Chemical Physics Letters 742 (2020) 137135

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

Research paper

Coarse-grained modeling of nanocube self-assembly system and transition network analyses

Kosuke Imamura[a], Takeshi Yamamoto[b], Hirofumi Sato[a,c,d,*]

[a] Department of Molecular Engineering, Graduate School of Engineering, Kyoto University, Kyoto 615-8510, Japan
[b] Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan
[c] Elements Strategy Initiative for Catalysts and Batteries (ESICB), Kyoto University, Kyoto 615-8520, Japan
[d] Fukui Institute for Fundamental Chemistry, Kyoto University, Kyoto 606-8103, Japan

HIGHLIGHTS

- Development of a coarse-grained model based on the data of all atom MD simulations.
- Implicit solvent model derived from the iterative Boltzmann inversion method.
- Reproduction of a self-assembled cubic structure (nanocube) with REMD simulation.
- Transition network of the self-assembly processes was studied by a cluster analysis.
- Suggestion of major routes reaching nanocube and a kinetic trapped structure.

ARTICLE INFO

Keywords:
Self-assembly
Coarse-grained model
Molecular dynamics simulation

ABSTRACT

An implicit-solvent coarse-grained (CG) model of a self-assembling molecular capsule called nanocube is developed based on all-atom molecular dynamics (MD) simulations and the iterative Boltzmann inversion method. The present model well describes the chemical specificity of constituent monomers (gear-shaped amphiphiles) and effective intermonomer interactions in aqueous methanol. Standard and replica exchange CG-MD simulations are performed and it is confirmed that the CG model can reproduce a highly ordered self-assembled structure similar to the AA model. Furthermore, the transition network is studied based on a cluster analysis, which suggests several major routes for the assembly via metastable bipyramidal states.

1. Introduction

Molecular capsules and cages have been extensively studied to explore new functions of molecular systems, such as a unique clathrate property and microscopic reaction field [1–5]. In the process of forming such structures, initial random structures reach a highly ordered structure and this process is known as self-assembly. In order to design these new molecular species, it is important to understand the whole picture of the self-assembly with atomic resolution.

In a series of experimental works [6,7], Hiraoka et al. demonstrated that gear-shaped amphiphilic (GSA) molecules self-assemble into a highly ordered, unique cubic structure (called nanocube) solely via van der Waals (vdW) interactions (i.e., without strong directional interactions such as metal coordination and H-bonding [3,4]). In this system, the meshing of the terminal pyridyl and methyl groups plays a crucial role in stabilizing the capsule structure via π–π and CH–π interactions [7–9]. Experimentally, a family of GSA-based capsule systems have been developed and their characteristics investigated [10,11]. Since it is experimentally difficult to detect transient structures, theoretical studies have also been carried out to study the reaction process. For example, a simple model of GSA molecules was developed to characterize the energy landscape via disconnectivity graphs [12], and more recently all-atom (AA) molecular dynamics (MD) simulations were performed with particular focus on solvent effects [13].

In this paper, we develop a new implicit-solvent coarse-grained (CG) model with the goal of obtaining a more comprehensive picture of nanocube self-assembly processes. CG models combine several atoms into one bead, thus reducing the number of interaction sites significantly and thus accelerating the statistical sampling [14]. A variety of CG models have been developed so far, including those for proteins [15], membranes [16], and viral capsids [17] (see Ref. [18] for reviews). Those models are often developed based on the Martini CG

* Corresponding author at: Department of Molecular Engineering, Graduate School of Engineering, Kyoto University, Kyoto 615-8510, Japan.
E-mail address: hirofumi@moleng.kyoto-u.ac.jp (H. Sato).

https://doi.org/10.1016/j.cplett.2020.137135
Received 10 December 2019; Received in revised form 17 January 2020; Accepted 19 January 2020
Available online 23 January 2020
0009-2614/

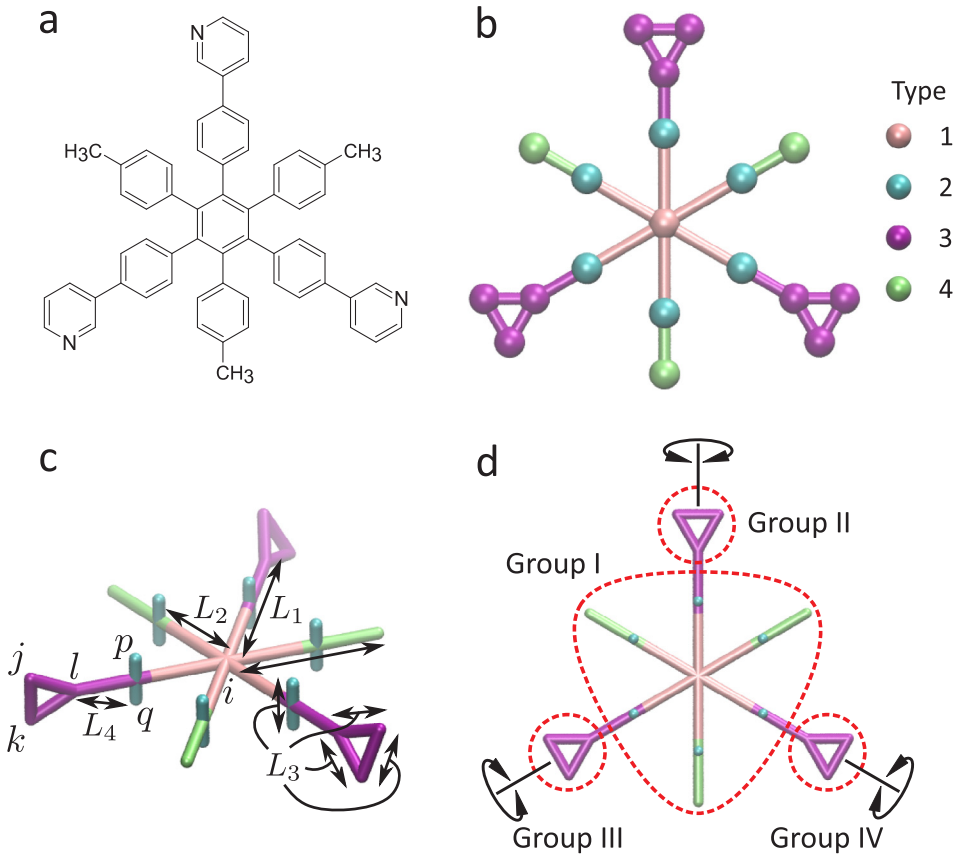

Fig. 1. (a) The structure of gear-shaped amphiphilic molecule. (b) The coarse-grained model of GSA molecule. The beads of type-2 are overlapped in this figure. (c) Topological parameters of our model, which are given by $L_1 = 7.266$ Å; $L_2 = 4.331$ Å; $L_3 = 2.416$ Å; $L_4 = \sqrt{(L_1 - L_2)^2 + (L_3/2)^2} = 3.175$ Å. (d) Definition of rigid groups and intramolecular degrees of freedom of the CG model.

force field [19]. On the other hand, it is not straightforward to apply such a general force field to the GSA molecules because of the highly specific intermonomer interactions. As such, we first develop a CG representation of GSA molecules that can describe the meshing of monomers reasonably well, while keeping the number of CG beads relatively small. Using the CG model thus obtained, we next perform standard and replica exchange molecular dynamics simulations to analyze the assembly process. In addition, following the idea of Markov state models (MSM) [20,21], we make a cluster analysis on the trajectory data and obtain insight into the transition network. We expect such a network analysis to be useful for our purpose, considering a recent application of MSM to the self-assembly of viral capsid models [21] and the qualitative similarity of assembly processes between synthetic and biological systems [22].

2. CG model of GSA molecules

2.1. Topology

Our CG model of GSA molecule is shown in Fig. 1. Following the previous studies [19,23,24], four types of CG beads were defined. First, the central benzene was represented as a single bead (labeled as type-1). In the following, six phenylene rings around the central benzene are called arm benzenes. Because of intramolecular repulsion, the arm benzenes are mostly perpendicular to the central benzene. Hence, two beads of type-2 were assigned to represent each arm benzene. To describe the π–π and CH–π interaction, three beads of type-3 were used for each pyridyl group. A methyl group attached to an arm benzene was represented as one bead. Consequently, 108 atoms were reduced to 25 CG beads. The position of the beads is shown in Fig. 1 b and c. Type-2 beads were placed at the centers of C–C bonds in the arm benzenes. To define type-3 beads, we selected two carbon atoms and one nitrogen atom, which are vertexes of an equilateral triangle, and replaced them with CG beads. The position of type-4 beads is the same as the position of the carbon atom of a methyl group. The mass of each bead is chosen as the total mass of the corresponding atomic group. For example, the mass of type-1 bead is 72 amu and that of type-2 is 38 amu (because the mass of one arm benzene is divided into two beads).

As shown in Fig. 1 d, the model is defined in terms of four rigid groups. This definition is motivated by the previous observation [13] that the monomer is rather rigid due to the steric hindrance among the arm benzenes, while the terminal pyridyl groups can rotate about the connecting bonds. The rigid group I consists of type-1, −2 and −4 beads and has D_{3h} symmetry. The groups II – IV are triangular units connected to the group I by a flexible bond, and the rigid groups are allowed to rotate around the latter bond. The rotational flexibility of the pyridyl groups is important for achieving the meshing with other monomers (and thus stabilizing the nanocube structure).

2.2. CG Force Field

The present CG force field is given by the sum of intra- and intermolecular potentials in the gas phase, plus an effective solvation potential V_{solv}:

$$U_{tot} = U_{intra} + U_{inter} + V_{solv}. \quad (1)$$

In this paper we used a two-step fitting procedure to determine the gas-phase potential ($U_{intra} + U_{inter}$) and the solvation potential (V_{solv}) based on reference AA-MD data. The reason for the above separation is that it allows us to study the solvation effect on nanocube formation by explicitly including or excluding the V_{solv} in a CG simulation. The details of

the fitting procedure are described below.

The intramolecular potentials consist of harmonic bonds, cosine-squared angles, and harmonic improper dihedrals:

$$U_{\mathrm{intra}} = \sum_{\mathrm{X=II,III,IV}} (U^{\mathrm{X}}_{\mathrm{bond}} + U^{\mathrm{X}}_{\mathrm{angle}} + U^{\mathrm{X}}_{\mathrm{improper}}), \tag{2}$$

$$U^{\mathrm{X}}_{\mathrm{bond}} = K_{\mathrm{bond}}(r^{\mathrm{X}}_{il} - L_1)^2 + K_{\mathrm{bond}}(r^{\mathrm{X}}_{pl} - L_4)^2 + K_{\mathrm{bond}}(r^{\mathrm{X}}_{ql} - L_4)^2, \tag{3}$$

$$U^{\mathrm{X}}_{\mathrm{angle}} = K_{\mathrm{angle}}(\cos\theta^{\mathrm{X}}_{ilj} - \cos 150^\circ)^2 + K_{\mathrm{angle}}(\cos\theta^{\mathrm{X}}_{ilk} - \cos 150^\circ)^2, \tag{4}$$

$$U^{\mathrm{X}}_{\mathrm{improper}} = K_{\mathrm{improper}}(\chi^{\mathrm{X}}_{ijkl})^2. \tag{5}$$

The summation in Eq. (2) is taken over the rigid groups II–IV in Fig. 1 d, and the subscripts used in Eqs. (3)–(5) are defined in Fig. 1 c. The role of Eq. (3) is to fix the position of point l. χ^{X}_{ijkl} represents the dihedral angle between the two planes formed by points i, j, k and j, k, l in the rigid group X.

The force constants in U_{intra} were taken directly from the Martini CG force field [19], except for K_{angle} which was modified to a larger value to keep the monomers sufficiently rigid. The resulting parameter values are given by K_{bond} = 23.9 kcal/(mol·Å^2); K_{angle} = 45.0 kcal/mol; K_{improper} = 23.9 kcal/(mol·rad^2).

The gas-phase intermolecular potential function (U_{inter} in Eq. (1)) is defined as the sum of Lennard-Jones (LJ) potentials that represent the vdW and the electrostatic interactions of the AA model in a coarse-grained manner. The U_{inter} is expressed as

$$U_{\mathrm{inter}} = \sum_{i,j} 4\varepsilon_{mn}\left\{\left(\frac{\sigma_{mn}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{mn}}{r_{ij}}\right)^{6}\right\}, \tag{6}$$

where m, n are bead type indices (1–4) and i, j are bead numbers in different monomers. Since there are 10 combinations of bead types, a total of 20 CG LJ parameters need to be determined. In this paper, we first determined guess CG parameters by fitting the value of U_{inter} to the interaction energy of the AA model. Specifically, we minimized the error function of the form

$$J(\mathbf{p}) = \frac{1}{N_f}\sum_{k}^{N_f} \{U_{\mathrm{inter}}(k, \mathbf{p}) - U_{\mathrm{AA}}(k)\}^2, \tag{7}$$

$$\mathbf{p} = (\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{44}, \sigma_{11}, \sigma_{12}, \ldots, \sigma_{44}). \tag{8}$$

Here, $U_{\mathrm{AA}}(k)$ is the interaction energy of a dimer in the gas phase obtained from the previous AA-MD simulation [13] (with k being the index of snapshots), and $\mathbf{p}$ represents a vector of LJ parameters. For each snapshot, the geometry of the CG model was determined from that of the AA model by using the correspondence shown in Fig. 1. N_f in Eq. (7) is the total number of snapshots chosen here as N_f = 2000. The minimization of Eq. (7) was performed via a least-squares fitting with the L-BFGS method.

Using the guess CG parameters obtained above, we next performed CG MD simulations for a dimer in the gas phase and made an additional manual tuning of the parameters so as to best reproduce the mean pairwise interaction energies. The final LJ parameters thus obtained are listed in Table 1, and the mean interaction energy for each pair of two bead types is shown in Fig. 2. We see from this figure that the mean interaction energies obtained from the CG and AA models are in good agreement with each other across all the pairs.

2.3. Solvation effect

We introduce an effective solvation potential (V_{solv} in Eq. (1)) to describe the solvent effect of a mixed solvent (specifically, a 3:1 (v/v) methanol/water mixture used in the experiments [6]). The same solvent condition was used in the previous AA MD study [13]. In this paper, we used the iterative Boltzmann inversion (IBI) method [25] to determine V_{solv} from radial distribution functions (RDFs) of the AA model. Specifically, we repeated the following iterations

$$V_0(r) = -k_B T \ln g^{\mathrm{ref}}(r), \tag{9}$$

$$V_{n+1}(r) = V_n(r) - \alpha k_B T \ln\left\{\frac{g^{\mathrm{ref}}(r)}{g^{\mathrm{CG}}_n(r)}\right\}, \tag{10}$$

where $V_n(r)$ denotes the solvation potential after n iterations, $g^{\mathrm{CG}}_n(r)$ is an RDF obtained from the n-th CG MD simulation, and α is a damping factor chosen as 0.1. We note that the above IBI procedure is not standard [25] in that it is used to determine the solvation term (V_{solv} in Eq. (1)) rather than the total intermolecular potential. This approach allows us to separately obtain $V_{\mathrm{solv}}(r)$ and thus perform CG simulations that correspond to gas and solution phases. In Eqs. (9) and (10), $g^{\mathrm{ref}}(r)$ is the reference RDF obtained from 1 μs AA MD simulation of GSA dimer in the solution phase performed in the previous study [13]. The short-range part of $g^{\mathrm{ref}}(r)$ (with $g^{\mathrm{ref}}(r) \ll 1$) was fit to an exponential form to deal with poor statistics. In the IBI procedure, each CG MD simulation was run for 3.4×10^6 steps (after an equilibration of 6×10^5 steps) with a cutoff distance of 12 Å. To remove noise, we also applied additional procedures to $g^{\mathrm{CG}}_n(r)$ after each IBI iteration. The solvation potential was smoothed via local averaging, such that

$$V_n(i) \leftarrow \frac{1}{4}\{V_n(i-1) + 2V_n(i) + V_n(i+1)\} \tag{11}$$

where i denotes the grid index for r. A total of 250 IBI iterations were performed until convergence was reached. Finally, the solvation potentials were shifted vertically to make them continuous at the cutoff distance.

According to the previous AA study [13], the GSA dimer in aqueous methanol tends to adopt an open/L-shaped structure due to the solvent-induced repulsion between the aromatic cores. To describe the latter repulsion, we introduced the solvation potential for CG pairs 1–1, 1–2,

Table 1
The LJ parameters for U_{inter} in Eq. (6).

pair (m–n)	ε_{mn} (kcal/mol)	σ_{mn} (Å)
1–1	0.4965	5.363
1–2	0.7049	4.560
1–3	0.6848	4.104
1–4	0.3679	4.798
2–2	0.6940	2.856
2–3	0.4525	3.501
2–4	0.5423	3.673
3–3	0.5384	3.409
3–4	0.3518	3.571
4–4	0.3097	3.627

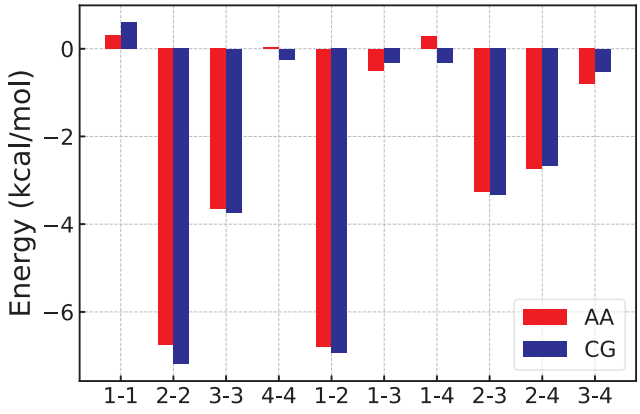

Fig. 2. The mean interaction energy of a bead pair obtained from CG MD simulations of a GSA dimer in the gas phase. The x axis represents a pair of CG bead types. The corresponding pairwise energies obtained from the AA MD simulations [13] are also shown for comparison.

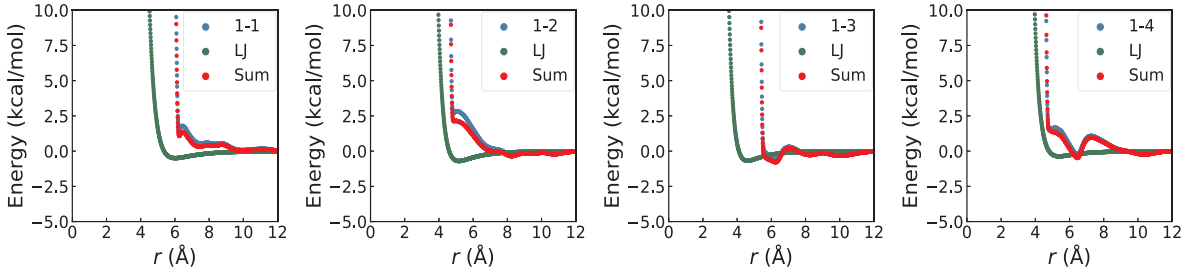

Fig. 3. Solvation potentials V_{solv} obtained from the IBI method. The x and y axes represent the distance between two CG beads and the potential energy, respectively. For comparison, the gas-phase intermolecular potential (U_{inter}) and the sum of U_{inter} and V_{solv} are also shown. Each panel displays the potentials corresponding to a given CG pair (1–1, 1–2, 1–3, and 1–4).

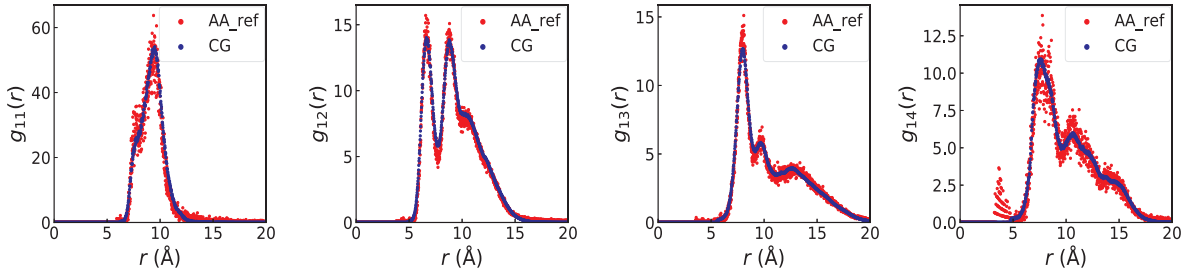

Fig. 4. Comparison of RDFs for a GSA dimer in solution obtained from AA and CG MD simulations. RDFs for bead types 1–1, 1–2, 1–3, and 1–4 are plotted against the distance between two CG beads. The reference RDFs obtained from the previous AA MD simulation [13] are shown in red, while the RDFs obtained from the present CG MD simulations are shown in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

1–3, and 1–4 in a pairwise manner. Here, we limited the number of CG pairs in V_{solv} so as to avoid possible overfitting issues.

Fig. 3 displays the solvation potentials (V_{solv}) obtained from the IBI scheme. For comparison, the gas-phase LJ potentials (U_{inter} in Eq. (1)) for the corresponding CG pairs are also shown. We see from Fig. 3 that V_{solv} becomes repulsive at larger distances than U_{inter} by reflecting the solvent-induced repulsion mentioned above [13]. Fig. 4 compares the RDFs obtained from CG-MD simulations based on the full CG potential in Eq. (1) and the reference AA RDFs. This figure shows that the peak position and height of the RDFs are in good agreement between the CG and AA models, suggesting that the present CG model describes the chemical specificity of the intermonomer interactions reasonably well.

3. Results and discussion

3.1. Standard CG MD simulations

We performed a set of standard MD simulations using the CG model obtained above. The simulation protocol used is as follows: All MD simulations were carried out using the LAMMPS program package [26]. The simulation box was a cubic cell with a side length of 70 Å, and periodic boundary condition was applied. We first performed NVE simulations of a GSA dimer using the gas-phase CG model to confirm that a time step of 10 fs is sufficient for good energy conservation. Subsequently, MD simulations based on the full CG model (including the solvation potential) were conducted under NVT condition at 300 K using the Langevin integrator with a friction constant of 5 ps^{-1}.

Fig. 5 displays the dominant structures obtained from the MD simulations. Each panel in this figure contains two structures, where the left and right sides depict a typical structure obtained from CG and AA simulations, respectively. Fig. 5 a and b compare the results obtained from the gas-phase and solution-phase MD simulations. It can be seen that both the CG and AA models adopt qualitatively similar structures. For example, a dimer is found to take stacked and staggered structures in the gas phase (Fig. 5 a), while it tends to adopt L-shaped structures in the solution phase (Fig. 5 b). The overall trend for the dimer geometry thus agrees well with the previous AA study, indicating that the present CG model describes the solvent effect in a reasonable manner.

We next performed 1 μs CG MD simulation of six monomers in the gas and solution phase to study the self-assembly process of nanocube. However, the cubic structure was not observed in both simulations and the latter simulation produced only metastable bipyramidal structures as shown in Fig. 5 c. That is, we were not able to obtain a cubic assembly structure from the standard CG MD simulations. This result suggests that the search for the global minimum structure is still challenging for the present CG model even if we add the solvation potential. The difficulty may be attributed to the meshing of GSA monomers that complicates the intermonomer interactions and hence the topography of the assembly energy landscape.

3.2. CG REMD simulations

Next, we performed replica exchange molecular dynamics (REMD) simulations to accelerate the sampling of the CG model [27]. Specifically, we used 12 replicas and set the minimum and maximum replica temperature to T_{min} = 300 K and T_{max} = 700 K, respectively. The temperature of the m-th replica was chosen geometrically as

$$T_m = T_{min}\left(\frac{T_{max}}{T_{min}}\right)^{m/11} \quad (m = 0, \ldots, 11). \tag{12}$$

A total of 12 REMD simulations were performed, where each simulation was run for 2 μs with the exchange attempt period of 10 ps. The initial configuration of each replica was prepared by randomly distributing a set of monomers in the simulation box.

Importantly, we found that the CG REMD simulations successfully produced a highly ordered nanocube structure. To illustrate the overall behavior of replicas, Fig. 6 displays the time evolution of the interaction energy (U_{inter} in Eq. (1)) calculated for the continuous REMD trajectory of each replica (labeled as Rep 0, ...,11). After rapidly forming an aggregated structure, the replicas exhibited a variety of structures with the interaction energy fluctuating from −140 to −20 kcal/mol. Typical transient structures are also shown in Fig. 6. Some of the replicas then transformed to a cubic assembly structure with the interaction energy of about −170 kcal/mol (Fig. 6 c and e). The latter structure is found to

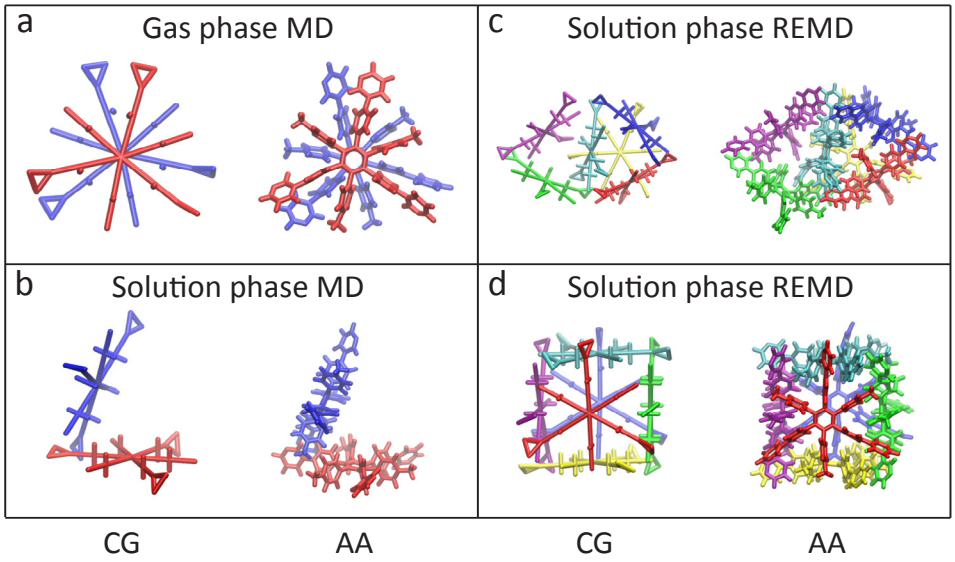

Fig. 5. Representative structures obtained from MD simulations. In each panel, the structures on the left and right sides correspond to typical structures obtained from CG and AA simulations, respectively. Panels (a) and (b) show the results of standard MD simulations (with no enhanced sampling), while panels (c) and (d) show the results of REMD simulations. For visual clarity, each monomer is displayed in different colors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

be very similar to that obtained experimentally [6]. We also observe a wide energy gap of 30 kcal/mol between the cubic and other transient states, which directly reflects the stabilization energy of the nanocube due to the meshing of GSA monomers. On the other hand, such stabilization also makes it difficult for the system to escape from the cubic structure (once the latter is formed). Consequently, the replicas reaching the nanocube state were not readily exchanged with other states. This makes a subsequent cluster analysis somewhat difficult

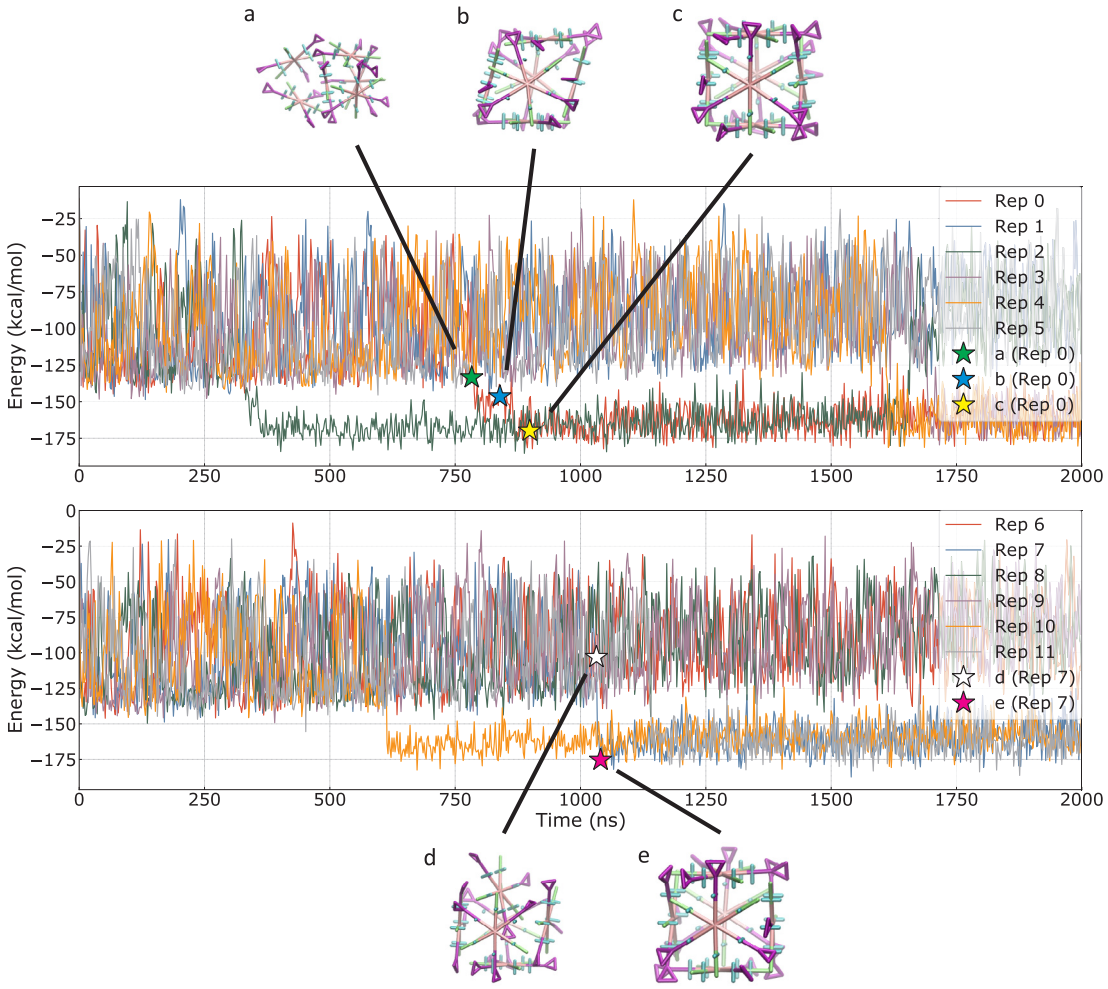

Fig. 6. The time evolution of U_{inter} in Eq. (6) obtained from one REMD simulation. Each line corresponds to the continuous trajectory of Rep 0~11. Trajectories were displayed every six replicas. The cubic structure was observed in some replicas. Structures in the processes of self-assembly were also displayed.

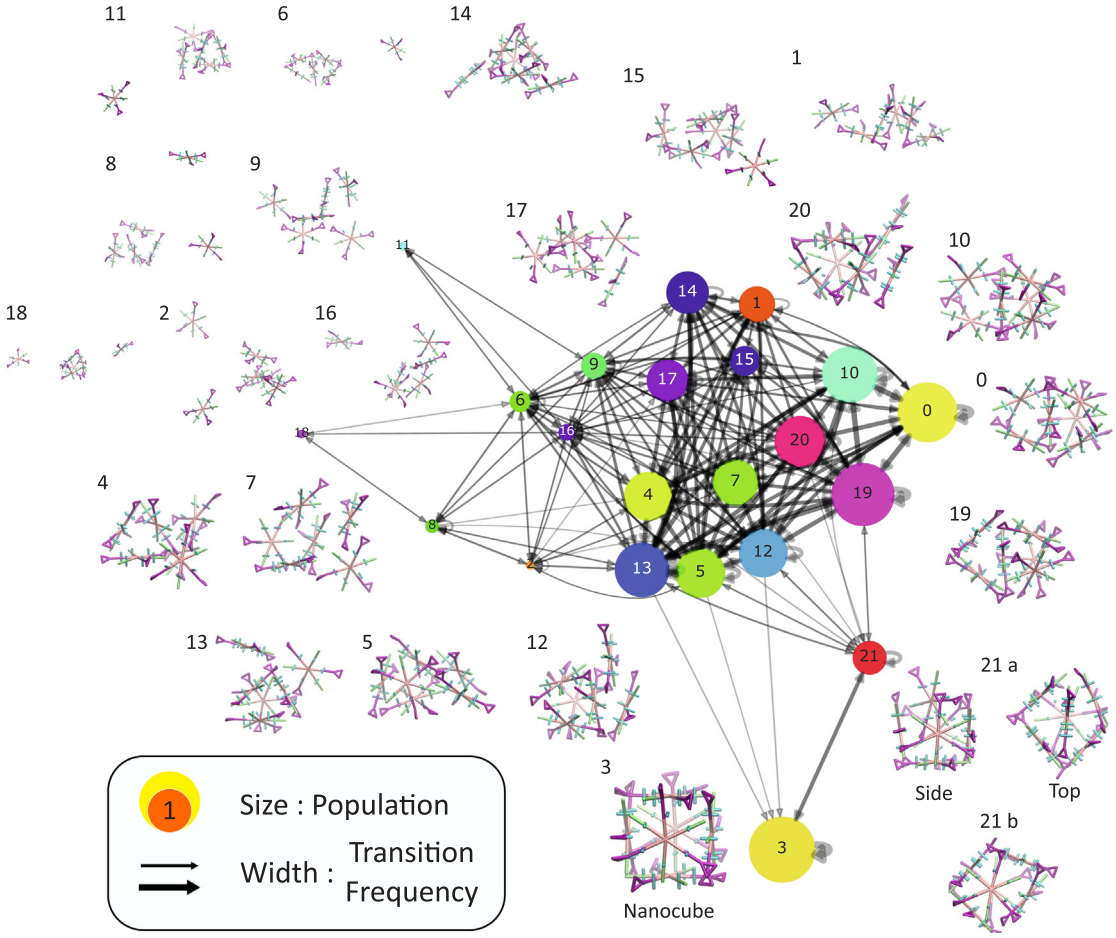

Fig. 7. Structural transition network for the hexamer system obtained from the cluster analysis. Circles and arrows represent the states defined by Ward's method and interstate transitions, respectively. For visual clarity, only the arrows with their widths greater than a certain threshold are shown.

because of the insufficient statistics of the transition between the cubic and other states (see Section 3.3). Anyway, the success of the REMD simulation suggests the capability of the present CG model for describing the specific intermonomer interactions and hence stabilizing the highly ordered cubic structure.

3.3. Trajectory analysis and transition network

We performed a cluster analysis of the REMD trajectories for the hexamer system to obtain more insight into structural transitions from the dispersed to aggregated states. In this paper we focused on the statistics and transition patterns of a continuous REMD trajectory involving all the replica temperatures; possible extensions to the analysis of temperature-specific REMD data are left for future study (see Section 4).

The procedure of our cluster analysis is as follows: First, the trajectory data were preprocessed by truncating them upon nanocube formation, which resulted in 52 trajectories with a total time of 71.7 μs. In each frame of a trajectory, the distances between the CG beads of type-1 (corresponding to the central benzene) were calculated and sorted in the descending order [28] to obtain a vector $\mathbf{v}_1 = (r_1^{11}, r_2^{11}, ..., r_{15}^{11})$ with $r_1^{11} > r_2^{11} > ... > r_{15}^{11}$ (where the system has ${}_6C_2 = 15$ distances for six monomers). Similarly, a vector $\mathbf{v}_4 = (r_1^{44}, r_2^{44}, ..., r_{153}^{44})$ was obtained for the distances between CG beads of type-4 (where the system has ${}_{3\times6}C_2 = 153$ distances). A feature vector $\mathbf{v}_f$ was then defined such that $\mathbf{v}_f = (\mathbf{v}_1\mathbf{v}_4)$ and used to characterize the system in the 168-dimensional space. A principal component analysis (PCA) was then performed for further dimensional reduction, and five principal components (PCs) were selected for cluster analyses. The mini-batch K-means method was applied to the obtained data in the PC space and all the structures were categorized into 1500 clusters. Subsequently, they were merged into 22 clusters with Ward's method [29]. The number of transitions between an arbitrary pair of two clusters was counted using a time window of 2 ns. Based on this counting, a diagram of transition network was obtained. All the analyses were performed with the MSMbuilder [30] and MDAnalysis [31] programs.

The result of our cluster analysis is shown in Fig. 7 as a network diagram. All the structures were grouped into the 22 states defined by Ward's method (labeled as 0, ...,21), with each circle corresponding to each state. The circles are depicted such that their diameters are proportional to the population of each state. The arrows represent the interstate transition with their width corresponding to the transition frequency. Representative structures of each state are also illustrated in Fig. 7.

From this figure we see that tetrahedral moieties frequently appear in the representative structures. For example, states 0, 10, and 19 contain two tetrahedral moieties and thus form a bipyramidal structure similar to those depicted in Fig. 5 c. Some other states (e.g., 5, 12–14, and 20) contain only one tetrahedral moiety with other two monomers attached in a more random manner. Structures with no tetrahedral units (e.g., states 4, 7, 15) are typically found in the high-temperature part of the trajectory. The entire diagram shows that a dense network is formed among different states, suggesting that interstate transitions occur frequently for those states.

State 3 in Fig. 7 corresponds to the cubic assembly state. It can be seen that the major routes reaching the cubic state are limited to only four routes, with the most dominant one coming from state 21. Typical structures corresponding to state 21 can be categorized into two types;

one is an open-box structure with one monomer trapped inside, and the other is a more strongly distorted cube-like structure (Fig. 6 b and d). It is also interesting to note that the bipyramidal structures (states 0, 10, and 19) are not directly connected to the final cubic state. In other words, the bipyramidal states need to first rearrange to other states for reaching the final state. This may be due to the stability of the bipyramidal states and suggests the presence of an energy barrier for the direct rearrangement to the cubic structure. The above result also seems consistent with the previous AA study that observed the kinetic trapping of similar bipyramidal states.

4. Conclusion

In this paper we have developed a new CG model for GSA molecules that self-assemble into a unique cubic structure in solution [6]. The GSAs are rather large amphiphilic molecules and possess highly specific interactions via the meshing of the terminal groups. This makes the CG modeling a challenge because of the need to accurately describe the specific intermonomer interactions. Here we constructed such a CG model based on the trajectory data obtained from the previous AA study [13] and also using the IBI method for including the solvent effect. The REMD simulations confirmed that the CG model successfully reproduces the nanocube formation, suggesting the capability of the present CG model. This is in contrast to standard MD, which failed to obtain such a structure and indicates a challenge for adequately sampling the present CG model. Nevertheless, we emphasize that the CG model has a significant advantage over the AA model in terms of computational efficiency (e.g., CG REMD simulations are several orders of magnitude cheaper than the corresponding AA simulations), which we expect to be useful for simulating greater numbers of monomers. We next studied the transition network of intermediate structures via a cluster analysis and identified several major routes connecting to the final cubic structure. Here we note that the REMD data contain samples of high temperatures, and it is preferable to make a similar analysis of temperature-specific data. In this regard, several studies recently proposed new approaches for extracting kinetic information from REMD simulations [32–34], and an application of such approaches may allow us to study temperature-specific transition pathways. It should also be noted that the present CG model is parametrized using the IBI method and the reference AA data for a dimer in aqueous methanol [13]. Since the AA samples of the dimer mainly consist of bound states [13], the present model has a limitation that V_{solv} in Eq. (1) accounts only for the solvent effects at relatively short distances. While the latter is useful for describing the rearrangement of a cluster in solution (e.g., an opening of stacked monomers into L-shaped ones), it is likely less accurate for describing monomer dissociation/association at larger distances. We therefore expect that further development of an improved CG model will be useful for studying a system containing a larger number of monomers, particularly for understanding transition networks associated with multiple nanocube formation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Dr. Toshifumi Mori (Institute for Molecular Science) for his insightful comments and kind advices. Professor Hiraoka (University of Tokyo) is grateful for the introduction of this interesting system. This work was supported by JSPS KAKENHI Grant No. JP17H03009. Theoretical computations were partly performed using Research Center for Computational Science, Okazaki, Japan.

References

[1] R. Chakrabarty, P.S. Mukherjee, P.J. Stang, Supramolecular coordination: self-assembly of finite two-and three-dimensional ensembles, Chem. Rev. 111 (2011) 6810–6918.
[2] S. Dong, B. Zheng, F. Wang, F. Huang, Supramolecular polymers constructed from macrocycle-based host-guest molecular recognition motifs, Acc. Chem. Res. 47 (2014) 1982–1994.
[3] M.M. Smulders, I.A. Riddell, C. Browne, J.R. Nitschke, Building on architectural principles for three-dimensional metallosupramolecular construction, Chem. Soc. Rev. 42 (2013) 1728–1754.
[4] K. Kobayashi, M. Yamanaka, Self-assembled capsules based on tetrafunctionalized calix [4] resorcinarene cavitands, Chem. Soc. Rev. 44 (2015) 449–466.
[5] M. Yoshizawa, L. Catti, Bent anthracene dimers as versatile building blocks for supramolecular capsules, Acc. Chem. Res. 52 (2019) 2392–2404.
[6] S. Hiraoka, K. Harano, M. Shiro, M. Shionoya, A self-assembled organic capsule formed from the union of six hexagram-shaped amphiphile molecules, J. Am. Chem. Soc. 130 (2008) 14368–14369.
[7] S. Hiraoka, T. Nakamura, M. Shiro, M. Shionoya, In-water truly monodisperse aggregation of gear-shaped amphiphiles based on hydrophobic surface engineering, J. Am. Chem. Soc. 132 (2010) 13223–13225.
[8] T. Mashiko, S. Hiraoka, U. Nagashima, M. Tachikawa, Theoretical study on substituent and solvent effects for nanocubes formed with gear-shaped amphiphile molecules, Phys. Chem. Chem. Phys. 19 (2017) 1627–1631.
[9] Y.-Y. Zhan, et al., Importance of molecular meshing for the stabilization of solvophobic assemblies, J. Org. Chem. 83 (2018) 5132–5137.
[10] Y.-Y. Zhan, et al., Induced-fit expansion and contraction of a self-assembled nanocube finely responding to neutral and anionic guests, Nat. Commun. 9 (2018) 4530.
[11] Y.-Y. Zhan, et al., Supramolecular fluorescence sensor for liquefied petroleum gas, Commun. Chem. 2 (2019) 1–7.
[12] Y. Yoshida, H. Sato, J.W. Morgan, D.J. Wales, Potential energy landscapes of tetragonal pyramid molecules, Chem. Phys. Lett. 664 (2016) 5–9.
[13] T. Yamamoto, H. Arefi, S. Shanker, H. Sato, S. Hiraoka, Self-assembly of nanocubic molecular capsules via solvent-guided formation of rectangular blocks, J. Phys. Chem. Lett. 9 (2018) 6082–6088.
[14] H.I. Ingólfsson, C.A. Lopez, J.J. Uusitalo, D.H. de Jong, S.M. Gopal, X. Periole, S.J. Marrink, The power of coarse graining in biomolecular simulations, Wiley Interdisc. Rev. Comput. Mol. Sci. 4 (2014) 225–248.
[15] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A.E. Dawid, A. Kolinski, Coarse-grained protein models and their applications, Chem. Rev. 116 (2016) 7898–7936.
[16] A.N. Leonard, E. Wang, V. Monje-Galvan, J.B. Klauda, Developing and testing of lipid force fields with applications to modeling cellular membranes, Chem. Rev. 119 (2019) 6227–6269.
[17] M.F. Hagan, R. Zandi, Recent advances in coarse-grained modeling of virus assembly, Curr. Opin. Virol. 18 (2016) 36.
[18] W.G. Noid, Perspective: Coarse-grained models for biomolecular systems, J. Chem. Phys. 139 (2013) 090901.
[19] S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, A.H.D. Vries, The MARTINI force field: coarse grained model for biomolecular simulations, J. Phys. Chem. B 111 (2007) 7812–7824.
[20] B.E. Husic, V.S. Pande, Markov state models: From an art to a science, J. Am. Chem. Soc. 140 (2018) 2386–2396.
[21] M.R. Perkett, M.F. Hagan, Using markov state models to study self-assembly, J. Chem. Phys. 140 (2014) 214101.
[22] S. Kai, T. Tateishi, T. Kojima, S. Takahashi, S. Hiraoka, Self-assembly of a pd4l8 double-walled square takes place through two kinds of metastable species, Inorg. Chem. 57 (2018) 13083–13086.
[23] V.S. Neverov, A.V. Komolkin, Coarse-grain model of the benzene ring with para-substituents in the molecule, J. Chem. Phys. 136 (2012) 094102.
[24] A. Villa, C. Peter, N.F. van der Vegt, Transferability of nonbonded interaction potentials for coarse-grained simulations: Benzene in water, J. Chem. Theory Comput. 6 (2010) 2434–2444.
[25] D. Reith, M. Pütz, F. Müller-Plathe, Deriving effective mesoscale potentials from atomistic simulations, J. Comput. Chem. 24 (2003) 1624–1636.
[26] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Comp. Phys. 117 (1995) 1–19.
[27] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chem. Phys. Lett. 314 (1999) 141–151.
[28] V.G. Grigoryan, M. Springborg, Structure and energetics of Ni clusters with up to 150 atoms, Chem. Phys. Lett. 375 (2003) 219–226.
[29] J.H. Ward Jr, Hierarchical grouping to optimize an objective function, J. Am. Stat. Assoc. 58 (1963) 236–244.
[30] K.A. Beauchamp, G.R. Bowman, T.J. Lane, L. Maibaum, I.S. Haque, V.S. Pande, MSMBuilder2: modeling conformational dynamics on the picosecond to millisecond scale, J. Chem. Theory Comput. 7 (2011) 3412–3419.
[31] N. Michaud-Agrawal, E.J. Denning, T.B. Woolf, O. Beckstein, MDAnalysis: a toolkit for the analysis of molecular dynamics simulations, J. Comput. Chem. 32 (2011) 2319–2327.
[32] H. Wu, F. Paul, C. Wehmeyer, F. Noé, Multiensemble markov models of molecular thermodynamics and kinetics, Proc. Natl. Acad. Sci. U.S.A. 113 (2016) E3221–E3230.
[33] L.S. Stelzl, G. Hummer, Kinetics from replica exchange molecular dynamics simulations, J. Chem. Theory Comput. 13 (2017) 3927–3935.
[34] C.T. Leahy, A. Kells, G. Hummer, N.-V. Buchete, E. Rosta, Peptide dimerization-dissociation rates from replica exchange molecular dynamics, J. Chem. Phys. 147 (2017) 152725.