Exploring Serine-lysine Peptide Aggregation in Water and Interaction with Inorganic Surfaces Using Molecular Dynamics

Kejia Wu

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2018

Committee:

Jim Pfaendtner

David Beck

Program Authorized to Offer Degree:

Chemical Engineering

University of Washington

Abstract

Exploring Serine-lysine Peptide Aggregation in Water and Interaction with Inorganic Surfaces Using Molecular Dynamics

Kejia Wu

Chair of the Supervisory Committee:

Jim Pfaendtner

Chemical Engineering

This project is a combined experimental and computational study of peptide aggregation mechanism and biomineralization phenomena among serine-lysine (S-K) peptides, a series of small artificial peptides composed of serine and lysine residues solely with varied sequences. Inspired by the richness of both lysine and serine amino acids in biomineralization activities, four different S-K peptides were designed to examine their specific qualities for aggregation in water as well as roles in titania (TiO_2) or silica (SiO_2) precipitation. To mimic the experimental counterparts, TiO_2 and SiO_2 under neutral condition are chosen as the inorganic surface models here. Considering the time limitation and possible driving force, we applied our new enhanced sampling method (PBmetaD-PF) with identical components to provide new molecular-level insight into the aggregation process and structures. In second part, we did reweighting and

clustering analysis to find out dominant states and structures for each system, and investigated their interaction with inorganic surfaces after testing their stability respectively. Different binding forces between the two metal oxide surfaces were studied and discussed in detail.

Introduction

Biomineralization refers to a widespread phenomenon among living organisms from prokaryotes to human, producing biominerals with hierarchically structured organic-inorganic composites and directing hard tissue formation. In recent years, researchers have been intrigued by this field mainly due to our limited knowledge of its biological intricacy, the exceptional control over crystallinity and morphology as well as mild conditions required during the formation process, where some unique proteins proved critical in bio-templated synthesis of those inorganic nanostructured phases[1]. Among varied minerals found in biology, the co-precipitation of SiO_2 (silica) and TiO_2 (titania) has been the most popular topic with multiple applications in pigments, insulators, ceramics, coatings, biosensors, etc. [2-6]

One of the most well-studied silica-precipitating system is the diatom, a marine alga taking in silicon in the form of silicic acid ($Si(OH)_4$) as precursor and precipitates silica (SiO_2) in return, which is incorporated into the cell wall at last. The proteins implicated turn out essential in the whole process, such as silaffin Sil1p from the species marine diatom Cylindrotheca fusiformis[7-9], producing silica nanospheres. The primary structure of the silaffin Sil1p contains a sequence of seven homologous repeats R1−R7 between residues 108 and 271, which are rich in both serines and lysines. Noticebly, all the lysines are required to be alkylated while serines phosphorylated during the post-translational modification for the further transport and incorporation into the biosilica.[10] Among these seven units, the 19-amino acid peptide R5 shows similar in vitro silica nanosphere formation activity at neutral pH and without the need for any post-translational modifications in native silaffin (buffer with phosphate ions, however, would be necessary for successful silica precipitation).[11-13] Being identified the significance in both silica-precipitating

and titania-precipitating activities of many silaffins by Kroger and co-workers[14-15], the lysine-rich motifs drew more detailed attention. For example, pentalysine clusters, a five-lysine 12-14 amino acid peptide motif with the sequence KxxKxxKyKxxK (y could represent 1−3 different residues) are found prominent within all silaffins and efficient in silica production activity, especially in the N-terminal and C-terminal domains.[16] Inspired by massive mutation studies of silaffins and their fragments, the spacing of lysine clusters and interposed serine residues were carefully investigated. Kroger then designed an artificial peptide named PLCart (KSSKSSKSSKSSK) which also proved desirable silica-precipitating activity.[16]

Considering the progress mentioned above and elsewhere, it was hypothesized that serine and lysine rich motifs play important role in both biomimetic silicification and titania formation procedures. To narrow down the research scope to the effects of primary structures of serine-lysine (S-K) peptides such as peptide length, lysine-serine ratio and the contiguity of lysine residues on TiO_2 formation, a series of short serine-lysine fragments were custom-made by Buckle and coworkers[17]. Interestingly, only three of those peptides produced consistent TiO2 nanospheres with Trizma hydrochloride (Tris HCl) buffer at neutral pH, which are KSSKK, SKSK3SKS and SK5SK5 respectively. It was also tested by Buckle that even with high silica production efficacy in phosphate buffer, the structurally similar PLCart peptide could not precipitate titania in Tris buffer at all.[17]

During the whole biomimetic approach of TiO_2 formation by S-K peptides, the possible peptide aggregation phenomenon triggered intense interest. Specifically, based on a pair of comparisons between experimental site-specific ^{13}C CS assignments for both the neat form and TiO_2-

embedded form of peptide KSSKK and $SKSK_3SKS$, Buckle predicted the changes experienced in backbone carbons for secondary structure as well as in side-chains for side-chain orientation, which could be possibly attributed to peptide-peptide aggregation. Besides, an equilibrium between monomer-aggregate might also be a plausible factor behind the difference in the precipitation curves of KSSKK and $SKSK_3SKS$, which with a distinguished percentage of the number of monomers versus that of aggregates.[17] In general, protein aggregation is often correlated with many serious human diseases such as Alzheimer's disease (AD), type II diabetes (T2D), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS) and prion diseases.[18] In order to gain insights into the biological mechanisms and aggregate structures of peptides and proteins, a variety of computational work and experimental studies have already been performed. However, most of them were focused on the mechanisms of model systems with uniform sequences or with the potential to form stable beta-sheets such as amyloid peptides[19-21]. Also, the homogenous nucleation-controlled aggregation has been considered widespread and explored the most in the relevant theoretical and computational work recently.[22-24] Therefore, we would like to take a further look at the atomistic details for a more random aggregate – serine-lysine peptides outside of those classic model systems. In face of the difficulty in efficiently sampling the whole conformational phase space for complex biomolecular systems, we applied a novel enhanced sampling method name Parallel Bias MetaD with Partitioned Families (PBMetaD-PF) first time, which was designed out in our group as an extension to the standard Parallel Bias MetaD (PBMetaD) approach to expedite convergence of free energy landscapes. As proved from Prakash's work on small particle Lennard-Jones systems, PBMetaD-PF method converges much faster than PBMetaD, meanwhile recovering the same free energy profile and stable minima.[25]

Herein, using this improved enhanced sampling scheme in MD, we investigated the aggregation mechanisms for four different serine-lysine peptide sequences in water, namely, KSSKK, $SKSK_3SKS$, SK_5SK_5 and KSSKSSKSSKSSK, which could precipitate uniform morphology of TiO_2 - nanosphere with Tris buffer at neutral pH. Specifically, we determined the possibility of each aggregation state per peptide system and obtained free energy landscape individually after reweighting. With the clustering analysis conducted within lowest energy states, we found out two to four most dominant structures for each system and tested their stability in a 100-ns unbiased MD simulation. The results from clustering and reweighting also directed our conclusion towards a correlated relationship between aggregates formation and the arrangement of alternating serine and lysine residues, i.e., pentalysine clusters with equally interposed serine residues seem to form dimers much more easily. Inspired by the discoveries about the role of phosphate/phosphorylation of silaffins in silica-precipitation, we also tested the effects of Tris and confirmed that Tris could help stabilize the peptide aggregates in solution. In addition, in each peptide system, we further took out the most stable conformations of both monomers and dimers and checked the binding affinity with TiO_2 and SiO_2, respectively. The binding mechanisms were investigated and compared between these two systems in molecular-level details, which would be discussed further below. For KSSKK and $SKSK_3SKS$ system, the backbone chemical shifts of peptide-neat and peptide-composite were also calculated and compared with the corresponding experimental values, indicating a changed secondary structure occurred during the co-precipitation process and some unchanged aggregation states while approaching inorganic surfaces. Especially without the presence of phosphate ions, the understanding of this unique peptide-peptide interactions and peptide-surface interactions would shed light on the nature of new biomolecular phenomena.

Methods

Peptide Aggregation

To mimic the experimental conditions, four peptide sequences which are, KSSKK, $SKSK_3SKS$, SK_5SK_5 and KSSKSSKSSKSSK without capping groups as well as Trizma hydrochloride (Tris HCl) were built within the open source software program Avogadro[26]. Considering the trade-off between more comprehensive aggregation pathways explored and less computational resources used, a cubic box of ~7.0 x 7.0 x 7.0 nm^3 containing four each kind of peptides was chosen as our simulation system, with five Tris molecules being added to match with their concentration in experiment. This box size, for another, is large enough in all three dimensions so as to avoid the possible interactions between the peptide itself and its neighboring mirror (i.e., within the potential distance cutoff) under periodic boundary conditions (PBC) in our simulations. All these systems were generated with water molecules in Visual Molecular Dynamics (VMD[27]), ensuring each peptide, which was placed in a random starting configuration, was within a relative short distance (no larger than 1.0 nm) between the side chains of each other. Afterwards, additional chloride ions (termed counter-ions with the number from 17 to 45) were added to each system to keep the whole box being neutral. Specifications of each simulation could be found in Table 1.

Table 1. Details on the setup of MD simulations for peptide aggregation.

Simulation	Peptide	Buffer	Peptide Charge	Number of Peptide	Collective Varibale	Cl^- Ions	Number of Particles
1	KSSKK	none	+3	4	PSD (Peptide COM)	12	34,714
2	KSSKK	Tris	+3	4	PSD (Peptide COM)	17	34,720
3	$SKSK_3SKS$	Tris	+5	4	PSD (Peptide COM), Rg	25	33,168

4	SK_5SK_5	Tris	+10	4	PSD (Peptide COM), Rg	45	33,036
5	KSSKSSKS SKSSK	Tris	+5	4	PSD (Peptide COM), Rg	25	33,026
6	KSSKSSKS SKSSK	none	+5	2	PSD (Peptide COM), Rg	10	12,145
7	KSSKSSKS SKSSK	Tris	+5	2	PSD (Peptide COM), Rg	12	12,154

[c]PSD refers to the peptide separation distance of peptides.

The GROMACS 5.1.2 MD engine[28] was used to perform all MD simulations with in conjunction with the PLUMED 2.0 plugin[29]. The peptides and ions were modeled using the Amber14-sb force field[40] while water model was selected as the TIP3P model as in most other current work[30]. Energy minimization was performed on each solvated system using a steepest descent algorithm for 40,000 steps. Next, these minimized systems were then equilibrated for 1 ns under the NPT ensemble at 300 K and 1 bar (semi-isotropic coupling scheme), with position restraints (1000 kJ mol^{-1} nm^{-1}) applied on all the heavy atoms of the peptides. During this equilibration, temperature coupling was performed with the Donadio-Bussi-Parrinello thermostat[31] while pressure coupling with Berendsen barostat[32], respectively. In a subsequent equilibration stage, the same thermostat was applied in another 100 ps NVT simulation at 300 K. Lennard-Jones and Coulombic interactions were calculated with a cutoff length of 1.0 nm, and long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME) method. Interactions between hydrogen and other heavy atoms were constrained using the LINCS algorithm[33] to allow for a 2 fs time step to be used in all simulations.

The parallel bias metadynamics with partitioned families method, or PBMetaD-PF, is a new enhanced sampling algorithm which could efficiently overcome the challenges in timescale limitations or conformational space sufficient sampling with faster convergence speed compared

with standard parallel bias metadynamics (PBMetaD). With the detailed scheme and protocol for building this method described elsewhere[25], an overview of the application of this method will be presented here in our system. For the shortest peptide, which is, four KSSKK systems, only six sets of collective variables (CV) were biased in MD simulation, namely, the orthogonal distance between each two peptides' center-of-mass (COM) – d12, d13, d14, d23, d24 and d34, while ten CVs including four radius of gyration for each single peptide were biased all together in the rest three systems. This method was designed based on some preliminary results, which showed a better sampling efficiency while biasing more CVs as radius of gyration meanwhile. The bias potential was initiated with a Gaussian hill height of 2.0 kJ/mol and a hill deposition stride of 1000 steps. In addition, the Gaussian hill width, or σ value, was selected as 0.02 nm for distance of COM and 0.01 nm for radius of gyration for every single simulation, determined by averaging the half of the standard deviation of the CV calculated in the equilibrium NVT simulations mentioned above. The adjustable bias factor was set to 30 for all simulations except for the one of KSSKSSKSSKSSK, with a larger bias factor as 55 to separate the much more stable aggregates in an acceptable running time. Generally, all simulations were carried out to about 1 μs and were deemed converged within this timescale. Convergence was usually confirmed as the time at which the change in free energy landscape – projected onto a new CV as coordination number – between the monomer states and aggregates states ceased to change dramatically beyond the level of thermal fluctuations (kT~2.5 at 300 K). Further analyses were conducted to prove sufficient sampling while the clustering results and probability plots were used to test the convergence for our purpose of exploring aggregation states of S-K peptide systems (Figure 3).

Peptide Surface Interaction

Each simulation system in this part was mainly composed of one dominant structure (either a stable monomer or a stable dimer from the previous clustering results) and one inorganic surface (either SiO_2 or TiO_2). Detailed system information was included in Table 2. The systems were solvated by water and neutralized by chloride ions in the same way as before, while Tris buffer were also added with the concentration in proportion to that of the new box size to mimic the experimental environment. A harmonic restraint between one surface atom and COM of the peptide was placed at 3.0 nm above the surface through all simulations. This was done to prevent the peptides from binding or being attracted to the underside of the surface, which was possible when periodic boundary conditions (PBC) was used in all our work. In order to test the effect of initial peptide side-chain orientation during the interaction process, each conformation being put slightly above the specific surface, was rotated 90 degree with x axis and simulated in parallel with each other. Same forcefield and parameters for peptides, Tris, water and ions were used as above.

Table 2. Details on the setup of MD simulations for TiO_2 interaction.

Simulation	Peptide	State	Peptide Charge	Orientation	Surface	Surface Charge	Cl^- Ions	Number of Particles
1	KSSKK	monomer	+3	0	TiO_2	0	5	13,951
2	KSSKK	monomer	+3	90	TiO_2	0	5	13,728
3	KSSKK	dimer	+3	0	TiO_2	0	8	13,726
4	KSSKK	dimer	+3	90	TiO_2	0	8	15,610
5	$SKSK_3SKS$	monomer	+5	0	TiO_2	0	7	15,942
6	$SKSK_3SKS$	monomer	+5	90	TiO_2	0	7	15,940
7	$SKSK_3SKS$	dimer	+5	0	TiO_2	0	12	14,111
8	$SKSK_3SKS$	dimer	+5	90	TiO_2	0	12	14,079
9	SK_5SK_5	monomer	+10	0	TiO_2	0	12	13,921
10	SK_5SK_5	dimer	+10	0	TiO_2	0	22	13,783

11	KSSKSSKS SKSSK	monomer	+5	0	TiO_2	0	7	14,015
12	KSSKSSKS SKSSK	dimer	+5	0	TiO_2	0	12	14,220

Table 3. Details on the setup of MD simulations for SiO_2 interaction.

Simulation	Peptide	State	Peptide Charge	Orientation	Surface	Surface Charge	Cl^- Ions	Number of Particles
1	KSSKK	monomer	+3	0	SiO_2	-20	5	13,726
2	KSSKK	monomer	+3	90	SiO_2	-20	5	13,728
3	KSSKK	dimer	+3	0	SiO_2	-20	8	13,726
4	KSSKK	dimer	+3	90	SiO_2	-20	8	15,610
5	$SKSK_3SKS$	monomer	+5	0	SiO_2	-20	7	15,942
6	$SKSK_3SKS$	dimer	+5	0	SiO_2	-20	12	14,111
7	SK_5SK_5	monomer	+10	0	SiO_2	-20	12	14,629
8	SK_5SK_5	dimer	+10	0	SiO_2	-20	22	14,310
9	KSSKSSKS SKSSK	monomer	+5	0	SiO_2	-20	7	13,997
10	KSSKSSKS SKSSK	monomer	+5	90	SiO_2	-20	7	13,728
11	KSSKSSKS SKSSK	dimer	+5	0	SiO_2	-20	12	13,256
12	KSSKSSKS SKSSK	dimer	+5	90	SiO_2	-20	12	13,530

The force field developed by Matsui and Akaogi was used to model TiO_2[34], whereas the INTERFACE package developed by Heinz and coworkers - integrated into the Amber14-sb force field - was used to model SiO_2 and surface Na^+ ions[35]. Following the protocol used in our past studies[36,37], the (100) surface of rutile TiO_2 was simulated with terminal oxygen atoms (i.e.,

without OH termination) to mimic a titanium surface in aqueous media at a neutral pH. The surface orientation and degree of cleavage were chosen to be consistent with recent simulations of TiO_2 conducted by Brandt et al.[36], with tri- and doubly-coordinated oxygen atoms, and with five- and six-coordinated titanium atoms in the top/bottom and bulk layers of the surface, respectively. TiO_2 unit cell structure was downloaded from the Crystallography Open Database.[37,38] A TiO_2 supercell of dimensions 4 x 6 x 9 nm was built from unit cell belonging to space group P42/mnm, with cell parameters of a = b = 4.594 Å and c = 2.959 Å. The final cell size used in our MD simulations was ~5.9 x 5.9 x 1.7 nm. For SiO_2, a quartz (001) surface was simulated with a Q2 surface environment[39,40]. The surface was prepared from a 7 x 4 x 3 nm supercell built from an α-quartz (001) rectangular cell provided in the INTERFACE force field[41], resulting in a final size of ~6.9 x 6.8 x 2.0 nm used in the simulations. All silicon atoms were hydroxylated with two silanol groups ($=Si(OH)_2$), resulting in 9.4 silanol groups/nm^2. Approximately 1.0 silanol group/nm^2 was ionized to SiO^- with requisite numbers of surface sodium ions added to maintain a neutral interface, mimicking the surface chemistry of quartz in a neutral pH solution.

As for the simulation protocol, the equilibration steps were exactly same as what was done in peptide aggregation part. Thereafter, a 100-ns unbiased MD was conducted for each simulation box. The MD trajectories were used to test if those dominant structures in water would be stable enough on surface too, where individual interacting residues or groups were identified. Furthermore, the changes in backbone position and the compactness of secondary structure for all the four S-K peptides on surface and in water were also discussed based on the clustering results of MD trajectories, and the result of KSSKK and $SKSK_3SKS$ was further used to be a comparison with that from experimentalists to verify the pattern we observed and summarized here.

Results and Discussion

Reweighting and Convergence of MetaD Simulations

When examining the convergence of the biased MD simulations, we usually check (1) the shape of free energy curve over time and (2) the free energy difference between the distinguished state 1 and the distinguished state 2 (for example, the solvated state and the absorbed state in the surface-binding systems).Typically, the convergence was established until the free energy profile doesn't change significantly (changes no more than 1 kT) and the energy difference cease changing over time at a noticeable rate. However, in our cases, the original free energy landscape which was projected onto the distance of COM between each two peptides turned out somewhat ambiguous for the aggregation phenomenon we are interested in. (Figure 1)

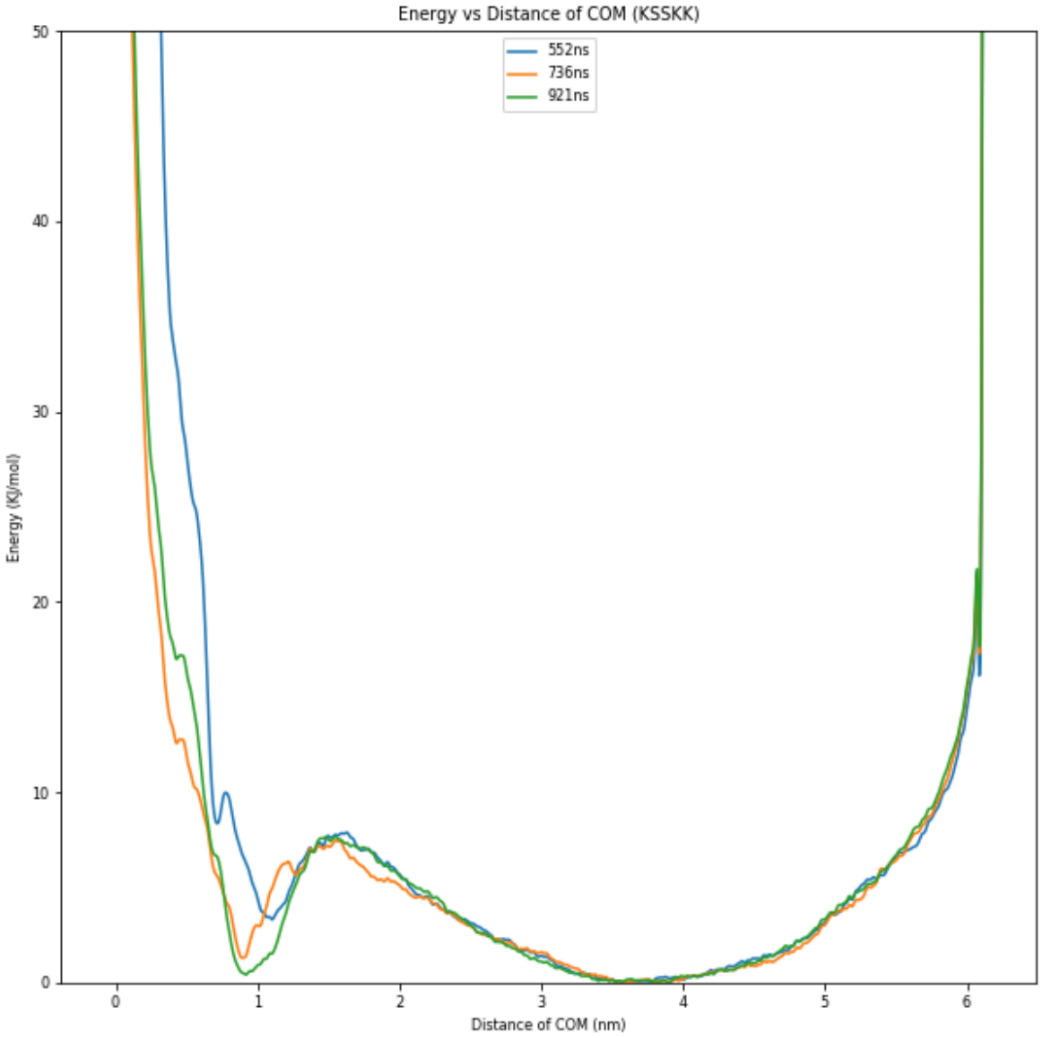

Figure 1. Aggregation of KSSKK with Tris: plot of the free energy projected onto the distance of center-of-mass between each two peptides.

More specifically, we are interested in finding out every possible aggregation states in our system as well as their corresponding potential energies. Despite we can clearly see there are low-energy state aggregates with distance of center-of-mass between 0.7nm to 1.4nm, it would be hard for us to identify the information for each individual aggregate state – which means, we are unable to determine ideally how many monomers form aggregates in this system. Therefore, we applied the reweighting analysis developed in PLUMED[42] to project our free energy profile onto a new CV –

COORDINATION between COM. A switching function was used to calculate the number of contacts between two custom-defined groups of atoms in this algorithm, which enables us to calculate the number of contacts within four peptides in one simulation box and obtain reweighted free energies for each of them. The function of COORDINATION is defined as follows:

$$\sum_{i \in A} \sum_{j \in B} sij$$

$$sij = \frac{1 - (\frac{rij - d0}{r0})^n}{1 - (\frac{rij - d0}{r0})^m}$$

where A and B denotes two group of atoms we defined, and the switching function makes sij differentiable. The default value for n parameter, m parameter and d0 parameter in switching function was used here, as 6, 12 and 0.0 respectively. After preliminary explorations about the appropriate parameters in the switching function, a r0 parameter as 1.5nm for KSSKK, 1.75nm for SKSK3SKS and 2.0nm for the other two peptides were finally determined for our work. Relevant free energy profiles are plotted in Figure 2, followed by a random sampling and checking the match between the value of corresponding new CV (COORDINATION) in COLVAR file and the distance between pairs of peptides in trajectories. With the purpose of investigating individual aggregation states, a probability plot of each aggregates was made to further confirm the convergence here (Figure 3), using a threshold of only considering two peptides are forming aggregates if half number of C alpha atoms in each peptide are counted as touching with a maximal distance of 0.5nm, 0.75nm, 0.85nm and 0.85nm for KSSKK, $SKSK_3SKS$, SK_5SK_5 and KSSKSSKSSKSSK respectively. All the distances were printed and selected using PLUMED. The convergence was assessed when the shape of probability profile remained unchanged during the last 30% simulation time.

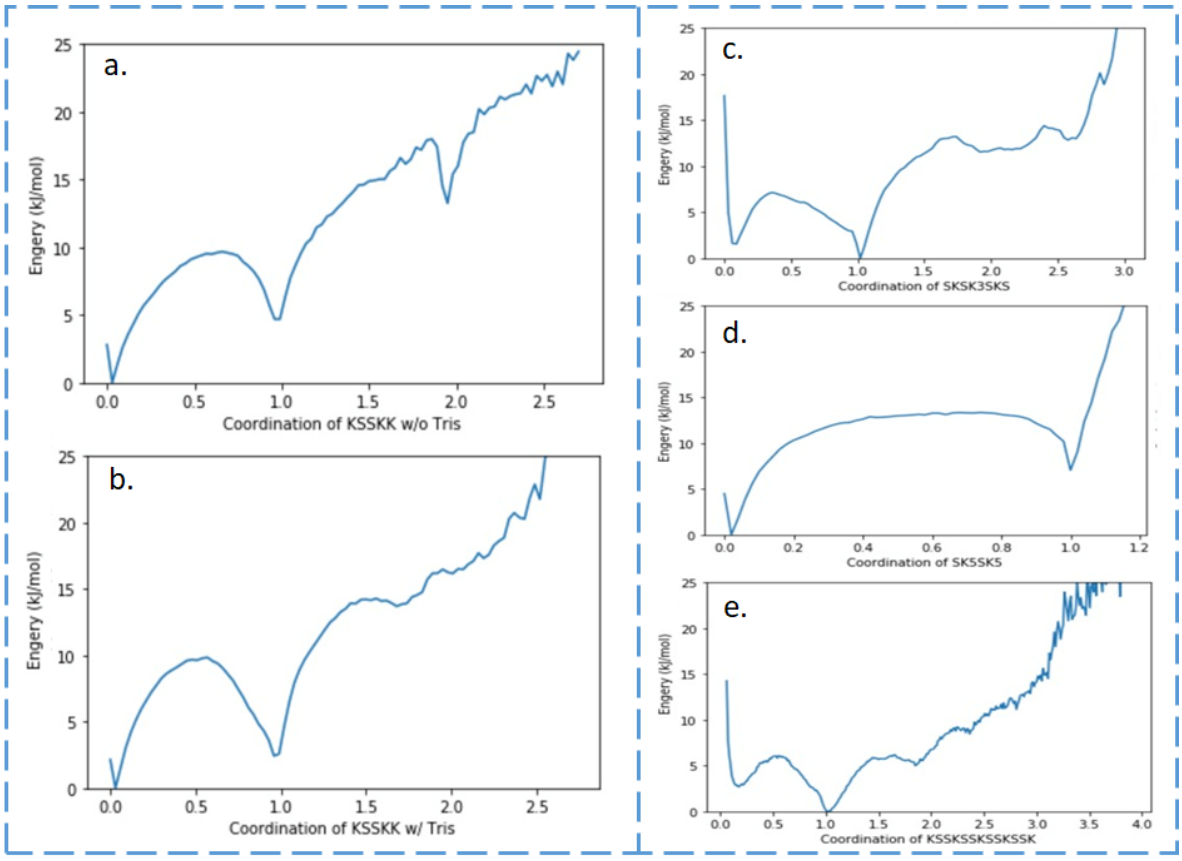

Figure 2. Free energy landscape for S-K peptides projected onto the coordination number of center-of-mass between each two peptides: (a) plot of the free energy surface for KSSKK without Tris system (r0=1.5nm); (b) plot of the free energy surface for KSSKK with Tris system (r0=1.5nm); (c) plot of the free energy surface for $SKSK_3SKS$ with Tris system (r0=1.75nm); (d) plot of the free energy surface for SK_5SK_5 with Tris system (r0=2.00nm); (e) plot of the free energy surface for KSSKSSKSSKSS with Tris system (r0=2.00nm).

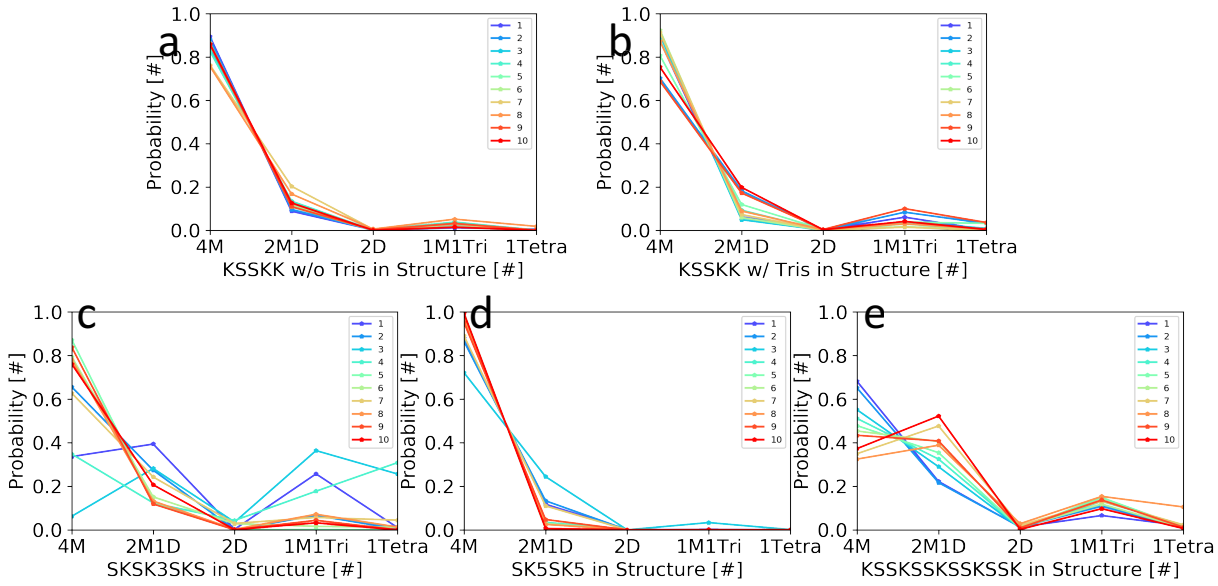

Figure 3. Probability plots for aggregation states of S-K peptides. 4M represents 4 monomers, 2M1D represents 2 monomers and 1dimer, 2D represents 2 dimers, 1M1Tri represents 1 monomer and 1 trimer while 1Tetra represents 1 tetramer. The right-top legend refers to the time evolution from 1 to 10 during the simulation. All probabilities were calculated from the biased MD without being reweighted yet and used to assess convergence of simulation.

Clustering and Dominant Structures

According to the free energy surfaces above, we divided the aggregation states of each system based on the kinks occurred in energy profiles and calculated the corresponding probability as shown in Table 4. On the basis of Table 4 and Figure 2, Tris was demonstrated to stabilize the dimers noticeably (increasing the percentage of dimer from around 22% to 40% and enhancing the binding affinity with about 3 kJ/mol), whose role was further visualized in detail from the next clustering result. Next, it was indicated that no aggregates formed in SK_5SK_5 peptide system, probably due to the comparative high energy barrier about 10 kJ/mol between monomers and

aggregates, which was obviously introduced by the strong electrostatic repulsion from the two clusters of five consecutive lysine residues within each one chain. On the contrary, peptide $SKSK_3SKS$ and KSSKSSKSSKSSK showed great potential to form clusters (mainly dimers) with the help of Tris buffer, with 56% and 65% respectively. It is also worth mentioning that except SK_5SK_5 peptide, monomers and dimers were almost always the two most favorable states for all these S-K peptides, without the tendency to grow into larger clusters anymore. As being discussed further below, this phenomenon might suggest a totally different aggregation mechanism than before.

Table 4. Probability table for each aggregation states.

Simulation	Peptide	% monomer	% dimer	% others
1	KSSKK w/o Tris	77	22	1
2	KSSKK w/ Tris	58	40	1
3	$SKSK_3SKS$ w/ Tris	44	54	2
4	SK_5SK_5 w/ Tris	91	6	3
5	KSSKSSKSSKSSK w/ Tris	35	63	2

To further explore the aggregation mechanisms and obtain dominant structures of S-K peptides, agglomerative clustering analyses were conducted within the energy minimum of both monomer and dimer for simulation 2, 3 and 5. For monomers with a coordination smaller than 0.87, the radius of gyration (Rg) was used as variables to be clustered as a good indicator of compactness or how the most popular monomers were folded in solution, which is defined as the root mean square distance from each atom of the protein to their centroid. For dimers with a coordination

ranging from 0.87 to 2.0, the coordination, especially the decimal part was used in clustering to find out the lowest energy state of dimer conformations. One representative structure was obtained from the largest cluster of each state for each peptide, with the details depicted below.

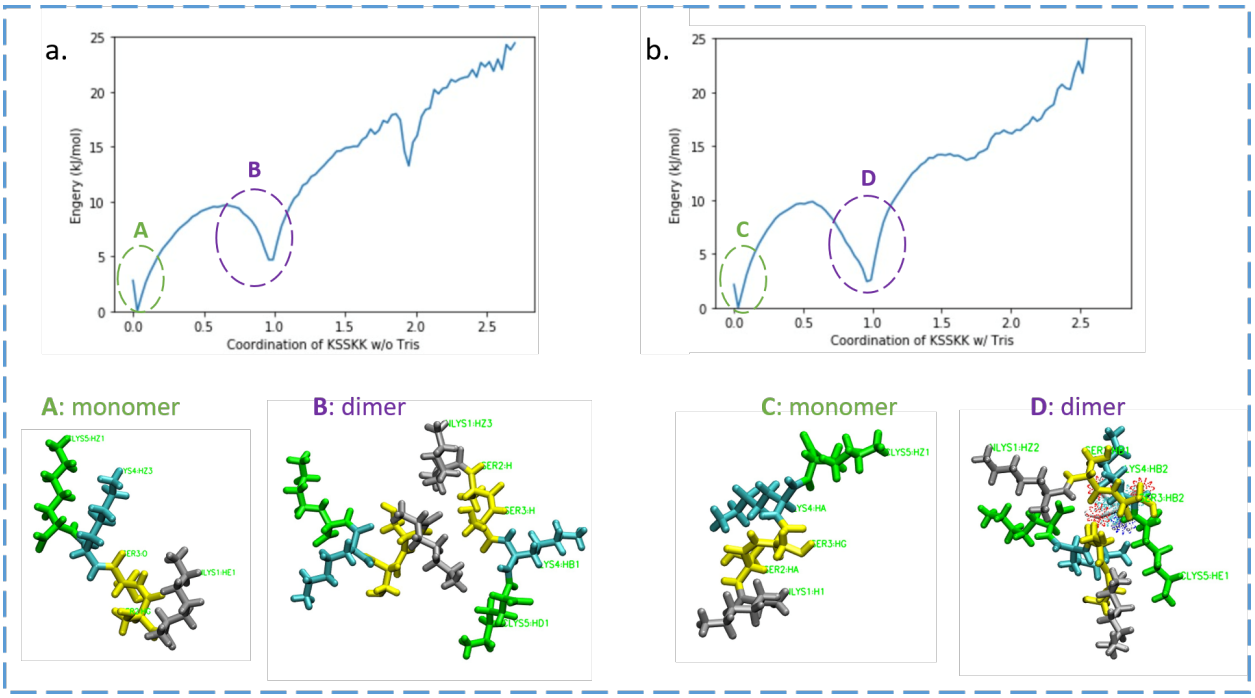

Figure 4. Dominant structures of KSSKK system for each energy minimum: (a) KSSKK without Tris system; (b) KSSKK with Tris system. Lysine residues, serine residues, N-terminal residues and C-terminal residues are shown in blue, yellow, gray and green coloring, respectively, and Tris is shown in color-dotted manner.

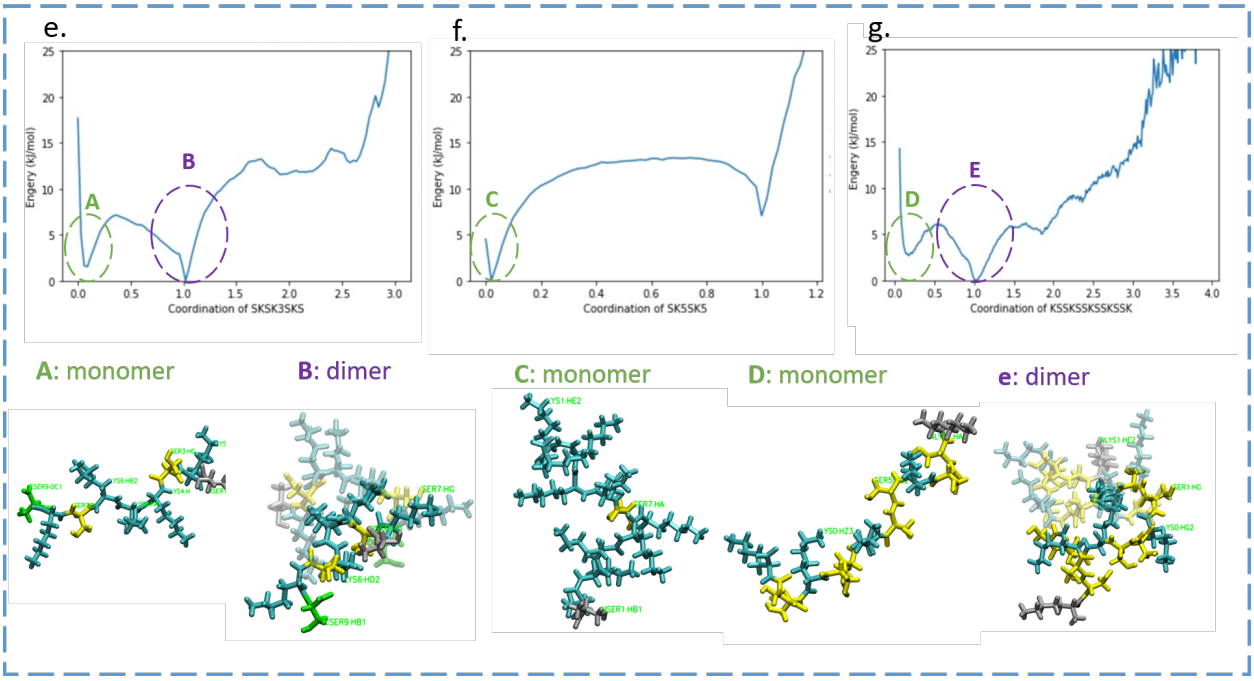

Figure 5. Dominant structures of $SKSK_3SKS$, SK_5SK_5 and KSSKSSKSSKSSK systems for each energy minimum: (c) $SKSK_3SKS$ with Tris system; (d) SK_5SK_5 with Tris system; (e) KSSKSSKSSKSSK with Tris system. Coloring is as described in Figure 4, with the modification of second-chain atoms in dimer shown in transparent representation for clarity.

From Figure 4 and Figure 5, no matter while Tris was present or not, it turned out the most stable conformation for KSSKK monomers was always extended, as predicted by the experimental chemical shifts results from Drobny's group.[17] Same phenomena were observed in the dominant monomer structures of $SKSK_3SKS$ and KSSKSSKSSKSSK systems, the former of which was also in agreement with Drobny's predictions. By contrast, SK_5SK_5 tended to fold to some degree rather than stay in an extended pose in solution, possibly owing to the strong repulsion from five continuous positively charged lysine residues again. Another interesting phenomenon was the

different dimer configurations of KSSKK without Tris and KSSKK with Tris. It was clearly seen that without Tris, KSSKK dimers were loosely associtaed with the positively charged side chain –NH_2 group of lysine and the side chain –OH group of serine with negative partial charge. However, this dimer was proved unstable in the subsequent 100-ns unbiased MD simulation in water. On the other hand, the KSSKK dimer with Tris as a bridge in middle, usually connecting the OH group of serine residues or the C-terminal carboxyl group of two chains to the hydrogen atoms of Tris together, turned out much more stable than its counterparts mentioned above. This unique Tris-bridged dimerization mechanism would inspire new findings and thoughts in protein aggregation, especially about the role of specific buffers and ions in it.

Interestingly, even though we also tested the dimerization process with and without Tris of the longest peptide, KSSKSSKSSKSSK, which has very similar fragments as KSSKK, and confirmed the similar dimer-stabilizing effects of Tris in this system, the aggregation mechanism seemed somewhat different. Considering the clustering results of $SKSK_3SKS$ and KSSKSSKSSKSSK, Tris was no longer a prerequisite for stable dimers (Figure 5). In fact, the existence of Tris in dimer was more likely a random choice in these two cases. Those dimer structures were, in some senses, folded and attached with each other with serine residues and lysine residues interposed throughout the whole pose, while serine residues with short side-chains somewhat buried inside and positively-charged lysine residues outside. One reasonable guess would be this alternating lysine-serine arrangement would be ideal for the aggregates. And the longer peptides with alternating pattern would have more freedom to fold themselves for that stable conformation, thereby decreasing the reliance on the small molecule Tris to move around and associate that structure. Furthermore, if this is the case, an intertwined dimer structure with many lysine residues pointing

outside with long side-chains would probably not welcome more aggregates, which could be an explanation for the dramatic higher energy and minimal percentage of trimers and tetramers in our simulations. Similarly, a peptide sequence SK_5SK_5 with two five-consecutive lysine residues in one chain would not be able to form any aggregates.

Peptide Interaction with SiO_2/TiO_2 Surfaces

Using all the stable structures mentioned above from clustering analysis, a set of 12 100-ns unbiased MD simulations were carried out with SiO_2 surface and TiO_2 surface respectively. A comparison between the binding affinity of S-K peptides towards TiO_2 and SiO_2 was discussed in detail, with a focus on binding residues or reasons behind not-binding phenomenon.

Generally, all these four S-K peptides (both monomers and dimers) bind strongly to the surface of SiO_2, no matter which orientation we chose to put the peptide on surface. (Figure 6) With a post-analysis about counting the frames in which the surface separation distance (SSD, which is the distance between the atom in the top layer of the SiO_2 slab and the COM of the peptide) was no greater than 1nm, S-K peptide bind to the SiO_2 slab throughout almost 95% of the simulation trajectories. Owing to the strong electrostatic interaction between the positively charged S-K peptides (with charge from +3 to +10 per peptide) and negatively charged SiO_2 surface (with charge of -20), peptide would maintain the initial orientation during the whole 100ns and bind tightly to the surface without being affected by the rotation. Also, adequate positively-charged binding sites would always be exposed and available in both monomer state and dimer state, therefore no significant difference has been observed between these two binding conditions.

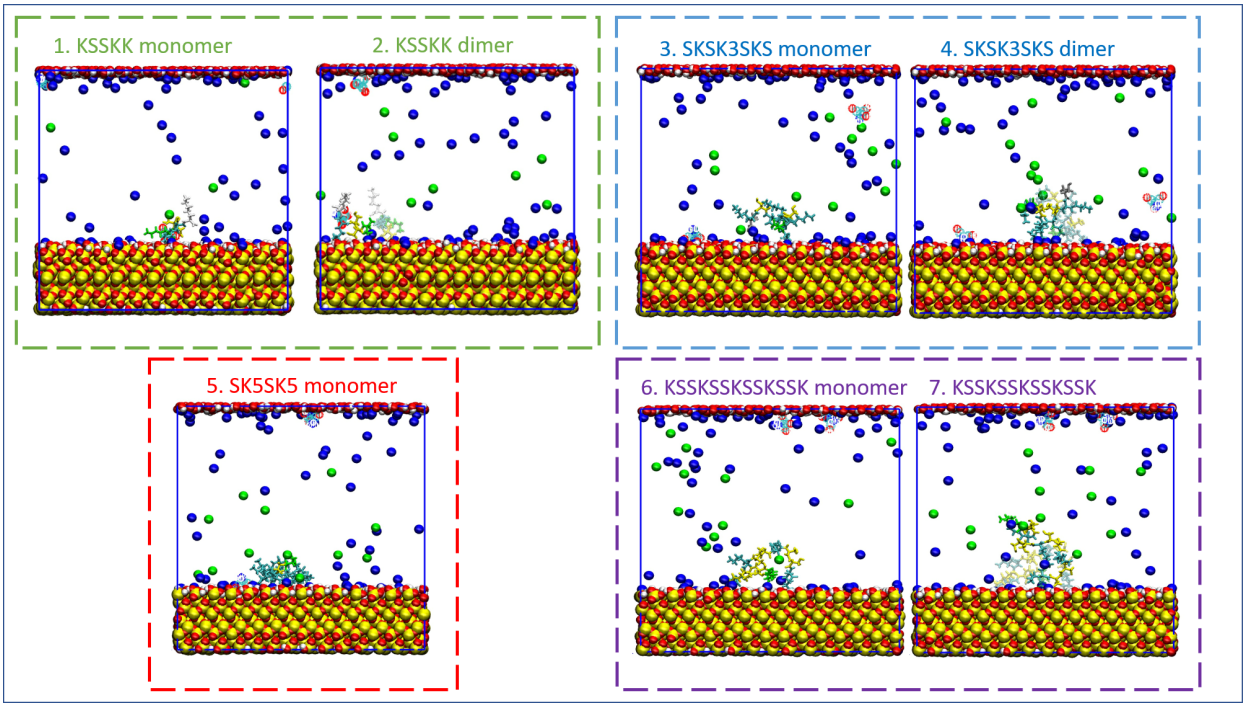

Figure 6. MD trajectory snapshots for the dominant structures on SiO_2 surface. Coloring is as described in Figure 4, with the modification of second-chain atoms in dimer shown in transparent representation, sodium ions as a part of counter-ions in SiO_2 surface is shown in dark blue while chloride ions shown in green. Water is not pictured for clarity.

Situation for peptide-TiO_2, however, was completely different. (Figure 7) Overall speaking, none of these S-K peptides could bind to the neutral TiO_2 slab without leaving it for a while. As was discussed in our draft paper, the electric field of TiO_2 would be much stronger than that of SiO_2 since the titanium atom in this model has a +2.2 charge, which is much higher than the +1.1-charged silicon atom in SiO_2 slab. With a lot of positive charges and complexity in structures, the S-K peptides would be highly likely compelled from the TiO_2 slab because of the strong electrostatic interaction and steric effects of the large biomolecules simultaneously. We also did a test of small positively-charged particles and SiO_2 surface interaction, proving sodium ions

(charge as +1) and Tris (charge as +1) could bind to SiO_2. One lysine residue was also demonstrated weak binding force onto the same TiO_2 model by Lyubartsev.[43] Nevertheless, the shortest KSSKK peptide could approach the surface while the OH group of Tris served as a bridge between the oxygen atom in the first layer of TiO_2 and the C-terminal carboxyl group of KSSKK monomer. The binding affinity seemed weak (happened only 40% of simulation time) and it would be less likely to happen for dimers with much larger overall positive charges and less degree of freedom for either Tris or serine residues to approach the oxygen atom in TiO_2 slab without destabilizing the aggregates. Same trend was visualized and recorded in $SKSK_3SKS$ and KSSKSSKSSKSSK systems.

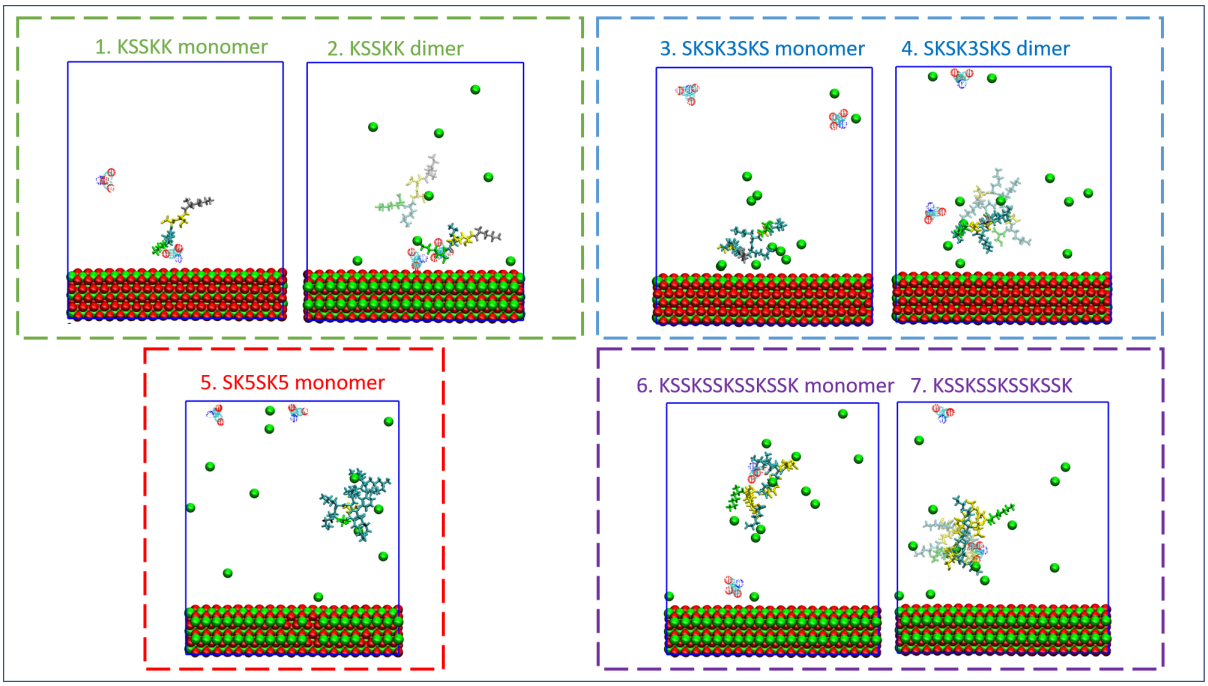

Figure 6. MD trajectory snapshots for the dominant structures on TiO_2 surface. Coloring is as described in Figure 4, with the modification of second-chain atoms in dimer shown in transparent representation, and the addition of chloride ions shown in green. Water is not pictured for clarity.

While S-K peptides show no preference for binding to the TiO_2 surface, the reason why they are able to precipitate TiO_2 nanospheres could be attributed to two main causes: (1) In vitro experiments, TiO_2 was usually produced with both particular protein/peptides and the precursor, which is TiBALDH providing the sources of titanium atoms. A highly likely hypothesis is that S-K peptides tend to interact with TiBALDH tightly in the first stage to catalyze and produce TiO_2, then be separated with them after the production process finish. (2) Considering the multiple possibilities of the protonation level of TiO_2, the interaction of S-K peptides and titania surface could also vary depending on the ionization of the surface. These hypotheses could be tested in our future studies.

Conclusions

This study investigated the molecular-level of serine-lysine peptides aggregation in water using the novel PBMetaD-PF enhanced sampling scheme as well as the specific binding interaction with SiO_2 and TiO_2 surfaces separately. In our simulations and analyses, we mainly focused on (1) establishing a protocol to effectively sample the conformational space of peptide aggregation and accurately assess the convergence of simulation; (2) exploring the relationship between the unique aggregates formation mechanisms and the alternating serine-lysine sequence arrangement of S-K peptides by identifying the dominant structures and states with clustering and visual analysis of simulation trajectories; (3) determining the role of Tris buffer during the S-K peptide aggregation and peptide-surface interaction; (4) studying and comparing the surface chemistry of TiO_2 and SiO_2 surface we used here and driving forces for peptide binding or not-binding behaviors. In general, seven simulations of peptide aggregation in water and twenty-four simulations of peptide-surface interaction were extensively studied using molecular dynamics simulations.

Inspired by the previous work conducted in our group, the application of the new PBMetaD-PF algorithm turned out to be efficient in sampling peptide aggregation states. Besides, the CV choice of coordination proved reliable in obtaining a much clearer free energy landscape. Next, based on lowest energy conformations from the clustering result, we realized a correlation between the increased aggregates forming tendency and a unique alternating serine-lysine arrangement of S-K peptides, for example, a sequence of pentalysine clusters interposed with roughly consistent number of serine residues as $SKSK_3SKS$ and KSSKSSKSSKSSK. This discovery inspires us to design more S-K peptides with similar sequence pattern but varied lengths to further test our hypothesis in future work. However, the absence of trimers and tetramers in our study also implies a special aggregation state for peptides composed solely of serine and lysine residues, which were inclined to expose the longer positively charged side-chains of lysine outside.

For another, it was demonstrated that Tris would help stabilize all those S-K peptides aggregates and even enlarge the binding affinity of S-K peptides and TiO_2 slab, both by being integrated into the OH group of serine side-chains or carboxyl group of C-terminal residues. The unique Tris-associated aggregate structure was also proposed in our work. More simulation boxes with a range of different Tris concentration could be conducted to determine the effect of Tris in detail. Also, if this is the case, future buffer selection for many biomolecular systems would need to be reconsidered.

In addition, the electrostatic forces between peptides and inorganic surfaces was further explored here. Comparing the systems of globally neutralized TiO_2 and negatively charged SiO_2, our results showed that for the metal oxide surfaces, the peptide's binding behavior was not only determined by the strength of electric filed, but also affected by the possible steric effects of the large biomolecules. In the potential next steps, thoughts could be extended to build peptides or even artificial molecules with same charge but different sizes and examine their behaviors on TiO_2 slab. And individual binding energies could be calculated from enhanced sampling methods for a much more comprehensive comparison. These findings would largely deepen our understanding of protein-mineral interactions as well as protein aggregation phenomenon and further upgrade our capabilities in improving current bio-medical technologies.

References

1. Krajina, A. B., Proctor, A. C., Schoen, A. P., Spakowitz, A. J., Heilshorn, S. C. Biotemplated synthesis of inorganic materials: An emerging paradigm for nanomaterial synthesis inspired by nature, *Progress in Materials Science*. **91**, 1-23 (2018).
2. Dickerson, M. B., Sandhage, K. H., Naik, R. R. Protein- and Peptide-Directed Syntheses of Inorganic Materials. *Chem. Rev.* **108**, 4935−4978 (2008).
3. Gordon, R., Losic, D., Tiffany, M. A., Nagy, S. S., Sterrenburg, F. A. S. The Glass Menagerie: Diatoms for Novel Applications in Nanotechnology. *Trends Biotechnol.* **27**, 116−127 (2009).
4. Losic, D., Mitchell, J. G., Voelcker, N. H. Diatomaceous Lessons in Nanotechnology and Advanced Materials. *Adv. Mater.* **21**, 2947−2958 (2009).
5. Naik, R. R., Whitlock, P. W., Rodriguez, F., Brott, L. L., Glawe, D. D., Clarson, S. J., Stone, M. O. Controlled Formation of Biosilica Structures in Vitro. *Chem. Commun.* **2**, 238−239 (2003).
6. Parker, A. R., Townley, H. E. Biomimetics of Photonic Nanostructures. *Nat. Nanotechnol.* **2**, 347−353 (2007).
7. Krö ger, N., Deutzmann, R., Bergsdorf, C., Sumper, M. SpeciesSpecific Polyamines from Diatoms Control Silica Morphology. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14133−14138 (2000).
8. Krö ger, N., Lorenz, S., Brunner, E., Sumper, M. Self-Assembly of Highly Phosphorylated Silaffins and Their Function in Biosilica Morphogenesis. *Science*. **298**, 584−586 (2002).

9. Poulsen, N., Krö ger, N. Silica Morphogenesis by Alternative Processing of Silaffins in the Diatom Thalassiosira pseudonana. *J. Biol. Chem.* **279**, 42993−42999 (2004).
10. Wenzl, S., Hett, R., Richthammer, P., and Sumper, M. Silacidins: Highly acidic phosphopeptides from diatom shells assist in silica precipitation in vitro. *Angew Chem. Int. Ed. Engl.* **47**, 1729–1732 (2008)
11. Knecht, M. R., Wright, D. W. Functional Analysis of the Biomimetic Silica Precipitating Activity of the R5 Peptide from Cylindrotheca fusiformis. *Chem. Commun.* **24**, 3038−3039 (2003).
12. Knecht, M. R., Wright, D. W. Amine-Terminated Dendrimers as Biomimetic Templates for Silica Nanosphere Formation. *Langmuir.* **20**, 4728−4732 (2004).
13. Rodríguez, F., Glawe, D. D., Naik, R. R., Hallinan, K. P., Stone, M. O. Study of the Chemical and Physical Influences upon In Vitro Peptide-Mediated Silica Formation. *Biomacromolecules*. **5**, 261−265 (2004).
14. Senior, L., Crump, M. P., Williams, C., Booth, P. J., Mann, S., Perriman, A. W., Curnow, P. Structure and Function of the Silicifying Peptide R5. *J. Mater. Chem. B.* **3**, 2607−2614 (2015).
15. Lechner, C. C., Becker, C. F. W. Modified Silaffin R5 Peptides Enable Encapsulation and Release of Cargo Molecules from Biomimetic Silica Particles. *Bioorg. Med. Chem.* **21**, 3533−3541 (2013).
16. Poulsen, N., Scheffel, A., Sheppard, V. C., Chesley, P. M., Krö ger, N. Pentalysine Clusters Mediate Silica Targeting of Silaffins in Thalassiosira Pseudonana. *J. Biol. Chem.* **288**, 20100−20109 (2013).

17. Buckle, E. L., Lum, J. S., Roehrich, A. M., Stote, R. E., Vandermoon, B., Dracinsky, M., Filocamo, S. F., Drobny, G. P. Serine–Lysine Peptides as Mediators for the Production of Titanium Dioxide: Investigating the Effects of Primary and Secondary Structures Using Solid-State NMR Spectroscopy and DFT Calculations. *J. Phys. Chem. B* **122**, 4708-4718 (2018)

18. Wang W. Protein aggregation and its inhibition in biopharmaceutics. *Int J Pharm.* **289**, 1–30 (2005).

19. Massimiliano, M., Giulia, M., Giorgio, C. Investigating the Mechanism of Peptide Aggregation: Insights from Mixed Monte Carlo-Molecular Dynamics Simulations. Biophys J. **94**, 4414–4426 (2008).

20. Riccardi, L., Nguyen, P. H., Stock, G. Construction of the Free Energy Landscape of Peptide Aggregation from Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **8**, 1471–1479 (2012).

21. Martín, C-P., Birgit, S., Advances in the Simulation of Protein Aggregation at the Atomistic Scale. *J. Phys. Chem. B*. **120**, 2991–2999 (2016).

22. Selkoe, D. J. Folding proteins in fatal ways. *Nature.* **426**, 900– 904 (2003).

23. Pepys, M. B. Amyloidosis. *Annu. Rev. Med.* **57**, 223–241 (2006).

24. Andrews, J. M., C. J. Roberts. A Lumry-Eyring nucleated polymerization model of protein aggregation kinetics. 1. Aggregation with pre-equilibrated unfolding. *J. Phys. Chem. B.* **111**, 7897–7913 (2007).

25. Prakash, A., Fu, C. D., Bonomi, M., Pfaendtner, J. Biasing Smarter, Not Harder, By Partitioning Collective Variables Into Familiesin Parallel Bias Metadynamics. [in preparation]

26. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4,** 17 (2012).

27. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).

28. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2,** 19–25 (2015).

29. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. (2013).

30. 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. Phys.* **79,** 926–935 (1983).

31. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 14101 (2007).

32. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).

33. Matsui, M. & Akaogi, M. Molecular Dynamics Simulation of the Structural and Physical Properties of the Four Polymorphs of TiO 2. *Mol. Simul.* **6,** 239–244 (1991).

34. Emami, F. S. *et al.* Force field and a surface model database for silica to simulate interfacial properties in atomic resolution. *Chem. Mater.* **26,** 2647–2658 (2014).

35. Brandt, E. G. & Lyubartsev, A. P. Molecular Dynamics Simulations of Adsorption of Amino Acid Side Chain Analogs and a Titanium Binding Peptide on the TiO2 (100) Surface. *J. Phys. Chem. C.* **119,** 18126–18139 (2015).

36. Healy, K. E. & Ducheyne, P. Hydration and preferential molecular adsorption on titanium in vitro. *Biomaterials.* **13,** 553–61 (1992).

37. Gražulis, S., Merkys, A., Vaitkus, A., Okulič-Kazarinas, M. & IUCr. Computing stoichiometric molecular composition from crystal structures. *J. Appl. Crystallogr.* **48,** 85–91 (2015).

38. Merkys, A. *et al.* COD::CIF::Parser: an error-correcting CIF parser for the Perl language. *J. Appl. Crystallogr.* **49,** 292–301 (2016).

39. Gražulis, S. *et al.* Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Res.* **40,** D420–D427 (2012).

40. Patwardhan, S. V. *et al.* Chemistry of aqueous silica nanoparticle surfaces and the mechanism of selective peptide adsorption. *J. Am. Chem. Soc.* **134,** 6244–6256 (2012).

41. Emami, F. S. *et al.* Force field and a surface model database for silica to simulate interfacial properties in atomic resolution. *Chem. Mater.* **26,** 2647–2658 (2014).

42. Massimiliano, B., Davide, B., Giovanni, B., Carlo, C., Davide, P., Paolo, R., Davide, D., Fabrizio, M., Fabio, P., Ricardo, A. B., Michele, P. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*. **180**, 1961–1972 (2009).

43. Brandt, E. G., Lyubartsev, A. P. Molecular Dynamics Simulations of Adsorption of Amino Acid Side Chain Analogues and a Titanium Binding Peptide on the TiO2 (100) Surface. *J. Phys. Chem. C*. **119**, 18126–18139 (2015).