Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 21, Issue Number 3, (2003)

A Comparison Between Elastic Network Interpolation and MD Simulation of 16S Ribosomal RNA

http://www.jbsdonline.com

Moon K. Kim[1]
Wen Li[2]
Bruce A. Shapiro[2]
Gregory S. Chirikjian[1,*]

[1]Department of Mechanical Engineering
The Johns Hopkins University
Baltimore, MD 21218, USA
[2]Laboratory of Experimental and Computational Biology
NCI Center for Cancer Research
National Cancer Institute Frederick
National Institute of Health
Building 469
Room 150
Frederick, MD 21702, USA

Abstract

In this paper a coarse-grained method called elastic network interpolation (ENI) is used to generate feasible transition pathways between two given conformations of the core central domain of 16S Ribosomal RNA (16S rRNA). The two given conformations are the extremes generated by a molecular dynamics (MD) simulation, which differ from each other by 10Å in root-mean-square deviation (RMSD). It takes only several hours to build an ENI pathway on a 1.5GHz Pentium with 512 MB memory, while the MD takes several weeks on high-performance multi-processor servers such as the SGI ORIGIN 2000/2100. It is shown that multiple ENI pathways capture the essential anharmonic motions of millions of timesteps in a particular MD simulation. A coarse-grained normal mode analysis (NMA) is performed on each intermediate ENI conformation, and the lowest 1% of the normal modes (representing about 40 degrees of freedom (DOF)) are used to parameterize fluctuations. This combined ENI/NMA method captures all intermediate conformations in the MD run with 1.5Å RMSD on average. In addition, if we restrict attention to the time interval of the MD run between the two extreme conformations, the RMSD between the closest ENI/NMA pathway and the MD results is about 1Å. These results may serve as a paradigm for reduced-DOF dynamic simulations of large biological macromolecules as well as a method for the reduced-parameter interpretation of massive amounts of MD data.

Key words: 16S ribosomal RNA, Elastic network interpolation, Intermediate conformation, Molecular dynamics, Normal mode analysis.

Introduction

Many macromolecular structures have been solved and posted in the Protein Data Bank (1). There we can often find multiple conformations of certain macromolecular structures. Some have "extended" and "compact" forms. The structure of a macromolecule is thought to be strongly related to its biological function. Such functions can include catalysis, regulation, transport, and binding of ligands (2). The study of conformational transitions among multiple forms is therefore important for understanding the relationship between structure and function. Namely, some motions are necessary for particular functions. Hence, comprehending conformational transitions can be useful for understanding biological mechanisms. This problem of elucidating transition pathways can be viewed as a more limited problem than the folding problem.

MD simulation is used for the prediction of conformational transitions. In MD simulation, atomic trajectories are calculated by the classical (Newtonian) equations of motion using structural data (such as that obtained from X-ray crystallography or NMR) as the input. MD simulation can provide realistic molecular motions including the effects of surrounding solvent and ions. However, the computational cost is very high and it is very difficult to obtain long-time-scale collective motions from even millions of MD timesteps because MD results resemble Brownian motion in which a time evolving conformation fluctuates rapidly. Alternatively, NMA is used

[*]Phone: 410-516-7127
Fax: 410-516-7254
Email: gregc@jhu.edu

to analyze global behaviors of macromolecules around a low energy equilibrium conformation (3, 4). In NMA, the molecular potential energy is approximated as a harmonic function using all atom empirical potentials. A generalized eigenvalue problem then results in eigenvalues and eigenvectors which are related to the frequencies and directions of corresponding motions, respectively. All atom NMA is much more computationally efficient than MD simulation but still has the limitation of data storage requirements in the case of large macromolecules. In addition, NMA is not able to predict large anharmonic motions and pathways.

As discussed above, MD simulation and NMA using all-atom empirical potentials are commonly used to follow the dynamics of macromolecules. However, the use of atomic approaches becomes computationally inefficient as the system size increases. To reduce this computational burden, many authors have demonstrated the utility of coarse-graining elastic network models by including, for example, only C_α atoms in a protein structure representing residues and using a simplified harmonic potential for considering internal interactions between neighboring residues. Such models are suitable to describe the global motions of large macromolecules (5-7). Recently, NMA associated with vector quantization has been applied for low-resolution structural data measured by cryogenic electron microscopy (cryo-EM) to elucidate large conformational changes (8-9).

On the other hand, it is also popular to generate pathways between the two end conformations and visualize those conformational transitions. Three of the most common methods are Cartesian interpolation, internal variable interpolation, and ENI. The Cartesian interpolation approach interpolates the atomic coordinates of the two end conformations linearly in Cartesian space (10). The disadvantages of this method are: (i) internal variables of solved intermediate conformations can be unrealistic; (ii) it can allow parts of macromolecules to pass through one another; (iii) the resulting transition pathway depends on the orientations of the two end conformations (i.e., it is not invariant under rigid-body displacements of the inputs).

An alternative interpolation approach is to use internal coordinates such as bond lengths, bond angles, and torsion angles instead (11, 12). If all internal variables were interpolated simultaneously, this would produce realistic bond lengths and torsion angles. However, as with Cartesian interpolation, some parts of the molecule could come unrealistically close to other parts in order to achieve a smooth simulated pathway in the process of generating intermediate conformations. This would produce highly unfavorable states in the sense of high-energy interactions or steric clashes. In ENI we do not interpolate Cartesian or internal coordinates, but rather the two sets of distances between spatially close atoms, which are modeled as being connected with linear springs (13, 14). Since we interpolate relative distance between spatially close things, unrealistic conformations and steric clashes become less likely.

In this paper, we generate a feasible pathway for the conformational transition of the core central domain of the 16S rRNA using the simplest potential and coarse-grained ENI. Our intermediates are compared with 5000 MD samples. We can approximate, to within a small error threshold, all MD conformations as fluctuations around our pathway using normal modes at each intermediate. That means the ENI method can be used to smooth out MD results and generate average MD pathways in a computationally efficient way. Elastic network interpolation can also be used to help interpret the vast amounts of data generated from MD simulations by identifying reaction coordinates.

Methods

NMA Using a Coarse-grained Elastic Network Model

In this section we derive a discrete mechanical model of the small conformational changes in the core central domain of the 16S rRNA around an equilibrium. In gen-

eral, coarse-grained models of proteins are built by including only C_α atoms as point masses. Each C_α atom represents an amino acid residue. Here we define a coarse-grained model for RNA chains somewhat differently. Figure 1 shows the assembly of nucleotides and the atomic numbering that is indicated in each nucleotide unit. P, (O3´-O5´), (C1´-C5´), C2, C4, C5, C6, C8, N1, N3, N7, and N9 (shown as bold characters and lines in Figure 1) are chosen as point masses for a coarse-grained representation. The structure of the core central domain of the 16S rRNA has been obtained at 2.6Å resolution from the PDB entry "1G1X" (15-17). Figure 2a shows the secondary structure and sequence of the 84 nucleotides which appear there. This coarse-graining method reduces the system size by about 50% so that we save substantial computational time. We label the mass of the i^{th} atom as m_i, and model the interaction between atoms i and j with a linear spring having stiffness $k_{i,j}$. Given the full set of masses, stiffnesses and equilibrium positions, we derive the global mass matrix and the global stiffness matrix.

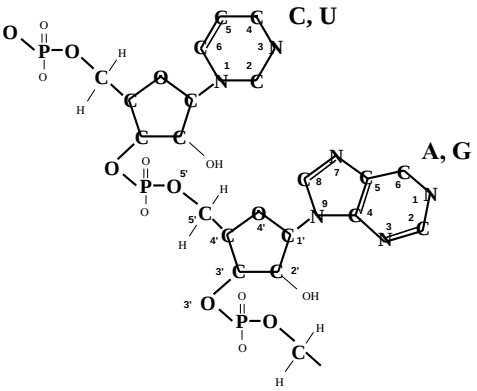


Figure 1: Fragment of an RNA chain with a pyrimidine (C or U), a purine (A or G), and a phosphodiester linkage. Atom numbering is indicated in each nucleotide unit. No distinction is made between A and G and between C and U in this coarse-grained model of RNA. P, (O3´-O5´), (C1´-C5´), C2, C4, C5, C6, C8, N1, N3, N7, and N9 (shown as bold characters and lines) are chosen as point masses for the coarse-grained representation. The other atoms are not included.

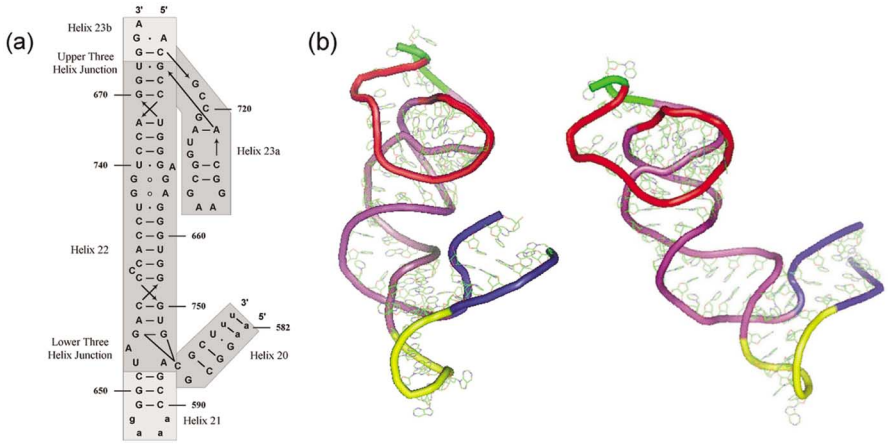


Figure 2: The secondary and the tertiary structures of the 16S rRNA. (**a**) The sequence of the core central domain of the 16S rRNA is shown. It is composed of 5 helical sections connected by two three-helix junctions at the ends of helix 22. Helix 20 makes an acute angle with helix 22 in the lower junction. The main conformational change appears here. The whole swing of helix 22 relative to helix 20 is about 10Å. In the upper junction, helix 23a lies parallel to helix 22 and part of helix 23b stacks on the top of helix 22. This figure is adopted from Agalarov *et al*., 2000. (**b**) The core central domain of the 16S rRNA is displayed with a tube representing its backbone with heavy atoms of the nucleotides. Helices 20, 21, 22, 23a, and 23b are blue, yellow, purple, red, and green, respectively. The extended form (right) is chosen from the MD results initiated from the compact form (left). This 3D figure is generated using PyMOL (26).

The position of the i^{th} atom at time t is denoted

$$\boldsymbol{x}_i(t) = [x_i(t), y_i(t), z_i(t)]^T \in R^3. \qquad [1]$$

The total kinetic energy in a network of N point masses can be defined as

$$T = \frac{1}{2}\sum_{i=1}^{N} m_i \,\|\dot{\boldsymbol{x}}_i(t)\|^2 = \frac{1}{2}\,\dot{\boldsymbol{\delta}}^T M\,\dot{\boldsymbol{\delta}}\,, \qquad [2]$$

where $\delta_{\,i}(t)$ is the small displacement vector of the i^{th} point mass,

$$\boldsymbol{x}_i(t) = \boldsymbol{x}_i(0) + \delta_{\,i}(t), \qquad [3]$$

$$\delta = [\delta_{\,1}^{\,T}, \ldots, \delta_{\,N}^{\,T}]^T \in R^{3N}, \qquad [4]$$

and the matrix M is the global mass matrix (14). The total potential energy has the form

$$V = \frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} k_{i,j}\{\|\boldsymbol{x}_i(t) - \boldsymbol{x}_j(t)\| - \|\boldsymbol{x}_i(0) - \boldsymbol{x}_j(0)\|\}^2\,, \qquad [5]$$

where $k_{i,j}$ is the (i, j) element of the "linking matrix" or "contact matrix", which is assumed to have a non-zero spring constant for all contacting pairs and zero for

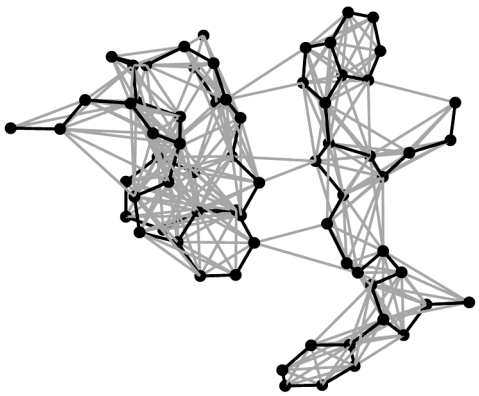


Figure 3: Representation of an RNA structure as an elastic network. For example, 4 nucleotides of 16S rRNA are shown as collections of black dots and lines. The spring connections between atoms within a cutoff distance of 4Å are indicated as black and grey lines. This network can be made either denser or sparser depending on the cutoff rule applied.

pairs not in contact spatially, regardless of the atom types concerned. Assuming that a base ring is structurally rigid, we can impose a larger stiffness value on the internal connections within each base in our elastic model. In the present case, we set the number 10 for an interaction within a base ring, while the other interactions have 1 as a spring constant (Figure 3). Only the ratios of stiffnesses are important in this formulation, and hence units are unimportant.

The springs represent interactions between close atoms in identical ways and a harmonic potential energy function defined by the elastic network model is appropriate to describe small deviations from equilibrium. In general, Eq. [5] is a nonlinear potential function even though the springs themselves are linear. However, when we assume that the deformations are small, V can be approximated as a classical quadratic function by using the Taylor expansion

$$V \approx V_0 + \frac{1}{2}\,\boldsymbol{\delta}^T K \boldsymbol{\delta}, \qquad [6]$$

where V_0 can be ignored without loss of generality, and the matrix K is the stiffness (Hessian) matrix for the whole network (14). Finally we get the equations of motion that describe harmonic motions of the 16S rRNA as

$$M\ddot{\delta} + K\delta = 0. \qquad [7]$$

Normal modes generated using this coarse-grained model can be used economically to predict feasible collective motions about an equilibrium conformation of the 16S rRNA.

Elastic Network Interpolation

The key idea of ENI is to interpolate two sets of distances between spatially close atoms (which are thought of as being connected by springs). One can generate intermediate conformations of the 16S rRNA by finding small changes in the positions of atoms induced by small changes in the distances between atoms.

Suppose that we have two extreme ("extended" and "compact") conformations of the 16S rRNA generated from MD simulation (Figure 2b). In the compact form, the 16S rRNA is bound to the ribosomal protein S15. Helix 22 and helix 21 stack coaxially and helix 20 lies close to helix 22 with an acute angle (18). In contrast, the absence of S15 induces extended conformations of the 16S rRNA in which helix 22 swings away from helix 20 with 10Å RMSD. Experimentally the interhelical angle between helix 22 and helix 20 is observed as ~120° (19) and the computer simulations obtain the value of 114° (17). The sets of Cartesian coordinates describing representative atoms in those two conformations are denoted as $\{\boldsymbol{x}_i\}$ and $\{\chi_i\}$, respectively. We introduce a cost function as follows

$$C(\boldsymbol{\delta}) = \frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} k_{i,j}\{\| \boldsymbol{x}_i + \boldsymbol{\delta}_i - \boldsymbol{x}_j - \boldsymbol{\delta}_j \| - l_{i,j}\}^2. \qquad [8]$$

Here δ is a $3N$ -dimensional vector of displacements with N being the number of atoms in each set. An intermediate conformation is defined by the value of δ that minimizes this cost when all other parameters are held constant. Here the linking matrix is formed as the "union" of the two linking matrices for $\{\boldsymbol{x}_i\}$ and $\{\chi_i\}$. The value $l_{i,j}$ is the desired distance between i and j, which can be chosen as

$$l_{i,j} = (1 - \alpha)\,\| \boldsymbol{x}_i - \boldsymbol{x}_j \| + \alpha\,\| \chi_i - \chi_j \|, \qquad [9]$$

where α is the coefficient specifying how far a given state is along the transition from $\{\boldsymbol{x}_i\}$ to $\{\chi_i\}$. For example, when $\alpha = 1.5$, the desired conformation would be expected to have the average value of inter-residue distances for conformations $\{\boldsymbol{x}_i\}$ and $\{\chi_i\}$.

We can find values of δ that minimize Eq. [8], which itself can be approximated for small values of $\| \delta_i \|$ and $\| \delta_j \|$ with the Taylor series approximation

$$C(\boldsymbol{\delta}) \approx \frac{1}{2}\,\boldsymbol{\delta}^T \Gamma \boldsymbol{\delta} + \frac{1}{2}\,\boldsymbol{\gamma}\boldsymbol{\delta} + B, \qquad [10]$$

where Γ is a $3N \times 3N$ matrix, γ is a $3N$ -dimensional row vector, and B is a constant (14).

We minimize $C(\delta)$ with respect to δ, which results in the following constraint equation:

$$\frac{\partial C(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} = \Gamma \boldsymbol{\delta} + \frac{1}{2}\,\boldsymbol{\gamma}^T = 0. \qquad [11]$$

The matrix Γ always has three zero eigenvalues corresponding to translation modes because a translated version of γ can also minimize the cost function. That is, the solution to Eq. [11] is not unique. One way we propose in this paper to address this issue is to add a weighted Cartesian interpolation to the cost function (Eq. [8]) in order to anchor structures in space such that

$$C'(\boldsymbol{\delta}) = C(\boldsymbol{\delta}) + \varepsilon \sum_{i=1}^{N} \| \boldsymbol{x}_i + \boldsymbol{\delta}_i - ((1-\alpha)\boldsymbol{x}_i + \alpha\,\boldsymbol{\chi}_i) \|^2 . \qquad [12]$$

This result also depends on the position and orientation of the initially given two end conformations. Therefore, we superimpose $\{\chi_i\}$ upon $\{\boldsymbol{x}_i\}$ before simulation. Here we choose $\varepsilon = 0.1$. The new Γ matrix built by Eq. [12] is no longer singular so that a unique δ can solve Eq. [11].

In our implementation, we calculate δ to be the solution of Eq. [12] when $\alpha = 0.01$. Then we obtain the first intermediate conformation by setting

$$\boldsymbol{x}_i \rightarrow \boldsymbol{x}_i + \delta_i. \qquad [13]$$

The remaining intermediate conformations are then obtained in an iterative way.

MD Simulations

MD simulations are utilized to investigate the dynamic behavior of the 16S rRNA in the absence of S15. Both the Particle-Mesh-Ewald (PME) method and the Generalized Born (GB/SA) model were used for computing electrostatic solvation energies (20, 21). The initial coordinates of the 16S rRNA were taken from the PDB entry "1G1X" (15-17). In this fragment, the native sequence of helix 20 and helix 21 were truncated and helix 21 was capped by adding a GAAA tetraloop (Figure 2a). The sequence numbering followed that of *Escherichia coli* (22). Hydrogen atoms were added to the crystal structure using the AMBER6/xleap editing program. The simulations were performed on SGI Origin 2100 and 2000 computers using the Cornell force field (23) and the Sander module in AMBER6 (21, 24). The production runs lasted for 5ns and the time step for all simulations using the GB/SA model was 1fs, while it was 2fs when using the PME method (17). 5000 conformations were extracted from the GB/SA data at equal time intervals for the comparison presented in the next section.

Simulation Results

Conformational Change of 16S rRNA

From the MD simulation for the core central domain of the 16S rRNA, we obtained 5000 sampled conformations. Figure 4 shows the RMSD between each of them and the initial conformation. We choose the interval between the two points indicated in Figure 4 as the end conformations for a transition pathway of the 16S rRNA, because conformation MD4698 is almost the same as the initial conformation and the swing motion of helix 22 appears clearly up to conformation MD4763. Such a motion may be represented as a smooth collective motion with superimposed fluctuations. Using these two end conformations, the ENI method generates smoothly evolving intermediate conformations.

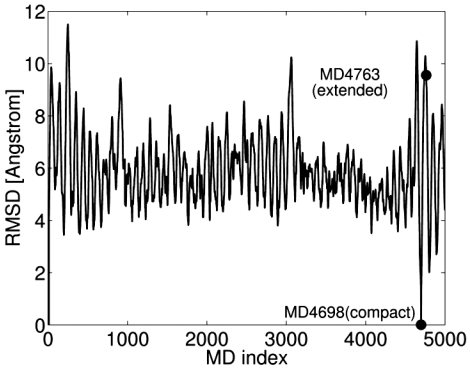


Figure 4: RMSD of 5000 MD results with respect to the initial conformation, which is indistinguishable from conformation MD4698. The fluctuation in RMSD resembles Brownian motion of the 16S rRNA. Hence, it is very hard to recognize collective motions from these time-involving MD results. Two extreme conformations are selected as the inputs of the ENI method proposed in this paper.

It is well known that the characteristics of an elastic network model depend on how the connectivity of the system is defined. There are two common ways. One is to restrain a cutoff distance from one atom to its neighbors. This method reflects the local packing density of a system well (7). However, the choice of this cutoff value affects the behavior of the model. A large cutoff makes a system very stiff with a correspondingly large computational complexity because there are many connections. On the other hand, a small cutoff has a computational advantage but simplifies a system too much to capture the collective motions in either NMA or network interpolation. Alternatively, we set the cutoff to be the number of neighbors assumed to be connected, regardless of the distance between atoms (13). This is suitable for generating a smooth pathway with a relatively small computational load when using ENI. In this context, a number cutoff of 20 is used for both end conformations. That is, we connect one atom to its neighbors by increasing the cutoff distance until 20 contacts are achieved. This ensures that all atoms are well connected. The linking matrix for the conformational transition is taken as the union of the linking matrices of each end conformation (i.e., a link is defined between any two atoms which are in contact in either end conformation).

Figure 5 shows the conformational change of the core central domain of the 16S rRNA. The left conformation is the initial compact form while the right one is the final extended form. In the middle, two intermediates are shown which are picked from 99 intermediate conformations generated by ENI. Helix 22 swings relative to helix 20. The RMSD is about 10Å during the transition. 3D movies showing this conformational change are posted on the web (25). One can compare the pathway generated by the network interpolation method with MD simulation results by observing those movies. Our interpolation gives a smooth pathway parameterized by RMSD while the MD simulation generates a pathway resembling Brownian motion. Each method has its own benefits; MD models the detailed mechanics of the conformational transitions, whereas ENI captures the essence of the anharmonic motions.

Figure 5: A diagram of the conformational transition of the core central domain of the 16S rRNA. A feasible and smooth pathway for the conformational transition of the 16S rRNA is obtained incrementally by using ENI. Two intermediate conformations are illustrated with the two end conformations (from left to right). The main changes appear around helix 22 which swings away from helix 20 associated with the RMSD of about 10Å. The colors of helices here are the same as those of Figure 2b.

Comparison Between MD Simulation and ENI

In this section we compare the 5000 MD simulation results with the 101 conformations (i.e. 2 end conformations + 99 intermediates) generated by ENI. First, we bin all the 5000 conformations according to how close each one is to the nearest of our intermediates in the sense of RMSD. Figure 6 shows the fluctuating motion of the MD results. This implies that MD conformations essentially move back and forth along the 1 DOF pathway generated by network interpolation. From this fact, we hypothesize that the massive set of MD conformations fall approximately along this pathway and the RMSD of each MD conformation from this pathway can be thought of as a fluctuation. The histogram in Figure 7a shows the population of MD conformations over elastic network intermediates. The RMSD of all the MD members in each bin is calculated with respect to the elastic network conformation for that bin and used to separate the histogram into three sections.

Next, we take each of our intermediates and do a normal mode analysis. Let $\{\boldsymbol{x}_i\}$ be the coordinates of the elastic network intermediate that defines a particular bin and $\{\boldsymbol{y}_i\}$ be the coordinates of an MD conformation in that bin after optimal rigid-

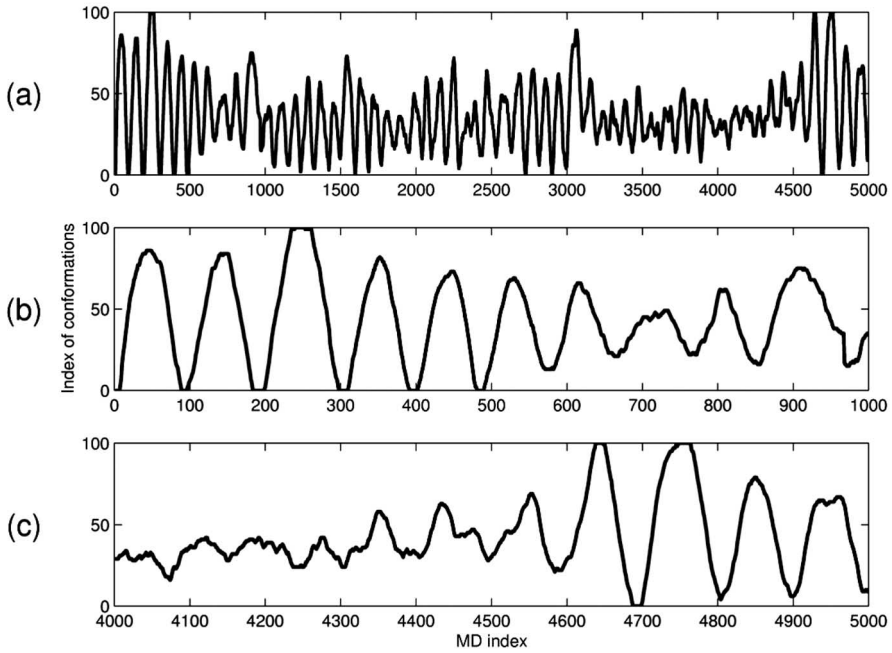


Figure 6: Fluctuating motion of MD results along the 1 DOF pathway generated from ENI. (**a**) All the 5000 MD conformations are binned according to how close one is to the nearest of the 101 intermediates predicted by network interpolation. On the vertical axis, index 0 indicates the initial compact conformation and index 100 indicates the final extended conformation. MD conformations appear to move back and forth quasi-periodically along the 1 DOF pathway generated by network interpolation in (**a**). The first and the last 1000 MD conformations are viewed at a finer scale in (**b**) and (**c**), respectively.

body superposition on $\{\boldsymbol{x}_i\}$. If we compute normal modes for $\{\boldsymbol{x}_i\}$ and truncate at a number m, which we take as 1% of the total number of degrees of freedom, then we can define a new conformation as

$$\boldsymbol{x}_i' = \boldsymbol{x}_i + \sum_{j=1}^{m} c_j \boldsymbol{v}_i^j, \qquad [14]$$

where $\boldsymbol{v}_i^j$ is the displacement vector of the i^{th} atom in the j^{th} mode. For each $\{\boldsymbol{y}_i\}$, $\{c_j\}$ can be determined to minimize the cost function

$$f(\boldsymbol{c}) = \sum_{i=1}^{N} \| \boldsymbol{x}_i' - \boldsymbol{y}_i \|^2, \qquad [15]$$

where $\boldsymbol{c} = [c_i, \ldots, c_m]^T$. In Figure 7b we recolor the histogram using the RMSD minimized over 1% of the modal coordinates. The magnitude of the RMSD is substantially reduced with 2.0Å RMSD on average. We observe that by parameterizing fluctuations about our 1 DOF pathway using 1% of the normal modes we can describe all MD conformations as fluctuations around our pathway. Increasing the number of normal modes will make this number better. In addition, if we restrict attention to only those MD conformations that occur between the times of the two extreme conformations (Figure 4), the RMSD of our pathway with the superimposed 1% normal modes is about 1Å from the MD results as shown in Figure 8. Figure 9 illustrates a conceptual diagram which represents the results of ENI as a 1 DOF pathway where MD data can be captured as fluctuations using a small fraction of the lower frequency normal modes.

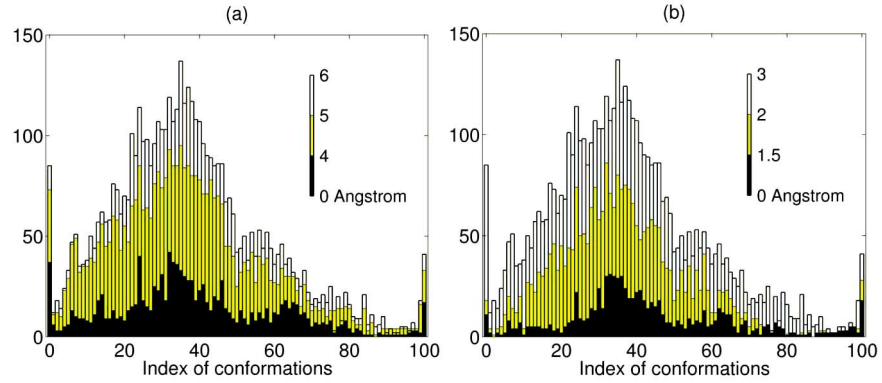


Figure 7: Histograms of MD conformations with respect to the nearest elastic network intermediate. (**a**) Histogram is made up of three different sections which are separated by RMSD values. The bar at the upper right corner indicates the RMSD range of each section. (**b**) Likewise, another histogram is generated between MD conformations and our intermediates. However, now 1% of the normal modes are used as coordinates to parameterize fluctuations around the elastic network pathway, and the RMSD values are substantially reduced to 2Å deviation on average.

Figure 8: Relationship between ENI and MD over a large motion that occurs during a short duration. (**a**) MD conformations from MD4698 (compact) to MD4763 (extended) are binned to the nearest elastic network intermediate. (**b**) Solid line indicates RMSD between the MD path and ours. The worst case is about 3Å. When we superimpose 1% of the normal modes on our intermediates, RMSD decreases to 1Å which is displayed as a solid-dotted line.

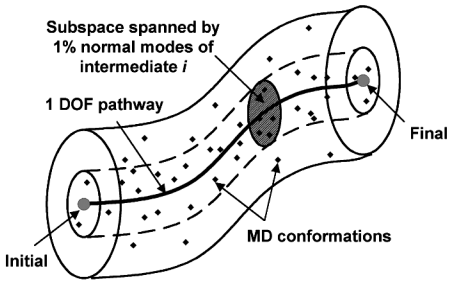
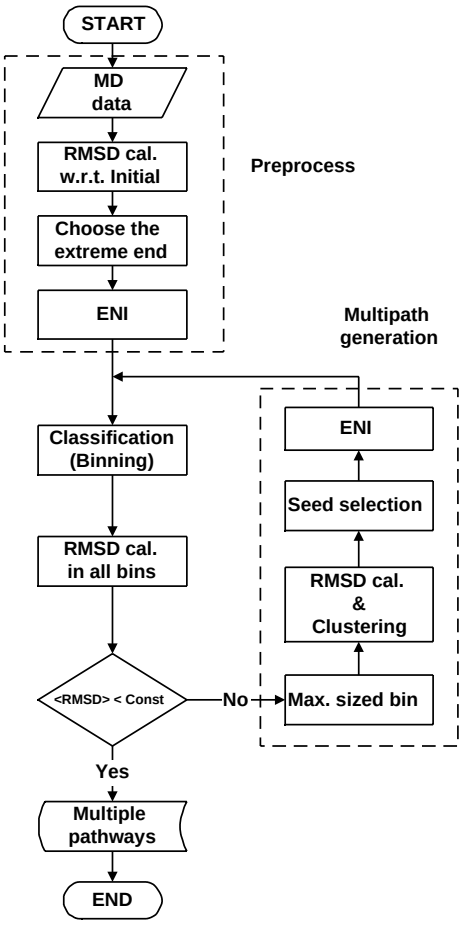
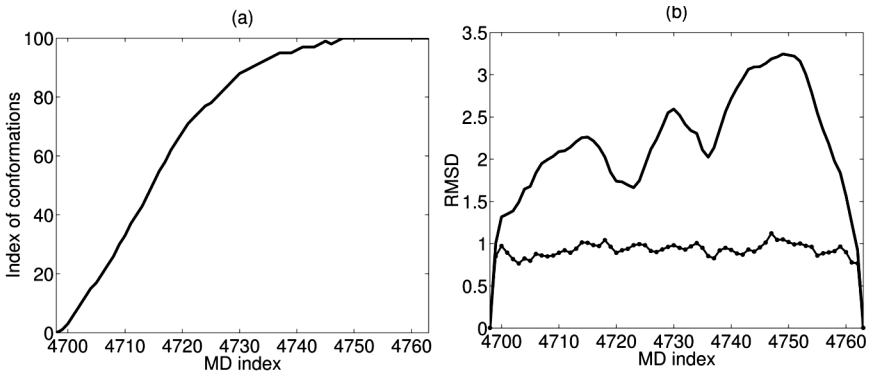


Figure 9: A conceptual tube diagram for conformational transitions. The ENI method generates a 1 DOF pathway between the two end conformations. With an average RMS error of 2Å, most MD simulation results can be captured as fluctuations in the subspace spanned by lower normal modes of intermediates (e.g. 1% in this context).





An Iterative Algorithm for Determining Multiple Pathways

ENI can generate a unique and smooth pathway between two extreme conformations chosen from the MD data. This produces an "average" conformational change which washes out fluctuations of the MD data. However, it is not true that all of the MD data always falls exactly along this 1-DOF pathway. In this case we can make a multi-pathway network which represents a variety of possible transition patterns. Figure 10 shows an iterative algorithm for determining multiple pathways between two states using the MD data and ENI. The algorithm for constructing this network of pathways is as follows:

(1) Preprocess:

Given sampled MD sampling data, calculate the RMSD with respect to the initial conformation in order to determine a final conformation (which is the furthest conformation from the initial as measured in RMSD). Generate a unique single pathway between the initial and final states using ENI.

(2) Decision:

Bin the MD data according to the nearest elastic network intermediates. In each bin, calculate the RMSD of the MD samples with respect to the representative conformation of that bin. If the average of the RMSD values over all bins is higher than a given threshold value, go through the iteration process below. Otherwise, stop here.

(3) Iteration:

Find the most heavily populated bin for the original single pathway (i.e., the bin which has the most MD samples). Construct a set of clusters in which no pair of MD samples has higher RMSD than a given value (i.e. users can adjust it to get the number of clusters as they want. Increasing the number of clusters results in a denser network of pathways). Select the MD point that has the lowest RMSD with others as seed in each cluster for generating multiple pathways by using ENI. Generate longitudinal pathways which start from the initial conformation and reach the final conformation while passing through the seeds. Network those pathways by connecting the MD seed conformations with conformations that are 50% along the newly calculated ENI pathways (see Figure 11). Go back to the decision process.

Using the original single pathway, the average RMSD value over all bins is 4.5Å. To illustrate the iteration process in this context, we arbitrarily set a threshold value equal to 3.5Å. Since the average RMSD is greater than this threshold, we move to the iteration process. We select the most heavily populated bin, bin35, in which we

Figure 10: A flowchart for generating multiple pathways. Given two extreme conformations (i.e. one is the initial conformation, the other is chosen from the MD data), ENI can generate a feasible pathway. We bin the MD data along this pathway and evaluate quantitatively if this path represents well the collective motions of the MD simulation. If not, we can iteratively build a multi-pathway network in order to capture all MD fluctuations with a threshold of the average RMSD.

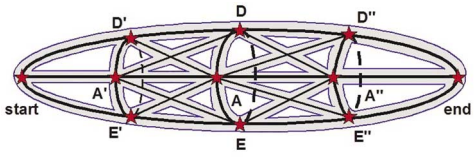


Figure 11: A cartoon depicting multiple pathways. Each tube around a branch in this network of pathways is analogous to that in Figure 9. The original ENI path is denoted as A. This can be thought of as the shortest path between two extremes in terms of RMSD. Two other pathways, D and E, are iteratively added to generate a multi-pathway network. The intermediates, A', D', and E' are chosen half way between the starting conformation and each MD seed conformation, whereas A'', D'', and E'' are chosen half way between each MD seed conformation and the ending conformation. More detail descriptions are provided in Table I.

can choose several seed points for new multiple pathways. Before choosing seeds, we first make clusters as follows: (i) For example, assume the diameter of a cluster to be 7.2Å. (ii) Make the first cluster with MD samples where the RMSD with respect to the representative conformation of the bin is less than 3.6Å. That is, think of this seed as the center of sphere and then take all MD samples within the sphere's radius. (iii) For the remaining MD samples in the bin, make other clusters using the rule as mentioned in the "iteration part" until nothing is left.

Including the first cluster, three different clusters are consequently built when applying this algorithm to our current data. The MD sample which has the lowest RMSD with others in the same cluster is chosen as a seed from the second and third clusters, respectively. Using those seeds, we generate two other longitudinal pathways and add them to the original single pathway. Figure 11 shows the multi-pathway network which consists of three longitudinal pathways denoted as A, D, and E and the additional pathways. They are networked between three seeds and six other points, denoted as A', D', E', A'', D'', and E''.

Next, we bin all of MD data again over this network of pathways. Table I presents the number of bins and that of MD samples in each pathway. We then recalculate the RMSD between the MD samples in each bin and its representative conformation. Figure 12a shows histograms of those RMSD values. The average RMSD is

Table I
List of multiple pathways

Path Name	# of Bins	# of MD Samples	References
A	101	87	The original ENI path. Start and end points are included
D	99	875	The second longitudinal path
E	99	714	The third longitudinal path
AD	49	206	Transverse path between A and D
AE	49	221	Transverse path between A and E
DE	49	317	Transverse path between D and E
A'D'	12	18	Transverse path between A' and D'
A'E'	12	20	Transverse path between A' and E'
D'E'	12	126	Transverse path between D' and E'
A''D''	12	0	Transverse path between A'' and D''
A''E''	12	31	Transverse path between A'' and E''
D''E''	12	39	Transverse path between D'' and E''
A'D	24	268	Longitudinal path between A' and D
A'E	24	121	Longitudinal path between A' and E
D'A	24	59	Longitudinal path between D' and A
D'E	24	481	Longitudinal path between D' and E
E'A	24	73	Longitudinal path between E' and A
E'D	24	611	Longitudinal path between E' and D
AD''	24	1	Longitudinal path between A and D''
AE''	24	71	Longitudinal path between A and E''
DA''	24	81	Longitudinal path between D and A''
DE''	24	142	Longitudinal path between D and E''
EA''	24	221	Longitudinal path between E and A''
ED''	24	217	Longitudinal path between E and D''

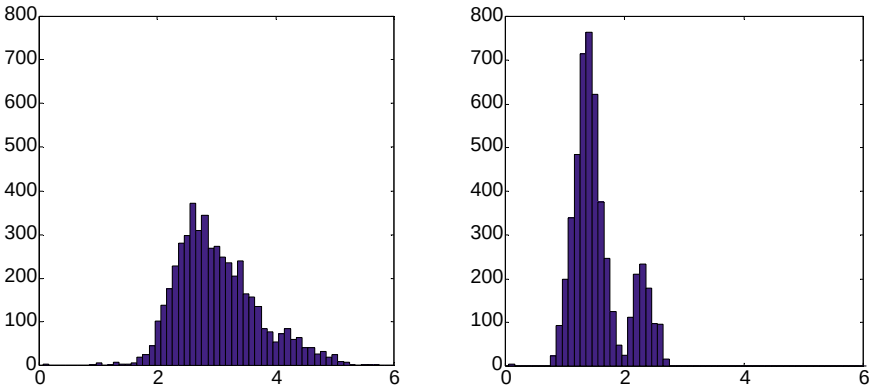


Figure 12: Histograms of RMSD between the MD data and multiple pathways. (**a**) The average is about 3Å, which is better than that of the single pathway by 1.5Å. (**b**) Multiple pathways superimposed with 1% of the normal modes deviate from all MD data with an average of 1.5Å.

reduced from 4.5Å to 3Å, which is below the given threshold. We can therefore stop here. A denser network of pathways would reduce this number further.

Table II
RMSD between the ENI paths and the MD samples*

	ENI	ENI/NMA
Single path	4.5Å(6.6Å)	2.0Å(3.1Å)
Multiple paths	3.0Å(5.7Å)	1.5Å(2.7Å)

* Average RMSD values are displayed. The value in parenthesis is the maximum.

In addition to the above procedure for generating pathways that capture anharmonic motions, we perform NMA on every conformation contained in the network of pathways. We then allow the conformations on the pathways the freedom to move in all the directions spanned by 1% of their normal modes. In this way, we can represent all MD data as fluctuations about this network of pathways at 1.5Å RMSD on average (Figure 12b). Table II summarizes the average RMSD values in both single and multi-pathway cases.

Conclusions

We have described a method that uses a coarse-grained elastic network model to generate feasible pathways for conformational transitions in the core central domain of the 16S Ribosomal RNA. Coarse-grained modeling and cutoffs in the number of nearest neighbors generate a sparse and uniformly dense linking matrix which permits efficient computations. This is a fast method for generating conformational transitions while still preserving steric constraints. Unlike MD in which the size of the timestep used is limited by the stiffest part of the structure, network interpolation is purely geometric and so intermediates are generated only by the difference in shape between the two conformations.

To compare network interpolation and MD, we take the 16S rRNA structure and run an MD simulation from which 5000 conformations are sampled. Then 99 intermediates between the two extreme conformations are generated using ENI. Simulation results illustrate that the ENI method presented here reliably generates sequences of feasible intermediate conformations of the 16S rRNA without steric clashes. Animations produced using this method are posted on the web (25). We also generate a multi-pathway network based on the original single ENI pathway and then bin all 5000 MD conformations according to how close each one is to the nearest of our intermediates in the sense of RMSD. By parameterizing fluctuations about our multiple pathways using only 1% of the normal modes in each bin we capture well all 5000 MD conformations as fluctuations. In addition, if we only concentrate on those conformations that occur between the two extreme conformations, the RMSD of our pathway with superimposed 1% normal modes is about 1Å away from the MD results. These results may play an important role in reduced-DOF dynamic simulations of large biological macromolecules as well as the reduced-parameter interpretation of massive amounts of MD data.

References and Footnotes

1. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne. *Nucleic Acids Res. 28*, 235-242 (2000).
2. S. Subbiah. *Protein Motions*. R.G. Landes Company, Austin, TX (1996).
3. B. Brooks and M. Karplus. *Proc. Natl. Acad. Sci. USA 80*, 6571-6575 (1983).
4. M. M. Tirion and D. Ben-Avraham. *Physica A 249*, 415-423 (1998).

5. I. Bahar and R. L. Jernigan. *J. Mol. Biol. 281*, 871-884 (1998).
6. I. Bahar, B. Erman, R. L. Jernigan, A. R. Atilgan and D. G. Covell. *J. Mol. Biol. 285*, 1023-1037 (1999).
7. A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin and I. Bahar. *Biophys. J. 80*, 505-515 (2001).
8. F. Tama, W. Wriggers and C. L. Brooks. *J. Mol. Biol. 321*, 297-305 (2002).
9. D. Ming, Y. Kong, M. A. Lambert, Z. Huang and J. Ma. *Proc. Natl. Acad. Sci. USA 99*, 8620-8625 (2002).
10. C. Vonrhein, G. J. Schlauderer and G. E. Schulz. *Structure 3*, 483-490 (1995).
11. G. J. Kleywegt and T. A. Jones. *Structure 3*, 535-540 (1995).
12. G. J. Kleywegt and T. A. Jones. *Structure 4*, 1395-1400 (1996).
13. M. K. Kim, R. L. Jernigan and G. S. Chirikjian. *Biophys. J. 83*, 1620-1630 (2002).
14. M. K. Kim, G. S. Chirikjian and R. L. Jernigan. *J. Mol. Graph. Model. 21*, 151-160 (2002).
15. S. C. Agalarov, G. S. Prasad, P. M. Funke, C. D. Stout and J. R. Williamson. *Science 288*, 107-112 (2000).
16. W. Li, B. Ma and B. A. Shapiro. *J. Biomol. Struct. Dyn. 19*, 381-396 (2001).
17. W. Li, B. Ma, and B. A. Shapiro. *Nucleic Acid Res. 31*, 629-638 (2003).
18. R. T. Batey and J. R. Williamson. *J. Mol. Biol. 261*, 550-567 (1996).
19. T. Ha, X. Zhuang, H. D. Kim, J. W. Orr, *et al. Proc. Natl Acad. Sci. USA 96*, 9077-9082 (1999).
20. H. G. Petersen. *J. Chem. Phy. 103*, 3668-3679 (1995).
21. P. K. Weiner and P. A. Kollman. *J. Comput. Chem. 2*, 287-303 (1981).
22. S. C. Agalarov and J. R. Wiliamson. *RNA 6*, 402-408 (2000).
23. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, *et al. J. Am. Chem. Soc. 117*, 5179-5197 (1995).
24. D. A. Case, D.A. Pearlman, J. W. Caldwell, T. E. Cheatham, III, W. S. Ross, C. Simmerling, et al. *AMBER6 Manual*. University of California, San Francisco, CA (1999).
25. http://custer.me.jhu.edu/proteins/rna.html
26. W. L. DeLano. *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA (2002).

Date Received: April 24, 2003

Communicated by the Editor Ramaswamy H Sarma