JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

Key Intermolecular Interactions in the *E. coli* 70S Ribosome Revealed by Coarse-Grained Analysis

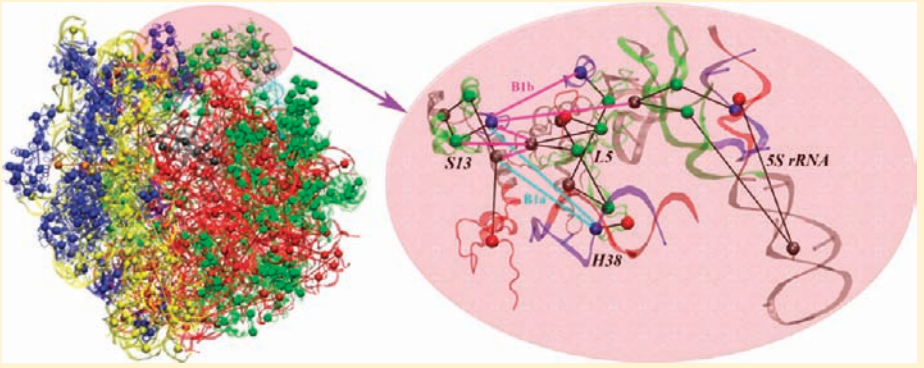
Zhiyong Zhang,[†,§] Karissa Y. Sanbonmatsu,[‡] and Gregory A. Voth*[,†]

[†]Department of Chemistry, James Franck Institute, Institute for Biophysical Dynamics, and Computation Institute, University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637, United States

[‡]Theoretical Biology and Biophysics, Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States

Supporting Information

ABSTRACT:



The ribosome is a very large complex that consists of many RNA and protein molecules and plays a central role in protein biosynthesis in all organisms. Extensive interactions between different molecules are critical to ribosomal functional dynamics. In this work, intermolecular interactions in the *Escherichia coli* 70S ribosome are investigated by coarse-grained (CG) analysis. CG models are defined to preserve dynamic domains in RNAs and proteins and to capture functional motions in the ribosome, and then the CG sites are connected by harmonic springs, and spring constants are obtained by matching the computed fluctuations to those of an all-atom molecular dynamics (MD) simulation. Those spring constants indicate how strong the interactions are between the ribosomal components, and they are in good agreement with various experimental data. Nearly all the bridges between the small and large ribosomal subunits are indicated by CG interactions with large spring constants. The head of the small subunit is very mobile because it has minimal CG interactions with the rest of the subunit; however, a large number of small subunit proteins bind to maintain the internal structure of the head. The results show a clear connection between the intermolecular interactions and the structural and functional properties of the ribosome because of the reduced complexity in domain-based CG models. The present approach also provides a useful strategy to map interactions between molecules within large biomolecular complexes since it is not straightforward to investigate these by either atomistic MD simulations or residue-based elastic network models.

INTRODUCTION

The ribosome is a large RNA−protein complex that synthesizes proteins in all cells, according to the sequence of messenger RNA (mRNA). In bacteria, the ribosome consists of a small subunit (30S) and a large subunit (50S), which together form the complete 70S ribosome. The small 30S subunit is formed from 16S ribosomal RNA (rRNA, ∼1550 nucleotides) and about 20 proteins (denoted as S1, S2, etc.). The large 50S subunit is composed of 23S rRNA (∼2900 nucleotides), 5S rRNA (∼120 nucleotides), and over 30 proteins (denoted as L1, L2, etc.). Therefore, the total molecular weight of the ribosome can be as large as 2.4 MDa.

For several decades it was difficult to infer structural details about the ribosome due to its large size and highly dynamic nature. Cryo-electron microscopy (cryo-EM) reconstructions provided low-resolution representations of the ribosome in different functional states.[1−8] Those studies, for the first time, indicated large-scale motion between the two subunits, that is, a ratchet-like rotation during translocation.[3] A major breakthrough was achieved when all-atom crystal structures of the isolated 30S and 50S subunits were solved.[9−13] These structures revealed atomic details about rRNAs, ribosomal proteins, and their extensive interactions, which provided insight into the mechanism of protein synthesis.[14−18] Progress has further been made in

Received: March 29, 2011
Published: September 12, 2011

obtaining the structure of the whole ribosome.[19−27] These high-resolution structures of the ribosome reveal the molecular-level mechanism of translation in different stages, including initiation, elongation cycle, and termination.[28−33]

All of the low- and high-resolution structures suggest that the ribosome is a highly dynamic RNA−protein complex with extensive functional interactions between its components. The large subunit associates with the small subunit to form the whole ribosome through a network of inter-subunit bridges.[4,20] The interface between the small and large subunits is occupied by molecules of transfer RNA (tRNA), which interact with mRNA through base-pairing between the codon and anticodon. The functional core of the ribosome contains only rRNAs, and thus S and L proteins are not directly involved in ribosomal functions. The majority of the proteins bind on the solvent-exposed side and periphery of the rRNA molecules, which are critical to maintain the RNA tertiary fold.[12,16,17] Therefore, it is important to study the interactions between ribosomal components in order to investigate their couplings, which are related to the functional dynamics of the whole ribosome.

Computer simulation techniques, such as atomistic molecular dynamics (MD) simulations, play an important role in the study of functional dynamics of biomolecules.[34−36] MD simulations of the ribosome have been performed and revealed ribosome dynamics at different functional stages.[37−47] They have also yielded valuable information about interactions in atomic detail, such as those between different molecules in the ribosome and those between the ribosome and its accessory factors[46] or drugs.[45] However, at the present time, MD simulations of the ribosome can only reach up to a time scale of a microsecond,[47] since it is a very large macromolecular assembly and it is computationally very expensive to simulate longer time scales. Furthermore, it provides an overly complex picture if one investigates all of the interactions in the ribosome at the atomic level.

Coarse-grained (CG) models[48−53] allow us to simulate very large biomolecular complexes like the ribosome[54] over much longer time scales than atomistic MD. The global functional dynamics of the ribosome have been investigated by residue-based elastic network models (ENMs), in which the positions of the C_α atoms for amino acids and P atoms for nucleotides are used to represent the ribosome.[55−61] These ENMs have described certain global motions[55,56] in the ribosome that are in agreement with the experimental data.[3,62,63] They also predict how the ribosome structure controls the dynamics of tRNA[57] and mRNA[58,60] to affect peptide synthesis, and how missing ribosomal protein will affect the motion of rRNAs.[59] However, one cannot determine directly how strong the interactions are between ribosomal components from these residue-based ENM models.

In this work, we combine MD data and CG models of the ribosome to investigate its intermolecular interactions. A previously developed essential dynamics coarse-graining (ED-CG) method[64−66] defines CG models with resolutions lower than one site per residue, which can preserve dynamic domains in the ribosome and capture its functional dynamics characterized by principal component analysis (PCA) of a MD trajectory[64] or ENM of a single structure.[65] CG sites in the ED-CG model are connected by effective harmonic interactions within a certain cutoff distance, and their spring constants are obtained by fluctuation-matching to those from the MD simulation.[67] Intermolecular interactions are then investigated at the CG level and compared to experimental data.

In the following sections, some details of the MD simulation of the ribosome are described, and the methods of ED-CG and fluctuation-matching are also briefly reviewed. Ribosome CG models are constructed by ED-CG from MD+PCA.[64] By analyzing dynamic domains in the ED-CG model, many functionally important regions of the ribosome are identified. Their interactions at the CG level are found to be in good agreement with results inferred from experimental data.

THEORY AND METHODS

Atomistic MD Simulation of the Ribosome. The initial starting structure of the 30S and 50S subunits and the P-tRNA anticodon stem loop were based on the structures 2I2P and 2I2T.[68] The A-tRNA and remainder of P-tRNA were modeled by aligning the structure 2J00.[26] The L7/L12 stalk was based on the structure 1MZP.[69] The MD simulation was performed with a parallel implementation of the GROMACS package (version 4.0.7),[70−73] using the AMBER99p force field.[74] The water and ion setup protocols were similar to our previous simulations.[45] The system contains the ribosome, 388 Mg^{2+}, 6272 K^+, 2831 Cl^-, and 602 587 SPCE waters, with 2 070 120 atoms in total. A 100 ns production run was conducted by using the Verlet integration scheme with a 2 fs time step.[75] The simulation was performed in a constant NPT ensemble, and the system was coupled to a temperature bath of 300 K through use of a Nosé-Hoover thermostat.[76,77] The pressure was adjusted to 1 bar with a relaxation time of 2.5 ps, and the compressibility was 4.5×10^{-5} bar^{-1}.[78] Covalent bonds were constrained using the LINCS algorithm,[79] while the cutoff distances for the van der Waals and Coulomb interactions were both chosen to be 0.9 nm. The long-range electrostatic interactions were treated by the PME algorithm,[80] with a tolerance of 10^{-5} and an interpolation order of 4.

ED-CG Ribosome Models. The details of the ED-CG methodology with MD+PCA have been described previously.[64] From the MD simulation of the ribosome, a trajectory of the center-of-mass (COM) of each residue (amino acids or nucleotides) was constructed. This trajectory, with a number of residues $n = 10\,986$, was used to perform PCA. After removal of the translational and rotational motion in the MD simulation by least-squares fitting each frame to a reference structure, a covariance matrix $\mathbf{C}\in\mathbb{R}^{3n}\times\mathbb{R}^{3n}$ of fluctuations of residues' COM was constructed. This matrix was then diagonalized to yield $3n$ eigenvectors $\mathbf{\Psi}_q\in\mathbb{R}^{3n}$ (PCA modes) and corresponding eigenvalues λ_q (magnitudes of mode fluctuations). From PCA, an essential subspace[81,82] can be defined that consists of the first few modes (denoted as $n_{ED} \ll 3n$) that are dominant in the fluctuations of the system (so-called essential dynamics).

In ED-CG, a CG map is optimized by minimizing the following variational residual,

$$\chi^2 = \frac{1}{3N}\sum_{I=1}^{N}\sum_{i\in I}\sum_{j\geq i\in I}\left\langle (\Delta\mathbf{r}_i^{ED})^2 - 2\Delta\mathbf{r}_i^{ED}\cdot\Delta\mathbf{r}_j^{ED} + (\Delta\mathbf{r}_j^{ED})^2\right\rangle \quad (1)$$

where N is the number of CG sites to be defined and $\langle(\Delta\mathbf{r}_i^{ED})^2\rangle$ is the mean-square fluctuation of residue i (represented by its COM) in the essential subspace. If two residues, i and j, move in a highly correlated fashion, their fluctuation difference would be very small, and χ^2 decreases if they are grouped into the same CG site I. By minimizing the residual, the ribosome is divided into N groups, and the residues in the same group move as a dynamic domain. It has been argued that essential dynamics describes functional collective motion of these domains.[83,84] Therefore, by taking the COM of these dynamic domains as CG sites, this N-site CG model can preserve the dynamic domains and capture their functional essential dynamics. From the essential PCA

modes, $\langle(\Delta \mathbf{r}_i^{\mathrm{ED}})^2\rangle$ can be calculated as

$$\langle(\Delta \mathbf{r}_i^{\mathrm{ED}})^2\rangle = \sum_{x=1}^{3}\sum_{q=1}^{n_{\mathrm{ED}}} \Psi_q^{i_x}\lambda_q\Psi_q^{i_x} \equiv \mathrm{tr}[(\mathbf{c}^{\mathrm{ED}})_{ii}] \quad (2)$$

where $\Psi_q^{i_x}$ is the component corresponding to the x (=1,2,3) coordinate of the COM of residue i in the eigenvector $\mathbf{\Psi}_q$. The term $(\mathbf{c}^{\mathrm{ED}})_{ii}\in\mathbb{R}^3\times\mathbb{R}^3$ refers to the ith superelement of a matrix $\mathbf{C}^{\mathrm{ED}}\in\mathbb{R}^{3n}\times\mathbb{R}^{3n}$, which is the covariance matrix in the essential subspace. tr[...] represents the trace. Therefore, eq 1 can be rewritten in the following form in the case of PCA:

$$\chi^2 = \frac{1}{3N}\sum_{I=1}^{N}\sum_{i\in I}\sum_{j\geq i\in I}(\mathrm{tr}[(\mathbf{c}^{ED})_{ii}] - 2\mathrm{tr}[(\mathbf{c}^{\mathrm{ED}})_{ij}] + \mathrm{tr}[(\mathbf{c}^{\mathrm{ED}})_{jj}]) \quad (3)$$

ED-CG ribosome models were defined as sequence-based models.[64,65] That is to say, CG sites are contiguous along the primary sequence, and the residues in each dynamic domain are contiguous as well. Therefore, one only needs to define boundaries to divide dynamic domains (a boundary is the last residue of a domain). The sequence-based ED-CG model defines a kind of "quasi-molecule" at the CG level, consistent with the underlying primary sequence. It should be noted that the number of CG sites in the sequence-based model needs to be enough to preserve the dynamic domains properly. If a sequence-contiguous domain is too large due to a too coarse model, the residues within the domain may not move collectively, and the residual (eq 3) would be very high. On the other hand, defining a large number of CG sites in the ribosome will make the ED-CG calculations very expensive, although the value of the residual (eq 3) is decreased. From our experience, the sequence-based ED-CG model will be optimal if one CG site represents about 20−30 residues (n/N) on average. At this resolution, the contiguous residues within the same dynamic domain are most likely correlated in their motions, which makes the residual (eq 3) fairly small.

Since the ribosome is a large complex with more than 60 molecules, a "divide and conquer" strategy was used.[65] ED-CG was applied to the whole ribosome first with hundreds of initial random boundary sets. After minimizing the residual (eq 3) of these sets respectively, the average number of CG sites in each molecule was obtained that determined the distribution of CG sites among ribosomal components. A total of 400 and 800 initial boundary sets were tried, respectively, and their CG-site distributions among the molecules were found to be the same. Therefore, in all the ED-CG ribosome models, only 400 initial boundary sets were used. By fixing the number of CG sites in each molecule, positions of these CG sites were variationally optimized within the molecule, separately, to further minimize the residual. This "divide and conquer" method was found to be efficient to find a robust ED-CG model, especially to a large biomolecular complex with many molecular components.[65]

Fluctuation Matching. In fluctuation matching,[67] CG sites are connected by effective harmonic bonds. These spring constants are optimized by minimizing the difference between the bond-length fluctuations calculated from the CG model and those from MD. After obtaining a CG map from the ED-CG method, a CG trajectory was constructed from the trajectory of the COM of amino acids/nucleotides in the ribosome. The average structure of the CG trajectory was computed by aligning all the frames to the reference CG structure and was used to build an ENM. Any two CG sites within a cutoff distance were bond-connected. Bond lengths in the average CG structure were used as the equilibrium values ($\langle R_{IJ}\rangle$). Mean-squared bond-length fluctuations were computed from the CG trajectory $\langle\Delta R^2_{IJ,\mathrm{MD}}\rangle = \langle(R_{IJ} - \langle R_{IJ}\rangle)^2\rangle$, where R_{IJ} is the bond-length between CG sites I and J in a single frame. From the ENM of the ED-CG model, the mean-squared bond-length fluctuations can also be obtained (denoted as $\langle\Delta R^2_{IJ,\mathrm{ENM}}\rangle$), which depend on the effective spring constants k_{IJ} between bonds. Starting from initial spring constants, their optimal values were determined by minimizing the difference between $\langle\Delta R^2_{IJ,\mathrm{ENM}}\rangle$ and $\langle\Delta R^2_{IJ,\mathrm{MD}}\rangle$. Two initial sets of spring constants were tried. In the first set, all the initial spring constants were chosen as 1, whereas in the second set, the initial spring constant between two CG sites was weighted by the inverse square of their distance. It turned out that, after minimization, the final spring constants from the two different initial cases were very similar; they were then used to estimate the effective interactions between ribosomal components at the CG level, thus providing a simplified picture of the ribosome motions.

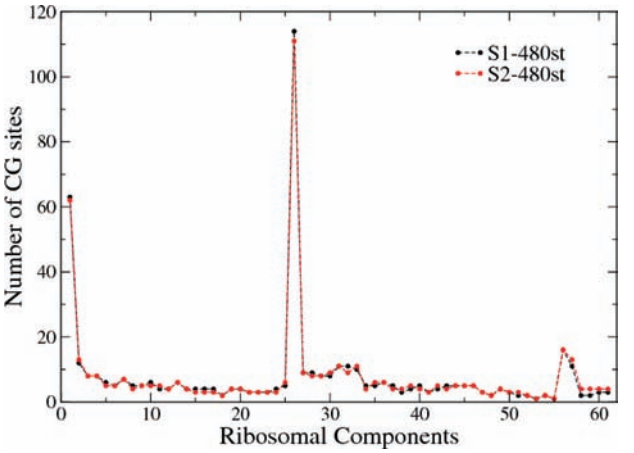


Figure 1. CG-site distribution among ribosomal components in the ED-CG 480-site models, from S1 (black) and S2 (red), respectively.

RESULTS AND DISCUSSION

Comparison of ED-CG Ribosome Models. *ED-CG Models Using Different Segments of the MD Trajectory.* Two 40 ns MD segments were taken from the 100 ns trajectory of the COM of ribosome residues. One was from 20 to 60 ns (denoted as S1), and the other was from 60 to 100 ns (denoted as S2). PCA was performed to S1 and S2, respectively. By defining the number of PCA modes that contribute 95% of the total fluctuation in the ribosome as the essential subspace, n_{ED} in eq 2 equals 420 in S1 and 228 in S2. An ED-CG 480-site ribosome model was built by using the two PCA basis sets, respectively (denoted as S1-480st and S2-480st). With 480 CG sites, the average size of dynamic domains is about 23 residues (10 986/480). As introduced in Theory and Methods, this size should be good for sequence-based ED-CG models. The CG-site distributions among 61 ribosomal components in the two ED-CG 480-site models are shown in Figure 1. Overall, the CG-site distributions between S1-480st (black) and S2-480st (red) look similar. After fluctuation matching, root-mean-square fluctuations (RMSFs) of the CG sites were calculated from the two ED-CG models, respectively (red curves in Figure 2a,b). For comparison, the RMSF values of the CG sites were also computed from S1 and S2, respectively (black curves in Figure 2a,b). Both S1-480st and S2-480st can well reproduce fluctuations from their respective MD trajectories. Some regions with large fluctuations (labeled in Figure 2b) are important in ribosomal functional dynamics, which will be discussed in more detail in the following sections. The major difference between the two MD segments S1 and S2 is that the protein L10 and the L7/L12 tetramer are more flexible in S2 (Figure 2b) than those in S1 (Figure 2a). L10 and the L7/L12

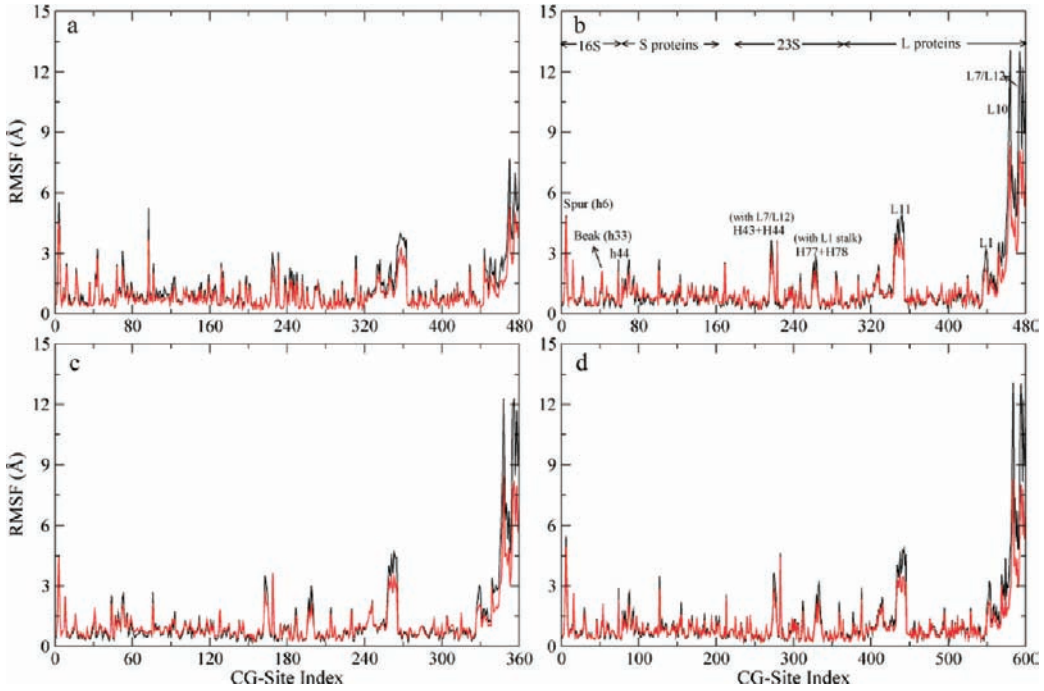


Figure 2. Root-mean-square fluctuations (RMSFs) of CG sites in the ribosome ED-CG models. (a) The S1-480st model. (b) The S2-480st model. 16S rRNA, S proteins, 23S rRNA, and L proteins are indicated. Some regions with large fluctuations, which may be functionally important, are labeled. (c) The S2-360st model. (d) The S2-600st model. The black curves are RMSFs computed from the MD, and the red curves are those from fluctuation matching.

tetramer are the last five components in the simulated ribosome system (Figure 1). More CG sites are allocated in these components in the S2-480st model (red in Figure 1) than in the S1-480st model (black in Figure 1). Since these components fluctuate more in S2 than in S1, the ED-CG method naturally assigns more CG sites to preserve their larger motions in the S2-480st model.

ED-CG Models with Different Resolution. By using the MD segment S2, two additional ED-CG models with 360 and 600 sites were defined (denoted as S2-360st and S2-600st, respectively) and compared with the S2-480st model. The average dynamic domain size in the S2-360st model is a little more than 30 residues (10986/360), and its residual (eq 3) has a value of 475.7. The average domain size in the S2-600st model is less than 20 residues, and the residual is as low as 133.5. The residual of the S2-480st model is 233.2. By comparing the RMSFs of the CG sites after fluctuation matching, the three ED-CG models with different resolution are very similar in describing fluctuations of the functional units in the ribosome (Figure 2b−d). In the following sections, only results from the S2-480st model are shown in the main text, while results from the other ED-CG models (S1-480st, S2-360st, and S2-600st) can be found in the Supporting Information (SI).

Functional Domains Described by the ED-CG Model. In the S2-480st ribosome model (Figure 3), there are 62 sites in 16S rRNA (Figure 4a), 111 sites in 23S rRNA (Figure 4b), 101 sites in the S proteins (Figure 5), three sites in A-site and P-site tRNA, respectively, three sites in mRNA (Figure 6), six sites in 5S rRNA (Figure 7), and 191 sites in the L proteins (Figure 8). Although 480 sites are very coarse compared to the size of the ribosome, it can already describe functional domains in the system quite well.

16S rRNA. 16S rRNA is the major component of the 30S small subunit, which contributes 14% of residues (22% mass) of the whole ribosome. There are four secondary structural domains in 16S rRNA along its sequence (Figure 4a): the 5′ domain, the central domain, the 3′ major domain, and the 3′ minor domain. The 62 sites in 16S rRNA are shown in Figure 4a, colored according to the four domains. The 5′, central, and 3′ major domains are compact domains, and their CG sites (blue, red, and magenta, respectively) are clearly separated. This CG-site distribution can describe the collective motions between these domains properly. Detailed analysis indicates that the CG sites identify many secondary structural motifs (such as helices) in 16S rRNA. For example, the spur region in the 5′ domain (body) corresponds to helix 6 (h6), and there are four CG sites defined in it. The beak region in the central domain (head) is comprised of h33, and four CG sites are located in this helix as well. Both the spur and beak regions, which are also indicated in Figure 2b, are flexible.[56] Therefore, high densities of CG sites are located in these regions in order to describe their dynamics. Unlike the other three domains, the 3′ minor domain is extended (yellow in Figure 4a), containing only two helices (h44 and h45). The two helices are located at the interface between the 30S and 50S subunits, and h44 is believed to act as a dynamic anchor for translocation.[56] There are six CG sites defined in h44 and only one site in h45.

23S rRNA. 23S rRNA is the largest component in the ribosome, containing 26% of the residues (42% mass) of the whole system.[30] The sequence of 23S rRNA can be divided into six secondary structural domains (from I to VI). The 111 CG sites in 23S rRNA are colored according to these sequential domains (Figure 4b). It is clear that the CG sites in the six sequential domains are significantly intertwined with each other, which is quite different from those in 16S rRNA (Figure 4a). The CG-site distribution indicates that the six sequential domains in 23S rRNA do not behave as dynamic domains with collective motions

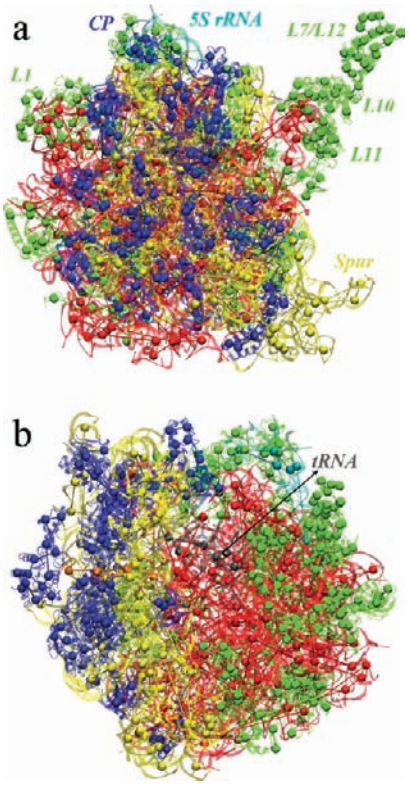


Figure 3. The S2-480st ribosome model. 16S rRNA, yellow; S proteins, blue; tRNAs (A-site and P-site, respectively), gray; mRNA, orange; 5S rRNA, cyan; 23S rRNA, red; L proteins, green. (a) View from the solvent-exposed side of 16S rRNA. The spur region, 5S rRNA, central protuberance (CP), and proteins L1, L10, L11, and L7/L12 are labeled. (b) Side view showing the subunit interface, with the 30S subunit on the left. tRNA molecules are labeled. Figures 3−10 were created using VMD.[94] In these figures, sequence-contiguous CG sites within each ribosomal component are connected by effective bonds.

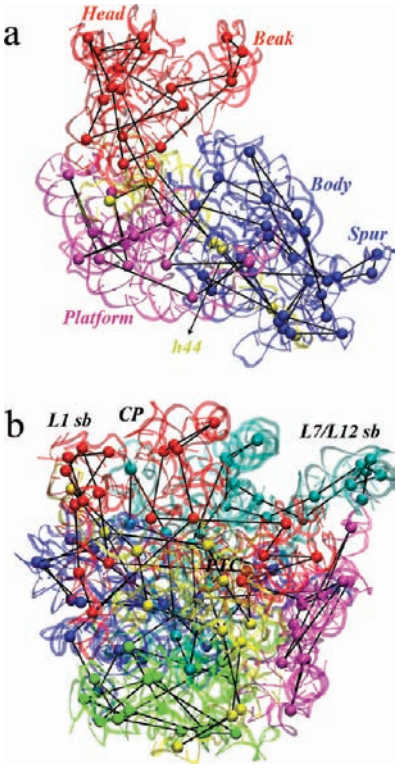


Figure 4. (a) The 62 CG sites in 16S rRNA (view from its solvent-exposed side). There are four secondary structural domains along the sequence: 5′ (blue), central (magenta), 3′ major (red), and 3′ minor (yellow). The spur region in the body, the beak region in the head, the platform, and h44 are labeled. (b) The 111 CG sites in 23S rRNA (interface view). There are six secondary structural domains along the sequence: I (blue), II (cyan), III (green), IV (yellow), V (red), and VI (magenta). Peptidyl transferase center (PTC), CP, L1, and L7/L12 stalk bases are labeled.

between them. Instead, the majority of 23S rRNA acts like a single rigid body.[56] From this massive body, a central protuberance (CP) and two stalks stretch out (Figures 3a and 4b). The CP interacts with 5S rRNA and some L proteins. The two stalks are L1 and L7/L12 stalks, respectively, which are dynamic elements in 23S rRNA and undergo movement in connection with their functional interactions. These flexible regions have higher densities of CG sites than the bulk body region (Figures 3a and 4b).

S Proteins. The present ribosome structure consists of 20 S proteins (S2−S21), which are involved in stabilizing the tertiary structure of 16S rRNA through extensive protein−RNA interactions.[12,16] The ED-CG method has defined CG models for all the S proteins with good convergence, and Figure 5 shows the CG sites of seven S proteins (S2, S4, S5, S12, S13, S15, and S20). The S proteins typically have one or more globular domains, plus extended internal loops or long N- or C-terminal tails.[16] The ED-CG models (Figure 5) can identify these extended regions, which may be functionally important to associate with the RNA inside the ribosome. Inside the globular domains, the CG sites preserve these dynamic domains and maintain the topology.

tRNAs. The ribosome has three binding sites for tRNA molecules: the A (aminoacyl), P (peptidyl), and E (exit) sites. In the simulated system, the A- and P-sites are each occupied by a tRNA. There are three CG sites in the A-tRNA (with sequence numbers 1−25, 26−47, and 48−76, respectively) and also three CG sites in the P-tRNA (with sequence numbers 1−23, 24−47, and 48−76, respectively). The three dynamic domains defined by ED-CG are almost the same between the two tRNA molecules (Figure 6). Especially the anticodon stem loop (ASL, from nucleotides 26−44) is almost exactly identical in both the A- and P-tRNAs (green in Figure 6).

5S rRNA. 5S rRNA is located at the top of the CP. There are only six CG sites in 5S rRNA since it is a small RNA molecule with 117 nucleotides (2−118) in the simulated system (Figure 7). It consists of three stems that project out from the loop A.[85] These stems bind to different regions in 23S rRNA and L proteins.[17,86] The six CG sites preserve the topology of the three stems quite well (Figure 7), which may suggest that collective domain motions exist between them.

L Proteins. There are 33 L proteins in the present 70S ribosome structure. The most striking feature of L proteins is that there are many extensions in the protein structures (Figure 8). It has been said that the L proteins generally contact sites in several domains that are often far apart along the sequence of the large subunit RNAs.[17] The structural extensions in the L proteins may help to stabilize these interdomain interactions that are necessary to maintain the structural integrity of the large subunit. The CG sites successfully defined these

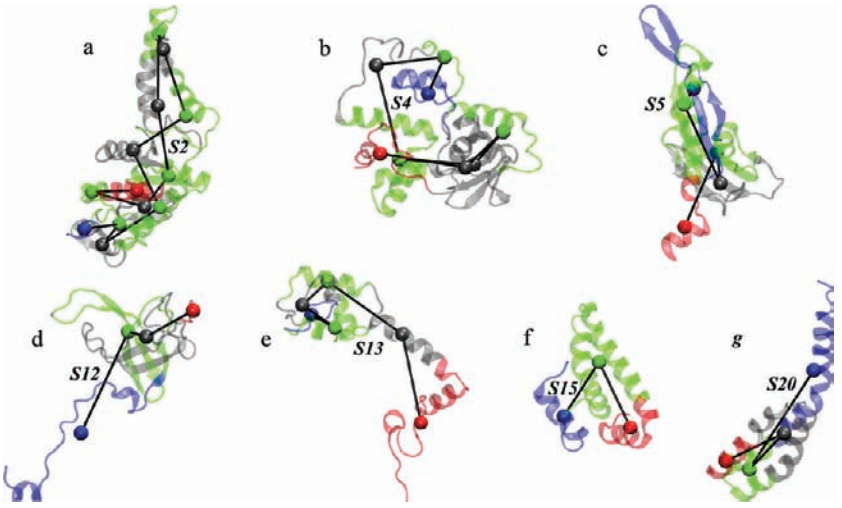


Figure 5. ED-CG models of S proteins: (a) S2 (13 sites), (b) S4 (eight sites), (c) S5 (five sites), (d) S12 (four sites), (e) S13 (six sites), (f) S15 (three sites), and (g) S20 (four sites). For each S protein, the N-terminal site is colored blue and the C-terminal site is colored red. Between them, there are alternating green and gray sites in order to indicate different dynamic domains.

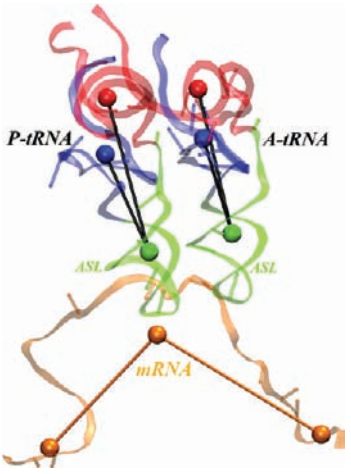


Figure 6. ED-CG models of mRNA and tRNAs (A-site and P-site, respectively). The mRNA is colored orange, and the anticodon stem loop (ASL) in each tRNA is colored green.

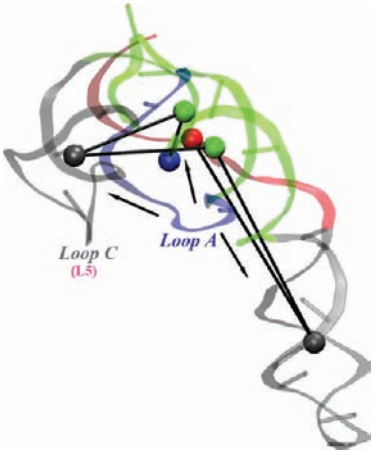


Figure 7. ED-CG six-site model of 5S rRNA. The 5′ site is colored blue, and the 3′ site is colored red. Between them, there are alternating green and gray sites in order to indicate different domains. Loop A is labeled, and three of its stems are indicated by arrows. Loop C that binds to protein L5 is also labeled.

extensions in the L proteins (Figure 8). Some proteins, such as L1, L10, and L7/L12, are highly flexible (Figure 2). Correspondingly higher densities of CG sites are located in these proteins (Figure 8a,d).

Intermolecular Interactions from the ED-CG Model. As introduced in the method of fluctuation matching, a large cutoff distance of 88 Å was used to avoid missing any important interactions over long distances. Thus, the CG sites in the S2-480st ribosome model are connected by 27 567 effective bonds in total, with each CG site bonded with at least other 40 sites. Spring constants of all these bonds were obtained by fluctuation matching. A large spring constant indicates that the bond between two CG sites is stiff, and their effective interaction can be strong if they are close in distance. The majority of the bonds have very soft (near-zero) spring constants, which are treated as no-bond connections. There are only 4883 bonds (about 18% of the total) that have spring constants $\geq 0.05 k_B T/\text{Å}^2$, and such a bond between two CG sites is defined as a CG interaction. Table 1 lists the distribution of CG interactions between ribosomal components in the S2-480st model, and the results from other ED-CG models can be found in Table S1 in the SI. Thus, intermolecular interactions in the ribosome can be investigated at the CG level.

CG Interactions between the Small and Large Subunits. There are 257 CG interactions between the 30S and the 50S subunits in the S2-480st model. The two subunits are associated through a network of inter-subunit bridges, which involve RNA−RNA, RNA−protein, and protein−protein interactions. These inter-subunit bridges have been investigated according to atomic structures of the ribosome.[20,22] Here we look into the bridge interactions at the CG level and also estimate how strong the bridges are on the basis of the number of CG interactions and their spring constants. The distribution of CG interactions between the small and large subunits is 61 CG interactions between 16S and 23S rRNA, 73 between 16S rRNA and L proteins, 60 between S proteins and 23S rRNA, and 62 between S proteins and L proteins (Table 1). Nearly all the inter-subunit

dx.doi.org/10.1021/ja2028487 |J. Am. Chem. Soc. 2011, 133, 16828–16838

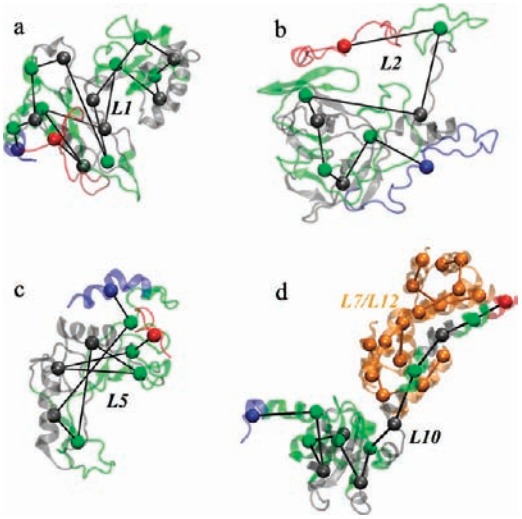


Figure 8. ED-CG models of L proteins: (a) L1 (16 sites), (b) L2 (nine sites), (c) L5 (nine sites), and (d) L10 (13 sites), with the L7/L12 tetramer (16 sites). For L proteins L1, L2, L5, and L10, the N-terminal site is colored blue, and the C-terminal site is colored red. Between them, there are alternating green and gray sites in order to indicate different dynamic domains. The CG sites in the L7/L12 tetramer are all colored orange.

Table 1. CG Interactions between the Ribosomal Components

	16S	S proteins	P-tRNA	mRNA	A-tRNA	5S	23S	L proteins
16S	320[a]	644	13	22	22	0	61	73
S proteins		512	11	31	18	1	60	62
P-tRNA			3	1	7	0	27	24
mRNA				1	1	0	2	0
A-tRNA					2	0	20	9
5S						12	21	76
23S							712	1197
L proteins								918

[a] If two CG sites are connected by an effective bond with a spring constant $\geq 0.05\ k_BT/Å^2$, they have a CG interaction.

bridges are identified by CG interactions (Table 2 for the S2-480st model, and Table S2 in SI for the other three ED-CG models). For example, bridges B1a and B1b are located between the head of the 30S subunit and the top of the 50S subunit. The bridge B1a consists of interactions between S13 and H38 in 23S rRNA (Figure 9), and there are three CG interactions between them (Table 2). The bridge B1b forms between S13 and L5 (Figure 9), which are connected by as many as eight CG interactions (Table 2). Many S and L proteins are located separately at the two opposite solvent-exposed surfaces in the ribosome (Figure 3b, blue and green). The bridge S13-L5 has a significant contribution to the interactions between the S and L proteins, with eight out of the total 62 CG interactions (Table 1). 5S rRNA in the large subunit has few interactions with the small subunit (Table 1, and Table S1 in SI). In the S2-480st model, 5S rRNA has only a single interaction of its loop C with the N-terminal of protein S13 (Figure 9). S13 is greatly involved in the bridges B1a and B1b, and protein L5 (involved in the bridge B1b) also makes close contacts with the loop C in 5S rRNA

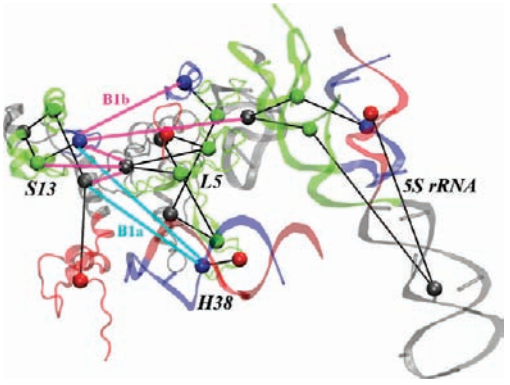


Figure 9. Inter-subunit bridges B1a and B1b between the head of the 30S subunit and the top of the 50S subunit. The bridges are described by interactions between CG sites. Those CG interactions with large spring constants are connected (colored cyan for the bridge B1a and magenta for the bridge B1b). The proteins S13 and L5 are colored as in Figures 5e and 8c; 5S rRNA is colored as in Figure 7. H38 contains two CG sites, which are colored blue (5′) and red (3′), respectively.

(Figure 7).[17,86] Therefore, it seems that 5S rRNA may also play a role in the inter-subunit bridges (Figure 9). The unit h44 is a long helix (also called a penultimate stem), which is the dominant structural component at the interface of the 30S subunit (Figure 4a). Therefore, it involves extensive bridge interactions with the 50S subunit, such as bridges B2a, B3, B5, and B6 (Figure 10). There are many CG interactions with large spring constants in these bridges (Table 2). All of them are formed between h44 and domain IV (located at the center of the 50S body; see regions colored yellow in Figure 4b). The surface of domain IV is structurally complementary to h44,[22] which may explain why they constitute the major inter-subunit bridges with many CG interactions. These CG interactions have generally larger spring constants than those in the head bridges (B1a and B1b, Table 2), which again indicate that the ribosome body is rigid, whereas the head is mobile and can rotate relative to the body.

CG Interactions between Domains within 16S rRNA. Among the 62 CG sites in 16S rRNA, there are 25 sites in the 5′ domain, 12 sites in the central domain, 18 sites in the 3′ major domain, and 7 sites in the 3′ minor domain. According to the distribution of these CG sites (Figure 4a), they can preserve the four domains that move relative to one another during protein synthesis. Therefore, it is interesting to count CG interactions between the four domains inside 16S rRNA (Table 3).

There are 320 CG interactions in total within 16S rRNA (Table 1), and over two-thirds of them (217) are intradomain CG interactions (Table 3). In all the interdomain CG interactions, the 3′ major domain (the head of 16S rRNA, red in Figure 4a) has the minimal number of 25 CG interactions with the other domains, although it contains the second largest number of 18 CG sites. The smallest number of CG interactions between the head and other domains makes it easier to move relative to the other domains, which is consistent with the head motion during the translocation process.[24,87] The 3′ minor domain (consisting of h44 and h45) is the smallest in all the four domains with the minimal number of seven CG sites; however, it has the largest number of 61 extensive CG interactions with the other domains. This result supports the notion that h44 may serve as a dynamic anchor for the motion of the 30S subunit.[56]

dx.doi.org/10.1021/ja2028487 |J. Am. Chem. Soc. 2011, 133, 16828–16838

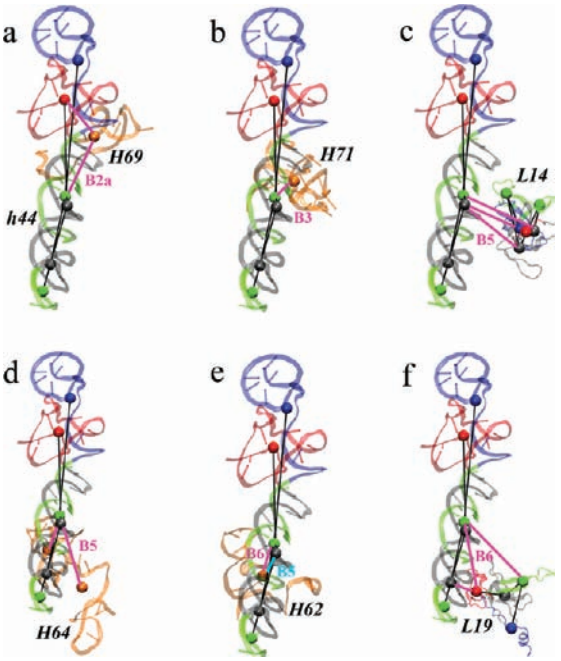


Figure 10. Intersubunit bridges between h44 in the 30S subunit and the 50S subunit. The bridges are described by effective interactions between CG sites. Those CG interactions with large spring constants are connected and are colored magenta, except those in panel e. (a) CG interactions between h44 and H69 that represent the bridge B2a. (b) CG interactions between h44 and H71 that represent the bridge B3. (c) CG interactions between h44 and L14 that belong to the bridge B5. (d) CG interactions between h44 and H64 that belong to the bridge B5. (e) CG interactions between h44 and H62 that belong to both the bridge B5 (cyan) and the bridge B6 (magenta). (f) CG interactions between h44 and L19 that belong to the bridge B6. For h44, the 5′ site is colored blue, and the 3′ site is colored red. Between them there are alternating green and gray sites in order to distinguish different CG sites. The proteins L14 and L19 are colored as in Figure 8, and all the helices in 23S rRNA are colored orange.

CG Interactions between 16S rRNA and S Proteins. There are 644 CG interactions between 16S rRNA and its S proteins (Table 1). According to experimental structures of the 30S subunit,[12,16] S proteins are not evenly distributed around 16S rRNA. Instead, many of them are located on the solvent-exposed side and periphery of the small subunit. There are also a larger number of proteins that bind to the head than to the other regions. Here CG interactions between S proteins and the four domains in 16S rRNA are computed. S proteins have 185 CG interactions with the 5′ domain, 135 with the central domain, 263 with the 3′ major domain, and only 61 with the 3′ minor domain, respectively. Consistent with the structural data, the head (3′ major domain) of 16S rRNA has the strongest CG interactions with S proteins. Since the head is highly mobile relative to the other three domains, more CG interactions with S proteins may help to stabilize and maintain the internal structure of the head domain. The 3′ minor domain has the minimal number of CG interactions with the S proteins because h44 and h45 are located at the interface between the subunits where not many S proteins are surrounded.

CG Interactions between 5S rRNA, 23S rRNA, and L Proteins in the 50S Subunit. It is not surprising that the largest number of CG interactions (1197) is between 23S rRNA and its L proteins since 23S rRNA is the largest ribosomal component that binds with the greatest number of proteins. The 50S subunit can be divided into four domains in space: the body, CP, L1 stalk, and L7/L12 stalk.[56] The body is very large and rigid. The 5S rRNA composes the top of the CP region, which has only one CG interaction with S13 in the small subunit and 97 CG interactions with the other parts in the large subunit (Table 1). The precise function of 5S rRNA has not yet been fully understood. As discussed above, 5S rRNA may involve the inter-subunit bridges B1a and B1b between the 30S head and the 50S CP and regulate the ratchet-like rotation between the subunits. This result supports the suggestion that 5S rRNA may function in the process of translocation.[86] The L1 and L7/L12 stalks move relative to the body with high flexibility. The L1 unit (Figure 8a) has 48 CG interactions with 23S rRNA, which form the base of L1 stalk. It has few CG interactions with other L proteins, which make L1 move relatively easily to control the exit of tRNA from its E-site.[58,88]

The L7/L12 stalk (Figure 8d) consists of a tetrameric form of L7/L12 proteins[89] in the present ribosome structure. It was found to be the most flexible region in the ribosome during the

Table 2. Bridge Interactions between the Small and Large Subunits in the Ribosome

bridge	30S subunit	50S subunit	CG interactions (k_{max})
B1a	S13[a]	H38[b]	3[c] (0.3[d])
B1b	S13	L5[e]	8 (1.0)
B2a	h44[f]	H69	4 (0.5)
B2b	h24	H67, H69	2 (1.4)
	h45	H69, H71	2 (2.2)
B2c	h24	H67	1 (0.8)
	h27	H67	1 (3.6)
B3	h44	H71	1 (13.3)
B4	h20	H34	1 (0.06)
	S15	H34	2 (1.8)
B5	h44	H64	2 (6.0)
	h44	L14	12 (3.7)
	h44	H62	2 (13.3)
B6	h44	H62	2 (13.3)
	h44	L19	8 (4.3)
B7a	h23	H68	0 (0.04)
B7b	h23	L2	5 (0.5)
	h24	L2	3 (1.4)
B8	h14	L14	5 (2.0)

[a] S13 means the S protein 13. [b] H38 means the helix 38 in the large subunit. [c] Number of CG interactions. [d] The largest spring constant (in $k_BT/Å^2$) in CG interactions that belong to a certain bridge. [e] L5 means the L protein 5. [f] h44 means the helix 44 in the small subunit

Table 3. CG Interactions between Domains in 16S rRNA

	5′	central	3′ major	3′ minor
5′	110[a]	28	8	30
central		36	6	20
3′ major			65	11
3′ minor				6

[a] If two CG sites are connected by an effective bond with a spring constant $\geq 0.05\ k_BT/Å^2$, they have a CG interaction.

MD simulation (Figure 2), which is consistent to the single-molecule FRET data.[63] Although each L7/L12 monomer contains only about 30 amino acids, it has a high density of four CG sites. The L7/L12 stalk has no CG interactions with 23S rRNA. It has 56 CG interactions with other L proteins, and 50 of them are with L10. The latter may form the base of the L7/L12 stalk. A long α-helix at the C-terminal of L10 reaches into the L7/L12 tetramer and makes extensive contacts (Figure 8d).

CG Interactions between mRNA and Other Components. mRNA has 53 CG interactions with the small subunit but almost no interactions with the large subunit (Table 1). To start protein synthesis, mRNA is initially recognized by the 30S subunit, and then the 50S subunit associates. That is to say, mRNA primarily binds with the small subunit that is supported by the data of CG interactions.

There are three CG sites in both the A- and P-tRNAs (Figure 6), and the anticodon stem loop (ASL) is exactly identified in both of them (green in Figure 5). The mRNA is short (36 nucleotides) in the simulated ribosome system, which is divided into three CG sites (orange in Figure 5). The mRNA has one CG interaction with the A- and the P-tRNA, respectively (Table 1). Both of them are between mRNA and ASL in the two tRNAs. Therefore, the CG interactions appear to reflect the interactions between anticodon and codon.

CONCLUSIONS

The *E. coli* 70S ribosome contains over 60 RNAs and proteins, and its functional dynamics is achieved through extensive interactions between these components. Atomistic MD simulations of the ribosome are computationally very expensive due to its large size.[37−47] Information about the intermolecular interactions is included in atomic MD data, but it is highly complex. Residue-based ENMs can predict collective dynamics between functional domains in the ribosome.[55−61] However, it is not straightforward and very complex to investigate how strong these interactions are by either the MD simulations or the residue-based ENMs.

This article focuses on the interactions between ribosomal components by mapping the MD fluctuations onto ribosome CG models. The ED-CG method[64−66] is used to build CG models of the ribosome. One of the advantages of the ED-CG models is that the CG sites can preserve the dynamic domains, and the resulting simplified, highly reduced resolution CG models are defined to capture the ribosomal functional dynamics. This idea of systematically identifying dynamic domains by best fitting to the essential PCA modes has also been suggested in other work.[90,91] Effective spring constants between the CG sites, obtained by fluctuation matching[67] to those from the MD simulation, can be used to estimate the strength of the interactions between those dynamic domains properly at the CG level.

Although all interactions between the CG sites are assumed to be effectively harmonic, the results are quite encouraging. Nearly all the bridges between the small and large subunits are identified by one or more CG interactions with large spring constants. Those CG interactions between the four domains in 16S rRNA (5′, central, 3′ major, and 3′ minor domains) support that its head (3′ major domain) is functionally most flexible since it has the smallest number of CG interactions with other domains. However, the head has the largest number of CG interactions with S proteins, which can be used to maintain the tertiary structure of the head. The L proteins at the two stalks (L1 and L7/L12 stalks) are found to be the most mobile components in the complex because they have very few CG interactions with the rest of the ribosome.

Therefore, the ED-CG method, together with fluctuation matching, provides a strategy to plot a simplified reduced resolution CG interaction map between structural components within the ribosome, derived from underlying atomistic MD data. These CG interactions are much simpler than those at the atomic level but are still quite informative because the interactions between dynamic domains are potentially very important to describe the functional dynamics of the whole complex. It should be noted that the ED-CG model and the effective interactions between the CG sites are generally obtained from a MD simulation at the time scale of nanoseconds, but the functional dynamics of the ribosome certainly occurs on much longer time scales. When using the ED-CG method, one should check carefully the extent to which the PCA modes can represent long-time ribosome dynamics. However, it should be emphasized that the essential subspace often exhibits good convergence if it includes enough modes.[92] In this article, the essential subspace consists of the PCA modes that contribute 95% of the total equilibrium fluctuations in the ribosome. The comparison between the S1- and S2-480st models supports the notion that the ED-CG models and CG interactions can effectively describe key functional couplings between the ribosomal components, such as the bridge interactions between the small and large subunits (Table 2, and Table S2 in SI).

Based on the same atomistic MD trajectory (i.e., the same PCA basis set), the CG interactions will largely depend on the different ED-CG models. By comparing three ED-CG models having different resolutions (from 360 to 480 CG sites), they are seen to all have a similar relative CG-site distributions among ribosomal components (Table 1, and Table S1 in SI), although the density of CG interactions is increased. All of the three ED-CG models can properly describe the inter-subunit bridge interactions, according to the number of CG interactions and their spring constants (Table 2, and Table S2 in SI), although individual CG interactions may vary. Since ED-CG models are optimized to capture the functional motions in the ribosome in the underlying MD data (Figure 2), these effective interactions between CG sites (i.e., dynamics domains) can be reasonably interpreted as physical and functional couplings among ribosomal components.

It must be emphasized that one should consider the optimal resolution of sequence-based ED-CG models. In a too coarse model, one CG site may represent a large number of contiguous residues that do not all move in a correlated fashion. This group of residues cannot be regarded as a dynamic domain with ignorable internal motion—in that case the residual would be very high (eq 3). In this case, the CG interactions may not be able to represent functional couplings among the ribosomal components appropriately. In this article, we choose average domain sizes from 20 to 30 residues, which make the residual of the sequence-based models fairly low, and the resulting ED-CG ribosome models thus perform well.

The combination of ED-CG and fluctuation matching can be used to analyze CG interactions between each ribosomal protein and rRNA, which may shed some light on assembly/disassembly of proteins in the 30S and 50S subunits.[16,17] Another subject of future research will be to obtain CG-interaction maps of the ribosome in different functional states in order to identify key changes during the process of protein synthesis. This strategy can generally be applied to any large biomolecular complexes to

determine effective harmonic interactions for CG elastic models, and the results may be compared with and even guide single-molecule experiments. Moreover, such an approach could yield insight into functional changes of large biomolecules upon ligand/substrate binding, which can help to uncover potential sites for experimental mutation studies or drug targets. Conformational transitions of the biomolecule can also be investigated by ENM-like models,[93] which is another area in which the ED-CG method may prove to be useful.

■ ASSOCIATED CONTENT

Supporting Information. Table S1, Table S2, and complete ref 87. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
gavoth@uchicago.edu

Present Addresses
[§]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China

■ ACKNOWLEDGMENT

This work is supported by a Collaborative Research in Chemistry grant from the National Science Foundation (NSF grant CHE-0628257 and CHE-1047323). K.Y.S. is grateful to NIH-GM072686 and LANL LDRD for generous support. We thank Dr. Andrea Grafmüller for helpful comments on the manuscript.

■ REFERENCES

(1) Agrawal, R. K.; Penczek, P.; Grassucci, R. A.; Li, Y. H.; Leith, A.; Nierhaus, K. H.; Frank, J. *Science* **1996**, *271*, 1000–1002.
(2) Agrawal, R. K.; Heagle, A. B.; Penczek, P.; Grassucci, R. A.; Frank, J. *Nat. Struct. Biol.* **1999**, *6*, 643–647.
(3) Frank, J.; Agrawal, R. K. *Nature* **2000**, *406*, 318–322.
(4) Gabashvili, I. S.; Agrawal, R. K.; Spahn, C. M. T.; Grassucci, R. A.; Svergun, D. I.; Frank, J.; Penczek, P. *Cell* **2000**, *100*, 537–549.
(5) Stark, H.; Rodnina, M. V.; Wieden, H. J.; van Heel, M.; Wintermeyer, W. *Cell* **2000**, *100*, 301–309.
(6) Valle, M.; Sengupta, J.; Swami, N. K.; Grassucci, R. A.; Burkhardt, N.; Nierhaus, K. H.; Agrawal, R. K.; Frank, J. *EMBO J.* **2002**, *21*, 3557–3567.
(7) Gao, H. X.; Sengupta, J.; Valle, M.; Korostelev, A.; Eswar, N.; Stagg, S. M.; Van Roey, P.; Agrawal, R. K.; Harvey, S. C.; Sali, A.; Chapman, M. S.; Frank, J. *Cell* **2003**, *113*, 789–801.
(8) Mitra, K.; Frank, J. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 299–317.
(9) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905–920.
(10) Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. *Nature* **2000**, *407*, 340–348.
(11) Schluenzen, F.; Tocilj, A.; Zarivach, R.; Harms, J.; Gluehmann, M.; Janell, D.; Bashan, A.; Bartels, H.; Agmon, I.; Franceschi, F.; Yonath, A. *Cell* **2000**, *102*, 615–623.
(12) Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M.; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T.; Ramakrishnan, V. *Nature* **2000**, *407*, 327–339.
(13) Harms, J.; Schluenzen, F.; Zarivach, R.; Bashan, A.; Gat, S.; Agmon, I.; Bartels, H.; Franceschi, F.; Yonath, A. *Cell* **2001**, *107*, 679–688.
(14) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 920–930.
(15) Ogle, J. M.; Brodersen, D. E.; Clemons, W. M.; Tarry, M. J.; Carter, A. P.; Ramakrishnan, V. *Science* **2001**, *292*, 897–902.
(16) Brodersen, D. E.; Clemons, W. M.; Carter, A. P.; Wimberly, B. T.; Ramakrishnan, V. *J. Mol. Biol.* **2002**, *316*, 725–768.
(17) Klein, D. J.; Moore, P. B.; Steitz, T. A. *J. Mol. Biol.* **2004**, *340*, 141–177.
(18) Schmeing, T. M.; Huang, K. S.; Strobel, S. A.; Steitz, T. A. *Nature* **2005**, *438*, 520–524.
(19) Cate, J. H.; Yusupov, M. M.; Yusupova, G. Z.; Earnest, T. N.; Noller, H. F. *Science* **1999**, *285*, 2095–2104.
(20) Yusupov, M. M.; Yusupova, G. Z.; Baucom, A.; Lieberman, K.; Earnest, T. N.; Cate, J. H. D.; Noller, H. F. *Science* **2001**, *292*, 883–896.
(21) Yusupova, G. Z.; Yusupov, M. M.; Cate, J. H. D.; Noller, H. F. *Cell* **2001**, *106*, 233–241.
(22) Tung, C. S.; Sanbonmatsu, K. Y. *Biophys. J.* **2004**, *87*, 2714–2722.
(23) Petry, S.; Brodersen, D. E.; Murphy, F. V.; Dunham, C. M.; Selmer, M.; Tarry, M. J.; Kelley, A. C.; Ramakrishnan, V. *Cell* **2005**, *123*, 1255–1266.
(24) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. D. *Science* **2005**, *310*, 827–834.
(25) Korostelev, A.; Trakhanov, S.; Laurberg, M.; Noller, H. F. *Cell* **2006**, *126*, 1065–1077.
(26) Selmer, M.; Dunham, C. M.; Murphy, F. V.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. *Science* **2006**, *313*, 1935–1942.
(27) Yusupova, G.; Jenner, L.; Rees, B.; Moras, D.; Yusupov, M. *Nature* **2006**, *444*, 391–394.
(28) Ramakrishnan, V. *Cell* **2002**, *108*, 557–572.
(29) Korostelev, A.; Noller, H. F. *Trends Biochem. Sci.* **2007**, *32*, 434–441.
(30) Steitz, T. A. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 242–253.
(31) Schmeing, T. M.; Ramakrishnan, V. *Nature* **2009**, *461*, 1234–1242.
(32) Yonath, A. *J. R. Soc. Interface* **2009**, *6*, S575–S585.
(33) Dunkle, J. A.; Wang, L. Y.; Feldman, M. B.; Pulk, A.; Chen, V. B.; Kapral, G. J.; Noeske, J.; Richardson, J. S.; Blanchard, S. C.; Cate, J. H. D. *Science* **2011**, *332*, 981–984.
(34) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
(35) Adcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, *106*, 1589–1615.
(36) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120–127.
(37) Sanbonmatsu, K. Y.; Joseph, S. *J. Mol. Biol.* **2003**, *328*, 33–47.
(38) Sanbonmatsu, K. Y.; Joseph, S.; Tung, C. S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15854–15859.
(39) Sanbonmatsu, K. Y. *Biochimie* **2006**, *88*, 1053–1059.
(40) Sanbonmatsu, K. Y.; Tung, C. S. *J. Struct. Biol.* **2007**, *157*, 470–480.
(41) Ishida, H.; Haywardy, S. *Biophys. J.* **2008**, *95*, 5962–5973.
(42) Petrone, P. M.; Snow, C. D.; Lucent, D.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16549–16554.
(43) Trobro, S.; Aqvist, J. *Biochemistry* **2008**, *47*, 4898–4906.
(44) Gumbart, J.; Trabuco, L. G.; Schreiner, E.; Villa, E.; Schulten, K. *Structure* **2009**, *17*, 1453–1464.
(45) Vaiana, A. C.; Sanbonmatsu, K. Y. *J. Mol. Biol.* **2009**, *386*, 648–661.
(46) Villa, E.; Sengupta, J.; Trabuco, L. G.; LeBarron, J.; Baxter, W. T.; Shaikh, T. R.; Grassucci, R. A.; Nissen, P.; Ehrenberg, M.; Schulten, K.; Frank, J. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 1063–1068.
(47) Whitford, P. C.; Geggier, P.; Altman, R. B.; Blanchard, S. C.; Onuchic, J. N.; Sanbonmatsu, K. Y. *RNA* **2010**, *16*, 1196–1204.

(48) Bahar, I.; Rader, A. J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586–592.
(49) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150.
(50) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192–198.
(51) Sherwood, P.; Brooks, B. R.; Sansom, M. S. P. *Curr. Opin. Struct. Biol.* **2008**, *18*, 630–640.
(52) Murtola, T.; Bunker, A.; Vattulainen, I.; Deserno, M.; Karttunen, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1869–1892.
(53) Voth, G. A., Ed. *Coarse-graining of condensed phase and biomolecular systems*; CRC Press: New York, 2009.
(54) Trylska, J.; Tozzini, V.; McCammon, J. A. *Biophys. J.* **2005**, *89*, 1455–1463.
(55) Tama, F.; Valle, M.; Frank, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9319–9323.
(56) Wang, Y. M.; Rader, A. J.; Bahar, I.; Jernigan, R. L. *J. Struct. Biol.* **2004**, *147*, 302–314.
(57) Wang, Y. M.; Jernigan, R. L. *Biophys. J.* **2005**, *89*, 3399–3409.
(58) Kurkcuoglu, O.; Doruker, P.; Sen, T. Z.; Kloczkowski, A.; Jernigan, R. L. *Phys. Biol.* **2008**, *5*, 14.
(59) Yan, A. M.; Wang, Y. M.; Kloczkowski, A.; Jernigan, R. L. *J. Chem. Theory Comput.* **2008**, *4*, 1757–1767.
(60) Kurkcuoglu, O.; Kurkcuoglu, Z.; Doruker, P.; Jernigan, R. L. *Proteins: Struct., Funct. Bioinf.* **2009**, *75*, 837–845.
(61) Kurkcuoglu, O.; Turgut, O. T.; Cansu, S.; Jernigan, R. L.; Doruker, P. *Biophys. J.* **2009**, *97*, 1178–1187.
(62) Ermolenko, D. N.; Majumdar, Z. K.; Hickerson, R. P.; Spiegel, P. C.; Clegg, R. M.; Noller, H. F. *J. Mol. Biol.* **2007**, *370*, 530–540.
(63) Aitken, C. E.; Petrov, A.; Puglisi, J. D. *Annu. Rev. Biophys.* **2010**, *39*, 491–513.
(64) Zhang, Z.; Lu, L.; Noid, W. G.; Krishna, V.; Pfaendtner, J.; Voth, G. A. *Biophys. J.* **2008**, *95*, 5073–5083.
(65) Zhang, Z. Y.; Pfaendtner, J.; Grafmüller, A.; Voth, G. A. *Biophys. J.* **2009**, *97*, 2327–2337.
(66) Zhang, Z. Y.; Voth, G. A. *J. Chem. Theory Comput.* **2010**, *6*, 2990–3002.
(67) Lyman, E.; Pfaendtner, J.; Voth, G. A. *Biophys. J.* **2008**, *95* 4183–4192.
(68) Berk, V.; Zhang, W.; Pai, R. D.; Cate, J. H. D. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15830–15834.
(69) Nikulin, A.; Eliseikina, I.; Tishchenko, S.; Nevskaya, N.; Davydova, N.; Platonova, O.; Piendl, W.; Selmer, M.; Liljas, A.; Drygin, D.; Zimmermann, R.; Garber, M.; Nikonov, S. *Nat. Struct. Biol.* **2003**, *10*, 104–108.
(70) Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
(71) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
(72) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
(73) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
(74) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472–2493.
(75) Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.
(76) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
(77) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
(78) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
(79) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
(80) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
(81) Garcia, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696–2699.
(82) Amadei, A.; Linnsen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct. Genet.* **1993**, *17*, 412–425.
(83) Kitao, A.; Go, N. *Curr. Opin. Struct. Biol.* **1999**, *9*, 164–169.
(84) Berendsen, H. J. C.; Hayward, S. *Curr. Opin. Struct. Biol.* **2000**, *10*, 165–169.
(85) Barciszewska, M. Z.; Szymanski, M.; Erdmann, V. A.; Barciszewski, J. *Acta Biochim. Pol.* **2001**, *48*, 191–198.
(86) Szymanski, M.; Barciszewska, M. Z.; Erdmann, V. A.; Barciszewski, J. *Biochem. J.* **2003**, *371*, 641–651.
(87) Ratje, A. H.; et al. *Nature* **2010**, *468*, 713–U143.
(88) Munro, J. B.; Altman, R. B.; Tung, C. S.; Cate, J. H. D.; Sanbonmatsu, K. Y.; Blanchard, S. C. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 709–714.
(89) Georgalis, Y.; Dijk, J.; Labischinski, H.; Wills, P. R. *J. Biol. Chem.* **1989**, *264*, 9210–9214.
(90) Potestio, R.; Pontiggia, F.; Micheletti, C. *Biophys. J.* **2009**, *96*, 4993–5002.
(91) Aleksiev, T.; Potestio, R.; Pontiggia, F.; Cozzini, S.; Micheletti, C. *Bioinformatics* **2009**, *25*, 2743–2744.
(92) Amadei, A.; Ceruso, M. A.; Di Nola, A. *Proteins: Struct., Funct. Genet.* **1999**, *36*, 419–424.
(93) Chu, J. W.; Voth, G. A. *Biophys. J.* **2007**, *93*, 3860–3871.
(94) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.