# Efficient prediction of protein conformational pathways based on the hybrid elastic network model

Sangjae Seo [a,1], Yunho Jang [b,1], Pengfei Qian [c], Wing Kam Liu [d], Jae-Boong Choi [a,c], Byeong Soo Lim [c], Moon Ki Kim [a,c,*]

[a] SKKU Advanced Institute of Nanotechnology, Sungkyunkwan University, Suwon 440-746, Republic of Korea
[b] Department of Mechanical and Industrial Engineering, University of Massachusetts, Amherst, MA 01003, USA
[c] School of Mechanical Engineering, Sungkyunkwan University, Suwon 440-746, Republic of Korea
[d] Department of Mechanical Engineering, Northwestern University, Evanston, IL 60208, USA

## ABSTRACT

Various computational models have gained immense attention by analyzing the dynamic characteristics of proteins. Several models have achieved recognition by fulfilling either theoretical or experimental predictions. Nonetheless, each method possesses limitations, mostly in computational outlay and physical reality. These limitations remind us that a new model or paradigm should advance theoretical principles to elucidate more precisely the biological functions of a protein and should increase computational efficiency. With these critical caveats, we have developed a new computational tool that satisfies both physical reality and computational efficiency. In the proposed hybrid elastic network model (HENM), a protein structure is represented as a mixture of rigid clusters and point masses that are connected with linear springs. Harmonic analyses based on the HENM have been performed to generate normal modes and conformational pathways. The results of the hybrid normal mode analyses give new physical insight to the 70S ribosome. The feasibility of the conformational pathways of hybrid elastic network interpolation (HENI) was quantitatively evaluated by comparing three different overlap values proposed in this paper. A remarkable observation is that the obtained mode shapes and conformational pathways are consistent with each other. Our timing results show that HENM has some advantage in computational efficiency over a coarse-grained model, especially for large proteins, even though it takes longer to construct the HENM. Consequently, the proposed HENM will be one of the best alternatives to the conventional coarse-grained ENMs and all-atom based methods (such as molecular dynamics) without loss of physical reality.

© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

The great developments of biological experiments have led to a rapid increase in the solved structures of macromolecules [1]. Also the development of computational models has enabled us to relate static structures of macromolecules to its functions by predicting conformational change [2,3]. Molecular dynamic (MD) simulation [4,5] is one of the most common computational tools utilized to elucidate macromolecular motions at the atomic level. However, although MD simulation provides time-dependent behaviors of a molecular system, including the effect of the surrounding solvent, current computer power limits the investigation of the relevant dynamics solely to an early stage [6]. In attempt to reduce computational cost, various coarse-grained (CG) approaches have been developed [7]. Among these CG approaches, the elastic network model (ENM) having a single-parameter or simplified potential function has proved to be a useful tool for investigating global dynamics of proteins [8,9]. The ENM represents the system as multiple degrees-of-freedom (DOF) linear mass-spring systems. For instance, a protein structure is modeled as a spring network among representative Cα atoms.

Early studies have shown that the normal mode analysis (NMA) based on ENM can effectively predict low frequency motions, which are relevant to the functional motion [10–13]. NMA basically describes thermal fluctuations of a macromolecule around its equilibrium energy. By combining several lowest normal modes, one can reproduce the early stage of the functional collective motions of the given structure. The results of NMA based on ENM correlate well with MD simulation results and experimentally observed

* Corresponding author at: SKKU Advanced Institute of Nanotechnology, Sungkyunkwan University, Suwon 440-746, Republic of Korea.
Tel.: +82 31 299 4840; fax: +82 31 290 5889.
*E-mail address:* mkkim@me.skku.ac.kr (M.K. Kim).
[1] These authors contributed equally to this work.

conformational changes in both the magnitude and direction of the motion [14–19]. Another advance in NMA technique is the consideration of rigidity of protein. The rigidity and flexibility are important aspects to account for the protein motions. The large-scale movements in macromolecules, such as domain and hinge-like motion, are highly engaged in relative motion among those rigid domains [20,21]. A rigid domain motion can be analyzed by rigid body motion; translation and rotation. The first attempt was the rotations–translations of blocks approach (RTB) proposed by Sanejouand group [22]. RTB considered consecutive amino-acid residues as a pure rigid body block and six DOF are projected onto the Hessian matrix. The block normal mode (BNM) method [23] was proposed to reduce the storage size in diagonalization step of RTB method by the projection matrix. Recently, a new domain decomposition method for RTB, called density-cluster RTB was developed [24]. The cluster NMA (cNMA) [25,26] was developed to overcome several limitations in RTB and BNM method. cNMA can directly convert Cartesian coordinates into rigid clusters without any projection, vice versa, whereas RTB and BNM methods transform the full Hessian matrix into a reduced subspace. One incomparable advantage of rigid-clustering is the dramatic reduction of computational cost. The computational complexity of the conventional NMA is the order of $n^3$ and the storage of a Hessian matrix is proportional to $n^2$. However, in case of rigid cluster based NMAs, the computational cost depends on the number of rigid domains into which the system is structurally decomposed [22].

In addition to the study of the intrinsic motion of macromolecules, investigating the transition pathway is also important to elucidate the relationship between structure and function. Firstly, elastic network interpolation (ENI) iteratively obtains the intermediate conformations by interpolating two corresponding sets of interatomic distances in each elastic network [27,28]. By comparing the transition pathway with molecular dynamics simulation result, ENI is also able to compromise between a physically oversimplified linear interpolation method and computationally expensive MD simulation [29]. Also Feng et al. simulated the pathway of adenylate kinase (AK) in atomic details by incorporating CHARMM force field to ENI [30]. Furthermore, various methods based on the free energy surface have been introduced. Plastic network model (PNM) generated pathway of AK based on the free energy surface of two ENM potentials [30] and mixed-ENM extended the free energy based technique by mixing the Boltzmann-weighted Go potentials [31]. To improve the accuracy of mixed-ENM, Tekpinar and Zheng generalized mixed-ENM by searching the saddle point of double-well potential functions [32]. Even though earlier studies have successfully predicted transition pathways, there has been no attempt to predict transition pathway holding rigidity information.

In order to preserve the rigidity and flexibility of protein and consider both for rigid clusters and point masses, there is clearly a need for methods that generally treat the macromolecules for NMA calculation and pathway generation. In this study, the hybrid ENM (HENM) is introduced for both the hybrid NMA (HNMA) and the hybrid ENI (HENI). The HENM consists of point-mass parts and rigid-cluster parts that correspond to flexible regions and rigid domains, respectively. Thus, the HENM was proposed as a tradeoff between Cα coarse graining and rigid-cluster modeling in which rigid clusters and point masses are linked to one another with linear springs [12,13,33]. HENM's computational advantage over the other methods is also discussed. The comparison of overlap of the computed pathways between the two end conformations with normal modes from each end conformation shows excellent consistency. Moreover, a comparison of the source code run time proves that the HENI method provides computational efficiency for large proteins (residue number > 300). Ribosome was chosen as the representative to test reliability of the proposed HENM as a simulator of protein conformational transitions. HNMA was applied to
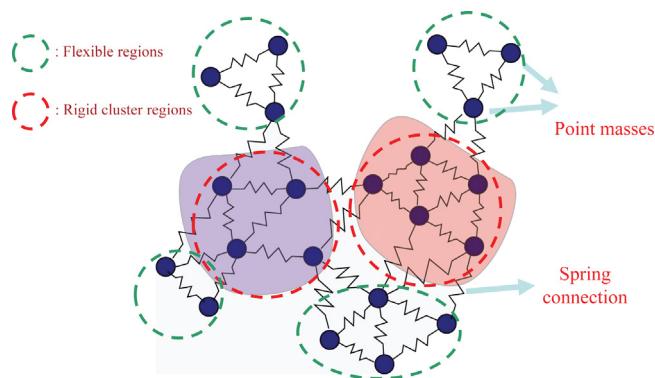


**Fig. 1.** Overview of HENM. Interactions between atoms are represented as a set of elastic springs. Rigid regions are described with rigid body motion, while the other flexible regions are represented as point masses.

the 70S ribosome and the first six normal modes showed its major dynamic motions, which facilitate the translocation of the tRNA inside the 70S ribosome. The generated translocation pathway of ribosome correlates well with the current understanding in the literature. Consequently, the proposed HENM can not only reduce computational cost but also generate reliable dynamic motions of large macromolecules without loss of generality.

## 2. Methods

In this section, we derive a full mathematical description of HENM. The HENM concept is based on the fact that most conformational changes in protein dynamics can be resolved into hinge and shear motions that are associated with the collective behavior of atoms in the rigid domains. An overview of HENM (Fig. 1) is as follows: (i) all representative atoms are networked with linear springs in a conventional way; (ii) rigid regions (e.g., secondary structures) are re-modeled as a set of rigid clusters; and (iii) the other regions, which are flexible (e.g., hinges and loops) are represented as point masses. The mathematical description of HENM is derived based on the potential energy between point to point mass, cluster to cluster, and point to cluster.

### 2.1. Determination of rigid clusters

Although various rigidity algorithms and theories [34–37] have been introduced, there is still no unique way to define rigid clusters and point masses with given structures. The secondary structures, such as alpha helices or beta sheets are usually assumed to be rigid domains. For the efficient and systematic detection of rigid regions, computational approaches are widely used. The graph-theoretical approach named FIRST identifies flexible and rigid regions of the proteins based on the mathematical rigidity theory [38]. The windowed root mean square deviation (WRMSD) method was developed from the conventional static comparison between two end conformations [39]. While the conventional root mean square deviation (RMSD) method calculates the RMSD value over the whole structure as a single representative metric to quantify how much two given structures are close to each other, the WRMSD method continuously computes the RMSD value of the local region captured by a finite window size along the backbone. The lower WRMSD value indicates that the corresponding region moves collectively like a rigid body during the conformational change. It also suggests that the window size should be small enough to avoid missing functionally important (even locally flexible) structures. For example, in the case of adenylate kinase (AK) presented in Fig. 2, there are four rigid clusters marked as green boxes when the window size is set to be 20 residues. After passing the local

**Table 1**
The list of example proteins selected to validate HENM.

| Protein | Open structure | Closed structure | Total residue number | Total DOF of Cα ENM | HENM | | |
|---|---|---|---|---|---|---|---|
| | | | | | Number of rigid clusters | Number of point masses | Total DOF of HENM |
| HIV-1 protease | 1HHP | 1AJX | 99 | 297 | 4 | 52 | 180 |
| Calmodulin | 1CFD | 1CFC | 148 | 444 | 5 | 49 | 177 |
| Adenylate kinase | 4AKE | 1AKE | 214 | 642 | 4 | 134 | 426 |
| Tyrosine phosphatase | 1YPT | 1YTS | 278 | 834 | 3 | 125 | 393 |
| Lactoferrin | 1LFH | 1LFG | 691 | 2073 | 5 | 40 | 150 |
| Ca2 ATPase | 1SU4 | 1KJU | 994 | 2982 | 9 | 359 | 1131 |

minima (red circles in Fig. 2), each rigid cluster is followed by the appearance of either a new rigid cluster or a flexible region. Likewise, the rigid domains and flexible regions for all other example proteins in this study are determined by WRMSD and summarized in Table 1. One can also perform rigid clustering by adopting the structural characteristics from previous literature or experimental observations.

### 2.2. Mathematical derivation of hybrid elastic network model

Once the rigid domains and flexible regions in a protein structure are predetermined, one can construct its HENM. Assuming that a given HENM consists of $M$ point masses and $N$ rigid clusters, the following nomenclature was used to derive a mathematical description of HENM for proteins that contain both rigid domains and flexible regions.

$M$: total number of point masses
$N$: total number of rigid clusters
$n(b)$: total number of point masses in cluster $b$
$a,a'$: point mass index
$b,b'$: rigid cluster index
$d$: cutoff distance.
$i,j$: point masses index in a rigid cluster.
$\|\cdot\|$: length of a vector.
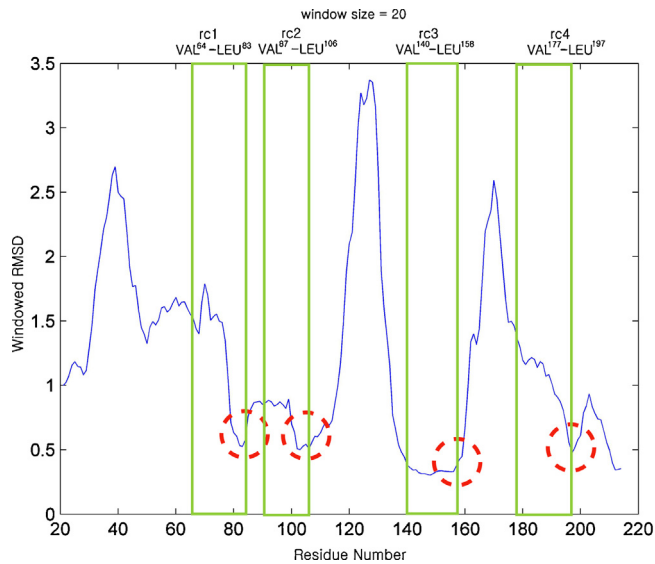


**Fig. 2.** Rigid clustering of AK by WRMSD. The window size is set to be 20 residues in this test. There is no unique way to determine the number of rigid clusters and their locations. However, given the WRMSD plot, one can perceptively identify several local minima indicated by red circles. In this case, we only take them from the WRMSD values under 0.5 Å. Then the corresponding residues within the given window size (green box region) are treated as a rigid body. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

$k_{a,a'}$: spring constant between residue $a$ and residue $a'$.
$k_{b,i,b',j}$: spring constant between residue $b$ of cluster $i$ and residue $b'$ of cluster $j$.
$k_{a,b,i}$: spring constant between residue $a$ and $b$ of cluster $i$.
$\vec{x}_a^p(t) \in \Re^3$: position of point mass $A$ at time $t$.
$\vec{x}_{b,i}^r(t) \in \Re^3$: position of the $i$th atom in rigid cluster $b$.
$\vec{x}_b^r(t) \in \Re^3$: position of the center of mass of rigid cluster $b$ at time $t$.

Here, the superscripts $p$ and $r$ represent point masses and rigid clusters, respectively.

The displacement for hybrid system is defined as

$$\vec{\delta}^h = [\vec{\delta}_1^{p^T}, \ldots, \vec{\delta}_M^{p^T}, \vec{\delta}_1^{r^T}, \ldots, \vec{\delta}_N^{r^T}]^T \in \Re^{(3M+6N)}, \tag{1}$$

where $\vec{\delta}_a^p(t) = \vec{x}_a^p(t) - \vec{x}_a^p(0)$ is the translational displacement for point mass $a$ and $\vec{\delta}_b^r(t) = [\vec{\chi}_b^{r^T}, \vec{\omega}_b^T]^T$ is the displacement of rigid cluster $b$. It should be noted that the displacement of rigid cluster is defined as a combination of translational displacement $\vec{\chi}_a^r(t) = \vec{x}_b^r(t) - \vec{x}_b^r(0)$ and the rotation vector $\vec{\omega}_b^T$, where $\vec{x}_b^r(t)$ is the position of the center of mass of cluster $b$ at time $t$.

Also, the position of residue $i$ within cluster $b$ at time $t$ is denoted by

$$\vec{x}_{b,i}^r(t) = R(\vec{\omega}_b(t))(\vec{x}_{b,i}^r(0) - \vec{x}_b^r(0)) + \vec{x}_b^r(0) + \vec{\chi}_b^r(t), \tag{2}$$

where the rotation is defined as

$$R(\vec{\omega}_b(t)) = \exp[mat(\vec{\omega}_b(t))]. \tag{3}$$

The '*mat*' operator converts a $3 \times 1$ vector into a skew-symmetric matrix such that

$$mat(\vec{x}) = \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & x_1 \\ -x_2 & x_1 & 0 \end{pmatrix}. \tag{4}$$

Now, the inertia matrix is derived from the quadratic expression of the kinetic energy. The total kinetic energy in a hybrid system is defined by

$$T(t) = \frac{1}{2}\dot{\vec{\delta}}^h(t)^T M \dot{\vec{\delta}}^h(t) \tag{5}$$

and

$$M = \begin{bmatrix} M^p & \tilde{0} \\ \tilde{0} & \tilde{0} \end{bmatrix} + \begin{bmatrix} \tilde{0} & \tilde{0} \\ \tilde{0} & M^r \end{bmatrix}, \tag{6}$$

where $M^p$ is a simple diagonal matrix of point masses, while $M^r$ is an inertia matrix of rigid clusters.

The total potential energy can be computed by the sum of three different potentials such as

$$V = V^p + V^r + V^h, \tag{7}$$

where $V^p$, $V^r$, and $V^h$ are potential energy for point to point interactions, cluster to cluster interactions, and cluster to point interactions, respectively. The stiffness matrix $K$ for the whole network is constructed such that

$$
\begin{aligned}
V &= \frac{1}{2}\vec{\delta}^{p\mathrm{T}}K^p\vec{\delta}^p + \frac{1}{2}\vec{\delta}^{r\mathrm{T}}K^r\vec{\delta}^r + \frac{1}{2}\vec{\delta}^{h\mathrm{T}}K^h\vec{\delta}^h \\
&= \frac{1}{2}[\,\vec{\delta}^{p\mathrm{T}}\ \ \vec{\delta}^{r\mathrm{T}}\,]\begin{bmatrix} K^p & \tilde{0} \\ \tilde{0} & \tilde{0} \end{bmatrix}\begin{bmatrix} \vec{\delta}^p \\ \vec{\delta}^r \end{bmatrix} \\
&\quad + \frac{1}{2}[\,\vec{\delta}^{p\mathrm{T}}\ \ \vec{\delta}^{r\mathrm{T}}\,]\begin{bmatrix} \tilde{0} & \tilde{0} \\ \tilde{0} & K^r \end{bmatrix}\begin{bmatrix} \vec{\delta}^p \\ \vec{\delta}^r \end{bmatrix} + \frac{1}{2}\vec{\delta}^{h\mathrm{T}}K^h\vec{\delta}^h \\
&= \frac{1}{2}\vec{\delta}^{h\mathrm{T}}\underbrace{\left( K^h + \begin{bmatrix} K^p & \tilde{0} \\ \tilde{0} & K^r \end{bmatrix} \right)}_{K}\vec{\delta}^h,
\end{aligned}
\tag{8}
$$

where $K^p$ is $3M \times 3M$, $K^r$ is $6N \times 6N$, and $K^h$ is $(3M+6N) \times (3M+6N)$. For more details on both mass matrix $M^h$ and stiffness matrix $K$, see Appendix A.

Given a quadratic form of kinetic and potential energies for a multiple DOF conservative system which consists of $M$ point masses, $N$ rigid clusters, and generalized coordinates $\vec{\delta}$, we can derive equation of motion (EOM) using Lagrange's equations such that

$$
\frac{d}{dt}\left( \frac{\partial L}{\partial \dot{\vec{\delta}}} \right) - \frac{\partial L}{\partial \vec{\delta}} = \vec{0},
\tag{9}
$$

where $L = T - V$. Finally we obtain the EOM for a hybrid CG protein model as

$$
M\ddot{\vec{\delta}} + K\vec{\delta} = \vec{0}.
\tag{10}
$$

### 2.3. Hybrid normal mode analysis

The solution of Eq. (10) can be obtained by the following steps. First, multiplication of $M^{-1/2}$ on both sides of Eq. (10) yields

$$
M^{1/2}\ddot{\vec{\delta}} + M^{-1/2}K\vec{\delta} = \vec{0}.
\tag{11}
$$

By letting $y = M^{1/2}\vec{\delta}$, one can get a mass weighted stiffness matrix $\bar{A}$ such as

$$
\ddot{y} + \underbrace{M^{-1/2}KM^{-1/2}}_{\bar{A}=\bar{A}^{\mathrm{T}}}y = \vec{0}.
\tag{12}
$$

Then, the eigenpairs of $\bar{A}$ which satisfy $\bar{A}\bar{V} = \overline{VD}$ can be calculated. The columns of $\bar{V}$ are the eigenvectors of $\bar{A}$ and the diagonal matrix $\bar{D}$ has the corresponding eigenvalues. The column vectors of $V = M^{-1/2}\bar{V}$ are the mode shapes transformed back into the original $\vec{\delta}$ coordinates. Now, $\vec{\delta}$ coordinates are orthogonal to each other. However, it is only for flexible regions for which $\vec{\delta}$ coordinates represent Cα displacements. For rigid domains, $\vec{\delta}$ coordinates represent mass-center displacement and orientation vector. So these $\vec{\delta}$ coordinates have to be transformed back to Cα displacements. After the transformation and projection step (Eq. (6)), corresponding Cα displacements are no longer orthonormal to one another. To conveniently use these results, we utilize Gram–Schmidt orthogonal process in which the set of orthonormal mode shapes $\bar{V}_c = \{\vec{v}_k^c\}$ are obtained by iteratively applying

$$
\vec{v}_k^c = \frac{\vec{v}_k^p - \sum_{l=1}^{k-1}(\vec{v}_k^p\vec{v}_l^c)\vec{v}_l^p}{||\vec{v}_k^p - \sum_{l=1}^{k-1}(\vec{v}_k^p\vec{v}_l^c)\vec{v}_l^p||}.
\tag{13}
$$

### 2.4. Hybrid elastic network interpolation

A cost function is proposed to calculate incremental conformational change of a hybrid system composed of $N$ rigid clusters and $M$ point masses such that

$$
\begin{aligned}
C &= C_{a,a'} + C_{b,b'} + C_{a,b} \\
&= \frac{1}{2}\sum_{a=1}^{M-1}\sum_{a'=a+1}^{M} k_{a,a'}\{||\vec{x}_a + \vec{\delta}_a^p - \vec{x}_{a'} - \vec{\delta}_{a'}^p|| - l_{a,a'}\}^2 \\
&\quad + \frac{1}{2}\sum_{b=1}^{N-1}\sum_{b'=b+1}^{N}\left\{ \sum_{i=1}^{n(b)}\sum_{j=1}^{n(b')} k_{b,i,b',j}(||\vec{x}_{b,i} + H_{b,i}\vec{\delta}_b^r - \vec{x}_{b',j} - H_{b',j}\vec{\delta}_{b'}^r|| - l_{b,i,b',j})^2 \right\} \\
&\quad + \frac{1}{2}\sum_{a=1}^{M}\sum_{b=1}^{N}\left\{ \sum_{i=1}^{n(b)} k_{a,b,i}(||\vec{x}_a + \vec{\delta}_a^p - \vec{x}_{b,i} - H_{b,i}\vec{\delta}_b^r|| - l_{a,b,i})^2 \right\}.
\end{aligned}
\tag{14}
$$

Same as the derivation of EOM for HENM, the cost function also consisted of point mass-point mass, rigid cluster-rigid cluster, and point mass-rigid cluster interaction denoted by $C_{a,a'}$, $C_{b,b'}$ and $C_{a,b}$, respectively. The first and second terms in Eq. (14) are the cost functions proposed elsewhere for Cα ENI and rigid-cluster ENI, respectively [25,27,28]. To consider the interactions between clusters and point masses, the last term is newly added in the cost function for HENI. The values $l$ is an estimate of the distance between two residues in an intermediate conformation, which can be chosen as,

$$
l_{a,b,i} = (1-\alpha)||\vec{x}_a - \vec{x}_{b,i}|| + \alpha||\vec{y}_a - \vec{y}_{b,i}||,
\tag{15}
$$

where $\alpha (0 < \alpha < 1)$ sets the coefficient specifying how far a given state is along the transitional pathway from $\{\vec{x}\}$ to $\{\vec{y}\}$.

The linearized cost function for HENI can be derived such as

$$
\begin{aligned}
C(\vec{\delta}) &\approx \frac{1}{2}(\vec{\delta}^{p\mathrm{T}}\Gamma^p\vec{\delta}^p + \vec{\delta}^{r\mathrm{T}}\Gamma^r\vec{\delta}^r + \vec{\delta}^{h\mathrm{T}}\Gamma^h\vec{\delta}^h) + \frac{1}{2}(\vec{\gamma}^p\vec{\delta}^p + \vec{\gamma}^r\vec{\delta}^r + \vec{\gamma}^h\vec{\delta}^h) + B \\
&= \frac{1}{2}[\,\vec{\delta}^{p\mathrm{T}}\ \ \vec{\delta}^{r\mathrm{T}}\,]\left( \begin{bmatrix} \Gamma^p & \tilde{0} \\ \tilde{0} & \tilde{0} \end{bmatrix} + \begin{bmatrix} \tilde{0} & \tilde{0} \\ \tilde{0} & \Gamma^r \end{bmatrix} + \Gamma^h \right)\begin{bmatrix} \vec{\delta}^p \\ \vec{\delta}^r \end{bmatrix} \\
&\quad + \frac{1}{2}([\,\vec{\gamma}^p\ \vec{0}\,] + [\,\vec{0}\ \vec{\gamma}^r\,] + \vec{\gamma}^h)\begin{bmatrix} \vec{\delta}^p \\ \vec{\delta}^r \end{bmatrix} + B,
\end{aligned}
\tag{16}
$$

where $\Gamma^p$: $3M \times 3M$ matrix, $\Gamma^r$: $6N \times 6N$ matrix, $\Gamma^h$: $(3M+6N) \times (3M+6N)$ matrix. $\vec{\gamma}^p$: $1 \times 3M$ row vector, $\vec{\gamma}^r$: $1 \times 6N$ row vector, $\vec{\gamma}^h$: $1 \times (3M+6N)$ row vector, and $B$: constant.

One can finally obtain $\vec{\delta}^h$, which minimizes the cost of Eq. (16) by setting its derivative zero,

$$
\frac{\partial C(\vec{\delta})}{\partial \vec{\delta}} = \Gamma\vec{\delta}^h + \frac{1}{2}\vec{\gamma}^{\mathrm{T}} = \vec{0},
\tag{17}
$$

where

$$
\Gamma = \begin{bmatrix} \Gamma^p & \tilde{0} \\ \tilde{0} & \Gamma^r \end{bmatrix} + \Gamma^h \quad \text{and} \quad \vec{\gamma} = [\,\vec{\gamma}^p\ \ \vec{\gamma}^r\,] + \vec{\gamma}^h.
\tag{18}
$$

Finally, one can generate a transitional pathway by iteratively updating $\vec{\delta}^h$ from one to the other conformation with an increment of $\alpha$. For further details of derivation, see Appendix A.

### 2.5. Overlap value calculation

The quality of HENM (including both HNMA and HENI) is evaluated by comparing the following three different overlap values. The first overlap value $O^i$ measures the angle difference between an eigenvector, $\bar{v}_i$, from each NMA and the displacement vector
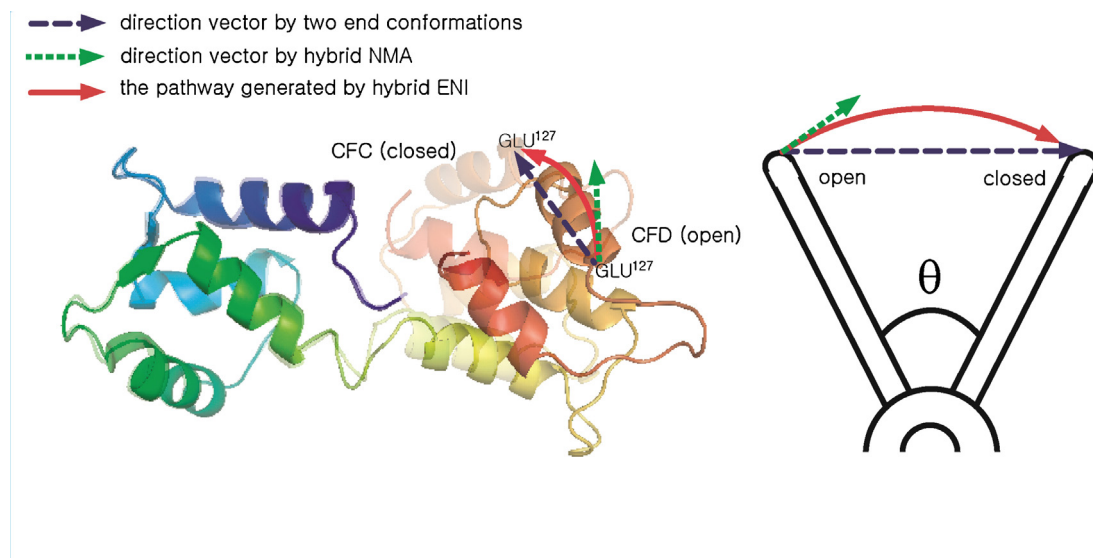
**Fig. 3.** Schematic of the variables in the overlap calculation. The open (CFD, dark color) and closed (CFC, transparent color) forms of Calmodulin are superimposed. The blue arrow is the position error between the two end conformations ($\Delta \bar{x}_{\exp}$), the green arrow is the eigenvector of HNMA ($\bar{v}_{H,i}$), and the red arrow is the pathway generated by HENI ($\Delta \bar{x}_{eni}$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

of the two end conformations in Cartesian space, $\Delta \bar{x}_{\exp} = \bar{x}_C - \bar{x}_O$, such that

$$O_i = \frac{|\bar{v}_i \cdot \Delta \bar{x}_{\exp}|}{(\bar{v}_i \cdot \bar{v}_i)^{1/2}(\Delta \bar{x}_{\exp} \cdot \Delta \bar{x}_{\exp})^{1/2}} \tag{19}$$

where $\bar{x}_C$ and $\bar{x}_O$ are the position vectors of the closed and the open forms, respectively.

The second overlap value, $O_{H,eni}^i$, is defined as the cosine value between $\bar{v}_{H,i}$ from HNMA and the displacement vector along the computed HENI pathway, $\Delta \bar{x}_{eni} = \bar{x}_1 - \bar{x}_O$, such that

$$O_{H,eni}^i = \frac{|\bar{v}_{H,i} \cdot \Delta \bar{x}_{eni}|}{(\bar{v}_{H,i} \cdot \bar{v}_{H,i})^{1/2}(\Delta \bar{x}_{eni} \cdot \Delta \bar{x}_{eni})^{1/2}} \tag{20}$$

where $\bar{x}_1$ is the first intermediate conformation computed by HENI.

Conventionally, $O^i$ was used to evaluate the quality of the conventional Cα NMA results. However, if a conformational change accompanies large collective motions such as hinge and translocation, the harmonic mode shapes ($\bar{v}_i$) computed by NMA at one local minimum state (open conformation in this context) are unable to predict an inharmonic pathway to the other minimum state (closed conformation in this context). Hence, the $O^i$ evaluation becomes less meaningful for pathway generation. Thus, we introduced $O_{H,eni}^i$ to validate the quality of HENI. Fig. 3 illustrates this idea graphically by adding three vectors, which are used to calculate the two different overlap values, on the superimposed Calmodulin structure.

## 3. Results and discussion

### 3.1. Assessment of hybrid elastic network model

To evaluate the qualities of HENM in terms of physical reality and computational efficiency, we have applied HENM to the various protein structures listed in Table 1. These six types of molecules possess several rigid domains and flexible regions defined by the WRMSD method. Table 1 shows the list of proteins examined here with their total DOF. By using the HENM modeling, their DOF dropped down by 7–66%. For objective and quantitative evaluation, we calculate the maximum overlap values and corresponding mode numbers of different ENMs, such as conventional CG ENM

[27], anisotropic network model (ANM), and rotations–translations of blocks (RTB). The eigenvectors of each ENM were obtained from currently available web servers [40–42]. The NMA calculation was conducted with default values on the web servers.

Table 2 presents the maximum overlap values for each protein and corresponding mode numbers. Compared with the point mass based ENMs; conventional CG ENM and ANM, the maximum overlap values of HNMA are bigger or at least similar. Also, the corresponding mode numbers are not shifted to higher modes. It indicates that although the DOF of HENM is relatively small, HNMA still successfully capture intrinsic motions of proteins. Moreover, the comparison with the RTB method, which is based on the rigid body motion as HENM does, shows much better results. The maximum overlap values are increased and the corresponding mode numbers are decreased. Significantly, the reduction of the system's DOF by rigid clustering does not degenerate the quality of NMA because functionally important motions are strongly related to a few lowest modes that are not sensitive to local dynamics in general.

Another validation is the comparison between the results of NMAs and $O_{H,eni}^i$. The overlap values $O_{H,eni}^i$ are very similar to the results of NMAs in all cases. It means that the HENI pathways are highly correlated with system's intrinsic mode shapes although HENI is not on the basis of vibration mechanics, but relying on only topological constraints.

### 3.2. Computational complexity of hybrid elastic network model

Table 1 shows that a large number of DOF has been reduced in HENM. It conjectures that HENM saves much computation time in principle. However, the transformation from individual residue's Cartesian coordinates to those of assembled rigid clusters requires additional computation time (Eq. (16)) and vice versa. We reasonably compare the total run time for projecting individual residues into rigid clusters, resolving the equation (Eq. (17)) and inverse projection to obtain new coordinates for each residue. Table 3 presents computational cost of both HENI and Cα ENI.

Two issues arise in Table 3. (i) Although HENM of lactoferrin has the fewer DOF than HIV-1, Calmodulin, and Adenylate Kinase (AK) in Table 1, its HENI run time is longer. (ii) The HENI run time of HIV-1, Calmodulin, and AK is longer than that of the corresponding

**Table 2**
Maximum overlap values of the example proteins obtained by different ENM methods.

| | HNMA | Conventional Cα NMA[b] | ANM[c] | RTB[d] | HENI ($O^i_{H,eni}$) |
|---|---|---|---|---|---|
| HIV-protease | 0.72 (2)[a] | 0.74 (2) | 0.75 (1) | 0.64 (4) | 0.74 (2) |
| Calmodulin | 0.78 (2) | 0.77 (2) | 0.70 (1) | 0.77 (3) | 0.9 (2) |
| Adenylate Kinase | 0.79 (1) | 0.8 (1) | 0.80 (1) | 0.62 (1) | 0.77 (1) |
| Tyrosine phosphatase | 0.45 (3) | 0.35 (1) | 0.41 (1) | 0.47 (9) | 0.48 (2) |
| Lactoferrin | 0.57 (1) | 0.49 (1) | 0.46 (3) | 0.46 (2) | 0.53 (1) |
| Ca2 ATPase | 0.42 (8) | 0.42 (2) | 0.34 (8) | 0.26 (9) | 0.34 (9) |

[a] The number inside the parentheses indicates the non-zero mode number at which the maximum overlap value is achieved.
[b] Conventional Cα NMA results are obtained from KOSMOS server [41].
[c] ANM results are obtained from ANM server [40].
[d] RTB results are obtained from ElNemo server [42].

Cα ENIs although HENM has definitely smaller DOF than Cα ENM. These two issues are mainly caused by the additional projection time for rigid clustering. This projection run time is influenced by the number of interactions of rigid domains. In Table 3, lactoferrin has a larger number of interactions of point mass–cluster and cluster–cluster compared to HIV-1, Calmodulin, and AK. For the second question, Table 3 also provides a comparison of the number of interactions between HENM and Cα ENM. For small proteins (e.g., HIV-1, Calmodulin, and AK), even though HENM eliminates the intraconnections within each rigid domain, the number of interactions does not reduce substantially. Even worse, in these cases, the projection time makes the total run time of HENI much longer than the Cα ENI case. However, for large proteins (e.g., tyrosine, lactoferrin, Ca2 ATPase), the total run time is reduced noticeably because the main ENI calculation time becomes much more dominant in these cases rather than pre- and post-processing time. From this, we deduce an empirical method to predict the total run time with respect to protein size (i.e., residue number). Fig. 4 presents the relationship between the total run time and residue number. HENI begins to run faster than Cα ENI when the residue number is larger than 300. Thus, we quantitatively conclude that HENI is computational meaningful when the residue number of a protein is greater than 300.

### 3.3. Case study: ribosome

As an example of HENM, we simulated ribosome with both HNMA and HENI. Ribosome (70S) is a large ribonucleoprotein complex that acts as a protein synthesis factory by translating the genetic information encoded in messenger RNA (mRNA) into a specific sequence of amino acids in a cell. This macromolecule consists of two major multi-component subunits called 50S (large) and 30S (small). The important structural units of the ribosome are described in many studies [43–48]. Thus, definitions of rigid clusters for the 30S subunit (2HGP) and 50S subunit (2HGQ) can be formulated.

The 30S structure consists of the 16S rRNA (1515 nucleotides, (nt)) and 21 proteins (2396 residues); this structure can be iden-
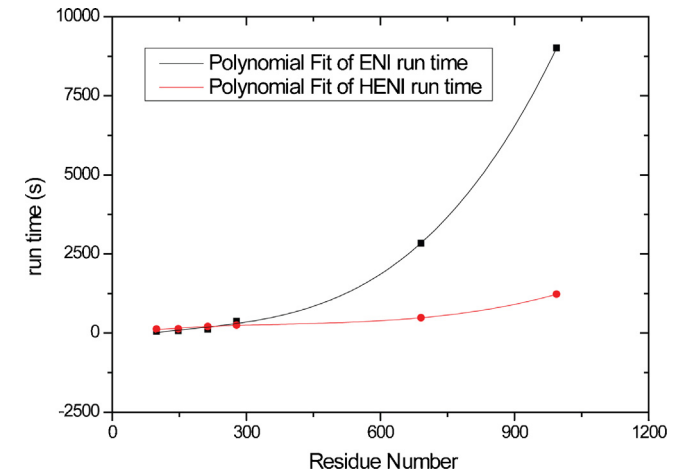


**Fig. 4.** The polynomial fit of the relationship between the total run time and residue number. The black and red curves represent the total run time of Cα ENI and HENI, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

tified by the head, body, and other important landmarks, such as beak, shoulder, platform, and spur. As each structure has a role in the translocation of the tRNA, these structures are treated as rigid clusters in the HENM. The A, P, and E site tRNAs (228 nt) are also provided in this structure, allowing all tRNAs to be point masses in the HENM. Frank and Agrawal [46] showed that the head and body of the 30S subunit are rotated with respect to the 50S subunit. The structures of the 30S subunit and tRNA are shown in Fig. 5(A).

The 50S subunit consists of the 23S rRNA (2889 nt), 5S rRNA (123 nt), and 34 proteins (3692 residues). Additionally, important structural units within the 50S subunit are the L1 stalk, L7/L12 stalk, and body. The L1 and L7/L12 stalks act as arms, helping the mRNA exit from the E site or enter the A site. In the 2HGY structure, the L7/L12 stalk does not exist; therefore, the rigid clusters are the L1 stalk, the head of the 50S, and the body of the 50S, as shown in Fig. 5(B). In the hybrid elastic model, the 70S ribosome is modeled as

**Table 3**
Timing results for HENM and Cα ENI[a] and the number of interactions in different parts of HENM.

| Protein | Run time (s) for Cα ENI | Run time (s) for HENI | Total number of interactions | | Number of interactions of HENM | | |
|---|---|---|---|---|---|---|---|
| | | | Cα ENM | HENM | Point mass–point mass | Cluster–cluster | Point mass–cluster |
| HIV-1 protease | 50.38 | 124.91 | 2080 | 1806 | 680 | 324 | 802 |
| Calmodulin | 67.16 | 130.13 | 3186 | 1798 | 536 | 232 | 1030 |
| Adenylate kinase | 115.94 | 206.11 | 4996 | 4036 | 2448 | 286 | 1302 |
| Tyrosine phosphatase | 373.61 | 248.28 | 5888 | 3972 | 1786 | 544 | 1642 |
| Lactoferrin | 2842.34 | 482.20 | 15,162 | 3634 | 256 | 2044 | 1334 |
| Ca2 ATPase | 9014.70 | 1229.83 | 24,556 | 14,402 | 6280 | 2644 | 5478 |

[a] All tests have been performed using Matlab on a 2.33-GHz Intel PC with 3.25 GB memory.
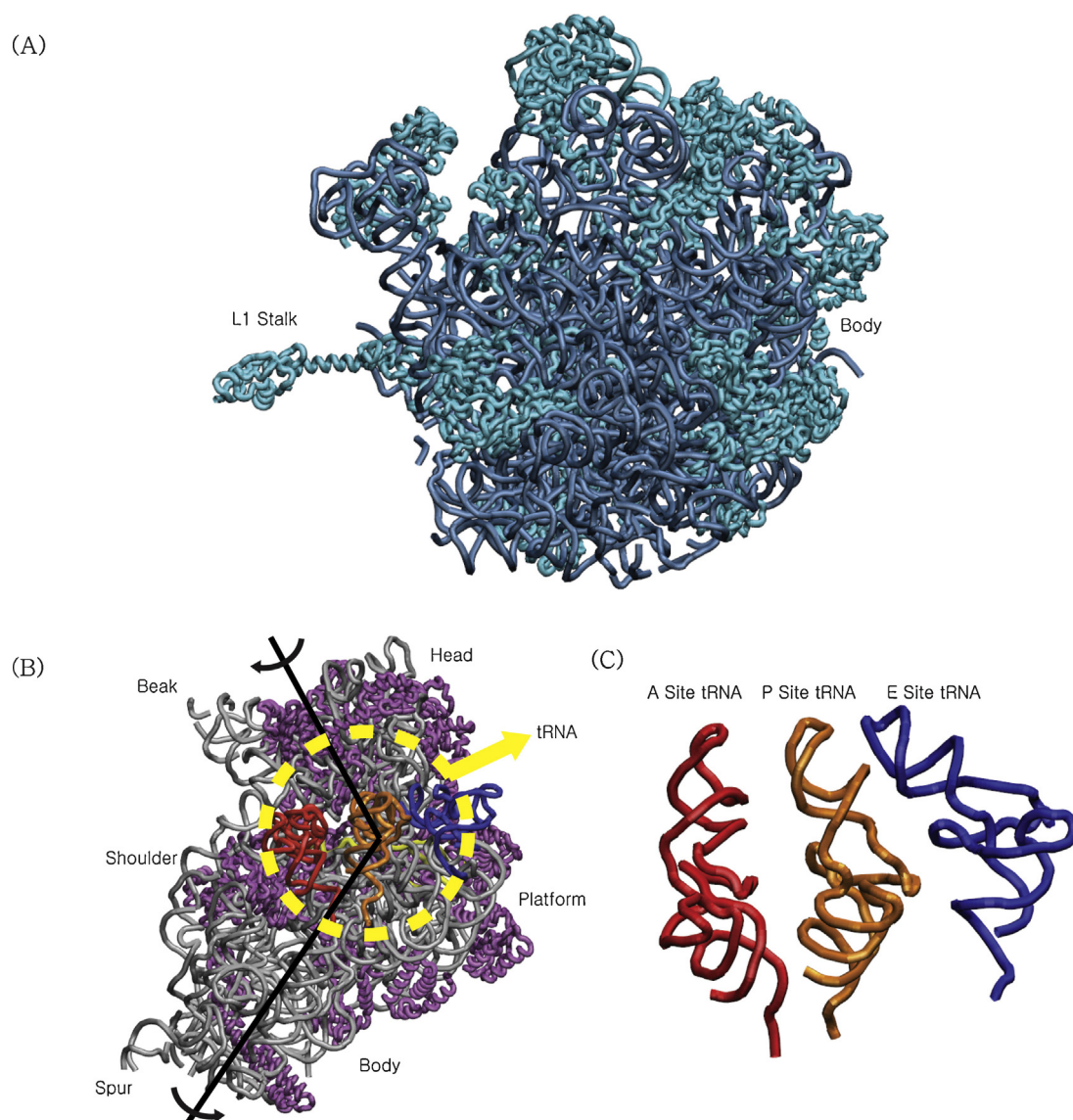
**Fig. 5.** Schematic of the 70S ribosome structure. (A) The structure of the 50S subunit tRNA is viewed from the 30S interface. The blue gray color represents the 23S and 5S rRNA, and cyan shows the 34 proteins in the 50S subunit. (B) The structure of the 30S subunit tRNA is viewed from the 50S interface. The purple color represents the 16S rRNA and the 21 proteins are colored in light purple. (C) A, P, and E site tRNAs are colored in red, yellow, and blue, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

**Table 4**
The components of the rigid clusters and point masses in HENM.

| | Nucleotides (nt) and residues in the 70S ribosome | Rigid clusters |
|---|---|---|
| 2HGP (30S subunit) | A site tRNA: 76 nt<br>P site tRNA: 76 nt<br>E site tRNA: 76 nt<br>mRNA: 50 nt<br>16S rRNA: 1515 nt<br>21 proteins: 2396 residues | **Cluster 1**: (Head of 30S)<br>a. U5–U921 (16S rRNA)<br>b. U1391–U1542 (16S rRNA)<br>c. Protein chains: G, H, I, K, N, O, R, S, T, U, W<br>**Cluster 2**: (Body of 30S)<br>a. G922–U1390 (16S rRNA)<br>b. Protein chains: E, F, J, L, M, P, Q, V, X |
| 2HGQ (50S subunit) | 23S rRNA: 2889 nt<br>5S rRNA: 123 nt<br>34 proteins: 3692 residues | **Cluster 3**: (L1 stalk)<br>a. Protein chain: C<br>**Cluster 4**: (Body of 50S)<br>a. 23S rRNA<br>b. 5S rRNA<br>c. Protein chains: D to 8 |
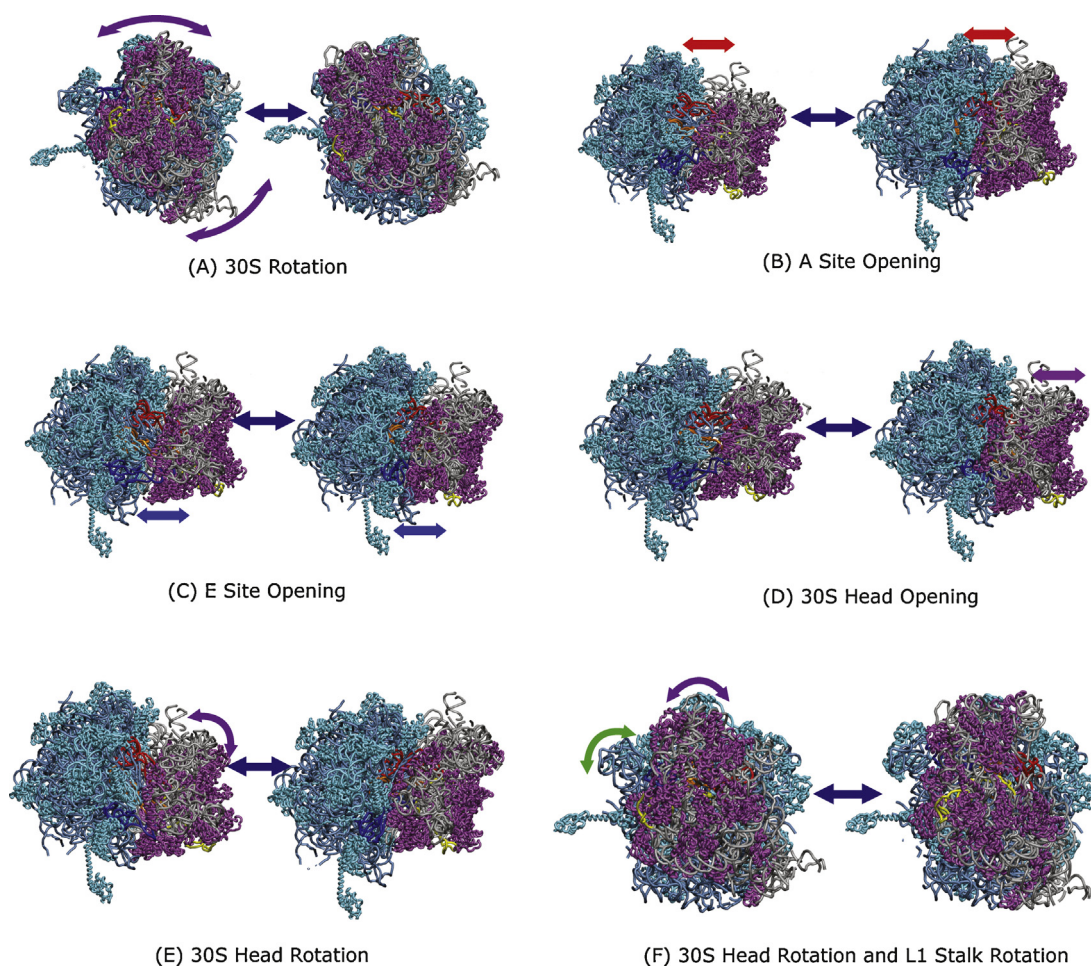
**Fig. 6.** The first six mode shapes of HNMA in ribosome. (A) The first mode shows a ratchet-like motion for the global translocation mechanism of the ribosome (viewed from the 30S subunit). (B) The second mode represents the facilitation of the tRNA docking into the A site (viewed from the top of the 70S structure). (C) The third mode shows the facilitation of the tRNA undocking from the E site (viewed from the top of the 70S structure) (D) The fourth mode is related to the opening motion of the head of the 30S subunit (viewed from the top of the 70S structure) (E) The fifth mode helps the translocation of tRNAs by the rotation of the 30S head (viewed from the top of the 70S structure). (F) The sixth mode demonstrates a synchronous rotation of the head of the 30S subunit and L1 stalk (viewed from the 30S structure).

4 large rigid clusters (the L1 stalk and body from the 50S subunit and the head and body from the 30S subunit) and 3322 point masses, including the tRNAs, mRNA, and interface regions between the two subunits. Table 4 shows how the rigid clusters are defined in the 30S and 50S subunits.

### 3.4. Hybrid normal mode analysis of ribosome

The results of the HNMA not only display a ratchet-like motion but also show tRNA translocation, which is one of the most cru-cial functional movements of ribosomes. However, this movement has not yet been fully understood. As shown in Fig. 6, the first six global modes essentially cooperate to enable the aforementioned movements during protein synthesis. The magnitude of fluctua-tion of residues corresponding each eigenvector is also depicted in Supplementary section. Although it is now equally distributed over the entire 70S ribosome structures, the global collective motions of ribosome are clearly observed in these low frequency modes. Conse-quently, HNMA classified the global motions of ribosome into three major categories: ratchet-like motion from mode 1; dock-ing and undocking motions from modes 2 and 3, respectively; and hinge motions of the head of the 30S subunit and L1 stalk from modes 4, 5, and 6.

The ratchet-like motion of the 70S ribosome clearly explains the global translocation mechanism of the ribosome in the first mode

(Fig. 6(A)). However, Tama et al. [49] and Wang et al. [50] reported slightly differently such that mode 1 only presented the rotation of the L1 and L7/L12 stalks while the ratchet-like motion was observed later in mode 3.

The second and third modes facilitate the tRNA docking into the A site and undocking from the E site, respectively. In mode 2, both the 30S and 50S subunits move in opposite directions, opening the A site tRNA to act as a gate that additionally helps the next tRNA dock into the A site position. As all atoms of the tRNA are modeled as point masses in HNMA, the elongation of the tRNA in the A site could be observed in atomic detail. Likewise, the E site tRNA elon-gates due to the undocking motion in mode 3. Both the docking and undocking motions only appear on the HNMA, which contradicts the studies by Tama and Wang. The fluctuation magnitudes corre-sponding to mode 2 and mode 3 show that they are the principal motions for gating A and E sites (Fig. 6(B) and (C)).

The fourth mode is related to the opening motion of the head of the 30S subunit, which allows extra room for tRNAs to easily shift from one site to the other (Fig. 6(D)). The fifth mode represents the rotation of the head of the 30S subunit and dominantly contributes to the translocation of the tRNA in the 30S subunit (Fig. 6(E)). The magnitude plot also shows that the movement of the 30S subunit is much greater than that of the 50S subunit. The opposite motion between the head and body of the 30S subunit is detected in the sixth mode, and this mode also shows movement of the L1 stalk
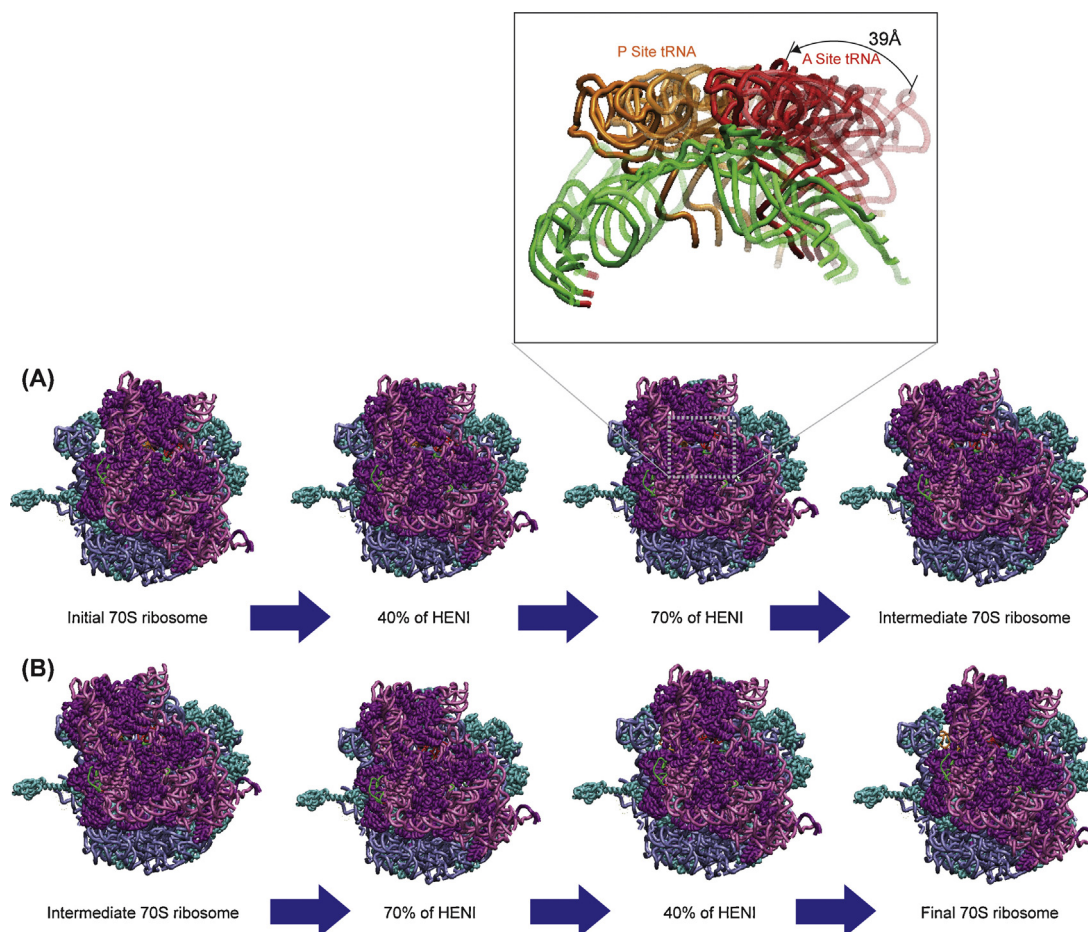
**Fig. 7.** Translocation pathway of ribosome. (A) HENI from initial to intermediate 70S ribosome structures (STEP1). This structure is viewed from the 30S subunit. The inset shows the translocation of tRNAs in A site (red) and P site (brown). The mRNA is colored in green. (B) HENI from intermediate to final 70S structures (STEP2). This structure is viewed from the 30S subunit. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

in the same direction as the 30S head moves (Fig. 6(F)). Although the movement of the 30S subunit is dominant throughout the 70S structure, a sharp and high peak of the L1 stalk (residues 7202–7429 or Chain C in 2HGQ) is highlighted in the 50S subunit. It has been reported by other literatures that the large motion of L1 stalk is closely related to the functional states of the ribosome [51–53].

The DOF of the proposed HENM are nearly one-third the total DOF of the CG ANM model [50] because the collective residues are defined as rigid clusters. It enables us to analyze large macromolecules efficiently. In contrast, rigid-block NMA is much simpler than HENM because Cα and phosphate atoms are sampled from every 5-residue rigid-blocks in the whole system [49]. It may lose local dynamics too much. For example, tRNA translocation in A, P, E sites, one of the most important functional motions in ribosome, was not clearly observed in rigid-block NMA. However, mode 2 and mode 3 of HNMA successfully capture local dynamics of A and E sites, respectively. Thus, one can say that HENM appropriately balances computational cost and physical accuracy.

### 3.5. Hybrid elastic network interpolation of ribosome

The transition pathway of ribosome is investigated as an example of HENI. The crystal structure of ribosome and its interaction with tRNAs were reported elsewhere [54–56]. The hybrid states of ribosome were also observed by time-resolved electron cryomicroscopy [57–59]. Although the experimental work successfully visualized the hybrid states, the atomic details of the intermediate states in long-time scale are still elusive. In this example, we applied

HENI to simulate the pathway of translocation. As a hypothesis, which can guide the transition pathway, we adopted a two-step mechanism for translocation proposed by Frank and Agrawal [46]. The first step (STEP1) is that the 30S subunit rotates with respect to the 50S subunit, and tRNA and mRNA advance from A and P sites to P and E sites, respectively. The next step for the translocation of tRNA (STEP2) is that only the 30S subunit rotates back to initial position called post translocation. It represents that the 30S releases the A and P site tRNAs, and comes back to its original position to complete one cycle of the translocation.

To simulate STEP1, a homogeneous transformation between A and P sites tRNAs is computed by minimizing RMSD, and it is applied to all of tRNAs, mRNA, and the 30S subunit to get the intermediate structure. Fig. 7(A) shows the initial structure and the obtained intermediate 70S ribosome structure. The inset shows the translocation of tRNA in A and P sites. As the tRNAs and 30S subunit are rotated to the intermediate position, the interactions between 50S rRNA and P site tRNA are disconnected and the new interactions between 50S rRNA and E site tRNA are created. Likewise, the atomic distances between 50S rRNA and A site tRNA increase and 50S rRNA and P site tRNA are decreases, which indicates the new interactions are formed. Then, HENI performs the interpolation from the initial to the intermediate structure resulting in a feasible transition pathway. As for STEP 2, the final position and orientation of translocated tRNAs are unknown, so we adopt the coordinates of P and E site tRNAs as the target of translocated A and P site tRNAs, respectively. Fig. 7(B) shows an HENI pathway from the intermediate to the post translocation state. It should be noted that the molecular

distances between tRNA and ribosome calculated by HENI provide a structural key to an understanding of the translocation mechanism. It clearly shows the change in the distance for the initial to the intermediate structures during STEP1, so all the interactions are disconnected and reconnected depending on the range of distance.

## 4. Conclusion

In this study, HENM is introduced as a highly realistic and inexpensive computational tool for generating conformational pathway as well as vibration mode shapes. It is hypothesized that protein structures can be represented by a mixture of rigid clusters and point masses during conformational changes and WRMSD is applied to determine rigid or flexible regions in the given protein structures.

The present analysis corroborates the following strengths of HENM: first, HENM not only generates functionally important global motions but also predicts feasible pathways between different conformations without loss of physical reality; second, it can save computational time for large proteins.

HNMA is performed to find the direction of global motions in an equilibrium state of proteins, while HENI generates feasible pathways between open and closed conformations. The quality of outcomes of HENM is judged by the comparison of two different overlap values which relate mode shapes computed from HNMA to the position error between two end conformations and the pathway generated by HENI. Although we model the system with less DOF in HENM compared to the conventional ENMs, HENM can still capture its functional mode shapes successfully. Higher overlap values implies that the predicted pathway is feasible and consistent with HNMA results. Also, the comparison of run time in both HENI and conventional Cα ENI quantitatively indicates that HENI has the advantage of computational efficiency for large proteins having more than 300 residues.

Consequently, HENM can play an important role in protein dynamics study as a compromise between computationally expensive MD simulation and physically oversimplified conventional CG ENM. For practical purpose, we also develop a protein dynamics webserver named KOSMOS (http://bioengineering.skku.ac.kr/kosmos), where HENM is fully implemented for both HNMA and HENI calculation [41]. To construct HENM, one can either directly assign rigid clusters through an interactive graphic user interface available on KOSMOS or upload rigid cluster information generated by other rigidity analysis servers. Then, KOSMOS automatically performs the requested analysis and visualizes the results in 3D interactive animations.

## Appendix A. Mass matrix

The mass matrix of rigid cluster $M_i^r$ is defined as

$$M_i^r = \begin{bmatrix} M_{trans}^r & 0_3 \\ 0_3 & M_{rot}^r \end{bmatrix}, \tag{A1}$$

$$M_{trans}^r = \begin{bmatrix} \bar{m}_b & 0 & 0 \\ 0 & \bar{m}_b & 0 \\ 0 & 0 & \bar{m}_b \end{bmatrix} \quad \text{and} \quad M_{rot}^r = I. \tag{A2}$$

Here $\bar{m}_b = \sum_{i=1}^{n(b)} m_{b,i}$ is the total mass of cluster $b$ and $I$ is its moment of inertia matrix such that $I = \sum_{i=1}^{n(b)} m_b(\hat{x}_{b,i}^{r\text{T}} \hat{x}_{b,i}^r I_3 - \hat{x}_{b,i}^r \hat{x}_{b,i}^{r\text{T}})$, where the coordinates of $i$th atom with respect to the center of mass of the corresponding cluster are given by $\hat{x}_{b,i}^r = x_{b,i}^r(0) - x_b^r(0)$. Finally, assembling sub-blocks $M_i^p$ and $M_i^r$ yields the following global inertia matrix. Superscript $h$ stands for a hybrid system

$$M^h = \begin{bmatrix} \overbrace{\begin{matrix} M_1^p & 0_3 & \cdots & 0_3 \\ 0_3 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_3 \\ 0_3 & \cdots & 0_3 & M_M^p \end{matrix}}^{3M \times 3M} & & & & \tilde{0} \\ & \underbrace{\begin{matrix} M_1^r & 0_6 & \cdots & 0_6 \\ 0_6 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_6 \\ 0_6 & \cdots & 0_6 & M_N^r \end{matrix}}_{6N \times 6N} \end{bmatrix} \tag{A3}$$

## Appendix B. Potential energy

$V^p$ is the potential energy for point to point interactions defined as

$$V^p = \frac{1}{2} \sum_{a=1}^{M-1} \sum_{a'=a+1}^{M} k_{a,a'} \left\{ ||\vec{x}_a^p(t) - \vec{x}_{a'}^p(t)|| - ||\vec{x}_a^p(0) - \vec{x}_{a'}^p(0)|| \right\}^2$$

$$= \frac{1}{2} \sum_{a=1}^{M-1} \sum_{a'=a+1}^{M} (\vec{\delta}_a^p(t) - \vec{\delta}_{a'}^p(t))^{\text{T}} G_{a,a'} (\vec{\delta}_a^p(t) - \vec{\delta}_{a'}^p(t)), \tag{A4}$$

where

$$k_{a,a'} = \begin{cases} 1 & \text{if } ||\vec{x}_a - \vec{x}_{a'}|| \le d \\ 0 & \text{if } ||\vec{x}_a - \vec{x}_{a'}|| > d \end{cases} \quad \text{and}$$

$$G_{a,a'} = k_{a,a'} \frac{(\vec{x}_a^p(0) - \vec{x}_{a'}^p(0))(\vec{x}_a^p(0) - \vec{x}_{a'}^p(0))^{\text{T}}}{||\vec{x}_a^p(0) - \vec{x}_{a'}^p(0)||^2}.$$

$V^r$ is the potential energy for cluster to cluster interactions defined as

$$V^r = \frac{1}{2} \sum_{b=1}^{N-1} \sum_{b'=b+1}^{N} \left[ \sum_{i=1}^{n(b)} \sum_{j=1}^{n(b')} k_{b,i,b',j} (||\vec{x}_{b,i}^\gamma(t) - \vec{x}_{b',j}^\gamma(t)|| - ||\vec{x}_{b,i}^\gamma(0) - \vec{x}_{b',j}^\gamma(0)||)^2 \right]$$

$$= \frac{1}{2} \sum_{b=1}^{N-1} \sum_{b'=b+1}^{N} \vec{\Delta}_{b,b'}(t)^{\text{T}} \left[ \sum_{i=1}^{n(b)} \sum_{j=1}^{n(b')} k_{b,i,b',j} S_{b,i,b',j} \right] \vec{\Delta}_{b,b'}(t) \tag{A5}$$

where $k_{b,i,b',j} = \begin{cases} 1 & \text{if } ||\vec{x}_{b,i} - \vec{x}_{b',j}|| \le d \\ 0 & \text{if } ||\vec{x}_{b,j} - \vec{x}_{b',j}|| > d \end{cases}$, $\vec{\Delta}_{b,b'} = [\vec{\delta}_b^{r\text{T}} \quad \vec{\delta}_{b'}^{r\text{T}}]^{\text{T}} \in \Re^{12}$, $S_{b,i,b',j} = Q_{b,i,b',j}^{\text{T}} Y_{b,i,b',j} Q_{b,i,b',j}$, and

$$Y_{b,i,b',j} = \frac{(\vec{x}_{b,i}^r(0) - \vec{x}_{b',j}^r(0))(\vec{x}_{b,i}^r(0) - \vec{x}_{b',j}^r(0))^{\text{T}}}{||\vec{x}_{b,i}^r(0) - \vec{x}_{b',j}^r(0)||^2}.$$

where $Q = [H_{i,b} - H_{i,b'}]$

$H_{b,i} = [I_3 - mat(\vec{x}^r_{b,i}(0) - \vec{x}^r_i(0))]$
$H_{b',j} = [I_3 - mat(\vec{x}^r_{b',i}(0) - \vec{x}^r_j(0))]$

$V^h$ is the potential energy for cluster to point interactions defined as

$$V^h = \frac{1}{2}\sum_{a=1}^{M}\sum_{b=1}^{N}\left\{\sum_{i=1}^{n(b)}k_{a,b,i}\{||\vec{x}^p_a(t) - \vec{x}^r_{b,i}(t)|| - ||\vec{x}^p_a(0) - \vec{x}^r_{b,i}(0)||\}^2\right\}$$

$$= \frac{1}{2}\sum_{a=1}^{M}\sum_{b=1}^{N}\left\{\sum_{i=1}^{n(b)}k_{a,b,i}\left\{||\underbrace{(\vec{x}^p_a(0) - \vec{x}^r_{b,i}(0))}_{W} + \underbrace{\delta^p_a(t) - H_{b,i}\delta^r_b(t)}_{Z}|| - \underbrace{||\vec{x}^p_a(0) - \vec{x}^r_{b,i}(0)||}_{W}\right\}^2\right\}, \tag{A6}$$

where

$$k_{a,b,i} = \begin{cases} 1 & \text{if } ||\vec{x}_a - \vec{x}_{b,i}|| \le d \\ 0 & \text{if } ||\vec{x}_a - \vec{x}_{b,i}|| > d \end{cases}.$$

The quadratic term in Eq. (A6) can be displayed as

$$|W + Z||^2 + ||W||^2 - 2||W + Z||||W||$$

$$\approx (2||W||^2 + 2WZ + ||Z||^2)$$

$$- 2\left(||W|| + \frac{WZ}{||W||} + \frac{Z^T\left(I_3 - \frac{WW^T}{||W||^2}\right)Z}{2||W||}\right)||W|| = Z^T\left[\frac{WW^T}{||W||^2}\right]Z. \tag{A7}$$

By defining,
$\bar{\Delta}_{a,b} = [\ \delta^{pT}_a \quad \delta^{rT}_b\ ]^T \in \Re^9$ and $Z_{a,b,i} = [\ I_3 \quad -H_{b,i}\ ]\bar{\Delta}_{a,b},$
Eq. (A6) is rewritten as

$$V^h = \frac{1}{2}\sum_{a=1}^{M}\sum_{b=1}^{N}\bar{\Delta}_{a,b}(t)^T\left[\sum_{i=1}^{n(b)}k_{a,b,i}S_{a,b,i}\right]\bar{\Delta}_{a,b}(t), \tag{A8}$$

where

$S_{a,b,i} = Z^T_{a,b,i}Y_{a,b,i}Z_{a,b,i}$ and

$$Y_{a,b,i} = \frac{(\vec{x}_a(0) - \vec{x}_{b,i}(0))(\vec{x}_a(0) - \vec{x}_{b,i}(0))^T}{||\vec{x}_a(0) - \vec{x}_{b,i}(0)||^2}.$$

## Appendix C. Cost function

$$C = C_{a,a'} + C_{b,b'} + C_{a,b}$$

$$= \frac{1}{2}\sum_{a=1}^{M-1}\sum_{a'=a+1}^{M}k_{a,a'}\{||\vec{x}_a + \vec{\delta}^p_a - \vec{x}_{a'} - \vec{\delta}^p_{a'}|| - l_{a,a'}\}^2$$

$$+ \frac{1}{2}\sum_{b=1}^{N-1}\sum_{b'=b+1}^{N}\left\{\sum_{i=1}^{n(b)}\sum_{j=1}^{n(b')}k_{b,i,b',j}(||\vec{x}_{b,i} + H_{b,i}\vec{\delta}^r_b - \vec{x}_{b',j} - H_{b',j}\vec{\delta}^r_{b'}|| - l_{b,i,b',j})^2\right\}$$

$$+ \frac{1}{2}\sum_{a=1}^{M}\sum_{b=1}^{N}\left\{\sum_{i=1}^{n(b)}k_{a,b,i}(||\vec{x}_a + \vec{\delta}^p_a - \vec{x}_{b,i} - H_{b,i}\vec{\delta}^r_b|| - l_{a,b,i})^2\right\}. \tag{A9}$$

As the solutions that minimize the first two terms of Eq. (A9) were already derived elsewhere [25,27,28], we linearize here only the last squared term such as

$$\left(||\underbrace{[\vec{x}_a - \vec{x}_{b,i}]}_{W} + \underbrace{[\vec{\delta}^h_a - H_{b,i}\vec{\delta}^h_b]}_{Z}|| \underbrace{-l_{a,b,i}}_{I}\right)^2$$

$$= (||W + Z|| - l)^2 \approx ||Z||^2 - \frac{l}{||W||}Z^T\left(I_3 - \frac{WW^T}{||W||^2}\right)Z$$

$$+ 2\left(1 - \frac{l}{||W||^2}\right)W^TZ + ||W||^2 - 2l||W|| + l^2. \tag{A10}$$

The second order terms in Eq. (A10) can be written in quadratic form

$$\bar{\Delta}^T_{a,b}\underbrace{[\ I_3 \quad -H_{b,i}\ ]^T\left\{I_3 - \frac{l}{||W||}\left(I_3 - \frac{WW^T}{||W||^2}\right)\right\}[\ I_3 \quad -H_{b,i}\ ]}_{G_{a,b,i}}\bar{\Delta}_{a,b}, \tag{A11}$$

where

$$\bar{\Delta}_{a,b} = \left[\ \delta^{pT}_a \quad \delta^{rT}_b\ \right]^T \in \Re^9 \quad \text{and} \quad Z = [\ I_3 \quad -H_{b,i}\ ]\bar{\Delta}_{a,b}.$$

$G_{a,b}$ can be defined as

$$G_{a,b} = \sum_{i=1}^{n(b)}k_{abi}G_{abi}. \tag{A12}$$

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

$$\underbrace{2\left(1 - \frac{l}{||W||^2}\right)W^T[I_3 \quad -H_{b,i}]\bar{\Delta}_{a,b}}_{g_{a,b,i}}, \tag{A13}$$

$$g_{a,b} = \sum_{i=1}^{n(b)}k_{a,b,i}g_{a,b,i}$$

The third term in Eq. (A9) consequently takes a simplified form with respect to $\bar{\Delta}_{a,b}$ as

$$C_{a,b} \approx \frac{1}{2}\sum_{a=1}^{M}\sum_{b=1}^{N}\{\bar{\Delta}^T_{a,b}G_{a,b}\bar{\Delta}_{a,b} + g_{a,b}\bar{\Delta}_{a,b} + B_{a,b}\}, \tag{A14}$$

where $B_{a,b}$ is the sum of constant terms derived in Eq. (A10) such that

$$B_{a,b} = \sum_{i=1}^{n(b)}k_{a,b,i}(||W||^2 - 2l||W|| + l^2). \tag{A15}$$

Eq. (A14) can be rewritten in terms of $\vec{\delta}^h(t) = [\vec{\delta}^{pT} \vec{\delta}^{rT}]^T \in \Re^{(3M+6N)}$ such as

$$C_{a,b}(\vec{\delta}^h) = \frac{1}{2} \vec{\delta}^{hT} \Gamma^h \vec{\delta}^h + \frac{1}{2} \vec{\gamma}^h \vec{\delta}^h + B_{a,b}. \tag{A16}$$

## Appendix D. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2013.10.009.

## References

[1] H.M. Berman, J. Westbrook, Z.K. Feng, G. Gilliland, T.N. Bhat, et al., The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[2] S. Subbiah, Protein Motions, Landes Bioscience, Austin, 1996.
[3] I. Bahar, T.R. Lezon, L.W. Yang, E. Eyal, Global dynamics of proteins: bridging between structure and function, Annu. Rev. Biophys. 39 (2010) 23–42.
[4] M. Levitt, Molecular dynamics of native protein – I. Computer simulation of trajectories, J. Mol. Biol. 168 (1983) 595–620.
[5] M. Levitt, Molecular dynamics of native protein – II. Analysis and nature of motion, J. Mol. Biol. 168 (1983) 621–657.
[6] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, et al., Atomic-level characterization of the structural dynamics of proteins, Science 330 (2010) 341–346.
[7] V. Tozzini, Coarse-grained models for proteins, Curr. Opin. Struct. Biol. 15 (2005) 144–150.
[8] M.M. Tirion, Large amplitude elastic motions in proteins from a single-parameter atomic analysis, Phys. Rev. Lett. 77 (1996) 1905–1908.
[9] I. Bahar, A.R. Atilgan, B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, Fold Des. 2 (1997) 173–181.
[10] A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, et al., Anisotropy of fluctuation dynamics of proteins with an elastic network model, Biophys. J. 80 (2001) 505–515.
[11] F. Tama, Y.H. Sanejouand, Conformational change of proteins arising from normal mode calculations, Protein Eng. 14 (2001) 1–6.
[12] L. Yang, G. Song, R.L. Jernigan, How well can we understand large-scale protein motions using normal modes of elastic network models? Biophys. J. 93 (2006) 920–929.
[13] I. Bahar, A.J. Rader, Coarse-grained normal mode analysis in structural biology, Curr. Opin. Struct. Biol. 15 (2005) 586–892.
[14] I. Bahar, T.R. Lezon, A. Bakan, I.H. Shrivastava, Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins, Chem. Rev. 110 (2010) 1463–1497.
[15] M. Feig, Z.F. Burton, RNA polymerase II flexibility during translocation from normal mode analysis, Proteins 78 (2010) 434–446.
[16] D. Tobi, I. Bahar, Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 18908–18913.
[17] O. Keskin, Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies, BMC Struct. Biol. 7 (2007) 31.
[18] A. Bakan, I. Bahar, The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 14349–14354.
[19] M. Gur, E. Zomot, I. Bahar, Global motions exhibited by proteins in micro- to milliseconds simulations concur with anisotropic network model predictions, J. Chem. Phys. 139 (2013) 121912.
[20] W.S. Bennett, R. Huber, J. Engel, Structural and functional aspects of domain motions in proteins, Crit. Rev. Biochem. 15 (1984) 291–384.
[21] S. Hayward, H.J.C. Berendsen, Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme, Proteins: Struct. Funct. Genet. 30 (1998) 144–154.
[22] F. Tama, F.X. Gadea, O. Marques, Y.H. Sanejouand, Building-block approach for determining low-frequency normal modes of macromolecules, Proteins 41 (2000) 1–7.
[23] G.H. Li, Q. Cui, A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca2+-ATPase, Biophys. J. 83 (2002) 2457–2474.
[24] O.N.A. Demerdash, J.C. Mitchell, Density-cluster NMA: a new protein decomposition technique for coarse-grained normal mode analysis, Proteins: Struct. Funct. Bioinforma. 80 (2012) 1766–1779.
[25] M.K. Kim, G.S. Chirikjian, R.L. Jernigan, Rigid cluster models of conformational transitions in macromolecular machines and assemblies, Biophys. J. 89 (2005) 43–55.
[26] A.D. Schuyler, G.S. Chirikjian, Normal mode analysis of proteins: a comparison of rigid cluster modes with C-alphacoarse graining, Graph. Model. 22 (2004) 183–193.
[27] M.K. Kim, G.S. Chirikjian, R.L. Jernigan, Elastic models of conformational transitions in macromolecules, J. Mol. Graph. Model. 21 (2002) 151–160.
[28] M.K. Kim, R.L. Jernigan, G.S. Chirikjian, Efficient generation of feasible pathways for protein conformational transitions, Biophys. J. 83 (2002) 1620–1630.
[29] M.K. Kim, Li Wen, B.A. Shapiro, G.S. Chirikjian, A comparison between elastic network interpolation and MD simulation of 16S ribosomal RNA, J. Biomol. Struct. Dyn. 21 (2003) 395–405.
[30] Y.P. Feng, L. Yang, A. Kloczkowski, R.L. Jernigan, The energy profiles of atomic conformational transition intermediates of adenylate kinase, Proteins: Struct. Funct. Bioinforma. 77 (2009) 551–558.
[31] W. Zheng, B.R. Brooks, G. Hummer, Protein conformational transitions explored by mixed elastic network models, Proteins 69 (2007) 43–57.
[32] M. Tekpinar, W.J. Zheng, Predicting order of conformational changes during protein conformational transitions using an interpolated elastic network model, Proteins: Struct. Funct. Bioinforma. 78 (2010) 2469–2481.
[33] A. Ahmed, H. Gohlke, Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory, Proteins 63 (2006) 1038–1051.
[34] A.S. Siddiqui, G.J. Barton, Continuous and discontinuous domains: an algorithm for the automatic generation of reliable protein domain definition, Protein Sci. 4 (1995) 872–884.
[35] K. Hinsen, Analysis of domain motions by approximate normal mode calculations, Proteins: Struct. Funct. Bioinforma. 33 (1999) 417–429.
[36] S. Kundu, D.C. Sorensen, G.N. Phillips Jr., Automatic domain decomposition of proteins by a Gaussian network model, Proteins: Struct. Funct. Bioinforma. 57 (2004) 725–733.
[37] V. Choi, N. Goyal, An algorithmic approach to the identification of rigid domains in proteins, Algorithmica 48 (2007) 343–362.
[38] D.J. Jacobs, A.J. Rader, L.A. Kuhn, M.F. Thorpe, Protein flexibility predictions using graph theory, Proteins 44 (2001) 150–165.
[39] R. Maiti, G.H.V. Domselaar, H.Y. Zhang, D.S. Wishart, SuperPose: a simple server for sophisticated structural superposition, Nucleic Acids Res. 32 (2004) W4–W590.
[40] E. Eyal, L.W. Yang, I. Bahar, Anisotropic network model: systematic evaluation and a new web interface, Bioinformatics 22 (2006) 2619–2627.
[41] S. Seo, M.K. Kim, KOSMOS: a universal morph server for nucleic acids, proteins and their complexes, Nucleic Acids Res. 40 (2012) W531–W536.
[42] K. Suhre, Y.H. Sanejouand, ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement, Nucleic Acids Res. 32 (2004) W4–W610.
[43] J.H. Cate, M.M. Yusupov, G.Z.H. Yusupova, T.N. Earnest, H.F. Noller, X-ray crystal structures of 70S ribosome functional complexes, Science 285 (1999) 2095–2104.
[44] I.S. Gabashvili, R.K. Agrawal, R. Grassucci, C.L. Squires, A.E. Dahlberg, et al., Major rearrangements in the 70S ribosomal 3D structure caused by a conformational switch in 16S ribosomal RNA, EMBO J. 18 (1999) 6501–6507.
[45] A.P. Carter, W.M. Clemons, D.E. Brodersen, R.J. Morgan-Warren, B.T. Wimberly, et al., Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics, Nature 407 (2000) 340–348.
[46] J. Frank, R.K. Agrawal, A ratchet-like inter-subunit reorganization of the ribosome during translocation, Nature 406 (2000) 318–322.
[47] N. Ban, P. Nissen, J. Hansen, P.B. Moore, T.A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution, Science 289 (2000) 905–920.
[48] H.X. Gao, J. Sengupta, M. Valle, A. Korostelev, N. Esxar, et al., Study of the structural dynamics of the E. coli 70S ribosome using real-space refinement, Cell 113 (2003) 789–801.
[49] F. Tama, M. Valle, J. Frank, C.L. Brooks, Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy, Proc. Natl. Acad. Sci. 100 (2003) 9319–9323.
[50] Y.M. Wang, A.J. Rader, I. Bahar, R.L. Jernigan, Global ribosome motions revealed with elastic network model, J. Struct. Biol. 147 (2004) 302–314.
[51] L.G. Trabuco, E. Schreiner, J. Eargle, P. Cornish, T. Ha, Z. Luthey-Schulten, et al., The role of L1 stalk-tRNA interaction in the ribosome elongation cycle, J. Mol. Biol. 402 (2010) 741–760.
[52] P.V. Cornish, D.N. Ermolenko, D.W. Staple, L. Hoang, R.P. Hickerson, H.F. Noller, et al., Following movement of the L1 stalk between three functional states in single ribosomes, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 2571–2576.
[53] J. Fei, P. Kosuri, D.D. MacDougall, R.L. Gonzalez, Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation, Mol. Cell. 30 (2008) 348–359.
[54] G. Yusupova, L. Jenner, B. Rees, D. Moras, M. Yusupov, Structural basis for messenger RNA movement on the ribosome, Nature 444 (2006) 391–394.
[55] M. Selmer, C.M. Dunham, F.V. Murphy, A. Weixlbaumer, S. Petry, A.C. Kelley, et al., Structure of the 70S ribosome complexed with mRNA and tRNA, Science 313 (2006) 1935–1942.
[56] M.M. Yusupov, G.Z. Yusupova, A. Baucom, K. Lieberman, T.N. Earnest, J.H.D. Cate, et al., Crystal structure of the ribosome at 5.5 angstrom resolution, Science 292 (2001) 883–896.
[57] N. Fischer, A.L. Konevega, W. Wintermeyer, M.V. Rodnina, H. Stark, Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy, Nature 466 (2010) 329–333.
[58] X. Agirrezabala, J.L. Lei, J.L. Brunelle, R.F. Ortiz-Meoz, R. Green, J. Frank, Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome, Mol. Cell. 32 (2008) 190–197.
[59] P. Julian, A.L. Konevega, S.H.W. Scheres, M. Lazaro, D. Gil, W. Wintermeyer, et al., Structure of ratcheted ribosomes with tRNAs in hybrid states, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 16924–16927.