MD Simulations of Sequence Tuned Peptides Supramolecular Structures on Graphene Interface

You-Hsin Chen

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Chemical Engineering

University of Washington

2021

Committee:

René Overney

Shaoyi Jiang

Program Authorized to Offer Degree:

Chemical Engineering

University of Washington

Abstract

MD Simulations of Sequence Tuned Peptides Supramolecular Structures on Graphene Interface

You-Hsin Chen

Chair of the Supervisory Committee:
Professor René Overney
Department of Chemical Engineering

Peptide conformations are pivotal for biomolecular interactions that determine to a great deal biofunctional activities. State-of-the-art experiments provide limited insight into molecular scale aspects of the weak interactions of these soft and complex macromolecules. This molecular dynamic simulation project utilizes microscopic data of directed evolution selected solid binding peptides to provide a fundamental understanding of the relationship between sequence, peptide structure and interfacial assembly on inorganic surfaces, such as graphene. Specifically, we use enhanced sampling computational methodologies to identify dominant structures of two peptides with similar sequence. With these dominant structures, we can identify the most stable peptide conformations on the mentioned inorganic substrate surfaces with relative orientational relationships and interactions that provide us with the necessary input to predict the patterning structure of single peptide phases, as well as immiscible

phase mixtures.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor, Professor René Overney, for his encouragement and guidance, Professor Shaoyi Jiang for being my committee member, my group members, Tyler Jorgenson, Huan-Jui Lee, Saransh Jain, Zhichen Nian, for their support and ideas, and my parents for their unconditional support and inspiration.

Chapter 1. INTRODUCTION

Since the discovery of graphene, an atomically thin sheet of sp^2 hybridized carbon[1], two-dimensional conducting materials have become of prominent interest for advanced bioelectronic and biomedical applications[2] such as diagnostics, drug delivery, and imaging. They possess exceptional electronic properties and are well suited for the development of highly sensitive and selective biosensors when paired with biomolecular adlayers[3]. Additionally, the molecular structure of the bio-nano interface is the key to define biodevice functionality. For instance, specific binding orientation provides signal transduction across the surface, while different orientations will block pathways and reduce the efficiency of the transduction. By manipulating the biomolecular orientation, conformation, and assembly structure of the adlayer, device functionality and performance can be fine-tuned[4].

A promising method to control molecular structure of the bioelectronic interfaces is self-assembly, which harnesses solid-binding peptides to self-assemble into ordered nanostructures. Christopher R. So *et al.* have discovered peptide assembly process on pyrolytic graphite[5]. The assembly mechanism can be separated into two sections, as shown in Figure 1. One is surface aggregation, which includes peptide binding and diffusion. The other is densification and ordering, involving self-assembly process.

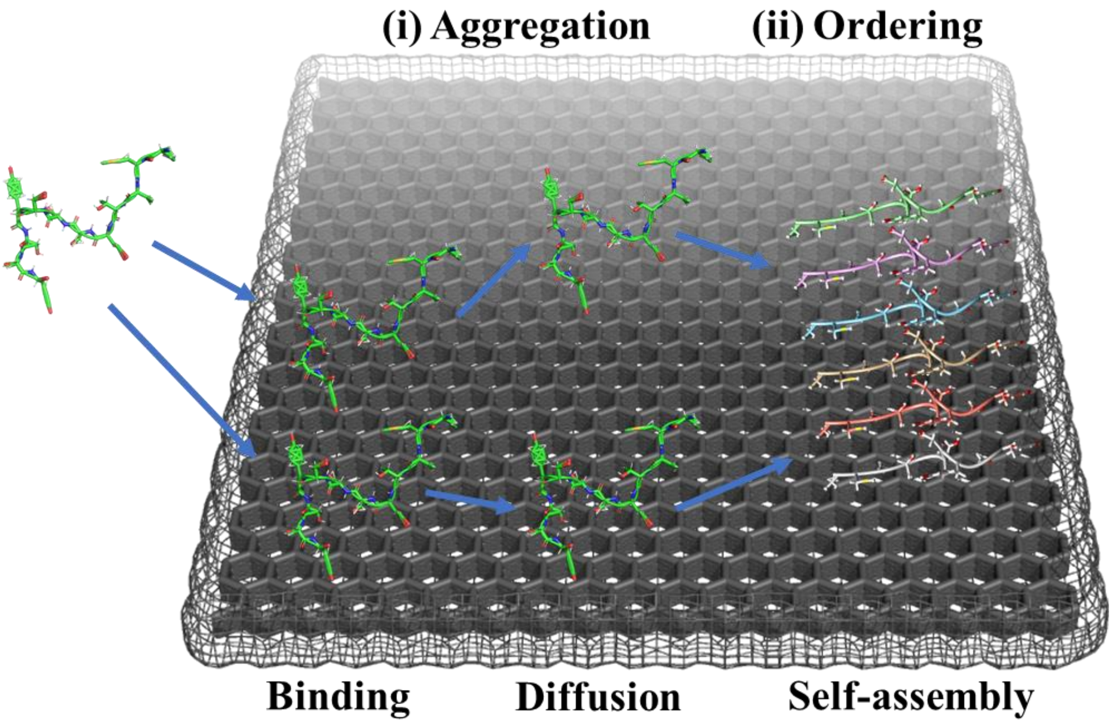


Figure 1. Schematics of the peptide self-assembly process on graphene: (i) aggregation involving binding, diffusion, and clustering processes and (ii) ordering involving self-assembly.

Polypeptides investigated in our group were GrBP5, selected combinatorically from millions of polypeptides for their relative binding strength to a wide variety of inorganic materials. For highly oriented pyrolytic graphite (HOPG), we found the wild-type sequence IMVTESSDYSSY (dubbed GrBP5-WT) one of the strongest binders of HOPG with a binding affinity and conformity that leads to organized self-assembly, as shown in Table 1. Inspired by this finding, we have exhaustively studied the mechanisms for conformationally mediated self-assembly of GrBP5-WT and its mutants, GrBP5-M2[6] to explore the effect of increased aromatic content, and, SS-GrBP5, which was chosen to explore assembly variations with only subtle chemical differences[7]. Based on our scanning force microscopy (SFM) studies, we found that GrBP5-WT and SS-GrBP5 exhibit distinctly different self-assembly structures, Figure 2. Based on its amino-acid sequence, SS-GrBP5 is more hydrophilic than GrBP5-WT while maintaining the amphiphilicity of GrBP5-WT. SFM scans, Figure 2, reveal similarly ordered, long nanowire-like self-assembly structures for both GrBP5-WT and SS-GrBP5 on HOPG.

Recently, the fabrication of binary assembled peptide has been attempted[8]. Two sequence differing peptides which can form ordered surface structures allow tuned mixed surface assembly with designed functionality[9]. As illustrated with GrBP5-WT and SS-GrBP5 peptides, Figure 2), one can infer that the wealth of peptide sequence space allows for the facile implementation of a great multitude of substrate and process tailored biomolecular self-assembly systems. As the co-organization of disparate materials into hierarchically assembled constructs is key to developing functional nano- and biotechnologies, it is only prudent to investigate co-assembly of peptides that have shown to be individually good assemblers for a chosen substrate. For that reason, we investigated the 50:50 mixture of GrBP5-WT and SS-GrBP5, and found them to co-assemble, Figure 2, however, revealing assembly patterns of distinct differences compared to the individual phase assemblies (Figure 2). The nanowires in the mixture are visibly shorter and narrower than that in the single phases. Our group member, Jorgenson, T. D., has observed that GrBP5-WT and SS-GrBP5 are immiscible and grow their long-range ordered structures competitively in the binary mixture[10]. We assume that these distinctions in structures result from different intermolecular interactions.

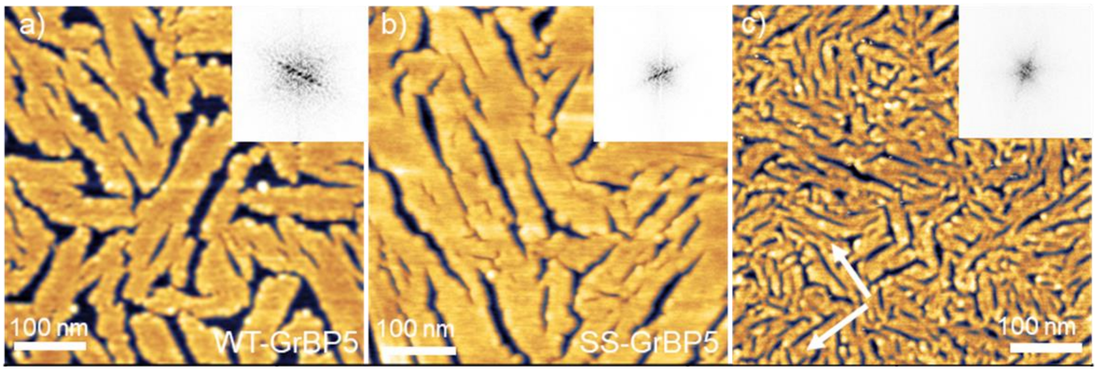


Figure 2. Self-assembled structure resulting from 2 μM (a) GrBP5-WT, (b) SS-GrBP5, and (c) 50:50 mixture of GrBP5-WT and SS-GrBP5 incubated on HOPG for 2 hours. All images are the same scale.

Table 1. Chemical information of GrBP5-WT and SS-GrBP5.

Name	**Sequence**	**MW**	**GRAVY**
GrBP5-WT	IMVTESSDYSSY	1381.5	-.242
SS-GrBP5	SSIMVTESSDYSSY	1555.6	-.321

The molecular weight (MW) in Daltons and the grand-average hydropathy (GRAVY). Color coding of the peptide sequence is as follows: green, hydrophobic; red, negatively charged; orange, hydrophilic; and blue, aromatic.

While the experimental study provides great insight into the binary peptide assembly of GrBP5-WT and SS-GrBP5, it leaves open details about the molecular co-arrangement that is essential to understand the interplay between substrate atomistic structure, peptide sequences and conformations for given solution and thermodynamic properties.

These open questions set the preamble for this thesis that is set out to investigate theoretically, based on molecular dynamic simulation, the molecular arrangement of the two identified polypeptides on graphene. We chose graphene over HOPG for reason of strong similarity regarding the surface atomistic structure, simplicity, as well as, its potential as extraordinary electronic material in the field of bioelectronics. Towards an improved understanding of the molecular self-assembly arrangement of GrBP5-WT and SS-GrBP5, we thoroughly investigate the molecular alignment of the peptides in both single phases, in correlation with the experimental SFM results. Specifically, we first determine the most preponderant peptide conformation in the solvent, then find the most likely orientation of the peptides in regards of graphene, followed by exploring the intermolecular interactions between peptides while binding, and consequently, determine the most likely angular relative arrangement of the peptides with the graphene lattice. Lastly, we explore the molecular assembly pattern that yields the angular relationships of the nanowire structures of the phase assemblies. The tools used for these tasks are *enhanced conformational sampling* for determining the most dominant peptide conformation, *classical molecular dynamics* (MD) simulation of single peptide system for proposing the most likely orientation of peptides, and classical MD simulation of duo-peptides system for investigating the intermolecular interactions between peptides.

Enhanced conformational sampling techniques are used explore the vast conformational space of a physical or chemical complex system. Thereby, metadynamics (MetaD)[11] is one of the most notable enhanced conformational sampling methods employed in computational physics, chemistry and biology. In MetaD, Gaussian potentials are added to the free-energy landscape repeatedly during the molecular dynamics simulation. The bias potential at time t is[12]:

$$V(s,t) = \omega \, exp\left[-\sum_{\alpha=1}^{N_{CV}} \frac{\left(s_\alpha - s_\alpha(t)\right)^2}{2\sigma_\alpha^2}\right] \qquad (1.1)$$

$$\omega = \frac{w_G}{\tau_G} \qquad (1.2)$$

where σ is the preassigned Gaussian width, ω is the deposition rate, w_G is the preassigned Gaussian height, and τ_G is the Gaussian deposit stride. When the bias potential is scaled down by a factor of $e^{-\frac{V(s(t),t)}{k_B\Delta T}}$, equation 1.1 can be modified to:

$$V(s,t) = \omega \, exp\left[-\frac{V(s(t),t)}{k_B\Delta T}\right] exp\left[-\sum_{\alpha=1}^{N_{CV}} \frac{\left(s_\alpha - s_\alpha(t)\right)^2}{2\sigma_\alpha^2}\right] \qquad (1.3)$$

where k_B is the Boltzmann constant, ΔT is an input parameter measured in temperature units. In equation 1.3, the bias potential is evaluated at the same point where the Gaussian is centered. Hence, the simulated system can spend more time on deeper wells and less time on the border of the explored region. This modified method is known as well-tempered metadynamics (WTMetaD)[13,14]. In MetaD, these added Gaussians discourage the system from revisiting visited configurations in the collective variable (CV) space. Thus, the simulated system can be pushed toward every metastable state in the free-energy landscape, determining the most stable conformation. With the most stable conformation as the starting point, classical MD simulation can be performed to estimate further peptide properties.

In light of what was said above, and discussed in greater detail in the following chapters, we begin our computation with sampling the peptide conformational space to acquire a valid starting point for MD simulations. In the literature, rigorous GrBP5-WT adsorption sampling has been performed using statistical analysis[15]. With WTMetaD sampling method, we investigate with high efficiency the most prominent peptide conformations in solution for the two peptides. Furthermore, to improve our understanding of the intricacy of the protein conformations in solution, small sequence differences and self-assembly behaviors, we examine two specific peptide properties, namely, the orientation and conformation. First, we determine the peptides' preferred orientations[16] with respect to graphene via classical MD simulations. Second, we investigated the stability of stable conformations of graphene-binding peptides using change of center of mass of peptides with time[17] and root mean square difference (RMSD) of atomic positions[18,19]. The experimental methodologies that were used in this thesis are described in greater detail in Chapter 2. The results and discussion follow in Chapter 3. Chapter 4 will summarize this work and provide and outlook on how to proceed based on the findings in this thesis.

Chapter 2. EXPERIMENTAL METHODS

2.1 Computational system

All simulations were performed via the GROMACS software[20], PLUMED plugin[21], CHARMM27 force field[22] for the peptides and buffering ions, and the TIP3P model[23] for water. Protein Data Bank (PDB) files were built through UCSF Chimera[24]. A graphene sheet is constructed to be 46×55 Å^2 using the built-in Nanostructures generator. All simulations were performed at 295 K, pH 7 and used right-angle parallelepiped simulation boxes with periodic boundary conditions. Integration was done using a Verlet algorithm[25]. Energy minimization of vacuum structure was done using a steepest descent algorithm for 5000 steps. After that, solvent was added to the peptide-graphene system followed by counterions addition which makes the net charge of the system neutral. The solvated system was relaxed again by performing 5000 iterations of steepest descent algorithm. The system was then equilibrated using NVT ensemble (constant number of particles, volume, and temperature) for 2.5 ps with the velocity rescaling thermostat[26] and time step of 2 fs followed by MD simulation.

2.1.1 WTMetaD simulation

WTMetaD simulations were performed using only radius of gyration of peptide as a collective variable. The peptide structure was oriented 10 Å above the graphene surface. The distance between the graphene sheet and the periodic boundary was 5 Å in the z-direction. WTMetaD simulations ran until convergence, which is 400 ns for WT and 300 ns for SS-GrBP5. A Gaussian is deposited every 500 time steps in WTMetaD. The Gaussian hill’s height and width is set to 2.0 kJ/mol and 0.01 nm respectively, whereas the biasfactor is set to 10.

2.1.2 Classical MD simulation

After WTMetaD reached convergence, the most stable adsorbed conformation on graphene was extracted and used as the starting point for classical MD simulations of single peptide system. The most dominant adsorbed structure obtained from single peptide simulation was then combined and applied to classical MD simulations of duo-peptides system. All classical MD simulations ran 100 ns.

2.2 Clustering algorithm

2.2.1 Gromos clustering algorithm

In WTMetaD simulations, we clustered peptide structures using the Gromos clustering algorithm[27] with a backbone RMSD cutoff of 2.0 Å. The Gromos algorithm takes a structure with the most neighbors within the given cutoff value as a cluster and identifies the center structure of each cluster. When the largest cluster is identified, it is eliminated from the pool of clusters. The process is repeated until all structures are distributed into clusters.

2.2.2 Ensemble clustering algorithm

For classical MD simulations, we applied an ensemble clustering algorithm using UCSF Chimera[28]. The ensemble clustering algorithm calculates RMSD by pairwise superposition of all structures. It merges two nearest clusters into a single cluster until reaching the minimum value of the penalty function, which is the cutoff point. Once the cutoff value is determined, it identifies the structure closest to the cluster centroid as the representative structure.

Chapter 3. RESULTS AND DISCUSSION

3.1 Conformational sampling

To identify representative conformation of both peptides as our starting point for classical MD simulation, we performed WTMetaD simulations using radius of gyration of both peptides as a collective variable. Because Gaussian potential was added to the energy landscape of the system repeatedly during the simulation, we had to determine when the WTMetaD simulations were converged. After WTMetaD simulations achieved convergence, we could extract the most stable adsorbed conformations for both peptides to run classical MD simulations.

3.1.1 Convergence analysis

When the system explores the full energy landscape, the change of free energy of adsorbed state approaches the value of thermal fluctuation as time evolves. In order to determine convergence of WTMetaD simulations, we calculated range of free energies for the last 10% of total simulation time. When WTMetaD simulations reached convergence, the range of free energies should be within the value of thermal energy k_BT (2.45 kJ/mol). Figure 3 shows free energy of adsorption as a function of simulation time for both peptides. As shown in Figure 3, range of free energies for the last 10% of total simulation is 2.06 kJ/mol for GrBP5-WT. The range is close to the value of thermal energy, meaning it could just achieve convergence. On the other hand, in Figure 3, range of free energies for the last 10% of total simulation is 0.29 kJ/mol for SS-GrBP5. The range is apparently smaller than the value of thermal energy, showing the simulation is not influenced by thermal effect greatly. Finally, the total simulation time for WT is 400 ns, and that for SS is 300 ns. We then clustered simulation results using the built-in Gromos clustering algorithm in the GROMACS.

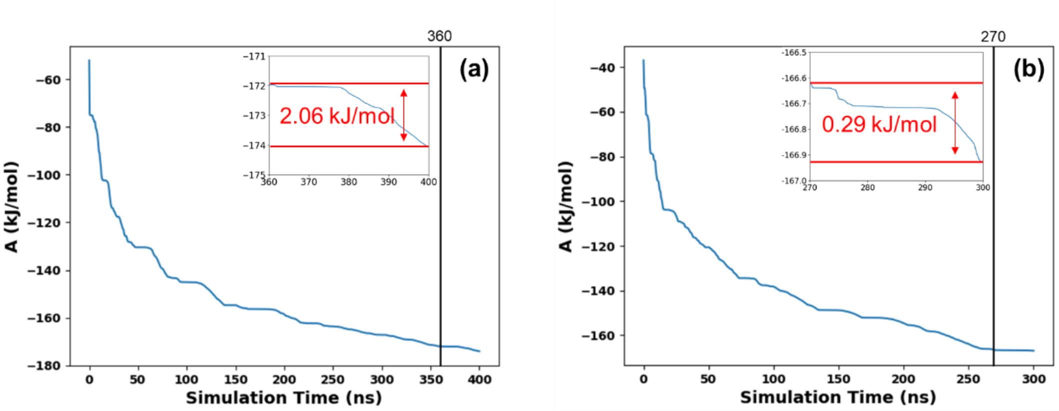


Figure 3. Time series of the free energy of adsorbed state at 295K, showing the convergence of (a) GrBP5-WT (b) SS-GrBP5. Insets show range of free energies for the last 10% of total simulation time.

3.1.2 Reweighted clustering analysis

Since both metadynamics simulations are converged, we can plot the free energy as a function of the collective variable, which is here the radius of gyration, as shown in Figure 4, and identify the most stable conformations for two peptides respectively. We observe that there are three evident wells for both peptides, implying that both peptides have three stable states. Theoretically, among these stable states, the most stable one should have the lowest free energy in the energy landscape, which is denoted by the black line in Figure 4.

For both peptides, we first clustered the conformational space to carry out a reweighted clustering analysis. The clusters with the lowest free energy in their adsorbed states are identified. Within these clusters, we then determined the representative peptide conformations. Clusters were identified for both peptides (Number of clusters: GrBP5-WT: 1378; SS-GrBP5: 1307) as shown in Appendix 2. GrBP5-WT having more clusters than SS-GrBP5 shows that GrBP5-WT is more flexible than SS-GrBP5. We computed free energies as a function of cluster index instead of radius of gyration of peptides so we could decide the cluster with the lowest free energy among all clusters. Figure 5 provides the free energy of each cluster for both peptides. As the cluster index increases, the number of structures within each cluster decreases. In Figure 5, we can identify the most stable conformations for GrBP5-WT in cluster number 23, whilst in Figure 5, the most stable conformations for SS-GrBP5 are found in cluster number 1. We extracted representative conformations of the two clusters and computed the radius of gyration of them, which are highlighted with orange lines in Figure 4. Clearly, the representative conformation with the lowest free energy (yellow line) is highly close to the stable conformation in theory (black line) for both peptides, suggesting the two conformations we obtained are the most stable ones. Figure 6 shows the most stable conformation for both peptides separately. Subsequently, we chose these two conformations as the starting point for classical MD simulations of single peptides to visit all possible peptide conformations.

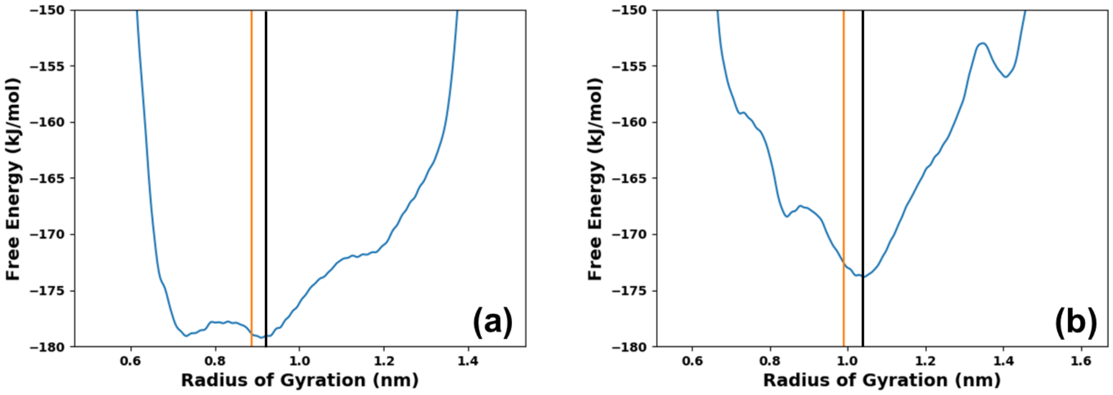


Figure 4. Free energy as a function of radius of gyration for (a) GrBP5-WT (b) SS-GrBP5. Black line: theoretically stable conformation; Orange line: stable conformation of the cluster with the lowest free energy.

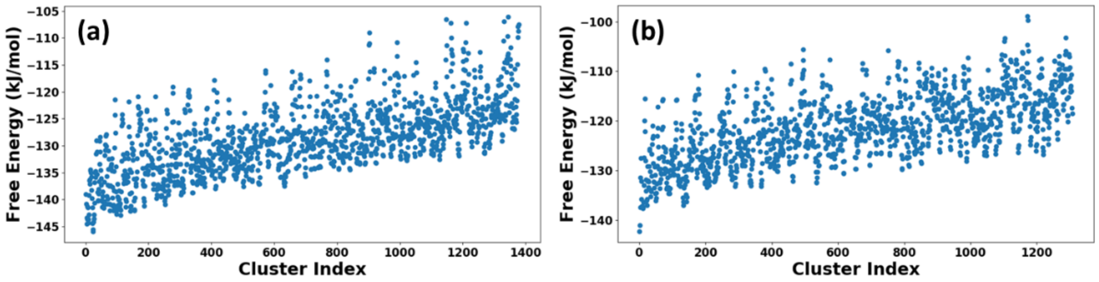


Figure 5. Free energy of (a) GrBP5-WT (b) SS-GrBP5 for each cluster index. Free energies were computed from metadynamics (MetaD).

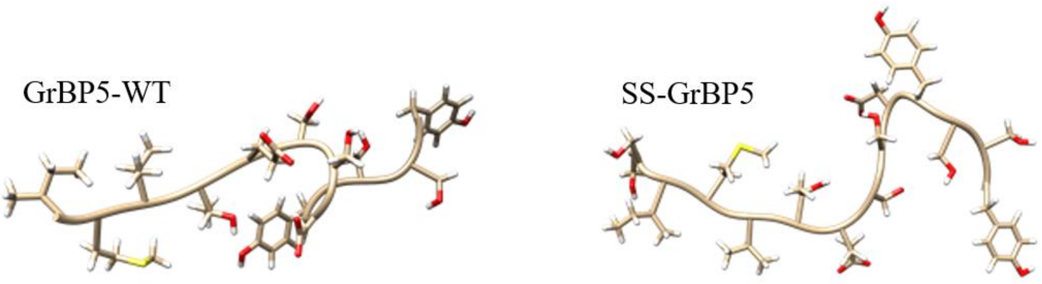


Figure 6. The most stable conformations for GrBP5-WT and SS-GrBP5.

3.2 Single-peptide system

We ran 100 ns classical MD simulations of single peptide systems for both peptides. Then, we clustered them using the built-in ensemble cluster method on Chimera to obtain peptide properties such as the dominant structures and peptide orientations. The information obtained from single peptide MD, was then used to investigate same peptide pairs (GrBP5-WT and SS-GrBP5). We further investigated peptide-peptide interactions from classical MD simulations of duo-peptides system.

3.2.1 Stability analysis

The time trajectory of RMSD illustrates the peptide structure deviation from a reference structure over time. Figure 7 shows time series of RMSD of atomic positions relative to the most stable structures for both peptides. Backbone RMSD for both peptides increases drastically at the beginning and keeps fluctuating in the classical MD simulation. That is, as soon as the simulations begin, the most stable structures of both peptides start undergoing conformational changes. More dramatic, as shown in Figure 7, the RMSD value of GrBP5-WT has a violent fluctuation for the first 10 ns, indicating that the starting conformation of GrBP5-WT simulation is not steady. We will discuss the unsteady conformation in terms of convergence of GrBP5-WT WTMetaD in greater detail later in the outlook section of Chapter 4.

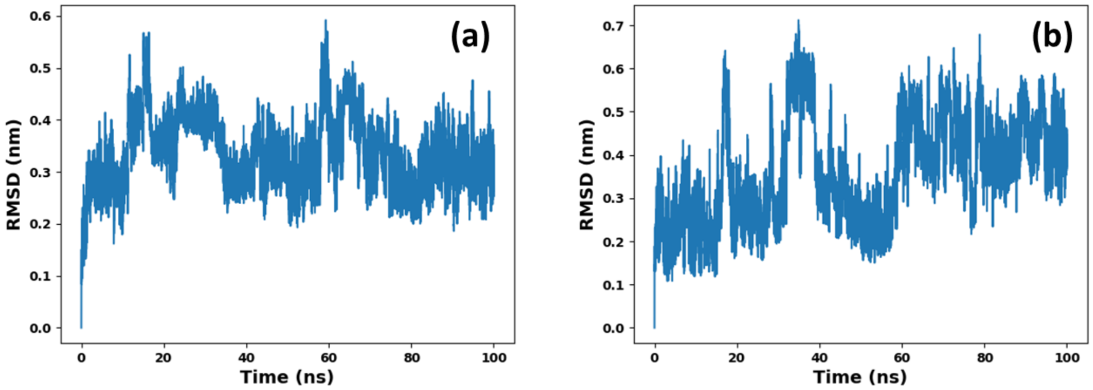


Figure 7. Backbone root-mean-square deviation (RMSD) of atomic positions relative to the structure present in the minimized and equilibrated system: (a) GrBP5-WT (b) SS-GrBP5.

3.2.2 Orientation analysis

The molecular orientation of the single peptides in respect to the atomistic structure of the graphene substrate is one of the pivotal parameters for the self-assembly organization into organized domains. In Figure 8, two graphene directions are highlighted, namely the *zig-zag* and the *armchair* direction. The nanowire-like orientations of the domains identified in the SFM plots of Figure 2 for GrBP5-WT and SS-GrBP5 are found to divert by $1^{\circ} \pm 7^{\circ}$ and $22^{\circ} \pm 6^{\circ}$ from the *zig-zag* direction, respectively. While GrBP5-WT nanowires assemble very closely along the *zig-zag* direction, SS-GrBP5 wires are more closely aligned, with a $8^{\circ} \pm 6^{\circ}$ offset angle, to the *armchair* direction. As the two peptides showed excellent co-assembly in the 50:50 mixture, Fig. 2c, and the finding that each assembly domain is pure (i.e., mixing does not occur on a molecular level[29]), the molecular orientation of the single molecules (GrBP5-WT and SS-GrBP5) prescribe the *complementary molecular assembly directions* of the two phase domains on graphene.

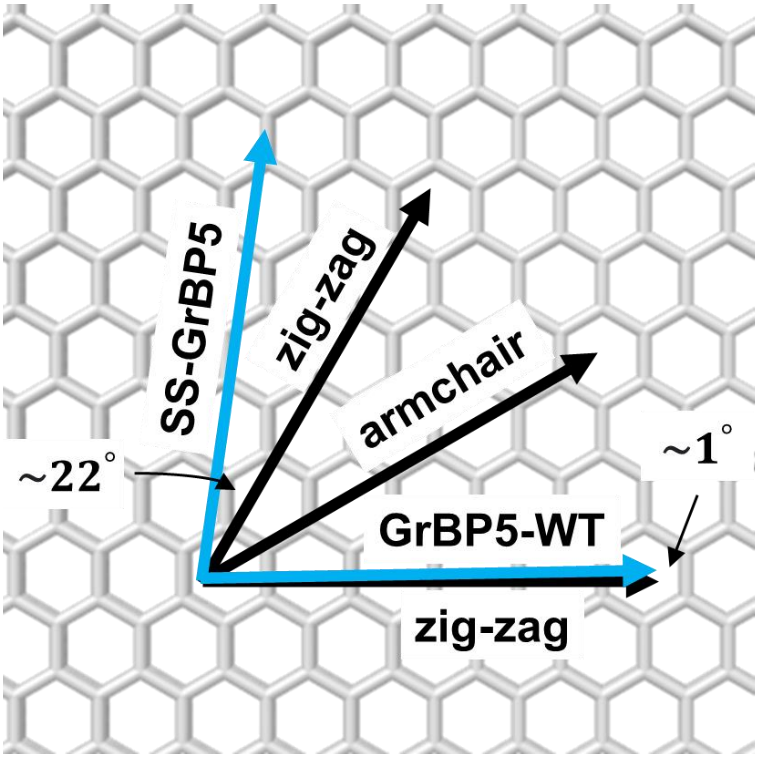


Figure 8. Nanowire-like domain orientations (blue) of GrBP5-WT and SS-GrBP5 contrasted to the *zig-zag* vector of graphene.

We set out to investigate the peptide orientations from a molecular perspective using classical MD simulations. By employing the ensemble clustering algorithm on UCSF

Chimera, we identified the average representative structures of the first two clusters for both peptides, as depicted in Figure 9. We then chose the long axis in the peptides, which is the N-C bond, to determine the peptide orientations with respect to the *zig-zag* vector of graphene, as revealed in Figure 9. The angle between the GrBP5-WT molecule and *zig-zag* vector of graphene is 37.9° for cluster 1, and 42.4° for the cluster 2, as indicated in Figure 9. On the other hand, the angle between the SS-GrBP5 molecule and *zig-zag* vector of graphene is 51.3° for cluster 1, and 39.5° for the cluster 2, as shown in Figure 9. Table 1 provides an orientation comparison between the SFM experiment and our simulation. As the N-C bond orientation was arbitrarily chosen to express the molecular orientation on the substrate, we cannot directly compare the experimental angles of the domains with the angles obtained from the molecular energy optimization presented in Figure 9.

To explore how our MD simulation fits the experimental domain structure, we inspected packing structures starting with surveying the periodic unit cell. We computed the average periodic unit length (5 nm) and width (5 nm) within peptide assembly from the SS-GrBP5 fast-Fourier Transform (FFT) filtered friction plot against nanowire distances, as shown in Figure 10. Using average structures of cluster 1 of both peptides, the most likely periodic unit cell structure for GrBP5-WT, based on aromatic-aromatic (yellow box) and hydrophobic (green box) interactions, and for

SS-GrBP5, based on aromatic-aromatic (yellow box) and hydrophilic (red box) interactions, and for the 50:50 mixture, based on aromatic-aromatic (yellow box), hydrophilic (red box) and hydrophobic (green box) interactions, were obtained, Figure 11. According to the peptide self-assembled structures (Figure 2) and the periodic units in Fig. 10, we provide in Figure 12 a conceivable patterning structures over an extended area (100 nm × 25 nm) for GrBP5-WT, SS-GrBP5. Considering that the 50:50 mixture of SS-GrBP5 and GrBP5-WT is immiscible on a molecular level, we present in Figure 13, a phase mixing pattern that matches the experimental data, while keeping the molecular orientations and unit cell structures intact. As we will discuss in the Chapter 4, the periodic units within these structures will require future multi-peptides simulations.

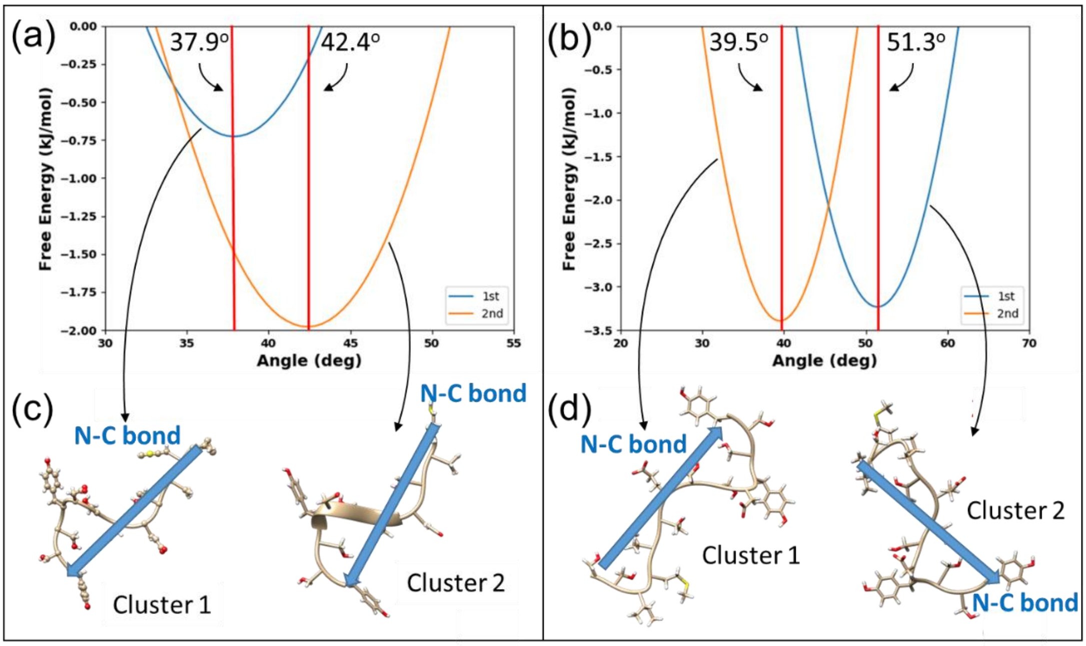


Figure 9. The orientation of N-C bond of the first two clusters with respect to the *zig-zag* direction of graphene: (a) GrBP5-WT (b) SS-GrBP5; (c) GrBP5-WT average structures of cluster 1 and cluster 2 (d) SS-GrBP5 average structures of cluster 1 and cluster 2.

Table 2. Orientation comparison between experiment and simulation

	GrBP5-WT (deg)	SS-GrBP5 (deg)	Difference
Experiment	1 +/- 7	22 +/- 6	23 +/- 13
Simulation (cluster 1)	37.9	51.3	13.4

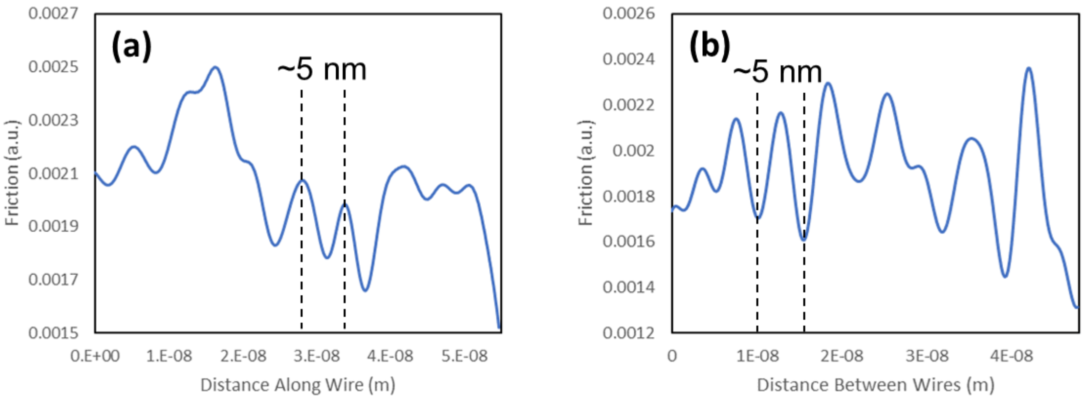


Figure 10. SS-GrBP5 FFT filtered friction as a function of distance (a) along nanowires (b) between nanowires. The distance between dashed lines corresponds to the average periodic unit length (width) within the peptide assembly.

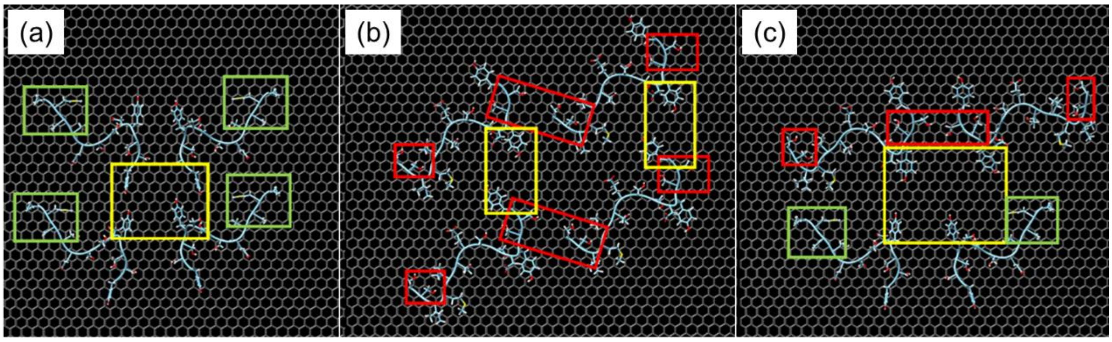


Figure 11. Proposed periodic unit of (a) GrBP5-WT (b) SS-GrBP5 (c) 50:50 mixture of SS-GrBP5 and GrBP5-WT within peptide assembly on graphene. Boxes show interactions between adjacent amino acids. Color coding of boxes is as follow: yellow: aromatic-aromatic interaction; green: hydrophobic interaction; red: hydrophilic interaction.

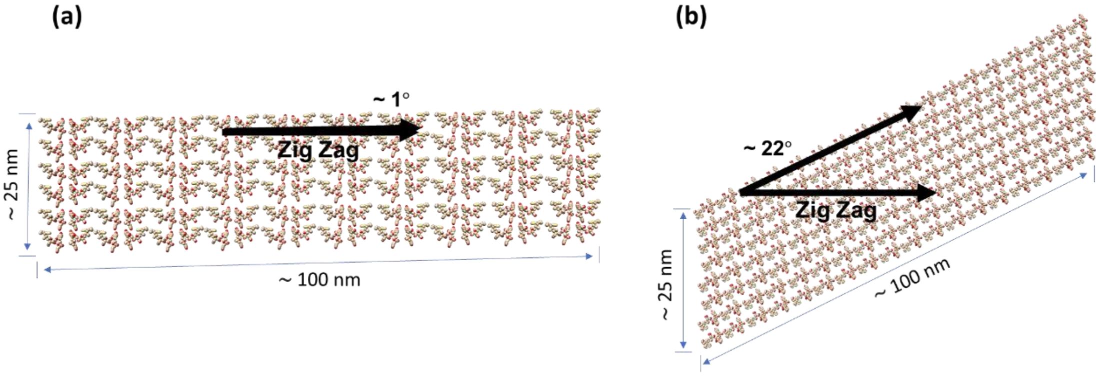


Figure 12. Proposed patterning structures of (a) GrBP5-WT (b) SS-GrBP5.

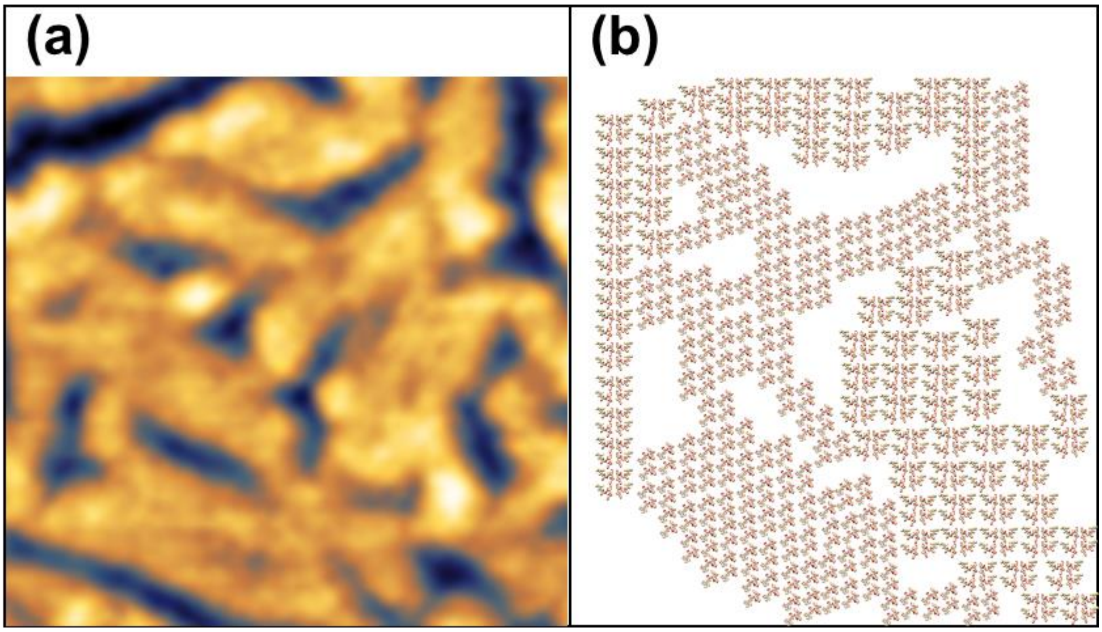


Figure 13. (a) AFM image (b) proposed patterning structure of co-assembly of SS-GrBP5 and GrBP5-WT.

3.3 Duo-peptides system

To investigate the interaction between peptides, we constructed three kinds of duo-peptides system, (1) GrBP5-WT and GrBP5-WT (WT-WT); (2) SS-GrBP5 and SS-GrBP5 (SS-SS); (3) SS-GrBP5 and GrBP5-WT (SS-WT), using structures of cluster 1 obtained from single peptide system, and then ran 100 ns classical MD simulations of duo-peptides system.

3.3.1 Intermolecular interaction analysis

We computed distance between center of mass (COM) of peptides in the duo-peptides system to identify interaction between peptides during the classical MD simulation, as shown in Figure 14. In Figure 14, the distance between COM of SS-SS illustrates a small fluctuation around 10 ns and the range from 60 ns to 80 ns (shown in the red boxes), indicating SS-SS interacts with each other while binding with the graphene substrate.

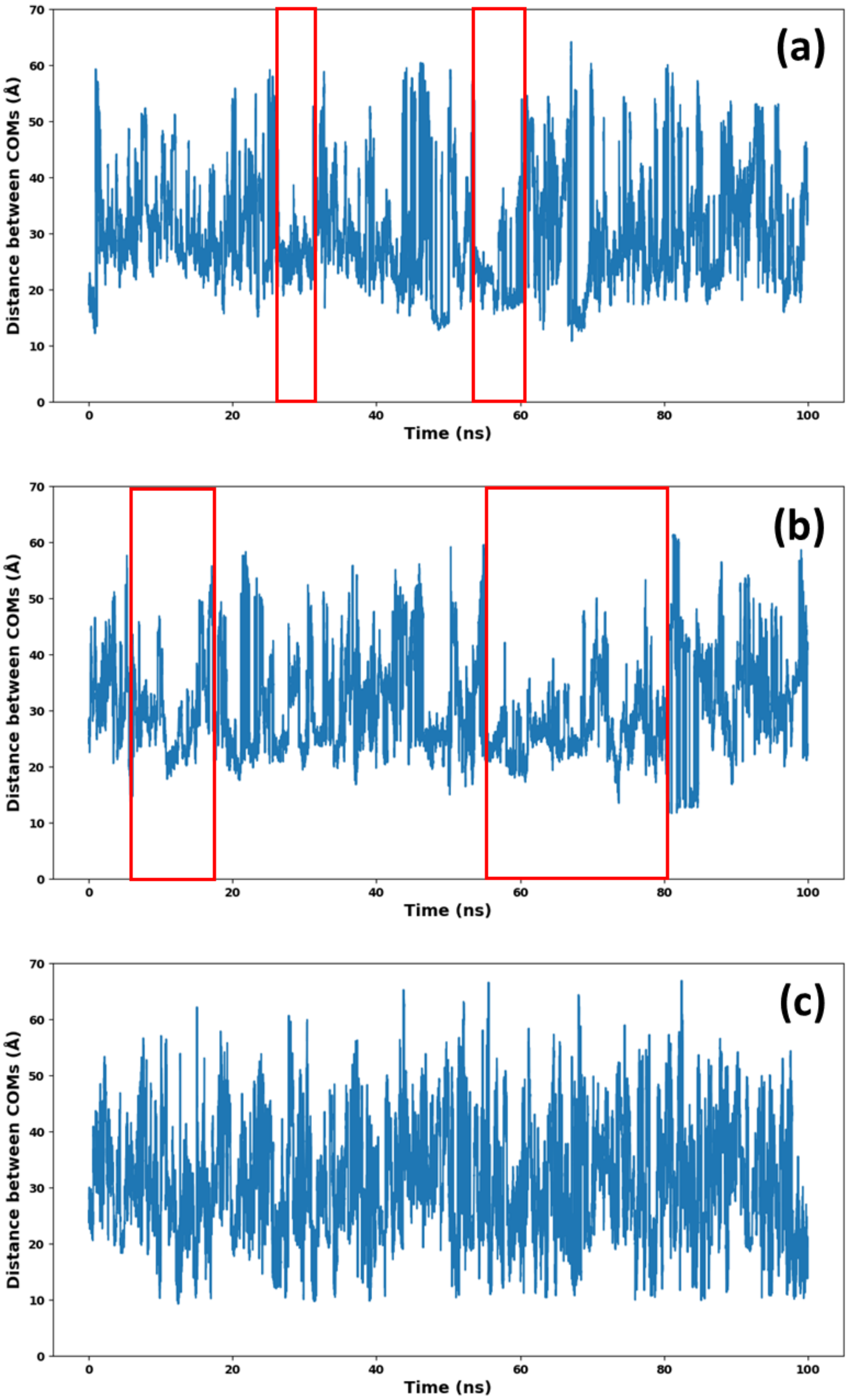


Figure 14. Distance between COMs in duo-peptides system: (a) WT-WT (b) SS-SS (c) SS-WT. Red squares show the area with small fluctuations.

Also, there are two small fluctuations for WT-WT around 30 ns and 60 ns (shown in the red box) in Figure 14, showing the WT-WT system exists intermolecular interactions. Comparing the COMs of the mixture to that of the single peptide system, there are always large fluctuations in the distance between COM of SS-WT during the whole simulations (Figure 14), suggesting the interactions between SS and WT are not obvious. The inconspicuous intermolecular interactions between the mixture indicate GrBP5-WT does not interact with SS-GrBP5 in the mixture. It implies interactions between WT-WT and between SS-SS are much stronger than interactions between WT-SS in the proposed patterning structure of mixture (Figure 12). Thus, the self-assembly process in the mixture can be divided into two parts, GrBP5-WT self-assembly and SS-GrBP5 self-assembly. These two assembly processes grow their own ordered structures competitively, supporting the shorter and narrower assembly structures of the SS-WT mixture (Figure 2c).

Chapter 4. CONCLUSION AND OUTLOOK

Motivated by experimental observations of different peptide conformations resulting from similar peptide sequence, we set out to investigate on a molecular scale of peptide properties and interactions between peptides. To summarize, we first used well-tempered metadynamics (WTMetaD) to enhance the conformation sampling of our simulations. We then applied the Gromos clustering algorithm to WTmetaD simulation data and demonstrated the peptide conformations of the most stable cluster for GrBP5-WT (WT) and SS-GrBP5 (SS). These most stable conformations were the starting point for classical molecular dynamics (MD) simulations of single peptide system. In the single peptide system, the varying values of RMSD of atomic positions during classical MD simulations mean that the two conformations we extracted from WTMetaD were unstable. We later clustered single peptide simulation data, chose structures of the first two clusters, and showed orientations of these structures with respect to *zig-zag* vector of graphene. The angle difference between WT and SS from experiments is $23^{\circ} \pm 13^{\circ}$, and that from simulations is 13.4°. Combining friction images (Figure 10) and average structures of cluster 1, where most conformations are (Figure 9), we proposed periodic units for WT, SS, and 50:50 mixture of SS and WT (Figure 11). Based on peptide self-assembled structures (Figure 2), nanowire orientations (Figure 8), and the proposed periodic units (Figure 11), we offered patterning structures for WT, SS and the mixture (Figure 12 and Figure 13). After that, we created duo-peptides system using the structures of cluster 1 acquired from single peptide system. To determine intermolecular interactions, we computed distance between center of mass (COM) of peptides in the duo-peptides system. For WT-WT and SS-SS, small fluctuations in distance between COMs at certain range of simulation time indicate there are interactions between the same peptides. However, the distance between COMs for SS-WT fluctuates widely over the course of the simulation, suggesting there is no apparent interaction between WT and SS. The unnoticeable intermolecular interactions between the mixture indicate the two self-assembly peptides grow their own ordered structures competitively in the mixture, as depicted in Figure 13, supporting the shorter and narrower assembly structures of the SS-WT mixture (Figure 2). According to the results of duo-peptides system simulations, we will further create the contact map to identify interactions between amino acid residues[30] in the future. Also, since the time trajectory of RMSD shows the GrBP5-WT conformation, we acquired from WTMetaD is unsteady, we will extend WTMetaD simulations to sample the whole CV space to obtain the most stable peptide conformations. In addition, to further distinguish interactions between WT-WT, SS-SS, and SS-WT while binding with graphene concurrently, we will run classical MD simulations of multi-peptides using the periodic unit (Figure 11) within our proposed patterning structures (Figure 12 and Figure 13).

REFERENCES

1. Ferrari, A. C. *et al.* Raman Spectrum of Graphene and Graphene Layers. *Physical review letters*, 97, 187401 (2006). doi: 10.1103/PhysRevLett.97.187401.
2. Sarikaya, M., Tamerler, C., Jen, AY., Schulten, K. & Baneyx, F. Molecular biomimetics: nanotechnology through biology. *Nature Materials, 2*, 577–585 (2003). doi: 10.1038/nmat964.
3. Hayamizu, Y., So, C. R., Dag, S., Page, T. S., Starkebaum, D. & Sarikaya, M. Bioelectronic interfaces by spontaneously organized peptides on 2D atomic single layer materials. *Scientific Reports, 6*, 33778 (2016). doi: 10.1038/srep33778.
4. Li, B., Fooksa, M., Heinze, S. & Meiler, J. Finding the needle in the haystack: towards solving the protein-folding problem computationally. *Critical Reviews in Biochemistry and Molecular Biology, 53*(1), 1-28 (2018). doi:10.1080/10409238.2017.1380596.
5. So, C. R., Hayamizu, Y., Yazici, H., Gresswell, C., Khatayevich, D., Tamerler, C. & Sarikaya, M. Controlling Self-Assembly of Engineered Peptides on Graphite by Rational Mutation. *ACS Nano, 6*(2), 1648–1656 (2012). doi: 10.1021/nn204631x.
6. Jorgenson, T. D., Milligan, M., Sarikaya, M. & Overney, R. M. Conformationally directed assembly of peptides on 2D surfaces mediated by thermal stimuli. *Soft Matter, 15*, 7360–7368 (2019). doi: 10.1039/C9SM00426B.
7. Jorgenson, T. D., Yucesoy, D. T., Sarikaya, M. & Overney, R. M. Thermal Selection of Aqueous Molecular Conformations for Tailored Energetics of Peptide Assemblies at Solid Interfaces. *Langmuir, 36*(1), 318–327 (2020). doi: 10.1021/acs.langmuir.9b02425.
8. Li, P., Sakuma, K., Tsuchiya, S., Sun, L. & Hayamizu, Y. Fibroin-like Peptides Self-Assembling on Two-Dimensional Materials as a Molecular Scaffold for Potential Biosensing. *ACS Appl. Mater. Interfaces, 11*, 20670–20677 (2019). doi: 10.1021/acsami.9b04079.
9. Sun, L., Narimatsu, T., Tsuchiya, S., Tanaka, T., Li, P. & Hayamizu, Y. Water stability of self-assembled peptide nanostructures for sequential formation of two-dimensional interstitial patterns on layered materials. *RSC Adv., 6*, 96889 (2016). doi: 10.1039/c6ra21244a.
10. Jorgenson, T. D. (2020). *Molecular Engineering of Peptide-Inorganic*

Interfaces: From Aqueous Conformations to Complex Surface Assemblies (Doctoral dissertation) (Order No. 28261638). Available from Dissertations & Theses @ University of Washington WCLP; ProQuest Dissertations & Theses Global. (2492245355).

11. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *WIREs: Computational Molecular Science, 1*(5), 826-843 (2011). doi: 10.1002/wcms.31.
12. Parrill, A. L. & Lipkowitz, K. B. *Reviews in Computational Chemistry 28.* (2015) doi:10.1002/9781118889886.
13. McCarty, J. & Parrinello, M. A variational conformational dynamics approach to the selection of collective variables in metadynamics. *J. Chem. Phys., 147*, 204109 (2017). doi: 10.1063/1.4998598.
14. Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters, 100*, 020603 (2008). doi: 10.1103/PhysRevLett.100.020603.
15. Penna, M. J., Mijajlovic, M., Tamerler, C. & Biggs, M. J. Molecular-level understanding of the adsorption mechanism of a graphite-binding peptide at the water/graphite interface. *Soft Matter, 11*, 5192 (2015). doi: 10.1039/c5sm00123d.
16. Chen, J. *et al.* Building two-dimensional materials one row at a time: Avoiding the nucleation barrier. *Science, 362*(6419), 1135–1139 (2018). doi: 10.1126/science.aau4146.
17. MacKerell Jr., A. D. Molecular Dynamics Simulation Analysis of a Sodium Dodecyl Sulfate Micelle in Aqueous Solution: Decreased Fluidity of the Micelle Hydrocarbon Interior. *J. Phys. Chem., 99*, 1846–1855 (1995). doi: 10.1021/j100007a011.
18. Daura, X., Gademann, K., Jaun, B., Seebach, D., Gunsteren, W. F. V. & Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem., 38*, 236–240 (1999). doi: 10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M.
19. Cheng, X. & Ivanov, I. Molecular Dynamics. *Methods in molecular biology, 929*, 243-285 (2012). doi: 10.1007/978-1-62703-050-2_11.
20. Hess, B., Kutzner, C., van der Spoel, D. & Lindhal, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput., 4*, 435–447 (2008). doi: 10.1021/ct700301q.
21. Bussi, G. & Tribello, G. A. Analyzing and Biasing Simulations with PLUMED. *Biomolecular Simulations, 2022*, 529–578 (2019). doi: 10.1007/978-1-4939-

9608-7_21.

22. Bjelkmar, P., Larsson, P., Cuendet, M. A., Hess, B. & Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput., 6*(2), 459-466 (2010). doi: 10.1021/ct900549r.
23. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Theory Comput., 79*, 926–935 (1983). doi: 10.1063/1.445869.
24. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem., 25*, 1605-1612 (2004). doi: 10.1002/jcc.20084.
25. Grubmüller, H., Heller, H., Windemuth, A. & Schulten, K. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. *Molecular Simulation, 6*, 121–142 (1991). doi: 10.1080/08927029108022142.
26. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys., 126*, 014101 (2007). doi: 10.1063/1.2408420.
27. Deighan, M. & Pfaendtner, J. Exhaustively Sampling Peptide Adsorption with Metadynamics. *Langmuir, 29*(25), 7999–8009 (2013). doi: 10.1021/la4010664.
28. Kelley, L. A., Gardner, S. P. & Sutcliffe, M. J. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng., 9*(11), 1063–1065 (1996). doi: 10.1093/protein/9.11.1063.
29. Jorgenson, T. D., Zareie, H. M., Sarikaya, M. & Overney, R. M. Symbiotic assembly of peptide nano-mosaics at solid interfaces. *Nanoscale, 13*(16), 7735-7743 (2021). doi: 10.1039/D0NR07208G.
30. Krall, A., Brunn, J., Kankanala, S. & Peters, M. H. A simple contact mapping algorithm for identifying potential peptide mimetics in protein–protein interaction partners. *Proteins, 82*(9), 2253–2262 (2014). doi: 10.1002/prot.24592.

APPENDIX 1

Briefly outlined here is the metadynamic simulation approach using the GROMACS package and PLUMED plugin. For more information about specific options, refer to the respective documentation at https://manual.gromacs.org/documentation and https://www.plumed.org/doc.

Software download link:
(1) GROMACS: https://manual.gromacs.org/documentation.
(2) PLUMED: https://www.plumed.org/download.

Step 1: Set up a peptide graphene system.

Create peptide.gro and peptide.top files using CHARMM27 force field and TIP3P water model.

```
gmx pdb2gmx -f peptide.pdb -o peptide.gro -p peptide.top -ignh
```

The two files only contain information about the peptide. Graphene information has to be added. For .top file, add graphene topology to the .top file.

```
#include “./charmm27_iff.ff/gra_iff_lg.itp”
```

Also, add graphene to the molecules present.

```
GRA            1
```

For .gro file, specify the location within the box of the peptide and graphene.

```
gmx editconf -f peptide.gro -o peptide_new.pdb -box 4.67307 5.53800 3.00 -center
2.336535 2.769 1.50
gmx editconf -f gra_iff.gro -o gra_new.pdb -box 4.67307 5.53800 3.00 -center
2.336535 2.769 0.50
```

Concatenate the two new .pdb file and then reconvert to a .gro file.

```
gmx editconf -f peptide_gra.pdb -o peptide_gra.gro -box 4.67307 5.538 3.00 -noc
```

Step 2: Energy minimization of vacuum system.

To ensure that there is no steric clashed or inappropriate geometry, we perform an energy minimization on the structure.

```
gmx grompp -v -f minim.mdp -c peptide_gra.gro -r peptide_gra.gro -p
peptide_gra.top -o pepgra-EM-vac.tpr
gmx mdrun -v -deffnm pepgra-EM-vac -c pepgra-EM-vac.gro
```

Step 3: Solvent and counterions addition.

Add solvent to the peptide-graphene system.

```
gmx solvate -cp pepgra_EM_vac.gro -cs spc216.gro -p peptide_gra.top -o
pepgra_sol.gro
```

Add counterions to make the solvated system neutral.

```
gmx grompp -f ions.mdp -c pepgra_sol.gro -r pepgra_sol.gro
-p peptide_gra.top -o ions.tpr
gmx genion -s ions.tpr -o pepgra_sol_ions.gro -p peptide_gra.top -pname NA -nname
CL -neutral
```

Step 4: Energy minimization of solvated system.

Implement an energy minimization on the solvated system.

```
gmx grompp -v -f minim.mdp –c pepgra_sol_ions.gro -r pepgra_sol_ions.gro -p
peptide_gra.top -o pepgra-EM-sol.tpr
gmx mdrun -v -deffnm pepgra-EM-sol -c pepgra-EM-sol.gro
```

Step 5: Relaxation of solvent.

Relax the solvent to assure the solvent configuration matches the peptide.

```
gmx grompp -v -f pr.mdp -c pepgra-EM-sol.gro -r pepgra-EM-sol.gro -p
peptide_gra.top -o pepgra-PR.tpr
gmx mdrun -v -deffnm pepgra-PR -c pepgra-PR.gro
```

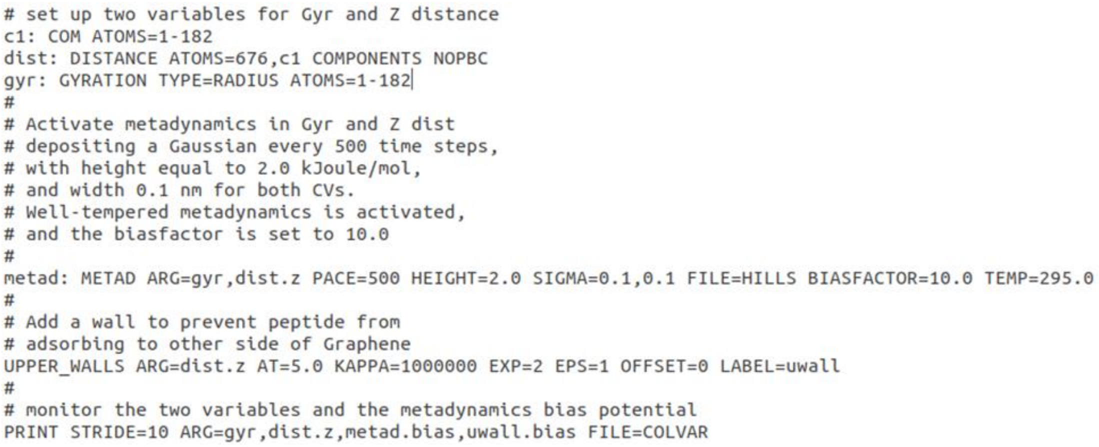
Step 6: Temperature coupling (NVT ensemble).

Conduct equilibration under NVT ensemble.

```
gmx grompp -v -f nvt.mdp -c pepgra-PR.gro -r pepgra-PR.gro -p peptide_gra.top -o
pepgra-NVT.tpr
gmx mdrun -v -deffnm pepgra-NVT -c pepgra-NVT.gro
```

Step 7: Metadynamics (Skip the step if run the classical MD simulation)

Generate .dat file using PLUMED commands to run metadynamics simulation. The following screenshot shows an example .dat file. For different peptides, these parameters must be changed.



```
# set up two variables for Gyr and Z distance
c1: COM ATOMS=1-182
dist: DISTANCE ATOMS=676,c1 COMPONENTS NOPBC
gyr: GYRATION TYPE=RADIUS ATOMS=1-182
#
# Activate metadynamics in Gyr and Z dist
# depositing a Gaussian every 500 time steps,
# with height equal to 2.0 kJoule/mol,
# and width 0.1 nm for both CVs.
# Well-tempered metadynamics is activated,
# and the biasfactor is set to 10.0
#
metad: METAD ARG=gyr,dist.z PACE=500 HEIGHT=2.0 SIGMA=0.1,0.1 FILE=HILLS BIASFACTOR=10.0 TEMP=295.0
#
# Add a wall to prevent peptide from
# adsorbing to other side of Graphene
UPPER_WALLS ARG=dist.z AT=5.0 KAPPA=1000000 EXP=2 EPS=1 OFFSET=0 LABEL=uwall
#
# monitor the two variables and the metadynamics bias potential
PRINT STRIDE=10 ARG=gyr,dist.z,metad.bias,uwall.bias FILE=COLVAR
```

Step 8: MD Simulation of equilibrated system.

Start the MD simulation.

```
gmx grompp -v -f md.mdp -c pepgra-NVT.gro -r pepgra-NVT.gro -p peptide_gra.top -o topol.tpr
gmx mdrun -plumed strongwall.dat -v
```

APPENDIX 2

Parameters used to run MD simulations:

1. minim.mdp

```
; Lines starting with ';' ARE COMMENTS
; Everything following ';' is also comment

title           = Energy Minimization   ; Title of run

; The following line tell the program the standard locations where to find certain files
cpp             = /lib/cpp      ; Preprocessor

; Define can be used to control processes
define          = -DFLEXIBLE

; Parameters describing what to do, when to stop and what to save
integrator      = steep         ; Algorithm (steep = steepest descent minimization)
emtol           = 10.0          ; Stop minimization when the maximum force < 1.0 kJ/mol
nsteps          = 5000          ; Maximum number of (minimization) steps to perform
nstenergy       = 1             ; Write energies to disk every nstenergy steps
energygrps      = System        ; Which energy group(s) to write to disk

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlist         = 1             ; Frequency to update the neighbor list
ns_type         = grid          ; Method to determine neighbor list (simple, grid)
coulombtype     = Reaction-Field ; Treatment of long range electrostatic interactions
epsilon_rf      = 78
rcoulomb        = 1.0           ; long range electrostatic cut-off
rvdw            = 1.0           ; long range Van der Waals cut-off
constraints     = none          ; Bond types to replace by constraints
pbc             = xyz           ; Periodic Boundary Conditions (yes/no)
periodic-molecules = yes
```

2. ions.mdp

```
; ions.mdp - used as input into grompp to generate ions.tpr
; Parameters describing what to do, when to stop and what to save
integrator  = steep         ; Algorithm (steep = steepest descent minimization)
emtol       = 1000.0        ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm
emstep      = 0.01          ; Minimization step size
nsteps      = 50000         ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlist         = 1         ; Frequency to update the neighbor list and long range forces
cutoff-scheme   = Verlet    ; Buffered neighbor searching
ns_type         = grid      ; Method to determine neighbor list (simple, grid)
coulombtype     = cutoff    ; Treatment of long range electrostatic interactions
rcoulomb        = 1.0       ; Short-range electrostatic cut-off
rvdw            = 1.0       ; Short-range Van der Waals cut-off
pbc             = xyz       ; Periodic Boundary Conditions in all 3 dimensions
```

3. pr.mdp

```
; VARIOUS PREPROCESSING OPTIONS
title                    = Position Restrained Molecular Dynamics
cpp                      = /lib/cpp
define                   = -DPOSRES

; RUN CONTROL PARAMETERS
integrator               = md
dt                       = 0.001 ; time step (ps)
nsteps                   = 2500  ; number of steps

; OUTPUT CONTROL OPTIONS
nstenergy                = 10


; NEIGHBORSEARCHING PARAMETERS
cutoff-scheme            = Verlet       ; Buffered neighbor searching
nstlist                  = 5
ns_type                  = grid
pbc                      = xyz
periodic-molecules       = yes

; OPTIONS FOR ELECTROSTATICS AND VDW
coulombtype              = PME
pme_order                = 4          ; cubic interpolation
fourierspacing           = 0.16       ; grid spacing for FFT
rcoulomb                 = 1.0
epsilon_rf               = 78
vdw-type                 = Cut-off
rvdw                     = 1.0

; Temperature coupling
tcoupl                   = v-rescale  ; Couple temperature according to modified Berendsen method
tc-grps                  = System
tau_t                    = 0.1        ; Coupling time constant, controlling strength of coupling
ref_t                    = 200         ; Temperature of heat bath

; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel                  = yes     ; Assign velocities to particles by taking them randomly from a Maxwell distribution
gen_temp                 = 200.0   ; Temperature to generate corresponding Maxwell distribution
gen_seed                 = 9999    ; Seed for (semi) random number generation. Different numbers give different sets of velocities

; OPTIONS FOR BONDS
constraints              = all-bonds ; All bonds will be treated as constraints (fixed length)
```

4. nvt.mdp

```
; VARIOUS PREPROCESSING OPTIONS
title                    = NVT simulation (constant number, volume and temperature)
cpp                      = /lib/cpp

; RUN CONTROL PARAMETERS
integrator               = md
dt                       = 0.002
nsteps                   = 1250

; OUTPUT CONTROL OPTIONS
nstxout                  = 0                       ; No output, except for last frame (coordinates)
nstvout                  = 0                       ; No output, except for last frame (velocities)
nstfout                  = 0                       ; No output, except for last frame (forces)
nstlog                   = 1                       ; Write every step to the log
nstenergy                = 10                      ; Write energies at every step
nstxtcout                = 0                       ; Do not write a compressed trajectory

; NEIGHBORSEARCHING PARAMETERS
cutoff-scheme            = Verlet       ; Buffered neighbor searching
nstlist                  = 10
ns-type                  = Grid
pbc                      = xyz
periodic-molecules       = yes

; OPTIONS FOR ELECTROSTATICS AND VDW
coulombtype              = PME
pme_order                = 4          ; cubic interpolation
fourierspacing           = 0.16       ; grid spacing for FFT
rcoulomb                 = 1.0
vdw-type                 = Cut-off
rvdw                     = 1.0
```

```
; Temperature coupling
tcoupl                  = v-rescale
tc-grps                 = System
tau_t                   = 0.1
ref_t                   = 295

; Pressure coupling
pcoupl                  = no

; OPTIONS FOR BONDS
constraints             = all-bonds
```

5. md.mdp

```
; VARIOUS PREPROCESSING OPTIONS
title                    = Production Simulation
cpp                      = /lib/cpp

; RUN CONTROL PARAMETERS
integrator               = md
tinit                    = 0       ; Starting time
dt                       = 0.002   ; 2 femtosecond time step for integration
nsteps                   = 50000000        ; 100ns

; OUTPUT CONTROL OPTIONS
nstxout                  = 50000 ; Writing full precision coordinates every nanosecond
nstvout                  = 50000 ; Writing velocities every nanosecond
nstfout                  = 0     ; Not writing forces
nstlog                   = 2500  ; Writing to the log file every step
nstenergy                = 2500  ; Writing out energy information every step
nstxtcout                = 2500  ; Writing coordinates every 5 ps

; NEIGHBORSEARCHING PARAMETERS
cutoff-scheme            = Verlet       ; Buffered neighbor searching
nstlist                  = 10
ns-type                  = Grid
pbc                      = xyz
periodic-molecules       = yes

; OPTIONS FOR ELECTROSTATICS AND VDW
coulombtype              = PME
pme_order                = 4         ; cubic interpolation
fourierspacing           = 0.16      ; grid spacing for FFT
rcoulomb                 = 1.0
vdw-type                 = Cut-off
rvdw                     = 1.0

; Temperature coupling is on
Tcoupl                   = v-rescale
tc-grps                  = System
tau_t                    = 0.1
ref_t                    = 295
; Pressure coupling is off
Pcoupl                   = no

; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel                  = no

; OPTIONS FOR BONDS
constraints              = all-bonds
constraint-algorithm     = Lincs
unconstrained-start      = yes
lincs-order              = 4
lincs-iter               = 1
lincs-warnangle          = 30
```

Clustering algorithm:

1. Gromos clustering algorithm

Backbone RMSD cutoff: 2.0 Å

Number of clusters: 1378 (WT); 1307 (SS)

Number of structures: 80001 (WT); 60001 (SS)

Number of structures in 1st cluster: 2530 (WT); 3346 (SS)

2. Ensemble clustering algorithm

Number of clusters: 43 (WT); 39 (SS)

Number of structures: 299 (WT); 299 (SS)

Number of structures in 1st cluster: 50 (WT); 60 (SS)

APPENDIX 3

Some alternative proposed periodic units and patterning structures are listed in Figure 15 and Figure 16. In Figure 15, we don't prefer this GrBP5-WT periodic unit because the interaction between amino acid in the left molecule and amino acid in the right molecule is not obvious. The proposed SS-GrBP5 periodic unit in Figure 16 is less preferred because its distances between amino acids are longer than those found in Figure 11, suggesting weaker hydrophilic and aromatic-aromatic interactions.

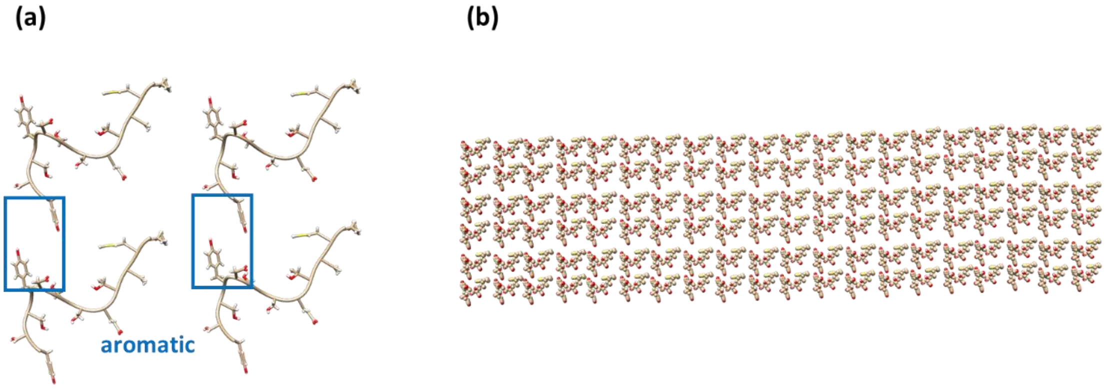


Figure 15. Alternative proposed (a) periodic unit (b) patterning structure of GrBP5-WT.

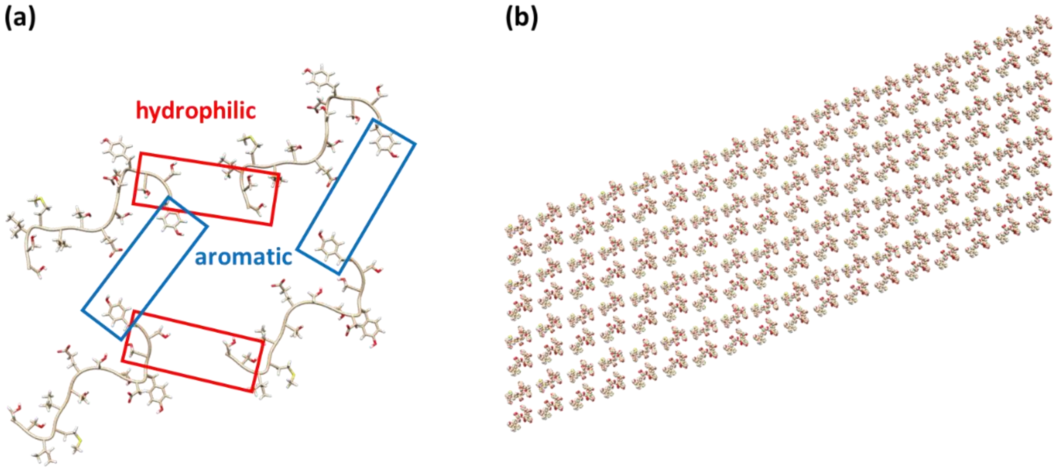


Figure 16. Alternative proposed (a) periodic unit (b) patterning structure of SS-GrBP5.