

Dynamics in the assembly of the 30S ribosomal subunit investigated by coarse-grained simulations

Xin Liu¹, and Zhiyong Zhang^{1,2} ✉

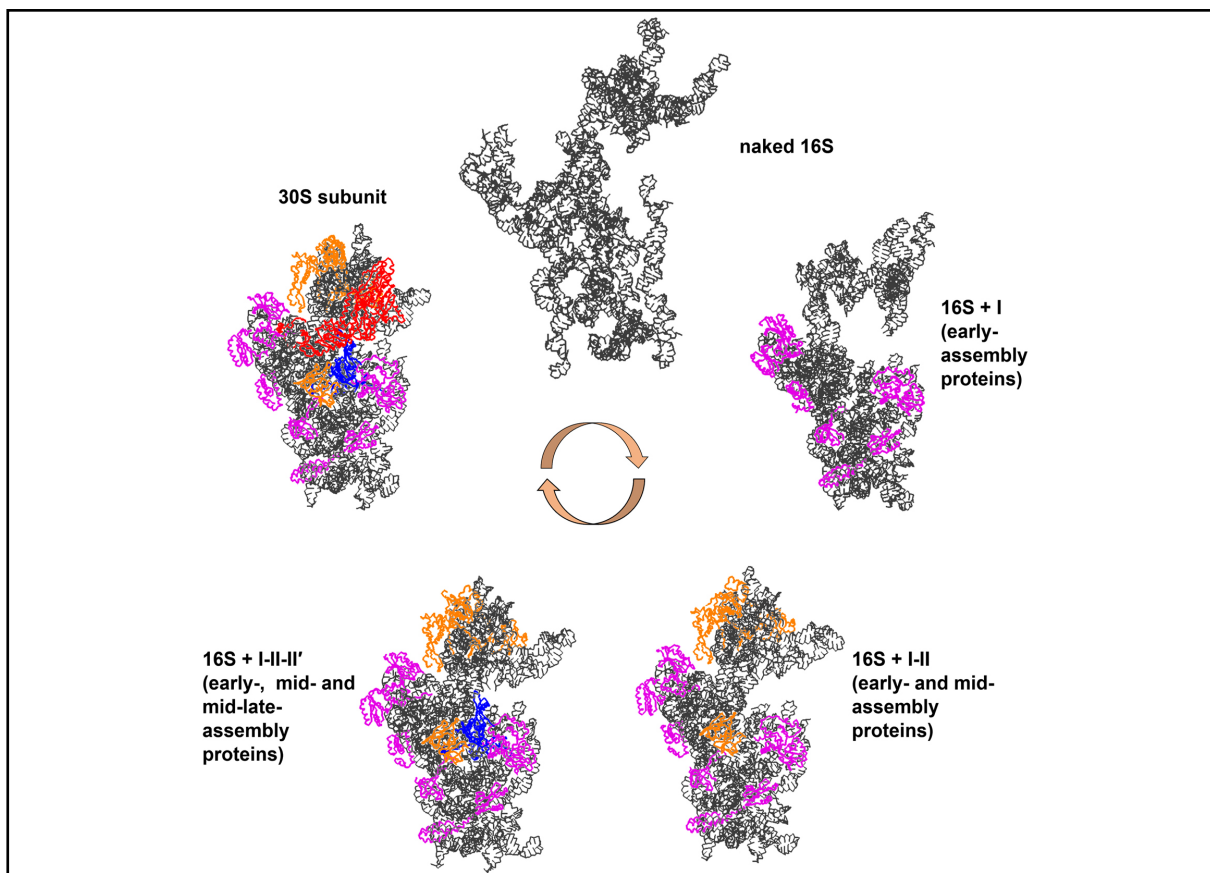
¹School of Data Science, University of Science and Technology of China, Hefei 230027, China;

²Department of Physics, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Zhiyong Zhang, E-mail: zzyzhang@ustc.edu.cn

© 2023 The Author(s). This is an open access article under the CC BY-NC-ND 4.0 license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Graphical abstract



As the ribosomal proteins are assembled sequentially at each stage, the 16S rRNA changes from an unfolded to a well-folded structure.

Public summary

- Our simulations provide some dynamic details during the assembly process of the 30S small subunit.
- S9 and S3 may have more important contributions to the assembly of the 30S subunit than other ribosomal proteins.
- Our strategy can be applied for studying the assembly dynamics of any large biomolecular complex as long as the assembly map is known.

Dynamics in the assembly of the 30S ribosomal subunit investigated by coarse-grained simulations

Xin Liu¹, and Zhiyong Zhang^{1,2} ✉

¹School of Data Science, University of Science and Technology of China, Hefei 230027, China;

²Department of Physics, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Zhiyong Zhang, E-mail: zzyzhang@ustc.edu.cn

© 2023 The Author(s). This is an open access article under the CC BY-NC-ND 4.0 license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Cite This: *JUSTC*, 2023, 53(9): 0906 (8pp)



Read Online

Abstract: The ribosome is a large biomolecular complex responsible for protein synthesis. In *Escherichia coli* (*E. coli*), a complete ribosome is composed of a 30S small subunit and a 50S large subunit. For approximately half a century, the 30S subunit has been a key model system for studying the in vitro assembly of the ribosome, and an assembly map has been proposed. However, structural details in the assembly of this protein–RNA complex remain elusive. In this paper, we conducted a series of coarse-grained simulations following the order of the assembly map to investigate conformational dynamics during the assembly process of the 30S subunit. It has been found that the tertiary structure of naked 16S rRNA is very unstable, which is the case after binding of early-assembly proteins. The mid-assembly proteins can significantly restrict the mobility of the 16S rRNA and make the latter close to the native structure. The final binding of the late-assembly proteins would fully obtain the collective motion of the 16S rRNA. In particular, proteins S9 and S3 may have more important contributions to the assembly of the 30S subunit than other S proteins. Our strategy of coarse-grained simulations can be generally used to study assembly dynamics of large biomolecular complexes as long as the assembly map is available.

Keywords: biomolecular complexes; assembly; ribosome; 30S subunit; coarse-grained simulation; principal component analysis

CLC number: Q617

Document code: A

1 Introduction

The ribosome is a large biomolecular complex that includes proteins and RNAs and is responsible for catalyzing protein synthesis^[1]. A complete ribosome contains a large subunit and a small subunit. In prokaryotes and archaea, the 50S large subunit is composed of a 23S rRNA, a 5S rRNA, and 33 proteins, whereas the 30S small subunit consists of a 16S rRNA and 21 proteins^[2].

The evolution of a large biomolecular complex can be seen as a long period of assembly, so the assembly can reflect the pathway of complex evolution^[3]. The assembly process of a complex is dynamic^[4], which is related to the order in which the same or different components come together^[3,5,6]. The assembly of the ribosome in vivo requires hundreds of assembly factors to work together under certain conditions, but the assembly in vitro can be done without any assembly factors^[7], which has been studied for a long time. As early as the 1960s, the in vitro spontaneous assembly ability of the 30S subunit in the *E. coli* ribosome was confirmed by the Nomura laboratory^[8], and subsequently, the assembly of the 50S subunit was confirmed by the Nierhaus laboratory^[9]. Later, the ability of ribosomes from other bacteria to assemble into catalytically active structures through separate natural RNA and proteins was also demonstrated^[10–15].

The 16S rRNA in the 30S ribosomal subunit has four independent domains: the 5' domain, the central domain, the 3' major domain, and the 3' minor domain. Each of the first three main domains can be independently assembled with the corresponding S proteins in vitro^[16–18]. Naked 16S rRNA is very active and requires the assembly of S proteins to form a stable ordered structure. The 30S subunit can be reconstituted in vitro with 16S rRNA as a starting point by adding the necessary S proteins to exhibit their relevant biochemical characteristics^[19]. Under reconstitution conditions, several methods of adding proteins to the 16S rRNA using combined binding allow the S proteins to be divided into three classes to form an assembly map^[20]. Primary binding proteins bind directly and independently to the 16S rRNA, which is thought to initiate the folding of each of the three main domains. Secondary binding proteins require at least one primary binding protein before binding to the 16S rRNA, and tertiary binding proteins require at least one protein from the first two stages to be assembled. Through in vitro assembly kinetic experiments, a kinetic assembly map was formed, and the S proteins were divided into early-, mid-, mid-late- and late-assembly proteins. The kinetic assembly data reflect various aspects of domain assembly and binding trends, indicating that the assembly ranges from the 5' domain to the 3' domain, which is consistent with the cotranscriptional assembly^[21].

In recent years, various experimental techniques have been used to study the assembly of the 30S subunit. Electrospray ionization mass spectrometry (MS), along with the latest cryo-EM techniques, can visualize intermediates at near-atomic resolution throughout the pathway of subunit construction and is therefore widely used to study the assembly pathway of complexes^[9]. While such data have proven to be a very useful starting point, they do not provide detailed information about dynamics in the assembly process. That is, only the structure can be obtained, and there is not enough information on how the complex forms the existing structure.

Molecular dynamics (MD) simulations have long been used in studying the dynamic processes of large biomolecules. They are generally atomic based on classical mechanics as kinetic principles, and molecular mechanical force fields are constantly improving^[22]. However, all-atom MD is very expensive in simulating large biomolecular complexes such as the ribosome, and it would be difficult to describe the complete and long time-scale kinetic process^[23]. In recent years, coarse-grained (CG) models have been used to reveal the physical mechanism of a series of protein/nucleic acid molecules^[24–28] because their computational efficiency is significantly higher than that of all-atom MD. CG simulations have been continuously optimizing and calibrating relevant force fields and methods due to their ability to display processes such as the assembly of large biomolecular complexes on a long time scale^[29]. There are many CG models available, such as the structure-based model^[30], the multiscale coarse-grained (MS-CG) method^[31], the MARTINI model^[32,33], and the SAFT model^[34].

Therefore, in this work, based on the kinetic assembly map, we study structural dynamics in the assembly process of the

30S subunit through CG simulations. The simulation data are compared with the existing experimental data of in vitro assembly, which can reveal the role of the specific S proteins in the assembly process.

2 Materials and methods

2.1 Simulated systems

The 30S ribosomal subunit of *E. coli* is composed of 16S rRNA and 21 ribosomal proteins (S proteins, named S1–S21 in decreasing order of molecular weight). An atomic model of the 30S subunit was taken from Ref. [35], in which the 16S rRNA contains 1536 nucleotides and 20 S proteins (S2–S21) are available (Fig. 1). The three main domains (5', central, and 3' major) of the 16S rRNA make up the body, platform, and head, respectively. The assembly of the S proteins is divided into four stages according to the kinetic map (Table 1). The early-assembly proteins include S4, S6, S11, S15, S16, S17, S18, and S20 (Fig. 1, magenta), the mid-assembly proteins include S7, S8, S9, S13, and S19 (Fig. 1, orange), the mid-late-assembly proteins include S5 and S12 (Fig. 1, blue), and the late-assembly proteins include S2, S3, S10, S14, and S21 (Fig. 1, red).

We built the following systems to run CG simulations. (i) The naked 16S rRNA is the starting state of the assembly. (ii) The 30S subunit is the end state of the assembly. (iii) Between the starting and end states, we added the S proteins one by one, following the order shown in Table 1.

2.2 The CG models

In the off-lattice Gō model of a protein, each amino acid

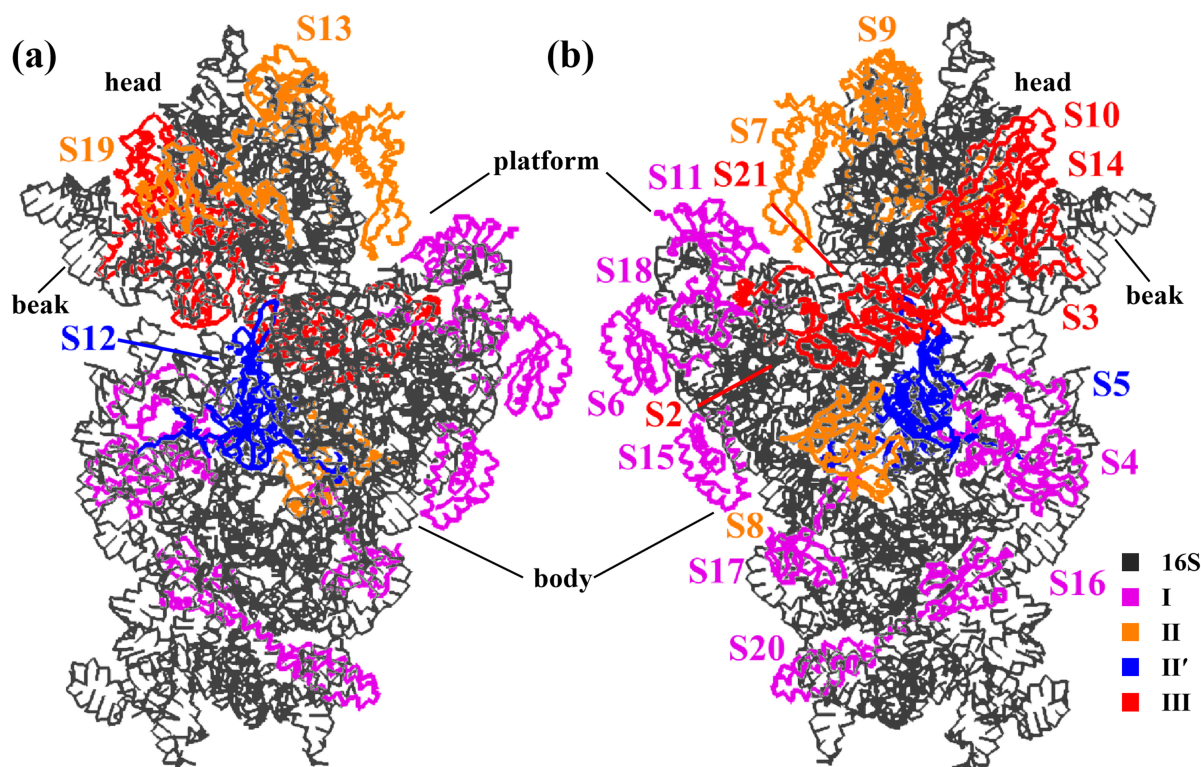


Fig. 1. Structure of the 30S ribosomal subunit. (a) The front side (the interface with the 50S subunit) and (b) the back. The 16S rRNA is colored black, and the S proteins are colored differently according to the kinetic assembly map (Table 1). The body, platform, head, and beak are labeled.

Table 1. Kinetic assembly map of the S proteins in the 30S ribosomal subunit.

Group	Protein	
I (early)	S4	
	S6	
	S11	
	S15	
	S16	
	S17	
	S18	
	S20	
	II (mid)	S7
		S8
S9		
S13		
S19		
II' (mid-late)	S5	
	S12	
III (late)	S2	
	S3	
	S10	
	S14	
	S21	

residue is represented by a CG particle, most often at the position of its C_α atom. The potential energy function is:

$$\begin{aligned}
 V(R|R_0) = & \sum_i k_b (r_{i,i+1} - r_{i,i+1,0})^2 + \sum_i k_\theta (\theta_i - \theta_{i,0})^2 + \\
 & \sum_i \{k_{\varphi_1} [1 - \cos(\varphi_i - \varphi_{i,0})] + k_{\varphi_3} [1 - \cos 3(\varphi_i - \varphi_{i,0})]\} + \\
 & \sum_{i < j - 3}^{\text{native}} \epsilon_{\text{go}} \left[5 \left(\frac{r_{ij,0}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{ij,0}}{r_{ij}} \right)^{10} \right] + \sum_{i < j - 3}^{\text{non-native}} \epsilon_{\text{cv}} \left(\frac{d}{r_{ij}} \right)^{12}. \quad (1)
 \end{aligned}$$

The first three terms represent potentials for virtual bond lengths, bond angles, and dihedral angles, respectively. The fourth term is the native contact potential between nonlocal residue pairs, where the summation is restricted to those pairs in the reference structure. The last term represents an excluded volume effect that is a penalty for nonnative contacts.

In this work, a Gō-like protein model called AICG2+^[36] was used:

$$V_{\text{AICG2+}}(R|R_0) = V_{\text{bond}} + V_{\text{loc}}^{\text{fb}} + V_{\text{loc}}^{\text{nat}} + V_{\text{nloc}}^{\text{nat}} + V_{\text{exv}}. \quad (2)$$

V_{bond} is the first term of Eq. (1). $V_{\text{loc}}^{\text{fb}}$ is a generic flexible local potential for the virtual bond angles and dihedral angles. $V_{\text{loc}}^{\text{nat}}$ is the structure-based local potential. $V_{\text{nloc}}^{\text{nat}}$ is the structure-based nonlocal contact potential, which is similar to the fourth term of Eq. (1) except that these ϵ_{go} are not uniform but determined from atomic interactions between specific residue pairs. V_{exv} is the last term of Eq. (1). Additionally, electrostatic forces were computed using the Debye–Hückel equation with a cutoff of 20 Å and ion strength of 0.15 mol/L. The

dielectric constant was set as 78.0 F/m. The off-lattice Gō model^[37] and its derivatives represent quasiharmonic fluctuations near native structures while achieving a perfect funnel energy landscape.

In the 16S rRNA, each nucleotide is represented by three CG particles, which are phosphate (P), sugar (S), and base (B). This model is a rational treatment for the chemical distribution of nucleotides. The potential is described by the following terms: the local potential, the nonlocal contact interaction, the excluded volume term, and the electrostatic interaction. Since the 16S rRNA is largely double-stranded, a PAIR_RNA potential needs to be included to maintain the stability of base pairs. This interaction has a similar form as the general contact interaction but uses different coefficients discriminated by the number of base-pairing hydrogen bonds.

For interactions between the 16S rRNA and the S proteins, we also used the Gō model with the addition of repulsion and electrostatic interactions.

2.3 CG simulations

All the CG simulations were conducted by CafeMol version 3.2.1^[30] using Langevin dynamics, with a time step of 0.3 in CafeMol time units. Each CG simulation was run for 10⁸ steps, and the temperature was set to 300 K. The overall translation and rotation of the entire molecule were allowed.

go_unit is an overall scaling factor of the pairwise interaction strength. For each native contact (intramolecule or intermolecule), the default value of go_unit is 1.0. We need to determine a proper value of go_unit to observe significant conformational dynamics of the 30S subunit while preserving its tertiary structure. Therefore, CG simulations of the 30S subunit were conducted by trying different values of go_unit, and three independent simulations were run for each go_unit. It has been found that a go_unit of 0.8 is appropriate. We then fixed the go_unit as 0.8 for both intra- and intermolecular contacts in the following CG simulations.

2.4 Principal component analysis

Principal component analysis (PCA) on a simulation trajectory of a large biomolecule is used to extract large-scale collective motion of the biomolecule from its small and random internal motion^[38]. This method comprises the following steps.

(i) All conformations in the trajectory are fitted to a reference structure to eliminate the overall translation and rotation of the biomolecule. (ii) A covariance matrix of positional fluctuation is constructed, $\sigma_{ij} = \langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle$, where r_i and r_j represent Cartesian coordinates of the particles selected for PCA, and $\langle \cdot \rangle$ means the coordinate average. (iii) The covariance matrix is diagonalized to obtain eigenvectors (PCA modes) and corresponding eigenvalues. The PCA modes with the largest eigenvalues generally describe the collective motion of the biomolecule^[39]. For each system, 1000 conformations were contained in the CG trajectory, and the P particles were used to construct the covariance matrix.

To clearly see the motion along certain PCA modes, one can project the conformations in the trajectory onto these modes individually. The root mean square inner product (RMSIP) between two sets of PCA modes can measure their dynamic similarity^[40]:

$$\text{RMSIP} = \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{v}_i^A \cdot \mathbf{v}_j^B)^2 \right)^{\frac{1}{2}}, \quad (3)$$

where \mathbf{v}_i and \mathbf{v}_j represent the i th and j th PCA modes obtained from two different sets of PCA modes (A and B), and n is the number of PCA modes used to compute RMSIP. Since the PCA modes are normalized, an RMSIP of 1 means that the two sets of PCA modes are identical, and an RMSIP of 0 indicates that they are orthogonal.

3 Results and discussion

3.1 Structural dynamics in the four-stage assembly of the 30S subunit

To measure conformational changes of the 16S rRNA in the CG simulations, the root mean squared deviations (RMSD) of all the P particles were calculated using the native structure of the 16S rRNA in the 30S subunit as the reference. In the CG simulation of the naked 16S rRNA, the system is very unstable, with an RMSD greater than 30.0 Å (Fig. 2, black). The tertiary structure of the 16S rRNA is unfolded. Both the platform and the head are expanded, and the junction between the head and the body is extremely mobile (Fig. 3a). The results indicate that, without the S proteins, the 16S rRNA itself cannot maintain its native structure.

We now want to investigate how the overall stability of the 16S rRNA changes after the S proteins at each stage are added sequentially. When only the early-assembly proteins are assembled to the 16S rRNA (denoted as 16S+I), the tertiary structure is still not stable, with RMSD values essentially between 20–30 Å (Fig. 2, magenta). These early-assembly proteins bind to the 5' and central domains (Fig. 3b, magenta), which make up the body and platform, respectively. Therefore, the body is stabilized, and the platform moves close to the body at the early stage compared to the naked 16S rRNA (Fig. 3a). However, no early-assembly protein is assembled to the 3' major domain that makes up the head, so the head is nearly as mobile as that in the naked 16S rRNA. This may explain why the RMSD values of the 16S+I system are still large. After the mid-assembly proteins are assembled

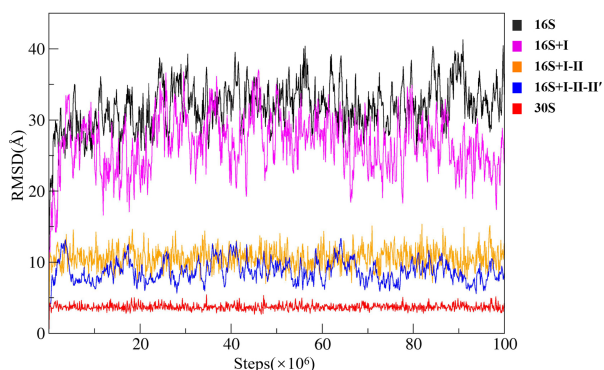


Fig. 2. Time evolution of RMSD during the CG simulations. Black: the naked 16S rRNA, magenta: the 16S rRNA with the early-assembly proteins, orange: the 16S rRNA with the early- and mid-assembly proteins, blue: the 16S rRNA with early-, mid-, and mid-late-assembly proteins, and red: the 30S subunit.

(denoted as 16S+I-II), the stability of the system is greatly increased. The RMSD values fluctuate by approximately 10.0 Å (Fig. 2, orange). According to the positions of the five mid-assembly proteins on the 16S rRNA (Fig. 3c, orange), four of them are located on the head that joins the two halves of the 3' major domain together; therefore, the mobility of the head is significantly decreased. Only one mid-assembly protein (S8) is located at the back of the body. The two mid-late assembly proteins are close to the central junction between the body and the head (Fig. 3d, blue). Once they are assembled (denoted as 16S+I-II-II'), the stability is increased slightly (Fig. 2, blue) compared to the 16S+I-II system (Fig. 2, orange). The 16S+I-II-II' system is also called the 21S reconstitution intermediate^[20], which lacks most of the head-body contacts, and the beak region in the head is flexible. The late-assembly proteins would make additional connections between the head and the body and restrict the motion of the beak (Fig. 3e, red). Therefore, in the final 30S subunit, the 16S rRNA is well folded with an average RMSD of 3.7 Å (Fig. 2, red).

The assembly order of the S proteins is basically from 5' to 3' of the 16S rRNA^[41], which does not mean that the early-assembly proteins contribute most to the stability of the system. It has been found that mid- and late-assembly proteins play the most important roles in the assembly process due to their locations on 16S rRNA.

We conducted PCA on the CG trajectory of the 30S subunit, and only P particles were used to construct the covariance matrix. Conformations in the trajectory were then projected onto a 2D subspace defined by the first two PCA modes (denoted as PC1 and PC2). The sampled region of the 30S subunit is centered at approximately (0 Å, 0 Å) (Fig. 4a, red). For each PCA mode, we took conformations with the most negative and the most positive projection values, superimposed them, and then drew arrows between the corresponding particles. Along PC1, the movement of the body tail (spur) is significant (Fig. 4b), whereas along PC2, the open/close movement between the head and the body is pronounced (Fig. 4c).

The other CG trajectories were also projected onto the same subspace. The naked 16S rRNA wildly samples a very large conformational space (Fig. 4a, black). With the S proteins at the different stages assembled, one can clearly see an “assembly path” on the subspace. At the early stage, although the sampled space is still quite large (Fig. 4a, magenta), it is somewhat narrower than that of the naked 16S rRNA (Fig. 4a, black). At the mid-stage, the conformational space is sharply decreased (Fig. 4a, orange), and at the mid-late-stage (Fig. 4a, blue), the conformational space approaches that of the folded 16S rRNA (Fig. 4a, red).

3.2 The role of specific S proteins in assembly

After investigating the S proteins at the four stages, another question is, what is the specific role of individual S proteins in the assembly of the 30S subunit? We carried out a series of CG simulations starting from the naked 16S rRNA, then adding the S proteins one by one following the order shown in Table 1, and finally ending with the 30S subunit. From S4

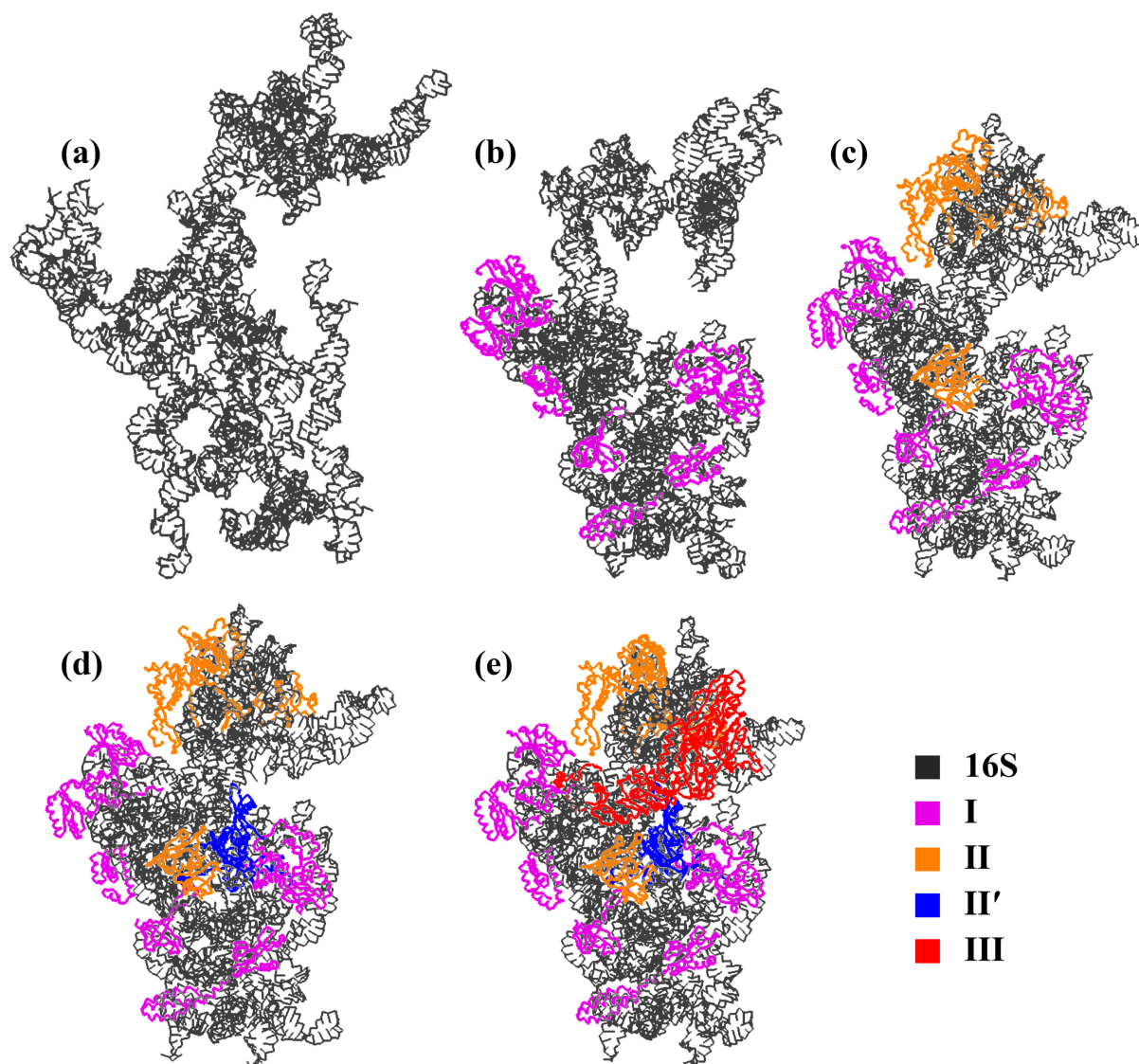


Fig. 3. Stability of the 16S rRNA after assembling the S proteins at each stage sequentially. (a) Naked 16S rRNA. (b) The 16S rRNA with only the early-assembly proteins. (c) The 16S rRNA with the early- and mid-assembly proteins. (d) The 16S rRNA with the early-, mid- and mid-late-assembly proteins. (e) The 30S subunit. At each stage, the final structure of the CG simulation is shown.

to S20, which are the early-assembly proteins, the systems are all very mobile, with RMSD values fluctuating approximately 30.0 Å (Fig. 5a, dashed curves in magenta). When S7 and S8, the first two added mid-assembly proteins, are assembled, the RMSD values are only slightly decreased (Fig. 5a, the two dashed curves in orange essentially between 20–25 Å). Once S9 is assembled, the system is significantly stabilized, with the RMSD values suddenly dropping to approximately 10.0 Å (the solid curve in orange). S9 has a long C-terminal tail inserting deeply into the head (Fig. 5b), which can join the two halves of the 3' major domain together and reduce the mobility of the head significantly. The assembly of the two mid-late-assembly proteins (S5 and S12) does not decrease the RMSD much (Fig. 5a, dashed curves in blue). When S2, the first added late-assembly protein, is assembled, the RMSD values are decreased to approximately 7.0 Å (Fig. 5a, the dashed curve in red). However, once S3 is assembled, the RMSD values again drop to approximately 4.0 Å (Fig. 5a,

the solid curve in red). S3 spans the gap between the beak in the head and the shoulder in the body (Fig. 5c), so it can stabilize the conformation of the beak and make contacts between the head and the body. The assembly of the last three late-stage proteins (S10, S14, and S21) no longer lowered the RMSD values (Fig. 5a, dashed curves in red). The separated RMSD values suggest key events in the assembly process of the 30S subunit. The first event is the folding of the head starting with the assembly of S9, and the second event is the restricted domain motion between the head (beak) and the body starting with the assembly of S3. That is, the two dividing S proteins (S9 and S3) play important roles in the assembly of the 30S subunit due to their unique locations on the 16S rRNA (Fig. 5b, c).

To quantify how the 16S rRNA obtains its collective motion in the 30S subunit during the assembly process, PCA was performed on every CG trajectory using the P particles to construct the covariance matrix. For each simulated system,

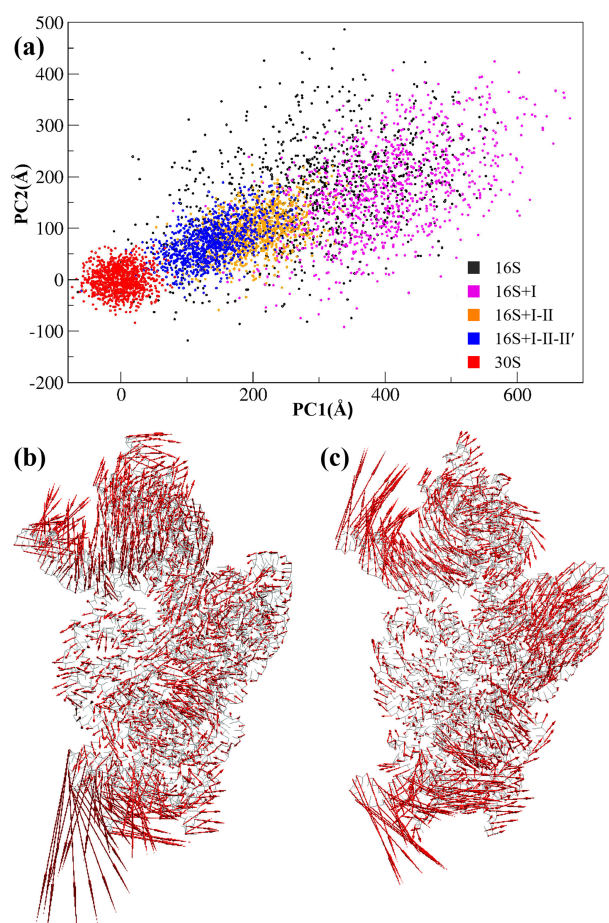


Fig. 4. An assembly path of the 30S subunit shown on a subspace defined by the PCA modes. PCA was conducted on the CG trajectory of the 30S subunit, and only P particles were used to construct the covariance matrix. The first two PCA modes with the largest eigenvalues (PC1 and PC2) were chosen to define the 2D subspace. (a) Projections of the different CG trajectories onto the subspace. (b) Collective motion along PC1. (c) Collective motion along PC2.

the first ten PCA modes with the largest eigenvalues were selected as a set. We used the PCA modes from the 30S subunit as a reference set and then computed RMSIP (Eq. (3)) between any other set of PCA modes and the reference (Fig. 6). The RMSIP between the PCA modes of the naked 16S rRNA and the reference is only 0.48, indicating that the conformational dynamics of the naked 16S rRNA are very different from those in the 30S subunit. The early-assembly proteins and the first two added mid-assembly proteins do not increase RMSIP much, remaining at approximately 0.55. Once S9 is assembled, the RMSIP is significantly increased to more than 0.8. The assembly of S9 would make the head essentially fold (Fig. 5b), and thus, the collective motion of the head relative to the body becomes dominant as that in the 30S subunit. The two mid-late-assembly proteins (S5 and S12) do not increase RMSIP much. After S3 is assembled, the RMSIP reaches 0.9, which means that the collective motion of the system is already very similar to that in the 30S subunit. The results here also support the key events during the assembly process, as shown in Fig. 5a.

3.3 Structural dynamics in a switched assembly order

What would happen if the assembly order of the S proteins is

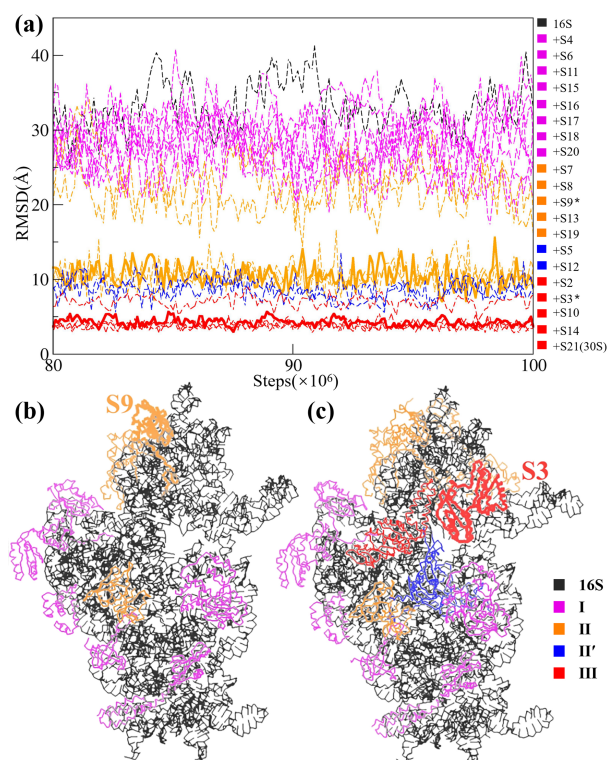


Fig. 5. The role of individual S proteins in the assembly of the 30S subunit. (a) Equilibrated (from 80 to 100 million steps of the CG simulations) RMSD values of all the simulated systems. Starting from the naked 16S rRNA, then adding the S proteins one by one following the order shown in Table 1, and finally ending with the 30S subunit. The RMSD curves at the four stages are colored differently. In the color bar on the right side, the two dividing proteins (S9 and S3) are labeled by *. The two corresponding RMSD curves are shown by solid lines, while other curves are shown by dashed lines. (b) The position of S9. (c) The position of S3.

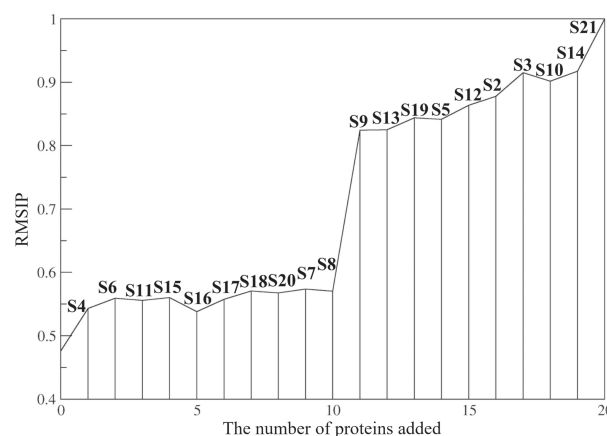


Fig. 6. Evolution of RMSIP in the assembly process of the 30S subunit. The x-axis represents the assembly order of the S proteins shown in Table 1. For each system, the RMSIP was calculated between the PCA modes of the system and those of the 30S subunit. The RMSIP ends at 1.0 because the system is the 30S subunit after S21 is assembled.

switched? To address this issue, we carried out another set of CG simulations by not following the kinetic map (Table 1) but assembling the S proteins in descending order of their molecular weights (S2, S3, ..., S21).

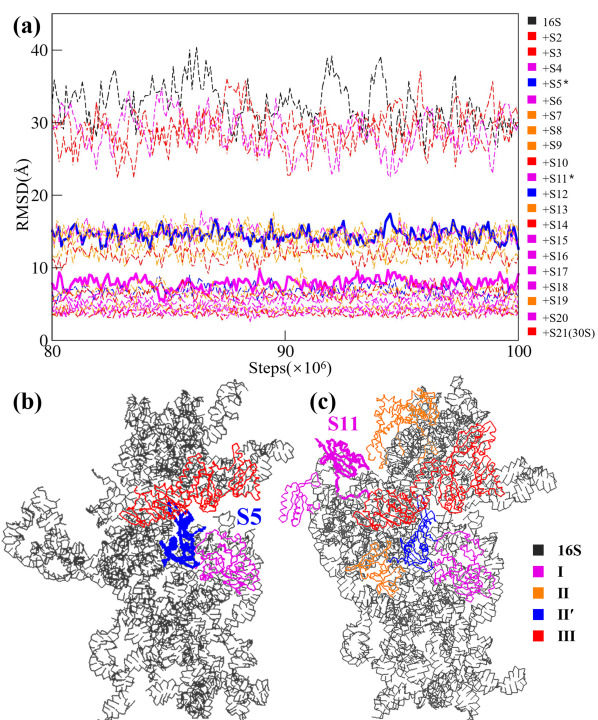


Fig. 7. The role of individual S proteins in a switched assembly order of the 30S subunit. (a) Equilibrated (from 80 to 100 million steps of the CG simulations) RMSD values of all the simulated systems. Starting from the naked 16S rRNA, then adding the S proteins one by one in descending order of their molecular weights, and finally ending with the 30S subunit. The RMSD curves are colored according to the stages to which the corresponding S proteins belong, as shown in Fig. 5a. In the color bar on the right side, the two dividing proteins (S5 and S11) are labeled by *. The two corresponding RMSD curves are shown by solid lines, while other curves are shown by dashed lines. (b) The position of S5. (c) The position of S11.

According to the RMSD curves shown in Fig. 7a, two dividing S proteins, S5 and S11, were detected in this switched assembly order. Once S5 is assembled, the RMSD values significantly drop from approximately 30.0 Å to approximately 15.0 Å (Fig. 7a, the solid curve in blue). S5 is located at the junction (Fig. 7b), which can restrict the motion between the head and the body. However, since there are only four proteins (S2 to S5) assembled at this time, the body and the head are not yet well folded, and the platform is still detached from the body (Fig. 7b). When S11 is assembled, the RMSD values decrease from more than 10.0 Å to approximately 8.0 Å (Fig. 7a, the solid curve in magenta). S11 binds to the platform, which can significantly stabilize the latter. At this time, S12 to S21 are not yet assembled, which makes the body still partially unfolded (Fig. 7c). The results indicate that the structural dynamics in the switched assembly order are different from those following the kinetic assembly map (Fig. 5). It should be noted that, without knowing the kinetic assembly map beforehand, the number of possible assembly orders is tremendous, and the CG simulations alone have a limited predictive power for determining which assembly order is reasonable. That is, the current work can only study structural dynamics during the assembly process of the 30S subunit based on its kinetic assembly map.

4 Conclusions

An increasing number of structures of large biomolecular complexes have been solved with the development of structural biology techniques, such as cryo-EM. However, with only one static structure, the dynamic process of how different components assemble into a complex is still unclear. As a complement to experiments, computer simulation may serve as an important tool for studying the assembly process of large biomolecular complexes. Due to the generally large size of a biomolecular complex and the long time scale of its assembly, CG simulation is an appropriate choice.

This paper works on the 30S small subunit of the *E. coli* ribosome, which is a model system for studying the assembly of biomolecular complexes. Based on the proposed kinetic assembly map, a series of CG simulations were carried out, from a naked 16S rRNA to a fully assembled 30S subunit. The CG simulations reveal conformational changes at the different assembly stages, and particularly those S proteins with important contributions to the assembly are detected. Our study has proven the rationality of the kinetic assembly map and provided structural details during the assembly of the 30S subunit.

Our strategy can be applied for studying structural dynamics in the assembly of any large biomolecular complex as long as the assembly map is available. However, due to the high complexity, assembly orders for many biomolecular complexes are unknown. In this case, one may obtain such information from advanced experimental techniques such as time-resolved cryo-EM^[42] or use additional computational tools to predict the assembly order of biomolecular complexes^[43].

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2021YFA1301504), the National Natural Science Foundation of China (91953101), and the Chinese Academy of Sciences Strategic Priority Research Program (XDB37040202). The Supercomputing Center of USTC provides partial computer resources for this project, and we are grateful to Mr. Yundong Zhang for his technical support.

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Xin Liu is currently a graduate student in the School of Data Science at the University of Science and Technology of China. She is under the supervision of Prof. Zhiyong Zhang. Her research mainly focuses on computer simulations of large biomolecular complex assemblies.

Zhiyong Zhang is currently a Professor in the Department of Physics at the University of Science and Technology of China (USTC). He received his Ph.D. degree in Biochemistry and Molecular Biology from USTC. His research interests include method development on multiscale modeling and integrative modeling of large biomolecular complexes.

References

- [1] Klinge S, Woolford J L. Ribosome assembly coming into focus. *Nat. Rev. Mol. Cell Biol.*, **2019**, *20* (2): 116–131.
- [2] Melnikov S, Ben-Shem A, de Loubresse N G, et al. One core, two shells: Bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.*, **2012**, *19* (6): 560–567.
- [3] Levy E D, Erba E B, Robinson C V, et al. Assembly reflects evolution of protein complexes. *Nature*, **2008**, *453* (7199): 1262–1265.
- [4] Marsh J A, Teichmann S A. Parallel dynamics and evolution: Protein conformational fluctuations and assembly reflect evolutionary changes in sequence and structure. *BioEssays*, **2014**, *36* (2): 209–218.
- [5] Marsh J A, Hernández H, Hall Z, et al. Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell*, **2013**, *153* (2): 461–470.
- [6] Friedman F K, Beychok S. Probes of subunit assembly and reconstitution pathways in multisubunit proteins. *Annu. Rev. Biochem.*, **1979**, *48*: 217–250.
- [7] Bunner A E, Nord S, Wikström M, et al. The effect of ribosome assembly cofactors on in vitro 30S subunit reconstitution. *J. Mol. Biol.*, **2010**, *398* (1): 1–7.
- [8] Traub P, Nomura M. Structure and function of *E. coli* ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **1968**, *59* (3): 777–784.
- [9] Nierhaus K H, Dohme F. Total reconstitution of functionally active 50S ribosomal subunits from *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **1974**, *71* (12): 4713–4717.
- [10] Green R, Noller H F. Reconstitution of functional 50S ribosomes from in vitro transcripts of *Bacillus stearothermophilus* 23S rRNA. *Biochemistry*, **1999**, *38* (6): 1772–1779.
- [11] Agalarov S, Yusupov M, Yusupova G. Reconstitution of functionally active *Thermus thermophilus* 30S ribosomal subunit from ribosomal 16S RNA and ribosomal proteins. In: Ennifar E, editor. *Nucleic Acid Crystallography: Methods and Protocols*. New York: Humana Press, **2016**: 303–314.
- [12] Londei P, Teuudò J, Acca M, et al. Total reconstitution of active large ribosomal subunits of the thermophilic archaeobacterium *Sulfolobus solfataricus*. *Nucleic Acids Res.*, **1986**, *14* (5): 2269–2285.
- [13] Sanchez M E, Urena D, Amils R, et al. In vitro reassembly of active large ribosomal subunits of the halophilic archaeobacterium *Haloferax mediterranei*. *Biochemistry*, **1990**, *29* (39): 9256–9261.
- [14] Fahnestock S R. Reconstitution of active 50S ribosomal subunits from *Bacillus licheniformis* and *Bacillus subtilis*. *Methods Enzymol.*, **1979**, *59*: 437–443.
- [15] Cohlberg J A, Nomura M. Reconstitution of *Bacillus stearothermophilus* 50S ribosomal subunits from purified molecular components. *J. Biol. Chem.*, **1976**, *251* (1): 209–221.
- [16] Weitzmann C J, Cunningham P R, Nurse K, et al. Chemical evidence for domain assembly of the *Escherichia coli* 30S ribosome. *FASEB J.*, **1993**, *7* (1): 177–180.
- [17] Agalarov S C, Zheleznyakova E N, Selivanova O M, et al. In vitro assembly of a ribonucleoprotein particle corresponding to the platform domain of the 30S ribosomal subunit. *Proc. Natl. Acad. Sci. U.S.A.*, **1998**, *95* (3): 999–1003.
- [18] Samaha R R, O'Brien B, O'Brien T W, et al. Independent in vitro assembly of a ribonucleoprotein particle containing the 3' domain of 16S rRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **1994**, *91* (17): 7884–7888.
- [19] Nierhaus K H. The assembly of prokaryotic ribosomes. *Biochimie*, **1991**, *73* (6): 739–755.
- [20] Mizushima S, Nomura M. Assembly mapping of 30S ribosomal proteins from *E. coli*. *Nature*, **1970**, *226* (5252): 1214–1218.
- [21] Culver G M. Assembly of the 30S ribosomal subunit. *Biopolymers*, **2003**, *68* (2): 234–249.
- [22] Tozzini V. Multiscale modeling of proteins. *Accounts. Chem. Res.*, **2010**, *43* (2): 220–230.
- [23] Zhang J, Li W, Wang J, et al. Protein folding simulations: From coarse-grained model to all-atom model. *IUBMB Life*, **2009**, *61* (6): 627–643.
- [24] Takada S. Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.*, **2012**, *22* (2): 130–137.
- [25] Tozzini V. Minimalist models for proteins: A comparative analysis. *Q. Rev. Biophys.*, **2010**, *43* (3): 333–371.
- [26] Li W F, Zhang J, Wang J, et al. Multiscale theory and computational method for biomolecule simulations. *Acta Phys. Sin.*, **2015**, *64* (9): 098701.
- [27] Yao X Q, Kenzaki H, Murakami S, et al. Drug export and allosteric coupling in a multidrug transporter revealed by molecular simulations. *Nat. Commun.*, **2010**, *1*: 117.
- [28] Yan A, Wang Y, Kloczkowski A, et al. Effects of protein subunits removal on the computed motions of partial 30S structures of the ribosome. *J. Chem. Theory Comput.*, **2008**, *4* (10): 1757–1767.
- [29] Wen B, Wang W, Zhang J, et al. Structural and dynamic properties of the C-terminal region of the *Escherichia coli* RNA chaperone Hfq: Integrative experimental and computational studies. *Phys. Chem. Chem. Phys.*, **2017**, *19* (31): 21152–21164.
- [30] Kenzaki H, Koga N, Hori N, et al. CafeMol: A coarse-grained biomolecular simulator for simulating proteins at work. *J. Chem. Theory Comput.*, **2011**, *7* (6): 1979–1989.
- [31] Izvekov S, Voth G A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B*, **2005**, *109* (7): 2469–2473.
- [32] Marrink S J, Risselada H J, Yefimov S, et al. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, **2007**, *111* (27): 7812–7824.
- [33] Souza P C T, Alessandri R, Barnoud J, et al. Martini 3: A general purpose force field for coarse-grained molecular dynamics. *Nat. Methods*, **2021**, *18* (4): 382–388.
- [34] Avendaño C, Lafitte T, Galindo A, et al. SAFT- γ force field for the simulation of molecular fluids. I. A single-site coarse grained model of carbon dioxide. *J. Phys. Chem. B*, **2011**, *115* (38): 11154–11169.
- [35] Zhang Z, Sanbonmatsu K Y, Voth G A. Key intermolecular interactions in the *E. coli* 70S ribosome revealed by coarse-grained analysis. *J. Am. Chem. Soc.*, **2011**, *133* (42): 16828–16838.
- [36] Li W, Wang W, Takada S. Energy landscape views for interplays among folding, binding, and allostery of calmodulin domains. *Proc. Natl. Acad. Sci. U.S.A.*, **2014**, *111* (29): 10550–10555.
- [37] Clementi C, Nymeyer H, Onuchic J N. Topological and energetic factors: What determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.*, **2000**, *298* (5): 937–953.
- [38] Amadei A, Linssen A B, Berendsen H J. Essential dynamics of proteins. *Proteins*, **1993**, *17* (4): 412–425.
- [39] de Groot B L, van Aalten D M, Amadei A, et al. The consistency of large concerted motions in proteins in molecular dynamics simulations. *Biophys. J.*, **1996**, *71* (4): 1707–1713.
- [40] Amadei A, Ceruso M A, Di Nola A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins*, **1999**, *36* (4): 419–424.
- [41] Powers T, Daubresse G, Noller H F. Dynamics of in vitro assembly of 16S rRNA into 30S ribosomal subunits. *J. Mol. Biol.*, **1993**, *232* (2): 362–374.
- [42] Amann S J, Keihlsler D, Bodrug T, et al. Frozen in time: Analyzing molecular dynamics with time-resolved cryo-EM. *Structure*, **2023**, *31* (1): 4–19.
- [43] Kurisaki I, Tanaka S. Computational prediction of heteromeric protein complex disassembly order using hybrid Monte Carlo/molecular dynamics simulation. *Phys. Chem. Chem. Phys.*, **2022**, *24* (17): 10575–10587.