

# Rigid-Cluster Models of Conformational Transitions in Macromolecular Machines and Assemblies

Moon K. Kim,\* Robert L. Jernigan,<sup>†</sup> and Gregory S. Chirikjian<sup>‡</sup>

\*Department of Mechanical and Industrial Engineering, University of Massachusetts at Amherst, Amherst, Massachusetts;

<sup>†</sup>Laurence H. Baker Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, Iowa; and

<sup>‡</sup>Department of Mechanical Engineering, The Johns Hopkins University, Baltimore, Maryland

**ABSTRACT** We present a rigid-body-based technique (called rigid-cluster elastic network interpolation) to generate feasible transition pathways between two distinct conformations of a macromolecular assembly. Many biological molecules and assemblies consist of domains which act more or less as rigid bodies during large conformational changes. These collective motions are thought to be strongly related with the functions of a system. This fact encourages us to simply model a macromolecule or assembly as a set of rigid bodies which are interconnected with distance constraints. In previous articles, we developed coarse-grained elastic network interpolation (ENI) in which, for example, only  $C^\alpha$  atoms are selected as representatives in each residue of a protein. We interpolate distance differences of two conformations in ENI by using a simple quadratic cost function, and the feasible conformations are generated without steric conflicts. Rigid-cluster interpolation is an extension of the ENI method with rigid-clusters replacing point masses. Now the intermediate conformations in an anharmonic pathway can be determined by the translational and rotational displacements of large clusters in such a way that distance constraints are observed. We present the derivation of the rigid-cluster model and apply it to a variety of macromolecular assemblies. Rigid-cluster ENI is then modified for a hybrid model represented by a mixture of rigid clusters and point masses. Simulation results show that both rigid-cluster and hybrid ENI methods generate sterically feasible pathways of large systems in a very short time. For example, the HK97 virus capsid is an icosahedral symmetric assembly composed of 60 identical asymmetric units. Its original Hessian matrix size for a  $C^\alpha$  coarse-grained model is  $>(300,000)^2$ . However, it reduces to  $(84)^2$  when we apply the rigid-cluster model with icosahedral symmetry constraints. The computational cost of the interpolation no longer scales heavily with the size of structures; instead, it depends strongly on the minimal number of rigid clusters into which the system can be decomposed.

## INTRODUCTION

Macromolecules and their assemblages (i.e., proteins, nucleic acids, carbohydrates, lipids, and complexes) play critical roles in living cells. Motions and conformational transitions are involved in the biological functions of the cell. Hence, comprehending the mechanics of such processes is imperative to understand the phenomena of life. During the last several decades, more than 30,000 macromolecule structures have been obtained experimentally and deposited in the Protein Data Bank (1). Due to this large amount of structural information, research areas such as computational biology, bioinformatics, and protein dynamics have been growing rapidly.

Molecular dynamics (MD) simulation and normal mode analysis (NMA) have been utilized for the dynamic analysis of macromolecules. Both of these are based on full-atom empirical potential functions describing chemical properties of macromolecules. These methodologies, however, become computationally inefficient or impractical as the size of the macromolecular systems of interest continues to increase. To reduce such a computational burden, Tirion (2) originally proposed an elastic network model in which a system is represented as a network of linear springs. Sophisticated

empirical potential models are replaced by a single-parameter Hookean potential in her elastic network model. Atilgan et al. (3) further simplified this elastic network model by reducing the number of degrees-of-freedom (DOF) by coarse-graining. For example, only  $C^\alpha$  atoms in a protein are treated as point masses and spatially proximal points are assumed to be linked with linear springs. Only structural (i.e., geometric) information is used to define a simple harmonic potential function. NMA based on this coarse-grained elastic network model has been performed to study the dynamics of the HK97 virus capsid (4). This is computationally more efficient than conventional approaches such as MD or even NMA using full-atom empirical potentials. Tama et al. (5) proposed the rotation-translation block method. In this approach, the macromolecule is divided into blocks, each of which consists of a few consecutive residues, and then the low-frequency normal modes are obtained as a linear combination of rigid-body motions of these blocks. NMA using the rotation-translation block method and the  $C^\alpha$  coarse-grained elastic network model has successfully approximated the low-frequency normal modes of large macromolecules and assemblies (6). Although efficient methods for obtaining normal modes can be quite important, the problem of obtaining conformational pathways addressed here is quite different.

Submitted April 14, 2004, and accepted for publication March 28, 2005.

Address reprint requests to Professor Gregory S. Chirikjian, Tel.: 410-516-7127; E-mail: gregc@jhu.edu.

© 2005 by the Biophysical Society

0006-3495/05/07/43/13 \$2.00

doi: 10.1529/biophysj.104.044347

To date, several approaches have been proposed to generate transition pathways using NMA. Mouawad and Perahia (7) found four possible intermediate conformations between the *T* and *R* states of human hemoglobin by NMA, followed by an energy minimization process. The *T* structure was deformed iteratively along several productive modes (i.e., the modes with the most important projection on the displacement vector from *T* to *R* structures) to the distance corresponding to the minimum root-mean-square deviation (RMSD) from *R*. Similarly, Xu et al. (8) scaled the size of deformation of the *T* structure induced by the lowest modes resulting in a single intermediate conformation. These studies do not yield a transition pathway between two end conformations, but generate a few feasible intermediate conformations.

Tama and Brooks (9) succeeded to create a putative pathway for the swelling process of an icosahedral virus by displacing the initial conformation along the direction of a single breathing mode. However, this linear method might fail to represent anharmonic nonlinear motions such as bending and twisting because a single normal mode only indicates infinitesimal motion along the direction of a more global transformation. Miyashita et al. (10) recently addressed a novel approach to reflect this nonlinearity of motions in macromolecules. They iteratively generate the nonlinear transition pathway between open and closed conformations of adenylyl kinase. In each step, NMA is performed for the current structure and then it is displaced along several normal modes that maximize advancement toward the final conformation (11–14).

Kim et al. (15) developed elastic network interpolation (ENI), which is a purely geometry-based technique. The essence of ENI is to uniformly interpolate the distances in two different conformations within the context of the coarse-grained elastic network model. ENI generates a feasible reaction pathway between two different conformations. It is suitable to describe the global motions of complex systems composed of small proteins or single proteins having up to several thousand residues within a reasonable time (i.e., a few hours on a desktop PC). In instances when only partial conformational data are obtained from experiments such as fluorescence resonance energy transfer or nuclear magnetic resonance, ENI can be utilized to incorporate that incomplete information in computer simulations (16). Moreover, ENI is used to interpret massive amounts of MD data by finding essential pathways (17).

A study of molecular dynamics has shown that most conformational changes in macromolecules can be classified into hinge and shear motions (18). For example, the hinge-bending motion of lactoferrin makes the iron-ion binding site open and close. The major conformational change in the HK97 viral capsid is the expansion of the capsid during the maturation process, which is induced by the shear motion between skewed trimers in each asymmetric unit (19). Since these motions are associated with the collective behavior of atoms, the structure of a macromolecule in which such

motions play an important role can often be treated as a set of rigid clusters. Note, however, that distributed nonlocalized motions such as overall stretching cannot be represented with this type of approach.

In this article, a new conformational interpolation method called rigid-cluster ENI is addressed, which is morphing (not NMA) of rigid clusters interconnected with distance constraints. When two conformations of a given protein assembly are reported, it is easy to verify which corresponding substructures in the two conformations are essentially the same up to rigid-body motion. In this case, these substructures are treated as rigid without reducing the quality of the data used as the input or making any assumptions that have not already been made in the process of obtaining the original structures. In this context our method is clearly appropriate, and is really the only tool available for use on a PC to morph very large structures in a way that observes realistic constraints on bond angles, bond lengths, and interresidue distances.

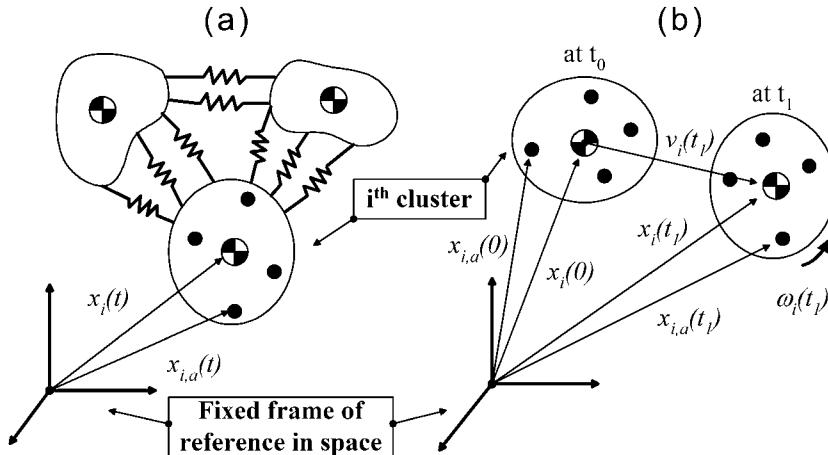
Since the main goal of this study is to provide handy meaningful tools to develop simple transition pathways as an alternative to the complexities of MD, there are possible choices of parameterization to obtain the desired combination between model resolution and computational efficiency. For an ideal hinge motion, rigid-cluster ENI is enough to catch its conformational change with negligible deviation from the target conformation. For a complex structure containing flexible regions such as loops, we could apply the hybrid method in which flexible regions are modeled with higher resolution than other rigid regions (20). This tradeoff can be adjusted by the user's preference between resolution and efficiency. A complete derivation is presented and then many example systems are used to illustrate and examine these methodologies.

## METHOD

### Rigid-cluster ENI

Schuylar and Chirikjian (21) recently proposed a rigid-cluster model for which low frequency normal modes had excellent agreement with those calculated from  $C^\alpha$  coarse-graining. The basic idea is to interconnect a set of rigid clusters with linear springs. It is equivalent to replacing point masses with rigid clusters in the conventional coarse-grained elastic network model. This approach is somewhat different from the substructure synthesis method (SSM) reported by Ming et al. (22). SSM synthesizes the normal modes for the assembled structure from those of its substructures by applying the Rayleigh-Ritz principle. It is much faster than directly solving the full eigenvalue problem. A set of constraints enforces the geometric compatibility at the interface between adjacent substructures. However, the rigid-cluster method assumes substructures as rigid bodies first and then NMA is performed by using standard approaches.

By adopting this rigid-cluster model, an incremental formulation is derived here to generate intermediate conformations along an anharmonic pathway between two conformations of the same macromolecule, which are denoted by sets of Cartesian coordinates  $\{\vec{x}_i\}$  and  $\{\vec{y}_i\}$ , respectively. One can build a rigid-cluster elastic network model for each of them as shown in Fig. 1 *a*.



Assume that each conformation can be treated as an assembly of  $N$  rigid bodies. The center of mass of the  $i^{\text{th}}$  cluster is

$$\vec{x}_i(t) = \frac{1}{m_i} \sum_{a=1}^{n(i)} m_{i,a} \vec{x}_{i,a}(t), \quad (1)$$

where  $m_i$  is the total mass of cluster  $i$ ,  $n(i)$  is the number of residues in cluster  $i$ ,  $m_{i,a}$  is the mass of residue  $a$  of cluster  $i$ , and  $\vec{x}_{i,a}(t)$  is the vector of Cartesian coordinates of residue  $a$  of cluster  $i$  at time  $t$ . Hence,

$$m_i = \sum_{a=1}^{n(i)} m_{i,a}. \quad (2)$$

Fig. 1 b shows rigid-body motion of the  $i^{\text{th}}$  cluster with respect to time  $t$ . The translational displacement of cluster  $i$  is defined as

$$\vec{v}_i(t) = \vec{x}_i(t) - \vec{x}_i(0), \quad (3)$$

where  $\vec{x}_i(0)$  is calculated from the reference crystal structure. The orientational displacement of cluster  $i$ ,  $\vec{\omega}_i(t)$ , is defined as the rotation matrix (see Appendix A),

$$R(\vec{\omega}_i(t)) = \exp[\text{mat}(\vec{\omega}_i(t))], \quad (4)$$

where  $\vec{\omega}_i(t)$  is parallel to the axis of rotation,  $\|\vec{\omega}_i(t)\|$  is the angle of rotation about that axis, and “mat” is the operator that converts a  $3 \times 1$  vector into a skew-symmetric matrix such that

$$\text{mat}(\vec{x}) = \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix}. \quad (5)$$

Therefore, the position of residue  $a$  of cluster  $i$  at time  $t$  is denoted by

$$\vec{x}_{i,a}(t) = R(\vec{\omega}_i(t))(\vec{x}_{i,a}(0) - \vec{x}_i(0)) + \vec{x}_i(0) + \vec{v}_i(t). \quad (6)$$

Assuming small rigid-body displacements from the initial conformation, the rotation matrix  $R(\vec{\omega}_i(t))$  is approximated such that

$$R(\vec{\omega}_i(t)) \approx I_3 + \text{mat}(\vec{\omega}_i(t)), \quad (7)$$

where  $I_3$  is the  $3 \times 3$  identity matrix. Substitution into Eq. 6 yields

$$\vec{x}_{i,a}(t) \triangleq \vec{x}_{i,a}(0) + H_{i,a} \vec{\delta}_i(t), \quad (8)$$

FIGURE 1 Schematic of a rigid cluster model. (a) In this model, one rigid cluster is connected to the others with distance constraints. This makes the system overconstrained. Conflicts are resolved with a quadratic penalty function as shown in Eq. 9, which is equivalent to viewing constraints as linear springs. (b) The pose of the  $i^{\text{th}}$  rigid cluster is presented. Small rigid-body motions of a rigid cluster are assumed in this context but exaggerated here for emphasis.

where the constant matrix  $H_{i,a} = [I_3 - \text{mat}(\vec{x}_{i,a}(0) - \vec{x}_i(0))] \in \mathcal{R}^{3 \times 6}$  and the displacement parameter  $\vec{\delta}_i(t) = [\vec{v}_i^T(t) \ \vec{\omega}_i^T(t)]^T \in \mathcal{R}^6$ . Note that the symbols  $\vec{v}_i$  and  $\vec{\omega}_i$  do not represent translational and angular velocities, but instead describe small displacements.

We propose a cost function to calculate incremental conformational changes of the system, which consists of  $N$  rigid clusters such that

$$C = \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left\{ \sum_{a=1}^{n(i)} \sum_{b=1}^{n(j)} k_{i,a,j,b} (\|\vec{x}_{i,a} + H_{i,a} \vec{\delta}_i - \vec{x}_{j,b} - H_{j,b} \vec{\delta}_j\| - l_{i,a,j,b})^2 \right\}, \quad (9)$$

where  $\|\cdot\|$  denotes the length of a vector and  $k_{i,a,j,b}$  is the spring constant between residue  $a$  of cluster  $i$  and residue  $b$  of cluster  $j$ . We simply set the value 1 for  $k_{i,a,j,b}$  whenever two residues are judged to be in contact in either conformation and 0 otherwise (for more details, see Computational Complexity, below). This cost function can be related to a classical pairwise potential function which is a simple harmonic function at an equilibrium state. We seek to establish a series of artificial equilibrium states between the two end conformations by perturbing this potential at an equilibrium state. Therefore, the goal at each iteration step is to let the system relax so that the calculated displacement vector causes the new conformation to settle at the new artificial equilibrium. In this way there is a penalty for the distance between interacting pairs to deviate after each incremental motion. The value  $l_{i,a,j,b}$  is an estimate of the distance between these two residues in an intermediate conformation, which can be chosen as

$$l_{i,a,j,b} = (1 - \alpha) \|\vec{x}_{i,a} - \vec{x}_{j,b}\| + \alpha \|\vec{y}_{i,a} - \vec{y}_{j,b}\|, \quad (10)$$

where  $\alpha \in [0, 1]$  is the coefficient specifying how far a given state is along the transition from  $\{\vec{x}_i\}$  to  $\{\vec{y}_i\}$  (15). We only consider interconnections among rigid clusters from the conventional elastic network model because intraconnections within a cluster are preserved under rigid-body motions.

Intermediate conformations are generated iteratively by altering the coefficient  $\alpha$  linearly from 0 to 1. Typically we generate 100 intermediate conformations, each of which is labeled with an index  $k$  to indicate that it is  $k\%$  along the path from  $\{\vec{x}_i\}$  to  $\{\vec{y}_i\}$ . For example, index 0 represents the initial conformation whereas index 100 stands for the final conformation. Unless the simulation breaks down, one can save computational time for generating the pathway by using a larger increment value of  $\alpha$ , which results in reducing the number of intermediate conformations. Often protein folding proceeds along multiple pathways. It can be anticipated that conformational transitions can likewise follow multiple pathways, particularly if the transitions are large ones.

The square-term of Eq. 9 can be linearized as

$$\begin{aligned}
(\| \underbrace{[\vec{x}_{i,a} - \vec{x}_{j,b}]}_X + \underbrace{[H_{i,a}\vec{\delta}_i - H_{j,b}\vec{\delta}_j]}_Y \| \underbrace{-l_{i,a,j,b}}_1)^2 &= (\| X + Y \| - l)^2 \approx \| Y \|^2 - \frac{l}{\| X \|} Y^T (I_3 - \frac{XX^T}{\| X \|^2}) Y \xleftarrow{2^{\text{nd}} \text{ order terms}} \\
&\quad + 2(1 - \frac{l}{\| X \|^2}) X^T Y \xleftarrow{1^{\text{st}} \text{ order terms}} + \| X \|^2 - 2l \| X \| + l^2 \xleftarrow{\text{constants}}. \quad (11)
\end{aligned}$$

The above second-order terms can be written in quadratic form

$$\underbrace{\vec{\Delta}_{i,j}^T [H_{i,a} - H_{j,b}]^T \{I_3 - \frac{l}{\| X \|^2} (I_3 - \frac{XX^T}{\| X \|^2})\} [H_{i,a} - H_{j,b}]}_{G_{i,a,j,b}} \vec{\Delta}_{i,j}, \quad (12)$$

where  $\vec{\Delta}_{i,j} = [\vec{\delta}_i^T \vec{\delta}_j^T]^T \in \mathcal{R}^{12}$  and  $Y = [H_{i,a} - H_{j,b}] \vec{\Delta}_{i,j}$ .  $G_{i,j}$  can be defined as

$$G_{i,j} = \sum_{a=1}^{n(i)} \sum_{b=1}^{n(j)} k_{i,a,j,b} G_{i,a,j,b}. \quad (13)$$

Likewise, the first-order terms can be expressed in a simple way such that

$$\underbrace{2(1 - \frac{l}{\| X \|^2}) X^T [H_{i,a} - H_{j,b}]}_{g_{i,a,j,b}} \vec{\Delta}_{i,j}, g_{i,j} = \sum_{a=1}^{n(i)} \sum_{b=1}^{n(j)} k_{i,a,j,b} g_{i,a,j,b}. \quad (14)$$

Eq. 9 consequently takes a simplified form with respect to  $\vec{\Delta}_{i,j}$  as

$$C = \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left\{ \vec{\Delta}_{i,j}^T G_{i,j} \vec{\Delta}_{i,j} + g_{i,j} \vec{\Delta}_{i,j} + B_{i,j} \right\}, \quad (15)$$

where  $G_{i,j}$  is a  $12 \times 12$  matrix,  $g_{i,j}$  is a  $1 \times 12$  row vector, and  $B_{i,j}$  is the sum of constant terms derived in Eq. 11 such that

$$B_{i,j} = \sum_{a=1}^{n(i)} \sum_{b=1}^{n(j)} k_{i,a,j,b} (\| X \|^2 - 2l \| X \| + l^2). \quad (16)$$

If we denote  $G_{i,j}$  and  $g_{i,j}$  such that

$$G_{i,j} = \begin{pmatrix} P_{i,j} & Q_{i,j} \\ Q_{j,i} & S_{i,j} \end{pmatrix}, \quad \vec{g}_{i,j} = [\vec{u}_{i,j} \vec{v}_{i,j}], \quad (17)$$

where  $P$ ,  $Q$ , and  $S$  are  $6 \times 6$  matrices,  $Q_{i,j} = Q_{j,i}^T$ , and  $\vec{u}$ ,  $\vec{v}$  are  $1 \times 6$  row vectors, then Eq. 15 can be expressed in terms of  $\vec{\delta}(t) = [\vec{\delta}_1^T, \dots, \vec{\delta}_N^T]^T \in \mathcal{R}^{6N}$ ,

$$C(\vec{\delta}) = \frac{1}{2} \vec{\delta}^T \Gamma \vec{\delta} + \frac{1}{2} \vec{g}^T \vec{\delta} + B, \quad (18)$$

where  $\Gamma$  is a  $6N \times 6N$  matrix,  $\vec{g}$  is a  $1 \times 6N$  row vector, and  $B$  is a constant. They can be constructed as

$$\begin{aligned}
\Gamma_{i,j} &= \begin{cases} \sum_{a=1}^{i-1} S_{a,i} + \sum_{b=i+1}^N P_{i,b} \\ \text{if } i = i, \text{ for } 1 \leq i, j \leq N, \\ Q_{i,j} \text{ if } i \neq j \end{cases} \\
\vec{g}_i &= \sum_{a=1}^{i-1} \vec{v}_{a,i} + \sum_{b=i+1}^N \vec{u}_{i,b}, \\
B &= \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^N B_{i,j}, \quad (19)
\end{aligned}$$

where  $\Gamma_{i,j}$  is the  $(i, j)$  submatrix of  $\Gamma$  and  $\vec{g} = [\vec{g}_1, \dots, \vec{g}_N]$ .

We finally obtain  $\vec{\delta}$ , which minimizes the cost of Eq. 18 by setting its derivative zero,

$$\frac{\partial C(\vec{\delta})}{\partial \vec{\delta}} = \Gamma \vec{\delta} + \frac{1}{2} \vec{g}^T = \vec{0}. \quad (20)$$

Substitution of  $\vec{\delta}$  in Eq. 6 yields a new conformation.

Although ENI uses the elastic network model and the quadratic cost function as discussed, this is basically not an harmonic analysis for a single conformation such as SSM, but rather a tool to generate an anharmonic pathway between the two end conformations. Matrix  $\Gamma$  and vector  $\vec{g}$  in Eq. 20 are physically analogous to the stiffness matrix and the external force vector, respectively.

## Computational complexity

We have observed that the dynamic behavior and computational efficiency of NMA based on elastic network models vary with the distance cutoff value defining interactions. ENI is also sensitive to cutoff values because it is basically derived from elastic network models. Extremely short cutoff values representing only local interactions will cause unrealistic results that lead to discontinuous pathways, whereas larger cutoff values will increase the number of connections (i.e., the density of linking matrices) so that greater computation time is required for generating intermediate transitions in large macromolecules.

We demonstrated a way to produce uniformly sparse linking matrices not only to increase computational efficiency for the whole interpolation process but also to guarantee realistic results (15,16). We can connect one residue to its neighboring residues by increasing distance until the limiting number of contacts (i.e., 20 in this context) is reached, regardless of the actual distance of the last connection. This enables the linking matrix to remain sparse and uniformly dense because all residues will have the same number of connections. This is a smoothed, more uniform representation of an elastic network for morphing of biomolecular structures.

Computational complexity varies widely depending on the choice of model parameters. In this article we have considered two advantages of the rigid-cluster ENI method regarding computational cost. First, we can reduce the number of interactions from the conventional  $C^\alpha$  ENI model (Fig. 2 a). Intraconnections within each rigid cluster can be neglected because the distance of any two points in a rigid cluster does not change (Fig. 2 b). Furthermore, we can eliminate most of the interconnections between every pair of clusters because only six linear spring connections are necessary to describe rigid-body motions of a cluster relative to another (Fig. 2 c). In the present case, we leave up to 10 of the closest interconnections between every pair of clusters. This number is still much more than needed to make a rigid-cluster network overconstrained. That is, it is enough to describe the structural change geometrically. However, structural details of the relatively flexible regions such as hinges and loops will be lost and the resulting morphed conformations could be energetically less accurate. To accommodate this problem, also addressed is a hybrid method representing a mixture of rigid clusters and point masses so as not to oversimplify the flexibility of systems.

Fig. 2 compares the density of the linking matrix based on  $C^\alpha$  ENI with that of rigid-cluster ENI for lactoferrin. Most spring connections are

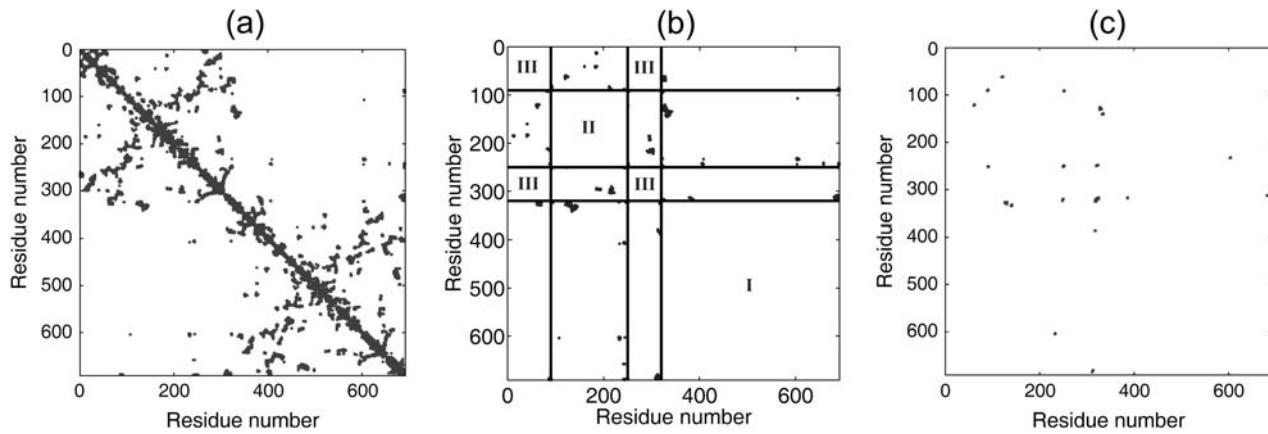


FIGURE 2 The sparseness patterns of the linking matrices for lactoferrin. (a) The contact number of 20 is used as a cutoff in  $C^\alpha$  ENI. This map displays the union of those constructed from the open and the closed conformations. The total number of interactions is 7562. (b) The lactoferrin structure has been simply modeled as three rigid clusters (see Fig. 3). Intraconnections within each cluster labeled as *I*, *II*, and *III* are eliminated and 409 interconnections remain to represent interactions between rigid clusters. (c) For further reduction, we leave up to 10 closest interconnections between every pair of clusters, and thus we have only 43 connections in the union-linking matrix of rigid-cluster ENI for the lactoferrin.

eliminated and only 0.6% of all the  $C^\alpha$ - $C^\alpha$  interactions are between rigid clusters in this particular case. This small number of connections can save a lot of computation time for constructing the ENI cost function in Eq. 18 from the atomic coordinates of the macromolecule provided (see  $T_1$  values in Table 1).

Second, the rigid-body representation needs very few DOF to describe system dynamics compared to an all-atom-based representation. Even  $C^\alpha$  coarse-grained approaches applied to very large macromolecular structures (e.g., GroEL-GroES complex, virus capsids, and ribosome) are still impracticable on a PC due to memory limitations. Rigid-clustering enables the size of  $\Gamma$  to be reduced. Hence, we can save a lot of time for the matrix inversion in Eq. 20.  $T_2$  values in Table 1 shows this relationship quantitatively. The sum of two parameters  $T_1$  and  $T_2$  is a total computation time per iteration. It is observed that rigid-cluster ENI is hundreds of times faster than the standard  $C^\alpha$  ENI in the lactoferrin and HK97 capsid. Hybrid ENI is also tens-of-times faster than symmetry-constrained ENI in the GroEL-GroES complex. Unless the system dynamics is changed too much, we can simplify large macromolecules as an assembly of rigid clusters, which can effectively reduce many costly problems that had required a super computer to needing only a single PC.

TABLE 1 Relationship among linking matrix density, DOF, and computational efficiency for sample proteins

Protein*	Method	Number of interactions	$T_1$ [s] <sup>†</sup>	DOF	$T_2$ [s] <sup>‡</sup>
Lactoferrin	[CG]	7562	9	2073	28
	[CG]+[RC]	43	0.11	18	<0.01
	[CG]+[SC]	23,764	272	5376	346
HK97	[CG]+[SC]+[RC]	1037	1.5	84	<0.01
	[CG]+[SC]	6910	40	1572	17
GroEL	[CG]+[SC]+[HB]	1457	2.3	162	0.08

[CG], [SC], [RC], and [HB] are the abbreviation of coarse-grained  $C^\alpha$  ENI, symmetry-constrained ENI, rigid-cluster ENI, and hybrid ENI, respectively.

\*All of the tests have been performed using MatLab on a 1.8-GHz Pentium PC with 2-GB memory. For the HK97 and GroEL, symmetry-constrained ENI is applied instead of the standard  $C^\alpha$  ENI because these proteins are too large to be implemented on a PC.

<sup>†</sup>The elapsed time for constructing  $\Gamma$  and  $\tilde{\gamma}$  of Eq. 18 in every iteration step.

<sup>‡</sup>The elapsed time for calculating the inverse of  $\Gamma$ . It entirely depends on the DOF of each method.

## SIMULATION RESULTS

Rigid-cluster ENI is tested for various systems that consist of several essentially rigid domains. For example, lactoferrin can be divided into three rigid clusters for its transition. Rigid-cluster ENI generates feasible pathways in lactoferrin. In addition, the conformational changes of symmetric systems such as the HK97 capsid and the GroEL-GroES complex are studied by using rigid-cluster ENI with symmetry constraints. Both rigid-clustering and symmetry constraints substantially reduce the computational cost of obtaining transition pathways for these large macromolecular assemblies. Hybrid ENI is also tested for the GroEL-GroES complex, which contains rigid and flexible domains together. The computational complexity of these methods is discussed at the end of this section.

## Lactoferrin

To apply the rigid-cluster ENI method, we have to choose the rigid clusters first. There is no unique way to define rigid clusters because in general macromolecules are not rigid structures. Rigid-clustering in this context starts with the static comparison of two end conformations based on structural information such as secondary structures and domains. If domains of a particular system are already defined in the literature, we can temporarily consider them as rigid clusters. Next, we calculate the windowed RMSD to finally define a set of rigid clusters minimizing RMSD between the two corresponding clusters obtained from each end conformation. Three rigid clusters have been chosen to represent the lactoferrin structure in Fig. 3. Clusters *I*, *II*, and *III* denote head (green), left (yellow), and right (red) lobes, respectively. Here Thr<sup>90</sup> and Val<sup>250</sup> bend like hinges between two lobes so that they can open and close the ion binding site. The RMSD

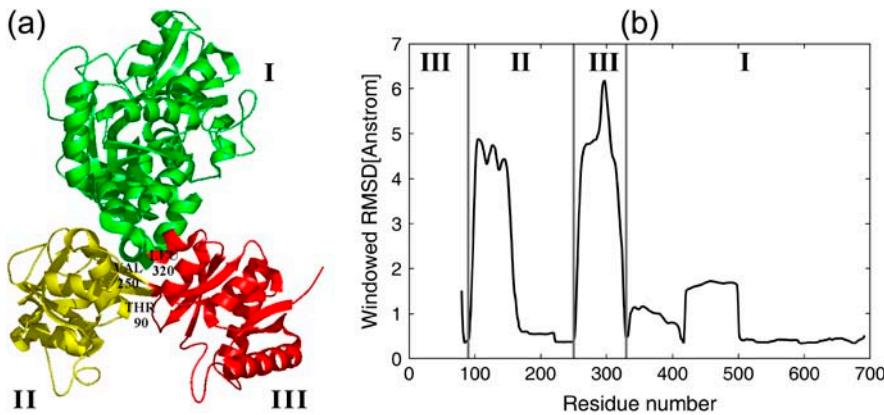


FIGURE 3 A rigid-cluster model of the lactoferrin structure. (a) Lactoferrin is assumed to have three rigid clusters: head (green), left (yellow), and right (red) lobes. Two lobes are opened and closed by the hinge motion around  $\text{Thr}^{90}$  and  $\text{Val}^{250}$ . RMSD between corresponding clusters in each conformation is displayed in Table 2. (b) For the open and closed forms of the lactoferrin, the windowed RMSD plot consecutively captures 80 residues per window. Three low peaks are chosen for the clustering in this context.

between the corresponding parts of each of the pairs of domains is  $\sim 1 \text{ \AA}$ . This value is relatively small compared with the large conformational change between the closed (1LFG) and open (1LFH) forms, which is  $\sim 7 \text{ \AA}$  in RMSD, so this is a good approximation.

Fig. 4 shows the pathway computed from rigid-cluster ENI. Clusters *II* and *III* separate so that the ion binding site is exposed. RMSD of all intermediates is measured with respect to the target (open) conformation. It decreases monotonically but there is 1  $\text{\AA}$  difference between the computed open conformation and the real open conformation because individual domains in the open conformation are not exactly the same as those in the closed conformation (see Table 2).

Since we are taking a coarse-grained approach, we cannot evaluate the detailed energetics as one usually does for all-atom-based models. Alternatively, we calculate the simple Hookean potential energy. Fig. 5 *a* shows the normalized Hookean potential energy of all intermediates for lactoferrin generated by rigid-cluster ENI. When we set the Hookean potential energy of the final conformation to be 0 (i.e., the final conformation is assumed to be an equilibrium state) and the relative potential of the initial conformation to be 1, the energy value monotonically converges to the offset arising from the error of the rigid cluster assumption because each of the rigid clusters in the two given conformations is not identical. Consequently, simple evaluation of the torsion-angle profile of the interesting residues provides us with an indirect metric for the quality of the computed pathway in

terms of the energetics. Fig. 5 *b* presents the changes of virtual torsion angles at the two hinge points  $\text{Thr}^{90}$  and  $\text{Val}^{250}$ . These monotonic angle changes indicate that the computed conformational transition seems to be naturally favorable and energetically feasible.

To check the atomic clash problem, the minimum distance among all pairs of  $C^\alpha$  atoms is measured for all intermediate conformations. The minimum is  $2.87 \text{ \AA}$  between  $\text{Cys}^{627}$  and  $\text{Pro}^{628}$  in Cluster *I*, and this intraconnection does not change during the transition within the context of rigid-cluster ENI. Therefore, there is no other distance shorter than this minimum intraconnection. This means that rigid-cluster ENI does not cause any steric problems among the rigid clusters in case of lactoferrin, which is true, at least, for this case.

The computational performance in rigid-cluster ENI is significantly better than that of  $C^\alpha$  coarse-grained elastic network interpolation ( $C^\alpha$  ENI). In this particular case, the DOF in the rigid-cluster model is only  $6 \times 3 = 18$ , whereas that of  $C^\alpha$  ENI is  $3 \times 691 = 2073$ .

#### HK97 Capsid: symmetry-constrained rigid-cluster ENI

In our previous article (4), the maturation process of the HK97 virus capsid was studied by using  $C^\alpha$  ENI with symmetry constraints. This icosahedral symmetric capsid is an assembly of 60 identical asymmetric units. Each unit consists of one hexamer plus an additional subunit from an

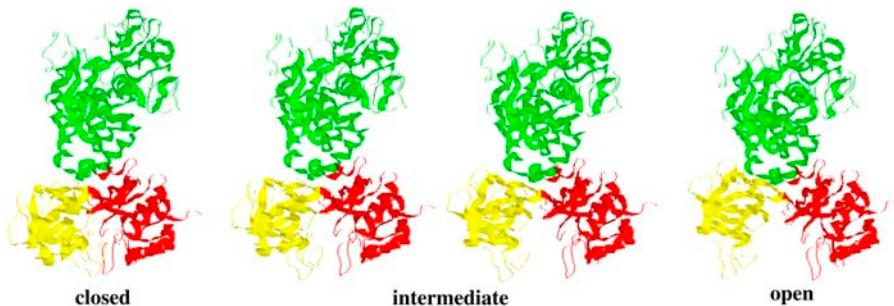


FIGURE 4 The conformational changes of lactoferrin generated by rigid-cluster ENI. The closed (left) and open (right) conformations are displayed with two intermediate conformations (middle) calculated from rigid-cluster ENI.

**TABLE 2** Three rigid clusters of lactoferrin

Cluster	Residue number	RMSD*
I	Gly <sup>321</sup> ~ Lys <sup>691</sup>	1.0
II	His <sup>91</sup> ~ Val <sup>250</sup>	0.5
III	Gly <sup>1</sup> ~ Thr <sup>90</sup> , Pro <sup>251</sup> ~ Leu <sup>320</sup>	1.2

\*RMSD between two corresponding rigid clusters that come from the closed and open conformations of lactoferrin, respectively, in units of Å.

adjacent pentamer as shown in Fig. 6. In this context, “subunit” means a capsid protein unit, which is the fundamental element from which to construct asymmetric units by copying itself. A subunit consists of the rigid core (*A* and *P* domains) and two motifs (*N*-arm and *E*-loop) named by Conway et al. (23). The rigid core (Asn<sup>182</sup>-Ser<sup>383</sup>) and *E*-loop (Leu<sup>148</sup>-Ala<sup>181</sup>) are each defined as rigid clusters. Since the pseudo-atomic model in the Prohead conformation (1IF0) has been obtained by the rigid-body mapping into the cryo-EM map (19), RMSD between corresponding clusters in the Prohead and Head (1FH6) conformations is obviously negligible (i.e., RMSD between the rigid cores is 0.1 Å and that of *E*-arms is 0.3 Å). However, the flexible *N*-arm (Gly<sup>128</sup>-Ala<sup>147</sup>) is discarded in this context only to test the rigid-cluster ENI method. Fig. 6 presents a rigid-cluster model for the HK97 capsid. An asymmetric unit is represented by 14 rigid clusters (i.e., 7 subunits × 2 clusters).

Given two end conformations, rigid-cluster ENI can be modified with symmetry constraints. It offers a significant computational advantage because it is not necessary to consider the whole structure but only a repeated asymmetric unit with symmetry constraints induced by the manner of assembly. The complete derivation is presented in the Appendix B. It should be noted that some slow modes of motion are not obtained in normal mode calculations when this symmetry condition is applied (24). Fig. 7 illustrates the conformational transition of the HK97 capsid generated with this methodology. During the simulation, no steric clashes occur between every possible pair of rigid clusters, regardless of their locations. This pathway is compared with that of  $C^\alpha$  symmetry-constrained ENI in Fig. 8. The DOF of the HK97 capsid in various models is calculated in Table 3. Both

rigid-clustering and symmetry constraints tremendously reduce the number of parameters so that one can save substantial computation time.

### GroEL-GroES complex: hybrid ENI

The GroEL-GroES complex assists protein folding with the consumption of ATP. X-ray crystallography has revealed that this chaperonin complex is formed by GroEL, GroES, and seven bound ADP molecules (PDB code: 1AON). The overall shape is elongated with the sevenfold symmetry as shown in Fig. 9. Seven GroEL units comprise a ring structure and then two rings are stacked back to back. The GroES caps one end of the GroEL ring which is tapered toward the GroEL-GroES interface (25,26). The GroEL and GroES share one rotation axis and each unit of the GroEL ring is composed of equatorial (Ala<sup>2</sup>-Pro<sup>137</sup> and Val<sup>411</sup>-Pro<sup>525</sup>), intermediate (Cys<sup>138</sup>-Gly<sup>192</sup> and Val<sup>376</sup>-Gly<sup>410</sup>), and apical (Met<sup>193</sup>-Gly<sup>375</sup>) domains. A shape change of the GroEL ring is associated with the GroES binding. When the GroES binds to the apical surface of the GroEL ring, it is observed that the apical and intermediate domains arise, which results in the expansion of the cavity volume. This is called the *cis* ring. In contrast, the release of the GroES from the *cis* ring triggers the conformational change back to the unliganded GroEL called the *trans* ring (27).

Since the equatorial domains show only a small movement to preserve the strong interface between the rings during the transition (28), it is not necessary that a rigid equatorial domain be modeled as a  $C^\alpha$  coarse-grained ENI. By contrast, oversimplification when using rigid-cluster ENI for other flexible domains may not represent such a system well. Hence, we treat rigid regions of each domain as rigid clusters, whereas relatively flexible regions that connect one domain to another domain remain as point masses in our elastic network model. Fig. 10 shows a schematic of the hybrid elastic network model in which rigid clusters and point masses are linked to one another with distance constraints (see Appendix C for the derivation). This hybrid ENI method compromises the computational benefit of low DOF parameterizations with the motion detail of high DOF parameterizations.

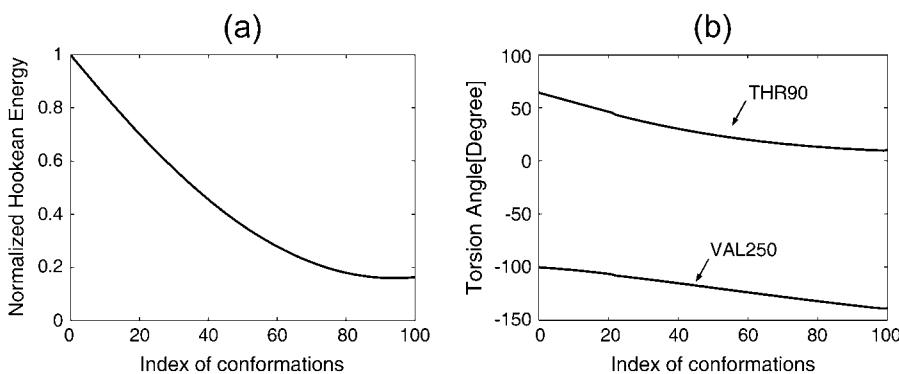


FIGURE 5 Characteristics of the simulated conformational transition in lactoferrin. (a) The normalized Hookean potential energy is calculated for each intermediate conformation. It decreases monotonically as the initial conformation follows the rigid-cluster pathway toward the final conformation. Here we note that the rigid-cluster pathway is energetically favorable. (b) Virtual torsion angles for Thr<sup>90</sup> and Val<sup>250</sup> vary monotonically during the transition. They act like hinges that open or close the two lobes. These results show that the rigid-cluster ENI is able to generate a feasible pathway observing distance constraints well.

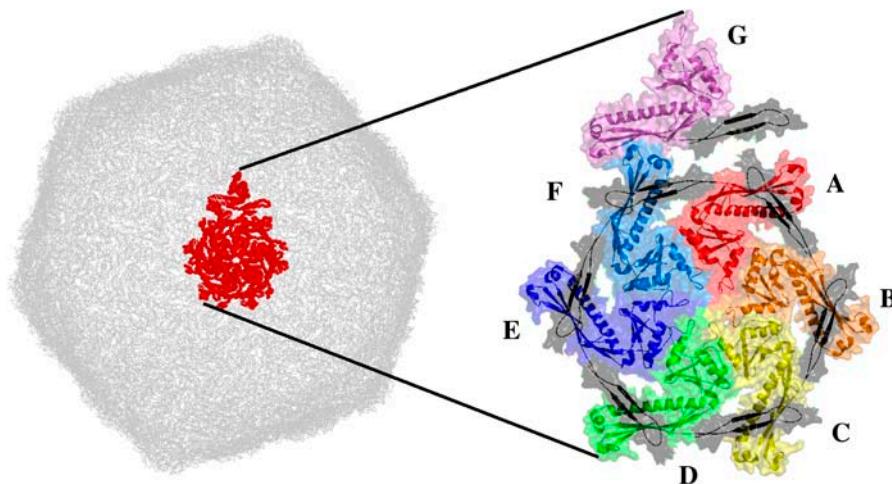


FIGURE 6 A rigid-cluster model of the HK97 capsid. An asymmetric unit of the icosahedral symmetric structure of the Head conformation is represented as 14 rigid clusters (i.e., seven *E*-arms, black; and seven rigid cores, rainbow).

Chain *A* and Chain *H* are selected from 1AON as representatives of *cis* and *trans* conformations of an asymmetric unit of the GroEL ring, respectively. A windowed RMSD is calculated to determine rigid and flexible regions as shown in Fig. 11. As a result, 11 rigid clusters and four flexible regions are chosen as shown in Table 4. RMSD between the corresponding clusters of *cis* and *trans* conformations is within 1.5 Å. This is relatively small when compared to the 12.4 Å conformational change of a GroEL subunit during its transition. In this particular case, Hybrid ENI with symmetry constraints reduces the number of DOF by 98.5% compared to conventional  $C^\alpha$  ENI (see Table 5). It also generates a feasible pathway between *cis* and *trans* conformations as illustrated in Fig. 12. All intermediate conformations are compared with those of  $C^\alpha$  ENI. No steric clashes are observed during the transition and the hybrid ENI pathway follows the  $C^\alpha$  ENI pathway to within a 1.7 Å deviation.

## CONCLUSIONS

In this article, the standard  $C^\alpha$  ENI method is extended to rigid-cluster systems. The conformational change of a rigid-cluster system can be represented by rigid-body motions such as hinge and shear motions. To represent these it is not necessary to model rigid domains of the structure in atomic detail. Only six parameters are required to capture the

translational and rotational motions of each rigid cluster. This strategy has a big impact on computational cost because it only depends on the number of rigid clusters modeled, regardless of the size of the molecules. Rigid-cluster ENI has been applied to lactoferrin. Even though the system is modeled with many fewer parameters than in  $C^\alpha$  ENI, the results are in good agreement with those of  $C^\alpha$  ENI. Furthermore, rigid-cluster ENI is exploited to be applicable to a symmetry-constrained system such as the HK97 capsid. Both rigid-clustering and symmetry constraints allow us to generate a feasible pathway of the HK97 capsid using <0.1% of the degrees of freedom of the  $C^\alpha$  ENI. For a system that has both rigid and flexible domains together,  $C^\alpha$  ENI and rigid-cluster ENI are combined as hybrid ENI. This is applied to the GroEL-GroES complex. The computed pathway shows that hybrid ENI not only simplifies the modeling of rigid domains, but also interpolates the motions of flexible parts at atomic detail. Animations produced by this work are posted on the web at [http://custer.me.jhu.edu/research/protein\\_kinematics/pathway\\_generation.html](http://custer.me.jhu.edu/research/protein_kinematics/pathway_generation.html). Rigid-cluster models may serve as a powerful and practical tool for the study of conformational transitions in large macromolecules where MD might not be feasible. Once again, the purpose of this study is not to generate the unique and fully detailed pathway, but to provide feasible and coarse-grained pathways at various resolutions by using several elastic network

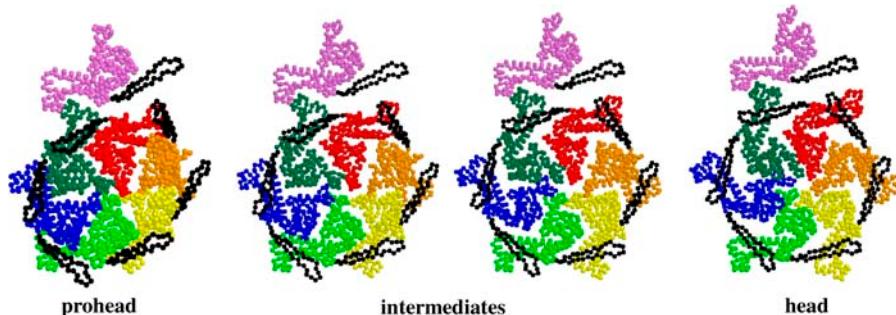


FIGURE 7 The transitional pathway of the HK97 capsid from the rigid-cluster ENI. The conformational change of an asymmetric unit is displayed as 14 rigid clusters. The lower DOF representation resulting from rigid-clustering saves substantial computational time as shown in Table 3, and produces acceptable results compared to those of the  $C^\alpha$  ENI method.

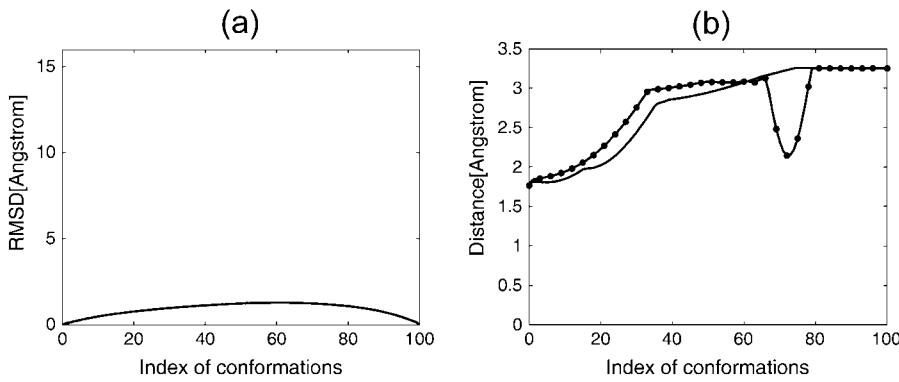


FIGURE 8 Comparison between  $C^\alpha$  ENI and rigid-cluster ENI for the HK97 capsid. (a) The minimum RMSD of all intermediates generated by  $C^\alpha$  ENI with respect to each intermediate conformation generated by rigid-cluster ENI is displayed. The rigid-cluster ENI pathway is quite close to that of  $C^\alpha$  symmetry-constrained ENI. It is within 1.3 Å during the swelling process of capsid with 16 Å RMSD. (b) The minimum distance between every pair of  $C^\alpha$  atoms of each intermediate conformation shows that rigid-cluster ENI (dotted line) also generates a feasible pathway without steric clashes as does  $C^\alpha$  ENI (solid line). Note that there is no other distance shorter than the minimum distance of the initial (Prohead) conformation.

models presented here (e.g., coarse-grained, rigid-cluster, and hybrid). The pathways we have generated thus far when applying these methods to a variety of structures have been largely similar to one another with little local deviation. However, it may certainly be the case that in some instances residues at critical locations can have major impact.

## APPENDIX A: RIGID-BODY KINEMATICS

A rigid-body motion preserves both distance and orientation between internal points before and after the motion. That is, if  $\vec{p}$  and  $\vec{q}$  are any two points in a rigid-body, then

$$\|g(\vec{p}) - g(\vec{q})\| = \|\vec{p} - \vec{q}\|, \quad (21)$$

where  $\vec{p}, \vec{q} \in \mathbb{R}^3$ ,  $g$  is an operator that represents a rigid-body transformation, and  $\|\cdot\|$  denotes the length of a vector. Fig. 13 a illustrates the relative orientation of a body-fixed frame, denoted as  $\{B\}$ , with respect to a space-fixed frame of reference, denoted as  $\{S\}$ . A rotation matrix is defined by stacking the coordinates of the principal axes of  $\{B\}$  relative to  $\{S\}$ , denoted as  $\vec{a}, \vec{b}$ , and  $\vec{c}$ , respectively (29):

$$R_S^B = [\vec{a} \ \vec{b} \ \vec{c}] \in \mathbb{R}^{3 \times 3}. \quad (22)$$

This matrix is interpreted as a coordinate transformation when the point  $p$  is represented by two different frames such that

$$\vec{p}_S = R_S^B \vec{p}_B, \quad (23)$$

where  $\vec{p}_S$  and  $\vec{p}_B$  are the coordinates of point  $p$  with respect to  $\{S\}$  and  $\{B\}$ , respectively.

TABLE 3 DOF of the HK97 capsid in various parameterizations

Method*	Asymmetric units	Residues	DOF	Total DOF	Fraction <sup>†</sup>
[CG]	60	1792	3	322,560	100
[CG]+[SC]	1	1792	3	5376	1.67
[CG]+[SC]+[RC]	1	14 clusters	6	84	0.03

The total DOF of the HK97 capsid in this particular modeling method.

\*All the notations follow those in Table 1.

<sup>†</sup>The fraction of the DOF of each method with respect to that of  $C^\alpha$  ENI is displayed as a percentage.

Many different parameterizations have been developed to describe rotations. In particular Fig. 13 b illustrates the axis-angle parameterization. A rotation matrix can be considered as an operator that rotates a vector resulting in a different vector under the same frame of reference (30). The orientational displacement,  $\vec{\omega}$ , in Eq. 4 can be decomposed into the rotation axis vector,  $\vec{n}$ , and the rotation angle,  $\theta$ , such that

$$\vec{\omega} = \theta \vec{n}, \quad (24)$$

where  $\vec{n} = \vec{\omega} / \|\vec{\omega}\| \in S^2$  and  $\theta = \|\vec{\omega}\| \in [-\pi, \pi]$ . Here  $S^2$  is the unit sphere whose center is at the origin. Using these parameters, a rotation matrix can be written as

$$R(\vec{\omega}) = I_3 + (\sin \theta) N + (1 - \cos \theta) N^2 = \sum_{k=0}^{\infty} \frac{(\theta N)^k}{k!} = \exp[\text{mat}(\vec{\omega})], \quad (25)$$

where  $N$  is the skew-symmetric matrix such that  $N = \text{mat}(\vec{n})$  (31). For infinitesimal rigid-body motions,  $\sin \theta \approx \theta$  and  $\cos \theta \approx 1$ . Hence, Eq. 25 is approximated such that

$$R(\vec{\omega}) \approx I_3 + \theta N = I_3 + \text{mat}(\vec{\omega}). \quad (26)$$

## APPENDIX B: SYMMETRY-CONSTRAINED RIGID-CLUSTER MODEL

Suppose that a symmetric system consists of  $m$  asymmetric units and each asymmetric unit is represented by  $N$  rigid clusters. That is, the whole system is treated as an assembly of  $m \times N$  rigid clusters. Any arbitrary residue  $a$  of cluster  $i$  in asymmetric unit  $j$  can be related to its corresponding residue in the reference asymmetric unit 1 by

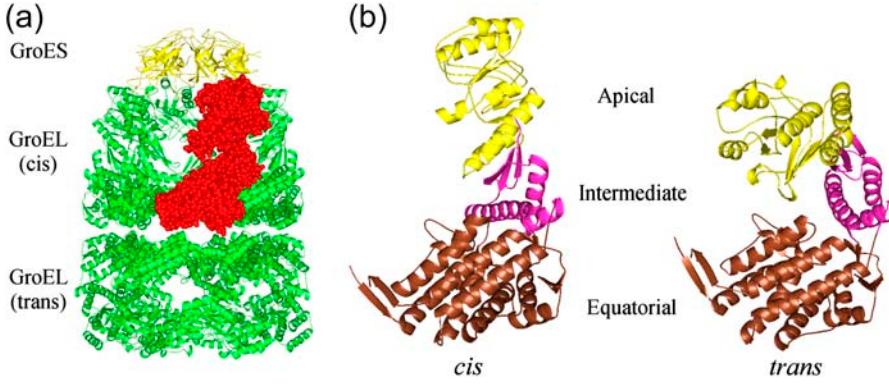
$$\vec{x}_{i,a}^j(t) = R^j \vec{x}_{i,a}^1(t), \quad (27)$$

where  $R^j$  is the rotation matrix from asymmetric unit 1 to asymmetric unit  $j$ . A superscript  $j$  indicates the  $j^{\text{th}}$  asymmetric unit, whereas a subscript  $i$  describes the  $i^{\text{th}}$  cluster in asymmetric unit  $j$ . In particular,  $i$  is 1–14 and  $j$  is 1–60 in the HK97 capsid. From Eq. 6,  $\vec{x}_{i,a}^1$  can be written as

$$\vec{x}_{i,a}^1(t) = R(\vec{\omega}_i^1(t))(\vec{x}_{i,a}^1(0) - \vec{x}_i^1(0)) + \vec{x}_i^1(0) + \vec{v}_i^1(t). \quad (28)$$

Multiplying by  $R^j$  on both sides of the equation yields

$$R^j \vec{x}_{i,a}^1(t) = R^j R(\vec{\omega}_i^1(t))(\vec{x}_{i,a}^1(0) - \vec{x}_i^1(0)) + R^j \vec{x}_i^1(0) + R^j \vec{v}_i^1(t). \quad (29)$$



Compared to Eq. 27, the left-hand side of the above equation is equal to  $\vec{x}_{i,a}^j(t)$  and the right-hand side of it can be expressed in a different way such that

$$\begin{aligned} \vec{x}_{i,a}^j(t) &= R^j R(\vec{\omega}_i^1(t)) R^{j^T} R^j (\vec{x}_{i,a}^1(0) - \vec{x}_i^1(0)) + R^j \vec{x}_i^1(0) + R^j \vec{v}_i^1(t) = R^j R(\vec{\omega}_i^1(t)) R^{j^T} (\vec{x}_{i,a}^j(0) - \vec{x}_i^j(0)) + \vec{x}_i^j(0) + R^j \vec{v}_i^1(t) \\ &= R^j \exp[m\vec{\omega}_i^1(t)] R^{j^T} (\vec{x}_{i,a}^j(0) - \vec{x}_i^j(0)) + \vec{x}_i^j(0) + R^j \vec{v}_i^1(t) = \exp[R^j m\vec{\omega}_i^1(t) R^{j^T}] (\vec{x}_{i,a}^j(0) - \vec{x}_i^j(0)) + \vec{x}_i^j(0) + R^j \vec{v}_i^1(t) \\ &= \exp[m\vec{\omega}_i^1(t)] (\vec{x}_{i,a}^j(0) - \vec{x}_i^j(0)) + \vec{x}_i^j(0) + R^j \vec{v}_i^1(t) = R(R^j \vec{\omega}_i^1(t)) (\vec{x}_{i,a}^j(0) - \vec{x}_i^j(0)) + \vec{x}_i^j(0) + R^j \vec{v}_i^1(t). \end{aligned} \quad (30)$$

When compared to Eq. 6, it is verified that

$$\begin{aligned} \vec{\omega}_i^j(t) &= R^j \vec{\omega}_i^1(t) \\ \vec{v}_i^j(t) &= R^j \vec{v}_i^1(t). \end{aligned} \quad (31)$$

From this fact, we can derive the relationship between  $\vec{\delta}_i^j(t)$  and  $\vec{\delta}_i^1(t)$  such that

$$\begin{aligned} \vec{\delta}_i^j(t) &= \begin{pmatrix} R^j \vec{v}_i^1(t) \\ R^j \vec{\omega}_i^1(t) \end{pmatrix} = \underbrace{\begin{pmatrix} R^j & 0_3 \\ 0_3 & R^j \end{pmatrix}}_{\tilde{R}^j} \begin{pmatrix} \vec{v}_i^1(t) \\ \vec{\omega}_i^1(t) \end{pmatrix} \\ &= \tilde{R}^j \vec{\delta}_i^1(t), \quad \text{for } i = 1, \dots, N. \end{aligned} \quad (32)$$

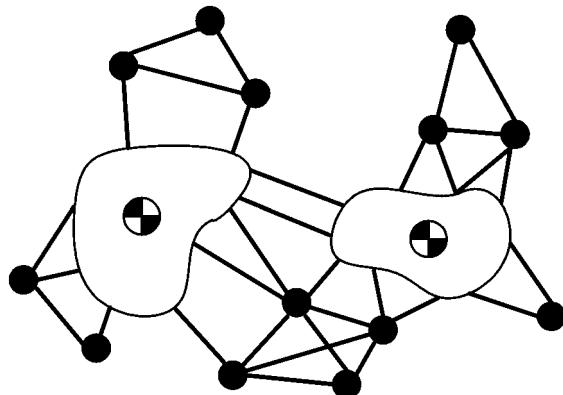


FIGURE 10 Schematic of the hybrid elastic network model. Rigid clusters and point masses are considered together in this model. Solid lines represent interconnections between objects with linear springs.

FIGURE 9 A cartoon of the GroEL-GroES chaperonin complex. (a) An asymmetric unit of GroEL is illustrated with a space-filling representation. Seven identical subunits comprise a GroEL ring structure. (b) The *cis* and *trans* conformations of GroEL are displayed. During the conformational transition between *cis* and *trans*, the equatorial (brown) domain acts like a rigid body. However, the intermediate (magenta) and apical (yellow) domains are flexible.

Let the whole displacement vector  $\vec{\delta}(t)$  be

$$\vec{\delta}(t) = [\vec{\Delta}^1(t)^T, \dots, \vec{\Delta}^m(t)^T]^T \in \mathcal{R}^{6mN}, \quad (33)$$

where  $\vec{\Delta}_i^j(t) = [\vec{\delta}_1^j(t)^T, \dots, \vec{\delta}_N^j(t)^T]^T \in \mathcal{R}^{6N}$  ( $j = 1, \dots, m$ ). Using the relationship in Eq. 32,  $\vec{\Delta}_i^j(t)$  can be related to  $\vec{\Delta}_i^1(t)$  as

$$\vec{\Delta}_i^j(t) = \hat{R}^j \vec{\Delta}_i^1(t), \quad (34)$$

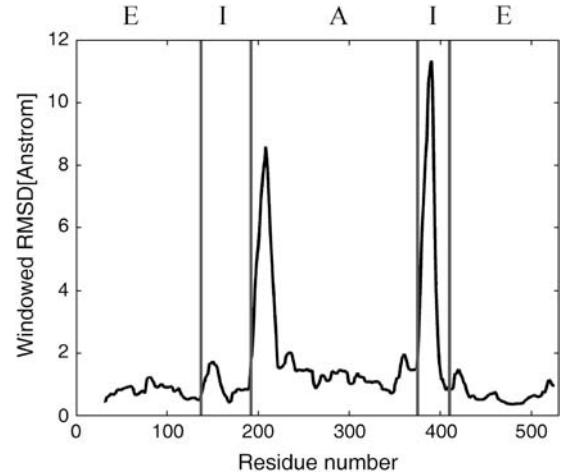


FIGURE 11 The windowed RMSD between *cis* and *trans* conformations of GroEL. Values shown are for 30 consecutive residues per window. Three domains are distinguished from one another by vertical lines. *A*, *I*, and *E* are the apical, intermediate, and equatorial domains, respectively. The two highest peaks are observed near the borders between intermediate and apical domains. We also find other small peaks near the interface between equatorial and intermediate domains. It shows these regions are much more flexible than the core regions of each domain. In the Hybrid ENI model, we keep those flexible regions as point masses, whereas rigid domains are treated as rigid clusters based on these windowed RMSD values.

**TABLE 4** Hybrid modeling of the GroEL ring structure

Residue number	Domain*	Cluster/point mass
Ala <sup>2</sup> ~ Asp <sup>64</sup>	Equatorial	Cluster 1 (1.0)
Lys <sup>65</sup> ~ Leu <sup>134</sup>	Equatorial	Cluster 2 (0.8)
Ser <sup>135</sup> ~ Asp <sup>140</sup>	Equatorial ↔ Intermediate	Point mass 1–6
Ser <sup>141</sup> ~ Lys <sup>168</sup>	Intermediate	Cluster 3 (0.4)
Val <sup>169</sup> ~ Asp <sup>185</sup>	Intermediate	Cluster 4 (1.0)
Glu <sup>186</sup> ~ Gly <sup>192</sup>	Intermediate ↔ Apical	Point mass 7–13
Met <sup>193</sup> ~ Leu <sup>222</sup>	Apical	Cluster 5 (1.5)
Ala <sup>223</sup> ~ Leu <sup>262</sup>	Apical	Cluster 6 (1.4)
Val <sup>263</sup> ~ Ile <sup>301</sup>	Apical	Cluster 7 (1.3)
Ser <sup>302</sup> ~ Gly <sup>337</sup>	Apical	Cluster 8 (1.1)
Glu <sup>338</sup> ~ Leu <sup>372</sup>	Apical	Cluster 9 (1.4)
Ala <sup>373</sup> ~ Thr <sup>385</sup>	Apical ↔ Intermediate	Point mass 14–26
Glu <sup>386</sup> ~ Glu <sup>408</sup>	Intermediate	Cluster 10 (0.7)
Glu <sup>409</sup> ~ Gly <sup>414</sup>	Intermediate ↔ Equatorial	Point mass 27–32
Gly <sup>415</sup> ~ Pro <sup>525</sup>	Equatorial	Cluster 11 (1.2)

RMSD between two corresponding rigid clusters is displayed in parenthesis in units of Å.

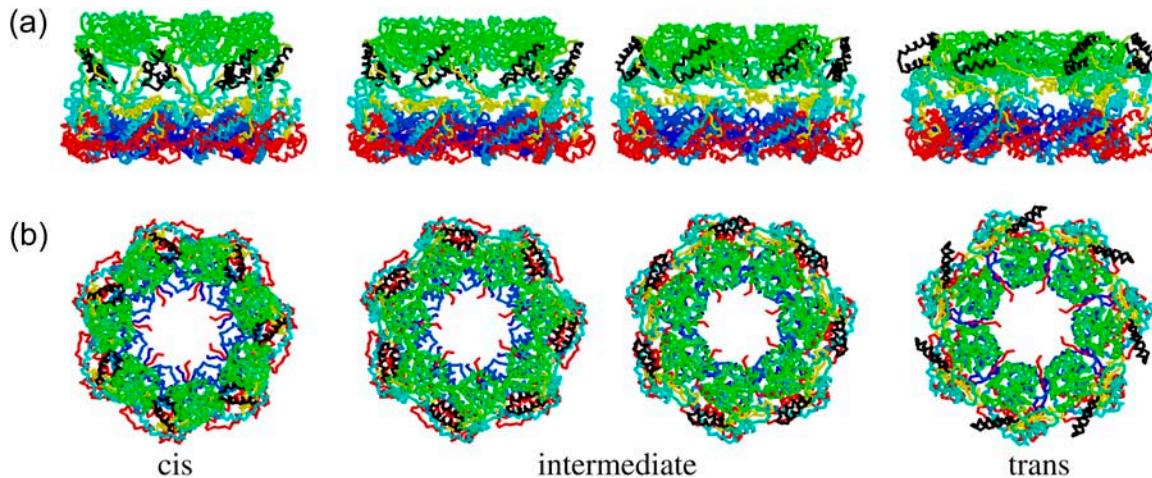
\*The symbol, ↔, indicates an interface between two domains listed. These flexible regions are modeled as point masses.

where

$$\hat{R}^j = \begin{bmatrix} \tilde{R}^j & 0_6 & \cdots & 0_6 \\ 0_6 & \tilde{R}^j & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_6 \\ 0_6 & \cdots & 0_6 & \tilde{R}^j \end{bmatrix} \in \mathcal{R}^{6N \times 6N}. \quad (35)$$

Combining Eq. 34 with Eq. 18 reduces the DOF of a system by the factor  $m$  so that

$$C(\vec{\Delta}^1(t)) = \frac{m}{2} \vec{\Delta}^1(t)^T \Gamma' \vec{\Delta}^1(t) + \frac{m}{2} \vec{\gamma}' \vec{\Delta}^1(t) + B, \quad (36)$$



**FIGURE 12** The transitional pathway of the GroEL ring structure in hybrid ENI. (a) Hybrid ENI is applied to generate a conformational transition from *cis* to *trans* conformations of an asymmetric unit of GroEL with symmetry constraints. Each intermediate conformation of the single ring structure is constructed by juxtaposing seven copies of a calculated asymmetric unit conformation. (b) A perpendicular view of the GroEL structure. During the transition from *cis* to *trans*, two helices (black) composed of Glu<sup>339</sup> through Ile<sup>353</sup> and Asp<sup>361</sup> through Lys<sup>371</sup> in Cluster 9 swing outward, resulting in the collapse of the apical domain by 13 Å along the rotation axis.

**TABLE 5** DOF of the GroEL ring structure in various parameterizations

Method*	Asymmetric		Residues	DOF	Total DOF	Fraction
	units					
[CG]	7	524	3	11,004	100	
[CG]+[SC]	1	524	3	1572	14.29	
[CG]+[SC]+[HB]	1	32	3	162	1.47	
		11 clusters	6			

\*All the notations follow those in Table 1.

where

$$\Gamma' = \sum_{j=1}^m \hat{R}^{1T} \Gamma_1 \hat{R}^j, \\ \vec{\gamma}' = \vec{\gamma}_1, \quad (37)$$

where  $\Gamma'$  is a  $6N \times 6N$  matrix and  $\vec{\gamma}'$  is a  $1 \times 6N$  row vector. After finding  $\vec{\Delta}'(t)$ , which minimizes the cost function in Eq. 36, the new conformation for the whole system can be constructed by copying the calculated reference unit  $m$  times.

### APPENDIX C: HYBRID MODEL

In this model, a point mass can be treated as a rigid cluster in which the center of mass is the position of the point mass itself and the orientation is not defined. Hence,  $H_{i,a}$  and  $\vec{\delta}_i(t)$  in Eq. 8 can be redefined as

$$H_{i,a} = [I_3 \ 0_3] \in \mathcal{R}^{3 \times 6} \\ \vec{\delta}_i(t) = [\vec{v}_i^T(t) \ \vec{0}_{1 \times 3}]^T \in \mathcal{R}^6, \quad (38)$$

where  $\vec{d}_{i,a}(0) - \vec{x}_i(0) = \vec{0}$  and  $\vec{\omega}_i^T(t) = \vec{0}_{1 \times 3}$ . Since the displacement vector of rigid cluster  $i$ , denoted as  $\vec{\delta}_i(t)$ , has a trivial (zero) part corresponding to orientation in the case of a point mass, the cost function of rigid-cluster ENI in Eq. 18 can be described in a reduced form with respect to only nontrivial terms.

For example, consider a system that consists of one point mass and one rigid cluster. If the point mass is treated as rigid cluster 1, then  $\vec{\delta}_1(t) = [\vec{v}_1^T(t) \ \vec{0}_{1 \times 3}]^T$ . On the other hand, the displacement of the second

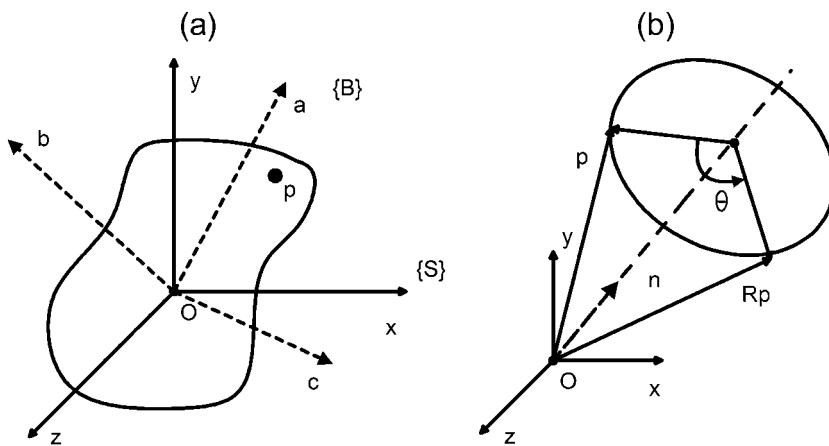


FIGURE 13 Rotation of a rigid body. (a) The body-fixed frame  $\{B\}$  is displayed as the dotted arrows. The point  $P$  in space can be represented by the coordinates of either space-fixed frame  $\{S\}$  or body-fixed frame  $\{B\}$ . (b) This is an axis-angle parameterization. The axis and angle of rotation are denoted by  $n$  and  $\theta$ , respectively.

(real) rigid cluster is denoted as  $\vec{\delta}_2(t) = [\vec{v}_2^T(t) \ \vec{\omega}_2^T(t)]^T$ . Using this notation, the cost function is written in component form as

$$C = \frac{1}{2} (\vec{v}_1^T \ \vec{0}_{1 \times 3} \ \vec{v}_2^T \ \vec{\omega}_2^T) \begin{bmatrix} \Gamma_{1,1}^a & \Gamma_{1,1}^b & \Gamma_{1,2}^a & \Gamma_{1,2}^b \\ \Gamma_{1,1}^c & \Gamma_{1,1}^d & \Gamma_{1,2}^c & \Gamma_{1,2}^d \\ \Gamma_{2,1}^a & \Gamma_{2,1}^b & \Gamma_{2,2}^a & \Gamma_{2,2}^b \\ \Gamma_{2,1}^c & \Gamma_{2,1}^d & \Gamma_{2,2}^c & \Gamma_{2,2}^d \end{bmatrix} \begin{pmatrix} \vec{v}_1 \\ \vec{0}_{3 \times 1} \\ \vec{v}_2 \\ \vec{\omega}_2 \end{pmatrix} + \frac{1}{2} (\vec{\gamma}_1^a \ \vec{\gamma}_1^b \ \vec{\gamma}_2^a \ \vec{\gamma}_2^b) \begin{pmatrix} \vec{v}_1 \\ \vec{0}_{3 \times 1} \\ \vec{v}_2 \\ \vec{\omega}_2 \end{pmatrix} + B. \quad (39)$$

After deleting rows and columns that multiply the zero vector (i.e., matrix elements written as bold letters), the reduced cost function for hybrid ENI is obtained as

$$C = \frac{1}{2} (\vec{v}_1^T \ \vec{v}_2^T \ \vec{\omega}_2^T) \begin{bmatrix} \Gamma_{1,1}^a & \Gamma_{1,2}^a & \Gamma_{1,2}^b \\ \Gamma_{2,1}^a & \Gamma_{2,2}^a & \Gamma_{2,2}^b \\ \Gamma_{2,1}^c & \Gamma_{2,2}^c & \Gamma_{2,2}^d \end{bmatrix} \begin{pmatrix} \vec{v}_1 \\ \vec{v}_2 \\ \vec{\omega}_2 \end{pmatrix} + \frac{1}{2} (\vec{\gamma}_1^a \ \vec{\gamma}_2^a \ \vec{\gamma}_2^b) \begin{pmatrix} \vec{v}_1 \\ \vec{v}_2 \\ \vec{\omega}_2 \end{pmatrix} + B. \quad (40)$$

## REFERENCES

- Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The protein data bank. *Nucleic Acids Res.* 28:235–242.
- Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.
- Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–515.
- Kim, M. K., R. L. Jernigan, and G. S. Chirikjian. 2003. An elastic network model of HK97 capsid maturation. *J. Struct. Biol.* 143:107–117.
- Tama, F., F. X. Gadea, O. Marques, and Y. H. Sanejouand. 2000. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins.* 41:1–7.
- Tama, F., W. Wriggers, and C. L. Brooks. 2002. Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. *J. Mol. Biol.* 321:297–305.
- Mouawad, L., and D. Perahia. 1996. Motions in hemoglobin studied by normal mode analysis and energy minimization: evidence for the existence of tertiary *T*-like, quaternary *R*-like intermediate structures. *J. Mol. Biol.* 258:393–410.
- Xu, C., D. Tobi, and I. Bahar. 2003. Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin *T*  $\leftrightarrow$  *R2* transition. *J. Mol. Biol.* 333:153–168.
- Tama, F., and C. L. Brooks. 2002. The mechanism and pathway of pH-induced swelling in Cowpea Chlorotic Mottle Virus. *J. Mol. Biol.* 318:733–747.
- Miyashita, O., J. N. Onuchic, and P. G. Wolynes. 2003. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc. Natl. Acad. Sci. USA.* 100:12570–12575.
- Marques, O., and Y. H. Sanejouand. 1995. Hinge-bending motion in citrate synthase arising from normal mode calculations. *Proteins.* 23:557–560.
- Tama, F., and Y. H. Sanejouand. 2001. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 14:1–6.
- Delarue, M., and Y. H. Sanejouand. 2002. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. *J. Mol. Biol.* 320:1011–1024.
- Valadie, H., J. J. Lacapere, Y. H. Sanejouand, and C. Etchebest. 2003. Dynamical properties of the Mscl of *Escherichia coli*: a normal mode analysis. *J. Mol. Biol.* 332:657–674.
- Kim, M. K., G. S. Chirikjian, and R. L. Jernigan. 2002. Elastic models of conformational transitions in macromolecules. *J. Mol. Graph. Model.* 21:151–160.
- Kim, M. K., R. L. Jernigan, and G. S. Chirikjian. 2002. Efficient generation of feasible pathways for protein conformational transitions. *Biophys. J.* 83:1620–1630.
- Kim, M. K., W. Li, B. A. Shapiro, and G. S. Chirikjian. 2003. A comparison between elastic network interpolation and MD simulation of 16S ribosomal RNA. *J. Biomol. Struct. Dyn.* 21:395–405.
- Gerstein, M., and W. Krebs. 1998. A database of macromolecular motions. *Nucleic Acids Res.* 26:4280–4290.
- Conway, J. F., W. R. Wikoff, N. Cheng, R. L. Duda, R. W. Hendrix, J. E. Johnson, and A. C. Steven. 2001. Virus maturation involving large subunit rotations and local refolding. *Science.* 292:744–748.
- Doruker, P., R. L. Jernigan, and I. Bahar. 2002. Dynamics of large proteins through hierarchical levels of coarse-grained structures. *J. Comput. Chem.* 23:119–127.

21. Schuyler, A. D., and G. S. Chirikjian. 2004. Normal mode analysis of proteins: a comparison of rigid cluster modes with  $C^\alpha$  coarse graining. *J. Mol. Graph. Model.* 22:183–193.
22. Ming, D., Y. Kong, Y. Wu, and J. Ma. 2003. Substructure synthesis method for simulating large molecular complexes. *Proc. Natl. Acad. Sci. USA*. 100:104–109.
23. Conway, J. F., R. L. Duda, N. Cheng, R. W. Hendrix, and A. C. Steven. 1995. Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system. *J. Mol. Biol.* 253:86–99.
24. Rader, A. J., D. H. Vlad, and I. Bahar. 2005. Maturation dynamics of bacteriophage HK97 capsid. *Structure*. 13:413–421.
25. Rye, H. S., S. G. Burston, W. A. Fenton, J. M. Beechem, Z. Xu, P. B. Sigler, and A. L. Horwich. 1997. Distinct actions of *cis* and *trans* ATP within the double ring of the chaperonin GroEL. *Nature*. 388:792–798.
26. Sigler, P. B., Z. Xu, H. S. Rye, S. G. Burston, W. A. Fenton, and A. L. Horwich. 1998. Structure and function in GroEL-mediated protein folding. *Annu. Rev. Biochem.* 67:581–608.
27. Xu, Z., A. L. Horwich, and P. B. Sigler. 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)<sub>7</sub> chaperonin complex. *Nature*. 388:741–750.
28. Xu, Z., and P. B. Sigler. 1998. GroEL/GroES: structure and function of a two-stroke folding machine. *J. Struct. Biol.* 124:129–141.
29. Murray, R. M., Z. Li, and S. S. Sastry. 1994. A Mathematical Introduction to Robotic Manipulation. CRC, Boca Raton, FL.
30. Spong, M. W., and M. Vidyasagar. 1989. Robot Dynamics and Control. John Wiley and Sons, New York.
31. Chirikjian, G. S., and A. B. Kyatkin. 2001. Engineering Applications of Noncommutative Harmonic Analysis. CRC, Boca Raton, FL.