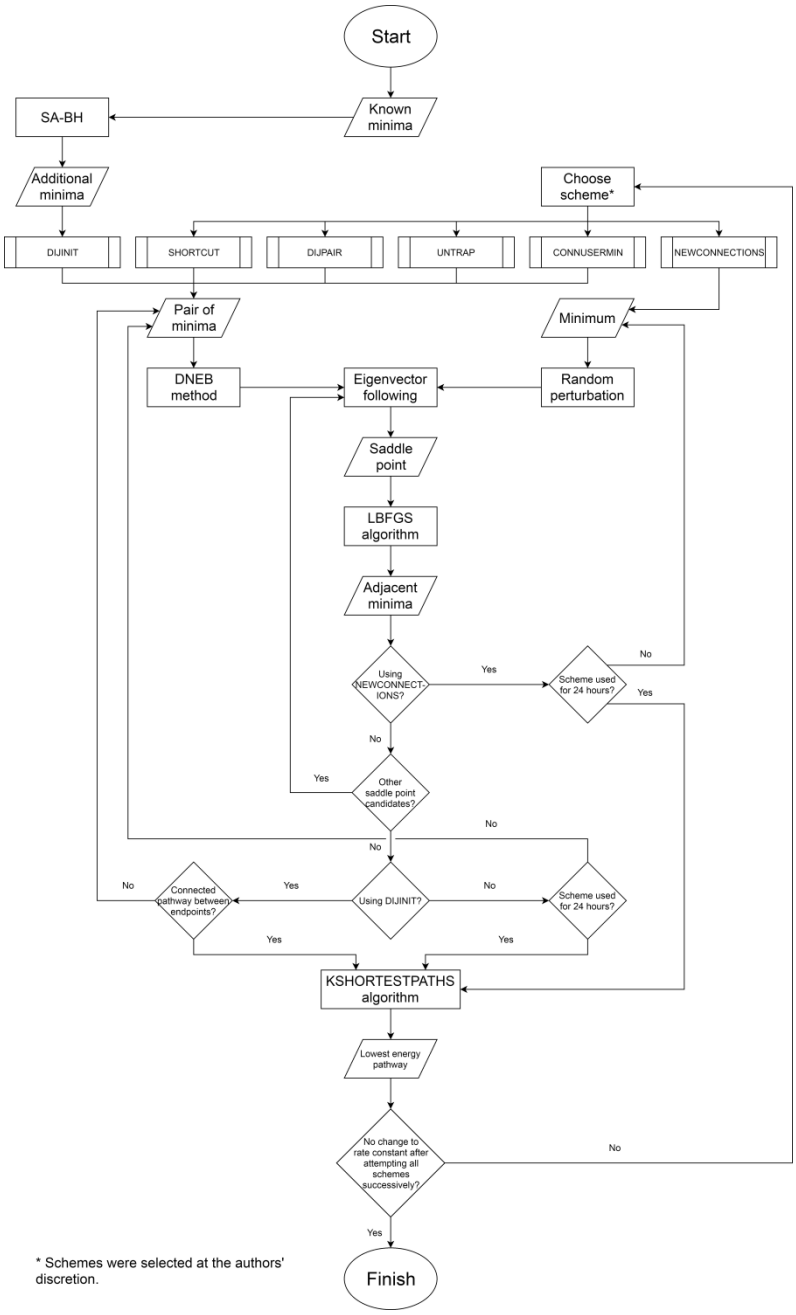

Figure A-1. Flow chart for the procedure described in Section 2.2 for expanding stationary point databases and finding the switching pathway.

A.2 Algorithms and Schemes Used

This section gives a brief summary of each of the mathematical algorithms and selection schemes used in the study as outlined in Section 4.2.2. References for each algorithm and scheme are given where available and should be consulted if greater detail is desired.

Limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm (Liu and Nocedal, 1989): An algorithm for the local minimization of a multivariable function, designed for optimal efficiency when used on functions with a very large number of variables.

Doubly nudged elastic band (DNEB) method (Trygubenko and Wales, 2004): An algorithm for estimating the minimum-energy path between a selected pair of minima and finding approximate locations for saddle points. A number of 'images' (intervening points) are placed between the endpoint minima, equidistantly in a straight line, and a harmonic spring potential is applied to each image in addition to the actual potential energy. The gradient of this combined potential, with modifications for computational stability, is then minimized using the LBFGS algorithm. This causes the chain of images to approximate the minimum energy path, provided the original linear assumption is good enough. Images with a higher potential energy than the preceding or following image are approximate saddle point candidates and used as starting points for eigenvector following.

Eigenvector following (Munro and Wales, 1999): An algorithm for precisely locating a saddle point by iteratively moving from a given starting point. On each iteration, an ‘uphill’ step is taken followed by minimization (using the LBFGS algorithm) in the space perpendicular to this uphill step.

CONNUSERMIN selection scheme: A scheme for selecting pairs of minima to be used as starting points for saddle point searches. CONNUSERMIN operates using a set of minima input by the user, and selects pairs of known minima that include at least one of these user-specified minima, in order of increasing Euclidean separation between the pair. We used this scheme to cross-check transition paths between the same structures at different surface energies. The minima on the putative pathway for one surface energy usually corresponded to local minima at the others, but these minima might not yet be known or be in an undersampled area of configurational space. Inputting these corresponding minima for use in this scheme ensured that any difference in the transition paths were genuine and not the result of undersampling.

DIJINIT selection scheme (Carr et al., 2005): A scheme for selecting pairs of minima to be used as starting points for saddle point searches. DIJINIT attempts to build a path of connected minima and saddle points between two specified minima on the surface. Based on the minima and saddle points already known, an ideal path with ‘gaps’ where further minima and saddle points might be is constructed, and the scheme selects pairs of minima on either side of these gaps. The scheme concludes when there are no more gaps; that is, a connected path between the desired minima has been constructed.

DIJPAIR selection scheme (Strodel et al., 2007): A scheme for selecting pairs of minima to be used as starting points for saddle point searches. DIJPAIR calculates optimal pathways between one of the endpoint minima and every other connected minimum, and selects pairs of minima separated by the highest saddle points. The purpose of this scheme is to free 'trapped' minima that might be on the true fastest pathway, but are currently not being considered because the only known connected saddle points are too high.

NEWCONNECTIONS selection scheme: A scheme for selecting minima to be used as starting points for saddle point searches. This is distinct from other selection schemes in that only a single minimum is chosen; a random perturbation is applied and eigenvector following is then used to find a saddle point. Because the DNEB method is skipped, saddle points may be found more quickly using NEWCONNECTIONS, but the saddle points found are less likely to be of consequence to the minimum energy pathway compared to other schemes. The minima chosen to be perturbed are left to the user's discretion. In this study, we applied this scheme iteratively to minima below the energy barrier of the putative pathway, and ensured the perturbation applied was small so as to maximize the chance of finding a saddle point near the original minimum in configurational space. This allowed us to find 'missing' saddles that had not been found by other schemes.

SHORTCUT selection scheme (Carr and Wales, 2005): A scheme for selecting pairs of minima to be used as starting points for saddle point searches. SHORTCUT operates on the current putative pathway between the endpoint

minima, and selects pairs of minima on the pathway that are separated by at least n saddle points, in order of increasing Euclidean separation between the pair. This scheme thus attempts to find 'shortcuts' on the path, cutting down the overall number of stationary points. We used $n = 1$, thus attempting every pair on the path.

UNTRAP selection scheme (Strodel et al., 2007): A scheme for selecting pairs of minima to be used as starting points for saddle point searches. UNTRAP calculates the energy barriers between one of the endpoint minima and every other connected minimum within a given threshold energy of the endpoint minimum, and selects the minima separated by the saddle point responsible for the highest energy barrier. UNTRAP serves the same purpose as the DIJPAIR scheme, but uses energy barrier height instead of pathway analysis. We chose 5.0 kcal/mol as the threshold, as we found that too great a threshold led to too many minima being included, which caused the scheme to run inefficiently.

KSHORTESTPATHS algorithm (Carr and Wales, 2008): An algorithm for computing the n paths with the greatest contribution to the steady-state rate constant. We thus used KSHORTESTPATHS to try to ensure the path with the best rate constant was found, by finding a very large number of paths ($n = 10^4$) and applying an energy threshold above which all saddle points were not considered, successively lowering it until no paths were available.

A.3 Development of Preferred Paths

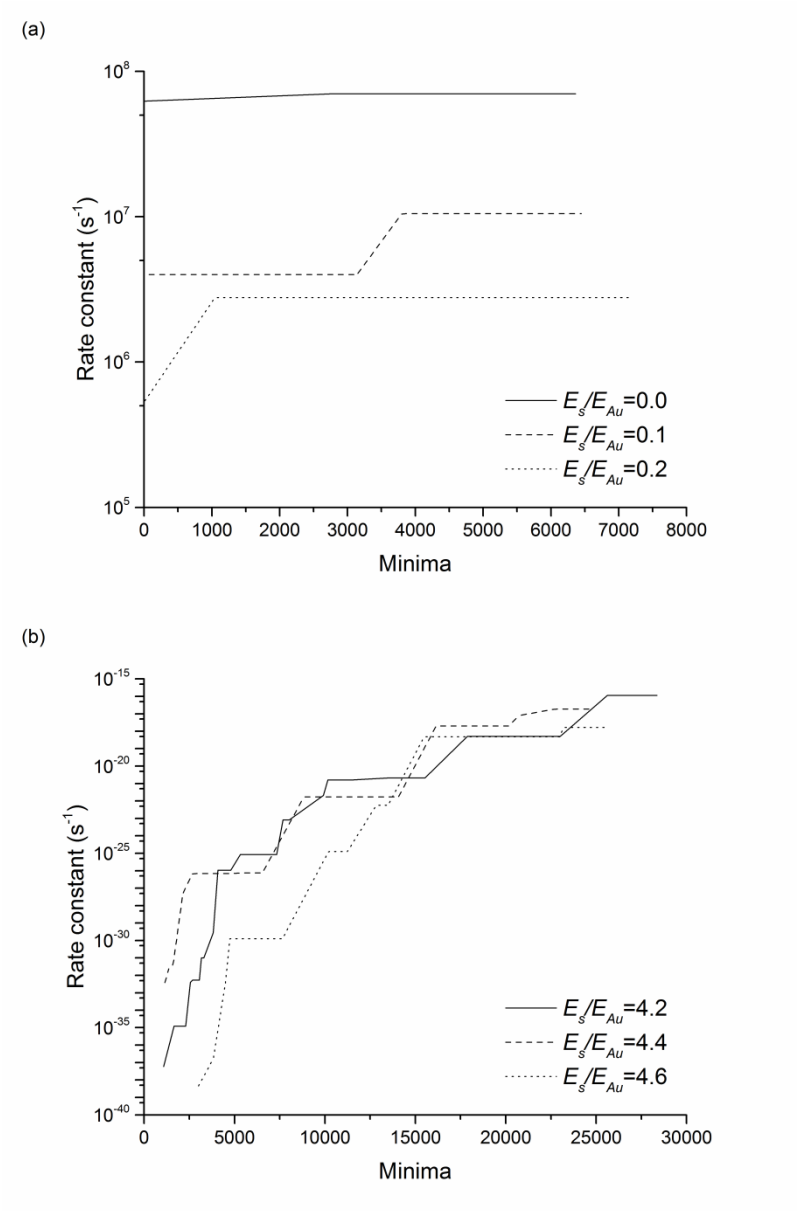

Figure A-2. Variation of the rate constant of the preferred transition pathway with the number of discovered minima for (a) the α→3_{10} transition at $E_s/E_{Au} = 0.0$, 0.1 and 0.2 and (b) the 3_{10}→2_7 transition at $E_s/E_{Au} = 4.2$, 4.4 and 4.6. The rate constants are calculated at a temperature of $k_BT = 0.3$ kcal/mol.

A.4 Additional Disconnectivity Graphs

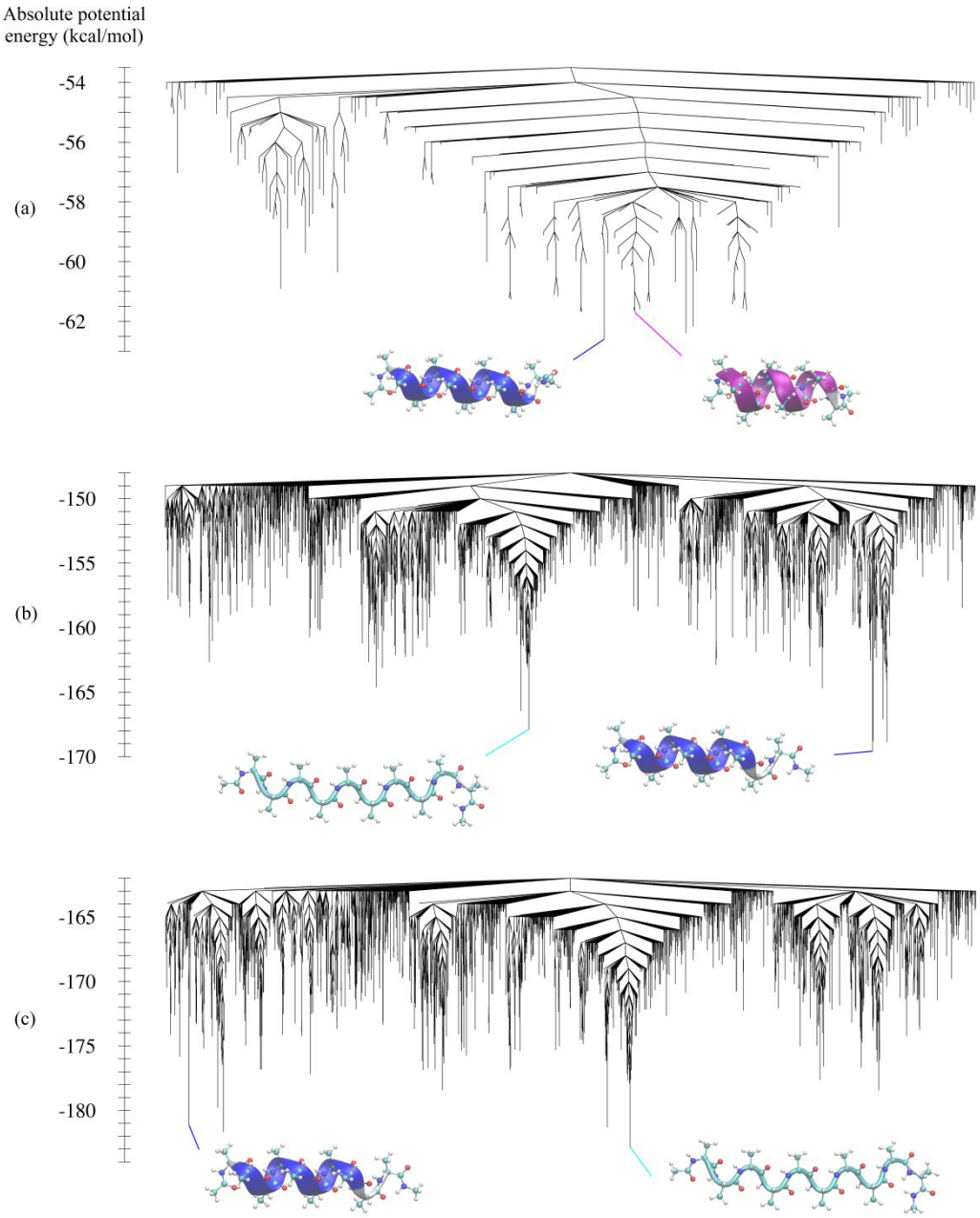

Figure A-3. Disconnectivity graphs showing the lowest minima found for 10-alanine at surface energy ratios of: (a) $E_s/E_{Au} = 0.2$, (b) $E_s/E_{Au} = 4.2$ and (c) $E_s/E_{Au} = 4.6$. The minima corresponding to the minimum-energy α- (magenta), 3_{10}- (blue) and 2_7-helices (cyan) are highlighted.

A.5 Additional Transition Pathways

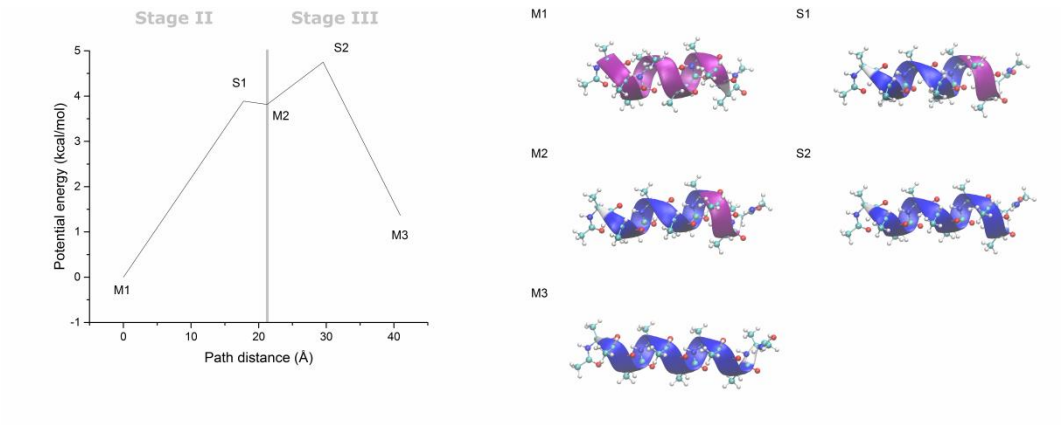

Figure A-4. Variation of the potential energy along the transition pathway between the α-helix (magenta) and 3_{10}-helix (blue) at $E_s/E_{Au} = 0.0$, and images of selected structures along the path, viewed from above the surface. The reference potential energy is that of the minimum-energy α-helix adsorbed on the surface. The path distance is the minimized Euclidean distance between neighboring stationary points. Lines between stationary points are provided as a guide to the eye only.

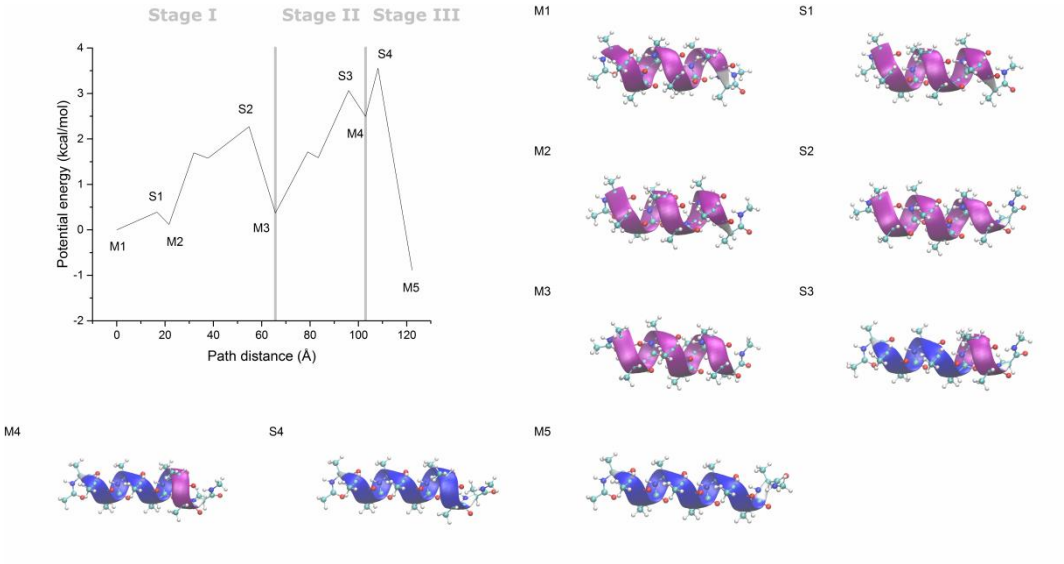

Figure A-5. Variation of the potential energy along the transition pathway between the α-helix (magenta) and 3_{10}-helix (blue) at $E_s/E_{Au} = 0.2$, and images of selected structures along the path, viewed from above the surface. The reference potential energy is that of the minimum-energy α-helix adsorbed on the surface. The path distance is the minimized Euclidean distance between neighboring stationary points. Lines between stationary points are provided as a guide to the eye only.

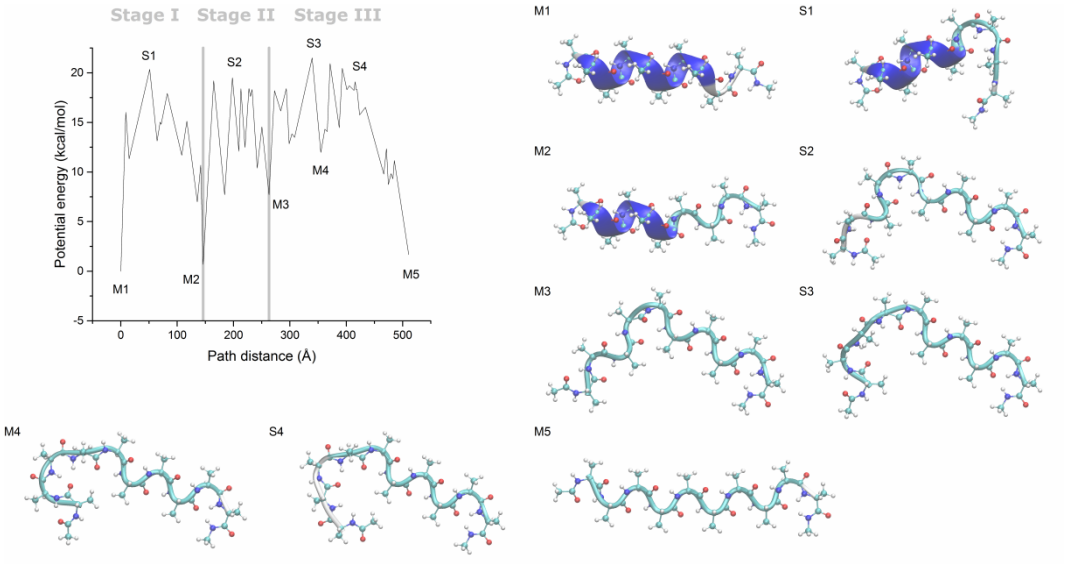

Figure A-6. Variation of the potential energy along the transition pathway between the 3_{10}-helix (blue) and 2_7-helix (cyan) at $E_s/E_{Au} = 4.2$, and images of selected structures along the path, viewed from above the surface. The reference potential energy is that of the minimum-energy 3_{10}-helix adsorbed on the surface. The path distance is the minimized Euclidean distance between neighboring stationary points. Lines between stationary points are provided as a guide to the eye only.

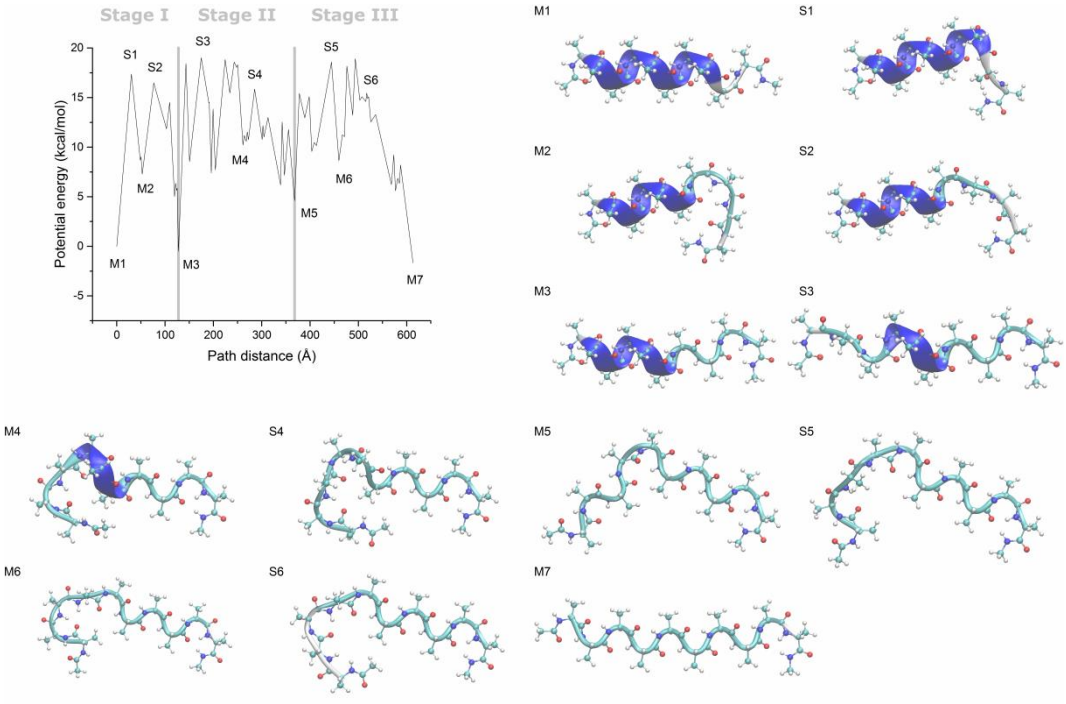

Figure A-7. Variation of the potential energy along the transition pathway between the 3_{10}-helix (blue) and 2_7-helix (cyan) at $E_s/E_{Au} = 4.6$, and images of selected structures along the path, viewed from above the surface. The reference potential energy is that of the minimum-energy 3_{10}-helix adsorbed on the surface. The path distance is the minimized Euclidean distance between neighboring stationary points. Lines between stationary points are provided as a guide to the eye only.

References

Wales Group Home Page [Online]. Available: https://www.ch.cam.ac.uk/group/wales/ [Accessed October 16, 2019].

ABASCAL, J. L. F. & VEGA, C. 2005. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.,* 123**,** 234505.

ABRAHAM, M. J. & GREADY, J. E. 2008. Ensuring Mixing Efficiency of Replica-Exchange Molecular Dynamics Simulations. *J. Chem. Theory Comput.,* 4**,** 1119-1128.

ABRAHAM, M. J., MURTOLA, T., SCHULZ, R., PÁLL, S., SMITH, J. C., HESS, B. & LINDAHL, E. 2015. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX,* 1-2**,** 19-25.

AKBULUT, H., YAVUZ, M., GULER, E., DEMIRKOL, D. O., ENDO, T., YAMADA, S., TIMUR, S. & YAGCI, Y. 2014. Electrochemical Deposition of Polypeptides: Bio-based Covering Materials for Surface Design. *Polym. Chem.,* 5**,** 3929-3936.

ALBRECHT, A. & MUNDLOS, S. 2005. The Other Trinucleotide Repeat: Polyalanine Expansion Disorders. *Curr. Opin. Genet. Dev.,* 15**,** 285-293.

ALEKSANDROV, A. & FIELD, M. 2012. A Hybrid Elastic Band String Algorithm for Studies of Enzymatic Reactions. *Phys. Chem. Chem. Phys.,* 14**,** 12544-12553.

ALHAT, D., LASRADO, V. & WANG, Y. A Review of Recent Phase Transition Simulation Methods: Saddle Point Search. ASME 2008 International

Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2008 Brooklyn, NY, U.S.A.: American Society of Mechanical Engineers Digital Collection, 103-111.

ANDERSON, E., BAI, Z., BISCHOF, C., BLACKFORD, S., DONGARRA, J., DU CROZ, J., GREENBAUM, A., HAMMARLING, S., MCKENNEY, A. & SORENSEN, D. 1999. *LAPACK Users' Guide,* Philadelphia, PA, U.S.A., Society for Industrial and Applied Mathematics.

AWASTHI, S. & NAIR, N. N. 2019. Exploring High-Dimensional Free Energy Landscapes of Chemical Reactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.,* 9**,** e1398.

BAKER, J. 1986. An Algorithm for the Location of Transition States. *J. Comput. Chem.,* 7**,** 385-395.

BAKER, J. 1987. An Algorithm for Geometry Optimization Without Analytical Gradients. *J. Comput. Chem.,* 8**,** 563-574.

BALL, K. D., BERRY, R. S., KUNZ, R. E., LI, F.-Y., PROYKOVA, A. & WALES, D. J. 1996. From Topographies to Dynamics on Multidimensional Potential Energy Surfaces of Atomic Clusters. *Science,* 271**,** 963-966.

BANERJEE, A., ADAMS, N., SIMONS, J. & SHEPARD, R. 1985. Search for Stationary Points on Surfaces. *J. Phys. Chem.,* 89**,** 52-57.

BANERJEE, R. & CUKIER, R. I. 2014. Transition Paths of Met-Enkephalin from Markov State Modeling of a Molecular Dynamics Trajectory. *J. Phys. Chem. B,* 118**,** 2883-2895.

BECKER, O. M. & KARPLUS, M. 1997. The Topology of Multidimensional Potential Energy Surfaces: Theory and Application to Peptide Structure and Kinetics. *J. Chem. Phys.,* 106**,** 1495-1517.

BEHN, A., ZIMMERMAN, P. M., BELL, A. T. & HEAD-GORDON, M. 2011a. Efficient Exploration of Reaction Paths via a Freezing String Method. *J. Chem. Phys.,* 135**,** 224108.

BEHN, A., ZIMMERMAN, P. M., BELL, A. T. & HEAD-GORDON, M. 2011b. Incorporating Linear Synchronous Transit Interpolation into the Growing String Method: Algorithm and Applications. *J. Chem. Theory Comput.,* 7**,** 4019-4025.

BELL, S. & CRIGHTON, J. S. 1984. Locating Transition States. *J. Chem. Phys.,* 80**,** 2464-2475.

BELLAICHE, M. M. J. & BEST, R. B. 2018. Molecular Determinants of Aβ42 Adsorption to Amyloid Fibril Surfaces. *J. Phys. Chem. Lett.,* 9**,** 6437-6443.

BELLUCCI, L., ARDÈVOL, A., PARRINELLO, M., LUTZ, H., LU, H., WEIDNER, T. & CORNI, S. 2016. The Interaction with Gold Suppresses Fiber-Like Conformations of the Amyloid β (16–22) Peptide. *Nanoscale,* 8**,** 8737-8748.

BELLUCCI, L. & CORNI, S. 2014. Interaction with a Gold Surface Reshapes the Free Energy Landscape of Alanine Dipeptide. *J. Phys. Chem. C,* 118**,** 11357-11364.

BERENDSEN, H. J. C., GRIGERA, J. R. & STRAATSMA, T. P. 1987. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.,* 91**,** 6269-6271.

BERENDSEN, H. J. C., POSTMA, J. P. M., VAN GUNSTEREN, W. F. & HERMANS, J. 1981. Interaction Models for Water in Relation to Protein Hydration. *In:* PULLMAN, B. (ed.) *Intermolecular Forces.* Dordrecht, The Netherlands: Springer Science+Business Media.

BERGELER, M., HERRMANN, C. & REIHER, M. 2015. Mode-Tracking Based Stationary-Point Optimization. *J. Comput. Chem.,* 36**,** 1429-1438.

BERMAN, H., HENRICK, K. & NAKAMURA, H. 2003. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.,* 10**,** 980.

BERRY, R. S. 1993. Potential Surfaces and Dynamics: What Clusters Tell Us. *Chem. Rev.,* 93**,** 2379-2394.

BERRY, R. S., DAVIS, H. L. & BECK, T. L. 1988. Finding Saddles on Multidimensional Potential Surfaces. *Chem. Phys. Lett.,* 147**,** 13-17.

BESALÚ, E. & BOFILL, J. M. 1998. On the Automatic Restricted-Step Rational-Function-Optimization Method. *Theor. Chem. Acc.,* 100**,** 265-274.

BEST, R. B. 2019. Atomistic Force Fields for Proteins. *In:* BONOMI, M. & CAMILLONI, C. (eds.) *Biomolecular Simulations: Methods and Protocols.* New York, NY, U.S.A.: Springer Science+Business Media.

BEST, R. B., MITTAL, J., FEIG, M. & MACKERELL, A. D., JR. 2012. Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity of α-helix and β-hairpin Formation. *Biophys. J.,* 103**,** 1045-1051.

BINNIG, G., QUATE, C. F. & GERBER, C. 1986. Atomic Force Microscope. *Phys. Rev. Lett.,* 56**,** 930-933.

BINNIG, G. & ROHRER, H. 1983. Scanning Tunneling Microscopy. *Surf. Sci.,* 126**,** 236-244.

BINNIG, G. & ROHRER, H. 2000. Scanning Tunneling Microscopy. *IBM J. Res. Dev.,* 44**,** 279-293.

BISWAS, P. K., VELLORE, N. A., YANCEY, J. A., KUCUKKAL, T. G., COLLIER, G., BROOKS, B. R., STUART, S. J. & LATOUR, R. A. 2012. Simulation of Multiphase Systems Utilizing Independent Force Fields to Control Intraphase and Interphase Behavior. *J. Comput. Chem.,* 33**,** 1458-1466.

BODANSZKY, M. 1988. *Peptide Chemistry,* Berlin, Germany, Springer-Verlag.

BOFILL, J. M. 1994. Updated Hessian Matrix and the Restricted Step Method for Locating Transition Structures. *J. Comput. Chem.,* 15**,** 1-11.

BOFILL, J. M., RIBAS-ARIÑO, J., VALERO, R., ALBAREDA, G., MOREIRA, I. D. P. & QUAPP, W. 2019. Interplay Between the Gentlest Ascent Dynamics Method and Conjugate Directions to Locate Transition States. *J. Chem. Theory Comput.,* 15**,** 5426-5439.

BOGE, L., HALLSTENSSON, K., RINGSTAD, L., JOHANSSON, J., ANDERSSON, T., DAVOUDI, M., LARSSON, P. T., MAHLAPUU, M., HÅKANSSON, J. & ANDERSSON, M. 2019. Cubosomes for Topical Delivery of the Antimicrobial Peptide LL-37. *Eur. J. Pharm. Biopharm.,* 134**,** 60-67.

BOHNER, M. U., MEISNER, J. & KÄSTNER, J. 2013. A Quadratically-Converging Nudged Elastic Band Optimizer. *J. Chem. Theory Comput.,* 9**,** 3498-3504.

BOWMAN, G. R., HUANG, X. & PANDE, V. S. 2010. Network Models for Molecular Kinetics and Their Initial Applications to Human Health. *Cell Res.,* 20**,** 622-630.

BRAUN, R., SARIKAYA, M. & SCHULTEN, K. 2002. Genetically Engineered Gold-Binding Polypeptides: Structure Prediction and Molecular Dynamics. *J. Biomater. Sci. Polym. Ed.,* 13**,** 747-757.

BRIGGS, B. D. & KNECHT, M. R. 2012. Nanotechnology Meets Biology: Peptide-Based Methods for the Fabrication of Functional Materials. *J. Phys. Chem. Lett.,* 3**,** 405-418.

BROOKS, B. R., BROOKS, C. L., III, MACKERELL, A. D., JR., NILSSON, L., PETRELLA, R. J., ROUX, B., WON, Y., ARCHONTIS, G., BARTELS, C., BORESCH, S., CAFLISCH, A., CAVES, L., CUI, Q., DINNER, A. R., FEIG, M., FISCHER, S., GAO, J., HODOSCEK, M., IM, W., KUCZERA, K., LAZARIDIS, T., MA, J., OVCHINNIKOV, V., PACI, E., PASTOR, R. W., POST, C. B., PU, J. Z., SCHAEFER, M., TIDOR, B., VENABLE, R. M., WOODCOCK, H. L., WU, X., YANG, W., YORK, D. M. & KARPLUS, M. 2009. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.,* 30**,** 1545-1614.

BROOKS, B. R., BRUCCOLERI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S. & KARPLUS, M. 1983. CHARMM: A Program

for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.,* 4**,** 187-217.

BURGER, S. K. & YANG, W. 2006. Quadratic String Method for Determining the Minimum-Energy Path Based on Multiobjective Optimization. *J. Chem. Phys.,* 124**,** 054109.

CAFLISCH, A. 2006. Network and Graph Analyses of Folding Free Energy Surfaces. *Curr. Opin. Struct. Biol.,* 16**,** 71-78.

CALVO, F., DOYE, J. P. K. & WALES, D. J. 2012. Energy Landscapes of Colloidal Clusters: Thermodynamics and Rearrangement Mechanisms. *Nanoscale,* 4**,** 1085-1100.

CARDOSO, M. H., RIBEIRO, S. M., NOLASCO, D. O., DE LA FUENTE-NÚÑEZ, C., FELÍCIO, M. R., GONÇALVES, S., MATOS, C. O., LIAO, L. M., SANTOS, N. C. & HANCOCK, R. E. W. 2016. A Polyalanine Peptide Derived from Polar Fish with Anti-Infectious Activities. *Sci. Rep.,* 6**,** 21385.

CARE, A., BERGQUIST, P. L. & SUNNA, A. 2015. Solid-Binding Peptides: Smart Tools for Nanobiotechnology. *Trends Biotechnol.,* 33**,** 259-268.

CARNAL, F., CLAVIER, A. & STOLL, S. 2016. Polypeptide-Nanoparticle Interactions and Corona Formation Investigated by Monte Carlo Simulations. *Polymers,* 8**,** 203.

CARR, J. M., TRYGUBENKO, S. A. & WALES, D. J. 2005. Finding Pathways Between Distant Local Minima. *J. Chem. Phys.,* 122**,** 234903.

CARR, J. M. & WALES, D. J. 2005. Global Optimization and Folding Pathways of Selected α-Helical Proteins. *J. Chem. Phys.*, 123**,** 234901.

CARR, J. M. & WALES, D. J. 2008. The Energy Landscape as a Computational Tool. *In:* CONNERADE, J.-P. & SOLOV'YOV, A. (eds.) *Latest Advances in Atomic Cluster Collisions: Structure and Dynamics from the Nuclear to the Biological Scale.* London, U.K.: Imperial College Press.

CARR, J. M., WHITTLESTON, C. S., WADE, D. C. & WALES, D. J. 2015. Energy Landscapes of a Hairpin Peptide Including NMR Chemical Shift Restraints. *Phys. Chem. Chem. Phys.*, 17**,** 20250-20258.

CERJAN, C. J. & MILLER, W. H. 1981. On Finding Transition States. *J. Chem. Phys.*, 75**,** 2800-2806.

CHAFFEY-MILLAR, H., NIKODEM, A., MATVEEV, A. V., KRÜGER, S. & RÖSCH, N. 2012. Improving upon String Methods for Transition State Discovery. *J. Chem. Theory Comput.*, 8**,** 777-786.

CHEN, J. & NUGEN, S. R. 2019. Detection of Protease and Engineered Phage-Infected Bacteria Using Peptide-Graphene Oxide Nanosensors. *Anal. Bioanal. Chem.*, 411**,** 2487-2492.

CHENG, C., MCGONIGAL, P. R., STODDART, J. F. & ASTUMIAN, R. D. 2015. Design and Synthesis of Nonequilibrium Systems. *ACS Nano*, 9**,** 8672-8688.

CHO, N.-H., CHEONG, T.-C., MIN, J. H., WU, J. H., LEE, S. J., KIM, D., YANG, J.-S., KIM, S., KIM, Y. K. & SEONG, S.-Y. 2011. A

Multifunctional Core-Shell Nanoparticle for Dendritic Cell-Based Cancer Immunotherapy. *Nat. Nanotechnol.,* 6**,** 675-682.

CHU, J.-W., TROUT, B. L. & BROOKS, B. R. 2003. A Super-Linear Minimization Scheme for the Nudged Elastic Band Method. *J. Chem. Phys.,* 119**,** 12708-12717.

COHAVI, O., CORNI, S., DE RIENZO, F., DI FELICE, R., GOTTSCHALK, K. E., HOEFLING, M., KOKH, D., MOLINARI, E., SCHREIBER, G., VASKEVICH, A. & WADE, R. C. 2010. Protein–Surface Interactions: Challenging Experiments and Computations. *J. Mol. Recognit.,* 23**,** 259-262.

COPPAGE, R., SLOCIK, J. M., BRIGGS, B. D., FRENKEL, A. I., HEINZ, H., NAIK, R. R. & KNECHT, M. R. 2011. Crystallographic Recognition Controls Peptide Binding for Bio-Based Nanomaterials. *J. Am. Chem. Soc.,* 133**,** 12346-12349.

CORNELL, W. D., CIEPLAK, P., BAYLY, C. I., GOULD, I. R., MERZ, K. M., FERGUSON, D. M., SPELLMEYER, D. C., FOX, T., CALDWELL, J. W. & KOLLMAN, P. A. 1995. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.,* 117**,** 5179-5197.

COSTA, D., PRADIER, C.-M., TIELENS, F. & SAVIO, L. 2015. Adsorption and Self-Assembly of Bio-Organic Molecules at Model Surfaces: A Route Towards Increased Complexity. *Surf. Sci. Rep.,* 70**,** 449-553.

COSTA, D., SAVIO, L. & PRADIER, C.-M. 2016. Adsorption of Amino Acids and Peptides on Metal and Oxide Surfaces in Water Environment: A Synthetic and Prospective Review. *J. Phys. Chem. B,* 120**,** 7039-7052.

CRAIK, D. J., FAIRLIE, D. P., LIRAS, S. & PRICE, D. 2013. The Future of Peptide-based Drugs. *Chem. Biol. Drug Des.,* 81**,** 136-147.

CREHUET, R., THOMAS, A. & FIELD, M. J. 2005. An Implementation of the Nudged Elastic Band Algorithm and Application to the Reaction Mechanism of HGXPRTase from Plasmodium falciparum. *J. Mol. Graph. Model.,* 24**,** 102-110.

CRIPPEN, G. M. & SCHERAGA, H. A. 1971. Minimization of Polypeptide Energy: XI. The Method of Gentlest Ascent. *Arch. Biochem. Biophys.,* 144**,** 462-466.

CULOT, P., DIVE, G., NGUYEN, V. H. & GHUYSEN, J.-M. 1992. A Quasi-Newton Algorithm for First-Order Saddle-Point Location. *Theor. Chim. Acta,* 82**,** 189-205.

CZERMINSKI, R. & ELBER, R. 1989. Reaction Path Study of Conformational Transitions and Helix Formation in a Tetrapeptide. *Proc. Natl. Acad. Sci. U.S.A.,* 86**,** 6963-6967.

CZERMINSKI, R. & ELBER, R. 1990. Reaction Path Study of Conformational Transitions in Flexible Systems: Applications to Peptides. *J. Chem. Phys.,* 92**,** 5580-5601.

DAUBER-OSGUTHORPE, P., ROBERTS, V. A., OSGUTHORPE, D. J., WOLFF, J., GENEST, M. & HAGLER, A. T. 1988. Structure and

Energetics of Ligand Binding to Proteins: Escherichia Coli Dihydrofolate Reductase-Trimethoprim, a Drug-Receptor System. *Proteins: Struct., Funct., Bioinf.,* 4**,** 31-47.

DE JONG, D. H., SINGH, G., BENNETT, W. F. D., ARNAREZ, C., WASSENAAR, T. A., SCHÄFER, L. V., PERIOLE, X., TIELEMAN, D. P. & MARRINK, S. J. 2012. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.,* 9**,** 687-697.

DELLAGO, C., BOLHUIS, P. & GEISSLER, P. L. 2002. Transition Path Sampling. *Adv. Chem. Phys.,* 123**,** 1-78.

DEMIR, B., YILMAZ, T., GULER, E., GUMUS, Z. P., AKBULUT, H., ALDEMIR, E., COSKUNOL, H., COLAK, D. G., CIANGA, I. & YAMADA, S. 2016. Polypeptide with Electroactive Endgroups as Sensing Platform for the Abused Drug 'Methamphetamine' by Bioelectrochemical Method. *Talanta,* 161**,** 789-796.

DENNIS, J. E., JR. & MORÉ, J. J. 1977. Quasi-Newton Methods, Motivation and Theory. *SIAM Rev.,* 19**,** 46-89.

DEWAR, M. J. S., HEALY, E. F. & STEWART, J. J. P. 1984. Location of Transition States in Reaction Mechanisms. *J. Chem. Soc. Faraday Trans. II,* 80**,** 227-233.

DHILLON, I. S. 1997. *A New O(n2) Algorithm for the Symmetric Tridiagonal Eigenvalue/Eigenvector Problem.* Doctor of Philosophy, University of California.

DIJKSTRA, E. W. 1959. A Note on Two Problems in Connexion with Graphs. *Numer. Math.,* 1**,** 269-271.

DILL, K. A. & MACCALLUM, J. L. 2012. The Protein-Folding Problem, 50 Years On. *Science,* 338**,** 1042-1046.

DOBSON, C. M. 2003. Protein Folding and Misfolding. *Nature,* 426**,** 884-890.

DOCTER, D., WESTMEIER, D., MARKIEWICZ, M., STOLTE, S., KNAUER, S. K. & STAUBER, R. H. 2015. The Nanoparticle Biomolecule Corona: Lessons Learned–Challenge Accepted? *Chem. Soc. Rev.,* 44**,** 6094-6121.

DOYE, J. P. K. & WALES, D. J. 1999. Structural Transitions and Global Minima of Sodium Chloride Clusters. *Phys. Rev. B: Condens. Matter Mater. Phys.,* 59**,** 2292-2300.

E, W., REN, W. & VANDEN-EIJNDEN, E. 2002. String Method for the Study of Rare Events. *Phys. Rev. B: Condens. Matter Mater. Phys.,* 66**,** 052301.

E, W., REN, W. & VANDEN-EIJNDEN, E. 2007. Simplified and Improved String Method for Computing the Minimum Energy Paths in Barrier-Crossing Events. *J. Chem. Phys.,* 126**,** 164103.

EATON, P. & WEST, P. 2010. *Atomic Force Microscopy,* New York, NY, U.S.A., Oxford University Press.

ELBER, R. & KARPLUS, M. 1987. A Method for Determining Reaction Paths in Large Molecules: Application to Myoglobin. *Chem. Phys. Lett.,* 139**,** 375-380.

EMAMI, F. S., PUDDU, V., BERRY, R. J., VARSHNEY, V., PATWARDHAN, S. V., PERRY, C. 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.

EVANS, D. A. & WALES, D. J. 2003. Free Energy Landscapes of Model Peptides and Proteins. *J. Chem. Phys.,* 118**,** 3891-3897.

EVANS, D. A. & WALES, D. J. 2004. Folding of the GB1 Hairpin Peptide from Discrete Path Sampling. *J. Chem. Phys.,* 121**,** 1080-1090.

FARRELL, J. D., LINES, C., SHEPHERD, J. J., CHAKRABARTI, D., MILLER, M. A. & WALES, D. J. 2013. Energy Landscapes, Structural Topologies and Rearrangement Mechanisms in Clusters of Dipolar Particles. *Soft Matter,* 9**,** 5407-5416.

FISCHER, S. & KARPLUS, M. 1992. Conjugate Peak Refinement: An Algorithm for Finding Reaction Paths and Accurate Transition States in Systems with Many Degrees of Freedom. *Chem. Phys. Lett.,* 194**,** 252-261.

FOSGERAU, K. & HOFFMANN, T. 2015. Peptide Therapeutics: Current Status and Future Directions. *Drug Discov. Today,* 20**,** 122-128.

FRENKEL, D. & SMIT, B. 2002. *Understanding Molecular Simulation: From Algorithms to Applications,* San Diego, CA, U.S.A., Academic Press.

FUKUNISHI, H., WATANABE, O. & TAKADA, S. 2002. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.,* 116**,** 9058-9067.

GALVÁN, I. F. & FIELD, M. J. 2008. Improving the Efficiency of the NEB Reaction Path Finding Algorithm. *J. Comput. Chem.,* 29**,** 139-143.

GFELLER, D., DE LOS RIOS, P., CAFLISCH, A. & RAO, F. 2007. Complex Network Analysis of Free-Energy Landscapes. *Proc. Natl. Acad. Sci. U.S.A.,* 104**,** 1817-1822.

GLADYTZ, A., JOHN, T., GLADYTZ, T., HASSERT, R., PAGEL, M., RISSELADA, H. J., NAUMOV, S., BECK-SICKINGER, A. G. & ABEL, B. 2016. Peptides@mica: From Affinity to Adhesion Mechanism. *Phys. Chem. Chem. Phys.,* 18**,** 23516-23527.

GLÄTTLI, A., DAURA, X. & VAN GUNSTEREN, W. F. 2002. Derivation of an Improved Simple Point Charge Model for Liquid Water: SPC/A and SPC/L. *J. Chem. Phys.,* 116**,** 9811-9828.

GOODROW, A., BELL, A. T. & HEAD-GORDON, M. 2008. Development and Application of a Hybrid Method Involving Interpolation and Ab Initio Calculations for the Determination of Transition States. *J. Chem. Phys.,* 129**,** 174109.

GOODROW, A., BELL, A. T. & HEAD-GORDON, M. 2010. A Strategy for Obtaining a More Accurate Transition State Estimate Using the Growing String Method. *Chem. Phys. Lett.,* 484**,** 392-398.

GOULD, N., ORTNER, C. & PACKWOOD, D. 2016. A Dimer-Type Saddle Search Algorithm with Preconditioning and Linesearch. *Math. Comp.,* 85**,** 2939-2966.

GRAY, J. J. 2004. The Interaction of Proteins with Solid Surfaces. *Curr. Opin. Struct. Biol.,* 14**,** 110-115.

GROSSFIELD, A. & ZUCKERMAN, D. M. 2009. Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations. *In:* WHEELER, R. A. (ed.) *Annual Reports in Computational Chemistry.* Amsterdam, The Netherlands: Elsevier.

GRUEBELE, M. 2002. Protein Folding: The Free Energy Surface. *Curr. Opin. Struct. Biol.,* 12**,** 161-168.

HALGREN, T. A. & LIPSCOMB, W. N. 1977. The Synchronous-Transit Method for Determining Reaction Pathways and Locating Molecular Transition States. *Chem. Phys. Lett.,* 49**,** 225-232.

HARDER, E., DAMM, W., MAPLE, J., WU, C., REBOUL, M., XIANG, J. Y., WANG, L., LUPYAN, D., DAHLGREN, M. K. & KNIGHT, J. L. 2015. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.,* 12**,** 281-296.

HAYAMIZU, Y., SO, C. R., DAG, S., PAGE, T. S., STARKEBAUM, D. & SARIKAYA, M. 2016. Bioelectronic Interfaces by Spontaneously Organized Peptides on 2D Atomic Single Layer Materials. *Sci. Rep.,* 6**,** 33778.

HE, J., CHEN, J., HU, G., WANG, L., ZHENG, J., ZHAN, J., ZHU, Y., ZHONG, C., SHI, X. & LIU, S. 2018. Immobilization of an Antimicrobial Peptide on Silicon Surface with Stable Activity by Click Chemistry. *J. Mater. Chem. B,* 6**,** 68-74.

HEINZ, H., LIN, T.-J., KISHORE MISHRA, R. & EMAMI, F. S. 2013. Thermodynamically Consistent Force Fields for the Assembly of

Inorganic, Organic, and Biological Nanostructures: The INTERFACE Force Field. *Langmuir,* 29**,** 1754-1765.

HEINZ, H. & RAMEZANI-DAKHEL, H. 2016. Simulations of Inorganic–Bioorganic Interfaces to Discover New Materials: Insights, Comparisons to Experiment, Challenges, and Opportunities. *Chem. Soc. Rev.,* 45**,** 412-448.

HELGAKER, T. 1991. Transition-State Optimizations by Trust-Region Image Minimization. *Chem. Phys. Lett.,* 182**,** 503-510.

HENKELMAN, G., JÓHANNESSON, G. & JÓNSSON, H. 2002. Methods for Finding Saddle Points and Minimum Energy Paths. *In:* SCHWARTZ, S. D. (ed.) *Theoretical Methods in Condensed Phase Chemistry.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

HENKELMAN, G. & JÓNSSON, H. 1999. A Dimer Method for Finding Saddle Points on High Dimensional Potential Surfaces Using Only First Derivatives. *J. Chem. Phys.,* 111**,** 7010-7022.

HENKELMAN, G. & JÓNSSON, H. 2000. Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *J. Chem. Phys.,* 113**,** 9978-9985.

HENKELMAN, G., UBERUAGA, B. P. & JÓNSSON, H. 2000. A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.,* 113**,** 9901-9904.

HERMANS, J., BERENDSEN, H. J. C., VAN GUNSTEREN, W. F. & POSTMA, J. P. M. 1984. A Consistent Empirical Potential for Water–Protein Interactions. *Biopolymers,* 23**,** 1513-1518.

HESTENES, M. R. & STIEFEL, E. 1952. Methods of Conjugate Gradients for Solving Linear Systems. *J. Res. Natl. Bur. Stand.,* 49**,** 409-436.

HEYDEN, A., BELL, A. T. & KEIL, F. J. 2005. Efficient Methods for Finding Transition States in Chemical Reactions: Comparison of Improved Dimer Method and Partitioned Rational Function Optimization Method. *J. Chem. Phys.,* 123**,** 224101.

HILLS, R. D., JR. & BROOKS, C. L., III 2009. Insights from Coarse-Grained Gō Models for Protein Folding and Dynamics. *Int. J. Mol. Sci.,* 10**,** 889-905.

HOARE, M. R. & PAL, P. 1971. Physical Cluster Mechanics: Statics and Energy Surfaces for Monatomic Systems. *Adv. Phys.,* 20**,** 161-196.

HOFFMANN, F. M. 1983. Infrared Reflection-Absorption Spectroscopy of Adsorbed Molecules. *Surf. Sci. Rep.,* 3**,** 107-192.

HOHENBERG, P. & KOHN, W. 1964. Inhomogeneous Electron Gas. *Phys. Rev.,* 136**,** B864-B871.

HORN, H. W., SWOPE, W. C., PITERA, J. W., MADURA, J. D., DICK, T. J., HURA, G. L. & HEAD-GORDON, T. 2004. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.,* 120**,** 9665-9678.

HOYOS-NOGUÉS, M., GIL, F. J. & MAS-MORUNO, C. 2018. Antimicrobial Peptides: Powerful Biorecognition Elements to Detect Bacteria in Biosensing Technologies. *Molecules,* 23**,** 1683.

HU, Q., WANG, P. & LASKIN, J. 2010. Effect of the Surface on the Secondary Structure of Soft Landed Peptide Ions. *Phys. Chem. Chem. Phys.,* 12**,** 12802-12810.

HUANG, J., RAUSCHER, S., NAWROCKI, G., RAN, T., FEIG, M., DE GROOT, B. L., GRUBMULLER, H. & MACKERELL, A. D., JR. 2017. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods,* 14**,** 71-73.

HUGHES, J., SMITH, T. W., KOSTERLITZ, H. W., FOTHERGILL, L. A., MORGAN, B. A. & MORRIS, H. R. 1975. Identification of Two Related Pentapeptides from the Brain with Potent Opiate Agonist Activity. *Nature,* 258**,** 577-580.

HUGHES, Z. E., KOCHANDRA, R. & WALSH, T. R. 2017. Facet-Specific Adsorption of Tripeptides at Aqueous Au Interfaces: Open Questions in Reconciling Experiment and Simulation. *Langmuir,* 33**,** 3742-3754.

HUGHES, Z. E. & WALSH, T. R. 2015. What Makes a Good Graphene-Binding Peptide? Adsorption of Amino Acids and Peptides at Aqueous Graphene Interfaces. *J. Mater. Chem. B,* 3**,** 3211-3221.

HUGHES, Z. E. & WALSH, T. R. 2018. Distinct Differences in Peptide Adsorption on Palladium and Gold: Introducing a Polarizable Model for Pd (111). *J. Phys. Chem. C,* 122**,** 19625-19638.

HUGHES, Z. E., WRIGHT, L. B. & WALSH, T. R. 2013. Biomolecular Adsorption at Aqueous Silver Interfaces: First-Principles Calculations, Polarizable Force-Field Simulations, and Comparisons with Gold. *Langmuir*, 29**,** 13217-13229.

HUMPHREY, W., DALKE, A. & SCHULTEN, K. 1996. VMD: Visual Molecular Dynamics. *J. Mol. Graphics,* 14**,** 33-38.

HUOTARI, A., XU, W., MÖNKÄRE, J., KOVALAINEN, M., HERZIG, K.-H., LEHTO, V.-P. & JÄRVINEN, K. 2013. Effect of Surface Chemistry of Porous Silicon Microparticles on Glucagon-Like Peptide-1 (GLP-1) Loading, Release and Biological Activity. *Int. J. Pharm.,* 454**,** 67-73.

HUQ, N. L., CROSS, K. J., MYROFORIDIS, H., STANTON, D. P., CHEN, Y.-Y., WARD, B. R. & REYNOLDS, E. C. 2018. Molecular Interactions of Peptide Encapsulated Calcium Phosphate Delivery Vehicle at Enamel Surfaces. *In:* ENDO, K., KOGURE, T. & NAGASAWA, H. (eds.) *Biomineralization: From Molecular and Nano-structural Analyses to Environmental Science.* Singapore: Springer Nature Singapore.

INOUE, Y., ONODERA, Y. & ISHIHARA, K. 2018. Initial Cell Adhesion onto a Phospholipid Polymer Brush Surface Modified with a Terminal Cell Adhesion Peptide. *ACS Appl. Mater. Interfaces,* 10**,** 15250-15257.

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.

JENA, K. C. & HORE, D. K. 2010. Water Structure at Solid Surfaces and Its Implications for Biomolecule Adsorption. *Phys. Chem. Chem. Phys.,* 12**,** 14383-14404.

JENSEN, F. 1995. Locating Transition Structures by Mode Following: A Comparison of Six Methods on the Ar8 Lennard-Jones Potential. *J. Chem. Phys.,* 102**,** 6706-6718.

JIANG, S., RHYKERD, C. L. & GUBBINS, K. E. 1993. Layering, Freezing Transitions, Capillary Condensation and Diffusion of Methane in Slit Carbon Pores. *Mol. Phys.,* 79**,** 373-391.

JIMÉNEZ, V. M. & MARZAL, A. Computing the K Shortest Paths: A New Algorithm and an Experimental Comparison. *In:* VITTER, J. S. & ZAROLIAGIS, C. D., eds. International Workshop on Algorithm Engineering, 1999 London, U.K.: Springer, 15-29.

JÓNSSON, H., MILLS, G. & JACOBSEN, K. W. 1998. Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions. *In:* BERNE, B. J., CICCOTTI, G. & COKER, D. F. (eds.) *Classical and Quantum Dynamics in Condensed Phase Simulations.* Singapore: World Scientific.

JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W. & KLEIN, M. L. 1983. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.,* 79**,** 926-935.

JORGENSEN, W. L., MAXWELL, D. S. & TIRADO-RIVES, J. 1996. Development and Testing of the OPLS All-Atom Force Field on

Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.,* 118**,** 11225-11236.

JORGENSEN, W. L. & 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.

JOSEPH, J. A., CHAKRABORTY, D. & WALES, D. J. 2019. Energy Landscape for Fold-Switching in Regulatory Protein RfaH. *J. Chem. Theory Comput.,* 15**,** 731-742.

JOSEPH, J. A., RÖDER, K., CHAKRABORTY, D., MANTELL, R. G. & WALES, D. J. 2017. Exploring Biomolecular Energy Landscapes. *Chem. Commun.,* 53**,** 6974-6988.

JOSEPH, J. A. & WALES, D. J. 2018. Intrinsically Disordered Landscapes for Human CD4 Receptor Peptide. *J. Phys. Chem. B,* 122**,** 11906-11921.

KAMINSKI, G. A., FRIESNER, R. A., TIRADO-RIVES, J. & JORGENSEN, W. L. 2001. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B,* 105**,** 6474-6487.

KARIM, R., LEPELTIER, E., ESNAULT, L., PIGEON, P., LEMAIRE, L., LÉPINOUX-CHAMBAUD, C., CLERE, N., JAOUEN, G., EYER, J. & PIEL, G. 2018. Enhanced and Preferential Internalization of Lipid Nanocapsules into Human Glioblastoma Cells: Effect of a Surface-Functionalizing NFL Peptide. *Nanoscale,* 10**,** 13485-13501.

KARPLUS, M. & MCCAMMON, J. A. 2002. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Mol. Biol.,* 9**,** 646-652.

KÄSTNER, J. 2011. Umbrella Sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.,* 1**,** 932-942.

KÄSTNER, J. & SHERWOOD, P. 2008. Superlinearly Converging Dimer Method for Transition State Search. *J. Chem. Phys.,* 128**,** 014106.

KHAIT, Y. G. & PUZANOV, Y. V. 1997. Search for Stationary Points on Multidimensional Surfaces. *J. Mol. Struct. THEOCHEM,* 398-399**,** 101-109.

KIM, S. N., KUANG, Z., SLOCIK, J. M., JONES, S. E., CUI, Y., FARMER, B. L., MCALPINE, M. C. & NAIK, R. R. 2011. Preferential Binding of Peptides to Graphene Edges and Planes. *J. Am. Chem. Soc.,* 133**,** 14480-14483.

KLIMEŠ, J., BOWLER, D. R. & MICHAELIDES, A. 2010. A Critical Assessment of Theoretical Methods for Finding Reaction Pathways and Transition States of Surface Processes. *J. Phys.: Condens. Matter,* 22**,** 074203.

KOCH, W. & HOLTHAUSEN, M. C. 2001. *A Chemist's Guide to Density Functional Theory*, Weinheim, Germany, Wiley-VCH Verlag GmbH.

KOHN, W. & SHAM, L. J. 1965. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.,* 140**,** A1133-A1138.

KOJIMA, S., NAGATA, F., KUGIMIYA, S. & KATO, K. 2018. Synthesis of Peptide-Containing Calcium Phosphate Nanoparticles Exhibiting Highly Selective Adsorption of Various Proteins. *Appl. Surf. Sci.,* 458**,** 438-445.

KOKH, D. B., CORNI, S., WINN, P. J., HOEFLING, M., GOTTSCHALK, K. E. & WADE, R. C. 2010. ProMetCS: An Atomistic Force Field for Modeling Protein−Metal Surface Interactions in a Continuum Aqueous Solvent. *J. Chem. Theory Comput.,* 6**,** 1753-1768.

KOLSBJERG, E. L., GROVES, M. N. & HAMMER, B. 2016. An Automated Nudged Elastic Band Method. *J. Chem. Phys.,* 145**,** 094107.

KOMATSUZAKI, T., HOSHINO, K., MATSUNAGA, Y., RYLANCE, G. J., JOHNSTON, R. L. & WALES, D. J. 2005. How Many Dimensions Are Required to Approximate the Potential Energy Landscape of a Model Protein? *J. Chem. Phys.,* 122**,** 084714.

KOSLOVER, E. F. & WALES, D. J. 2007. Comparison of Double-Ended Transition State Search Methods. *J. Chem. Phys.,* 127**,** 134102.

KRESSE, G. & HAFNER, J. 1993. Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B: Condens. Matter Mater. Phys.,* 47**,** 558.

KRIVOV, S. V. & KARPLUS, M. 2002. Free Energy Disconnectivity Graphs: Application to Peptide Models. *J. Chem. Phys.,* 117**,** 10894-10903.

KUMAR, A., MA, H., ZHANG, X., HUANG, K., JIN, S., LIU, J., WEI, T., CAO, W., ZOU, G. & LIANG, X.-J. 2012. Gold Nanoparticles Functionalized with Therapeutic and Targeted Peptides for Cancer Treatment. *Biomaterials,* 33**,** 1180-1189.

KUMEDA, Y., WALES, D. J. & MUNRO, L. J. 2001. Transition States and Rearrangement Mechanisms from Hybrid Eigenvector-Following and Density Functional Theory. Application to C10H10 and Defect Migration in Crystalline Silicon. *Chem. Phys. Lett.,* 341**,** 185-194.

KUNZ, R. E. & BERRY, R. S. 1995. Statistical Interpretation of Topographies and Dynamics of Multidimensional Potentials. *J. Chem. Phys.,* 103**,** 1904-1912.

KUSHIMA, A., LIN, X., LI, J., EAPEN, J., MAURO, J. C., QIAN, X., DIEP, P. & YIP, S. 2009. Computing the Viscosity of Supercooled Liquids. *J. Chem. Phys.,* 130**,** 224504.

LAERA, S., CECCONE, G., ROSSI, F., GILLILAND, D., HUSSAIN, R., SILIGARDI, G. & CALZOLAI, L. 2011. Measuring Protein Structure and Stability of Protein–Nanoparticle Systems with Synchrotron Radiation Circular Dichroism. *Nano Lett.,* 11**,** 4480-4484.

LAIO, A. & PARRINELLO, M. 2002. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.,* 99**,** 12562-12566.

LATOUR, R. A. 2008. Molecular Simulation of Protein-Surface Interactions: Benefits, Problems, Solutions, and Future Directions (Review). *Biointerphases,* 3**,** FC2-FC12.

LAZARIDIS, T. & KARPLUS, M. 1999. Effective Energy Function for Proteins in Solution. *Proteins: Struct., Funct., Bioinf.,* 35**,** 133-152.

LEIMKUHLER, B. & MATTHEWS, C. 2015. *Molecular Dynamics with Deterministic and Stochastic Numerical Methods,* Cham, Switzerland, Springer International Publishing.

LEMPESIS, N., BOULOUGOURIS, G. C. & THEODOROU, D. N. 2013. Temporal Disconnectivity of the Energy Landscape in Glassy Systems. *J. Chem. Phys.,* 138**,** 12A545.

LI, B., ZHANG, R. & SHI, X. 2019. Aggregation of Amyloid Peptides into Fibrils Driven by Nanoparticles and Their Curvature Effect. *Phys. Chem. Chem. Phys.,* 21**,** 1784-1790.

LI, Z. & SCHERAGA, H. A. 1987. Monte Carlo-Minimization Approach to the Multiple-Minima Problem in Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.,* 84**,** 6611-6615.

LIANG, P., LI, Q., WU, Z., JIANG, J.-H. & YU, R.-Q. 2016. Graphene Oxide–Peptide Nanoassembly as a General Approach for Monitoring the Activity of Histone Deacetylases. *Analyst,* 141**,** 3989-3992.

LIDE, D. R. (ed.) 2009. *CRC Handbook of Chemistry and Physics,* Boca Raton, FL, U.S.A.: CRC Press.

LINDGREN, P., KASTLUNGER, G. & PETERSON, A. A. 2019. Scaled and Dynamic Optimizations of Nudged Elastic Bands. *J. Chem. Theory Comput.,* 15**,** 5787-5793.

LIU, D. C. & NOCEDAL, J. 1989. On the Limited Memory BFGS Method for Large Scale Optimization. *Math. Prog.,* 45**,** 503-528.

LIU, L., LI, Q., ZHANG, S., WANG, X., HOFFMANN, S. V., LI, J., LIU, Z., BESENBACHER, F. & DONG, M. 2016. Identification of a Novel Parallel β-Strand Conformation within Molecular Monolayer of Amyloid Peptide. *Adv. Sci.*, 3**,** 1500369.

LIU, P., KIM, B., FRIESNER, R. A. & BERNE, B. J. 2005. Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proc. Natl. Acad. Sci. U.S.A.*, 102**,** 13749-13754.

LUO, M., GAO, Y., YANG, S., QUAN, X., SUN, D., LIANG, K., LI, J. & ZHOU, J. 2019. Computer Simulations of the Adsorption of an N-Terminal Peptide of Statherin, SN15, and Its Mutants on Hydroxyapatite Surfaces. *Phys. Chem. Chem. Phys.*, 21**,** 9342-9351.

LYMAN, E. & ZUCKERMAN, D. M. 2006. Ensemble-Based Convergence Analysis of Biomolecular Trajectories. *Biophys. J.*, 91**,** 164-172.

MACKERELL, A. D., JR., BASHFORD, D., BELLOTT, DUNBRACK, R. L., JR., EVANSECK, J. D., FIELD, M. J., FISCHER, S., GAO, J., GUO, H., HA, S., JOSEPH-MCCARTHY, D., KUCHNIR, L., KUCZERA, K., LAU, F. T. K., MATTOS, C., MICHNICK, S., NGO, T., NGUYEN, D. T., PRODHOM, B., REIHER, W. E., III, ROUX, B., SCHLENKRICH, M., SMITH, J. C., STOTE, R., STRAUB, J., WATANABE, M., WIÓRKIEWICZ-KUCZERA, J., YIN, D. & KARPLUS, M. 1998. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B,* 102**,** 3586-3616.

MAHAFFY, R., BHATIA, R. & GARRISON, B. J. 1997. Diffusion of a Butanethiolate Molecule on a Au {111} Surface. *J. Phys. Chem. B,* 101**,** 771-773.

MAHMOUDI, M., LYNCH, I., EJTEHADI, M. R., MONOPOLI, M. P., BOMBELLI, F. B. & LAURENT, S. 2011. Protein−Nanoparticle Interactions: Opportunities and Challenges. *Chem. Rev.*, 111**,** 5610-5637.

MAIER, J. A., MARTINEZ, C., KASAVAJHALA, K., WICKSTROM, L., HAUSER, K. E. & SIMMERLING, C. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters From ff99SB. *J. Chem. Theory Comput.,* 11**,** 3696-3713.

MAŁOLEPSZA, E., STRODEL, B., KHALILI, M., TRYGUBENKO, S., FEJER, S. N. & WALES, D. J. 2010. Symmetrization of the AMBER and CHARMM Force Fields. *J. Comput. Chem.,* 31**,** 1402-1409.

MANNOOR, M. S., TAO, H., CLAYTON, J. D., SENGUPTA, A., KAPLAN, D. L., NAIK, R. R., VERMA, N., OMENETTO, F. G. & MCALPINE, M. C. 2012. Graphene-Based Wireless Bacteria Detection on Tooth Enamel. *Nat. Commun.,* 3**,** 763.

MAPLE, J. R., HWANG, M. J., JALKANEN, K. J., STOCKFISCH, T. P. & HAGLER, A. T. 1998. Derivation of Class II Force Fields: V. Quantum Force Field for Amides, Peptides, and Related Compounds. *J. Comput. Chem.,* 19**,** 430-458.

MARRUECOS, D. F., SCHWARTZ, D. K. & KAAR, J. L. 2018. Impact of Surface Interactions on Protein Conformation. *Curr. Opin. Colloid Interface Sci.*, 38**,** 45-55.

MARTIN, L., BILEK, M. M., WEISS, A. S. & KUYUCAK, S. 2016. Force Fields for Simulating the Interaction of Surfaces with Biological Molecules. *Interface Focus,* 6**,** 20150045.

MAS-MORUNO, C. 2018. Surface Functionalization of Biomaterials for Bone Tissue Regeneration and Repair. *In:* BARBOSA, M. A. & MARTINS, M. C. L. (eds.) *Peptides and Proteins as Biomaterials for Tissue Regeneration and Repair.* Duxford, U.K.: Elsevier.

MAURO, J. C., LOUCKS, R. J. & BALAKRISHNAN, J. 2005. A Simplified Eigenvector-Following Technique for Locating Transition Points in an Energy Landscape. *J. Phys. Chem. A,* 109**,** 9578-9583.

MCIVER, J. W. & KOMORNICKI, A. 1972. Structure of Transition States in Organic Reactions. General Theory and an Application to the Cyclobutene-Butadiene Isomerization Using a Semiempirical Molecular Orbital Method. *J. Am. Chem. Soc.,* 94**,** 2625-2633.

MCLAUGHLIN, P. J. & ZAGON, I. S. 2014. Novel Treatment for Triple-Negative Breast and Ovarian Cancer: Endogenous Opioid Suppression of Women's Cancers. *Expert Rev. Anticancer Ther.,* 14**,** 247-250.

MÉTHIVIER, C., CRUGUEL, H., PRADIER, C.-M. & HUMBLOT, V. 2017. Supramolecular Chiral Self-Assemblies of Gly–Pro Dipeptides on Metallic fcc (110) Surfaces. *Faraday Discuss.*, 204**,** 69-81.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. 1953. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.,* 21**,** 1087-1092.

MEYERS, S. R. & GRINSTAFF, M. W. 2011. Biocompatible and Bioactive Surface Modifications for Prolonged in Vivo Efficacy. *Chem. Rev.,* 112**,** 1615-1632.

MIJAJLOVIC, M. & BIGGS, M. J. 2007. Study of Conformational Switching in Polyalanine at Solid Surfaces Using Molecular Simulation. *J. Phys. Chem. C,* 111**,** 15839-15847.

MIJAJLOVIC, M., PENNA, M. J. & BIGGS, M. J. 2011. Free Energy of Adsorption of Proteins at Fluid/Solid Interfaces Using Molecular Simulation. *Chemeca 2011: Engineering a Better World.* Sydney Hilton Hotel, NSW, Australia: Engineers Australia.

MIJAJLOVIC, M., PENNA, M. J. & BIGGS, M. J. 2013. Free Energy of Adsorption for a Peptide at a Liquid/Solid Interface via Nonequilibrium Molecular Dynamics. *Langmuir,* 29**,** 2919-2926.

MONTICELLI, L., KANDASAMY, S. K., PERIOLE, X., LARSON, R. G., TIELEMAN, D. P. & MARRINK, S.-J. 2008. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.,* 4**,** 819-834.

MORTENSON, P. N., EVANS, D. A. & WALES, D. J. 2002. Energy Landscapes of Model Polyalanines. *J. Chem. Phys.,* 117**,** 1363-1376.

MU, Y. 2011. Effects of Surface Hydrophobicity on the Conformational Changes of Polypeptides of Different Length. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.,* 84**,** 031906.

MU, Y., TANG, B. & YU, M. 2014. Length-Dependent β-Sheet Growth Mechanisms of Polyalanine Peptides in Water and on Hydrophobic Surfaces. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.,* 89**,** 032711.

MÜLLER, K. & BROWN, L. D. 1979. Location of Saddle Points and Minimum Energy Paths by a Constrained Simplex Optimization Procedure. *Theor. Chim. Acta,* 53**,** 75-93.

MUNRO, L. J. & WALES, D. J. 1999. Defect Migration in Crystalline Silicon. *Phys. Rev. B: Condens. Matter Mater. Phys.,* 59**,** 3969-3980.

MURRELL, J. N. & LAIDLER, K. J. 1968. Symmetries of Activated Complexes. *Trans. Faraday Soc.,* 64**,** 371-377.

NAKANISHI, K., SAKIYAMA, T. & IMAMURA, K. 2001. On the Adsorption of Proteins on Solid Surfaces, a Common but Very Complicated Phenomenon. *J. Biosci. Bioeng.,* 91**,** 233-244.

NAKAYAMA, T., SAKURABA, T. & YAMAMOTO, Y. Self-Assembly and Adsorption Properties of Fmoc-Substituted Short Peptide Bearing Charged Side Chains. AIP Conference Proceedings, 2015. 090068.

NEELAMRAJU, S., GOSAVI, S. & WALES, D. J. 2018. Energy Landscape of the Designed Protein Top7. *J. Phys. Chem. B,* 122**,** 12282-12291.

NEELAMRAJU, S., OAKLEY, M. T. & JOHNSTON, R. L. 2015. Chiral Effects on Helicity Studied via the Energy Landscape of Short (D, L)-Alanine Peptides. *J. Chem. Phys.,* 143**,** 165103.

NEL, A. E., MADLER, L., VELEGOL, D., XIA, T., HOEK, E. M. V., SOMASUNDARAN, P., KLAESSIG, F., CASTRANOVA, V. & THOMPSON, M. 2009. Understanding Biophysicochemical Interactions at the Nano-Bio Interface. *Nat. Mater.,* 8**,** 543-557.

NEWMAN, M. E. J. 2003. The Structure and Function of Complex Networks. *SIAM Rev.,* 45**,** 167-256.

NICHOLS, J., TAYLOR, H., SCHMIDT, P. & SIMONS, J. 1990. Walking on Potential Energy Surfaces. *J. Chem. Phys.,* 92**,** 340-346.

NIKODEM, A., MATVEEV, A. V., ZHENG, B.-X. & RÖSCH, N. 2012. Efficient Two-Step Procedures for Locating Transition States of Surface Reactions. *J. Chem. Theory Comput.,* 9**,** 588-599.

NOCEDAL, J. 1980. Updating Quasi-Newton Matrices with Limited Storage. *Math. Comp.,* 35**,** 773-782.

NOCEDAL, J. 1996. Large Scale Unconstrained Optimization. *In:* DUFF, I. S. & WATSON, G. A. (eds.) *The State of the Art in Numerical Analysis.* Oxford, U.K.: Oxford University Press.

NOCEDAL, J. & WRIGHT, S. 2006. *Numerical Optimization,* New York, NY, U.S.A., Springer Science+Business Media.

NOÉ, F. & FISCHER, S. 2008. Transition Networks for Modeling the Kinetics of Conformational Change in Macromolecules. *Curr. Opin. Struct. Biol.,* 18**,** 154-162.

O'NEAL, D., TAYLOR, H. & SIMONS, J. 1984. Potential Surface Walking and Reaction Paths for C2v Be + H2 <-- BeH2 --> Be + 2H (1A1). *J. Phys. Chem.,* 88**,** 1510-1513.

OAKLEY, M. T. & JOHNSTON, R. L. 2012. Exploring the Energy Landscapes of Cyclic Tetrapeptides with Discrete Path Sampling. *J. Chem. Theory Comput.,* 9**,** 650-657.

OKAMOTO, Y. & HANSMANN, U. H. E. 1995. Thermodynamics of Helix-Coil Transitions Studied by Multicanonical Algorithms. *J. Phys. Chem.,* 99**,** 11276-11287.

OLSEN, R. A., KROES, G. J., HENKELMAN, G., ARNALDSSON, A. & JÓNSSON, H. 2004. Comparison of Methods for Finding Saddle Points Without Knowledge of the Final States. *J. Chem. Phys.,* 121**,** 9776-9792.

ONUCHIC, J. N. & WOLYNES, P. G. 2004. Theory of Protein Folding. *Curr. Opin. Struct. Biol.,* 14**,** 70-75.

OOSTENBRINK, C., VILLA, A., MARK, A. E. & VAN GUNSTEREN, W. F. 2004. 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.

ORTIZ-HERNANDEZ, M., RAPPE, K., MOLMENEU, M., MAS-MORUNO, C., GUILLEM-MARTI, J., PUNSET, M., CAPARROS, C., CALERO, J.,

FRANCH, J. & FERNANDEZ-FAIREN, M. 2018. Two Different Strategies to Enhance Osseointegration in Porous Titanium: Inorganic Thermo-Chemical Treatment Versus Organic Coating by Peptide Adsorption. *Int. J. Mol. Sci.,* 19**,** 2574.

OZBOYACI, M., KOKH, D. B., CORNI, S. & WADE, R. C. 2016a. Modeling and Simulation of Protein–Surface Interactions: Achievements and Challenges. *Q. Rev. Biophys.,* 49**,** e4.

OZBOYACI, M., KOKH, D. B. & WADE, R. C. 2016b. Three Steps to Gold: Mechanism of Protein Adsorption Revealed by Brownian and Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.,* 18**,** 10191-10200.

PAGEL, M. & BECK-SICKINGER, A. G. 2017. Multifunctional Biomaterial Coatings: Synthetic Challenges and Biological Activity. *Biol. Chem.,* 398**,** 3-22.

PARK, C. & GODDARD, W. A. 2000. Stabilization of α-Helices by Dipole-Dipole Interactions Within α-Helices. *J. Phys. Chem. B,* 104**,** 7784-7789.

PASUPULETI, M., SCHMIDTCHEN, A. & MALMSTEN, M. 2012. Antimicrobial Peptides: Key Components of the Innate Immune System. *Crit. Rev. Biotechnol.,* 32**,** 143-171.

PENNA, M. J., MIJAJLOVIC, M. & BIGGS, M. J. 2014. Molecular-Level Understanding of Protein Adsorption at the Interface Between Water and a Strongly Interacting Uncharged Solid Surface. *J. Am. Chem. Soc.,* 136**,** 5323-5331.

PENNA, M. J., MIJAJLOVIC, M., TAMERLER, C. & BIGGS, M. J. 2015. Molecular-Level Understanding of the Adsorption Mechanism of a Graphite-Binding Peptide at the Water/Graphite Interface. *Soft Matter*, 11**,** 5192-5203.

PETERS, B., HEYDEN, A., BELL, A. T. & CHAKRABORTY, A. 2004. A Growing String Method for Determining Transition States: Comparison to the Nudged Elastic Band and String Methods. *J. Chem. Phys.*, 120**,** 7877-7886.

PFAENDTNER, J. 2019. Metadynamics to Enhance Sampling in Biomolecular Simulations. *In:* BONOMI, M. & CAMILLONI, C. (eds.) *Biomolecular Simulations: Methods and Protocols.* New York, NY, U.S.A.: Springer Science+Business Media.

PHILLIPS, J. C., BRAUN, R., WANG, W., GUMBART, J., TAJKHORSHID, E., VILLA, E., CHIPOT, C., SKEEL, R. D., KALE, L. & SCHULTEN, K. 2005. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.*, 26**,** 1781-1802.

PLASENCIA GUTIÉRREZ, M., ARGÁEZ, C. & JÓNSSON, H. 2016. Improved Minimum Mode Following Method for Finding First Order Saddle Points. *J. Chem. Theory Comput.*, 13**,** 125-134.

PLIMPTON, S. 1995. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.*, 117**,** 1-19.

POLAK, E. 1971. *Computational Methods in Optimization: A Unified Approach,* New York, NY, U.S.A., Academic Press.

POLLING, S., ORMSBY, A. R., WOOD, R. J., LEE, K., SHOUBRIDGE, C., HUGHES, J. N., THOMAS, P. Q., GRIFFIN, M. D. W., HILL, A. F. & BOWDEN, Q. 2015. Polyalanine Expansions Drive a Shift into α-Helical Clusters Without Amyloid-Fibril Formation. *Nat. Struct. Mol. Biol.,* 22**,** 1008-1015.

POUDEL, L., TAMERLER, C., MISRA, A. & CHING, W.-Y. 2017. Atomic-Scale Quantification of Interfacial Binding Between Peptides and Inorganic Crystals: the Case of Calcium Carbonate Binding Peptide on Aragonite. *J. Phys. Chem. C,* 121**,** 28354-28363.

POVIMONSKY, A. G. & RAPAPORT, H. 2017. Peptide Coating Applied on the Spot Improves Osseointegration of Titanium Implants. *J. Mater. Chem. B,* 5**,** 2096-2105.

PRENTISS, M. C., WALES, D. J. & WOLYNES, P. G. 2010. The Energy Landscape, Folding Pathways and the Kinetics of a Knotted Protein. *PLoS Comput. Biol.,* 6**,** e1000835.

QIAN, J., HENTSCHKE, R. & KNOLL, W. 1997. Superstructures of Cyclodextrin Derivatives on Au (111): A Combined Random Planting− Molecular Dynamics Approach. *Langmuir,* 13**,** 7092-7098.

QIU, R., XIAO, J. & CHEN, X. D. 2017. Further Understanding of the Biased Diffusion for Peptide Adsorption on Uncharged Solid Surfaces That Strongly Interact with Water Molecules. *Colloids Surf., A,* 518**,** 197-207.

QU, Y., MA, M., WANG, Z., ZHAN, G., LI, B., WANG, X., FANG, H., ZHANG, H. & LI, C. 2013. Sensitive Amperometric Biosensor for

Phenolic Compounds Based on Graphene–Silk Peptide/Tyrosinase Composite Nanointerface. *Biosens. Bioelectron.,* 44**,** 85-88.

QUAN, X., SUN, D. & ZHOU, J. 2019. Molecular Mechanism of HIV-1 TAT Peptide and Its Conjugated Gold Nanoparticles Translocating Across Lipid Membranes. *Phys. Chem. Chem. Phys.,* 21**,** 10300-10310.

RABE, M., VERDES, D. & SEEGER, S. 2011. Understanding Protein Adsorption Phenomena at Solid Surfaces. *Adv. Colloid Interface Sci.,* 162**,** 87-106.

RAMAKRISHNAN, S. K., ZHU, J. & GERGELY, C. 2017. Organic–Inorganic Interface Simulation for New Material Discoveries. *Wiley Interdiscip. Rev. Comput. Mol. Sci.,* 7**,** e1277.

RAMEZANI-DAKHEL, H., RUAN, L., HUANG, Y. & HEINZ, H. 2015. Molecular Mechanism of Specific Recognition of Cubic Pt Nanocrystals by Peptides and of the Concentration-Dependent Formation from Seed Crystals. *Adv. Func. Mater.,* 25**,** 1374-1384.

RAO, F. & CAFLISCH, A. 2004. The Protein Folding Network. *J. Mol. Biol.,* 342**,** 299-306.

RATHORE, N., CHOPRA, M. & DE PABLO, J. J. 2005. Optimal Allocation of Replicas in Parallel Tempering Simulations. *J. Chem. Phys.,* 122**,** 024111.

REN, W. & VANDEN-EIJNDEN, E. 2013. A Climbing String Method for Saddle Point Search. *J. Chem. Phys.,* 138**,** 134105.

RÖDER, K., JOSEPH, J. A., HUSIC, B. E. & WALES, D. J. 2019. Energy Landscapes for Proteins: From Single Funnels to Multifunctional Systems. *Adv. Theory Simul.,* 2**,** 1800175.

RÖDER, K. & WALES, D. J. 2018. Energy Landscapes for the Aggregation of Aβ17–42. *J. Am. Chem. Soc.,* 140**,** 4018-4027.

ROGOSNITZKY, M., FINEGOLD, M. J., MCLAUGHLIN, P. J. & ZAGON, I. S. 2013. Opioid Growth Factor (OGF) for Hepatoblastoma: A Novel Non-Toxic Treatment. *Invest. New Drugs,* 31**,** 1066-1070.

ROSE, J. P. & BERRY, R. S. 1992. Towards Elucidating the Interplay of Structure and Dynamics in Clusters: Small KCl Clusters as Models. *J. Chem. Phys.,* 96**,** 517-538.

ROSS-NAYLOR, J. A., MIJAJLOVIC, M. & BIGGS, M. J. 2020. Energy Landscape Mapping and Replica Exchange Molecular Dynamics of an Adsorbed Peptide. *J. Phys. Chem. B,,* 124**,** 2527-2538.

ROSS-NAYLOR, J. A., MIJAJLOVIC, M., ZHANG, H. & BIGGS, M. J. 2017. Characterizing the Switching Transitions of an Adsorbed Peptide by Mapping the Potential Energy Surface. *J. Phys. Chem. B,* 121**,** 11455-11464.

ROSSI, M., BLUM, V., KUPSER, P., VON HELDEN, G., BIERAU, F., PAGEL, K., MEIJER, G. & SCHEFFLER, M. 2010. Secondary Structure of Ac-Ala n-LysH+ Polyalanine Peptides (n= 5, 10, 15) in Vacuo: Helical or Not? *J. Phys. Chem. Lett.,* 1**,** 3465-3470.

RYLANCE, G. J., JOHNSTON, R. L., MATSUNAGA, Y., LI, C.-B., BABA, A. & KOMATSUZAKI, T. 2006. Topographical Complexity of Multidimensional Energy Landscapes. *Proc. Natl. Acad. Sci. U.S.A.,* 103**,** 18551-18555.

SAAD, Y., CHELIKOWSKY, J. R. & SHONTZ, S. M. 2010. Numerical Methods for Electronic Structure Calculations of Materials. *SIAM Rev.,* 52**,** 3-54.

SACCHI, M., WALES, D. J. & JENKINS, S. J. 2017. Energy Landscapes and Dynamics of Glycine on Cu (110). *Phys. Chem. Chem. Phys.,* 19**,** 16600-16605.

SALAPAKA, S. M. & SALAPAKA, M. V. 2008. Scanning Probe Microscopy. *IEEE Control Systems Magazine.*

SALOMON-FERRER, R., CASE, D. A. & WALKER, R. C. 2013. An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.,* 3**,** 198-210.

SAURA, P., RÖPKE, M., GAMIZ-HERNANDEZ, A. P. & KAILA, V. R. 2019. Quantum Chemical and QM/MM Models in Biochemistry. *In:* BONOMI, M. & CAMILLONI, C. (eds.) *Biomolecular Simulations: Methods and Protocols.* New York, NY, U.S.A.: Springer Science+Business Media.

SAWLE, L. & GHOSH, K. 2016. Convergence of Molecular Dynamics Simulation of Protein Native States: Feasibility vs Self-Consistency Dilemma. *J. Chem. Theory Comput.,* 12**,** 861-869.

SCHAFER, L., NEWTON, S. Q., CAO, M., PEETERS, A., VAN ALSENOY, C., WOLINSKI, K. & MOMANY, F. A. 1993. Evaluation of the Dipeptide Approximation in Peptide Modeling by Ab Initio Geometry Optimizations of Oligopeptides. *J. Am. Chem. Soc.,* 115**,** 272-280.

SCHERMANN, J.-P. 2008. *Spectroscopy and Modelling of Biomolecular Building Blocks,* Amsterdam, The Netherlands, Elsevier.

SCHLEGEL, H. B. 2011. Geometry Optimization. *Wiley Interdiscip. Rev. Comput. Mol. Sci.,* 1**,** 790-809.

SCHMID, N., EICHENBERGER, A. P., CHOUTKO, A., RINIKER, S., WINGER, M., MARK, A. E. & VAN GUNSTEREN, W. F. 2011. Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. *Eur. Biophys. J.,* 40**,** 843-856.

SCHWAMINGER, S., BLANK-SHIM, S. A., BORKOWSKA-PANEK, M., ANAND, P., FRAGA-GARCÍA, P., FINK, K., WENZEL, W. & BERENSMEIER, S. 2018. Experimental Characterization and Simulation of Amino Acid and Peptide Interactions with Inorganic Materials. *Eng. Life Sci.,* 18**,** 84-100.

SEK, S., TOLAK, A., MISICKA, A., PALYS, B. & BILEWICZ, R. 2005. Asymmetry of Electron Transmission Through Monolayers of Helical Polyalanine Adsorbed on Gold Surfaces. *J. Phys. Chem. B,* 109**,** 18433-18438.

SEKER, U. O. S. & DEMIR, H. V. 2011. Material Binding Peptides for Nanotechnology. *Molecules,* 16**,** 1426-1451.

SHANG, C. & LIU, Z.-P. 2010. Constrained Broyden Minimization Combined with the Dimer Method for Locating Transition State of Complex Reactions. *J. Chem. Theory Comput.,* 6**,** 1136-1144.

SHAO, J., TANNER, S. W., THOMPSON, N. & CHEATHAM, T. E. 2007. Clustering Molecular Dynamics Trajectories: 1. Characterizing the

Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.,* 3**,** 2312-2334.

SHARPE, D. J. & WALES, D. J. 2019. Identifying Mechanistically Distinct Pathways in Kinetic Transition Networks. *J. Chem. Phys.,* 151**,** 124101.

SHEPPARD, D., TERRELL, R. & HENKELMAN, G. 2008. Optimization Methods for Finding Minimum Energy Paths. *J. Chem. Phys.,* 128**,** 134106.

SHIBA, K. 2017. Programmable Bio-surfaces for Biomedical Applications. *In:* SUNNA, A., CARE, A. & BERGQUIST, P. L. (eds.) *Peptides and Peptide-based Biomaterials and their Biomedical Applications.* Cham, Switzerland: Springer International Publishing.

SIMONS, J., JOERGENSEN, P., TAYLOR, H. & OZMENT, J. 1983. Walking on Potential Energy Surfaces. *J. Phys. Chem.,* 87**,** 2745-2753.

SKWARCZYNSKI, M. & TOTH, I. 2016. Peptide-Based Synthetic Vaccines. *Chem. Sci.,* 7**,** 842-854.

SLOCIK, J. M., DENNIS, P. B., GOVOROV, A. O., BEDFORD, N. M., REN, Y. & NAIK, R. R. 2019. Chiral Restructuring of Peptide Enantiomers on Gold Nanomaterials. *ACS Biomater. Sci. Eng.***,** In press at time of submission.

SLOCIK, J. M., GOVOROV, A. O. & NAIK, R. R. 2011. Plasmonic Circular Dichroism of Peptide-Functionalized Gold Nanoparticles. *Nano Lett.,* 11**,** 701-705.

SLOCIK, J. M. & NAIK, R. R. 2017. Sequenced Defined Biomolecules for Nanomaterial Synthesis, Functionalization, and Assembly. *Curr. Opin. Biotechnol.,* 46**,** 7-13.

SMEETON, L. C., OAKLEY, M. T. & JOHNSTON, R. L. 2014. Visualizing Energy Landscapes with Metric Disconnectivity Graphs. *J. Comput. Chem.,* 35**,** 1481-1490.

SMIATEK, J. & HEUER, A. 2011. Calculation of Free Energy Landscapes: A Histogram Reweighted Metadynamics Approach. *J. Comput. Chem.,* 32**,** 2084-2096.

SMIDSTRUP, S., PEDERSEN, A., STOKBRO, K. & JÓNSSON, H. 2014. Improved Initial Guess for Minimum Energy Path Calculations. *J. Chem. Phys.,* 140**,** 214106.

SOLTANI, N. & GHOLAMI, M. R. 2017. Increase in the β-Sheet Character of an Amyloidogenic Peptide upon Adsorption onto Gold and Silver Surfaces. *ChemPhysChem,* 18**,** 526-536.

SOMANI, S. & WALES, D. J. 2013. Energy Landscapes and Global Thermodynamics for Alanine Peptides. *J. Chem. Phys.,* 139**,** 121909.

SOUM, C., RUBIO-ALBENQUE, S., FERY-FORGUES, S., DÉLÉRIS, G., ALOUINI, M.-A. & BERTHELOT, T. 2015. Supramolecular Peptide/Surface Assembly for Monitoring Proteinase Activity and Cancer Diagnosis. *ACS Appl. Mater. Interfaces,* 7**,** 16967-16975.

SOUSA, S. F., FERNANDES, P. A. & RAMOS, M. J. 2007. General Performance of Density Functionals. *J. Phys. Chem. A,* 111**,** 10439-10452.

SPRENGER, K. G., PRAKASH, A., DROBNY, G. & PFAENDTNER, J. 2018. Investigating the Role of Phosphorylation in the Binding of Silaffin Peptide R5 to Silica with Molecular Dynamics Simulations. *Langmuir,* 34**,** 1199-1207.

STEELE, W. A. 1973. The Physical Interaction of Gases with Crystalline Solids: I. Gas-Solid Energies and Properties of Isolated Adsorbed Atoms. *Surf. Sci.,* 36**,** 317-352.

STEELE, W. A. 1978. The Interaction of Rare Gas Atoms with Graphitized Carbon Black. *J. Phys. Chem.,* 82**,** 817-821.

STILLINGER, F. H. & WEBER, T. A. 1982. Hidden Structure in Liquids. *Phys. Rev. A: At., Mol., Opt. Phys.,* 25**,** 978-989.

STILLINGER, F. H. & WEBER, T. A. 1984. Packing Structures and Transitions in Liquids and Solids. *Science,* 225**,** 983-989.

STRODEL, B., LEE, J. W. L., WHITTLESTON, C. S. & WALES, D. J. 2010. Transmembrane Atructures for Alzheimer's Aβ1−42 Oligomers. *J. Am. Chem. Soc.,* 132**,** 13300-13312.

STRODEL, B. & WALES, D. J. 2008. Free Energy Surfaces from an Extended Harmonic Superposition Approach and Kinetics for Alanine Dipeptide. *Chem. Phys. Lett.,* 466**,** 105-115.

STRODEL, B., WHITTLESTON, C. S. & WALES, D. J. 2007. Thermodynamics and Kinetics of Aggregation for the GNNQQNY Peptide. *J. Am. Chem. Soc.,* 129**,** 16005-16014.

SUGITA, Y., KAMIYA, M., OSHIMA, H. & RE, S. 2019. Replica-Exchange Methods for Biomolecular Simulations. *In:* BONOMI, M. & CAMILLONI, C. (eds.) *Biomolecular Simulations: Methods and Protocols.* New York, NY, U.S.A.: Springer Science+Business Media.

SUGITA, Y. & OKAMOTO, Y. 1999. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.,* 314**,** 141-151.

SUN, H. 1998. COMPASS: An Ab Initio Force-field Optimized for Condensed-phase Applications Overview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B,* 102**,** 7338-7364.

SUN, Y., DOMINY, B. N. & LATOUR, R. A. 2007. Comparison of Solvation-Effect Methods for the Simulation of Peptide Interactions with a Hydrophobic Surface. *J. Comput. Chem.,* 28**,** 1883-1892.

SUN, Y. & LATOUR, R. A. 2006. Comparison of Implicit Solvent Models for the Simulation of Protein–Surface Interactions. *J. Comput. Chem.,* 27**,** 1908-1922.

SUNNA, A., CARE, A. & BERGQUIST, P. L. (eds.) 2017. *Peptides and Peptide-based Biomaterials and their Biomedical Applications,* Cham, Switzerland: Springer International Publishing.

SWENDSEN, R. H. & WANG, J.-S. 1986. Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.,* 57**,** 2607-2609.

SWOPE, W. C., ANDERSEN, H. C., BERENS, P. H. & WILSON, K. R. 1982. A Computer Simulation Method for the Calculation of Equilibrium

Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.,* 76**,** 637-649.

SZWEDA, R., TRZEBICKA, B., DWORAK, A., OTULAKOWSKI, L., KOSOWSKI, D., HERTLEIN, J., HALADJOVA, E., RANGELOV, S. & SZWEDA, D. 2016. Smart Polymeric Nanocarriers of Met-Enkephalin. *Biomacromolecules,* 17**,** 2691-2700.

TAO, P., LARKIN, J. D. & BROOKS, B. R. 2012. Reaction Path Optimization and Sampling Methods and Their Applications for Rare Events. *In:* PAHLAVANI, M. R. (ed.) *Some Applications of Quantum Mechanics.* Rijeka, Croatia: InTech.

TOPOL, I. A., BURT, S. K., DERETEY, E., TANG, T.-H., PERCZEL, A., RASHIN, A. & CSIZMADIA, I. G. 2001. α-and 310-Helix Interconversion: A Quantum-Chemical Study on Polyalanine Systems in the Gas Phase and in Aqueous Solvent. *J. Am. Chem. Soc.,* 123**,** 6054-6060.

TORRIE, G. M. & VALLEAU, J. P. 1977. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.,* 23**,** 187-199.

TRINDADE, R., ALBREKTSSON, T., TENGVALL, P. & WENNERBERG, A. 2016. Foreign Body Reaction to Biomaterials: On Mechanisms for Buildup and Breakdown of Osseointegration. *Clin. Implant Dent. Relat. Res.,* 18**,** 192-203.

TRUCANO, P. & CHEN, R. 1975. Structure of Graphite by Neutron-Diffraction. *Nature,* 258**,** 136-137.

TRYGUBENKO, S. A. & WALES, D. J. 2004. A Doubly Nudged Elastic Band Method for Finding Transition States. *J. Chem. Phys.,* 120**,** 2082-2094.

TRYGUBENKO, S. A. & WALES, D. J. 2006. Kinetic Analysis of Discrete Path Sampling Stationary Point Databases. *Mol. Phys.,* 104**,** 1497-1507.

TSAI, C. J. & JORDAN, K. D. 1993. Use of an Eigenmode Method to Locate the Stationary Points on the Potential Energy Surfaces of Selected Argon and Water Clusters. *J. Phys. Chem.,* 97**,** 11227-11237.

VALSSON, O., TIWARY, P. & PARRINELLO, M. 2016. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annu. Rev. Phys. Chem.,* 67**,** 159-184.

VERLET, L. 1967. Computer "Experiments" On Classical Fluids. I. Thermodynamical Properties of Lennard-jones Molecules. *Phys. Rev.,* 159**,** 98-103.

VOGEL, Z. & ALTSTEIN, M. 1977. The Adsorption of Enkephalin to Porous Polystyrene Beads: A Simple Assay for Enkephalin Hydrolysis. *FEBS Lett.,* 80**,** 332-336.

WALES, D. J. 1990. Structural and Topological Consequences of Anisotropic Interactions in Clusters. *J. Chem. Soc. Faraday Trans.,* 86**,** 3505-3517.

WALES, D. J. 1994. Rearrangements of 55-Atom Lennard-Jones and (C60)55 Clusters. *J. Chem. Phys.,* 101**,** 3750-3762.

WALES, D. J. 2002. Discrete Path Sampling. *Mol. Phys.,* 100**,** 3285-3305.

WALES, D. J. 2003. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses,* Cambridge, U.K., Cambridge University Press.

WALES, D. J. 2004. Some Further Applications of Discrete Path Sampling to Cluster Isomerization. *Mol. Phys.,* 102**,** 891-908.

WALES, D. J. 2006. Energy Landscapes: Calculating Pathways and Rates. *Int. Rev. Phys. Chem.,* 25**,** 237-282.

WALES, D. J. 2009. Calculating Rate Constants and Committor Probabilities for Transition Networks by Graph Transformation. *J. Chem. Phys.,* 130**,** 204111.

WALES, D. J. 2013. Surveying a Complex Potential Energy Landscape: Overcoming Broken Ergodicity Using Basin-Sampling. *Chem. Phys. Lett.,* 584**,** 1-9.

WALES, D. J. 2015. Perspective: Insight into Reaction Coordinates and Dynamics from the Potential Energy Landscape. *J. Chem. Phys.,* 142**,** 130901.

WALES, D. J. 2017. Decoding Heat Capacity Features from the Energy Landscape. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.,* 95**,** 030105.

WALES, D. J. 2018. Exploring Energy Landscapes. *Annu. Rev. Phys. Chem.,* 69**,** 401-425.

WALES, D. J. & BOGDAN, T. V. 2006. Potential Energy and Free Energy Landscapes. *J. Phys. Chem. B,* 110**,** 20765-20776.

WALES, D. J. & CARR, J. M. 2012. Quasi-Continuous Interpolation Scheme for Pathways Between Distant Configurations. *J. Chem. Theory Comput.,* 8**,** 5020-5034.

WALES, D. J., DISNEY, M. D. & YILDIRIM, I. 2019. Computational Investigation of RNA A-Bulges Related to the Microtubule-Associated Protein Tau Causing Frontotemporal Dementia and Parkinsonism. *J. Phys. Chem. B,* 123**,** 57-65.

WALES, D. J. & DOYE, J. P. K. 1997. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A,* 101**,** 5111-5116.

WALES, D. J. & HEAD-GORDON, T. 2012. Evolution of the Potential Energy Landscape with Static Pulling Force for Two Model Proteins. *J. Phys. Chem. B,* 116**,** 8394-8411.

WALES, D. J., MILLER, M. A. & WALSH, T. R. 1998. Archetypal Energy Landscapes. *Nature,* 394**,** 758-760.

WALES, D. J. & WALSH, T. R. 1996. Theoretical Study of the Water Pentamer. *J. Chem. Phys.,* 105**,** 6957-6971.

WALSH, T. R. 2017. Pathways to Structure–Property Relationships of Peptide–Materials Interfaces: Challenges in Predicting Molecular Structures. *Acc. Chem. Res.,* 50**,** 1617-1624.

WALSH, T. R. & KNECHT, M. R. 2017. Biointerface Structural Effects on the Properties and Applications of Bioinspired Peptide-Based Nanomaterials. *Chem. Rev.,* 117**,** 12641-12704.

WANG, D.-M., JIAO, X., PLOTNIKOFF, N. P., GRIFFIN, N., QI, R.-Q., GAO, X.-H. & SHAN, F.-P. 2017. Killing Effect of Methionine Enkephalin on Melanoma In Vivo and In Vitro. *Oncol. Rep.*, 38**,** 2132-2140.

WANG, J., CIEPLAK, P. & KOLLMAN, P. A. 2000. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.*, 21**,** 1049-1074.

WEINER, P. K. & KOLLMAN, P. A. 1981. AMBER: Assisted Model Building with Energy Refinement. A General Program for Modeling Molecules and Their Interactions. *J. Comput. Chem.*, 2**,** 287-303.

WESTERBERG, K. M. & FLOUDAS, C. A. 1999. Locating All Transition States and Studying the Reaction Pathways of Potential Energy Surfaces. *J. Chem. Phys.*, 110**,** 9259-9295.

WOODY, R. W. 1995. Circular Dichroism. *Methods Enzymol.*, 246**,** 34-71.

WRIGHT, L. B., PALAFOX-HERNANDEZ, J. P., RODGER, P. M., CORNI, S. & WALSH, T. R. 2015. Facet Selectivity in Gold Binding Peptides: Exploiting Interfacial Water Structure. *Chem. Sci.*, 6**,** 5204-5214.

WRIGHT, L. B., RODGER, P. M., CORNI, S. & WALSH, T. R. 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.

WRIGHT, L. B. & WALSH, T. R. 2013. Efficient Conformational Sampling of Peptides Adsorbed onto Inorganic Surfaces: Insights from a Quartz Binding Peptide. *Phys. Chem. Chem. Phys.,* 15**,** 4715-4726.

XU, Z., YANG, X., WEI, Q., ZHAO, W., CUI, B., YANG, X. & SAHAI, N. 2018. Quantitatively Identifying the Roles of Interfacial Water and Solid Surface in Governing Peptide Adsorption. *Langmuir,* 34**,** 7932-7941.

YANG, Y., YANG, B., LI, M., WANG, Y., YANG, X. & LI, J. 2017. Salivary Acquired Pellicle-Inspired DpSpSEEKC Peptide for the Restoration of Demineralized Tooth Enamel. *Biomed. Mater.,* 12**,** 025007.

YIN, X., LI, B., LIU, S., GU, Z., ZHOU, B. & YANG, Z. 2019. Effect of the Surface Curvature on Amyloid-β Peptide Adsorption for Graphene. *RSC Adv.,* 9**,** 10094-10099.

YOUNG, D. C. 2001. *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems,* New York, NY, U.S.A., John Wiley & Sons, Inc.

ZAGON, I. S. & MCLAUGHLIN, P. J. 2014. Opioid Growth Factor and the Treatment of Human Pancreatic Cancer: A Review. *World J. Gastroenterol.,* 20**,** 2218-2223.

ZHANG, J. & DU, Q. 2012. Shrinking Dimer Dynamics and Its Applications to Saddle Point Search. *SIAM J. Numer. Anal.,* 50**,** 1899-1921.

ZHANG, L., DU, Q. & ZHENG, Z. 2016. Optimization-Based Shrinking Dimer Method for Finding Transition States. *SIAM J. Sci. Comput.,* 38**,** A528-A544.

ZHANG, L. & HERMANS, J. 1994. 310 Helix Versus α-Helix: A Molecular Dynamics Study of Conformational Preferences of Aib and Alanine. *J. Am. Chem. Soc.,* 116**,** 11915-11921.

ZHANG, L. & SUN, Y. 2018. Charged Surface Regulates the Molecular Interactions of Electrostatically Repulsive Peptides by Inducing Oriented Alignment. *Langmuir,* 34**,** 4390-4397.

ZHANG, M., YIN, B.-C., WANG, X.-F. & YE, B.-C. 2011. Interaction of Peptides with Graphene Oxide and Its Application for Real-Time Monitoring of Protease Activity. *Chem. Commun.,* 47**,** 2399-2401.

ZHANG, X., BHATT, D. & ZUCKERMAN, D. M. 2010. Automated Sampling Assessment for Molecular Simulations Using the Effective Sample Size. *J. Chem. Theory Comput.,* 6**,** 3048-3057.

ZIMMERMAN, P. M. 2013. Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. Theory Comput.,* 9**,** 3043-3050.

ZIMMERMAN, P. M. 2015. Navigating Molecular Space for Reaction Mechanisms: An Efficient, Automated Procedure. *Mol. Simulat.,* 41**,** 43-54.

ZOZULIA, O., DOLAN, M. A. & KORENDOVYCH, I. V. 2018. Catalytic Peptide Assemblies. *Chem. Soc. Rev.,* 47**,** 3621-3639.