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

*James A. Ross-Naylor†, Milan Mijajlovic‡ and Mark J. Biggs§**

†School of Chemical Engineering, The University of Adelaide, Adelaide, South Australia, 5005 Australia

‡School of Engineering, Newcastle University, NE1 7RU, United Kingdom

§College of Science and Technology, Nottingham Trent University, Nottingham, NG11 8NS, United Kingdom

*Corresponding author. Phone: +44 (0)115 848 3840. E-mail: mark.biggs@ntu.ac.uk.

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ABSTRACT

Adsorption of peptides at the interface between a fluid and a solid occurs widely in both nature and applications. Knowing the dominant conformations of adsorbed peptides and the energy barriers between them is of interest for a variety of reasons. Molecular dynamics (MD) simulation is a widely used technique that can yield such understanding. However, the complexity of the energy landscapes of adsorbed peptides means comprehensive exploration of the energy landscape by MD simulation is challenging. An alternative approach is energy landscape mapping (ELM), which involves the location of stationary points on the potential energy surface, and its analysis to determine, for example, the pathways and energy barriers between them. In the study reported here, comparison is made between this technique and replica exchange molecular dynamics (REMD) for met-enkephalin adsorbed at the interface between graphite and the gas phase: the first ever direct comparison of these techniques for adsorbed peptides. Both methods yield up the dominant adsorbed peptide conformations. Unlike REMD, however, ELM also readily allows the identification of the connectivity and energy barriers between the favored conformations, transition paths and structures between these conformations, and the impact of entropy. It also permits the calculation of the constant volume heat capacity, although the accuracy of this is limited by the sampling of high-energy minima. Overall, compared to REMD, ELM provides additional insight into the adsorbed peptide system provided sufficient care is taken to ensure key parts of the landscape are adequately sampled.

1 INTRODUCTION

Peptides are often used in the functionalization of solid surfaces due to their high degree of programmability.[1] Applications of material-binding peptides include the formation and

functionalization of nanomaterials,[2-4] protective or bioactive coatings for medical implants and other biomaterials,[5-6] biosensing,[7-8] cancer diagnosis[9] and drug delivery.[10] 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.[11-12] 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.[13-15]

Computational studies have led to an improved understanding of peptide-surface interactions in recent years.[16-18] 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.[17] 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.[19] 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.[16] However, this occurs at the cost of the model's accuracy, and the development of reliable coarse-grained models remains a work in progress.[18]

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).[20-23] 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,[24] understanding the nature of the PES and FES through disconnectivity graphs[25-26] and transition mechanisms between stable configurations *via* discrete path sampling (DPS).[27]

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.[28] 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.[29] The system considered was met-enkephalin adsorbed at the gas/graphite interface, which has been previously studied by some of the authors.[30] 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.

2 METHODOLOGY

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.[31] 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.[32] 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.[33] In ELM calculations, peptide-surface interactions were modeled by the Steele potential:[34]

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

where l = 0, …, 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 1.

Table 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] Ref. 33

[b] Ref. 34

[c] Derived from literature bond length and unit cell height data.[35]

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.

2.2 METHODS

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.[36] 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.[37] 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.[36] 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.[38-39] 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[40] 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,[25] 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[27] 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[41] was used to determine the fastest transition path,[42] and double-ended transition state searches were applied between pairs of minima on this path, selected in order of increasing Euclidean separation.[43] A double-ended transition state search

involves the use of the doubly nudged elastic band (DNEB) method[44] 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.[21,45] 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.[24] Additionally, optimal transition pathways between selected low PE minima were determined using the KSHORTESTPATHS algorithm,[46] 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.[47] 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.[48]

ELM was implemented using the software of Wales and co-workers, which is freely available on their website.[49] This software was interfaced with the CHARMM program for PE calculations,[50] which was modified to include the Steele potential.

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 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 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 1(a).

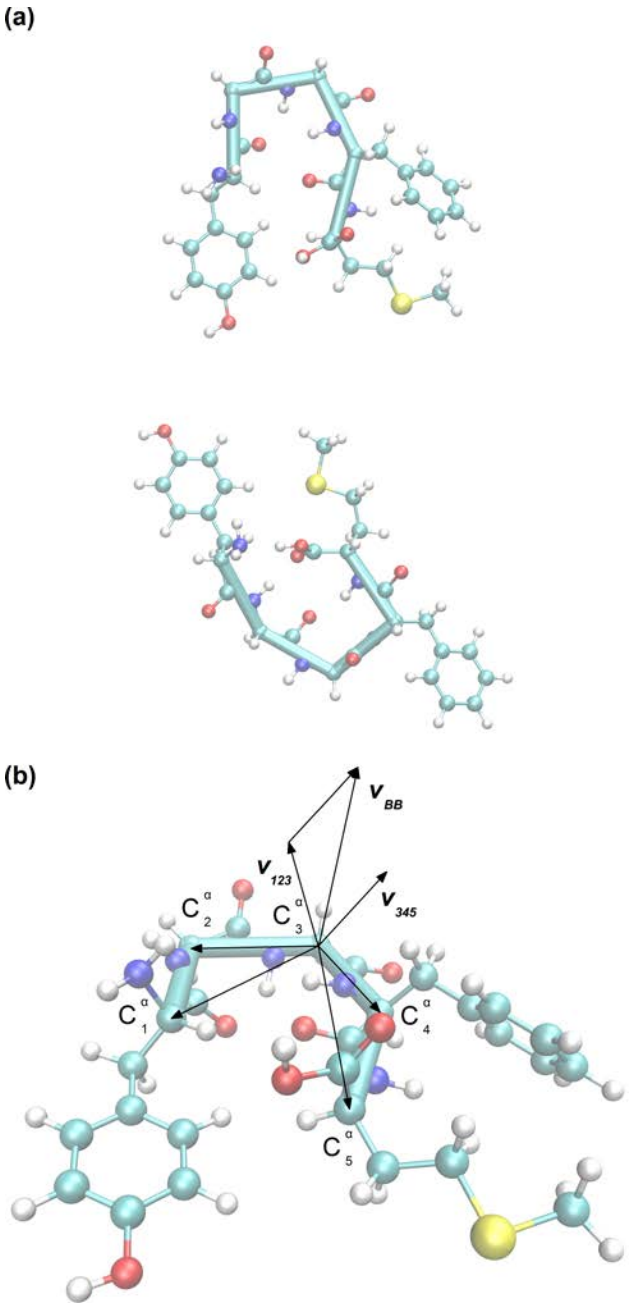

Figure 1. The orientation of met-enkephalin to the graphite surface: (a) the two predominant adsorbed orientations viewed from directly above the surface, which is not shown; and (b) the definition of the backbone normal vector, $\boldsymbol{v}_{\mathbf{BB}}$, that is used to define the orientation of the molecule to the surface, which is the sum of $\boldsymbol{v}_{\mathbf{123}}$ and $\boldsymbol{v}_{\mathbf{345}}$ 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.

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 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, 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,[51] 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[52] 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.[53] 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.[52] 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:

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

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.[54]

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:

$$C_V = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2}$$

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).[55] 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).[56]

2.3 STUDY DETAILS

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 $kcal^2/mol^2Å^2$, and a convergence criterion of the RMS gradient not exceeding 10^{-7} kcal/molÅ.

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 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.*,[57] 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.[58] 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.[58] Accordingly, an exchange period of 5 ps and 40,000 exchange attempts were chosen, yielding a total simulation length of 200 ns.

3 RESULTS AND DISCUSSION

3.1 PERFORMANCE OF REMD SIMULATION

Figure 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,[57] although the variation does indicate that the temperatures chosen for the replicas are not distributed in the most computationally efficient way.

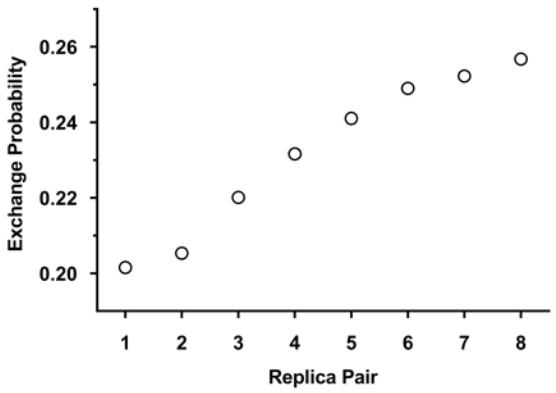

Figure 2. The REMD exchange probability for each pair of replicas where pairs are ordered by increasing temperature: (1) 300 K and 354 K, (2) 354 K and 418 K, and so forth.

Figure 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 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.[54] However, given the steps taken to ensure sufficient mixing of the replicas, we are confident that this possibility has been minimized as far as reasonably practicable.

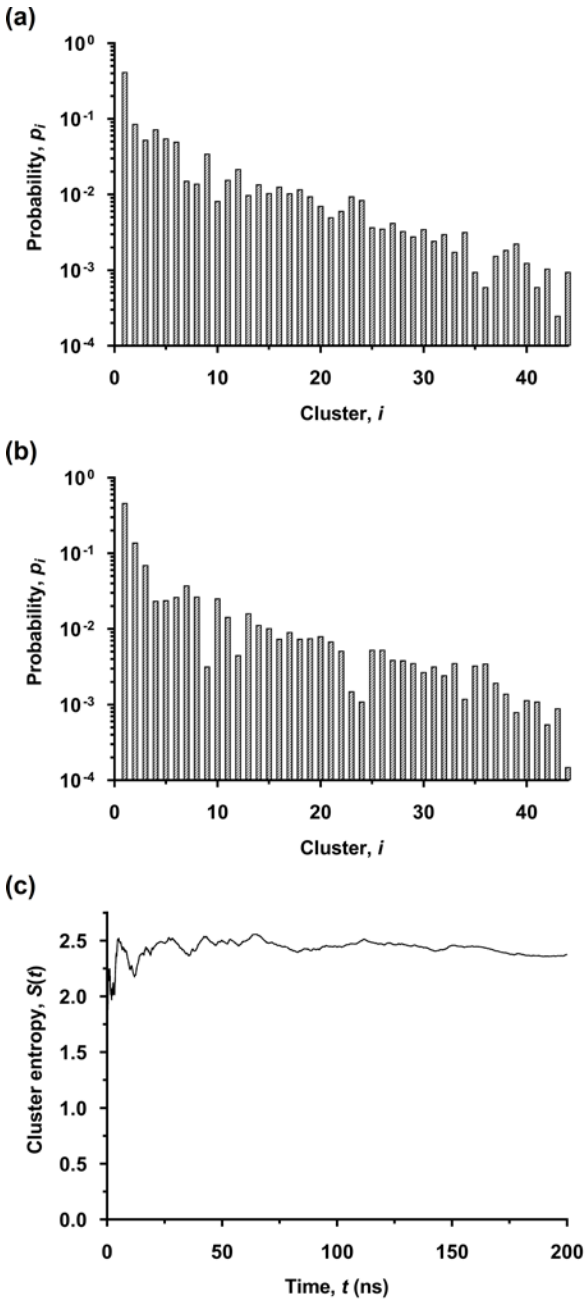

Figure 3. Results illustrating the self-consistency of the REMD simulation: (a) Cluster probability distribution of the 300 K replica in the first 100 ns of the REMD simulation; (b) cluster probability distribution of the 300 K replica in the second 100 ns of the simulation; and (c) variation of the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

3.2 REMD CLUSTER ANALYSIS

Figure 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 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 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.

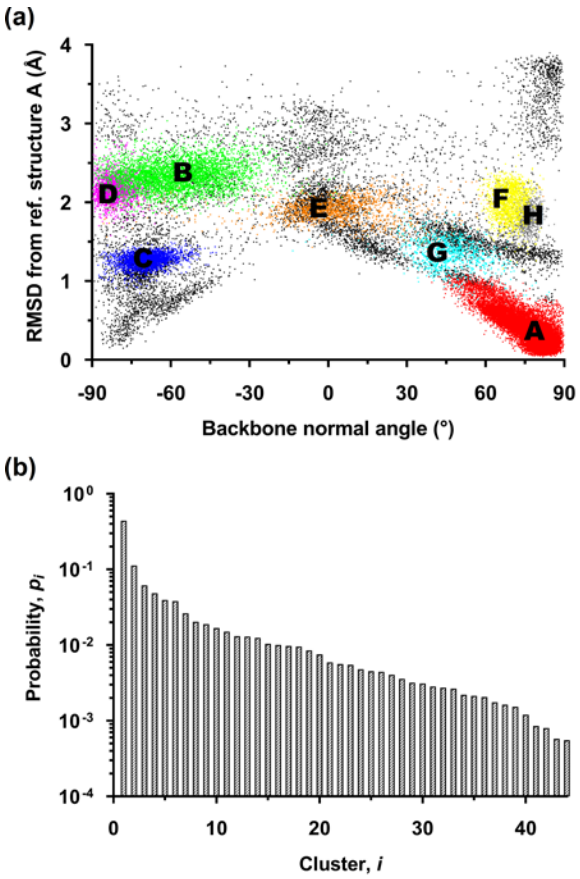

Figure 4. Results of the REMD clustering algorithm: (a) all adsorbed snapshots taken during the REMD simulation, plotted according to their backbone normal angle and RMSD from the reference structure of cluster A, with snapshots belonging to clusters A through H colored accordingly; and (b) cluster probability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

Figure 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 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 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,[30] indicating good agreement with that prior work. Structure E also is of 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 1) pointing in opposite directions with respect to the surface, accounting for the near-zero backbone normal angle observed in Figure 4(a).

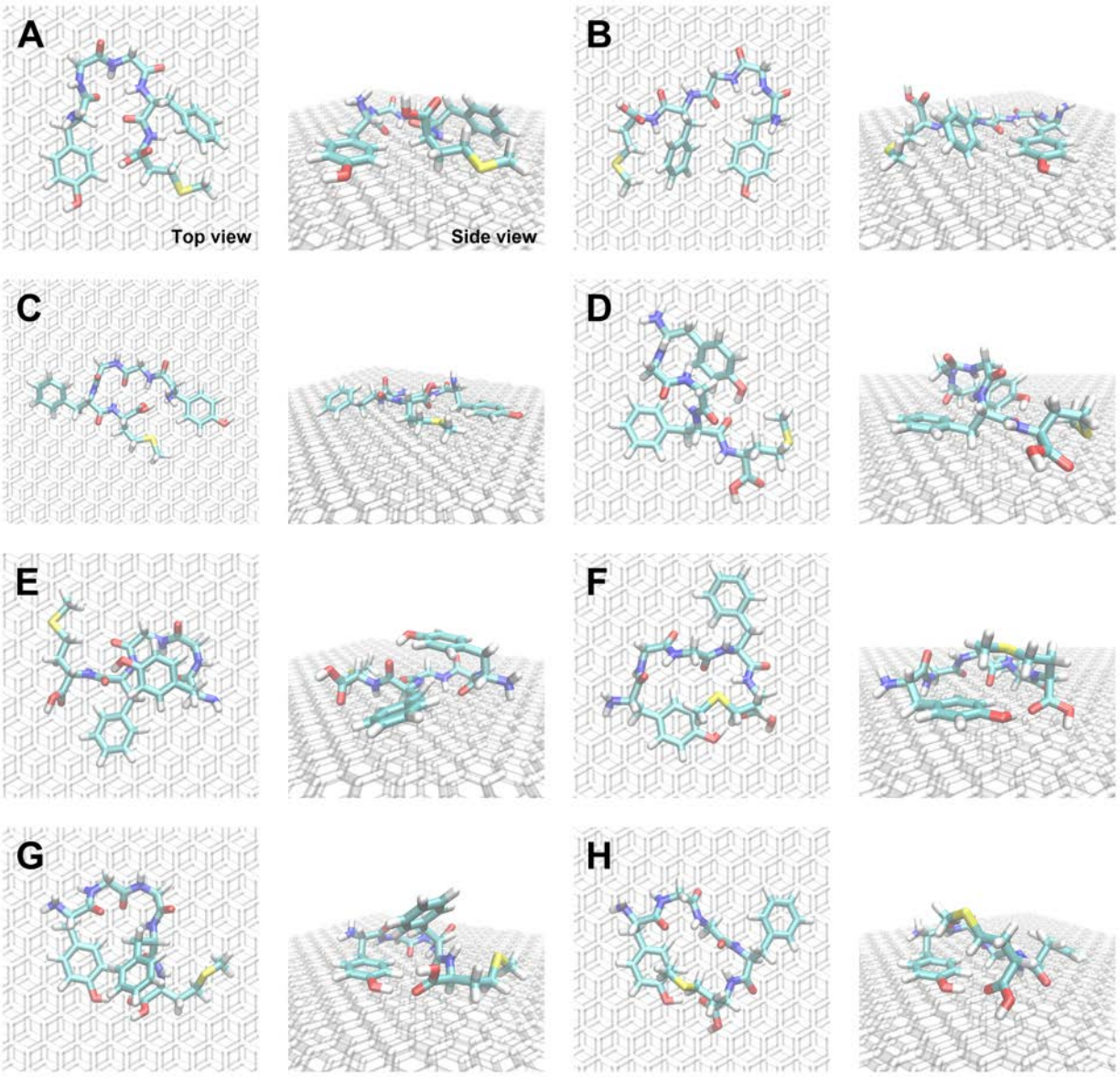

Figure 5. Reference structures for clusters A through H.

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.[28,59] 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,[21] the evaluation of these points provides an indicator of how extensively ELM has probed the PES. Figure 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.

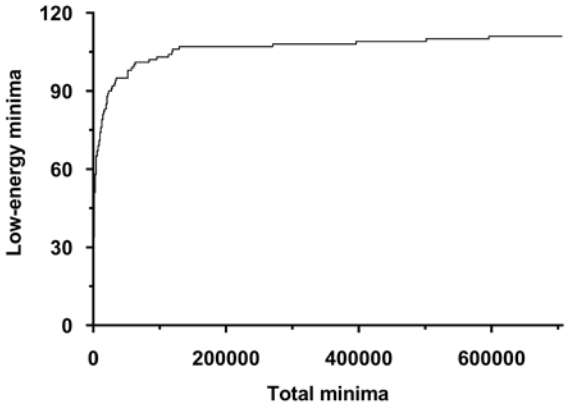

Figure 6. Variation of the number of low-lying minima (i.e. within 2.5 kcal/mol of the global minimum) with the total number of minima found through the ELM simulation.

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 signficantly 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 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 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 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.[25] 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.[21] 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, while the presence of the graphite surface may provide a significant geometric obstacle to some conformational transitions.

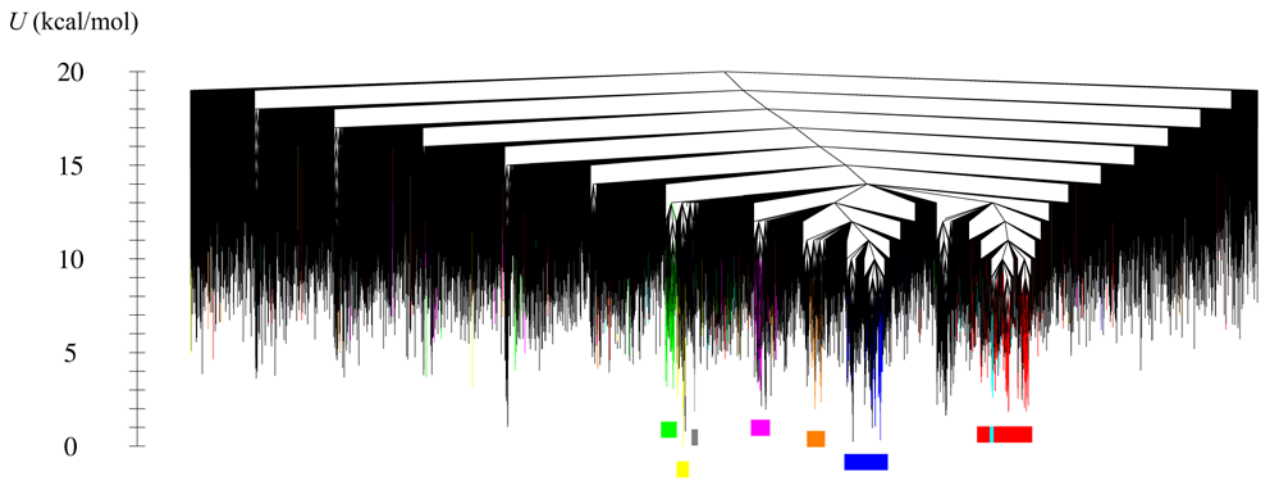

Figure 7. PE disconnectivity graph. All minima connected to the global minimum by a maximum-energy saddle point below U = 20 kcal/mol are shown. Colored minima are the associated minima of REMD snapshots from the most prominent clusters: A (red), B (green), C (blue), D (pink), E (orange), F (yellow),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

The colored minima in Figure 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 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 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 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.[60-61] 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 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[62] that the ratio of the sampling probabilities 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:

$$\frac{p_i}{p_j} = e^{-\frac{\Delta G_{ij}}{k_B T}}$$

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 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 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,[63] and is therefore unlikely to account for any observable discrepancies in results.

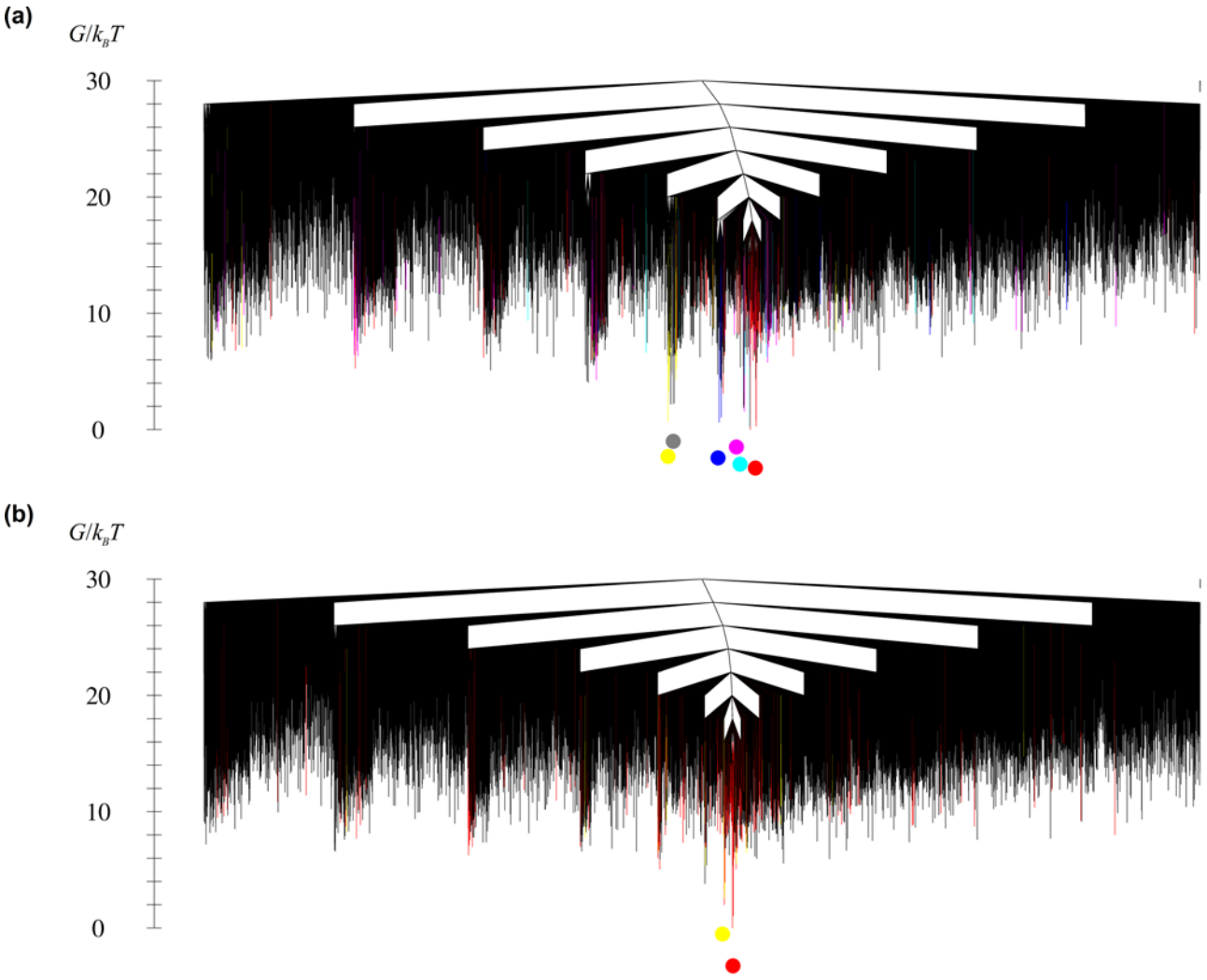

Figure 8. FE disconnectivity graphs at: (a) 300 K and (b) 350 K. All minima connected to the global minimum by a maximum-energy saddle point below $G = 30k_BT$ are shown. Colored minima are the associated minima of REMD snapshots from the most prominent clusters: A

(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.

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 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 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 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.[47] 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 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.

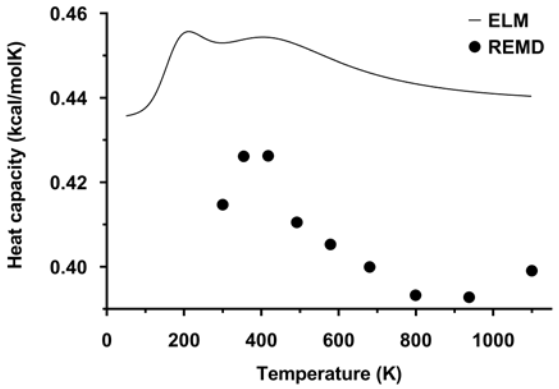

Figure 9. Constant volume heat capacities, as a function of temperature, calculated from the stationary point database obtained from ELM, and separately from the REMD simulation.

Figure 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 to the system's dynamics as the shown path. As displayed in the inset structures on Figure 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.[64] 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.

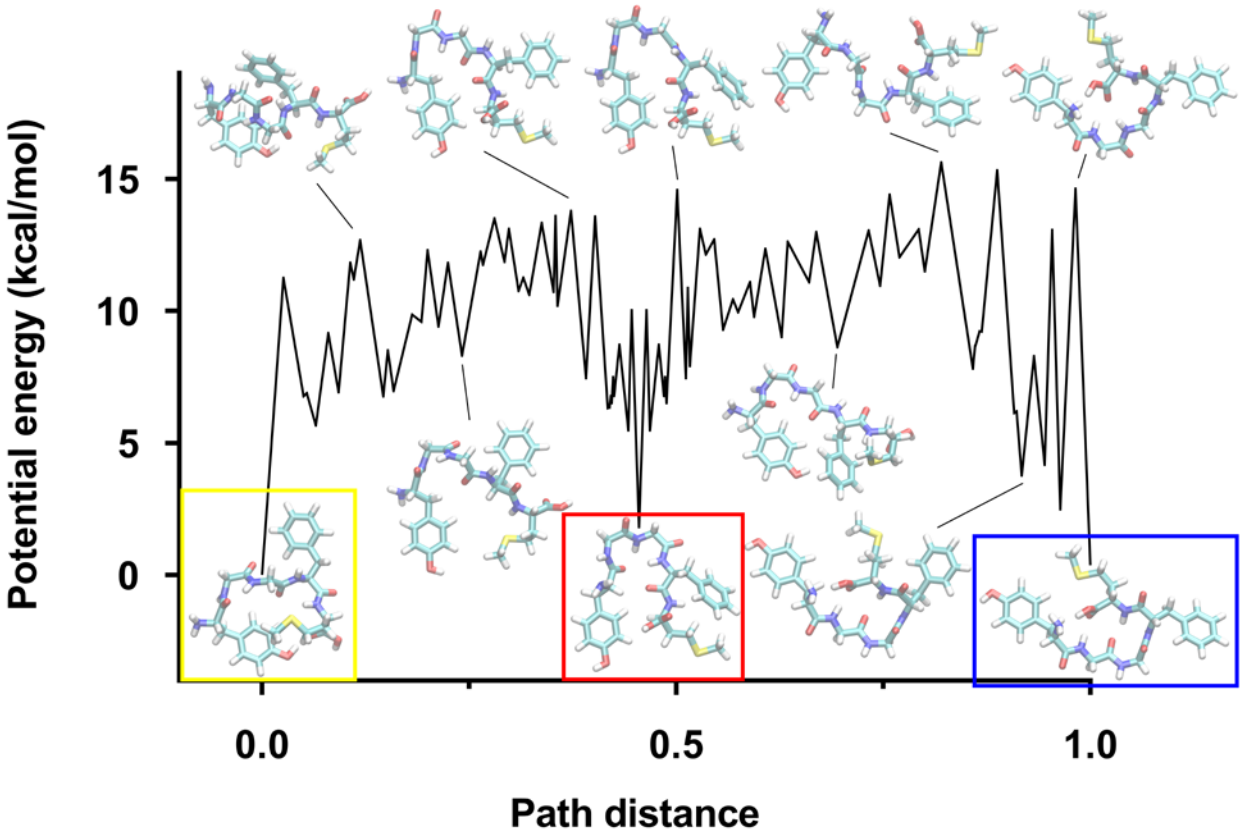

Figure 10. Variation of the PE along the transition pathway between the reference structures for cluster F (left, yellow), A (center, red) and C (right, blue). Selected structures along the pathway are shown as insets. The reference PE is that of structure F, the global PE minimum. The path distance 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.

Rate constants for the transition path as shown in Figure 10 at 300 K, 350 K and 400 K are displayed in Figure 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 the results for both REMD and free energy analysis of the ELM-derived stationary point database. However, the rate constants shown in Figure 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 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 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.

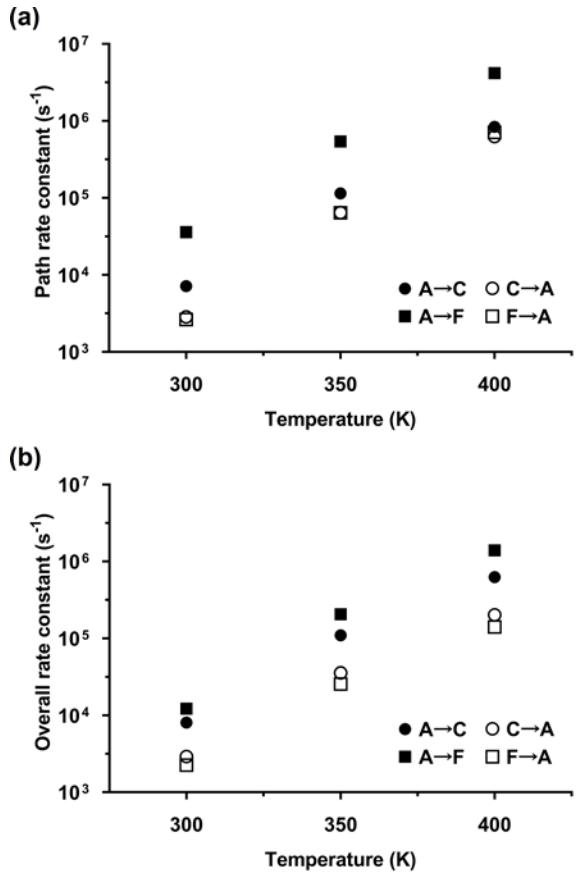

Figure 11. (a) Variation of the rate constants with temperature for the transition paths between structures F, A and C as shown in Figure 10, and (b) variation of the overall phenomenological rate constants between F, A and C with temperature.

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,[65] however, provide a promising means of addressing this.[47]

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.[62] 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.[21,66-67] 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.[19] 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,[2-6] the adsorption of multiple peptides at a surface is a topic of interest, and is the subject of ongoing work.

ACKNOWLEDGMENT

J.A.R. is grateful to the University of Adelaide for the receipt of an Australian Postgraduate Award. The support of the Australian Research Council Discovery Program (DP130101714) is gratefully acknowledged. Computational resources were provided by the Phoenix HPC service at the University of Adelaide.

REFERENCES

1. Shiba, K. In *Peptides and Peptide-based Biomaterials and their Biomedical Applications;* Sunna, A.; Care, A.; Bergquist, P. L., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp 1-20.

2. Care, A.; Bergquist, P. L.; Sunna, A. Solid-Binding Peptides: Smart Tools for Nanobiotechnology. *Trends Biotechnol.* **2015,** *33*, 259-268.

3. Slocik, J. M.; Naik, R. R. Sequenced Defined Biomolecules for Nanomaterial Synthesis, Functionalization, and Assembly. *Curr. Opin. Biotechnol.* **2017,** *46,* 7-13.

4. Walsh, T. R.; Knecht, M. R. Biointerface Structural Effects on the Properties and Applications of Bioinspired Peptide-Based Nanomaterials. *Chem. Rev.* **2017,** *117,* 12641-12704.

5. Pagel, M.; Beck-Sickinger, A. G. Multifunctional Biomaterial Coatings: Synthetic Challenges and Biological Activity. *Biol. Chem.* **2017,** *398,* 3-22.

6. Mas-Moruno, C. In *Peptides and Proteins as Biomaterials for Tissue Regeneration and Repair;* Barbosa, M. A.; Martins, M. C. L., Eds.; Elsevier: Duxford, U.K., 2018; pp 73-100.

7. Demir, B.; Yilmaz, T.; Guler, E.; Gumus, Z. P.; Akbulut, H.; Aldemir, E.; Coskunol, H.; Colak, D. G.; Cianga, I.; Yamada, S. Polypeptide with Electroactive Endgroups as Sensing Platform for the Abused Drug 'Methamphetamine' by Bioelectrochemical Method. *Talanta* **2016,** *161,* 789-796.

8. Liang, P.; Li, Q.; Wu, Z.; Jiang, J.-H.; Yu, R.-Q. Graphene Oxide–Peptide Nanoassembly as a General Approach for Monitoring the Activity of Histone Deacetylases. *Analyst* **2016,** *141,* 3989-3992.

9. Soum, C.; Rubio-Albenque, S.; Fery-Forgues, S.; Déléris, G.; Alouini, M.-A.; Berthelot, T. Supramolecular Peptide/Surface Assembly for Monitoring Proteinase Activity and Cancer Diagnosis. *ACS Appl. Mater. Interfaces* **2015,** *7,* 16967-16975.

10. Szweda, R.; Trzebicka, B.; Dworak, A.; Otulakowski, L.; Kosowski, D.; Hertlein, J.; Haladjova, E.; Rangelov, S.; Szweda, D. Smart Polymeric Nanocarriers of Met-enkephalin. *Biomacromolecules* **2016,** *17,* 2691-2700.

11. Docter, D.; Westmeier, D.; Markiewicz, M.; Stolte, S.; Knauer, S. K.; Stauber, R. H. The Nanoparticle Biomolecule Corona: Lessons Learned–Challenge Accepted? *Chem. Soc. Rev.* **2015,** *44,* 6094-6121.

12. Trindade, R.; Albrektsson, T.; Tengvall, P.; Wennerberg, A. Foreign Body Reaction to Biomaterials: On Mechanisms for Buildup and Breakdown of Osseointegration. *Clin. Implant Dent. Relat. Res.* **2016,** *18,* 192-203.

13. Bellucci, L.; Ardèvol, A.; Parrinello, M.; Lutz, H.; Lu, H.; Weidner, T.; Corni, S. The Interaction with Gold Suppresses Fiber-Like Conformations of the Amyloid β (16–22) Peptide. *Nanoscale* **2016,** *8,* 8737-8748.

14. Liu, L.; Li, Q.; Zhang, S.; Wang, X.; Hoffmann, S. V.; Li, J.; Liu, Z.; Besenbacher, F.; Dong, M. Identification of a Novel Parallel β-Strand Conformation within Molecular Monolayer of Amyloid Peptide. *Adv. Sci.* **2016,** *3,* 1500369.

15. Soltani, N.; Gholami, M. R. Increase in the β-Sheet Character of an Amyloidogenic Peptide upon Adsorption onto Gold and Silver Surfaces. *ChemPhysChem* **2017,** *18,* 526-536.

16. Heinz, H.; Ramezani-Dakhel, H. Simulations of Inorganic–Bioorganic Interfaces to Discover New Materials: Insights, Comparisons to Experiment, Challenges, and Opportunities. *Chem. Soc. Rev.* **2016,** *45,* 412-448.

17. Ozboyaci, M.; Kokh, D. B.; Corni, S.; Wade, R. C. Modeling and Simulation of Protein–Surface Interactions: Achievements and Challenges. *Q. Rev. Biophys.* **2016,** *49,* e4.

18. Ramakrishnan, S. K.; Zhu, J.; Gergely, C. Organic–Inorganic Interface Simulation for New Material Discoveries. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017,** *7,* e1277.

19. Walsh, T. R. Pathways to Structure–Property Relationships of Peptide–Materials Interfaces: Challenges in Predicting Molecular Structures. *Acc. Chem. Res.* **2017,** *50,* 1617-1624.

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

21. Wales, D. J. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses;* Cambridge University Press: Cambridge, U.K., 2003.

22. Wales, D. J.; Bogdan, T. V. Potential Energy and Free Energy Landscapes. *J. Phys. Chem. B* **2006,** *110,* 20765-20776.

23. Wales, D. J. Exploring Energy Landscapes. *Annu. Rev. Phys. Chem.* **2018,** *69,* 401-425.

24. Evans, D. A.; Wales, D. J. Free Energy Landscapes of Model Peptides and Proteins. *J. Chem. Phys.* **2003,** *118,* 3891-3897.

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

26. Wales, D. J.; Miller, M. A.; Walsh, T. R. Archetypal Energy Landscapes. *Nature* **1998,** *394,* 758-760.

27. Wales, D. J. Discrete Path Sampling. *Mol. Phys.* **2002,** *100,* 3285-3305.

28. Ross-Naylor, J. A.; Mijajlovic, M.; Zhang, H.; Biggs, M. J. Characterizing the Switching Transitions of an Adsorbed Peptide by Mapping the Potential Energy Surface. *J. Phys. Chem. B* **2017,** *121,* 11455-11464.

29. Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999,** *314,* 141-151.

30. Mijajlovic, M.; Penna, M. J.; Biggs, M. J. Free Energy of Adsorption of Proteins at Fluid/Solid Interfaces Using Molecular Simulation. In *Chemeca 2011: Engineering a Better World*, Engineers Australia: Sydney Hilton Hotel, NSW, Australia, 2011; pp 1221-1231.

31. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017,** *14,* 71-73.

32. Małolepsza, E.; Strodel, B.; Khalili, M.; Trygubenko, S.; Fejer, S. N.; Wales, D. J. Symmetrization of the AMBER and CHARMM Force Fields. *J. Comput. Chem.* **2010,** *31,* 1402-1409.

33. Braun, R.; Sarikaya, M.; Schulten, K. Genetically Engineered Gold-Binding Polypeptides: Structure Prediction and Molecular Dynamics. *J. Biomater. Sci. Polym. Ed.* **2002,** *13,* 747-757.

34. Steele, W. A. The Physical Interaction of Gases with Crystalline Solids: I. Gas-Solid Energies and Properties of Isolated Adsorbed Atoms. *Surf. Sci.* **1973,** *36,* 317-352.

35. Trucano, P.; Chen, R. Structure of Graphite by Neutron-Diffraction. *Nature* **1975,** *258,* 136-137.

36. Wales, D. J.; Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **1997,** *101,* 5111-5116.

37. Liu, D. C.; Nocedal, J. On the Limited Memory BFGS Method for Large Scale Optimization. *Math. Prog.* **1989,** *45,* 503-528.

38. Strodel, B.; Lee, J. W. L.; Whittleston, C. S.; Wales, D. J. Transmembrane Atructures for Alzheimer's Aβ1−42 Oligomers. *J. Am. Chem. Soc.* **2010,** *132,* 13300-13312.

39. Joseph, J. A.; Wales, D. J. Intrinsically Disordered Landscapes for Human CD4 Receptor Peptide. *J. Phys. Chem. B* **2018,** *122,* 11906-11921.

40. Munro, L. J.; Wales, D. J. Defect Migration in Crystalline Silicon. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999,** *59,* 3969-3980.

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

42. Evans, D. A.; Wales, D. J. Folding of the GB1 Hairpin Peptide from Discrete Path Sampling. *J. Chem. Phys.* **2004,** *121,* 1080-1090.

43. Carr, J. M.; Wales, D. J. Global Optimization and Folding Pathways of Selected α-Helical Proteins. *J. Chem. Phys.* **2005,** *123,* 234901.

44. Trygubenko, S. A.; Wales, D. J. A Doubly Nudged Elastic Band Method for Finding Transition States. *J. Chem. Phys.* **2004,** *120,* 2082-2094.

45. Stillinger, F. H.; Weber, T. A. Packing Structures and Transitions in Liquids and Solids. *Science* **1984,** *225,* 983-989.

46. Carr, J. M.; Wales, D. J. In *Latest Advances in Atomic Cluster Collisions: Structure and Dynamics from the Nuclear to the Biological Scale;* Connerade, J.-P.; Solov'yov, A., Eds.; Imperial College Press: London, U.K., 2008; pp 321-330.

47. Wales, D. J. Decoding Heat Capacity Features from the Energy Landscape. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2017,** *95,* 030105.

48. Trygubenko, S. A.; Wales, D. J. Kinetic Analysis of Discrete Path Sampling Stationary Point Databases. *Mol. Phys.* **2006,** *104,* 1497-1507.

49. Wales Group Home Page. https://www.ch.cam.ac.uk/group/wales/ (accessed September 29, 2018).

50. 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., et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009,** *30,* 1545-1614.

51. Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007,** *3,* 2312-2334.

52. Lyman, E.; Zuckerman, D. M. Ensemble-Based Convergence Analysis of Biomolecular Trajectories. *Biophys. J.* **2006,** *91,* 164-172.

53. Mijajlovic, M.; Penna, M. J.; Biggs, M. J. Free Energy of Adsorption for a Peptide at a Liquid/Solid Interface via Nonequilibrium Molecular Dynamics. *Langmuir* **2013,** *29,* 2919-2926.

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

55. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005,** *26,* 1781-1802.

56. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996,** *14*, 33-38.

57. Rathore, N.; Chopra, M.; de Pablo, J. J. Optimal Allocation of Replicas in Parallel Tempering Simulations. *J. Chem. Phys.* **2005,** *122*, 024111.

58. Abraham, M. J.; Gready, J. E. Ensuring Mixing Efficiency of Replica-Exchange Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2008,** *4,* 1119-1128.

59. Wright, L. B.; Walsh, T. R. Efficient Conformational Sampling of Peptides Adsorbed onto Inorganic Surfaces: Insights from a Quartz Binding Peptide. *Phys. Chem. Chem. Phys.* **2013,** *15*, 4715-4726.

60. Wales, D. J.; Head-Gordon, T. Evolution of the Potential Energy Landscape with Static Pulling Force for Two Model Proteins. *J. Phys. Chem. B* **2012,** *116,* 8394-8411.

61. Wales, D. J. Perspective: Insight into Reaction Coordinates and Dynamics from the Potential Energy Landscape. *J. Chem. Phys.* **2015,** *142*, 130901.

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

63. Jiang, S.; Rhykerd, C. L.; Gubbins, K. E. Layering, Freezing Transitions, Capillary Condensation and Diffusion of Methane in Slit Carbon Pores. *Mol. Phys.* **1993,** *79,* 373-391.

64. Marruecos, D. F.; Schwartz, D. K.; Kaar, J. L. Impact of Surface Interactions on Protein Conformation. *Curr. Opin. Colloid Interface Sci.* **2018,** *38,* 45-55.

65. Wales, D. J. Surveying a Complex Potential Energy Landscape: Overcoming Broken Ergodicity Using Basin-Sampling. *Chem. Phys. Lett.* **2013,** *584,* 1-9.

66. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications;* 2nd ed.; Academic Press: San Diego, CA, U.S.A., 2002.

67. Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002,** *116,* 9058-9067.

TOC GRAPHIC