

RESEARCH ARTICLE | JUNE 20 2023

Graphic contrastive learning analyses of discontinuous molecular dynamics simulations: Study of protein folding upon adsorption

Special Collection: [Accelerate Materials Discovery and Phenomena](#)

Size Zheng  ; Yong Wei  ; Yuewei Lin  ; Tao Wei  



Appl. Phys. Lett. 122, 253701 (2023)

<https://doi.org/10.1063/5.0157933>



View
Online



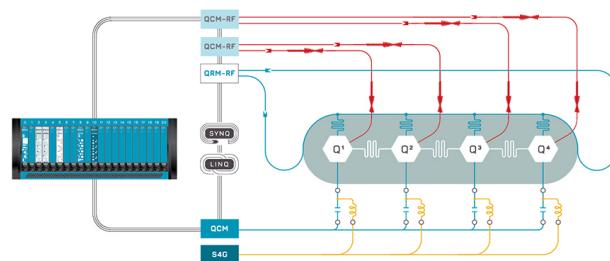
Export
Citation

CrossMark



Integrates all
Instrumentation + Software
for Control and Readout of

Superconducting Qubits
NV-Centers
Spin Qubits



Superconducting Qubit Setup

[find out more >](#)

Graphic contrastive learning analyses of discontinuous molecular dynamics simulations: Study of protein folding upon adsorption

Cite as: Appl. Phys. Lett. **122**, 253701 (2023); doi: [10.1063/5.0157933](https://doi.org/10.1063/5.0157933)

Submitted: 12 May 2023 · Accepted: 1 June 2023 ·

Published Online: 20 June 2023



View Online



Export Citation



CrossMark

Size Zheng,¹  Yong Wei,²  Yuewei Lin,³  and Tao Wei^{4,a)} 

AFFILIATIONS

¹College of Materials and Chemistry & Chemical Engineering, Chengdu University of Technology, Chengdu, Sichuan 610059, People's Republic of China

²Department of Computer Science, High Point University, High Point, North Carolina 27268, USA

³Computational Science Initiative, Brookhaven National Laboratory, Upton, New York 11973, USA

⁴Department of Chemical Engineering, Howard University, Washington, D.C. 20059, USA

Note: This paper is part of the APL Special Collection on Accelerate Materials Discovery and Phenomena.

^{a)} Author to whom correspondence should be addressed: tao.wei@howard.edu

ABSTRACT

A comprehensive understanding of the interfacial behaviors of biomolecules holds great significance in the development of biomaterials and biosensing technologies. In this work, we used discontinuous molecular dynamics (DMD) simulations and graphic contrastive learning analysis to study the adsorption of ubiquitin protein on a graphene surface. Our high-throughput DMD simulations can explore the whole protein adsorption process including the protein structural evolution with sufficient accuracy. Contrastive learning was employed to train a protein contact map feature extractor aiming at generating contact map feature vectors. Subsequently, these features were grouped using the k-means clustering algorithm to identify the protein structural transition stages throughout the adsorption process. The machine learning analysis can illustrate the dynamics of protein structural changes, including the pathway and the rate-limiting step. Our study indicated that the protein-graphene surface hydrophobic interactions and the π - π stacking were crucial to the seven-stage adsorption process. Upon adsorption, the secondary structure and tertiary structure of ubiquitin disintegrated. The unfolding stages obtained by contrastive learning-based algorithm were not only consistent with the detailed analyses of protein structures but also provided more hidden information about the transition states and pathway of protein adsorption process and structural dynamics. Our combination of efficient DMD simulations and machine learning analysis could be a valuable approach to studying the interfacial behaviors of biomolecules.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0157933>

Protein interfacial behavior, including conformational changes and orientation, is critical in various applications, such as biomaterials,¹ antibacterial/antifouling coatings,^{2–4} industrial catalysts,⁵ biosensors,^{6–9} biofuel cells,¹⁰ polymer membrane,¹¹ and bioremediation.^{12,13} Protein adsorption occurs at the water-surface interface over a wide time range, spanning from nanoseconds to seconds.^{8,14–16} It is difficult to probe the dynamic process from the atomistic and molecular scales by experiments alone, particularly the initial stage of adsorption.¹⁷ Molecular dynamics (MD) simulations at both atomistic^{3,4,6,11,18–21} and coarse-grained (CG) scales^{8,9,22} have been widely used to study protein interfacial behaviors. However, MD simulations with explicit solvent are computationally expensive, making it impractical to present the entire process.

To accurately model protein adsorption and maintain computational efficiency, we incorporated a Gō-like model^{23–25} for protein-surface interactions into discontinuous molecular dynamics (DMD) simulations,^{26,27} which can efficiently operate at both the atomistic and CG scales. DMD, originally proposed by Alder and Wainwright,²⁸ has been further developed to study biomolecular systems.^{29–33} Unlike MD simulations driven by time,^{34,35} DMD simulations are event-driven, allowing for adaptive and larger time steps and up to 100 times greater computational efficiency.³⁶ We employed DMD to study the folding, aggregation, and adsorption of peptides and proteins.^{32,37–39}

The structure transition processes of protein molecules observed in DMD simulations can be represented by a sequence of contact maps.^{40,41} To elucidate the pathway of the protein unfolding/folding

on surfaces, we applied unsupervised autoencoders and clustering algorithms to identify stages of protein structure transitions by generating low-dimensional feature vectors of a contact map and clustering them.^{33,42} Contrastive learning⁴³ is a state-of-the-art self-supervised learning model that has proven effective in extracting deep representations from images for downstream tasks, such as clustering and classification.^{44–47} It is independent of the dataset and has been successfully applied to study protein structures.^{48,49} In this study, we utilized contrastive learning to train a contact map feature extractor for protein structural changes upon adsorption on a graphene surface. The feature vectors were then grouped by the k-means clustering algorithm to identify the protein transition stages during these processes.

Ubiquitin, a small protein composed of 76 amino acids, is present in all eukaryotic cells and performs its functions by conjugating with a wide range of target proteins.^{50–59} Ubiquitin can be implicated in neurodegenerative diseases associated with proteostasis dysfunction, such as Alzheimer's disease, motor neuron disease, Huntington's disease, and Parkinson's disease.^{60,61} Atomistic-level conformational changes of ubiquitin adsorbed on different surfaces remain unrevealed despite various experimental studies.^{62–64} This study aims to investigate the structural changes of ubiquitin adsorbed on a graphene surface using DMD simulations and graph clustering analyses.

In this study, we used our in-house developed program package, sDMD.^{26,27} Our DMD simulations present protein intramolecular and intermolecular interactions at the atomistic scale, while using a coarse-grained scale to describe protein-surface interactions. This allows for an accurate representation of protein folding/unfolding dynamics and the impact of surface on protein structure simultaneously. In the simulation, the interaction between the protein and the surface is modeled using a Gō-like model.^{23–25,27,33} This model effectively captures the major effects of the graphene surface on protein adsorption, including hydrophobic interactions and π - π stacking. More details about the model and simulation parameters are illustrated in the supplementary material (Figs. S1 and S2 and Table S1). In this study, we adopted the

small protein ubiquitin as a model system. Our method of the combined contrastive learning analysis and DMD simulations can be also applied to study the adsorption behaviors of other types of proteins.

In the DMD simulations, the solvent was treated implicitly using the Lazaridis-Karplus solvation model.⁶⁵ It is noteworthy that due to the event-driven nature of DMD and the implicit solvent treatment, it is challenging to correlate simulation time and temperature with their physical counterparts.^{66,67} Therefore, we used reduced units, including the time step t^* and temperature $T^* = T/T_s$, where T_s was determined by setting $Nk_B T_s = E$, where N is the Avogadro's constant, k_B is the Boltzmann's constant, and E is a unit of energy equal to 1 kcal/mol. For simplicity, the superscript “*” will be omitted, as all the future temperatures are dimensionless. The value of T_s was found to be 503.2 K, and we set $T = 0.57$, which was about 286.8 K lower than the folding temperature of ubiquitin.⁶⁸ At $T = 0.57$, our simulation showed that the ubiquitin would not undergo structural denaturation in the bulk water without a substrate surface.

An ubiquitin protein molecule (PDB code: 2MOR) was initially placed 2.0 nm from an implicit graphene surface to study adsorption. Simulation systems were enclosed in a $10 \times 10 \times 10 \text{ nm}^3$ box with the periodic boundary conditions (PBCs) in the X , Y , and Z directions (Fig. S4 in the supplementary material). The simulation protocol included an energy minimization procedure, followed by multiple short runs of 20 000 timesteps each, progressively increasing the system's temperature from 0.4 to 0.57. The production simulations were performed in the NVT ensemble at $T = 0.57$ until the system reached equilibrium, which occurred after approximately 3.0×10^6 timesteps.

As illustrated in Fig. 1, the center of the algorithm is to use contrastive learning to train a contact map feature extractor by maximizing the agreement of positive pairs $(\tilde{x}_i, \tilde{x}_j)$ in which $\tilde{x}_i$ and $\tilde{x}_j$ are augmented contact maps obtained from the same contact map x . The augmentation functions that generate the augmented contact maps belong to a set of candidate augmentations U . In this work, U consists of random resized cropping, Sobel filtering, and random horizontal

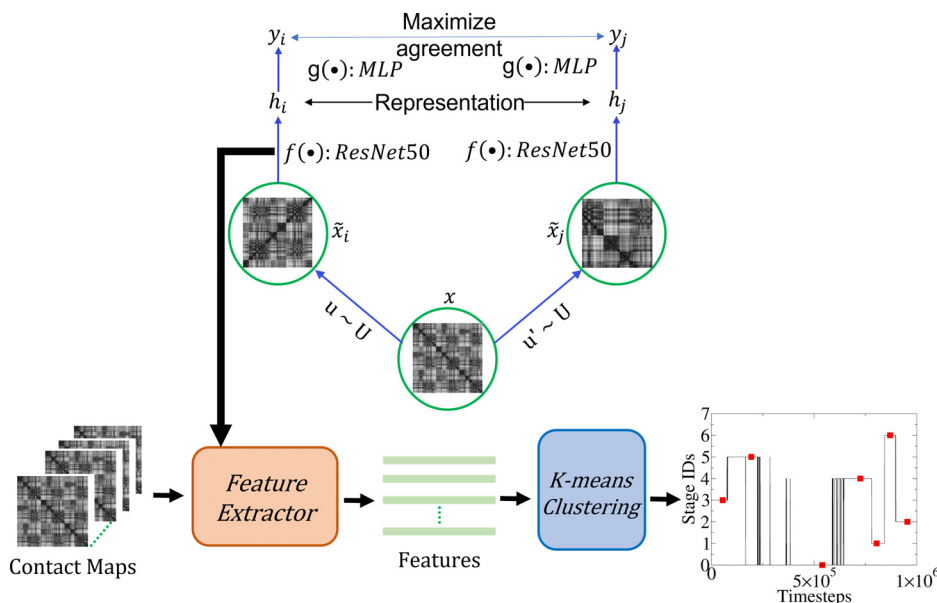


FIG. 1. The overall structure of contrastive learning-based protein structure transition stage identification. $(\tilde{x}_i, \tilde{x}_j)$ are in a positive pair if $\tilde{x}_i$ and $\tilde{x}_j$ are augmented contact maps from the same contact map x , generated by randomly applied augmentations $u \sim U$ and $u' \sim U$, in which U is the set of candidate augmentation functions. $f(\bullet) = \text{ResNet50}$ is the feature extractor, and $g(\bullet)$ is a multilayer perceptron (MLP) with one hidden layer that projects the contact map representations for the loss function application.

flipping. These operations are applied in the aforementioned sequential order stochastically with a probability of 0.5 each. The backbone contact map feature extractor (*ResNet50*) and the project head [multi-layer perceptron (MLP) with one hidden layer] are trained to maximize the agreement between augmented contact maps belonging to positive pairs. The loss function used in the training is defined as

$$loss_{i,j} = -\log \frac{\exp(\text{sim}(x_i, x_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(x_i, x_j))},$$

where $\text{sim}(x_i, x_j)$ is the cosine similarity between a pair of augmented contact maps (x_i and x_j). τ is a temperature parameter, which is determined empirically. After the training in contrastive learning is done, the sequence of contact maps obtained from the MD simulations is fed into the backbone feature extractor. Features of contact maps are vectors of size (2048×1) . These feature vectors are then grouped into clusters by the k-means clustering algorithm to form stages of protein molecular structure transitions.

In this work, we first simulated the entire adsorption process using our high-throughput sDMD program.^{26,27} We carried out five independent runs with different initial configurations of the ubiquitin on the graphene surface. Upon visual inspection using VMD,⁶⁹ the simulation trajectories showed that ubiquitin protein adsorbed onto the graphene surface and its structure collapsed, regardless of the initial landing sites or orientations. Our simulations encompassed 3.0×10^6 timesteps and involved approximately 25×10^9 bond events and 7.2×10^9 collision events. Based on Marchut and Hall's estimation,⁷⁰ the timescale of our simulations was approximately 250 μs . To conduct the contrastive learning analyses, we selected one representative simulation trajectory. Since the ubiquitin protein's structure

disintegrated into random coils after 1.0×10^6 timesteps, providing no useful information on transition states, our contrastive learning analysis focused only on the data from the first 1.0×10^6 timesteps.

Figure 2 shows that the adsorbed ubiquitin protein underwent significant structural transformations. At $t = 0.086 \times 10^6$, the residue Leu50 reached the graphene surface due to the hydrophobic interaction, followed shortly by Leu56 and Ile61. Then, these three hydrophobic residues acted as anchors on the graphene surface. The protein oscillated around the anchor sites and finally landed on the surface with the hydrophobic residues of Ile23-Val26 in an α -helix and Leu15-Val17 in a β -sheet at $t = 0.307 \times 10^6$. In this state, the β -sheet was perpendicular to the surface due to the strong residue-surface hydrophobic interactions, but it gradually collapsed and became a random coil at $t = 0.800 \times 10^6$ [Figs. 2(a) and 2(c)]. The further structural collapse caused the attachment of hydrophobic residues (Ile3-Val5, Leu8, Ile13, Leu15, Val17, Leu43-Phe45, Ile61, Leu67, Leu69-Leu71, and Leu73). Residues (Tyr59 and His68) were also on the graphene surface due to their π - π interactions with the surface. The protein's tertiary structure also collapsed, causing a significant increase in the root-mean-squared deviation (RMSD) [Fig. 2(b)]. After $t = 2.602 \times 10^6$, β -strands reformed within Gln2-Gly11, Thr14-Glu18, and Glu41-Leu50 residues, but they were parallel to the graphene surface instead of perpendicular (see the full time evolution of the secondary structure in Fig. S4 in the supplementary material). Additionally, the α -helix (Ile23-Glu34) started to randomize at $t = 1.600 \times 10^6$ and did not recover until the end of the simulation. Our results agree with previous experimental measurements,⁶³ which found significant changes in ubiquitin's secondary structure on graphene quantum dots. However, these changes cannot be observed in MD simulations,⁶³ due to the limited simulation time.

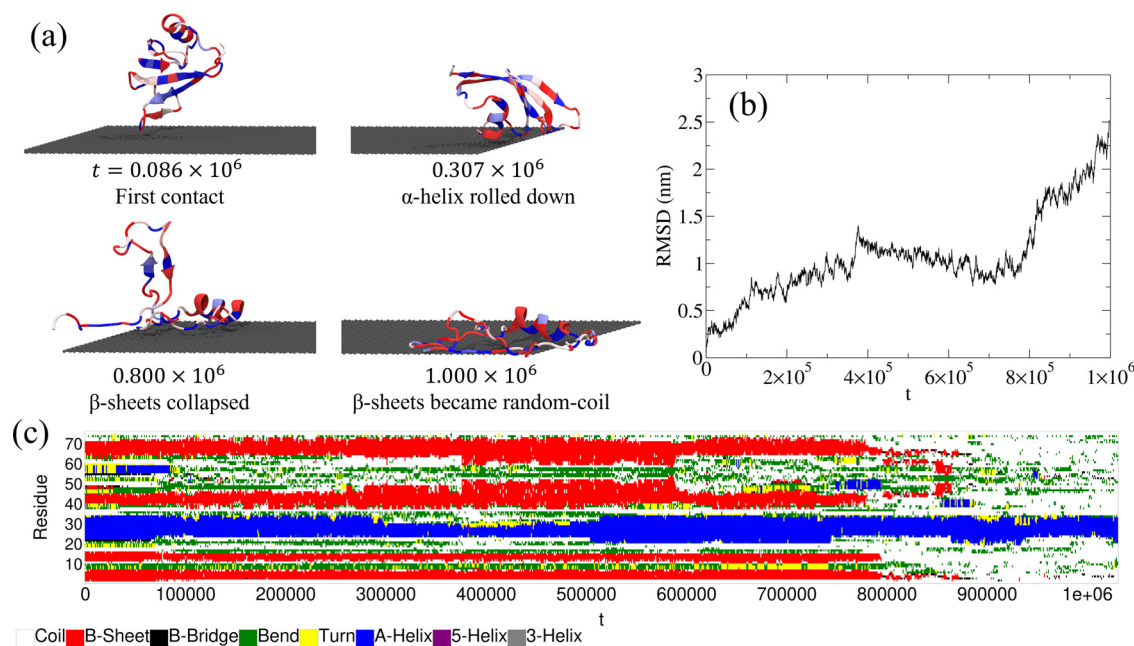


FIG. 2. (a) Zoom-in snapshots of the key events during the adsorption process. (b) RMSD of ubiquitin's tertiary structure during the adsorption process. (c) Time evolutions of the secondary structures. For the snapshots of ubiquitin, the hydrophobic residues were colored blue and the hydrophilic residues are red.

Next, to identify the pathway of protein structure unfolding upon adsorption, we analyzed the DMD simulation data in the first 1.0×10^6 timesteps, using the contrastive learning-based analysis and the simulated protein conformations. A *ResNet50* contact map feature extractor is trained by contrastive learning. The last fully connected layer in *ResNet50* is replaced by an identity layer with 2048 output dimensions. The mini-batch size is 500, and the optimizer used in training is stochastic gradient descent. The temperate parameter $\tau = 0.2$. The base learning rate of a linear warmup cosine annealing scheduler is set to 1.0×10^{-6} . The learning rate follows a linear warmup schedule between 0.0 and the base learning rate followed by a cosine annealing schedule. The maximum number of epochs is 550. To prevent overfitting, training is stopped when there is no improvement of epoch loss for 50 validation checks. The k-means clustering algorithm is used to group the sequence of contact map features into clusters. To determine the optimal number of clusters, we varied the cluster number (k) from 2 to 16 and calculated the corresponding silhouette scores⁷¹ for each of them [Fig. 3(a)]. It was found that $k = 2$

yields the highest silhouette score. The two stages represent the protein in bulk and on the surface, respectively. Obviously, grouping the contact maps into two clusters does not reveal enough details about the adsorption process of ubiquitin protein on a graphene surface. $k = 7$ yields the second-highest silhouette score; thus, seven is chosen as the number of clusters for the k-means clustering algorithm [Fig. 3(b)].

The contact map in which the feature vector is the closest to the centroid of the cluster is used as a representative of the stage [see the red squares in Fig. 3(b)]. The representative contact maps for the seven stages are at timesteps of 0.056×10^6 , 0.195×10^6 , 0.540×10^6 , 0.725×10^6 , 0.804×10^6 , 0.870×10^6 , and 0.953×10^6 [Fig. 3(c)]. They correspond well to key events during the adsorption process, including the protein's native state (stage 1), residue anchoring (stage 2), protein rollover and bending to achieve a full landing (stages 3 and 4), β -sheet collapsing (stages 5 and 6), and random coil formation (stage 7). From the time span of each stage as shown in Fig. 3(b), it was observed that the protein's full landing on the surface (i.e., rolling and bending in stages 3 and 4) was the most time-limiting step in the

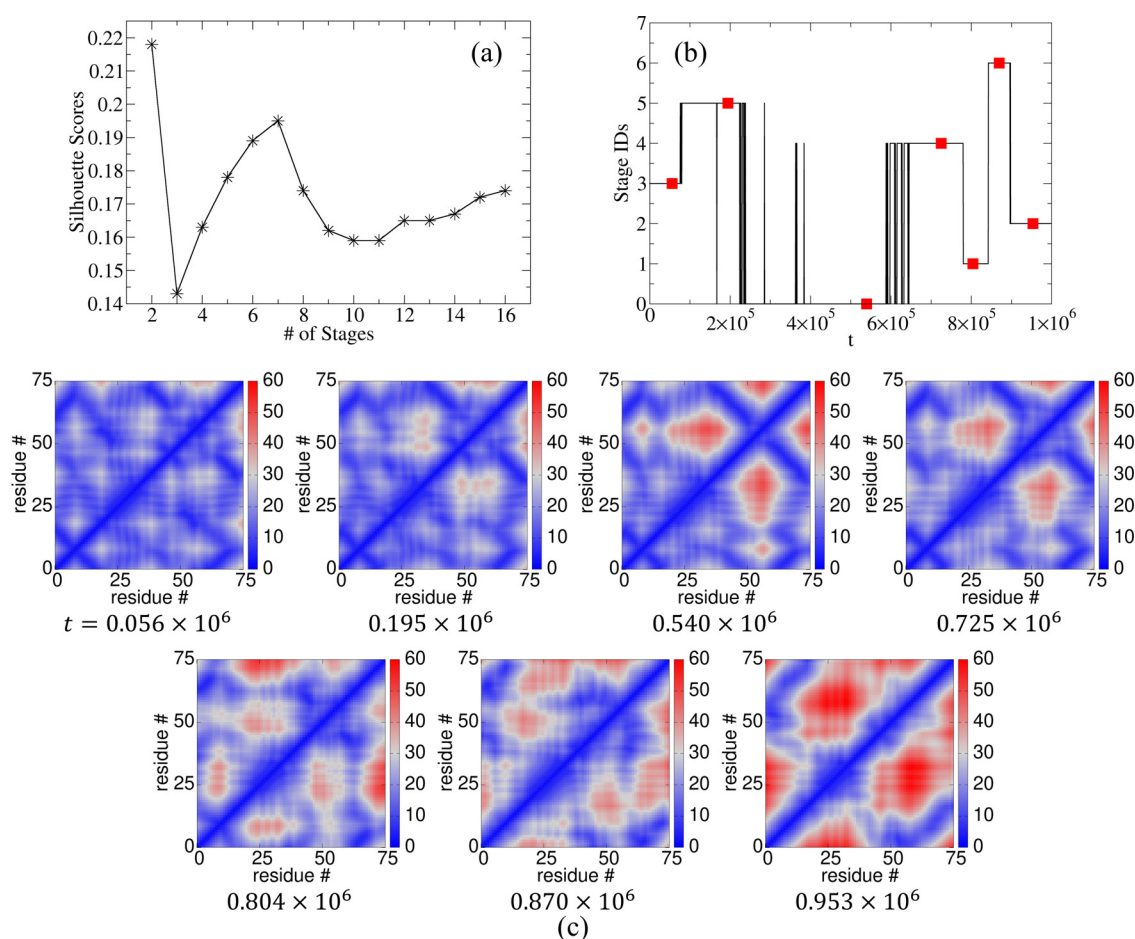


FIG. 3. (a) The Silhouette scores with a different number of stages, ranging from 1 to 16. Number 7 had the best score except for number 2. (b) Seven stages of the ubiquitin adsorption process identified from the contrastive learning analyses. The stage IDs were rearranged to form seven stages in a chronological sequence. The red dots represent the positions that were in closest proximity to their respective stage centers. (c) The representative contact maps of ubiquitin corresponding to each of the stage centers. The distance unit of the color bar of contact maps is Å. Note: For the purpose of visualization effect, the contact maps here are presented with colormaps.

process of protein adsorption (i.e., 1.0×10^6 steps). In our DMD simulations, the dehydration of the surface and protein was taken into account (see the method of DMD in the supplementary material). This observation was consistent with our previous studies^{8,9,18,72–74} of atomistic and CG MD simulations with explicit water, which showed that, in order to achieve maximum contact with the surface at the equilibrium state, the interfacial water molecules needed to be depleted, and the protein rotated on the surface accompanied by slow structural changes. Once the protein fully contacted the surface (in stage 5 and after), a significant portion of the secondary structures quickly and continuously randomized [Fig. 2(c)].

It is worth noting that the contact map feature extractor is crucial for obtaining lower-dimensional characteristic representations of contact maps, which ensures the success of k-means clustering analysis. The length of the ubiquitin protein amino acid sequence is 76; thus, the dimension of a contact map of it is $76 \times 76 = 5776$, storing the pair-wise distance of the amino acids. The freedom of contact maps is reduced to 2048 by the feature extractor. When the sequence of contact maps becomes longer, it is helpful to reduce the representation dimension further to ensure that the k-means clustering results are stable. As shown in the supplementary material, to analyze simulation data in 3.0×10^6 timesteps, the representation feature vector dimension is reduced to 512 by replacing the last fully connected layer of the *ResNet50* with a linear layer of 2048 input and 512 output dimensions, respectively. By comparing the molecular structure transition stages in Figs. 3(b) and S4 in the supplementary material, both results align well with each other within the first 1.0×10^6 timesteps. However, once the protein structure eventually transformed into random coils, the resulting contact maps provided only arbitrary and insignificant structural details, leading to the detection of rapid flipping movement by the contrastive-learning-based analysis [see Fig. S4(b) in the supplementary material]. This also demonstrates the reliability of the contrastive learning method in analyzing protein structural transitions.

In this work, we employed graphic contrastive learning to analyze the DMD simulation trajectory of the ubiquitin adsorption on the graphene surface. Our DMD simulations with a Gō-like CG model can simulate protein–surface adsorption with large temporal scale and structural changes. Upon adsorption, ubiquitin's structure randomized due to strong hydrophobic and π - π interactions. Contrastive learning analysis can detect protein folding/unfolding and elucidates the entire pathway of structural changes and the rate-limit step. Our study revealed a seven-stage unfolding pathway of ubiquitin on graphene. Combining this machine learning approach with efficient DMD simulations could be an essential tool for accelerating materials discovery and phenomena for biomolecular systems.

See the supplementary material for the all-atom protein model and implicit solvation model (Figs. S1 and S2), Gō-like model for the protein–surface interaction (Table S1), initial simulation system (Fig. S3), and comprehensive analysis results for the entire simulation of 3.0×10^6 timesteps (Fig. S4).

T. Wei acknowledges the grant support from the National Science Foundation (Nos. NSF 2118099 and 1943999).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Size Zheng and Yong Wei contributed equally to this work.

Size Zheng: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Yong Wei:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Yuwei Lin:** Conceptualization (equal); Formal analysis (equal); Methodology (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Tao Wei:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- A. K. Leonardi and C. K. Ober, *Annu. Rev. Chem. Biomol. Eng.* **10**, 241–264 (2019).
- Q. Li, C. Wen, J. Yang, X. Zhou, Y. Zhu, J. Zheng, G. Cheng, J. Bai, T. Xu, and J. Ji, *Chem. Rev.* **122**(23), 17073–17154 (2022).
- H. Huang, C. Zhang, R. Crisci, T. Lu, H. C. Hung, M. S. J. Sajib, P. Sarker, J. Ma, T. Wei, S. Jiang, and Z. Chen, *J. Am. Chem. Soc.* **143**(40), 16786–16795 (2021).
- P. Sarker, G. T. Chen, M. S. J. Sajib, N. W. Jones, and T. Wei, *Colloids Surf., A* **653**, 129943 (2022).
- Q. Li, Y. Chen, S. Bai, X. Shao, L. Jiang, and Q. Li, *Colloids Surf., B* **188**, 110812 (2020).
- T. Zhang, T. Wei, Y. Han, H. Ma, M. Samieegohar, P. W. Chen, I. Lian, and Y. H. Lo, *ACS Cent. Sci.* **2**(11), 834–842 (2016).
- M. Yi, C. H. Lau, S. Xiong, W. Wei, R. Liao, L. Shen, A. Lu, and Y. Wang, *ACS Appl. Mater. Interfaces* **11**(17), 15698–15708 (2019).
- M. S. Jahan Sajib, P. Sarker, Y. Wei, X. Tao, and T. Wei, *Langmuir* **36**(44), 13356–13363 (2020).
- P. Sarker, M. S. J. Sajib, X. Tao, and T. Wei, *J. Phys. Chem. B* **126**(3), 601–608 (2022).
- D. S. Kim, H. S. Choi, X. Yang, J. H. Yang, J. H. Lee, H. Y. Yoo, J. Lee, C. Park, and S. W. Kim, *J. Ind. Eng. Chem.* **81**, 108–114 (2020).
- M. S. Jahan Sajib, Y. Wei, A. Mishra, L. Zhang, K. I. Nomura, R. K. Kalia, P. Vashishta, A. Nakano, S. Murad, and T. Wei, *Langmuir* **36**(26), 7658–7668 (2020).
- T. Wei, H. Ma, and A. Nakano, *J. Phys. Chem. Lett.* **7**(5), 929–936 (2016).
- W. Guo, X. Zou, H. Jiang, K. J. Koebke, M. Hoarau, R. Crisci, T. Lu, T. Wei, E. N. G. Marsh, and Z. Chen, *J. Phys. Chem. B* **125**(28), 7706–7716 (2021).
- T. Wei, S. Kaewtathip, and K. Shing, *J. Phys. Chem. C* **113**(6), 2053–2062 (2009).

- ¹⁵M. Samieegohar, H. Ma, F. Sha, M. S. Jahan Sajib, G. I. Guerrero-García, and T. Wei, *Appl. Phys. Lett.* **110**(7), 073703 (2017).
- ¹⁶W. Guo, T. Lu, R. Crisci, S. Nagao, T. Wei, and Z. Chen, *Chem. Sci.* **14**(11), 2999–3009 (2023).
- ¹⁷Y. Li, T. L. Ogorzalek, S. Wei, X. Zhang, P. Yang, J. Jasensky, C. L. Brooks, E. N. G. Marsh, and Z. Chen, *Phys. Chem. Chem. Phys.* **20**(2), 1021–1029 (2018).
- ¹⁸C. M. Nakano, H. Ma, and T. Wei, *Appl. Phys. Lett.* **106**(15), 153701–153704 (2015).
- ¹⁹J. Zheng, L. Li, S. Chen, and S. Jiang, *Langmuir* **20**(20), 8931–8938 (2004).
- ²⁰L. Li, S. Chen, J. Zheng, B. D. Ratner, and S. Jiang, *J. Phys. Chem. B* **109**(7), 2934–2941 (2005).
- ²¹B. Qiao and M. Olvera de la Cruz, *ACS Nano* **14**(8), 10616–10623 (2020).
- ²²G. Yu, J. Liu, and J. Zhou, *J. Phys. Chem. B* **118**(17), 4451–4460 (2014).
- ²³S. Wei and T. A. Knotts, *J. Chem. Phys.* **139**(9), 095102–095111 (2013).
- ²⁴S. Wei, X. Zou, J. Tian, H. Huang, W. Guo, and Z. Chen, *J. Am. Chem. Soc.* **141**(51), 20335–20343 (2019).
- ²⁵S. Wei, X. Zou, K. Cheng, J. Jasensky, Q. Wang, Y. Li, C. Hussal, J. Lahann, C. L. Brooks, and Z. Chen, *J. Phys. Chem. C* **120**(34), 19078–19086 (2016).
- ²⁶S. Zheng, L. Javidpour, M. Sahimi, K. S. Shing, and A. Nakano, *Comput. Phys. Commun.* **247**, 106873–106886 (2020).
- ²⁷S. Zheng, M. S. J. Sajib, Y. Wei, and T. Wei, *J. Chem. Theory Comput.* **17**(3), 1874–1882 (2021).
- ²⁸B. J. Alder and T. E. Wainwright, *J. Chem. Phys.* **31**(2), 459–466 (1959).
- ²⁹A. Voegler Smith and C. K. Hall, *Proteins* **44**(3), 344–360 (2001).
- ³⁰F. Ding, N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich, *J. Mol. Biol.* **324**(4), 851–857 (2002).
- ³¹D. Feng, J. M. Borreguero, S. V. Buldyrev, H. E. Stanley, and N. V. Dokholyan, *Proteins* **53**(2), 220 (2003).
- ³²S. Zheng, A. Sahimi, K. S. Shing, and M. Sahimi, *Biophys. J.* **120**(1), 64–72 (2021).
- ³³J. Chen, E. Xu, Y. Wei, M. Chen, T. Wei, and S. Zheng, *Langmuir* **38**(35), 10817–10825 (2022).
- ³⁴C. L. Rountree, R. K. Kalia, E. Lidorikis, A. Nakano, L. Van Brutzel, and P. Vashishta, *Annu. Rev. Mater. Res.* **32**(1), 377–400 (2002).
- ³⁵T. Wei and C. Ren, in *Polymer Science and Innovative Applications* (Elsevier, 2020), pp. 207–228.
- ³⁶L. Hernandez de la Pena, R. van Zon, J. Schofield, and S. B. Opps, *J. Chem. Phys.* **126**(7), 074106 (2007).
- ³⁷S. Zheng, L. Javidpour, K. S. Shing, and M. Sahimi, *J. Chem. Phys.* **145**(13), 134306 (2016).
- ³⁸S. Zheng, K. S. Shing, and M. Sahimi, *J. Chem. Phys.* **148**(10), 104305 (2018).
- ³⁹S. Zheng, A. Sahimi, K. S. Shing, and M. Sahimi, *J. Chem. Phys.* **150**(14), 144307 (2019).
- ⁴⁰W. Zheng, Y. Li, C. Zhang, R. Pearce, S. M. Mortuza, and Y. Zhang, *Proteins* **87**(12), 1149–1164 (2019).
- ⁴¹Y. Li, C. Zhang, E. W. Bell, W. Zheng, X. Zhou, D. J. Yu, and Y. Zhang, *PLoS Comput. Biol.* **17**(3), e1008865 (2021).
- ⁴²D. Bhowmik, S. Gao, M. T. Young, and A. Ramanathan, *BMC Bioinf.* **19**, 47–48 (2018).
- ⁴³T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, in *Proceedings of the 2020 International Conference on Machine Learning* (PMLR, 2020), pp. 1597–1607.
- ⁴⁴T. Park, A. A. Efros, R. Zhang, and J. Y. Zhu, in *Proceedings of the Part IX Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020* (Springer International Publishing, 2020).
- ⁴⁵K. Chaitanya, E. Erdil, N. Karani, and E. Konukoglu, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2020), Vol. 33, pp. 12546–12558.
- ⁴⁶Y. Li, P. Hu, Z. Liu, D. Peng, J. T. Zhou, and X. Peng, in *Proceedings of the AAAI Conference on Artificial Intelligence* (AAAI, 2021), Vol. 35.
- ⁴⁷M. R. Karim, O. Beyan, A. Zappa, I. G. Costa, D. Rebholz-Schuhmann, M. Cochez, and S. Decker, *Briefings Bioinf.* **22**(1), 393–415 (2021).
- ⁴⁸A. X. Lu, H. Zhang, M. Ghassemi, and A. Moses, “Self-supervised contrastive learning of protein representations by mutual information maximization,” *bioRxiv*:283929 (2020).
- ⁴⁹P. Hermosilla and T. Ropinski, “Contrastive representation learning for 3D protein structures,” *arXiv:2205.15675* (2022).
- ⁵⁰M. H. Glickman and A. Ciechanover, *Physiol. Rev.* **82**(2), 373–428 (2002).
- ⁵¹C. Hoege, B. Pfander, G. L. Moldovan, G. Pyrowolakis, and S. Jentsch, *Nature* **419**(6903), 135–141 (2002).
- ⁵²J. D. Schnell and L. Hicke, *J. Biol. Chem.* **278**(38), 35857–35860 (2003).
- ⁵³L. Hicke and R. Dunn, *Annu. Rev. Cell Dev. Biol.* **19**, 141–172 (2003).
- ⁵⁴C. M. Pickart, *Cell* **116**(2), 181–190 (2004).
- ⁵⁵C. M. Pickart and M. J. Eddins, *Biochim. Biophys. Acta.* **1695**(1–3), 55–72 (2004).
- ⁵⁶D. Mukhopadhyay and H. Riezman, *Science* **315**(5809), 201–205 (2007).
- ⁵⁷V. Joshi, A. Amanullah, A. Upadhyay, R. Mishra, A. Kumar, and A. Mishra, *Front. Mol. Neurosci.* **9**, 93 (2016).
- ⁵⁸T. E. T. Mevissen, Y. Kulathu, M. P. C. Mulder, P. P. Geurink, S. L. Maslen, M. Gersch, P. R. Elliott, J. E. Burke, B. D. M. van Tol, M. Akutsu, F. E. Oualid, M. Kawasaki, S. M. V. Freund, H. Ovaa, and D. Komander, *Nature* **538**(7625), 402–405 (2016).
- ⁵⁹E. Perrody, L. Abrami, M. Feldman, B. Kunz, S. Urbe, and F. G. van der Goot, *Elife* **5**, e19083 (2016).
- ⁶⁰D. Popovic, D. Vucic, and I. Dikic, *Nat. Med.* **20**(11), 1242–1253 (2014).
- ⁶¹J. Y. Yebury, N. E. Farrawell, and L. McAlary, *Trends Neurosci.* **43**(5), 274–284 (2020).
- ⁶²S. Mondal, R. Thirupathi, L. P. Rao, and H. S. Atreya, *RSC Adv.* **6**(58), 52539–52548 (2016).
- ⁶³G. Fang, B. Luan, C. Ge, Y. Chong, X. Dong, J. Guo, C. Tang, and R. Zhou, *Carbon* **121**, 285–291 (2017).
- ⁶⁴S. A. Malik, Z. Mohanta, C. Srivastava, and H. S. Atreya, *Nanoscale Adv.* **2**(5), 1904–1912 (2020).
- ⁶⁵T. Lazaridis and M. Karplus, *Proteins* **35**(2), 133–152 (1999).
- ⁶⁶D. Shirvanyants, F. Ding, D. Tsao, S. Ramachandran, and N. V. Dokholyan, *J. Phys. Chem. B* **116**(29), 8375–8382 (2012).
- ⁶⁷Y. Wang, Q. Shao, and C. K. Hall, *J. Biol. Chem.* **291**(42), 22093–22105 (2016).
- ⁶⁸G. I. Makhatazde, M. M. Lopez, J. M. Richardson III, and S. T. Thomas, *Protein Sci.* **7**(3), 689–697 (1998).
- ⁶⁹W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graphics* **14**(1), 33–38, 27–38 (1996).
- ⁷⁰A. J. Marchut and C. K. Hall, *Proteins* **66**(1), 96–109 (2006).
- ⁷¹P. J. Rousseeuw, *J. Comput. Appl. Math.* **20**, 53–65 (1987).
- ⁷²T. Wei, M. A. Carignano, and I. Szleifer, *J. Phys. Chem. B* **116**(34), 10189–10194 (2012).
- ⁷³G. T. Chen, P. Sarker, B. Qiao, and T. Wei, *J. Nanopart. Res.* **25**(6), 108 (2023).
- ⁷⁴T. Wei, M. A. Carignano, and I. Szleifer, *Langmuir* **27**(19), 12074 (2011).