HHS Public Access

Author manuscript

J Chem Inf Model. Author manuscript; available in PMC 2020 June 05.

Published in final edited form as:

J Chem Inf Model. 2019 September 23; 59(9): 3899–3909. doi:10.1021/acs.jcim.9b00514.

A Multiscale Model for the Self-assembly of Coat Proteins in Bateriophage MS2

Bo Wang[1], **Junjie Zhang**[2], **Yinghao Wu**[1,*]

[1]Department of Systems and Computational Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY, 10461

[2]Department of Biochemistry and Biophysics, Center for Phage Technology, Texas A&M University, College Station, TX 77843

Abstract

The self-assembly of viral capsids is an essential step to the formation of infectious viruses. Elucidating the kinetic mechanisms of how a capsid or virus-like particle assembles could advance our knowledge about the viral lifecycle, as well as the general principles in self-assembly of biomaterials. However, current understanding of capsid assembly remains incomplete for many viruses due to the fact that the transient intermediates along the assembling pathways are experimentally difficult to be detected. In this paper, we constructed a new multiscale computational framework to simulate the self-assembly of virus like particles. We applied our method to the coat proteins of Bacteriophage MS2 as a specific model system. This virus-like particle of Bacteriophage MS2 has a unique feature that its 90 sequence-identical dimers can be classified into two structurally various groups: one is the symmetric CC dimer; the other is the asymmetric AB dimer. The homotypic interactions between AB dimers result in a 5-fold symmetric contact, while the heterotypic interactions between AB and CC dimers result in 6-fold symmetric contact. We found that the assembly can be described as a physical process of phase transition that is regulated by various factors such as concentration and specific stoichiometry between AB and CC dimers. Our simulations also demonstrate that heterotypic and homotypic interfaces play distinctive roles in modulating the assembling kinetics. The interaction between AB and CC dimers is much more dynamic than the interaction between two AB dimers. We therefore suggest that the alternate growth of viral capsid through the heterotypic dimer interactions dominates the assembling pathways. This is, to the best of our knowledge, the first multiscale model to simulate the assembling process of coat proteins in Bacteriophage MS2. The generality of this approach opens the door to its further applications in assembly of other viral capsids, virus-like particles and novel drug delivery systems.

[*]Corresponding authors: Yinghao Wu, Phone: (718) 678-1232, Fax: (718) 678-1018, yinghao.wu@einstein.yu.edu.

Author Contributions

J.Z. and Y.W. designed research; B.W. and Y.W. performed research; B.W. and Y.W. analyzed data; B.W., J.Z. and Y.W. wrote the paper.

Competing financial interests: The authors declare no competing financial interests.

Introduction

Viruses are small infectious agents that need host cells to reproduce[1]. They contain genomes, which are packaged in a protein coat called viral capsid[2]. A capsid contains multiple copies of oligomeric protein subunits that can self-assemble into a highly symmetric supramolecular structure[3]. The research on viral capsids or virus-like particles (VLPs) is an extremely active area, not only because it is one of the most important model systems for understanding the general principles in self-assembly of biological materials[4], but also because of its potential applications in medicine and bio-nanotechnology[5], for example, the development of cancer vaccines[6]. Among different morphologies, the icosahedron is the most common structure for viral capsids[7]. Icosahedral capsids contain 12 pentagonal faces and $10\times(T-1)$ hexagonal faces, in which T is called the "triangulation number" that represents the size and complexity of the capsids[8]. For instance, the capsid of bacteriophage MS2 has the icosahedral structure with T=3. Bacteriophage MS2 is a single-stranded RNA (ssRNA) virus that infects the bacterium *Escherichia coli*[9]. The structure of the virus in complex with the *Escherichia coli* F-pilus has recently be discovered by single-particle cryo-electron tomography (cryo-EM)[10]. The VLP that resembles its capsid consists of 180 sequence-identical coat protein monomers, which further form, 90 structurally stable dimers[11], compared with the 89 dimers and a maturation protein in the original capsid. Interestingly, unlike most other T=3 icosahedral capsids, the 90 dimers from the VLP of MS2 are classified into two structurally different groups. One group contains 60 asymmetric dimers (called AB dimers), while the other one contains 30 symmetric dimers (called CC dimers). This alternate pattern between AB and CC dimers is a unique feature of MS2 and few other ssRNA viruses such as phage Qβ[12]. However, the functional role of this spatial organization in regulating the kinetic pathways of capsid assembly is not fully understood.

Comparing with the experimental techniques that are difficult to characterize the kinetic pathways of capsid assembly, computational modeling is much less time-consuming and offers mechanistic details that are currently inaccessible in the laboratory[13]. A large variety of models have been developed to simulate the assembling processes of viral capsids[14]. Mathematical analysis based on rate equation models was first used to describe the time evolution of capsid intermediates[15–17]. Later, particle-based simulation approaches were generated through adding the spatial resolution and special geometry to the assembly[18–23]. These approaches were further improved by constructing the coarse-grained models for each protein monomer in the capsids[24]. They have been successfully applied to understand the polymorphism in capsids of icosahedral[25] and HIV viruses[26–29]. However, in these studies, the scaled units were used to calibrate binding between pseudo-atoms in two interacting monomers. These homotypic and artificially adjusted parameters are difficult to capture the detailed energetics in viral capsids. On the other hand, all-atom molecular dynamic (MD) or Brownian dynamic (BD) simulations are based on physics-based potential functions[30–32]. Unfortunately, due to the intense consumption for computational resources, these algorithms were so far limited to assessing the binding stability and rates between individual subunits of the viral capsid[33–35], and are still not realistic to simulate their assembly pathways[36, 37]. In order to overcome the restrictions in different computational methods, multiscale modeling is becoming a promising technique that can balance the energetic details and computational

efficiency[38–43]. As a result, the implementation of this technique, which will reveal the kinetic mechanisms of how capsids assemble, is highly demanded.

In this work, a multiscale framework is developed to simulate self-assembly of coat proteins in MS2. In this framework, each protein dimer in the VLP of MS2 was represented by a rigid-body (RB) in a simplified model so that the assembling pathways can be effectively traced with spatial details. The RB model was specifically designed to capture sufficient geometric features in the icosahedral structure of the MS2 capsid. As illustrated by the structural model (Figure 1), there are two types of binding interfaces in the VLP: the heterotypic binding interface between AB and CC dimers and the homotypic binding interface between two CC dimers. The binding rates of these two interfaces in the RB-based model were characterized from high-resolution simulations, which incorporate structural and energetic information between individual dimers. As a result, this method is able to approach the spatial-temporal scale of assembling process, while retain the basic binding energetics encoded at the interfaces between different structural dimers. We found that assembly can be described as a physical process of phase transition that is regulated by various factors such as concentration and specific stoichiometry between AB and CC dimers. Our simulations indicate that heterotypic and homotypic interfaces play distinctive roles in modulating the dynamics of assembly. The association between AB and CC dimers is faster but less stable than the association between two AB dimers, although the structures of these two dimers are highly similar, indicating that the binding between AB and CC dimers is more dynamic. Therefore, we suggest that the alternate growth of viral capsid controlled by the dynamics of heterotypic dimer interactions dominates the assembling pathways. Altogether, this is, to the best of our knowledge, the first multiscale diffusion-reaction model to study the self-assembly of viral capsid and VLP. The generality of the method paves the way for its applications to viral capsids and VLPs with any other symmetries.

Results

Construct the rigid-body-based model for the MS2 viral capsid

Based on the icosahedral symmetry observed in the capsid of MS2[11], we built a coarse-grained model for its VLP in which each of the 90 structural dimers is represented by a spherical rigid body (Figure 1a). This is a reasonable simplification, because each dimer forms a very compact structure in which the β-strands of two subunits in the dimer form a continuous sheet, and the C-terminal helix of each subunit fits into the groove between the N-terminal hairpin and C-terminal helix of its neighboring subunit in the dimer (Figure 1c). The diameter of each rigid body equals 21.0Å, which is set to fit the size of a dimer in the capsid. Rigid bodies of 60 AB dimers are shown in blue, while rigid bodies of 30 CC dimers are shown in red. The capsid contains two specific types of interactions between dimers. The homotypic interactions between AB dimers result in the formation of homogeneous pentamers. In contrast, the heterotypic interactions between AB dimers and CC dimers lead to the formation of heterogeneous hexamers that are alternately occupied by three AB dimers and three CC dimers. These two types of interactions are resulted from the structural difference between the AB and CC dimers. The loops between the F and G strands (FG

loop) in a CC dimer are identical, while the FC loop in one subunit of a AB dimer is structurally more disorder and more closely associated with the rest of the protein (Figure 1c)[44]. As a result, the 5-fold contacts are made by AB dimers with their disordered FG loops, while the 6-fold contacts are made between AB and CC dimers with their extended FG loops.

Because each dimer in the rigid-body model of the capsid has four structural neighbors, we assign four binding sites on the surface of each rigid body to delineate the interactions between dimers and their structural neighbors, as shown by the yellow spots in Figure 1a. The index of these binding sites is listed in Figure 1b. These binding sites are designed to be distinctive with each other, so that the structural different between AB and CC dimers can be further incorporated into the coarse-grained model. Additionally, specific rules were made for the interactions between different binding sites of various dimers, so that they can aggregate with specific folds of symmetry. For instance, the binding site “1” of one AB dimer only form contact with the binding site “2” of another AB dimer. Through this interaction with 5-fold symmetry, AB dimers can oligomerize into homo-pentamers. Similarly, the binding sites “3” or “4” in an AB dimer can form contact with binding sites in CC dimers. Through this heterotypic interaction with 6-fold symmetry, AB and CC dimers can oligomerize into hexamers. Considering that the 6-fold contacts can also be formed between two AB dimers or between two CC dimers using their extended FG loops, the binding sites “3” or “4” between two AB dimers or are allowed to form interactions with 6-fold symmetry that are not presented in the capsid but still can theoretically exist. Likewise, the binding sites “1” or “2” in one CC dimer are also allowed to form interactions with 6-fold symmetry with the binding sites “3” or “4” in another CC dimer. However, the 5-fold contacts can only be formed between the binding sites “1” or “2” of AB dimers due to the potential stereo-chemical clashes between disordered and extended FG-loops.

The above rigid body model, together with the dimer stoichiometry and the spatial arrangement of binding sites incorporates sufficient geometric information for the assembly of the viral capsid. Followed by the rules of interactions between specific pairs of binding sites, dimers can be assembled together with predefined fold of symmetry from a random initial configuration, while the assembling kinetics is further determined by the diffusion constant of each dimer and binding rates between two specific types of dimers. Details of the diffusion reaction algorithm used to guide the assembling process are described in the *Methods*.

Characterize the general dynamic features of capsid assembly

In order to understand the kinetic mechanism of how capsid of MS2 forms, we simulated the assembly of capsid dimers using the RB-based model described above. Specifically, 150 spherical rigid bodies of capsid dimers were randomly placed in a 3-dimensional cubic box with a volume of $100\times100\times100$ nm^3 as an initial configuration (Figure 2b). The ratio between AB and CC dimers were fixed as 2:1, as observed in the native state of the MS2 capsid. Following the initial setup, both AB and CC dimers undergo stochastic diffusions in the simulation box. The diffusion constants for both types of dimers equal $10Å^2$/ns. Two dimers will further associate into homogeneous or heterogeneous complexes if their distance

is below the cutoff (10Å), and *vice versa*, a complex can also dissociate into two separate dimers, depending on the rates of association and dissociation. As the first step, the values of both association rate (r_{on}) and dissociation rate (r_{off}) for both homotypic and heterotypic interactions were given equally ($r_{on}=0.05ns^{-1}$; $r_{off}=0.0025ns^{-1}$) to characterize the general topology of capsid assembly. The kinetics profiles of assembly are plotted in Figure 2a, based on which we suggest that the capsid formation is a two-step process.

In detail, the first stage is from 0 to around 2.7×10^8ns, as shown by the yellow panel in Figure 2a. Within this stage, the number of oligomers in the system (black curve in Figure 2a) increases from 0 to about 20. In contrast, the size of these oligomers grows very slowly (red curve in Figure 2a). Only small oligomers consisting of less than 5 dimers were formed in this stage. For instance, a hetero-trimer formed by two AB dimers and one CC dimer is highlighted in the snapshot at 1.0×10^8ns (pointed by the orange arrow in Figure 2c). Therefore, we propose that this stage is called “seeding” phase in which small oligomers with different combinations of AB and CC dimers were dynamically formed to serve as seeds for further growth. Following the first stage, the sizes of oligomers start growing fast, while the total number of oligomer decreases, as shown by the blue panel in Figure 2a. We propose that this stage is called “growing” phase in which small oligomers are merged and form more stable larger oligomers. One representative snapshot at 5.0×10^8ns is plotted as Figure 2d. At the end of the simulation, there are less than 10 oligomers left in the system. The largest oligomer formed contains more than 50 dimers (Figure 2e). We noticed that the specific symmetry with the triangulation number T equals 3 is observed in the partially assembled capsid, in which 5-fold contacts are formed by AB dimers and 6-fold contacts are formed among AB and CC dimers with alternate spatial patterns.

To illustrate the effect of stoichiometry between AB and CC dimers on capsid assembly, we fixed the total number of dimers in the system and the size of the simulation box. The percentage of AB dimers was then changed from 0 to 1. At the end of simulation for each value of tested percentage, we counted the maximal size of oligomer formed in the system. The histogram of largest formed oligomer versus the percentage of AB dimers in the system is plotted in Figure 3a. The final snapshots at some representative percentage values are inserted as red panels. The figure shows that only small oligomers with less than 10 dimers were formed when the system is dominated by either AB dimers (blue) or CC (red) dimers, as indicated by the right and left panels in Figure 3a, respectively. Large capsid-like oligomers can only be formed within a narrow window of percentage. The peak is located at the value of 0.6 when the ratio of AB and CC dimers equals 2:1, leading into the alternate growth of large oligomers as plotted in the middle panel of Figure 3a. This result indicates that the effective assembly of MS2 viral capsid requires the presence of dimers with specific stoichiometry.

To further test the concentration dependence of assembly, we fixed the size of the simulation box and stoichiometry between AB and CC dimer at 2:1. We changed the total number of dimers in the system from 0 to 300. At the end of the simulation at each concentration, we counted the maximal size of oligomers formed in the system. The correlation between the molecular concentration and largest size of formed oligomers is plotted in Figure 3b. The figure indicates that when the total number of dimers increases from 0 to 40, the system

undergoes a phase transition. Specifically, when the system contains lower than 30 dimers, no oligomer is observed during simulations. However, a small increase of dimer number from 30 to 40 led to the assembly of oligomers that contain more than 10 dimers, as shown by the left panel in Figure 3b. Further increase of concentration facilitates the assembly of larger oligomers. For instance, when the system contains 200 dimers, capsid-like large capsids with more than 60 dimers were obtained from the simulation (the middle panel of Figure 3b). Interestingly, we found that the further increase of system concentration disfavor the assembly of large oligomers, as shown by the right panel in Figure 3b. We speculate that the non-native dimer-dimer contacts under high local concentrations prevent these miss-assembled small oligomers from further growth.

Taken together, our simulation results suggest that the assembly of MS2 viral capsid is characterized by its multistep kinetic process, and is highly sensitive to the molecular concentration and stoichiometry in the system.

Evaluate the kinetic binding rates of homotypic and heterotypic interactions

As shown in Figure 1, the binding interfaces between two dimers in the viral capsid are classified into two geometrical groups: the 5-fold homotypic binding interface mediates the interactions between two AB dimers, while the 6-fold heterotypic binding interface mediates the interactions between AB dimers and CC dimers. In the last section, we investigated the general mechanism of capsid assembly by assuming that the values of both association and dissociation rates are identical for both binding interfaces. In order to analyze the difference of functions between homotypic and heterotypic interfaces in regulating the capsid assembly, we further changed their rates into different values.

In specific, we separately evaluated the functions of association and dissociation. We first focused on association by fixing the dissociation rates for both binding interfaces at the value of $0.0025ns^{-1}$. Three different values were then assigned to r_{on} of both interfaces: $0.05ns^{-1}$, $0.01ns^{-1}$, and $0.002ns^{-1}$. All nine combinations were tested and their simulation results are summarized as the three-dimensional bars in Figure 4a. The values of association rates for homotypic and heterotypic binding interfaces are indexed along the x-axis and the y-axis, while the largest size of oligomers formed along each simulation is illustrated as the height of bars along the z-axis. The figure indicates that large oligomers can only be formed when the association rates of heterotypic interactions are as high as $0.05ns^{-1}$ (grey bars). When the r_{on} of heterotypic interactions equals $0.002ns^{-1}$ (white bars) and $0.01ns^{-1}$ (striped bars), only small oligomers containing less than 10 dimers were found. Interestingly, large oligomers can still be formed in the simulations even if the association rate of homotypic interactions is very weak. For instance, an oligomer with more than 30 dimers was found in the system when the r_{on} of heterotypic interactions equals $0.05ns^{-1}$, while the r_{on} of homotypic interactions is as weak as $0.0025ns^{-1}$. This result strongly suggests that the assembly of capsid-like large complexes is kinetically regulated by the association between AB and CC dimers.

In attempt to further understand how stability of interactions between different dimers affects capsid assembly, we further fixed the rate of association at $0.05ns^{-1}$ and tuned r_{off} into different values. Five specific values were adopted for both types of interfaces, ranging

from $0.01ns^{-1}$ to $0.000625ns^{-1}$. All different combinations were tested and their simulation results are summarized as the two-dimensional matrix in Figure 4b. The values of dissociation rates for heterotypic and homotypic binding interfaces are indexed along the x-axis and the y-axis, while the largest size of oligomers formed along each simulation is coded by different colors as indicated on the right panel. As shown by the blue units in the figure, only small oligomers were formed under weak interactions (large values of r_{off}). However, a small decrease of r_{off} leads to the sudden appearance of large capsid-like oligomers, indicating that there is a threshold in the strength of dimer interactions to trigger the phase transition of capsid assembly. Interestingly, when we further increase the strength of dimer interactions by decreasing the values of r_{off}, largest size of oligomers start to reduce, as shown by the green units. We speculate that this could be due to the fact that the growth of viral capsid was trapped by the strong interactions between dimers. Moreover, Figure 4b suggests an asymmetric pattern between heterotypic and homotypic interactions. As illustrated by the lower-left region of the matrix, the threshold-like transition was observed under strong homotypic but weak heterotypic interactions. Such behavior does not exist when homotypic interaction is weak and heterotypic interaction is strong, as illustrated by the upper-right region of the matrix. In the meantime, oligomers are easier to trap under strong heterotypic interaction than strong homotypic interaction. These results suggest that the stability of homotypic interaction play a more important role in regulating the dynamics of capsid assembly.

In summary, we demonstrated that the assembly of the MS2 VLP is a phase transition that is closely regulated by the inter-molecular binding between different dimers. The homotypic and heterotypic binding interfaces have distinctive functions. While the association kinetics of heterotypic interactions facilitates the formation of large capsid-like complexes, the dynamics of assembly is more closely modulated by the stability of homotypic interactions between AB dimers.

Explore the energetic basis of viral assembly from high-resolution simulations

We have shown the difference of homotypic and heterotypic binding interfaces in the regulation of viral assembly based on a geometric model. In order to further explore the energetic origin of this difference, higher-resolution atomic simulations were used to evaluate these two types of interactions in which the structural details of both dimers were presented. Specifically, the atomic coordinates for both homotypic interaction between two AB dimers and heterotypic interaction between an AB dimer and a CC dimer are taken from the PDB id 1BMS[45]. We first applied our recently developed residue-based kinetic Monte-Carlo (kMC) method to estimate the association rates of binding between two different dimers. Two systems were built to respectively test: 1) the binding between two AB dimers through the homotypic interface (Figure 5a), and 2) the binding between an AB dimer and a CC dimer through the heterotypic interface (Figure 5b).

For each system, 10^3 simulation trajectories were generated. In the initial conformation of each trajectory, two dimers with residue-based representation were placed with a random position relative to each other in which the distance between their binding interfaces is fallen within a cutoff value d_c, as described in the *Methods*. During the kMC simulations, the

diffusions of both dimers in the system are guided by their inter-molecular interactions. The interactions combine both hydrophobic effect and electrostatic interactions of two interacting dimers. At the end of each trajectory, two dimers either form an encounter complex through their pre-defined interface, or diffuse away from each other. Based on the simulation results collected from all the trajectories, we counted how many times an encounter complex can be formed given a specific value of d_c. We systematically tested different values of distance cutoff d_c, ranging from 10Å to 15Å. The relation between d_c and number of formed encounter complexes for both systems is plotted in Figure 5c.

The figure shows that probabilities of association drop for both binding interfaces when the distance cutoff increases, suggesting that encounter complexes are more difficult to form if two dimers are initially separated farther from each other. More importantly, the comparison of the curves between homotypic and heterotypic binding shows that the association rate of binding through homotypic interface is much lower than through the heterotypic interface, indicating that the binding between an AB dimer and a CC dimer is much faster than the binding between two AB dimers. Interestingly, in our rigid-body simulation, we revealed that assembly is kinetically controlled by the association rate of heterotypic interactions in capsid. Therefore, the residue-based kMC simulations provide the energetic basis to the rigid body simulation results, and demonstrated that the fast association between AB and CC dimers through heterotypic interface is not only required to geometrically facilitate the assembly of viral capsid with specific high-order symmetry, but also is naturally designed in their primary protein sequences.

In addition to the rate of association, we also considered the stability of interactions between dimers on the atomic level. Specifically, the unbiased discrete molecular dynamic (DMD) simulations were carried out to two systems, one contains two AB dimers and the other contains an AB dimer and a CC dimer. We performed DMD simulations at three temperatures: 275, 287.5 and 300K. Five independent replicas were generated for each temperature. All replicas were initiated from different random seed, and the length of each replica is 50ns. As a result, a sum of 250 ns was obtained for each system at different temperatures. Detailed simulation setups are described in the *Methods*. We calculated the percentage of remaining native contacts between two dimers along each simulation trajectory. The percentage of remaining native contacts is defined as the ratio of native contact number at current simulation step versus the total native contact number at the beginning[46]. Residues from two dimers form a native contact during simulations if the distance between any atoms on the sidechains of these two residues was less than 2Å from the distance in the native conformation.

The calculated percentage of remaining native contacts between two AB dimers is plotted for one of the representative replica at 275K (Figure 5d). The figure shows that the percentage dropped after the beginning and fluctuated along the simulation, indicating the recursive dissociation and re-association between two dimers. The final percentage was stabilized around 10%. We also compare the initial conformation of the complex with the final structure at the end of the 50ns trajectory. In the inserted panel of Figure 5d, the final conformation after simulation (yellow) is aligned with the native conformation (blue). The structural comparison shows that two dimers were partially dissociated after 50ns of

simulation. Their binding orientation was opened about 10° relative to the native complex. We further calculated the average percentage of remaining native contacts over the last 5 ns of all five replicas. The results are plotted as histogram with error bars in Figure 5e for both AB/AB and AB/CC systems at three temperatures. The homotypic interaction between two AB dimers is represented by the grey columns, while the heterotypic interaction between AB and CC dimers is represented by the striped columns. The figure firstly shows that in both AB/AB and AB/CC systems, there are less remaining native contacts when the simulation temperatures increase. This is resulted from the fact that protein complexes are more stable at low temperatures. More importantly, the histogram shows that the remaining native contacts in the AB/CC system are consistently lower than the AB/AB system at all temperatures. Therefore, our simulations suggest the complexes formed through the homotypic interactions are thermodynamically more stable than the heterotypic interactions.

Integrate the atomic simulation results with the rigid-body-based model

Our rigid body model suggests that the formation of large capsid-like complexes is facilitated by the fast association of heterotypic interactions, while the dynamics of assembly is more closely modulated by the stability of homotypic interactions between AB dimers. Interestingly, the simulations at higher resolution further confirmed that binding through homotypic interface is much slower but more stable than through the heterotypic interface. Using the information from the atomic simulations as reference, we determined the final values for association and dissociation rates and fed them into the rigid-body model to simulate the assembly of capsid. As a result, the value of r_{on} for homotypic interface is $0.025ns^{-1}$and r_{off} equals $0.0025ns^{-1}$, and the value of r_{on} for heterotypic interface is $0.05ns^{-1}$and r_{off} equals $0.000625ns^{-1}$. We practically adopted fast rates for association and dissociation, but keeping their relative differences between homotypic and heterotypic interactions. This can assure that the simulation can be computationally accessible and in the meanwhile the kinetics of the assembly process still remains qualitatively unchanged. Other simulation parameters were listed as follows. 200 capsid dimers were included in a cubic box with a volume of 100×100×100 nm^3 with the ratio between AB and CC dimers fixed as 2:1.

Using these parameters in the rigid-body model, the dynamics of capsid assembly was simulated again. Starting from a random configuration, the total number of oligomers in the system is plotted in Figure 6a along with the simulation time. The figure shows an initial increase and consecutive decrease in the total number of oligomers, indicating that the system follows the similar two-step kinetic profile as obtained before. The snapshot of the final configuration is shown in Figure 6b. The large capsid-like oligomers with the specific symmetry of the triangulation number equals 3 are observed in the figure, indicating that the capsid of Bacteriophage MS2 can be assembled under the realistic condition of protein interactions. We further plot the size of the largest oligomer observed along the simulation (black curve in Figure 6c), and decomposed the oligomer into the number of CC dimers (red curve) and the number of AB dimers (blue curve) in the complex. The figure shows that there are about 50 dimers finally included in the largest oligomer. Moreover, while the numbers of both AB and CC dimers in the oligomer keep growing along the simulation, the total number of AB dimers always doubles the number of CC dimers. This suggests an

alternate growing pattern during the assembly. Some representative structures of the oligomer are selected along its growth. They give a close look at how viral capsid was assembled. Interestingly, it is worth mentioning that non-native contacts can be formed in the initial stage of the assembly. For instance, three AB dimers and three CC dimers formed a non-native complex with their 6-fold symmetric contacts (Figure 6d). However, it was later replaced by a more stable native-like spatial arrangement (Figure 6e), partially due to the reason that binding between AB and CC dimers is more dynamic and interactions between AB dimers are stronger. The size of the oligomer then kept growing while more dimers were added (Figure 6f).

Based on these simulation results, we propose a kinetic pathway to describe the assembly of MS2 capsid (Figure 6g). Specifically, because of the fast association through the heterotypic interface, a trimer containing two AB dimers and one CC dimer can be formed at the beginning of assembly and it is further stabilized through the homotypic AB/AB interaction. Multiple copies of these trimers are aggregated together using their 6-fold symmetric contacts, leading into the formation of a large kinetic intermediate complex at the end of the seeding stage. This intermediate complex contains three open interfaces with 5-fold symmetry and three open interfaces with 6-fold symmetry on its edge. This gives large opportunity for more copies of both AB and CC dimers to alternately join the oligomer. Therefore, the size of the oligomer can keep growing in the follow-up stage, until the formation of final capsid.

Discussions

Self-assembly is a process for many biomolecules to carry out their functions in cells by forming high-order complexes[47]. The assembly of viral capsids is one of the most important model systems for understanding self-assembling mechanism, due to their highly symmetric structures. While it is difficult to experimentally observe the nanometer-scale rapid reactions of capsid assembly, simulation methods have been developed to study the detailed dynamics of these systems. Here, using Bacteriophage MS2 as an example, we constructed a new multiscale computational framework to simulate the self-assembly of viral capsid. The capsid of Bacteriophage MS2 is particularly interesting, because its coat proteins are classified into two sequence-identical but structurally various groups: one is symmetric CC dimers; the other is asymmetric AB dimers. The homotypic interactions between AB dimers result in a 5-fold symmetric contact, while the heterotypic interactions between AB and CC dimers result in 6-fold symmetric contact. Using our multiscale simulations, we found that the assembly of capsid-like large oligomers is highly sensitive to the concentration and the stoichiometry between AB and CC dimers in the system. We further show that the functions of homotypic and heterotypic binding interfaces are not identical during the multistep kinetic process of capsid assembly. More specifically, the heterotypic binding between AB and CC dimers is much more dynamic, and plays a more important role in regulating capsid assembly. Our simulations are consistent with the previous experimental study, in which a combination of mass spectrometry and kinetic modelling suggests the preferential binding between CC and AB dimers[48]. We therefore suggest that the assembling pathway of MS2 viral capsid is dominated by the alternate growth of heterogeneous oligomers consisting of AB and CC dimers with specific stoichiometry. Our method which complements the

energetic details and computational efficiency is the first multiscale model to study capsid assembly of Bacteriophage MS2.

Current method provides the foundation for its further extension to study more biologically relevant processes of Bacteriophage MS2. For instance, although experimental evidences have shown that RNA is not required for the self-assembly of MS2 capsid, it was recently proposed that the specific interactions between capsid proteins and short sequences in the RNA genome, called packaging sites, can enhance the rate and efficiency of assembly[49–51]. Specifically for Bacteriophage MS2, binding of a stem loop from the RNA genome has been shown to function as an allosteric effector switching the structure of coat proteins from one conformation (CC dimer) to the other (AB dimer) so that assembly can be largely facilitated[52]. The RNA genome of Bacteriophage MS2 is not modeled in our method. However, the effect of RNA-protein interactions is implicitly included in the system by assuming an initial ratio of AB and CC dimers. By extending the multiscale simulation framework developed in the work to a system which contains both capsid protein dimers and RNA genome as a polymer, the conformational transition from CC to AB dimer after RNA binding can be directly modeled. We will be able to understand how the cooperative effects between RNA and coat proteins promote capsid assembly. Moreover, in the capsid of Bacteriophage MS2, there is a single copy of maturation protein in addition to the coat proteins[12]. The maturation protein is required during the attachment between MS2 and the side of the host bacterial pili. However, for model simplification, the maturation protein is not presented in the current stage of simulations. We mainly focused on how coat proteins assemble into capsid-like oligomers. In the future, the interactions between maturation protein and other coat proteins can be explicitly added into the system.

The generality of this multiscale simulation approach opens the door to its applications in assembly of other viral capsids, which are classified on the basis of their morphology. Except icosahedral symmetry, the coat proteins in some other viruses can self-assemble into arrays with helical symmetry surrounding the nucleic acids. For an example, tobacco mosaic virus (TMV) is one of the most extensively studied helical viruses. The capsid of TMV consists of a right-handed single helix of coat protein subunits. Experiments showed that self-assembly of TMV is initiated by the nucleation, in which coat proteins are sub-assembled into a two-ring structure called the ‘disk’[53]. By applying our multiscale method, we can estimate the pairwise binding rates and stability between TMV coat protein using atomic simulations, and further understand the kinetic mechanism of ‘disk’ formation and the downstream assembling pathway by rigid-body-based simulation. Ultimately, the goal of our study is to design capsid-like large protein oligomers which can be used as drug-delivery nanomaterials[54]. The development of traditional synthetic drug delivery systems, such as liposomes and polymers, is hindered by the fact that only few of them can finally reach the target tissues. In contrast, viruses have been naturally evolving to be able to deliver genetic materials into their host cells in a highly efficient manner. By computationally redesigning the binding interfaces in the natural viral capsids, our method will be able to regulate the assembling and dis-assembling kinetics of these bio-inspired nanomaterials. This will bring new insights to improve the efficacy of current drug delivery systems.

Methods

A rigid-body based diffusion-reaction algorithm for simulating capsid assembly

The first section of Results and Figure 1 describe the geometric representation for each dimer in the capsid, the spatial arrangement of binding sites on the surface of each dimer and the rules of interactions between dimers. Given the model representation, a large number of rigid bodies are generated and randomly distributed in a 3D simulation box (Figure 2b). After the initial set-up, simulations are followed by a diffusion-reaction algorithm[55]. The algorithm details were delineated in our previous works[56, 57]. In brief, each simulation step consists of two independent scenarios of movements. The first scenario is called diffusion, in which dimers are randomly diffused along their three translational and three rotational degrees of freedom. The amplitude of diffusions for each dimer is determined by its corresponding diffusion coefficient.

Diffusions are followed by the reaction scenario in which the kinetics of association and dissociation between dimers is modeled. Association between two dimers will be triggered with the probability of $P_{on}=r_{on}\times\Delta t_{RB}$, if the distance between their specific binding sites is smaller than the effective distance d_{on}. The length of each simulation time step (Δt_{RB}) is 10 nanosecond, while r_{on} and d_{on} are estimated from the residue-based kinetic Monte-Carlo simulations, as introducing in the next section. In parallel with the association, a complex has the probability to dissociate into separate dimers within each time step, which is calculated by $P_{off}=r_{off}\times\Delta t_{RB}$. The values of dissociation rate r_{off} between dimers were estimated based on the information derived from the higher-resolution structure-based simulations. After dissociation, all dimers from the original complex separate with each other and diffuse alone until they encounter others dimers in the following simulation steps. After both diffusion and reaction scenarios are completed within a corresponding time step, the new configuration is saved and the simulation time is updated. The kinetics of the system thus evolves in both Cartesian and compositional spaces through this iterative process.

A residue-based kMC algorithm for calculating the association rates between dimers

For a given pair of capsid dimers, the association rate r_{on} through a homotypic or heterotypic binding interface can be calculated by a previously developed kinetic Monte-Carlo (KMC) simulation[58]. A coarse-grained model of dimers is used in the simulation. Each residue in the model is simplified by two representative sites; one is the position of its Cα atom, while the other is the functional center of a side-chain selected based on the specific properties of a given amino acid. The simulation starts from an initial conformation, in which two dimers are placed randomly and the corresponding binding interface is separated within a given distance cutoff d_c. Following the initial conformation, both dimers undergo random diffusions. The probability to accept the diffusional movements within each simulation step is determined by the Metropolis criterion, based on the difference of binding energy before and after diffusions. The binding energy is calculated by a physics-based scoring function which contains electrostatic interaction and hydrophobic effect. More specifically, the electrostatic interaction is adopted from the Kim-Hummer model[59, 60], while the score in the hydrophobic interaction between a pair of contact residues are taken from a previous study by Kyte and Doolittle[61]. At the end of each simulation step, if the new conformation is

accepted, we will further evaluate whether these two dimers can form an encounter complex by counting how many native contacts are restored. If an encounter complex is formed through the corresponding interface based on predefined criteria, the current simulation trajectory will be terminated. Otherwise, the simulation continues until the maximal time duration is reached.

In order to effectively estimate the association rate for a given binding interface, simulations were performed under different values of distance cutoff. For a specific distance cutoff d_c, multiple trajectories (10^3) are carried out. Each trajectory starts from a relatively different initial conformation, but the initial distances between the corresponding interfaces of two domains in all trajectories are below this d_c. Complexes are successfully formed at the end of some trajectories. In contrast, two domains diffuse far away from each other at the end of other trajectories. Consequently, the value of the association rate r_{on} and effective distance of association d_{on} are derived from the statistical analysis of these trajectories, as described in our previous study. These effective rate and distance are integrated in the rigid-body based model[56, 62, 63] to simulate capsid assembly. In order to connect residue based and rigid-body based simulations with the same time-scale, the maximal time duration to terminate each trajectory of residue-based simulations is fixed to Δt_{RB}, which is the length of time step in the rigid-body simulation. As a result, each trajectory of residue-based simulation consists of 10^3 steps and each step is 0.01nanosecond, so that the total simulation time for each trajectory is 10ns, which equals the value of Δt_{RB}.

All atom DMD simulations for estimating the stability between dimers

DMD is an optimized atomic molecular dynamic featured with the much-enhanced sampling[64, 65]. DMD have indicated a proven power in accurate prediction of the protein stability, conformations and intermolecular orientations[66, 67]. More importantly, it provides alternative way in investigating the dynamics of macromolecules which typically inaccessible to conventional MD settings. The detailed description of the DMD algorithm can be found elsewhere[68, 69]. In brief, all the heavy atoms of the proteins and polar hydrogens are explicated modeled by the united-atom model. The continuous interatomic interaction potentials are break into step-wise potential functions imbedded with the molecular mechanics-based Medusa force filed[70]. The bonded interactions include van der Waals, solvation, hydrogen bond, and electrostatics. The solvation energy is modeled by the Lazaridis-Karplus implicit solvent model[71]. Considering the complexities of model system, an implicit solvent approach is appropriate for studying the dissociation between two dimers. Screened electrostatic interactions are computed by the Debye-Huckel approximation. A Debye length of 1 nm is used by assuming a water dielctric constant of 80 and a monovalent electrolyte concentration of 0.1 M. The Anderson's thermostat is used to maintain constant temperature and a periodic boundary condition with a simulation boxsize 20×20×20nm is applied.

Conformations of symmetric and asymmetric capsid dimers were obtained from RCSB Protein Data Bank with the PDB id 1BMS[45]. The basic and acidic residues for the proteins were assigned charges corresponding to their titration states under physiological conditions (pH = 7.4), i.e., Arg and Lys residues were assigned +1e while Asp and Glu were assigned

−1e, and His was neutral. Counter ions (Cl-) were added accordingly to maintain the net charge of each system as zero and to account for possible counter-ion condensation. Energy minimizations were performed for the system prior to the productive simulations. Two specific systems were studied. The first system is a tetramer that contains two AB dimers, while the other system contains an AB and a CC dimer. For each system, five independent replicas of unbiased DMD simulations were carried out at three separate temperatures: 275, 287.5 and 300K. All replicas were initiated from random seed that lasted 50ns, where the last 5 ns were used for the statistical analysis. As a result, a sum of 250 ns was obtained for each system at different temperatures.

Acknowledgement

This work was supported by the National Institutes of Health under Grant Numbers R01GM120238 and R01GM122804. The work is also partially supported by a start-up grant from Albert Einstein College of Medicine. Computational support was provided by Albert Einstein College of Medicine High Performance Computing Center. J.Z. is supported by start-up funding from the Department of Biochemistry and Biophysics at Texas A&M University and the Center for Phage Technology, jointly sponsored by Texas AgriLife and Texas A&M University. J.Z acknowlege the NIH grants R21AI137696, P01AI095208, the Welch Foundation grants A-1863, the Texas A&M Triads for Transformation grant, the TAMU School of Science Strategic Transformative Research Program and the CST*R Foundation.

References

1. Cohen FS, How Viruses Invade Cells. Biophys. J 2016, 110, 1028–1032. [PubMed: 26958878]
2. Roos WH; Ivanovska IL; Evilevitch A; Wuite GJL, Viral Capsids: Mechanical Characteristics, Genome Packaging and Delivery Mechanisms. Cell. Mol. Life Sci 2007, 64, 1484–1497. [PubMed: 17440680]
3. Katen S; Zlotnick A The Thermodynamics of Virus Capsid Assembly In Methods in Enzymology: Biothermodynamics,Vol 455*, Part A*, Johnson ML; Holt JM; Ackers GK, Eds.; Elsevier Academic Press Inc: San Diego, 2009; Vol. 455, pp 395–417.
4. McManus JJ; Charbonneau P; Zaccarelli E; Asherie N, The Physics of Protein Self-Assembly. Curr. Opin. Colloid Interface Sci 2016, 22, 73–79.
5. Aniagyei SE; DuFort C; Kao CC; Dragnea B, Self-Assembly Approaches to Nanomaterial Encapsulation in Viral Protein Cages. J. Mater. Chem 2008, 18, 3763–3774. [PubMed: 19809586]
6. Chen CC; Xing L; Stark M; Ou TW; Holla P; Xiao K; Kamita SG; Hammock BD; Lam K; Cheng RH, Chemically Activatable Viral Capsid Functionalized for Cancer Targeting. Nanomedicine 2016, 11, 377–390. [PubMed: 26786134]
7. Mannige RV; Brooks CL, Periodic Table of Virus Capsids: Implications for Natural Selection and Design. Plos One 2010, 5, 7.
8. Prasad BVV; Schmid MF Principles of Virus Structural Organization In Viral Molecular Machines, Rossmann MG; Rao VB, Eds.; Springer-Verlag Berlin: Berlin, 2012; Vol. 726, pp 17–47.
9. Salmond GPC; Fineran PC, A Century of the Phage: Past, Present and Future. Nat. Rev. Microbiol 2015, 13, 777–786. [PubMed: 26548913]
10. Meng R; Jiang M; Cui Z; Chang JY; Yang K; Jakana J; Yu X; Wang Z; Hu B; Zhang J, Structural Basis for the Adsorption of a Single-Stranded Rna Bacteriophage. Nature communications 2019, 10, 3130.
11. Toropova K; Basnak G; Twarock R; Stockley PG; Ranson NA, The Three-Dimensional Structure of Genomic Rna in Bacteriophage Ms2: Implications for Assembly. J. Mol. Biol 2008, 375, 824–836. [PubMed: 18048058]
12. Cui Z; Gorzelnik KV; Chang JY; Langlais C; Jakana J; Young R; Zhang J, Structures of Qbeta Virions, Virus-Like Particles, and the Qbeta-Mura Complex Reveal Internal Coat Proteins and the Mechanism of Host Lysis. Proc Natl Acad Sci U S A 2017, 114, 11697–11702. [PubMed: 29078304]

13. Perlmutter JD; Hagan MF Mechanisms of Virus Assembly In Annual Review of Physical Chemistry, *Vol* 66, Johnson MA; Martinez TJ, Eds.; Annual Reviews: Palo Alto, 2015; Vol. 66, pp 217–239.
14. Hagan MF Modeling Viral Capsid Assembly In Advances in Chemical Physics, Vol 155, Rice SA; Dinner AR, Eds.; John Wiley & Sons Inc: Hoboken, 2014; Vol. 155, pp 1–67. [PubMed: 25663722]
15. Zlotnick A, To Build a Virus Capsid - an Equilibrium-Model of the Self-Assembly of Polyhedral Protein Complexes. J. Mol. Biol 1994, 241, 59–67. [PubMed: 8051707]
16. Zlotnick A, Theoretical Aspects of Virus Capsid Assembly. Journal of Molecular Recognition 2005, 18, 479–490. [PubMed: 16193532]
17. Zlotnick A; Johnson JM; Wingfield PW; Stahl SJ; Endres D, A Theoretical Model Successfully Identifies Features of Hepatitis B Virus Capsid Assembly. Biochemistry 1999, 38, 14644–14652. [PubMed: 10545189]
18. Schwartz R; Shor PW; Prevelige PE; Berger B, Local Rules Simulation of the Kinetics of Virus Capsid Self-Assembly. Biophys. J 1998, 75, 2626–2636. [PubMed: 9826587]
19. Hagan MF; Chandler D, Dynamic Pathways for Viral Capsid Assembly. Biophys. J 2006, 91, 42–54. [PubMed: 16565055]
20. Hicks SD; Henley CL, Irreversible Growth Model for Virus Capsid Assembly. Phys. Rev. E 2006, 74, 17.
21. Johnston IG; Louis AA; Doye JPK, Modelling the Self-Assembly of Virus Capsids. *J. Phys.-Condes*. Matter 2010, 22, 9.
22. Rapaport DC, Self-Assembly of Polyhedral Shells: A Molecular Dynamics Study. Phys. Rev. E 2004, 70.
23. Rapaport DC, Role of Reversibility in Viral Capsid Growth: A Paradigm for Self-Assembly. Phys. Rev. Lett 2008, 101.
24. Hagan MF; Zandi R, Recent Advances in Coarse-Grained Modeling of Virus Assembly. Curr. Opin. Virol 2016, 18, 36–43. [PubMed: 27016708]
25. Nguyen HD; Reddy VS; Brooks CL, Invariant Polymorphism in Virus Capsid Assembly. J. Am. Chem. Soc 2009, 131, 2606–2614. [PubMed: 19199626]
26. Grime JMA; Dama JF; Ganser-Pornillos BK; Woodward CL; Jensen GJ; Yeager M; Voth GA, Coarse-Grained Simulation Reveals Key Features of Hiv-1 Capsid Self-Assembly. Nature Communications 2016, 7.
27. Grime JMA; Voth GA, Early Stages of the Hiv-1 Capsid Protein Lattice Formation. Biophys. J 2012, 103, 1774–1783. [PubMed: 23083721]
28. Qiao X; Jean J; Weber J; Zhu FQ; Chen B, Mechanism of Polymorphism and Curvature of Hiv Capsid Assemblies Probed by 3d Simulations with a Novel Coarse Grain Model. Biochimica Et Biophysica Acta-General Subjects 2015, 1850, 2353–2367.
29. Zhu FQ; Chen B, Monte Carlo Simulations of Hiv Capsid Protein Homodimer. Journal of Chemical Information and Modeling 2015, 55, 1361–1368. [PubMed: 26107886]
30. Wieczorek G; Zielenkiewicz P, Influence of Macromolecular Crowding on Protein-Protein Association Rates--a Brownian Dynamics Study. Biophysical journal 2008, 95, 5030–5036. [PubMed: 18757562]
31. Ermakova E, Lysozyme Dimerization: Brownian Dynamics Simulation. Journal of molecular modeling 2005, 12, 34–41. [PubMed: 16133093]
32. Haddadian EJ; Gross EL, A Brownian Dynamics Study of the Interactions of the Luminal Domains of the Cytochrome B6f Complex with Plastocyanin and Cytochrome C6: The Effects of the Rieske Fes Protein on the Interactions. Biophysical journal 2006, 91, 2589–2600. [PubMed: 16844750]
33. Dykeman EC; Stockley PG; Twarock R, Dynamic Allostery Controls Coat Protein Conformer Switching During Ms2 Phage Assembly. J. Mol. Biol 2010, 395, 916–923. [PubMed: 19913554]
34. ElSawy KM; Caves LSD; Twarock R, The Impact of Viral Rna on the Association Rates of Capsid Protein Assembly: Bacteriophage Ms2 as a Case Study. J. Mol. Biol 2010, 400, 935–947. [PubMed: 20562027]

35. Perkett MR; Mirijanian DT; Hagan MF, The Allosteric Switching Mechanism in Bacteriophage Ms2. J. Chem. Phys 2016, 145, 15.
36. Perilla JR; Schulten K, Atomistic Characterization of the Hiv Capsid from Molecular Dynamics Simulations. Biophys. J 2015, 108, 209A–209A.
37. Perilla JR; Schulten K, Physical Properties of the Hiv-1 Capsid from All-Atom Molecular Dynamics Simulations. Nature Communications 2017, 8.
38. Noid WG; Chu JW; Ayton GS; Krishna V; Izvekov S; Voth GA; Das A; Andersen HC, The Multiscale Coarse-Graining Method. I. A Rigorous Bridge between Atomistic and Coarse-Grained Models. J. Chem. Phys 2008, 128, 244114. [PubMed: 18601324]
39. Ayton GS; Noid WG; Voth GA, Multiscale Modeling of Biomolecular Systems: In Serial and in Parallel. Current Opinion in Structural Biology 2007, 17, 192–198. [PubMed: 17383173]
40. Sherwood P; Brooks BR; Sansom MSP, Multiscale Methods for Macromolecular Simulations. Current Opinion in Structural Biology 2008, 18, 630–640. [PubMed: 18721882]
41. Ramis-Conde I; Drasdo D; Anderson ARA; Chaplain MAJ, Modeling the Influence of the E-Cadherin-Beta-Catenin Pathway in Cancer Cell Invasion: A Multiscale Approach. Biophys. J 2008, 95, 155–165. [PubMed: 18339758]
42. Chakrabarti A; Verbridge S; Stroock AD; Fischbach C; Varner JD, Multiscale Models of Breast Cancer Progression. Annals of Biomedical Engineering 2012, 40, 2488–2500. [PubMed: 23008097]
43. Krobath H; Rozycki B; Lipowsky R; Weikl TR, Binding Cooperativity of Membrane Adhesion Receptors. Soft Matter 2009, 5, 3354–3361.
44. Ni CZ; Syed R; Kodandapani R; Wickersham J; Peabody DS; Ely KR, Crystal Structure of the Ms2 Coat Protein Dimer: Implications for Rna Binding and Virus Assembly. Structure 1995, 3, 255–263. [PubMed: 7788292]
45. Stonehouse NJ; Valegard K; Golmohammadi R; vandenWorm S; Walton C; Stockley PG; Liljas L, Crystal Structures of Ms2 Capsids with Mutations in the Subunit Fg Loop. J. Mol. Biol 1996, 256, 330–339. [PubMed: 8594200]
46. Su Z; Wu Y, Computational Studies of Protein-Protein Dissociation by Statistical Potential and Coarse-Grained Simulations: A Case Study on Interactions between Colicin E9 Endonuclease and Immunity Proteins. Phys Chem Chem Phys 2019, 21, 2463–2471. [PubMed: 30652698]
47. Stupp SI; Zha RH; Palmer LC; Cui H; Bitton R, Self-Assembly of Biomolecular Soft Matter. Faraday Discuss 2013, 166, 9–30. [PubMed: 24611266]
48. Morton VL; Dykeman EC; Stonehouse NJ; Ashcroft AE; Twarock R; Stockley PG, The Impact of Viral Rna on Assembly Pathway Selection. Journal of molecular biology 2010, 401, 298–308. [PubMed: 20621589]
49. Perlmutter JD; Hagan MF, The Role of Packaging Sites in Efficient and Specific Virus Assembly. J. Mol. Biol 2015, 427, 2451–2467. [PubMed: 25986309]
50. Dykeman EC; Stockley PG; Twarock R, Packaging Signals in Two Single-Stranded Rna Viruses Imply a Conserved Assembly Mechanism and Geometry of the Packaged Genome. J. Mol. Biol 2013, 425, 3235–3249. [PubMed: 23763992]
51. Stockley PG; Twarock R; Bakker SE; Barker AM; Borodavka A; Dykeman E; Ford RJ; Pearson AR; Phillips SEV; Ranson NA; Tuma R, Packaging Signals in Single-Stranded Rna Viruses: Nature's Alternative to a Purely Electrostatic Assembly Mechanism. Journal of Biological Physics 2013, 39, 277–287. [PubMed: 23704797]
52. Stockley PG; Rolfsson O; Thompson GS; Basnak G; Francese S; Stonehouse NJ; Homans SW; Ashcroft AE, A Simple, Rna-Mediated Allosteric Switch Controls the Pathway to Formation of a T=3 Viral Capsid. J. Mol. Biol 2007, 369, 541–552. [PubMed: 17434527]
53. Butler PJG, Self-Assembly of Tobacco Mosaic Virus: The Role of an Intermediate Aggregate in Generating Both Specificity and Speed. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 1999, 354, 537–550.
54. Somiya M; Liu Q; Kuroda S, Current Progress of Virus-Mimicking Nanocarriers for Drug Delivery. Nanotheranostics 2017, 1, 415–429. [PubMed: 29188175]
55. Xie Z-R; Chen J; Wu Y, A Coarse-Grained Model for the Simulations of Biomolecular Interactions in Cellular Environments. J. Chem. Phys 2014, 140, 054112. [PubMed: 24511927]

56. Chen J; Wu Y, A Multiscale Computational Model for Simulating the Kinetics of Protein Complex Assembly. Methods in molecular biology (Clifton, N.J.) 2018, 1764, 401–411.
57. Chen J; Almo SC; Wu Y, General Principles of Binding between Cell Surface Receptors and Multi-Specific Ligands: A Computational Study. PLoS computational biology 2017, 13, e1005805. [PubMed: 29016600]
58. Xie ZR; Chen J; Wu Y, Predicting Protein-Protein Association Rates Using Coarse-Grained Simulation and Machine Learning. Sci Rep 2017, 7, 46622. [PubMed: 28418043]
59. Kim YC; Hummer G, Coarse-Grained Models for Simulations of Multiprotein Complexes: Application to Ubiquitin Binding. Journal of molecular biology 2008, 375, 1416–1433. [PubMed: 18083189]
60. Ravikumar KM; Huang W; Yang S, Coarse-Grained Simulations of Protein-Protein Association: An Energy Landscape Perspective. Biophys J 2012, 103, 837–845. [PubMed: 22947945]
61. Kyte J; Doolittle RF, A Simple Method for Displaying the Hydropathic Character of a Protein. Journal of molecular biology 1982, 157, 105–132. [PubMed: 7108955]
62. Wang B; Xie ZR; Chen J; Wu Y, Integrating Structural Information to Study the Dynamics of Protein-Protein Interactions in Cells. Structure 2018.
63. Xie ZR; Chen J; Wu Y, Multiscale Model for the Assembly Kinetics of Protein Complexes. J Phys Chem B 2016, 120, 621–632. [PubMed: 26738810]
64. Proctor EA; Ding F; Dokholyan NV, Discrete Molecular Dynamics. Wiley Interdisciplinary Reviews-Computational Molecular Science 2011, 1, 80–92.
65. Shirvanyants D; Ding F; Tsao D; Ramachandran S; Dokholyan NV, Discrete Molecular Dynamics: An Efficient and Versatile Simulation Method for Fine Protein Characterization. Journal of Physical Chemistry B 2012, 116, 8375–8382.
66. Chen YW; Ding F; Nie HF; Serohijos AW; Sharma S; Wilcox KC; Yin SY; Dokholyan NV, Protein Folding: Then and Now. Archives of Biochemistry and Biophysics 2008, 469, 4–19. [PubMed: 17585870]
67. Sharma S; Ding F; Dokholyan NV, Probing Protein Aggregation Using Discrete Molecular Dynamics. Frontiers in Bioscience-Landmark 2008, 13, 4795–4807.
68. Ding F; Tsao D; Nie H; Dokholyan NV, Ab Initio Folding of Proteins with All-Atom Discrete Molecular Dynamics. Structure (London, England : 1993) 2008, 16, 1010–1018.
69. Shirvanyants D; Ding F; Tsao D; Ramachandran S; Dokholyan NV, Discrete Molecular Dynamics: An Efficient and Versatile Simulation Method for Fine Protein Characterization. The journal of physical chemistry. B 2012, 116, 8375–8382. [PubMed: 22280505]
70. Yin S; Biedermannova L; Vondrasek J; Dokholyan NV, Medusascore: An Accurate Force Field-Based Scoring Function for Virtual Drug Screening. Journal of chemical information and modeling 2008, 48, 1656–1662. [PubMed: 18672869]
71. Lazaridis T; Karplus M, Effective Energy Function for Proteins in Solution. Proteins 1999, 35, 133–152. [PubMed: 10223287]

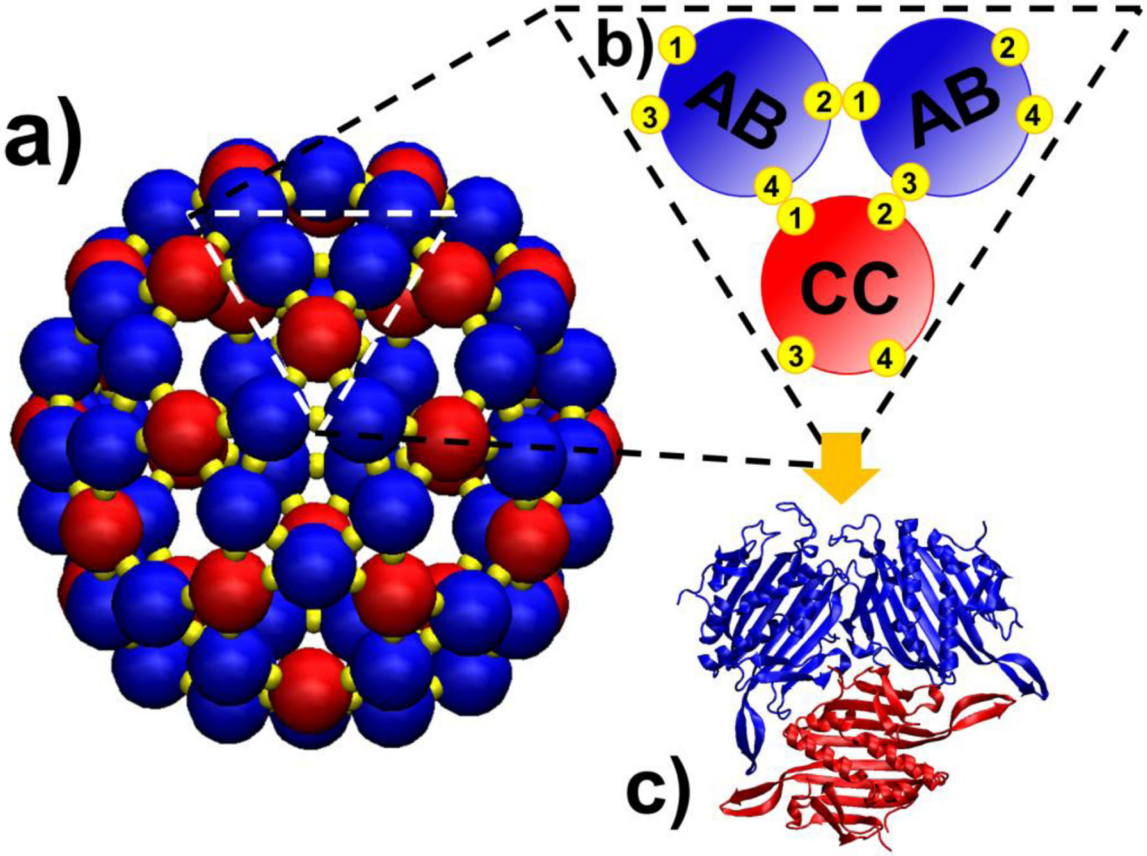

Figure 1:

A coarse-grained model was constructed for the capsid of Bacteriophage MS2 based on the symmetry of the icosahedral structure with the "triangulation number" equals 3. Each of the 90 structurally stable dimers in the capsid is represented by a spherical rigid body. The Symmetric CC dimers are shown in red, while the asymmetric AB dimers are shown in blue **(a)**. Each dimer in the capsid has four structural neighbors through two types of binding interfaces. Therefore, four binding sites are assigned on the surface of each dimer to characterize the interactions with its neighbors **(b)**. The structural details of AB and CC dimers and their homotypic or heterotypic binding interfaces are shown in **(c)**, with the same color code as the rigid-body model.

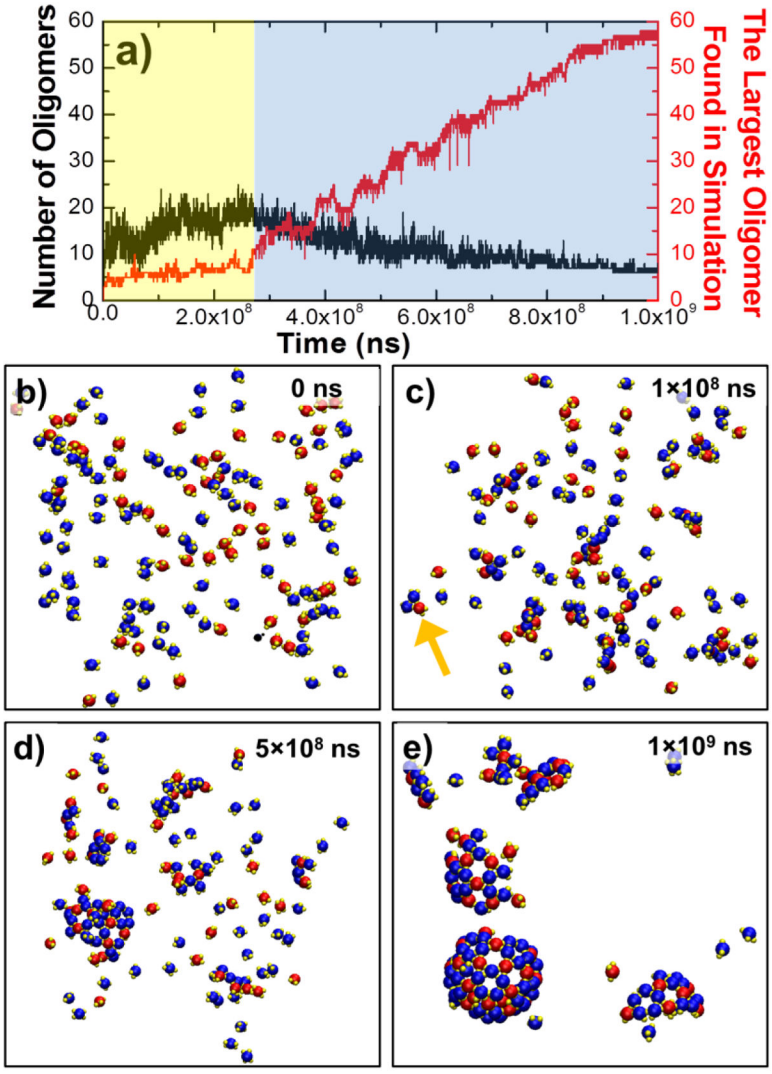

Figure 2:

We started our study from simulating the assembly of MS2 capsid with the rigid-body based model. The kinetics profile of capsid assembly is plotted in **(a)**. Along with the simulation time, we show the total number of formed oligomers (black curve) and the largest oligomer found in the system (red curve). We suggest that there are two stages involved in the assembly. The first "seeding" stage is shown by the yellow panel, while the second "growing" stage is shown by the blue panel in the figure. The simulation was started from an initial configuration **(b)**. Two representative snapshots in the middle of the assembling pathway are plotted in **(c)** and **(d)**, as indicated by the recorded time in the upper right corner of each figure. The final configuration at the end of the simulation trajectory is shown in **(e)**. A capsid-like large oligomer with more than 50 dimers was observed in the plot.

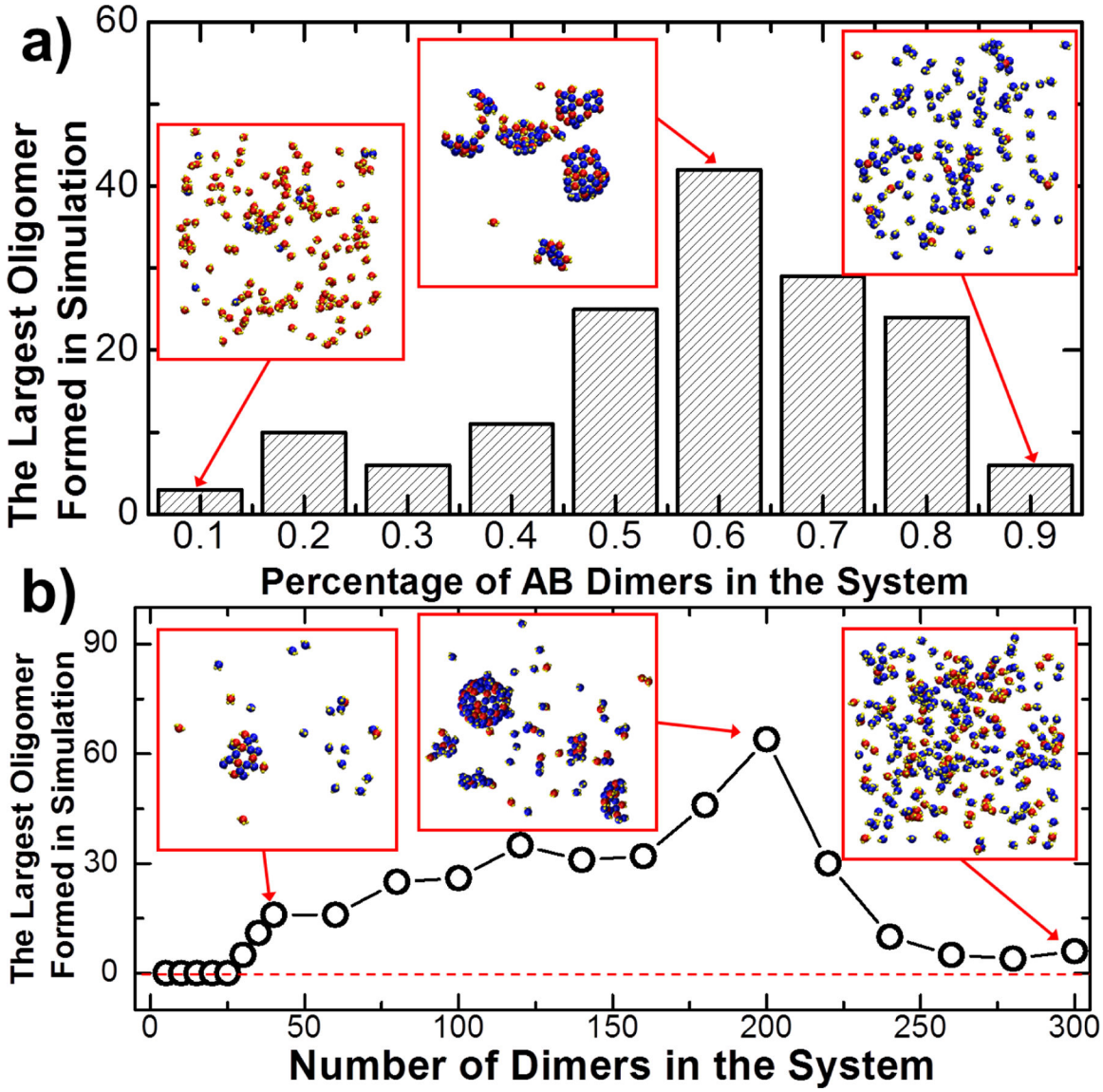

Figure 3:

We fixed the total number of dimers in the system and the size of the simulation box and changed the percentage of AB dimers from 0 to 1 to illustrate the effect of stoichiometry between AB and CC dimers on capsid assembly. The histogram of largest formed oligomer versus the percentage of AB dimers in the system is plotted in **(a),** with the final snapshots at some representative percentage values inserted as red panels. The figure suggests that the effective assembly of MS2 viral capsid requires the presence of dimers with specific stoichiometry. To further test the concentration dependence of assembly, we fixed the size of the simulation box and stoichiometry between AB and CC dimer at 2:1. We changed the total number of dimers in the system from 0 to 300. The correlation between the concentration and maximal size of formed filament is plotted in **(b)**, with the final snapshots at some representative concentrations inserted as red panels.

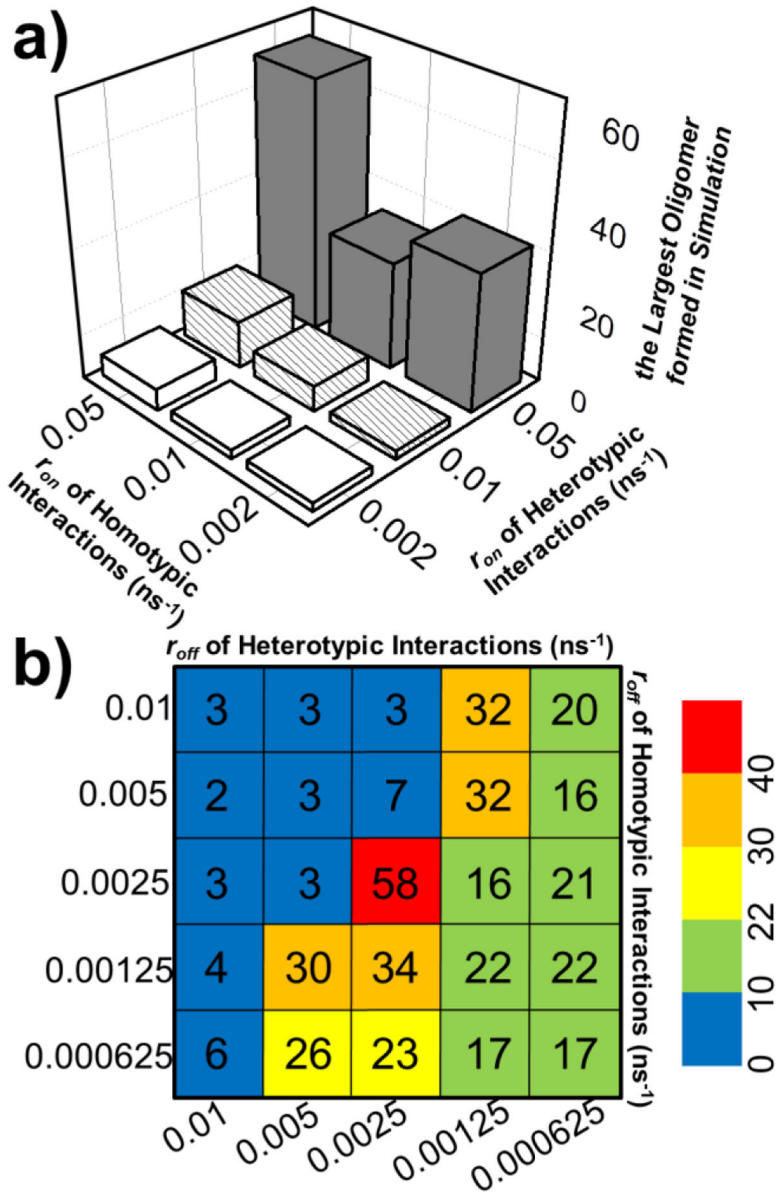

Figure 4:

We evaluated the functions of association and dissociation in regulating capsid assembly. We first fixed the dissociation rates and assigned three different values to the association rate of both homotypic and heterotypic binding interfaces. The largest sizes of oligomers formed for all nine combinations are plotted as the three-dimensional bars **(a)**, where the values of association rates for homotypic and heterotypic binding interfaces are indexed along the x-axis and the y-axis. We further fixed association rates and tuned dissociation rates into different values. Five specific values were adopted for both types of interfaces and simulation results of all combinations are summarized as the two-dimensional matrix in **(b)**. The size of largest oligomers formed for each combination is indexed by the color bar in the right panel.

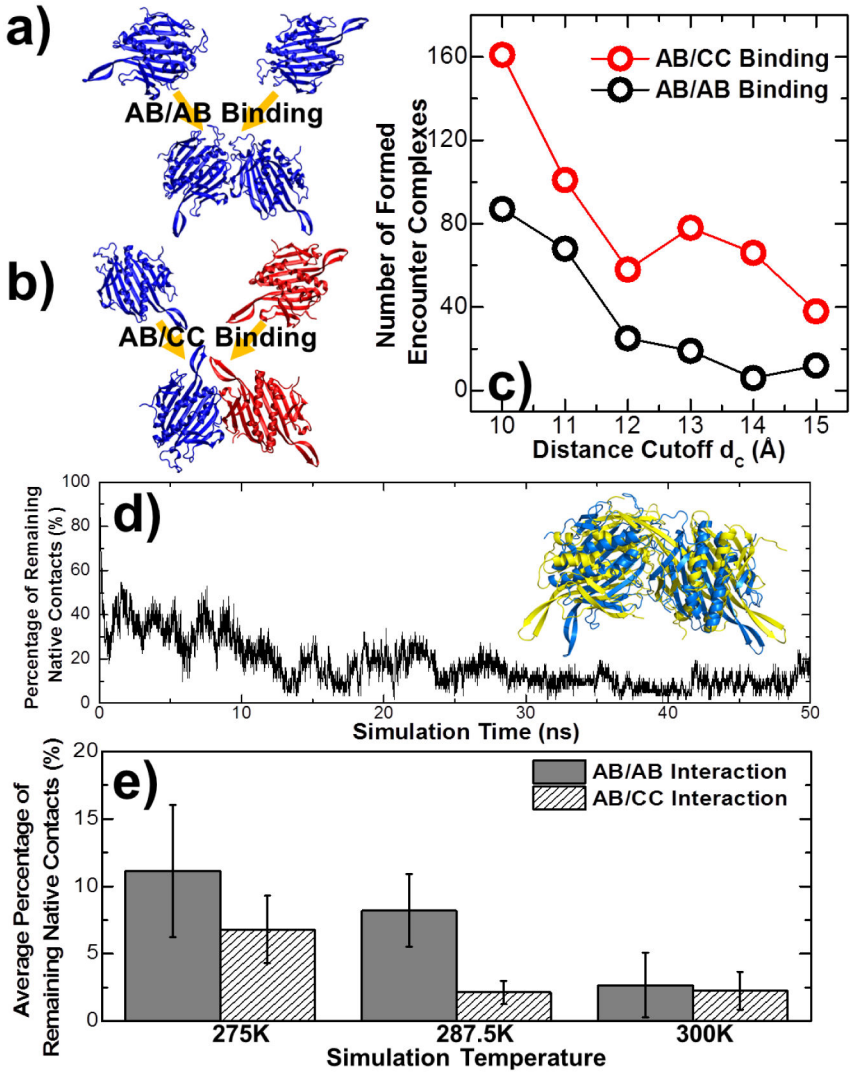

Figure 5:

Residue-based kinetic Monte-Carlo (kMC) simulation was applied to estimate the association rates between dimers through homotypic **(a)** and heterotypic **(b)** binding interfaces. The relation between distance cutoff and the frequency of forming encounter complexes for both interfaces is plotted in **(c)**. We further applied all-atom DMD simulations to estimate the stability of homotypic and heterotypic interactions between dimers. Five 50ns-long independent replicas were generated at three different temperatures. We calculated percentage of remaining native contacts between dimers along the simulations. One of the representative trajectories at 275K is plotted in **(d)**. The comparison of structures between the initial and the final complex is shown as an insert. Finally, the average percentage of remaining native contacts was calculated over the last 5 ns of all five replicas. The results are plotted as histogram with error bars in **(e)** for both homotypic and heterotypic binding systems at three temperatures.

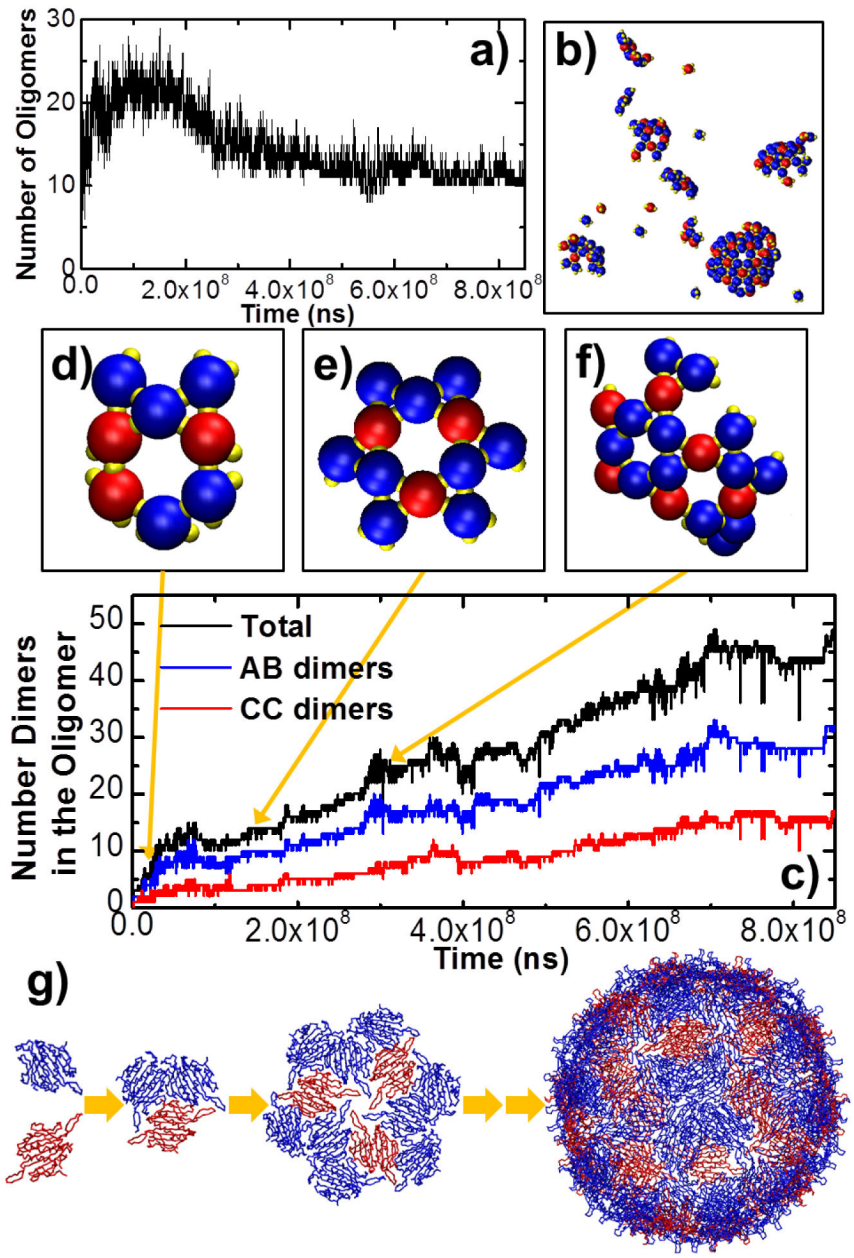

Figure 6:

Using the information derived from the atomic simulations as parameters for the association and dissociation rates, we simulated the capsid assembly again by the rigid-body based model. The total number of formed oligomers in the system is plotted in **(a)** along the simulation time, while the final configuration at the end of the trajectory is shown in **(b)**. In **(c)**, we plotted the largest oligomer found in the system (black curve), as well as the number of CC dimers (red curve) and AB dimers (blue curve) in the largest oligomer, while some representative snapshots of the largest oligomer formed along the assembly trajectory are inserted in **(d)**, **(e)** and **(f)**, respectively. Finally, we propose a kinetic pathway of capsid

assembly that consists of multiple steps and the alternate growth between AB and CC dimers, as a schematic shown in **(g)**.