Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Molecular Graphics and Modelling 24 (2006) 296–306
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

A connection rule for α-carbon coarse-grained elastic network models using chemical bond information

Jay I. Jeong [a], Yunho Jang [b], Moon K. Kim [b,*]

[a] Department of Mechanical Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA
[b] Department of Mechanical and Industrial Engineering, University of Massachusetts at Amherst, Amherst, MA 01003, USA

Received 6 June 2005; received in revised form 21 September 2005; accepted 21 September 2005
Available online 11 November 2005

Abstract

A sparser but more efficient connection rule (called a bond-cutoff method) for a simplified α-carbon coarse-grained elastic network model is presented. One of conventional connection rules for elastic network models is the distance-cutoff method, where virtual springs connect an α-carbon with all neighbor α-carbons within predefined distance-cutoff value. However, though the maximum interaction distance between α-carbons is reported as 7Å, this cutoff value can make the elastic network unstable in many cases of protein structures. Thus, a larger cutoff value (>11Å) is often used to establish a stable elastic network model in previous researches. To overcome this problem, a connection rule for backbone model is proposed, which satisfies the minimum condition to stabilize an elastic network. Based on the backbone connections, each type of chemical interactions is considered and added to the elastic network model: disulfide bonds, hydrogen bonds, and salt-bridges. In addition, the van der Waals forces between α-carbons are modeled by using the distance-cutoff method. With the proposed connection rule, one can make an elastic network model with less than 7Å distance cutoff, which can reveal protein flexibility more sharply. Moreover, the normal modes from the new elastic network model can reflect conformational changes of a given protein better than ones by the distance-cutoff method. This method can save the computational cost when calculating normal modes of a given protein structure, because it can reduce the total number of connections. As a validation, six example proteins are tested. Computational times and the overlap values between the conformational change and infinitesimal motion calculated by normal mode analysis are presented. Those animations are also available at UMass Morph Server (http://biomechanics.ecs.umass.edu/umms.html).

Keywords: Bond-cutoff; Distance-cutoff; Elastic network model; Normal mode analysis; Network stability

1. Introduction

The normal mode analysis (NMA) is a useful tool to analyze the dynamic characteristics for a protein structure in the frequency domain, and it is often used to calculate the flexibility of protein structures, conformational changes, or temperature factors. Basically, in NMA, one treats the protein structure as a mass-spring system, namely, an elastic network model as depicted in Fig. 1. Each atom is treated as a point mass and the interactions between atoms as virtual springs. Mathematically, we can derive the dynamic equation of this mass-spring system using Lagrange's equation such that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\delta}_i}\right) - \frac{\partial L}{\partial \delta_i} = \mathbf{0}, \tag{1}$$

where $L = T - V$ [1]. T is the kinetic energy, and V is the potential energy of an elastic network model. δ_i is the i th component of generalized deviation vector $\boldsymbol{\delta} \in \mathbb{R}^{3N}$, when N numbers of point-masses are considered for the elastic network.

One can calculate normal modes of the given protein structure by using all-atom empirical potential function. The

* Corresponding author. Tel.: +1 413 545 2357; fax: +1 413 545 1027.
E-mail address: mkkim@ecs.umass.edu (M.K. Kim).

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.09.006

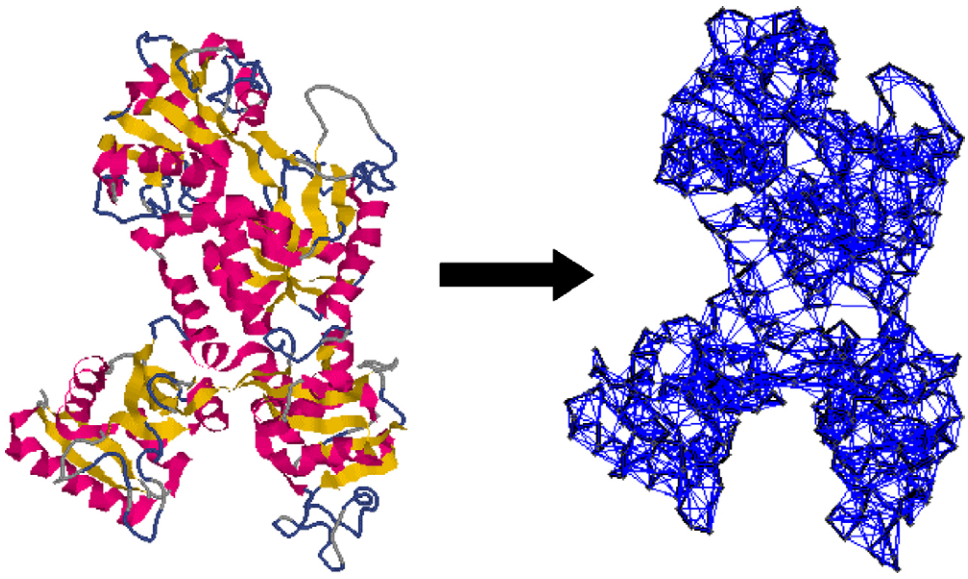

Fig. 1. The representation of an elastic network model. The interactions (i.e., the potential between atoms) are modeled as virtual springs and the atoms as point masses. The backbone of protein structure is presented as the thick black lines, and interactions between α-carbons as the blue lines.

empirical potential function includes the energy from internal and external energy terms such that

$$V_{\text{total}} = \underbrace{\sum_{\text{bonds}} K_r(r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)]}_{\text{Internal energy}} + \underbrace{\sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] + \sum_{\text{H-bonds}}\left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right]}_{\text{External energy}}. \tag{2}$$

The internal energy terms describe the energy associated with changes in bond lengths, bond angles, and torsion angles. On the other hand, the external energy terms include salt-bridges, hydrogen bonds, and van der Waals interactions between atoms. The empirical potential function reflects all the energy of the given protein structure. Usually, software packages of empirical energy model are used to derive the energy of a given structure [2,3].

However, the computational cost to derive an all-atom based elastic network is very high or sometimes impractical. Moreover, one should execute simulated annealing for the given X-ray data to find an equilibrium state as an initial conformation for NMA, which also needs high computational cost. Practically, the limitation of data storage especially for large proteins is another problem [1,4–7].

To reduce such a computational burden, one can make an elastic network model only with α-carbons of the backbone for a given protein structure, where α-carbon atoms are treated as representatives of the corresponding residues. This simplified model is called as α-carbon coarse-grained elastic network model. In the folding of a protein structure, most conformational changes take place in backbone dihedral angles and side chain articulations. Therefore, this approach is sufficient to globally reveal the backbone dynamics of a given protein structure, because the low-frequency modes from NMA represent only the collective motions without any atomic detail [8–11].

To apply a simplified coarse-grained elastic network model to a protein structure, one should define a simplified potential which represents the interactions between α-carbons with virtual springs. Tirion [12]suggested the simplified potential model as an alternative to the empirical potential model such that

$$E(\mathbf{r}_a, \mathbf{r}_b) = \tfrac{1}{2}\gamma(|\mathbf{r}_{a,b}| - |\mathbf{r}^0_{a,b}|)^2, \tag{3}$$

where γ is a scaling constant to match the theoretical result to experimental data. This simplified potential means that the potential energy of the elastic network is treated as a function of distance change between two atoms. Tirion also showed that this simplified potential yields a good result to calculate several lowest normal modes.

Regarding the simplified potential, it is important to determine the connectivity between α-carbons, which represents the interactions between α-carbon atoms in a protein structure. In the literature, several connection rules have been proposed for the simplified elastic network model. One of the well-known and widely used connection rules is the distance-cutoff method, where the α-carbon atoms within a distance-cutoff are connected to each other with virtual springs, and the stiffness values of each virtual spring is set as a constant value [5–8,12,13]. Bahar and Jernigan suggested 7Å as the distance-

cutoff of a coarse-grained model for proteins [14]. This distance-cutoff is reported reliable for Gaussian network model (GNM) [5], where the connectivity between atoms could reflect the vibration of each atom in a given protein structure, so that one can predict temperature factors of a protein structure from NMA. However, this cutoff value did not work well with the anisotropic network model (ANM), which can represent directional fluctuations of each atom in three dimensional space [8]. They had to use the bigger value such as 12–15Å so as not to generate more than six-zero eigenvalues that exceed the number of rigid-body motions of a given protein structure. This is the main problem of the distance-cutoff method, namely it can not guarantee the stability of the elastic network model when a small distance-cutoff value is used. Although the maximum distance of interactions between α-carbons is reported as 7Å, a distance-cutoff of less than 10Å may result in more than six zero eigenvalues. Thus, simplified elastic network models are usually built using a distance-cutoff value larger than 11Å [8,7,13,15].

Alternatively, one can use a cutoff-number of connections instead of a distance-cutoff method [16]. In this method, the number of connections for each atom is limited by a maximum constant value. Typically, 20 connections for each atom is used to build an elastic network model. Even though the elastic network generated by cutoff-number method is more stable and uniform than that of the distance-cutoff method, it is only good for Elastic Network Interpolation (ENI), which is designed for interpolation between two different conformations of the same protein [17–19]. Such a model may destroy local flexibility too much, since the same number of connections per atoms is applied.

In this work, we newly present a bond-cutoff method which includes the chemical interactions of a given protein in its simplified elastic network model. This method makes the network model much sparser, but computation much faster than the conventional rules such as distance-cutoff and cutoff-number methods. In previous work, Jacobs et al. [20] assessed chemical interactions in the protein structures to predict the flexibility by using graph theory. By establishing a bond-bending network and checking the hydrogen bonds and salt-bridges as constraint of the network, they can reveal flexible and rigid parts of the given protein. They showed these interactions play an important role in protein structure flexibility. In contrast to finding flexibility of protein structures, however, one must stabilize the elastic network firstly in order to adapt these chemical interactions and to obtain normal modes. Hence, we propose the virtual spring connections along the backbone chain, which can guarantee the stability of a simplified elastic network model. Then, the connectivity of external energy terms are considered.

Tama and Sanejouand [10] tested overlap values between normal modes and conformational changes of 20 example proteins, when using a simplified elastic network model with the distance-cutoff method. They observed that the collective motions obtained by several lowest normal modes can represent the global conformational changes. In this paper, we also use the overlap calculation as a validation tool for the proposed method.

In Section 2 we present a new methodology to generate elastic network models, in which a network is stabilized by backbone modeling as well as additional spring connections based on chemical bond information. As a validation, we test six example proteins by comparing overlap values between infinitesimal motions calculated by NMA and direction vectors of conformational changes in Section 3. Various features of the bond-cutoff method are discussed in Section 4.

2. Methodology

2.1. Backbone modeling

When building an elastic network model with α-carbons, each α-carbon has three degrees of freedom (DOF) in the direction of X, Y, and Z axes. Therefore, in the case of elastic network system with N number of α-carbons, the total number of degrees of freedom of the system is $3N$, which includes the six rigid-body motions and $3N-6$ number of relative deformations (i.e., non-rigid body normal modes). These $3N-6$ constraints can be categorized into the three internal coordinate representation, which is widely used to describe the conformational changes of a polymer: $N-1$ bond lengths (prismatic joints), $N-2$ bond angles (revolute joints at each pivot point), and $N-3$ torsion angles (revolute joints in the middle of each link).

In order to stabilize an elastic network, two conditions must be kept. First, there must be more than three virtual springs per each point mass to constrain the degrees of freedom of each point mass. Second, total number of constraint must exceed $3N-6$ [21]. One simple but efficient way to satisfy these constraints is to make the spring connections from one residue to more than three neighbors along the backbone of a given protein such that

$$K_{i,j}^{\mathrm{cov}} = \begin{cases} \gamma \times k_{\mathrm{cov1}} & \text{if} \quad |i-j| = 1 \\ \gamma \times k_{\mathrm{cov2}} & \text{if} \quad 2 \le |i-j| \le B_{\mathrm{c}}, \\ 0 & \text{if} \quad |i-j| > B_{\mathrm{c}} \end{cases} \tag{4}$$

where B_{c} is the "bond-cutoff" number, and γ is a constant value to adjust spring stiffness. The k_{cov1} and k_{cov2} are the ratio constants, which represent the stiffness of interactions along the backbone atoms in a given protein. In the case that the value of bond-cutoff value (B_{c}) is 3, the total number of spring connections is equal to the number of minimum constraints $3N-6$, and the each α-carbon has more than three virtual springs with its neighbor α-carbons. This means that the elastic network model with backbone modeling satisfies the mechanical stability, so that it can reflect all the changes through the backbone.

2.2. Chemical interactions: disulfide bonds, hydrogen bonds, salt-bridges, and van der Waals forces

In addition to the backbone modeling, we can now consider the effects of the chemical interactions. First, disulfide bonds are modeled. For the relationship of disulfide bonds, one can

check the position of all sulfide atoms of cysteine residues for the given protein and connect two sulphur atoms with a disulfide bridge, when the distance between them is below 3Å [22]. Then, this connectivity is represented as a linkage of α-carbon pairs in a simplified coarse-grained model. One can make the linking matrix with respect to residues such that

$$K_{i,j}^{\text{SSbonds}} = \begin{cases} \gamma \times k_{\text{SSbonds}} & \text{if} \quad \|\vec{s}_i - \vec{s}_j\| \le d_s \\ 0 & \text{if} \quad \|\vec{s}_i - \vec{s}_j\| > d_s \\ 0 & \text{if} \quad i \text{ th or } j \text{ th residue is not cysteine} \end{cases} \tag{5}$$

where $\vec{s}_i$ is a position vector for a sulphur atom of the i th residue. d_s is distance cutoff of disulfide bonds, which is set as 3Å. k_{SSbonds} represents the ratio constant of the disulfide connection, which is set as 100.

Second, hydrogen bonds are modeled, which play a major role in stabilizing secondary structures such as α-helices and β-strands. We use the HBPLUS program to generate the connections representing hydrogen bonds for a given protein structure [23]. This program automatically positions missing hydrogen atoms in PDB files, and then predicts all the possible hydrogen bonds. From output file of HBPLUS, we obtain the connectivity information and convert it into virtual links between residues. One can write the equation which can represent hydrogen bonds such that

$$K_{i,j}^{\text{Hbonds}} = \begin{cases} \gamma \times k_{\text{Hbonds}} & \text{if} \quad H_{i,j} = 1 \\ 0 & \text{if} \quad H_{i,j} = 0 \end{cases}, \tag{6}$$

where $H_{i,j}$ is the connection relationship between residues calculated by HBPLUS. Each component of $H_{i,j}$ is set as one when a hydrogen bond exists between i th and j th residues, otherwise it is set as zero. k_{Hbonds} is the stiffness ratio for hydrogen bonds, which is set as 10.

Third, ionic bonds or salt-bridges are modeled, which are interactions between charged atoms [24]. For the modeling of salt-bridge, we use a similar rule used in the modeling of disulfide bonds. Namely, if anionic atoms of charged amino acids such as aspartic acid and glutamic acid are placed near the cationic atoms of lysine and arginine within 4Å, one can assume that a salt-bridge exists between the specified residues [25]. This connectivity is also converted into residue-based linkages in the elastic network model. As a result, one can model ionic interactions such that

$$K_{i,j}^{\text{SaltBridge}} = \begin{cases} \gamma \times k_{\text{SaltBridge}} & \text{if} \quad \|\vec{b}_i - \vec{b}_j\| \le d_{\text{ion}} \\ 0 & \text{if} \quad \|\vec{b}_i - \vec{b}_j\| > d_{\text{ion}} \\ 0 & \text{if} \quad i \text{ th or } j \text{ th residue is not a charged amino acid} \end{cases}, \tag{7}$$

where b_i is the position of the ionic atoms of i th residue. The cutoff distance of ionic interaction d_{ion} is set as 4Å, and the stiffness ratio $k_{\text{SaltBridge}}$ is set as 10.

Finally, van der Waals forces are modeled to represent the dispersion force between α-carbons. To model this effect, we adopt distance-cutoff model, where the linking matrix can be derived such that

$$K_{i,j}^{\text{vdw}} = \begin{cases} \gamma \times k_{\text{vdw}} & \text{if} \quad \|\vec{x}_i - \vec{x}_j\| \le R_{\text{c}} \\ 0 & \text{if} \quad \|\vec{x}_i - \vec{x}_j\| > R_{\text{c}} \end{cases} \tag{8}$$

where R_{c} is a distance-cutoff value. k_{vdw} is ratio constant for van der Waals force, which is set as one.

The linking matrix $K_{i,j}$ for the given protein structure is set as union of $K_{i,j}^{\text{Covalent}}$, $K_{i,j}^{\text{SSbonds}}$, $K_{i,j}^{\text{Hbonds}}$, $K_{i,j}^{\text{SaltBridge}}$, and $K_{i,j}^{\text{vdw}}$. Each component of $K_{i,j}$ has the biggest value among those of the linking matrices for each type of chemical interactions. Namely, it has a non-negative value when the corresponding residue has covalent bonds, hydrogen bonds, ionic bonds, van der Waals force, or no connection.

2.3. Overlap of normal modes

The overlap is an index to compare the similarity between the calculated normal modes and the conformational changes of a given protein. In this work, the overlap value defined by Marques and Sanejouard [26] is used such that

$$I_j = \frac{\left|\sum_{i=1}^{3N} A_{ij}\Delta r_i\right|}{\left[\sum_{i=1}^{3N} A_{ij}^2 \sum_{i=1}^{3N} \Delta r_i^2\right]^{1/2}}, \tag{9}$$

where I_j is the overlap index for the j th normal mode, which is the cosine of the angle between the conformational change vector and the j th normal mode vector. The A_{ij} is a displacement by the j th normal mode at the position of i th α-carbon. After superimposing the open form on the closed form, we can obtain the difference vector ΔR which represents the conformational change between two conformations. Here Δr_i is the i th component of ΔR corresponding to the j th α-carbon. The overlap value would be from zero to one. If the calculated normal mode is the exactly same as the direction of conformational change, the overlap value will be ''one''.

3. Results

As a validation of the proposed connection rule, six proteins are tested: LAO binding protein [27], maltodextrin binding protein [28,29], adenylate kinase [30,31], enolase [32,33], lactoferrin [34], and calcium ATPase [35,36]. These protein structures have at least two different conformations such as ''open'' and ''closed'' forms with more than 3Å root mean squared deviation (RMSD) value. We select those proteins from ''Molecular Movement Database'' [37], and obtain the structural information from the Brookhaven PDB [38]. We present the protein names, the PDB codes, and the number of residues in Table 1.

The elastic network models for test proteins are built by using the bond-cutoff method, which includes backbone modeling, disulfide bridges, hydrogen bonds, salt-bridges and van der Waals interactions as mentioned above. In all cases, the value of bond-cutoff (B_{c}) is set as three, and the value

Table 1
Proteins used for NMA with the bond-cutoff method

Protein name	PDB codes	No. of residues
LAO binding protein	2LAO, 1LST	238
Maltodextrin binding protein	1OMP, 1ANF	370
Adenylate kinase	4AKE, 1AKE	421
Enolase	1ONE, 4ENL	436
Lactoferrin	1LFH, 1LFG	691
Calcium ATPase	1SU4, 1KJU	994

of distance-cutoff (R_c), which represents van der Waals interaction, are varied as 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, and 8.0Å. In addition, these new models are compared with four conventional models generated by distance-cutoff rule, where R_c is set to be 11.0, 13.0, 15.0, and 17.0Å, and the stiffness of virtual springs is set as all the same value as one [1,4,8].

In Fig. 2, we present the overlap results of two different types of network models together. One is generated by the bond-cutoff rule proposed here and the other by the conventional distance-cutoff rule. The maximum overlap value are presented as bar graphs, and the rank of the corresponding normal mode (rigid body modes are not taken into account in the numbering) is placed at upper side of the bar graph.

One can observe that the elastic network models by the bond-cutoff method can reflect the conformational change better than distance-cutoff model in the cases of the open forms. In the case of the open form of maltodextrin binding protein (1OMP), the maximum overlap by the bond-cutoff method is 0.93 when B_c is three and R_c is 7.0Å, whereas the distance-cutoff rule achieves only 0.83 when R_c is 15.0Å. The minimum overlap values are 0.67 when B_c is three and R_c is 6.0Å by the bond-cutoff method and 0.70 by the distance-cutoff method, respectively. In the case of LAO binding protein, the elastic networks by the bond-cutoff method shows better performance than the conventional connection rules. The overlap values by the distance-cutoff rule are about 0.80. However, the new method improves them near or more than 0.90 except when the R_c is 4.5Å. One can also observe that in calcium ATPase are the overlap values by the new method better than the distance-cutoff cases, though the normal mode with the maximum overlap value has been shifted from the first to the second lowest non-rigid body mode.

The open form of lactoferrin (1LFH) shows impressive improvements in overlap values. The maximum overlap value is improved from 0.53 (distance-cutoff model when R_c is 11.0Å) to 0.82 (bond-cutoff model when B_c is 3, and R_c is 4.5Å). In Fig. 3, the mode shapes by the bond-cutoff method and by distance-cutoff method are presented, respectively. For comparison, the direction vectors along the conformational change are also displayed by arrow graph. One can observe that the main conformational change between 1LFH and 1LFG is the relative rotational motion of N2 domain [39], which is depicted as Fig. 3(a). In Fig. 3(b), the normal mode by the bond-cutoff method dominantly shows the rotational motion of N2 domain. In contrast, the normal mode from the distance-cutoff model represents relatively small motion of N2 domain, but large twisting motion of C domain as shown in Fig. 3(c). The animation of these motions is presented under http://biomechanics.ecs.umass.edu/umms.html. On the animation, the ENI (Elastic Network Interpolation) method is used to show conformational transition between 1LFH and 1LFG [16]. The ENI method is an interpolation technique to minimize the RMSD value of two given structures, which also presents the rotational motion of N2 domain.

In contrast that most of open forms of test proteins get the higher overlap values when the bond-cutoff method is applied, the overlap values for the closed conformations by the bond-cutoff method are largely similar to those by the distance-cutoff method. For example, the maximum overlap values of the closed form of lactoferrin (1LFG) are 0.65 (the bond-cutoff method when B_c is three and R_c is 7.5Å) and 0.62 (the distance-cutoff method when R_c is 17.0Å), respectively. Only adenylate kinase (1AKE) and LAO binding protein (1LST) exceptionally shows the substantial improvement of the overlap values by the bond-cutoff method even in the closed conformations (see Fig. 2).

NMA results for six example proteins and the animations are available at UMass Morph Server which is a protein dynamics web server based on elastic network models. One cannot only obtain the protein motions graphically, but also download those numeric data files from http://biomechanics.ecs.umass.edu/umms.html.

4. Discussion

4.1. Various stiffness values for chemical interactions

We use different stiffness values for each type of chemical interactions. The force constant of the spring linking between two consecutive α-carbons is set to be 100, whereas the constant for van der Waals interactions is set as one. We determine values of the ratio constants for virtual springs based on the order of each bonding energy for chemical interactions [40].

The virtual springs between i th and $(i+1)$ th residues along the protein backbone represent the virtual bond lengths. The relative distance between two consecutive α-carbons should be about 3.8Å consistently for most protein structures. Therefore, these connections are modeled as stronger compared to other chemical interactions. The connections from i th to $(i+2)$ th residues are required to satisfy virtual bond angle constraints. Moreover, the connections between i th and $(i+3)$ th residues represent the virtual torsion angle constraints. The energy needed to change bond angles of the native protein is much higher than that of torsion angles. However, in a simplified coarse-grained model, these two angle changes turn out to be a mixture of changes in both virtual bond and torsion angles. Thus we set a same stiffness constant value of one for connections between i th and $(i+2)$ th residues, and between i th and $(i+3)$ th residues in this context.

For the other interactions, the accurate stiffness value for each interaction is not available. We just adopt the order of

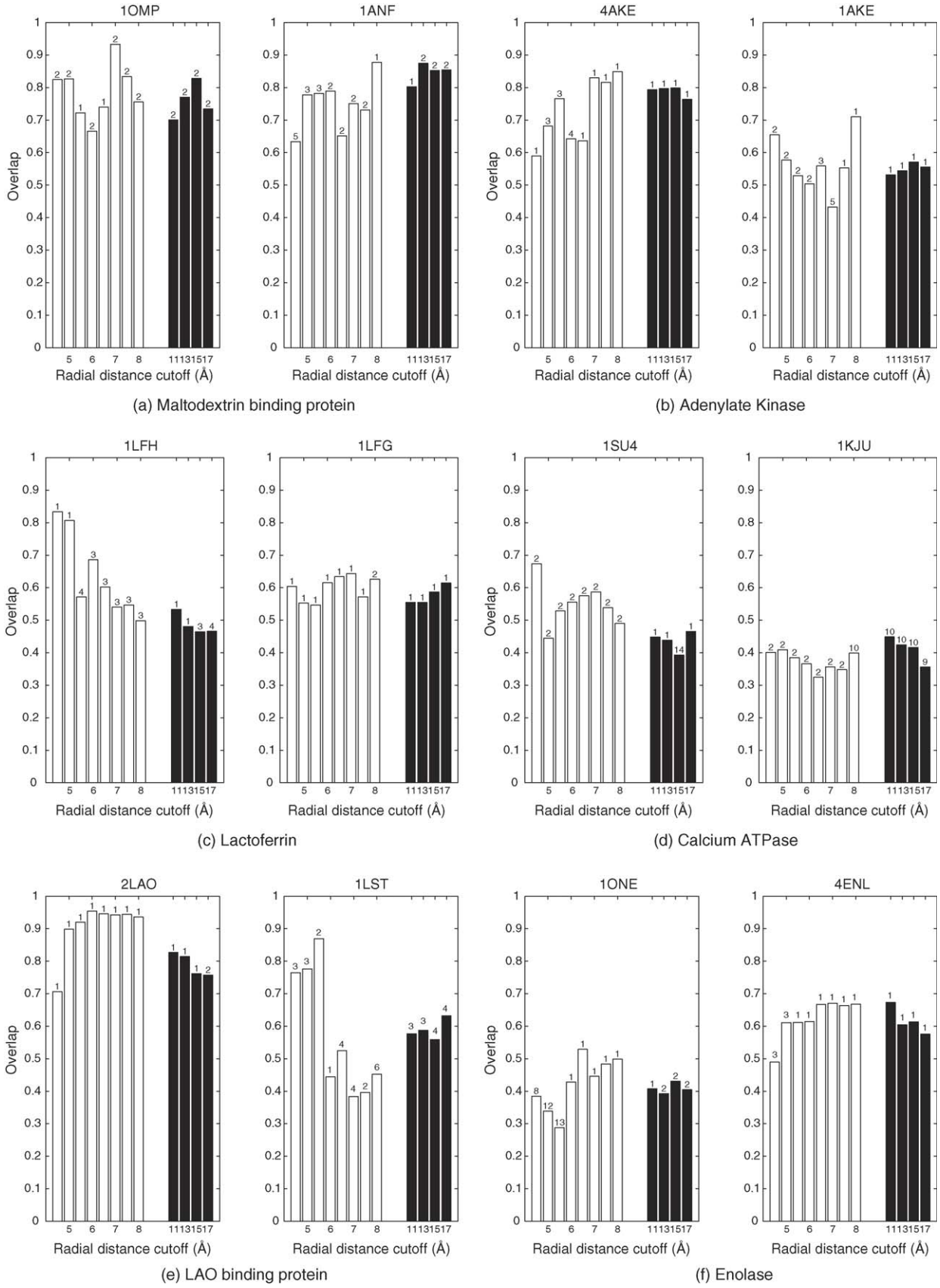

Fig. 2. The comparison of the maximum overlap between the bond-cutoff model (white bar) and the distance-cutoff model (black bar).

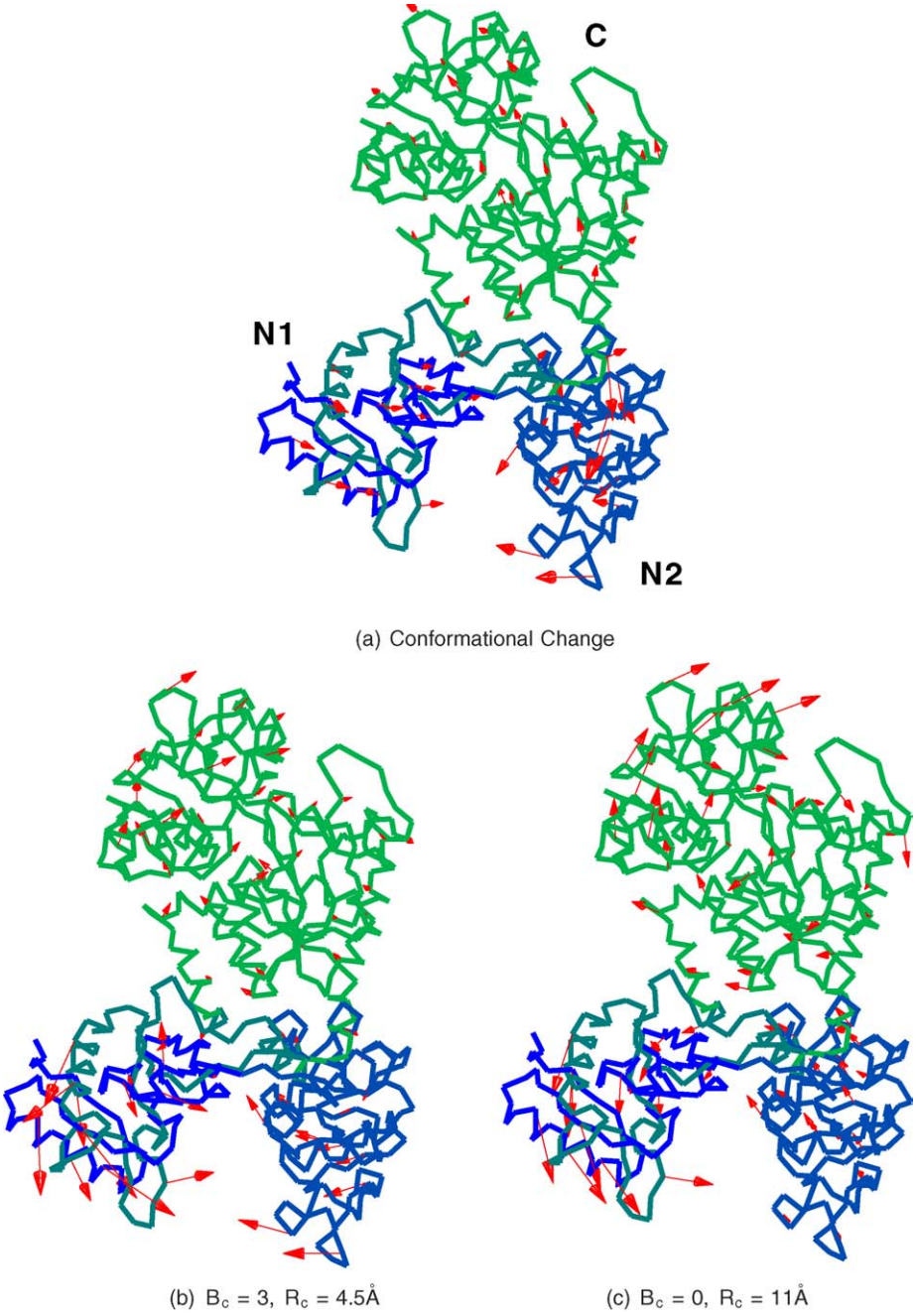

Fig. 3. The comparison of the lowest normal mode of the open form of lactoferrin (1LFH) generated by the bond-cutoff method with that of the distance-cutoff method. (a) The conformational change vector from the open form (1LFH) to the closed form (1LFG). (b) The normal mode shape when the bond-cutoff value (B_c)is three and the distance-cutoff value (R_c)is 4.5Å. (c) The normal mode shape by the distance-cutoff only model when R_c is 11.0Å.

average energy for various interaction as ratio constants. Table 2 summarizes the stiffness ratio with respect to the types of interactions.

4.2. Minimum cutoff value for the distance-cutoff rule

Next, we discuss the reason why 11Å is used as the minimum stable cutoff distance, when one uses the distance-cutoff rule to make the elastic network model for a given protein structure. In Fig. 4, the distances from i th residues to

Table 2
The stiffness ratio

Coefficient	Stiffness ratio	Average energy (kcal/mol)
k_{cov1}	100	100
k_{cov2}	1	
$k_{SSbonds}$	100	100
k_{Hbonds}	10	5
$k_{SaltBrige}$	10	5
k_{vdw}	1	< 1.0

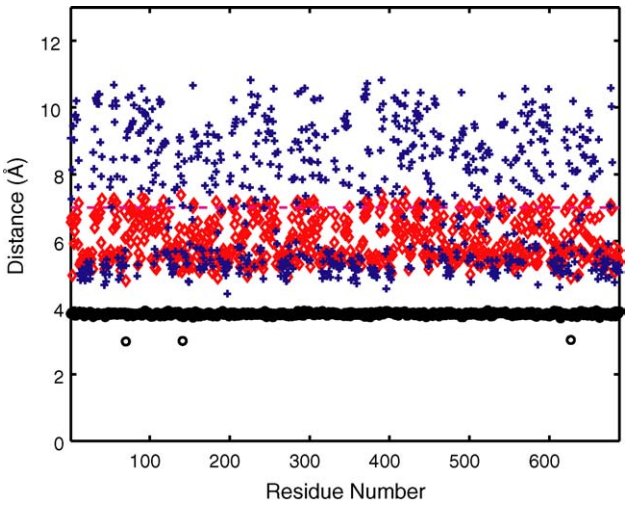

Fig. 4. The distance between neighbors of α-carbons of lactoferrin (1LFH). The distance between i th and $(i+1)$ th α-carbon atoms is presented as symbol 'o', between i th and $(i+2)$ th as ' ◇', and between i th and $(i+3)$ th as ' +' respectively. The maximum distance of i th and $(i+3)$ th backbone atom is 10.8Å. This result shows the minimum distance-cutoff of 1LFH should exceed 10.8Å to make the elastic network stable when the distance-cutoff method is used for lactoferrin. The total number of points is 2067, which is the same as the number of constraint $3N-6$ ($N=691$).

$(i+1)$ th, $(i+2)$ th, and $(i+3)$ th residues are presented for the open form of lactoferrin (1LFH). One can easily observe that the maximum distance between i th and $(i+3)$ th is about 11Å. This result explains well why people have empirically used the cutoff values larger than 11Å to date [1,4,8]. Namely, one has to choose a larger value than 11Å as a distance-cutoff value in order to satisfy the stability condition of an elastic network model mentioned in Section 2.1.

In addition, it is worth emphasizing that the virtual spring connections among three consecutive neighbors along the backbone are not *necessary* condition, but *sufficient* condition with a minimum number of connections to stabilize a given elastic network system. Even if the distance-cutoff method with a cutoff value smaller than 11Å is able to build an elastic network model, it is limited to only small globular proteins due to the network stability. In contrast, one can always obtain a stable elastic network model by the proposed bond-cutoff method regardless of the size and shape of a given protein.

4.3. Making sensitive elastic networks with respect to R_c

In the cases of the distance-cutoff method, one can observe that the change in the maximum overlap value is negligible when the distance-cutoff value R_c varies from 11 to 17Å. As of the open forms of lactoferrin (1LFH) and calcium ATPase (1SU4), the maximum overlap values are about 0.61 and 0.45, respectively, the difference between highest and lowest overlap values is only 0.06 and 0.03. This result implies that the distance-cutoff method is good for making robust elastic networks insensitive to the range of cutoff distance. In fact, the wider range of distance-cutoff value over 11Å has been used to predict reliable temperature factors from NMA based on elastic network models [5,8].

In contrast, the bond-cutoff method shows a different feature. The overlap values are widely varied depending on R_c. Using the bond-cutoff method with B_c of three, the lactoferrin (1LFH) has the maximum overlap value of 0.82 with 4.5Å (R_c), whereas the minimum value of 0.46 is obtained with 8.0Å (R_c). This minimum value is similar to the lowest overlap value of 0.47 obtained by the distance-cutoff model. Therefore, we conclude that compared to the bond-cutoff method, the bond-cutoff method generates much more sensitive (or flexible) elastic network models with respect to R_c. Namely, the elastic network generated by the bond-cutoff method has less damping characteristics than that by the distance-cutoff method.

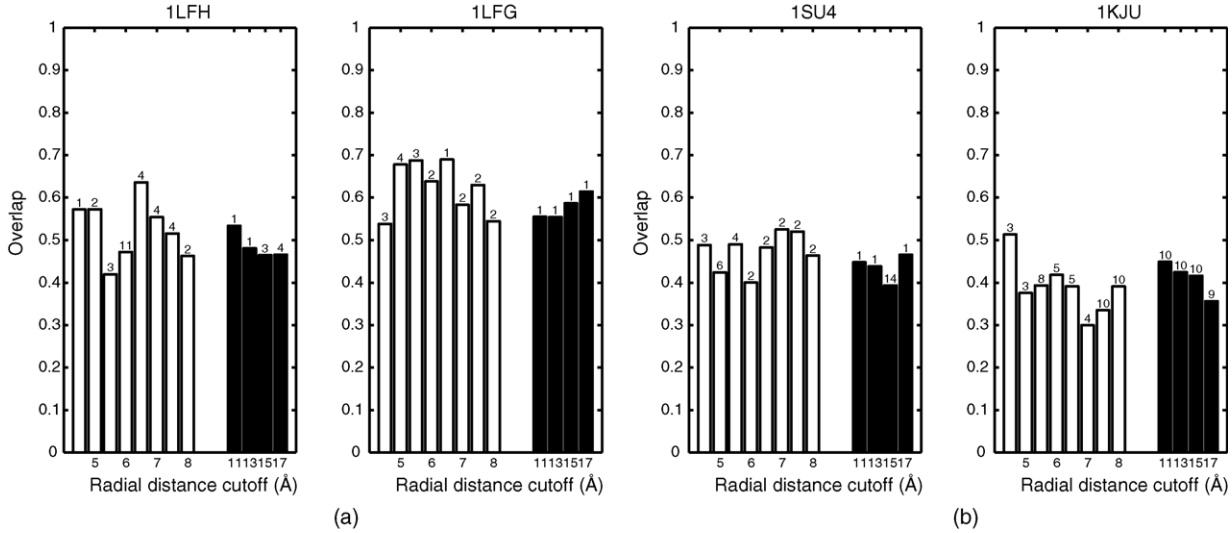

Fig. 5. The maximum overlap of the distance-cutoff model and the bond-cutoff model without hydrogen bond and ionic bond information, for lactoferrin and calcium ATPase.

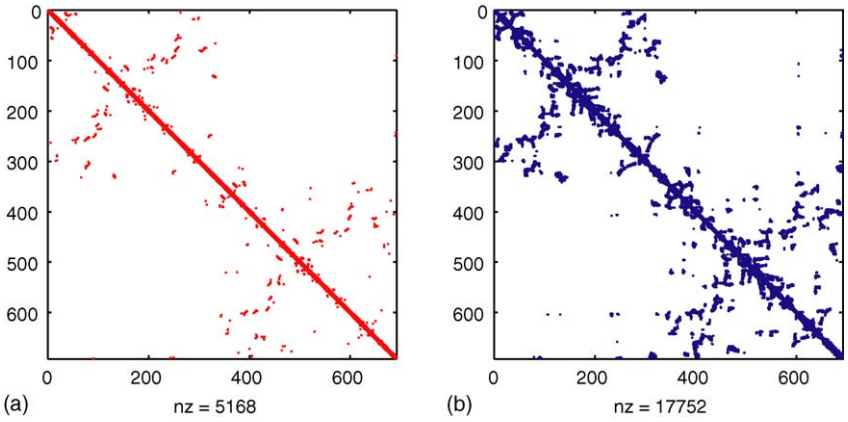

Fig. 6. The sparsity pattern of linking matrices. The linking matrices of the open form of lactoferrin (1LFH) calculated by the bond-cutoff method and the distance-cutoff method are presented, respectively. (a) The linking matrix from the bond-cutoff method. The backbone modeling and the interactions such as hydrogen bonds, salt-bridges, and van der Waals forces ($R_c = 5.0$Å) are considered to build this sparse linking matrix. (b) The linking matrix from the distance-cutoff method. The distance-cutoff R_c is 11.0Å. The total number of connections from the bond-cutoff method (= 5168) is less than half the connection from the distance-cutoff method (= 17,752).

4.4. Elastic network model without virtual springs representing hydrogen bonds and salt-bridges

Hydrogen bonds and salt-bridges play an important role in maintaining folded structures of native proteins, as well as the conformational changes between meta-stable conformations. To demonstrate those roles in the proposed elastic network model, we temporarily exclude the virtual springs which represent hydrogen bonds and salt-bridges from our model so that only the backbone and van der Waals interactions are considered. Fig. 5 presents the overlap results about lactoferrin (1LFH) and calcium ATPase (1SU4). The maximum overlap value for lactoferrin (calcium ATPase) falls down from 0.82 (0.67) to 0.63 (0.52).

We observe two important facts from this experiment. First, protein dynamics is ruled by hydrogen bond and salt-bridge interactions rather than van der Waals interactions. Without modeling the effects of hydrogen bonds and salt-bridges in our model, we cannot capture the nature of proteins well. Second, there exist unnecessary virtual springs in the elastic network model by the distance-cutoff method. When a distance-cutoff over 11Å is applied to a protein, the linking matrix includes most of links corresponding to hydrogen bonds and salt-bridges because those interaction ranges are shorter than the cutoff distance. Why are not the overlap values improved, even getting worse? Although the distance-cutoff method is able to generate a robust elastic network model including all chemical interactions by exaggerating the van der Waals range, such an assumption produces so many unnecessary spring connections in the network model so that the stiffness is overestimated and the generality of the system is destroyed too much. Consequently, the proper distribution of virtual spring connections determined by chemical bond information is more important than the number of springs in the elastic network model.

4.5. Computational cost

The proposed bond-cutoff method usually makes less number of connections than the distance-cutoff method does.

Table 3
The elapsed time to calculate first 40 normal modes

B_c	3	3	3	3	3	3	3	3	0	0	0	0
R_c(Å)	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	11.0	13.0	15.0	17.0
1OMP	1.69	1.68	1.70	1.68	1.78	2.02	2.43	2.56	4.07	8.45	8.00	9.90
1ANF	1.27	1.56	1.66	1.90	1.78	1.87	2.44	2.43	4.09	7.71	8.74	10.24
4AKE	0.61	0.64	0.75	0.76	0.81	0.86	0.91	0.93	2.03	2.78	3.48	4.16
1AKE	0.70	0.72	1.01	0.83	0.91	0.98	1.02	1.07	2.07	3.13	3.98	4.87
1LFH	3.45	3.74	3.66	3.79	4.14	4.14	4.31	5.39	8.63	16.95	26.76	23.93
1LFG	3.53	3.47	3.94	3.73	4.04	3.94	4.71	5.46	8.20	15.31	29.78	26.21
1SU4	5.34	5.33	5.48	6.18	6.82	6.97	6.60	7.55	12.89	19.07	36.84	61.44
1KJU	6.44	6.40	6.62	7.43	7.22	8.88	9.01	9.58	14.32	25.26	50.86	57.51
2LAO	0.93	0.75	0.74	0.92	1.04	1.09	1.05	1.10	2.40	3.72	4.06	4.73
1LST	0.76	0.74	0.91	1.09	1.04	1.18	1.22	1.15	2.53	4.16	4.52	5.36
1ONE	2.09	2.24	2.39	3.28	3.67	2.97	3.08	3.53	5.80	11.58	10.65	13.69
4ENL	2.01	2.08	2.39	2.61	2.49	3.11	3.26	3.67	5.52	10.24	10.05	12.99

The bond-cutoff value zero means the distance-cutoff only. The calculation time by the bond-cutoff model is faster than that of the conventional model as from half to one tenth. We use the Pentium 4 computer with 1.6 GHz CPU, 1 GByte RAM, and Linux operating system. (Unit: second).

Therefore, the computational cost to solve eigenvalue problem of the stiffness matrix is also reduced.

Fig. 6 shows the linking matrices of lactoferrin which is built by the bond-cutoff method and the distance-cutoff method, respectively. The total number of connections by the new method are 5168 when B_c is three and R_c is 5.0Å, whereas the distance-cutoff method creates 17,752 links when R_c is set as 11.0Å. The density of the linking matrix of the new model is less than half that of the conventional model built by the distance-cutoff method. The computation time for NMA is also reduced from 4.83 to 1.50 seconds.

The CPU time for other proteins with various elastic network models is in Table 3. One can observe the computational improvement in all cases. For this simulation, we utilize a PC with 1.6 GHz Pentium 4 processor, 1 GB RAM, and Linux operating system. Function ''eigs'' of the Matlab package has been used to solve the eigenvalue problem [41].

5. Conclusions

In this article, a sparse and more efficient connection rule for α-carbon coarse-grained elastic network models is proposed as an alternative of the conventional distance-cutoff method. The connection rules are basically governed by the chemical bond information. First, the virtual spring connections along three consecutive backbone residues are established. This backbone modeling itself is enough to make the network analytically stable and also explains the reason why the conventional distance-cutoff model needs a cutoff value larger than 11Å in order to perform NMA properly. Second, hydrogen bond pairs are calculated by the HBPLUS software and the corresponding residues are connected to each other. Likewise, the virtual springs for salt-bridges are modeled between residue bases when the distance between ionic atoms is less than 4.0Å. We show that the connections by hydrogen bonds and ionic bonds play an important role in representing protein dynamics with infinitesimal motions set by normal modes. In addition, van der Waals interactions are considered with a radial distance-cutoff model. One can build more reliable elastic network models by varying the van der Waals interaction range from 4.5 to 8.0Å. This relatively short cutoff range is sometimes impracticable in the distance-cutoff method.

As a validation, we perform NMA for six example proteins, each of which has two different conformations (i.e., open and closed). The overlap values are improved in our model and the computational cost is also diminished as compared with the conventional approaches. Consequently, the bond-cutoff method proposed here is physically more accurate and numerically more efficient than the distance-cutoff method because of the significant reduction of unnecessary spring connections in elastic network models.

Acknowledgements

The authors thank Professor G.S. Chirikjian for his contributions including valuable discussions. This work was financially supported by the Department of Energy of Unite States (DE-FG02-04ER25626), and supported by the Korea Research Foundation Grant funded by Korea Government (MOEHRD) (grant no. KRF-2003-214-D00321).

References

[1] M.K. Kim, Elastic network models of biomolecular structure and dynamics. Ph.D. Thesis, Johns Hopkins University, Baltimore, MA, USA, 2004.
[2] D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM, A program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.
[3] D.A. Pearlmanv, D.A. Case, J.W. Caldwell, W.S. Rossv, T.E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kollman, AMBER, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules, Comput. Phys. Commun. 91 (1995) 1–41.
[4] A.D. Schuyler, G.S. Chirikjian, Normal mode analysis of proteins: a comparison of rigid cluster modes with C-alpha coarse graining, J. Mol. Graph. Model 22 (2004) 183–193.
[5] 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.
[6] G.H. Li, Q. Cui, A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca^{2+}-ATPase, Biophys. J. 83 (2002) 2457–2474.
[7] 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.
[8] A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, I. Bahar, Anisotropy of fluctuation dynamics of proteins with an elastic network model, Biophys. J. 80 (2001) 505–515.
[9] K. Hinsen, Analysis of domain motions by approximate normal mode calculations, Proteins 33 (1998) 417–429.
[10] F. Tama, Y.H. Sanejouand, Conformational change of protein arising from normal mode calculations, Protein Eng. 14 (2001) 1–6.
[11] K. Suhre, Y.H. Sanejouand, Elnemo: a normal mode web-server for protein movement analysis and the generation of templates for molecular replacement, Nucl. Acids Res. 32 (2004) 610–614.
[12] M.M. Tirion, Large amplitude elastic motions in proteins from a single-parameter atomic analysis, Phys. Rev. Lett. 77 (1996) 1905–1908.
[13] O. Kurkcuoglu, R.L. Jernigan, P. Doruker, Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions, Polymer 45 (2004) 649–657.
[14] I. Bahar, R.L. Jernigan, Inter-residue protential in globular proteins and the dominance of highly specific hydrophilic interactions at close separation, J. Mol. Biol. 266 (1997) 195–214.
[15] Z. Jiang, L. Zhang, J. Chen, A. Xia, D. Zhao, Effect of amino acid on forming residue-residue contacts in proteins, Polymer 43 (2002) 6037–6047.
[16] M.K. Kim, R.L. Jernigan, G.S. Chirikjian, Efficient generation of feasible pathways for protein conformational transitions, Biophys. J. 83 (2002) 1620–1630.
[17] M.K. Kim, G.S. Chirikjian, R.L. Jernigan, Elastic models of conformational transitions in macromolecules, J. Mol. Graph. Model 21 (2002) 151–160.
[18] M.K. Kim, R.L. Jernigan, G.S. Chirikjian, An elastic network model of HK97 capsid maturation, J. Struct. Biol. 143 (2003) 107–117.
[19] M.K. Kim, W. Li, 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.
[20] D.J. Jacobs, A.J. Rader, L.A. Kuhn, M.F. Thorpe, Protein flexibility predictions using graph theory, Proteins 44 (2001) 150–165.
[21] H. Yan, A.R. Day, M.F. Thorpe, Stability of networks under tension and pressure, Phys. Rev. B 38 (1988) 6876–6880.
[22] R. Sayle, E.J. Milner-White, RasMol: biomolecular graphics for all, Trends Biochem. Sci. 20 (1995) 374.

[23] I.K. McDonald, J.M. Thornton, Satisfying hydrogen-bonding potential in proteins, J. Mol. Biol. 238 (1994) 777–793.
[24] C. Branden, J. Tooze, Introduction to Protein Structure, 2nd ed., Garland, New York, 1999.
[25] E. Martz, Protein explorer: easy yet powerful macromolecular visualization, Trends Biochem. Sci. 27 (2002) 107–109, http://www.proteinexplorer.org.
[26] O. Marques, Y.H. Sanejouand, Hinge-bending motion in citrate synthase arising from normal mode calculations, Proteins 23 (1995) 557–560.
[27] B.H. Oh, J. Pandit, C.H. Kang, K. Nikaido, S. Gokcen, G.F. Ames, S.H. Kim, Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand, J. Biol. Chem. 268 (1993) 11348.
[28] A.J. Sharff, L.E. Rodseth, J.C. Spurlino, F.A. Quiocho, Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis, Biochemistry (US) 31 (1992) 10657.
[29] F.A. Quiocho, J.C. Spurlino, L.E. Rodseth, Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor, Structure 5 (1997) 997.
[30] C.W. Mueller, G.J. Schlauderer, J. Reinstein, G.E. Schulz, Adenylate kinase motions during catalysis, an energetic counterweight balancing substrate binding, Structure 4 (1996) 147–156.
[31] C.W. Muller, G.E. Schulz, Structure of the complex between adenylate kinase from *Escherichia coli*and the inhibitor Ap5A refined at 1.9 angstrom resolution. A model for a catalytic transition state, J. Mol. Biol. 224 (1992) 159–177.
[32] T.M. Larsen, J.E. Wedekind, I. Rayment, G.H. Reed, A carboxylate oxygen of the substrate bridges the magnesium ions at the active site of enolase: structure of the yeast enzyme complexed with the equilibrium mixture of 2-phosphoglycerate and phosphoenolpyruvate at 1.8 Åresolution, Biochemistry (US) 35 (1996) 4349–4358.
[33] L. Lebioda, B. Stec, Crystal structure of holoenzyme refined at 1.9 Åresolution: trigonal-bipyramidal geometry of the cation binding site, J. Am. Chem. Soc. 111 (1989) 8511.
[34] G.E. Norris, B.F. Anderson, E.N. Baker, Molecular replacement solution of the structure of apolactoferrin, a protein displaying large-scale conformational change, Acta Crystallogr. B 47 (1991) 998.
[35] C. Toyoshima, M. Nakasako, H. Nomura, H. Ogawa, Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Åresolution, Nature 405 (2000) 647.
[36] C. Xu, W.J. Rice, W. He, D.L. Stokes, A structural model for the catalytic cycle of Ca^{2+}-ATPase, J. Mol. Biol. 316 (2002) 201.
[37] W.G. Krebs, V. Alexandrov, C.A. Wilson, N. Echols, H.Y. Yu, M. Gerstein, Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic, Proteins 48 (2002) 682–695.
[38] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P. Bourne, The protein data bank, Trends Biochem. Sci. 28 (2000) 235–242.
[39] B.F. Anderson, G.E. Norris, S.V. Rumball, D.H. Thomas, E.N. Baker, Lactoferrin: structure and function. Chapter A. Comparison of the three-dimensional structures of human lactoferrin in its iron free and iron saturated forms, in: Advances in Experimental Medicine and Biology, Plenum Press, New York, 1994, pp. 227–230.
[40] J.M. Berg, L. Stryer, J.L. Tymoczko, Biochemistry, 5th ed., W.H. Freeman and Company, 2002.
[41] Matlab, http://www.mathworks.com.