Energy landscapes for clusters of hexapeptides

Special Collection: Monte Carlo methods, 70 years after Metropolis et al. (1953)

Nicy ; John W. R. Morgan ; David J. Wales

J. Chem. Phys. 161, 054112 (2024)
https://doi.org/10.1063/5.0220652
Energy landscapes for clusters of hexapeptides

Cite as: J. Chem. Phys. 161, 054112 (2024); doi: 10.1063/5.0220652
Submitted: 27 May 2024 • Accepted: 17 July 2024 • Published Online: 2 August 2024

Nicy, John W. R. Morgan, and David J. Wales

AFFILIATIONS
Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Note: This paper is part of the JCP Special Topic on Monte Carlo methods, 70 years after Metropolis et al. (1953).

ABSTRACT
We present the results for energy landscapes of hexapeptides obtained using interfaces to the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) program. We have used basin-hopping global optimization and discrete path sampling to explore the landscapes of hexapeptide monomers, dimers, and oligomers containing 10, 100, and 200 monomers modeled using a residue-level coarse-grained potential, Mpipi, implemented in LAMMPS. We find that the dimers of peptides containing amino acid residues that are better at promoting phase separation, such as tyrosine and arginine, have melting peaks at higher temperature in their heat capacity compared to phenylalanine and lysine, respectively. This observation correlates with previous work on the same uncapped hexapeptide monomers modeled using atomistic potential. For oligomers, we compare the variation in monomer conformations with radial distance and observe trends for selected angles calculated for each monomer. The LAMMPS interfaces to the GMIN and OPTIM programs for landscape exploration offer new opportunities to investigate larger systems and provide access to the coarse-grained potentials implemented within LAMMPS.

I. INTRODUCTION

The potential energy landscape (PEL) of a system defines how the potential energy changes as a function of the degrees of freedom. For a system of $N$ atoms, the PEL lies in a $3N+1$ dimensional space. Molecular properties are encoded in the underlying landscape, which forms the basis for structure prediction, calculation of thermodynamics, and analysis of global dynamics in terms of barriers and rates. To explore the multidimensional landscape, we coarse-grain in terms of stationary points, specifically local minima, which define mechanically stable states, and first-order saddle points, which define transition states. Steepest-descent pathways connect the minima and transition states, and a discrete path is defined using triples of minimum-transition-state-minimum. The stationary points and pathways are calculated using geometry optimization techniques, providing a complementary approach to conventional Monte Carlo and molecular dynamics simulations. Thermodynamic and kinetic properties are obtained by post-processing the kinetic transition network defined by the database of stationary points using standard methods of statistical mechanics and unimolecular rate theory.

We employ three main programs to implement the computational side of this energy landscapes approach, namely, GMIN for global optimization and structure prediction; OPTIM to characterize transition states and pathways; and PATHSAMPLE to manage parallel OPTIM searches, expand stationary point databases using discrete path sampling, and perform post-processing to calculate thermodynamic and kinetic properties. This framework has been applied to a wide variety of systems in molecular science, as reviewed elsewhere, and it can be extended to explore more abstract solution landscapes, for example, in machine learning, electronic structure, and quantum computing.

Usually, the first step in exploring a landscape is to find the global minimum and other low-lying states. Various approaches have been employed to identify low-lying structures separated by large barriers, including conformational flooding, paving and deforming landscapes, and using optimal Tsallis-based weights to maximize acceptance of replica exchanges. Here, we employ basin-hopping global optimization, which has proved to be a powerful tool for such surveys, with a wide range of applications to atomic and molecular clusters, condensed matter, and biomolecules, as well as the abstract solution landscapes mentioned above. The basic algorithm generates a chain of local minima by proposing a
change in the current minimum, minimizing the objective function of interest, and accepting or rejecting the new minimum. Hence basin-hopping shares some features of the Monte Carlo approach,23,26,27 celebrated in this anniversary issue of the journal. However, there is an important difference between basin-hopping and Monte Carlo schemes that aim to sample thermodynamic properties because detailed balance is not required for global optimization. In particular, it is more efficient to reset the coordinates to the values for the current minimum before proposing a move, which breaks the detailed balance. Basin-hopping moves can be orders of magnitude larger than the coordinate perturbations usually employed in thermodynamic scaling because the algorithm is exploring the landscape of local minima and the accept/reject condition is applied after minimization. The coordinate steps can, therefore, generate very high energy initial configurations because the effective potential is the value after minimization. This freedom to propose large moves through the configuration space enables low-lying minima to be located rapidly.

Enhanced sampling techniques, such as parallel tempering and replica exchange, are not competitive for global optimization. However, hybrid methods that combine basin-hopping to overcome broken ergodicity with enhanced sampling schemes for higher energy regions of the landscape have been used to calculate thermodynamic properties efficiently via the total energy density of states. Previous work, e.g., employed basin-hopping in combination with both Wang-Landau and parallel tempering. Knowledge of the low-lying minima has also been used to solve the lockout problem that can cause problems in nested sampling.

Numerous variants of basin-hopping have been described, which all share the key principle of moves between local minima. Lamarckian genetic algorithms, where populations of local minima are considered, could be described in terms of basin-hopping with genetic moves. Since the accept/reject procedure is applied to local minima, atoms can pass through each other in moves proposed for basin-hopping. In fact, analysis of the transformed landscape shows that occupation probabilities for competing low energy structures may be broadened as a function of temperature, which facilitates sampling of alternative morphologies.

The usual Metropolis accept/reject scheme is often used in basin-hopping, but threshold acceptance and Tsallis statistics have also been tested. The most efficient move scheme is likely to be system-dependent, and we discuss the methods used for hexapeptide clusters in the present contribution in Sec. II. For atomic clusters, it is helpful to pay special attention to the treatment of surface atoms and defects, incorporate memory in the step-taking, and consider symmetrized perturbations. It is also possible to sample based on a local free energy or a grand potential, and parallel basin-hopping runs that exchange replicas between different temperatures have been employed to treat oligomers of the Aβ12 peptide. Generalized basin-hopping schemes treat optimization in continuous and discrete metric spaces and have been used to explore the landscape for nanoalloy clusters, mutant peptides, and electronic structures.

An important consideration for exploration of the landscape using geometry optimization is the scaling with system size, which depends on the computational cost of the energy and gradient, as well as the convergence of various algorithms to the target stationary points. Hence, to treat large systems, it is helpful to use coarse-grained potentials. In the present contribution, we describe the results obtained by interfacing our software for landscape exploration with the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS). LAMMPS has implementations for many interatomic, many-body, machine learning, and coarse-grained potentials, and it offers the flexibility to develop and extend these potentials for molecular modeling.

As an initial application, we consider monomers, dimers, and clusters containing 10, 100, and 200 hexapeptides modeled using the Mpipi potential implemented in LAMMPS. Mpipi is a multiscale residue-level coarse-grained potential, which has been used to model liquid–liquid phase separation (LLPS) of proteins. The phase separation of proteins and nucleic acids enables cells to form membraneless organelles. These biomolecular condensates exhibit both a high density phase and a low density phase, which can be formed via density phase transition or phase separation coupled to percolation. Intrinsic disorder proteins, encoding aromatic–aromatic and cation–aromatic interactions commonly exhibit such phase separation behavior. In particular, systems containing tyrosine are more likely to separate better than counterparts containing phenylalanine, and arginine is better than lysine at promoting phase separation. As it is difficult to experimentally investigate the conformations of such intrinsically disordered proteins, various computational simulations using atomistic and coarse-grained potentials have been performed to gain greater insight into how the sequence of amino acids encodes the phase separation behavior.

A recent exploration of energy landscapes for hexapeptide monomers modeled using an atomistic potential revealed that the peptides encoding interactions promoting phase separation often have more low-temperature heat capacity features and more frustrated energy landscapes for capped hexapeptides. For uncapped hexapeptides, the highest temperature heat capacity peak was observed to appear at higher temperature for the peptides that contain amino acids that are better promoters of phase separation. However, phase separation is clearly a collective property, and further investigation of how phase separation propensity is encoded in the landscape is required for larger oligomers. Atomistic simulations then become computationally expensive, and coarse-grained potentials are often employed in this regime. The current contribution aims to complement previous work by exploring the energy landscape of larger systems for the same hexapeptides, now represented using a coarse-grained potential. Although the main aim is to study oligomers containing hundreds of monomers, we begin by exploring the landscape of monomers, dimers, and trimers to benchmark the parameters required for the basin-hopping moves and geometry optimization.

One interesting result (Sec. III) is the observation of correlation between some of the features in the heat capacity of monomers modeled using the atomistic potential and dimers modeled using the coarse-grained residue-level Mpipi representation. We also define geometric order parameters in the form of two different angles for each monomer in the oligomer. We employ these geometric parameters to investigate variation of the monomer conformations with radial distance in the low energy oligomer structures. These predictions may be testable in future experiments and
could help understand the underlying basis for phase separation propensity.

II. METHODS

The current study builds on the previous work\[102] in which atomistic energy landscapes of monomers were explored for various hexapeptide sequences. Here, we explore the energy landscapes of oligomers of the same hexapeptides modeled using the Mpiipi potential. Interfaces to LAMMPS were implemented in our GMIN and OPTIM programs, which were then used to explore the energy landscapes of these hexapeptides. The technical details for building the interface can be found elsewhere.\[103]

A. Peptide model using Mpiipi in LAMMPS

The hexapeptide monomers, dimers, and oligomers containing 10, 100, and 200 monomers are modeled as intrinsically disordered proteins using a coarse-grained residue-level Mpiipi potential implemented using standalone LAMMPS. Various studies using Mpiipi have shown that it can be used to obtain quantitative estimates of single molecule properties, such as radius of gyration, and collective behavior properties, such as phase diagrams.\[60,104–106] In the Mpiipi model, the potential energy is defined using a sum of harmonic bond energies, Debye–Hückel, and short-ranged Wang–Frenkel terms,\[107] which have a short-range excluded volume repulsion and a long-ranged attraction that gradually decays to zero. This representation assumes that the pairwise amino acid interactions are sufficient to model the phase separation propensity,\[26,99,100,108,109] and that the aromatic-aromatic (π–π)\[75–83] and cation–aromatic (cation–π)\[81,84–86] interactions predominantly drive the transition. The model neglects higher order energy terms, which could be important to account for cooperative interactions in the crowded intracellular environment, and the interaction strengths do not include explicit temperature dependence.\[60]

B. Global optimization by basin-hopping using LMPGMIN

The LAMMPS-GMIN (LMPGMIN) interface was used for global optimization of hexapeptides using the basin-hopping algorithm.\[22–24] For monomers, only Cartesian steps were employed to propose structural perturbations. However, for dimers, decamers, and larger oligomers, rigid body moves were also used after every few Cartesian steps. The rigid body moves include expansion, rotation, translation, and compression in that sequence. Translation is performed in Cartesian coordinates, and rotation is performed using angle-axis coordinates.\[110–112] Before implementing translation and rotation, the system needs to be expanded to avoid clashes between monomers, and after the moves are performed, the system needs to be compressed to prevent dissociation of monomers (Fig. 1), followed by full relaxation with the true interaction potential.

For monomers, a total of 10 000 basin-hopping steps were performed with the temperature parameter in the Metropolis accept/reject condition set to 300 K. Cartesian perturbations with a uniform distribution and a maximum value of 1.2 Å were used for every coordinate. The large maximum step size of 1.2 Å was found to be optimal after testing different values in the range of 0.1–3.0 Å in increments of 0.1 Å. A root-mean-square convergence criterion of 10−6 kcal mol−1 Å−1 was found to be optimal after testing different values in the range of 0.1–3.0 Å in increments of 0.1 Å. A root-mean-square convergence criterion of 10−6 kcal mol−1 Å−1 for the gradient was employed for the gradient. The amount of rotation and translation employed was roughly independent of system size. In general, small rotation moves work better than larger moves. The optimal Cartesian step size decreases with the size of the system, as judged by the lowest minima obtained in benchmarking runs. However, it may be useful to start with a large maximum step size and, as the run progresses, decrease the step size to obtain the lowest energy structures. Similarly, the temperature is another very important parameter that needs to be chosen.
carefully to obtain low energy structures reliably. For larger oligomers, a lower temperature for the accept/reject test was found to give lower lying structures. During these calculations, we also compared basin-hopping runs using Cartesian steps alone and with a combination of Cartesian steps and rigid body moves. Overall, the rigid body moves proved to be very helpful in efficiently obtaining lower energy structures.

C. Discrete path sampling using LMPOPTIM

For monomers and dimers, we used discrete path sampling with OPTIM interfaced to LAMMPS (LMPOPTIM) to obtain pathways between the local minima and the global minimum found using basin-hopping. A pathway is explored using the doubly nudged elastic-band (DNEB) algorithm to generate candidate transition states. Hybrid eigenvector-following (HEF) is then employed to refine the candidate transition states, and approximate steepest-descent paths are characterized with the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) approach to find the local minima connected by the transition state. The missing connection algorithm was used to choose additional pairs of minima for further connection attempts using the same series of algorithms until a fully connected pathway was obtained. The network of stationary points was expanded using a scheme that attempts connections between low energy local minima in the connected set and local minima in the unconnected set that are close in terms of Euclidean distance.

D. Disconnectivity graphs

Disconnectivity graphs are used to visualize the organization of the energy landscape. These graphs preserve the information about the highest barrier that needs to be overcome for the interconversion of different minima. The vertical axis represents the potential or free energy, and the nodes along the horizontal axis represent superbasins. The individual branches represent the local minima and terminate at the potential energy of local minima. Local minima lying in the same superbasin (the branches connected to the same node) can interconvert via a barrier below the threshold energy on the vertical axis. The local minima making the largest contributions to the heat capacity of dimers, structures in which any monomers were dissociated are expected for the residue-level coarse-grained potential. Only two secondary structure of pentapeptides can be context-dependent; we have three reasons for studying hexapeptides. First, the secondary structure of pentapeptides can be context-dependent; second, the amino acid sequence patterns for peptides longer than eight amino acids are less reliable; and third, there are several hexapeptides that form amyloids, and some of these fibrils attenuate neuroinflammation. Overall, it may be useful to compare the thermodynamic properties of hexapeptides that can form amyloids and the hexapeptides that contain phase separation promoting interactions. Having chosen the length of peptides and amino acids, we arranged them following the “stickers-and-spacers” framework. Previous studies have shown that dipeptides may be used as “stickers,” and for the present work, we chose FF, YY, KY, YR, and YK as dipeptide stickers, separated by a flexible GG spacer to prevent the formation of ordered structures. We explored the energy landscapes of monomers and dimers of AAGGAA, FFGGFF, YYGGYY, YYGGY, YKGYK, YRGGRY, and YKGGKY using both basin-hopping and discrete path sampling. We also performed basin-hopping global optimization using a combination of Cartesian steps and rigid body moves to obtain low energy structures for oligomers containing 10, 100, and 200 monomers for some of these hexapeptides, including AAGGAA, FFGGFF, and YYGGYY. Detailed analysis of the lowest energy structure for these oligomers reveals specific trends in the conformations of monomers as the radial distance increases. Some of these trends agree with a previous report exploring the conformations of peptides in dense phase, dilute phase, and the interface between them.

A. Monomer and dimer energy landscapes

The landscapes of hexapeptide monomers are simple, as expected for the residue-level coarse-grained potential. Only two local minima were located for FFGGFF, YYGGYY, and KYGGYK.
FIG. 2. Disconnectivity graphs for hexapeptide monomers modeled using the Mpipi potential. The scale bar represents 0.1 kcal mol\(^{-1}\). The end-to-end distance between the first and sixth amino acids in the hexapeptide is labeled here. Reproduced from Nicy, Ph.D thesis, Apollo - University of Cambridge Repository, 2023 with the permission of AIP Publishing.

monomers and a single local minimum for RYGGYR and YRGGRY. Several local minima were found for the control peptide AAGGAA (Fig. 2). For all the hexapeptides, the end-to-end distance, defined as the distance between the first and sixth amino acid, is larger in the global minimum compared to the other local minima. In particular, the end-to-end distance for the global minimum of YYGGYY is slightly larger compared to FFGGFF. Here, we are not considering the permutational isomers as separate local minima.

The energy landscapes for the dimers reveal similar monomer structures as for the global minimum of FFGGFF, YYGGYY, RYGGYR, KYGGYK, YRGGRY, and YKGGKY (Figs. 3 and 4). The global minimum for AAGGAA contains S-shaped monomers, whereas the global minimum of other peptides contains two C-shaped monomers oriented in approximately perpendicular planes (Figs. 3 and 4). The comparison of end-to-end distance of monomers in the global minimum reveals that both the monomers are slightly more expanded for YYGGYY compared to FFGGFF and for RYGGYR compared to KYGGYK. Even for the global minimum of YRGGRY and YKGGKY dimers, one of the monomers of YRGGRY is more expanded compared to both the monomers in YKGGKY. This expanded structure for the monomers within the dimers for peptides containing Y and R compared to F and K, respectively, may be correlated with the better phase separation propensity encoded by Y and R.

The heat capacity analysis for the dimers reveals that the dimers containing residues that are better promoters of phase separation have the highest temperature heat capacity (\(C_V\)) peak at higher temperatures (Fig. 5). The temperature at which the highest \(C_V\) peak (melting peak) is obtained lies in the order of AAGGAA < FFGGFF < YYGGYY (Table I). Clearly, this result correlates with the fact that alanine does not act as a sticker and does not promote phase separation, whereas tyrosine is better at promoting phase separation compared to phenylalanine. A similar trend is observed for dimers featuring cation–aromatic interactions, where the highest \(C_V\) peak lies at higher temperature for RYGGYR compared to KYGGYK. For the YRGGRY dimer and uncapped peptide monomer, we observe a peak at low temperature, and instead of a melting peak, we see an inflection point at high temperature (Table I). Hence, if we compare the highest temperature peaks of YRGGRY and YKGGKY, the peak for YKGGKY lies at a higher temperature. However, if we take into account the higher temperature inflection point observed for YRGGRY, it lies at a higher temperature than the melting peak of YKGGKY for both the dimer and uncapped peptide monomers. Interestingly, this trend in the melting peak of dimers modeled using a physics-based coarse-grained potential correlates with the melting peak of uncapped monomers for the same hexapeptides modeled using an atomistic potential. In the previous report,\(^{102}\) we did not mention that the high temperature heat capacity feature (the melting peak) for the uncapped peptides exhibits a trend. In fact, the peak occurs at higher temperatures for uncapped peptides (and most of the capped peptides) encoding interactions that are better at promoting phase separation. All the peptides considered in the current work are consistent with this trend. The precise position of the melting peak in \(C_V\) depends on the
FIG. 3. Disconnectivity graphs for hexapeptide dimers modeled using the Mpipi potential: (top) AAGGAA, (bottom) FFGGFF, and YYGGYY. The scale bar represents 0.1 kcal mol$^{-1}$. Reproduced from Nicy, Ph.D thesis, Apollo - University of Cambridge Repository, 2023 with the permission of AIP Publishing.
The convergence of the database in terms of the number of minima sampled. We did not attempt to enumerate all the high energy minima for the atomistic potential in that study. The harmonic approximation used to estimate heat capacity involves a systematic error that will usually shift the melting peak to a higher temperature. The higher energy minima populated at higher temperature tend to have more low frequency normal modes, where anharmonicity means that the density of states is underestimated in the harmonic approximation. Incomplete sampling of these high energy minima also leads to underestimation of the landscape entropy, and therefore, the melting peak shifts to a higher temperature. These effects can be quantified using full anharmonic sampling, which entails a significant computational cost. In the present work, it is sufficient to analyze $C_V$ in the harmonic approximation to deduce features of the underlying landscape. For the hexapeptide monomers modeled using the atomistic potential, and for the dimers modeled using the residue-level potential, we believe that the position of the peak is reasonably well converged for the current databases (Fig. 5) in terms of the landscape entropy. The remaining systematic error corresponds to anharmonic effects.

FIG. 4. Disconnectivity graphs for hexapeptide dimers modeled using the Mpipi potential. The scale bar represents 0.1 kcal mol$^{-1}$. Reproduced from Nicy, Ph.D thesis, Apollo - University of Cambridge Repository, 2023 with the permission of AIP Publishing.
Heat capacity, $C_V(T)$, for hexapeptide dimers modeled using the Mpi-p potential. Adapted from the Ph.D thesis of Nicy. The dashed and dotted lines represent $C_V$ obtained using the first 50% and 75% of the minima in the order that they were sampled, respectively. We employ energy units for the temperature on the horizontal axis, which is convenient for calculations because it allows us to use dimensionless quantities everywhere in the calculations. We therefore express $RT$ in kcal mol$^{-1}$.

TABLE I. Melting peaks in $C_V$ observed for different hexapeptide dimers modeled using the Mpi-p potential in the current work and monomers modeled using FF99IDPs and FF19SB atomistic force fields in previous work. For the AAGGAA dimer, we report the temperature (in energy units) at which the highest value of $C_V$ is observed. For YRGGRY, we report the positions of low temperature $C_V$ peaks (represented using an asterisk, "*") in the table below. We note that for YRGGRY, an inflection point occurs at 0.133, 0.877, and 0.674 kcal mol$^{-1}$ for the dimer modeled using Mpi-p and uncapped peptide monomers modeled using FF99IDPs and FF19SB force fields, respectively.

<table>
<thead>
<tr>
<th>Hexapeptide</th>
<th>Dimer</th>
<th>FF99IDPs</th>
<th>FF19SB</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAGGAA</td>
<td>Mpipi</td>
<td>0.01</td>
<td>0.354</td>
</tr>
<tr>
<td></td>
<td>Capped</td>
<td>0.256</td>
<td>0.494</td>
</tr>
<tr>
<td></td>
<td>Uncapped</td>
<td>0.352</td>
<td></td>
</tr>
<tr>
<td>FFGGFF</td>
<td>0.042</td>
<td>0.622</td>
<td>0.444</td>
</tr>
<tr>
<td></td>
<td>0.695</td>
<td>0.536</td>
<td></td>
</tr>
<tr>
<td>YYGGYY</td>
<td>0.062</td>
<td>0.630</td>
<td>0.607</td>
</tr>
<tr>
<td></td>
<td>0.721</td>
<td>0.702</td>
<td></td>
</tr>
<tr>
<td>KYGGYK</td>
<td>0.05</td>
<td>0.734</td>
<td>0.628</td>
</tr>
<tr>
<td></td>
<td>0.724</td>
<td>0.611</td>
<td></td>
</tr>
<tr>
<td>RYGGYR</td>
<td>0.157</td>
<td>0.836</td>
<td>0.693</td>
</tr>
<tr>
<td></td>
<td>0.928</td>
<td>0.670</td>
<td></td>
</tr>
<tr>
<td>YKGGKY</td>
<td>0.125</td>
<td>0.862</td>
<td>0.694</td>
</tr>
<tr>
<td></td>
<td>1.078</td>
<td>0.299</td>
<td></td>
</tr>
<tr>
<td>YRGGRY</td>
<td>0.05*</td>
<td>0.389</td>
<td>0.526*</td>
</tr>
<tr>
<td></td>
<td>0.445</td>
<td>0.256*</td>
<td></td>
</tr>
</tbody>
</table>

Melting peak in $C_V$ (kcal mol$^{-1}$)

B. Oligomer energy landscapes

Basin-hopping global optimization with rigid body moves and Cartesian perturbations was used to obtain candidates for the lowest energy structure in oligomers containing 10, 100, and 200 monomers of the same hexapeptide sequence. Clearly, the lowest energy structures of oligomers of AAGGAA, FFGGFF, and YYGGYY exhibit heterogeneous (compact and extended) conformations for the component monomers (Fig. 6). In previous work, trends in conformational heterogeneity in different phases and interfaces were reported, with intermolecular interactions favored in the dense phase and intramolecular interactions in the dilute phase. Although it is not possible to observe the dense and dilute phases for clusters in the size range considered here, we did note that oligomer structures containing dissociated
monomers exhibit intermolecular interactions for the oligomer and intramolecular interaction for dissociated monomers.

We use geometric order parameters to search for trends in monomer conformations as a function of radial distance from the center of the oligomer. In particular, we define the end-to-end distance of the hexapeptide monomer as the separation between the first and sixth amino acid and the radial distance as the separation between the center of geometry of the oligomer and the center of geometry of each monomer. The angle \( \theta \) is defined similarly to previous work.\(^{149} \) Here, \( \theta \) is the angle between the vectors pointing from the first or last amino acid (whichever is nearer to the center of geometry of oligomer) of the monomer to the center of geometry of the oligomer and to the first or last amino acid of the monomer, respectively (Fig. 7). We also define another angle, \( \alpha \), between the vectors pointing from the center of geometry of the oligomer to the first amino acid and sixth amino acid of each monomer (Fig. 7).

We observe that the end-to-end distance of the monomers is not correlated with the radial distance (Fig. 8). Although we might have expected to see monomers with smaller end-to-end distance at smaller radial distances and monomers with larger end-to-end distance at larger radial distance, there was no such trend visible. Instead, we observe some extended monomer conformations at small distances and some compact monomer conformations at large distances. Similar to previous work,\(^{149} \) we observe a smaller value for \( \cos^2 \theta \) at large radial distances for the largest sized oligomer considered here, containing 200 monomers (Fig. 9). Interestingly, the angle \( \alpha \) does exhibit a clear trend for oligomers containing 100 and 200 monomers. \( \alpha \) decreases as the radial distance increases (Fig. 10), whereas \( \alpha \) increases as the end-to-end distance increases for a majority of monomers (Fig. 11). However, the end-to-end distance is not correlated with radial distance. We also observe that the oligomers containing Y or F are more expanded compared to A. This difference is evident from the comparison of the largest radial distance for monomers in the oligomers of the same size (Fig. 8–10).

![Schematic diagram to define angles](image)

**FIG. 7.** Schematic diagram to define angles. Here, “C” represents the center of geometry of the oligomer. The amino acids in the hexapeptide monomer are numbered from 1 to 6. (a) \( \theta \) is defined as the angle between the vectors pointing from the first amino acid (or last amino acid, whichever is nearer the center of geometry of the oligomer) of a monomer to the center of geometry of the oligomer and the last amino acid (or first amino acid). (b) \( \alpha \) is defined as the angle between the vectors pointing from the center of geometry to the first and last amino acid of the hexapeptide.

![Correlation between extension (or expansion) of monomers and distance of the monomer from the center of geometry](image)

**FIG. 8.** Correlation between extension (or expansion) of monomers and distance of the monomer from the center of geometry. Here, the scatter plots are overlaid with bin averages for the variable along the horizontal axis. The legends indicate the sequence of hexapeptide and the number of monomers in the oligomer.
FIG. 9. Correlation between $\cos^2 \theta$ defined for each monomer and the distance of the monomer from the center of geometry of oligomer. Here, the scatter plots are overlaid with bin averages for the variable along the horizontal axis.

FIG. 10. Correlation between the angle $\alpha$ defined for each monomer and the distance of the monomer from the center of geometry of oligomer. Here, the scatter plots are overlaid with bin averages for the variable along the horizontal axis.
IV. CONCLUSIONS

The LAMMPS-GMIN and LAMMPS-OPTIM interfaces provide useful tools to explore the energy landscapes of large systems using the coarse-grained potentials implemented within LAMMPS. Rigid body moves, such as expansion, rotation, and translation, coupled with subsequent compression and full relaxation, proved helpful in obtaining low energy structures for all the clusters. The energy landscape framework can be used to estimate the heat capacity as a function of temperature using the harmonic superposition approximation. Features in the heat capacity can provide insight into the physicochemical properties. In particular, we find that the interactions promoting phase separation may exhibit a higher temperature melting peak in the heat capacity, both for monomers modeled using atomistic potentials in previous work and dimers modeled using the coarse-grained potential in the current work. Larger oligomers exhibit conformational heterogeneity that agrees with previous work on larger systems explored using a different method. In the future, we can also use the LAMMPS-GMIN interface to train and test machine learning potentials by minimizing the cost function and obtaining alternative fits to training data for a cost function, as defined by the local minima in the energy landscape.

ACKNOWLEDGMENTS

The authors thank Professor Jerelle A. Joseph and Dr. Aleks Reinhardt for help with implementing the Mpiipi potential. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) (D.J.W., Grant No. EP/N035003/1); the Allen, Meek and Read Cambridge International Scholarship; the Santander fund, St Edmund’s College, University of Cambridge; and the Trinity-Henry Barlow Honorary Award (Nicy).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Nicy: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Investigation (equal); Methodology (equal); Software (supporting); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). John W. R. Morgan: Methodology (equal); Software (supporting); Writing – review & editing (equal). David J. Wales: Formal analysis (supporting); Funding acquisition (lead); Investigation (equal); Methodology (equal); Project administration (lead); Resources (lead); Software (lead); Supervision (lead); Writing – review & editing (equal).

DATA AVAILABILITY

The data that supports the findings of this study are openly available in Apollo - University of Cambridge Repository at https://doi.org/10.17863/CAM.101792, Ref. 154. A tutorial

FIG. 11. Correlation between the angle \( \alpha \) and the end-to-end distance defined for each monomer. Here, the scatter plots are overlaid with bin averages for the variable along the horizontal axis.
for creating the database, examples of input files, and coordinates of lowest energy structure of oligomers is available on https://github.com/nicy-nicy/LAMMPS_Mpipi_energy_landscape.

REFERENCES


