LANGMUIR

pubs.acs.org/Langmuir Article

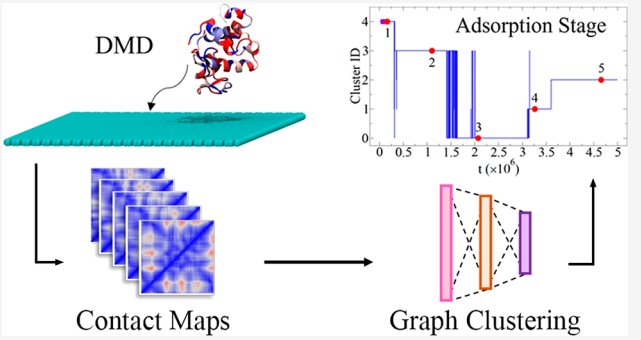
Graph Clustering Analyses of Discontinuous Molecular Dynamics Simulations: Study of Lysozyme Adsorption on a Graphene Surface

Jing Chen, Enze Xu, Yong Wei,* Minghan Chen,* Tao Wei,* and Size Zheng*

Cite This: *Langmuir* 2022, 38, 10817−10825 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Understanding the interfacial behaviors of biomolecules is crucial to applications in biomaterials and nanoparticle-based biosensing technologies. In this work, we utilized autoencoder-based graph clustering to analyze discontinuous molecular dynamics (DMD) simulations of lysozyme adsorption on a graphene surface. Our high-throughput DMD simulations integrated with a Gō-like protein−surface interaction model makes it possible to explore protein adsorption at a large temporal scale with sufficient accuracy. The graph autoencoder extracts a low-dimensional feature vector from a contact map. The sequence of the extracted feature vectors is then clustered, and thus the evolution of the protein molecule structure in the absorption process is segmented into stages. Our study demonstrated that the residue−surface hydrophobic interactions and the $\pi-\pi$ stacking interactions play key roles in the five-stage adsorption. Upon adsorption, the tertiary structure of lysozyme collapsed, and the secondary structure was also affected. The folding stages obtained by autoencoder-based graph clustering were consistent with detailed analyses of the protein structure. The combination of machine learning analysis and efficient DMD simulations developed in this work could be an important tool to study biomolecules' interfacial behaviors.



1. INTRODUCTION

The interaction between proteins and solids (e.g., nanoparticles, semiconductors, minerals, oxides, and metals) has played a critical role in many fields, such as antibacterial/antifouling coating,[1] industrial catalysts,[2,3] biosensors,[4,5] and drug delivery.[6] Much experimental work has been carried out on the adsorption of proteins on solid materials, but the presentation of the microscopic details of the molecular structure and adsorption process remains unclear and very challenging.[7−9] Atomistic molecular dynamics (MD) simulation, on the other hand, is capable of providing time-evolution spatial details of a protein's structure at the atomic level.[9−14] However, due to the complexity of the protein adsorption process, the computational cost would be prohibitively expensive for the complete simulation of the whole process, especially performed with explicit solvent and all-atom molecular models.[9−15] Coarse-grained (CG) MD simulation can offer a more efficient way to simulate the same system by sacrificing the model's resolution and the computing accuracy through grouping a number of atoms into CG particles and simplifying the intra- and intermolecular interaction functions[16] but cannot present the secondary structural changes of proteins.[9,17−19]

To achieve sufficient computational efficiency without compromising accuracy and to present protein's secondary structural changes during the simulation of the adsorption process, we integrated a Gō-like model into our in-house-developed framework of the discontinuous molecular dynamics (DMD) simulation program.[20] DMD is a fast and effective simulation method derived from conventional MD (CMD) at the atomistic and coarse-grained scales. It was first proposed by Alder and Wainwright[21] in 1959 to simulate hard-sphere interactions and then further developed to simulate protein folding and aggregation.[22−24] As opposed to the time-driven nature (explicit continuous dynamics) of CMD, DMD is event-driven with an adaptive and large time step. Thus, a DMD simulation can be up to 100 times more efficient than an atomistic CMD simulation in implicit solvent, as reported in previous research.[25] In such a way, the trajectory and potential of a complex biological system can be simulated discontinuously with spatial and time scales beyond the limit of CMD. We used DMD to successfully simulate the folding and aggregation[26−29] of various peptides/proteins as well as surface adsorption,[20] which covered the entire dynamic process from ab initio to the final equilibrium state while requiring less computational time compared to CMD. The results were also

Received: May 24, 2022
Revised: August 12, 2022
Published: August 24, 2022

https://doi.org/10.1021/acs.langmuir.2c01331
Langmuir 2022, 38, 10817−10825

consistent with the experimental measurements. In the simulation of peptide/protein adsorption on the surface, an all-atom force field[30] was employed to compute the intra- and intermolecular interactions of proteins precisely and efficiently at the atomistic level, and a Gō-like model[31] was applied to calculate the peptide/protein interactions with the surface at the CG level.

Many attempts have been made to characterize and predict protein folding using deep learning approaches since it has demonstrated remarkable success in many applications involving complex data.[32−35] With tools such as multitask learning and autoencoders, deep learning offers new possibilities over traditional methods. The sparse autoencoder was first proposed by Ng et al.[36] to learn features from unlabeled data. Based on the prior work, Tian et al.[37] integrated the sparse autoencoder into the fields of graph clustering, referred to as a graph encoder, and obtained better clustering accuracy than the mainstream spectral clustering algorithm. In recent studies of protein folding, Bhowmik et al.[38] used a convolutional variational autoencoder (CVAE) to learn low-dimensional latent features from long-time-scale protein folding simulations and then cluster those results. Similarly, Liu et al.[39] proposed a data-driven generative conformation clustering method based on the adversarial autoencoder (AAE) to evaluate the transition rates and cluster protein behaviors. In this work, inspired by the work of Tian et al.,[37] we developed a graph autoencoder stratification method to segment the protein adsorption process and identify the representative protein structures in each stage. The evolution process of protein molecule structure obtained in the simulation was represented by a sequence of contact maps. A contact map is a two-dimensional matrix in which the value at (i, j) is the spatial distance between the ith and jth amino acids of a protein molecule at a particular moment. The graph autoencoder extracted a low-dimensional feature vector from each contact map. The sequence of the feature vectors was clustered, and thus the evolution process of protein molecule structure was segmented into stages.

Lysozyme, as an anti-inflammatory enzyme, has great biological significance. Although many experimental studies have been conducted on the adsorption of lysozyme on different surfaces,[7,8,40−43] few presented the conformational changes at the atomistic scale, and thus the mechanism details underlying the surface's effects on the adsorption on the time scale of microseconds from the mesoscale level are still unclear. In this work, we present the simulation results to study the adsorption behavior and conformational changes of lysozyme on the graphene surface, combined with graph clustering analyses to segment the adsorption process. The structure of the rest of this article is as follows: section 2 describes the details of the implementation of DMD simulations and graph clustering analyses; section 3 presents and discusses the simulation and analysis results; and section 4 gives conclusions and summarizes the article.

2. METHOD

2.1. DMD Simulations. The simulations were carried out using our in-house-developed program package, *sDMD*.[20,44] We used an all-atom force field[30] for the intra- and interprotein interactions and a CG Gō-like model[31] for the peptide/protein interactions with the surface. The solvent in the simulations was implicitly represented by the Lazaridis−Karplus solvation model.[30,45] It is noteworthy that because the DMD is event-driven and the solvent is implicit, it is difficult to correlate the simulation time and temperature with the actual physical time and temperature.[46,47] Therefore, we used the time step, t^*, and the reduced temperature, $T^* = T/T_s$, rather than the real unit to represent the time scale and temperature in the simulation. To determine T_s, we set $Nk_BT_s = E$, where N represents Avogadro's constant, k_B is Boltzmann's constant, and E is a unit of energy, which is taken as 1 kcal/mol. For the sake of simplicity, however, we will delete the superscript "*", knowing that all the future temperatures are dimensionless. We thus obtained T_s = 503.2 K and set T = 0.566, which was slightly below the folding temperature of the lysozyme. The corresponding real temperature, therefore, is 285 K, at which the lysozyme showed no denaturing.

The simulations were first performed in a bulk environment without a solid surface; then, a lysozyme (PDB code: 1AKI) was placed about 2.0 nm above an implicit surface, which mimicked the graphene surface, for the adsorption simulations. Both systems were in a box with dimensions of 15 × 15 × 15 nm^3, with periodic boundary conditions (PBC) in the X, Y, and Z directions (Figure 1). The simulations started with an energy minimization, followed by a series of short runs, each lasting 20 000 time steps, during which the system's temperature was gradually increased from 0.4 to 0.566. Then the production simulations were carried out in the *NVT* ensemble at T = 0.566 until the system reached the equilibrium state at about 5.0 × 10^6 time steps.

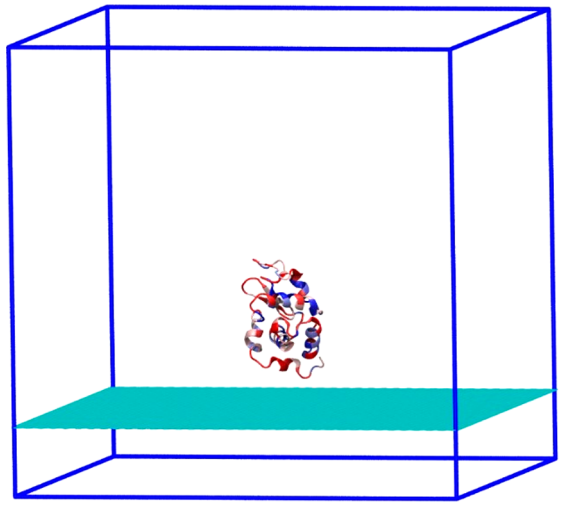


Figure 1. Snapshot of the initial configuration of the system: graphene (cyan) on the X−Z plane (the Y axis is normal to the plane) and lysozyme (the hydrophobic residues were colored blue and the hydrophilic residues were colored red). The PDB code of the protein model is 1AKI.

As mentioned in section 1, the intra- and intermolecular interactions of proteins are represented by an all-atom molecular model within implicit water. A brief description of the model is given in the Supporting Information, and the full details can be found in the literature.[30,44] The interaction between protein and a surface, however, is represented by a Gō-like model.[31,48] It can present the main effects of the graphene surface on protein adsorption: hydrophobicity interaction and π−π stacking. The potential energy function is as follows

https://doi.org/10.1021/acs.langmuir.2c01331
Langmuir 2022, 38, 10817−10825

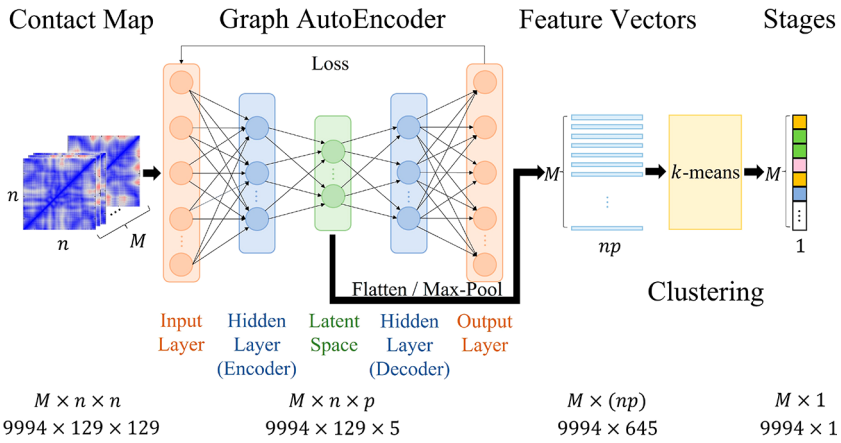


Figure 2. Overflow of graph autoencoder stratification for the detection of the stages of protein adsorption.

Table 1. Graph Autoencoder Stratification for Protein Adsorption Stage Detection

Input

M matrices, each of which is a $n \times n$ symmetric matrix S_i, $i = 1, \ldots, M$.

For $i = 1$ to M:

1. Normalize S_i and take $X^{(1)} = D^{-1}S_i$ as the input of the sparse encoder. $D = diag\{d_1, d_2, \ldots, d_n\}$ and $d_j = \max_i\{\sum_{k=1}^{n} S_{i(j,k)}\}$, where $S_{i(j,k)}$ represents the entry in the j^{th} row and k^{th} column of the matrix S_i.

2. Sparse autoencoder.
   For $j = 1$ to Γ:

   (a) Build a three-layer encoder with input data $X^{(j)}$.
   (b) Train the encoder by optimizing with back-propagation. Obtain the hidden layer activations $h^{(j)}$.
   (c) Set $X^{(j+1)} = h^{(j)}$. Note that $X^{(j)} \in R^{n \times r^j}$. Let $p = r^{\Gamma}$.

   End

3. Flatten the output matrix $X^{\Gamma} \in R^{n \times p}$ into a pn-dimension vector V_i

End

Run k-means on feature vectors $[V_i]_{i=1}^{M}$.

Output

Final clustering label.

$$V_{\text{surface}} = \sum_{i}^{N} \left\{ \pi\rho{\sigma_i}^3 \epsilon_i \left[\theta_1 \left(\frac{\sigma_i}{z_{is}}\right)^9 - \theta_2 \left(\frac{\sigma_i}{z_{is}}\right)^7 + \theta_3 \left(\frac{\sigma_i}{z_{is}}\right)^3 \right] - [\theta\chi_s + \theta_p(\chi_{ip} + 2\delta)] \left(\frac{\sigma_i}{z_{is}}\right)^3 \right\} \quad (1)$$

where N is the residue's number in the protein; z_{is} is the distance between residue i and the surface; σ_i and ϵ_i are van der Waals parameters; θ_s and χ_s in the fourth term control the hydrophobicity of the surface; θ_p and χ_{ip} control the hydrophobicity of each amino acid in the protein; and δ controls the planarity of the side chain. Thus, the first three terms $\left(\left(\frac{\sigma_i}{z_{is}}\right)^x, x = 9, 7, \text{and } 3\right)$ of the above function can be used to calculate the interactions between any types of amino acids and a specific surface, and the fourth and fifth terms mainly demonstrate the hydration/dehydration effects and the

https://doi.org/10.1021/acs.langmuir.2c01331
Langmuir 2022, 38, 10817−10825

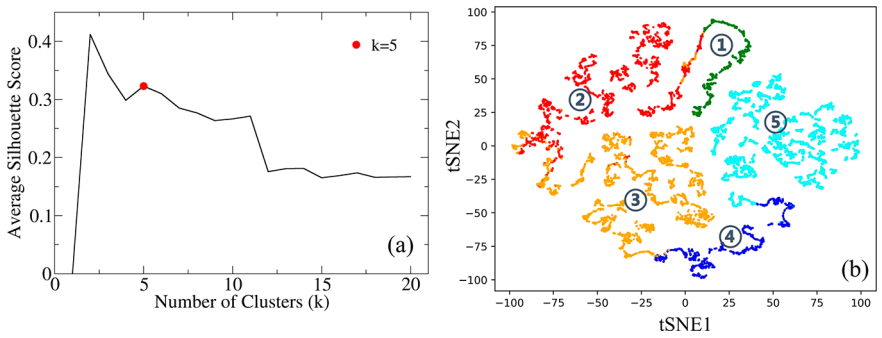


Figure 3. (a) Average silhouette scores with different numbers of clusters, ranging from 1 to 20. k = 5 is chosen as the best number of clusters. (b) t-SNE plot of graph autoencoder stratification. The color codes of the stages are green (stage 1), red (stage 2), orange (stage 3), blue (stage 4), and cyan (stage 5).

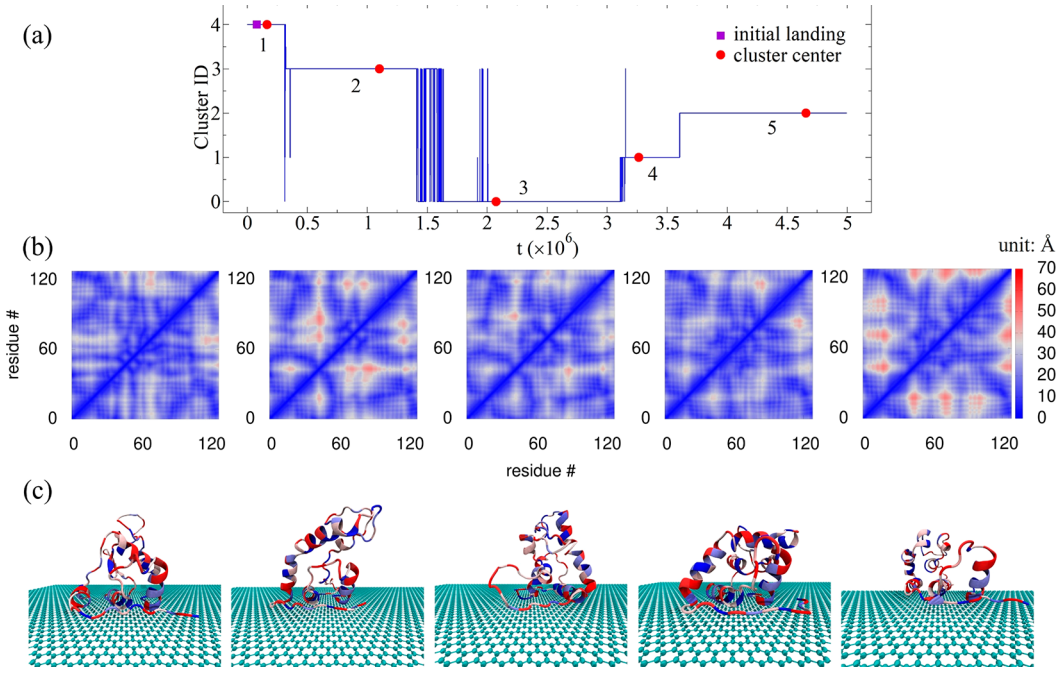


Figure 4. (a) Five clustered stages of the lysozyme adsorption process. Cluster IDs were remapped into the five stages in chronological order. The purple dot is the initial landing point, and the red dots are the positions that are the closest to the corresponding cluster centers. (b) Five representative contact maps that are the closest to the corresponding cluster centers in terms of Euclidian distance with a unit of Å. (c) Five snapshots of the protein molecule structure closest to the cluster centers.

$\pi-\pi$ stacking interactions between a protein and the graphene surface due to the side chain's planarity. In this work, the graphene surface was simulated as a relatively hydrophobic (RPH) surface with χ = 1.5, as in the previous studies.[20,48,49] The values of hydropathy indices of all 20 types of amino acids and 3 types of surfaces, as well as the other parameters used in the above function, can be found in Table S1 in the Supporting Information.

Within the Gō-like model, the atoms in each residue were grouped into a single CG bead with its center of mass located on the α-carbon in the all-atom model. Thus, the surface's force acting on each CG bead was assigned to the α-carbon of a residue and then dissipated to its surrounding atoms by future interactions, during which the whole process is energy- and momentum-conserved. In DMD, the continuous potential profile between protein/peptide and the surface was discretized and represented by step functions. The protocols used in this work during discretization are as follows: the minimum distance between the protein/peptide residues and the surface is where the potential reaches positive 1.0 kcal/mol; the cutoff distance was set at 10.0 Å; in between, the discretized function was added a step once the distance between two adjacent steps reached 2 Å or the distance between one step and the position of the lowest potential reached 1 Å. The parameters of the discretized potential for the interactions of protein residues with the graphene surface are listed in the Supporting Information (Table S2). Detailed information on the discretization of protein−protein interactions in DMD simulations was reported elsewhere.[44]

2.2. Machine Learning Analyses of the Folding Stages. The molecular structure of a protein in three dimensions can be represented by the spatial distances of all possible amino acid residue pairs, formulating a symmetric matrix S, also called the contact map. S is an $n \times n$ matrix, where n is the total number of residues in the protein, and the value at (i, j) is the distance between the α-carbons at the ith and jth residues in the chain. Therefore, the spatial evolution of a protein structure during the adsorption process can be

https://doi.org/10.1021/acs.langmuir.2c01331
Langmuir 2022, 38, 10817−10825

characterized by a series of contact maps collected at each time step. Here we have full data sets of contact maps depicting the structural evolution of lysozyme during the complete process of adsorption. Because the lysozyme has 129 residues, the corresponding contact maps are matrices of size 129 × 129 storing the distance between each pair of residues in the protein.

We proposed a graph autoencoder stratification method, an autoencoder-based graph clustering model, to segment the protein adsorption process. As shown in Figure 2, the inputs of our model were M contact maps collected during the protein adsorption process, represented as $[S_i]_{i=1}^{M}$ where S_i is a symmetric matrix of size $n \times n$ at time step i. For each S_i, we fed the normalized training set $D^{-1}S_i$ into a sparse encoder (feed-forward neural networks) consisting of three fully connected layers. The output feature from the encoder became a matrix of size $n \times p$, often referred to as a latent space. Further flattening the feature matrices, the original M contact maps can be represented as M feature vectors, denoted as $[V_i]_{i=1}^{M}$. Finally, the k-means method was used to cluster the feature vectors into different absorption stages. In the case of the lysozyme adsorption process, $M = 9994$, $n = 129$, and $p = 5$. The pseudocode of the graph autoencoder stratification method is summarized in Table 1. We ran our model on a RedHat Linux system using Python 3.6 and Torch 1.10.

3. RESULTS

We performed five independent adsorption simulations on the graphene surface, starting with various initial configurations. A visual inspection of the trajectories by VMD[50] showed that the lysozyme in all runs could adsorb on the graphene surface eventually, despite undergoing different conformational and orientational changes. We presented a representative simulation trajectory in which lysozyme firmly adsorbed on the graphene surface and utilized it for the graph clustering analysis. The entire simulation time was 5.0×10^6 time steps, including about 7.5 billion bond events and 2.5 billion collision events, for which the time scale of our simulations would be about 75 μs based on Marchut and Hall's estimation.[51] Our simulations showed that the temporal profile of the lysozyme's potential reached a plateau for more than 1.0×10^6 time steps in the last period of simulation, which demonstrated that our simulation system had reached equilibrium (Figure S3 in the Supporting Information).

Figures 3 and 4 show the results of autoencoder-based graph clustering analysis of 5.0×10^6 time steps in the simulation sample. To find the optimal number of clusters, we used different numbers of clusters ranging from 1 to 20 and evaluated the silhouette scores (Figure 3a). It was found that $k = 2$ has the highest silhouette score, which is 0.412, but obviously, it is not precise enough to present the adsorption process. If $k = 2$, there are only two stages, i.e., protein in the bulk and protein on the surface, which cannot offer useful information about the structural evolution of lysozyme protein. Since the silhouette scores of $k = 3$, 4, and 5 are very close to each other, we also applied two other metrics, the Davies–Bouldin score and the elbow method, to select the best number of clusters (Figures S4 and S5 in the Supporting Information). According to the Davies–Bouldin score, the clustering is better if its score is lower. It was found that $k = 5$ with the lower Davies–Bouldin score exhibits better clustering than $k = 3$ and 4 (Figure S4 in the Supporting Information). Using the elbow method, the best value is the first k after which the line becomes linear. Figure S5 in the Supporting Information shows that $k = 5$ is the transition point, after which the score curve becomes linear. Figure 3b illustrates the t-distributed stochastic neighbor embedding (t-SNE) of our stratification result, which projected 645 (129 × 5)-dimensional features into a 2-dimensional plane. The contact maps of the same stage formed clear clusters, and there was an obvious distinction among the five stages, indicating that our stratification method had a good understanding of the data and was capable of distinguishing between the contact maps. Notably, the autoencoder played a significant role in reducing the degree of freedom, which extracted representative features from each contact map and guaranteed the success of the subsequent k-means analysis. As shown in Figure 2, the original contact map of lysozyme protein contained 16 641 degrees of freedom, and after the autoencoder, it was reduced to a vector of 645 features. The feature extraction not only accelerated the efficiency of the K-means clustering algorithm but also improved the stability of the final stratification results.

As shown in Figure 4a, the transitions of folding stages occurred at time steps of 0.32×10^6, 1.50×10^6, 3.12×10^6, and 3.60×10^6, respectively, suggesting that there were significant changes in the protein structure. To facilitate our discussion, in chronological order, we rearranged the cluster IDs into five stages (1−5) of the protein adsorption process. We further calculated the Euclidean center of clusters and selected the contact maps closest to the corresponding cluster centers as the representative protein structure for each stage. The contact maps closest to the cluster centers were at 0.165×10^6, 1.102×10^6, 2.074×10^6, 3.263×10^6, and 4.656×10^6 time steps, as shown in Figure 4a. These contact maps and the corresponding molecular structures are shown in Figure 4b,c, respectively. It was observed that the contact maps and the configurations of stages 1, 4, and 5 showed noticeable pattern differences, whereas stages 2 and 3 displayed similarities to a certain extent. The similarity between stages 2 and 3 is also evidenced by the back and forth jumps around the time step of 1.50×10^6 in Figure 4a.

In stage 1 at $t = 0.080 \times 10^6$ (Figure 5a, also the purple dot in Figure 4a), the residues of Arg128 and hydrophobic Leu129

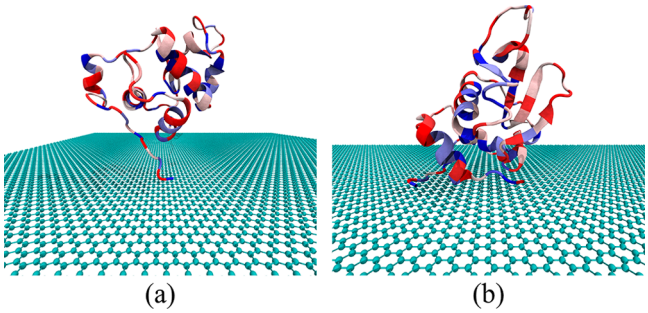


Figure 5. System snapshots at (a) $t = 0.080 \times 10^6$ and (b) $t = 0.135 \times 10^6$.

on the C-terminal of lysozyme came into contact with the graphene surface. Shortly after that, at $t = 0.135 \times 10^6$, which was about the time of the cluster center of stage 1 at $t = 0.165 \times 10^6$, the residues of Ala122-Cys127, which contain Trp123 with an aromatic group of strong π–π interactions with the graphene surface, further landed on the surface. In the meantime, the residues of Val2-Cys6 on the N-terminal were also adsorbed, including the hydrophobic Val2, Phe3 (also

https://doi.org/10.1021/acs.langmuir.2c01331
Langmuir 2022, 38, 10817−10825

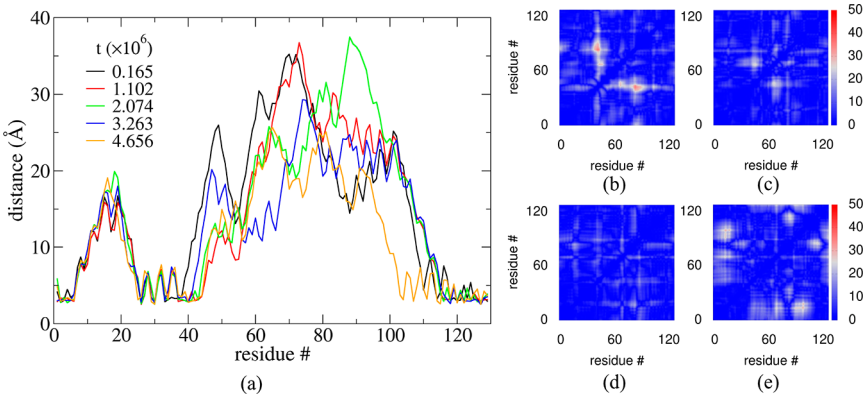


Figure 6. (a) Distance between each residue in lysozyme to the graphene surface at the five cluster centers. RMSD of the contact maps between the neighboring cluster centers: (b) between stages 1 and 2, (c) 2 and 3, (d) 3 and 4, and (e) 4 and 5. The unit of distance is Å.

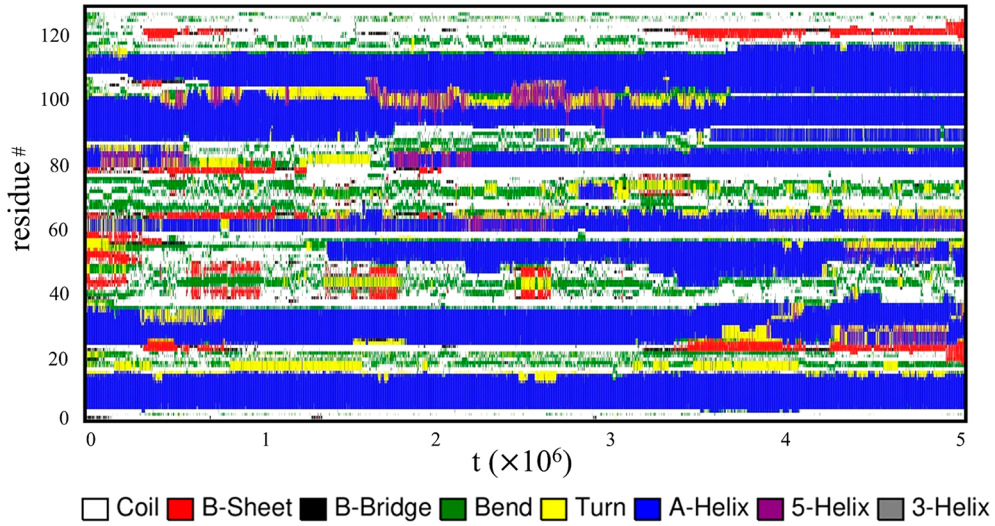


Figure 7. Time evolutions of the secondary structure of lysozyme during the adsorption process.

aromatic), and Cys6 (Figure 5b). Interestingly, it was found that the C- and N-terminals can behave like "anchors", which locked the lysozyme on the surface hereafter. As a result, the residues surrounding the "anchors" quickly fell on the surface in sequence, which could lead to the transition from stage 1 to stage 2.

Figure 6a demonstrates the distance of each residue of the lysozyme to the substrate surface, incorporating the overall effect of the protein's translational motion and rotational motion and the internal structural alternations. To clearly elucidate the internal structural changes upon adsorption and decouple the translational and rotational motions, we also analyzed the root-mean-square deviation (RMSD) of the contact maps for the neighboring stages (Figure 6(b–e)). At $t = 1.25 \times 10^6$, which was slightly after the time of cluster center of stage 2, the residues of Lys1-Cys6, Gly26-Ala42, and Cys115-Leu129 were adsorbed on the surface. Compared with stage 1, the residues of Thr40-Asn65 moved closer to the surface, while Cys80-Ala95 diffused in the opposite direction (Figure 6a). Correspondingly, Figure 6b showed the hot spots of large RMSD values in the areas of two clusters (Asn39-Asn44 and Cys80-Ser91), both of which moved apart. In stages 2–4, more residues (i.e., Asn59-Leu75) moved toward the surface, while the residues of Cys80-Ala95 were moving back and forth from the substrate surface during the process (Figure 6a). However, as shown in Figure 6c,d, relatively small internal structural changes among those three stages (stages 2–4) were observed. This was also in agreement with the fluctuations at the transitions of 1.50×10^6 and 3.12×10^6 (Figure 4a). Several areas of large RMSD in the contact map were detected at stage 5 (the white spots of large RMSD values: Lys1-Gly22 and Asn37-Trp108 in Figure 6e), indicating that the tertiary structure of lysozyme had significantly collapsed. By tracking the simulation trajectory, more details could be observed to explain large structural changes at stage 5. At the beginning of stage 5 at $t = 3.71 \times 10^6$, Asn103-Thr118 with the α-helix structure fell on the surface, including the hydrophobic residues (Met105, Ala107, Val109, Ala110, and Cys115). In addition, the residues of Ala95-Gly102 also moved closer to the surface.

Figure 7 shows the time evolution of the secondary structures of lysozyme during the adsorption process. It was found that the β-sheet had more significant changes compared with the helical structures. The helical region of Cys80-Leu84 unfolded for some time but restored to its original structure later. Also, the residues in Gly49-Leu56 formed a new α-helix during the simulation. Compared to the significant collapse in the tertiary structure, the protein's secondary structure displayed smaller changes.

https://doi.org/10.1021/acs.langmuir.2c01331
Langmuir 2022, 38, 10817–10825

4. CONCLUSIONS

Continuous or discontinuous molecular dynamics simulations at the atomistic or coarse-grained scale have been intensively applied to the study of protein folding/unfolding in the bulk solution[16,29,52,53] or on substrate surfaces[9,11,17,19,20,27,31,54] for many biomedical and industrial applications. However, an analysis of the structural evolution using simulation trajectories remains difficult. Further quantitative analysis of simulation data can provide critical insights into the underlying biophysics and elucidate protein's structure−function relationship. In this work, we successfully applied the method of autoencoder-based graph clustering to analyze DMD simulation trajectories of protein adsorption using a model system that consisted of a lysozyme protein and a graphene surface in implicit water. Our in-house-developed DMD program, *sDMD*, by combining an all-atom molecular model for protein−protein interactions and a Gō-like CG model for protein−surface interactions, can effectively elucidate the dynamics of protein folding/unfolding and also accurately describe protein−surface interactions, particularly protein−surface hydrophobic interactions and $\pi-\pi$ stacking. Moreover, the sparse autoencoder-based graph clustering model was used to analyze the protein contact maps to identify the folding/unfolding states and illustrate the protein adsorption path.

The DMD simulations showed that a lysozyme protein could land on a hydrophobic graphene surface with its terminal groups, followed by subsequent structural changes due to the strong protein−surface hydrophobic interactions and $\pi-\pi$ stacking in Cys30-THR40. It was also found that the lysozyme experienced a collapse in its tertiary structure and mild changes in its secondary structure. The analyses of autoencoder-based graph clustering showed a five-stage unfolding path. By comparing with detailed analyses of protein structure, our study showed that the autoencoder-based graph clustering can provide sensitive and accurate detection of a protein's structural changes. The combination of deep graph clustering analyses and efficient DMD simulations developed in this work could be an important tool to provide insights into biomolecules' interfacial behaviors. It is also notable that the autoencoder-based graph clustering can be applicable to the analyses of other types of simulation trajectories, such as continuous atomistic MD and coarse-grained MD.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.langmuir.2c01331.

Schematic diagram of the all-atom protein model; potential profiles in the all-atom model; values of the other parameters used in the equation of $V_{surface}$; discretized potentials as a function of the distance of amino acid residues to the surface; temporal profile of the lysozyme's potential energy on the graphene surface; Davies−Bouldin score; and score calculated by the elbow method (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Yong Wei − *Department of Computer Science, High Point University, High Point, North Carolina 27268, United States*; Email: ywei@highpoint.edu

Minghan Chen − *Department of Computer Science, Wake Forest University, Winston-Salem, North Carolina 27109, United States*; Email: chenm@wfu.edu

Tao Wei − *Department of Chemical Engineering, Howard University, Washington, D.C. 20059, United States*; orcid.org/0000-0001-6888-1658; Email: tao.wei@howard.edu

Size Zheng − *College of Materials and Chemistry & Chemical Engineering, Chengdu University of Technology, Chengdu, Sichuan 610059, P. R. China*; orcid.org/0000-0001-5719-1918; Email: zhengsize19@cdut.edu.cn

Authors

Jing Chen − *College of Materials and Chemistry & Chemical Engineering, Chengdu University of Technology, Chengdu, Sichuan 610059, P. R. China*

Enze Xu − *Department of Computer Science, Wake Forest University, Winston-Salem, North Carolina 27109, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.langmuir.2c01331

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work used the Extreme Science and Engineering Discovery Environment (XSEDE) Bridges2-AI at the Pittsburgh Supercomputing Center through allocations IRI200024 and CIS200010. T.W. acknowledges grant support from the National Science Foundation (NSF 1943999) and is grateful for computational resources from the program of Extreme Science and Engineering Discovery Environment (XSEDE) and the Texas Advanced Computing Center (TACC). S.Z. acknowledges the financial support from the Natural Science Foundation of Sichuan Province (2022NSFSC1274).

■ REFERENCES

(1) Kolesnyk, I.; Konovalova, V.; Kharchenko, K.; Burban, A.; Knozowska, K.; Kujawski, W.; Kujawa, J. Improved Antifouling Properties of Polyethersulfone Membranes Modified with A-Amylase Entrapped in Tetronic® Micelles. *J. Membr. Sci.* **2019**, *570−571*, 436−444.

(2) Li, Q.; Chen, Y.; Bai, S.; Shao, X.; Jiang, L.; Li, Q. Immobilized Lipase in Bio-Based Metal-Organic Frameworks Constructed by Biomimetic Mineralization: A Sustainable Biocatalyst for Biodiesel Synthesis. *Colloids Surf., B* **2020**, *188*, 110812−110818.

(3) Liese, A.; Hilterhaus, L. Evaluation of Immobilized Enzymes for Industrial Applications. *Chem. Soc. Rev.* **2013**, *42*, 6236−6249.

(4) Nguyen, H. H.; Lee, S. H.; Lee, U. J.; Fermin, C. D.; Kim, M. Immobilized Enzymes in Biosensor Applications. *Materials (Basel, Switzerland)* **2019**, *12* (1), 121−155.

(5) Talbert, J. N.; Goddard, J. M. Enzymes on Material Surfaces. *Colloids Surf., B* **2012**, *93*, 8−19.

(6) Cui, J.; Yan, Y.; Such, G. K.; Liang, K.; Ochs, C. J.; Postma, A.; Caruso, F. Immobilization and Intracellular Delivery of an Anticancer Drug Using Mussel-Inspired Polydopamine Capsules. *Biomacromolecules* **2012**, *13* (8), 2225−2228.

(7) Wu, X.; Narsimhan, G. Effect of Surface Concentration on Secondary and Tertiary Conformational Changes of Lysozyme Adsorbed on Silica Nanoparticles. *Biochim. Biophys. Acta, Proteins Proteomics* **2008**, *1784* (11), 1694−1701.

(8) Li, S.; Mulloor, J. J.; Wang, L.; Ji, Y.; Mulloor, C. J.; Micic, M.; Orbulescu, J.; Leblanc, R. M. Strong and Selective Adsorption of

Lysozyme on Graphene Oxide. *ACS Appl. Mater. Interfaces* **2014**, *6* (8), 5704−5712.
(9) Wei, T.; Carignano, M. A.; Szleifer, I. Molecular Dynamics Simulation of Lysozyme Adsorption/Desorption on Hydrophobic Surfaces. *J. Phys. Chem. B* **2012**, *116* (34), 10189−10194.
(10) Jahan Sajib, M. S.; Wei, Y.; Mishra, A.; Zhang, L.; Nomura, K.-I.; Kalia, R. K.; Vashishta, P.; Nakano, A.; Murad, S.; Wei, T. Atomistic Simulations of Biofouling and Molecular Transfer of a Cross-Linked Aromatic Polyamide Membrane for Desalination. *Langmuir* **2020**, *36* (26), 7658−7668.
(11) Zhang, T.; Wei, T.; Han, Y.; Ma, H.; Samieegohar, M.; Chen, P.-W.; Lian, I.; Lo, Y.-H. Protein−Ligand Interaction Detection with a Novel Method of Transient Induced Molecular Electronic Spectroscopy (Times): Experimental and Theoretical Studies. *ACS Cent. Sci.* **2016**, *2* (11), 834−842.
(12) Wei, T.; Ma, H.; Nakano, A. Decaheme Cytochrome Mtrf Adsorption and Electron Transfer on Gold Surface. *J. Phys. Chem. Lett.* **2016**, *7* (5), 929−936.
(13) Wei, T.; Sajib, M. S. J.; Samieegohar, M.; Ma, H.; Shing, K. Self-Assembled Monolayers of an Azobenzene Derivative on Silica and Their Interactions with Lysozyme. *Langmuir* **2015**, *31* (50), 13543−13552.
(14) Nakano, C. M.; Ma, H.; Wei, T. Study of Lysozyme Mobility and Binding Free Energy During Adsorption on a Graphene Surface. *Appl. Phys. Lett.* **2015**, *106* (15), 153701−153704.
(15) Wei, T.; Ren, C. In *Polymer Science and Innovative Applications*; Al-Maadeed, M. A. A.; Ponnamma, D.; Carignano, M. A., Eds.; Elsevier: 2020; Chapter 6, pp 207−228.
(16) Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press: 2008.
(17) Jahan Sajib, M. S.; Sarker, P.; Wei, Y.; Tao, X.; Wei, T. Protein Corona on Gold Nanoparticles Studied with Coarse-Grained Simulations. *Langmuir* **2020**, *36* (44), 13356−13363.
(18) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The Martini Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812−7824.
(19) Sarker, P.; Sajib, M. S. J.; Tao, X.; Wei, T. Multiscale Simulation of Protein Corona Formation on Silver Nanoparticles: Study of Ovispirin-1 Peptide Adsorption. *J. Phys. Chem. B* **2022**, *126* (3), 601−608.
(20) Zheng, S.; Sajib, M. S. J.; Wei, Y.; Wei, T. Discontinuous Molecular Dynamics Simulations of Biomolecule Interfacial Behavior: Study of Ovispirin-1 Adsorption on a Graphene Surface. *J. Chem. Theory Comput.* **2021**, *17* (3), 1874−1882.
(21) Alder, B. J.; Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **1959**, *31* (2), 459−466.
(22) Voegler Smith, A.; Hall, C. K. Alpha-Helix Formation: Discontinuous Molecular Dynamics on an Intermediate-Resolution Protein Model. *Proteins: Struct., Funct., Bioinf.* **2001**, *44* (3), 344−360.
(23) Ding, F.; Borreguero, J. M.; Buldyrey, S. V.; Stanley, H. E.; Dokholyan, N. V. Mechanism for the Alpha-Helix to Beta-Hairpin Transition. *Proteins: Struct., Funct., Bioinf.* **2003**, *53* (2), 220−228.
(24) Ding, F.; Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. Molecular Dynamics Simulation of the Sh3 Domain Aggregation Suggests a Generic Amyloidogenesis Mechanism. *J. Mol. Biol.* **2002**, *324* (4), 851−857.
(25) Hernández de la Peña, L.; van Zon, R.; Schofield, J.; Opps, S. B. Discontinuous Molecular Dynamics for Rigid Bodies: Applications. *J. Chem. Phys.* **2007**, *126* (7), 074106−074112.
(26) Zheng, S.; Javidpour, L.; Shing, K. S.; Sahimi, M. Dynamics of Proteins Aggregation. I. Universal Scaling in Unbounded Media. *J. Chem. Phys.* **2016**, *145* (13), 134306−134313.
(27) Zheng, S.; Shing, K. S.; Sahimi, M. Dynamics of Proteins Aggregation. Ii. Dynamic Scaling in Confined Media. *J. Chem. Phys.* **2018**, *148* (10), 104305−104314.
(28) Zheng, S.; Sahimi, A.; Shing, K. S.; Sahimi, M. Molecular Dynamics Study of Structure, Folding, and Aggregation of Poly-Glycine-Alanine (Poly-Ga). *J. Chem. Phys.* **2019**, *150* (14), 144307−144312.
(29) Zheng, S.; Sahimi, A.; Shing, K. S.; Sahimi, M. Molecular Dynamics Study of Structure, Folding, and Aggregation of Poly-Pr and Poly-Gr Proteins. *Biophys. J.* **2021**, *120* (1), 64−72.
(30) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. Ab Initio Folding of Proteins with All-Atom Discrete Molecular Dynamics. *Structure* **2008**, *16* (7), 1010−1018.
(31) Wei, S.; Knotts, T. A. A Coarse Grain Model for Protein-Surface Interactions. *J. Chem. Phys.* **2013**, *139* (9), 095102−095111.
(32) Wang, S.; Sun, S.; Li, Z.; Zhang, R.; Xu, J. Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. *PLoS Comput. Biol.* **2017**, *13* (1), e1005324.
(33) Xu, J. Distance-Based Protein Folding Powered by Deep Learning. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (34), 16856−16865.
(34) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Zidek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577* (7792), 706−710.
(35) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with Alphafold. *Nature* **2021**, *596* (7873), 583−589.
(36) Ng, A. Sparse Autoencoder. *CS294A Lecture Notes* **2011**, *72*, 1−19.
(37) Tian, F.; Gao, B.; Cui, Q.; Chen, E.; Liu, T.-Y. Learning Deep Representations for Graph Clustering. *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*; AAAI Press: Québec City, Québec, Canada, 2014; pp 1293−1299.
(38) Bhowmik, D.; Gao, S.; Young, M. T.; Ramanathan, A. Deep Clustering of Protein Folding Simulations. *BMC Bioinf.* **2018**, *19* (Suppl 18), 484−495.
(39) Liu, Y.; Amzel, L. M. Conformation Clustering of Long Md Protein Dynamics with an Adversarial Autoencoder. *arXiv* 2018, https://arxiv.org/abs/1805.12313.
(40) Smith, S. C.; Ahmed, F.; Gutierrez, K. M.; Frigi Rodrigues, D. A Comparative Study of Lysozyme Adsorption with Graphene, Graphene Oxide, and Single-Walled Carbon Nanotubes: Potential Environmental Applications. *Chem. Eng. J.* **2014**, *240*, 147−154.
(41) Kubiak-Ossowska, K.; Cwieka, M.; Kaczynska, A.; Jachimska, B.; Mulheran, P. A. Lysozyme Adsorption at a Silica Surface Using Simulation and Experiment: Effects of Ph on Protein Layer Structure. *Phys. Chem. Chem. Phys.* **2015**, *17* (37), 24070−24077.
(42) Bai, Y.; Ming, Z.; Cao, Y.; Feng, S.; Yang, H.; Chen, L.; Yang, S.-T. Influence of Graphene Oxide and Reduced Graphene Oxide on the Activity and Conformation of Lysozyme. *Colloids Surf., B* **2017**, *154*, 96−103.
(43) Komorek, P.; Walek, M.; Jachimska, B. Mechanism of Lysozyme Adsorption onto Gold Surface Determined by Quartz Crystal Microbalance and Surface Plasmon Resonance. *Bioelectrochemistry* **2020**, *135*, 107582−107591.
(44) Zheng, S.; Javidpour, L.; Sahimi, M.; Shing, K. S.; Nakano, A. Sdmd: An Open Source Program for Discontinuous Molecular Dynamics Simulation of Protein Folding and Aggregation. *Comput. Phys. Commun.* **2020**, *247*, 106873−106886.
(45) Lazaridis, T.; Karplus, M. Effective Energy Function for Proteins in Solution. *Proteins: Struct., Funct., Bioinf.* **1999**, *35* (2), 133−152.
(46) Shirvanyants, D.; Ding, F.; Tsao, D.; Ramachandran, S.; Dokholyan, N. V. Discrete Molecular Dynamics: An Efficient and Versatile Simulation Method for Fine Protein Characterization. *J. Phys. Chem. B* **2012**, *116* (29), 8375−8382.

https://doi.org/10.1021/acs.langmuir.2c01331
Langmuir 2022, 38, 10817−10825

(47) Wang, Y.; Shao, Q.; Hall, C. K. N-Terminal Prion Protein Peptides (Prp(120−144)) Form Parallel in-Register B-Sheets Via Multiple Nucleation-Dependent Pathways. *J. Biol. Chem.* **2016**, *291* (42), 22093−22105.
(48) Zou, X.; Wei, S.; Jasensky, J.; Xiao, M.; Wang, Q.; Brooks Iii, C. L.; Chen, Z. Molecular Interactions between Graphene and Biological Molecules. *J. Am. Chem. Soc.* **2017**, *139* (5), 1928−1936.
(49) Wei, S.; Zou, X.; Tian, J.; Huang, H.; Guo, W.; Chen, Z. Control of Protein Conformation and Orientation on Graphene. *J. Am. Chem. Soc.* **2019**, *141* (51), 20335−20343.
(50) Humphrey, W.; Dalke, A.; Schulten, K. Vmd: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33−38.
(51) Marchut, A. J.; Hall, C. K. Effects of Chain Length on the Aggregation of Model Polyglutamine Peptides: Molecular Dynamics Simulations. *Proteins: Struct., Funct., Bioinf.* **2007**, *66* (1), 96−109.
(52) van Der Munnik, N. P.; Sajib, M. S. J.; Moss, M. A.; Wei, T.; Uline, M. J. Determining the Potential of Mean Force for Amyloid-Beta Dimerization: Combining Self-Consistent Field Theory with Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2018**, *14* (5), 2696−2704.
(53) Nakano, C. M.; Byun, H. S.; Ma, H.; Wei, T.; El-Naggar, M. Y. A Framework for Stochastic Simulations and Visualization of Biological Electron-Transfer Dynamics. *Comput. Phys. Commun.* **2015**, *193*, 1−9.
(54) Huang, H.; Zhang, C.; Crisci, R.; Lu, T.; Hung, H. C.; Sajib, M. S. J.; Sarker, P.; Ma, J.; Wei, T.; Jiang, S.; Chen, Z. Strong Surface Hydration and Salt Resistant Mechanism of a New Nonfouling Zwitterionic Polymer Based on Protein Stabilizer Tmao. *J. Am. Chem. Soc.* **2021**, *143* (40), 16786−16795.

https://doi.org/10.1021/acs.langmuir.2c01331
Langmuir 2022, 38, 10817−10825