electronic reprint

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449
Editors: **E. N. Baker and Z. Dauter**

A method for finding candidate conformations for molecular replacement using relative rotation between domains of a known structure

Jay I. Jeong, Eaton E. Lattman and Gregory S. Chirikjian

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A method for finding candidate conformations for molecular replacement using relative rotation between domains of a known structure

Jay I. Jeong,[a] Eaton E. Lattman[b] and Gregory S. Chirikjian[c]*

[a]School of Mechanical and Automotive Engineering, Kookmin University, Seoul 136702, South Korea, [b]Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, USA, and [c]Department of Mechanical Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA

Correspondence e-mail: gregc@jhu.edu

Received 2 June 2005
Accepted 18 January 2006

This paper presents a methodology to obtain candidate conformations of multidomain proteins for use in molecular replacement. For each separate domain, the orientational relationship between the template and the target structure is obtained using standard molecular replacement. The orientational relationships of the domains are then used to calculate the relative rotation between the domains in the target conformation by using pose-estimation techniques from the field of robotics and computer vision. With the angle of relative rotation between the domains as a cost function, iterative normal-mode analysis is used to drive the template structure to a candidate conformation that matches the X-ray crystallographic data obtained for the target conformation. The selection of the correct intra-protein domain orientations from among the many spurious maxima in the rotation function (including orientations obtained from domains in symmetry mates rather than within the same copy of the protein) presents a challenge. This problem is resolved by checking R factors of each domain, measuring the absolute value of relative rotation between domains, and evaluating the cost value after each candidate conformation is driven to convergence with iterative NMA. As a validation, the proposed method is applied to three test proteins: ribose-binding protein, lactoferrin and calcium ATPase. In each test case, the orientation and translation of the final candidate conformation in the unit cell are generated correctly from the suggested procedure. The results show that the proposed method can yield viable candidate conformations for use in molecular replacement and can reveal the structural details and pose of the target conformation in the crystallographic unit cell.

1. Introduction

Molecular replacement (MR; Rossmann, 1972, 1990, 2001) often fails for multidomain protein structures because of the flexibility of the structures. Most of this flexibility is concentrated in the linking regions between domains. If large conformational changes occur in the protein, the phase information of the template protein cannot be adopted as that of the target protein. This is often the case in ligand-bound proteins even if the sequences of the template and the target are identical (Suhre & Sanejouand, 2004*b*). Thus, it would be useful to have methods to 'morph' the template structure into candidate conformations which more closely match the X-ray data when MR cannot be applied directly, specifically in the case of multidomain proteins.

The MR method can be used with separate subunits to find the crystal structure of multidomain proteins (Cygler & Anderson, 1988*a*,*b*; Bernstein & Hol, 1997). Because each domain remains more or less rigid during conformational

changes, the position and orientation of each domain can sometimes be obtained and assembled into the whole structure. Practically, however, the translation function often fails to find the exact position of each corresponding domain, even though the rotation function can frequently find the exact orientation of the domain (Rossmann, 1990; Giacovazzo *et al.*, 1998). Brünger proposed Patterson correlation refinement, which can adjust the flexible regions between domains of a multidomain protein and find a better conformation for use in computation of the translation function (Brünger, 1990, 1997). This method can also be used with the orientation of subunits, each of which is separately obtained by direct rotation-function evaluation, to improve the search (DeLano & Brünger, 1995). However, Brünger also reported that this method is limited by the radius of convergence. In the present work, we propose a new method to substantially deform the template conformation into the target conformation based on iterative normal-mode analysis and the relative rotation between domains.

Many of the largest conformational changes in multidomain proteins appear to arise from rigid-body motion between domains: flexible regions between domains dominate conformational changes while the domains remain rigid. For instance, in the case of lactoferrin, which has an open form (PDB code 1lfh) and a closed form (PDB code 1lfg), the root-mean-square deviation (r.m.s.d.) between these two conformations is 6.4 Å when comparing their α-carbon traces. In contrast, the r.m.s.d. value in each corresponding domain is less than 0.6 Å (Norris *et al.*, 1991). This also means that the collective motions of domains dominate the conformational changes, not the fluctuations of each residue.

Normal-mode analysis (NMA) can be used to predict the conformational changes of multidomain proteins by calculating the collective motions. Using NMA, harmonic motions of a given protein structure around an equilibrium conformation (Brooks *et al.*, 1995; Hinsen, 1998; Moritsugu & Kidera, 2004) and collective motions and dynamic fluctuations of given protein structures (Bahar *et al.*, 1997; Atilgan *et al.*, 2001; Li & Cui, 2002, 2004; Kurkcuoglu *et al.*, 2004; Schuyler & Chirikjian, 2004) can be calculated. The collective motions obtained by NMA can represent the dominant motion of any given structure, which is also related to conformational changes (Marques & Sanejouand, 1995; Tama & Sanejouand, 2001; Tama *et al.*, 2004; Suhre & Sanejouand, 2004*b*; Schuyler & Chirikjian, 2005). Krebs *et al.* (2002) also reported that a few low-frequency normal modes dominate the conformational change for about 50% of the protein structures deposited in the Protein Data Bank (Berman *et al.*, 2000).

Recently, Suhre and Sanejouand proposed a novel method to obtain candidate structures for MR by using NMA (Suhre & Sanejouand, 2004*a*,*b*). After calculating normal-mode shapes from the elastic network model (Tirion, 1996; Bahar *et al.*, 1997; Atilgan *et al.*, 2001; Hinsen, 1998; Moritsugu & Kidera, 2004; Jeong *et al.*, 2006), they applied arbitrary scale factors to the conformational deviations calculated from NMA. By superimposing these deviations onto the original conformation, they obtained candidate conformations. MR is then performed for all candidate conformations in order to judge the exact conformation from among the candidates by NMA. Their method successfully revealed the crystal structures of several proteins: maltodextrin-binding protein, HIV-1 protease and glutamine-binding protein. When a large or complex conformational change is to be analyzed, however, linear amplification of modes may not be sufficient to deform the template into the target conformation. Since the normal modes guarantee only infinitesimal deviation of a given protein structure, excessive amplification can produce physically unrealistic candidate conformations for the target protein.

In order to make large conformational changes and at the same time to avoid nonphysical conformations, we use iterative normal-mode analysis in this article (Tama *et al.*, 2004; Hinsen *et al.*, 2005). After obtaining an intermediate conformation by adding small motions from NMA, another NMA is applied to the newly obtained conformation until the target conformation is approached. At each iteration, the amplitude and directions of the participating normal modes should be determined in order to drive the template structure to the target, since the normal-mode shapes can only be obtained without amplitude and direction. Thus, a cost function for choosing the proper conformation is required to judge the closest conformation from the candidate conformations at each step. Tama and coworkers applied an iterative normal-mode method to morph the known protein structure into a low-resolution electron-density map from the cryo-EM method (Tama *et al.*, 2004). They iterated with a gradient search about the intermediate structure to match the target electron density. Here, we propose a different procedure of iterative NMA by using a result from the statistical mechanics of macromolecules (Kim, 2004).

In this work, the relative rotation between domains is used as a cost function to drive iterative NMA. After obtaining the rotation function of each domain by applying existing MR software to each separate domain, we convert the rotation function of each domain into the relative rotation between the domains in the target conformation using pose-estimation methods from the field of robotics and computer vision (Chirikjian & Kyatkin, 2001). Using the relative rotation between domains, we can drive iterative NMA and obtain good candidates which are similar to the target protein structure. The candidate conformations can then be used in MR to match the diffraction pattern of the target protein. By morphing the template conformation into the target conformation using only the rotation function, our method makes it easier to find the translation function for candidate conformations of the whole protein.

We explain our methodology in §2. In this section, we present the elastic network modeling and iterative normal-mode methods and a cost function for driving iterative NMA with relative rotation between domains. In §3, we demonstrate the results produced by the proposed method for three protein structures: ribose-binding protein, lactoferrin and calcium ATPase, which have open and closed forms. We discuss the results and methods in §4. Finally, we conclude this work in §5.

2. Methods

2.1. Elastic network modeling

Normal modes of a given protein structure can be calculated based on the elastic network model, which analyzes an equivalent mass-spring system in static equilibrium as a representation of the protein structure (Brooks *et al.*, 1995). If only a few lowest normal modes are required, the simplified α-carbon coarse-grained elastic network model can be used, in which only C^{α} atoms are used to represent the corresponding residues (Tirion, 1996; Bahar *et al.*, 1997; Kim, Chirikjian *et al.*, 2002; Kim, Jernigan *et al.*, 2002). The simplified elastic network model is frequently used for many applications because of its cost-effective features in numerical calculation (Atilgan *et al.*, 2001; Kim, Jernigan *et al.*, 2003; Kim, Li *et al.*, 2003; Kim *et al.*, 2005; Bahar & Jernigan, 1997; Kurkcuoglu *et al.*, 2004). A Hookean pairwise potential for the simplified elastic network model can be written as

$$c(\boldsymbol{\delta}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_{i,j}(\|\mathbf{x}_i + \boldsymbol{\delta}_i - \mathbf{x}_j - \boldsymbol{\delta}_j\| - R_{i,j})^2, \qquad (1)$$

where $\mathbf{x}_i$ and $\boldsymbol{\delta}_i$ are the position and the deviation of the ith α-carbon. Here, $R_{i,j}$ is the relative distance between the ith and jth residues. The normal modes can be obtained by solving the eigenvalue/eigenvector problem associated with (1).

In this work, we use bond-cutoff connection rules in order to build an elastic network model (Jeong *et al.*, 2006). This method guarantees the stability of the elastic network model with any distance-cutoff value (Yan *et al.*, 1988). Based on backbone modeling, we apply the distance-cutoff method to model interactions between C^{α} atoms which are not located sequentially but within a distance cutoff. Different distance-cutoff (R_c) values are used for intradomain and interdomain residues, respectively. A 10 Å distance-cutoff value is used for intradomain residues, whereas a 5 Å distance-cutoff value is used for interdomain residues. This strategy increases the rigidity of each domain of a test protein structure. The non-rigid-body motion calculated from NMA then mostly reflects the relative motion between domains with more connections inside the domain.

2.2. Iterative procedure to simulate large conformational changes

When calculating the collective harmonic motions for a protein structure using NMA, only normal-mode shapes without magnitude and with plus or minus directional ambiguity can be obtained. Thus, these two properties of the collective motion should be determined in order to drive the template conformation towards the target. The magnitude of each normal mode is determined by using ideas from the statistical mechanics of protein structures. According to Kim (2004), the root-mean-square fluctuation of each normal mode is inversely proportional to the frequency of the given mode. Thus, we multiply this inverse frequency by each corresponding normal mode to obtain the collective motions of protein structures. This procedure emphasizes the lowest non-rigid normal modes over the other normal modes.

Next, we determine the direction at each step from the template to the goal. Because we have already set the amplitude, we only have to determine the direction of the normal mode (plus or minus) for the intermediate conformation to approach the target. This procedure can be performed by setting the cost function which represents the conformational changes and reducing its value to zero (see the following section for details of the cost function). We present the algorithm of the iterations as a flow chart in Fig. 1.

When defining the ith intermediate conformation as $\mathbf{X}_i$, its deviation $\boldsymbol{\Delta}_{i,k}$ calculated by the kth normal mode can be given as

$$\boldsymbol{\Delta}_{i,k} = \mathbf{v}_{i,k}/\omega_{i,k}, \qquad (2)$$

where $\omega_{i,k}$ is the natural frequency and $\mathbf{v}_{i,k}$ is the mode shape of the kth normal mode after the ith iteration, respectively.

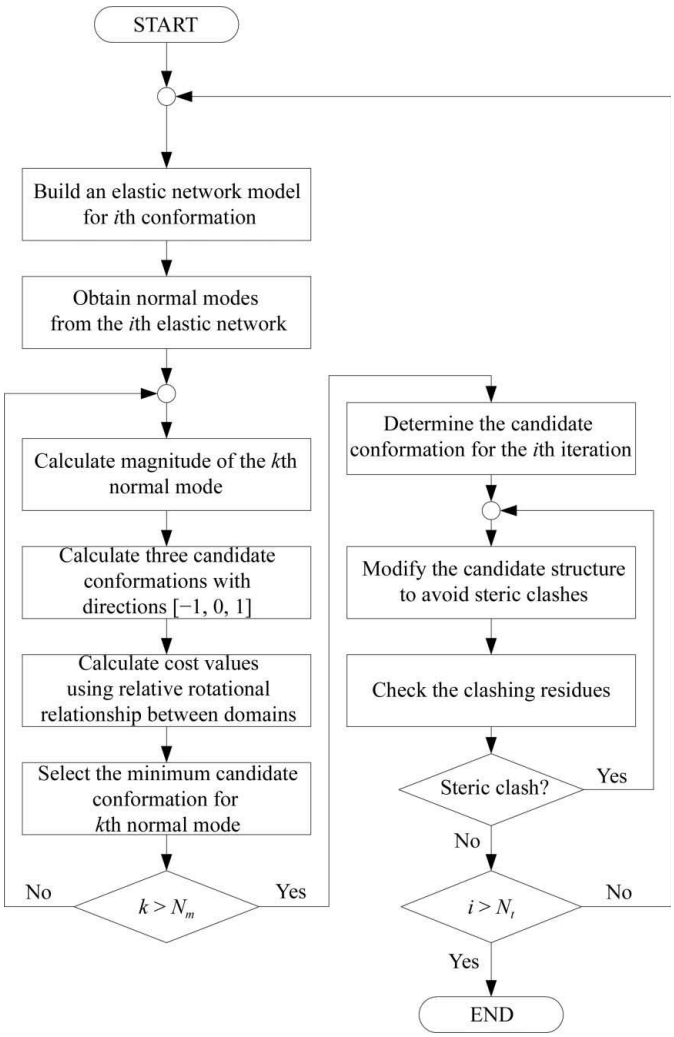

Figure 1
Flowchart of iterative normal-mode analysis. By using normal-mode shape and the relative rotation between domains, the template conformation can be morphed to near the target. Moreover, the elastic network minimization algorithm is used to retain the geometric constraints of the protein structure during iteration.

The index i is the iteration number and takes values from 1 to the maximum number of iterations N_n. k is the index of non-rigid-body normal modes calculated from one to N_m.

The direction of this deviation should then be decided. We set the weight value w_j as -1, 0 or 1 to represent the plus or minus direction or to exclude the specific normal mode from the perturbation. This means that there are three candidate conformations for each normal mode. At the ith iteration, the frequency ω_{i-1} and mode $\mathbf{v}_{i-1}$ can be calculated by solving an eigenvalue/eigenvector problem. The candidate conformations $\mathbf{C}_{i,j,k}$ can then be derived as

$$\mathbf{C}_{i,j,k} = \mathbf{X}_{i-1} + \varepsilon w_j \mathbf{\Delta}_{i-1,k}, \tag{3}$$

where $\mathbf{X}_{i-1}$ is the $(i-1)$th intermediate conformation. The parameter ε is a dimensionless scaling factor to adjust the magnitude of all the normal modes. After the candidate conformations have been calculated at the ith iteration and the kth normal mode, three candidate conformations with weight values can be obtained.

The cost functions of these three conformations should be evaluated in order to select the closest candidate conformation to the target. After obtaining the direction which minimizes the cost function, the candidate conformation $\mathbf{C}_{i,j_{\min},k}$ can be determined for the kth normal mode. We then repeat the same procedure for the $(k+1)$th normal mode. This procedure can be performed for all normal modes under consideration. When the minimum cost function search for all the normal modes up to N_m is finished, the ith pathway conformation $\mathbf{X}_i$ can be obtained.

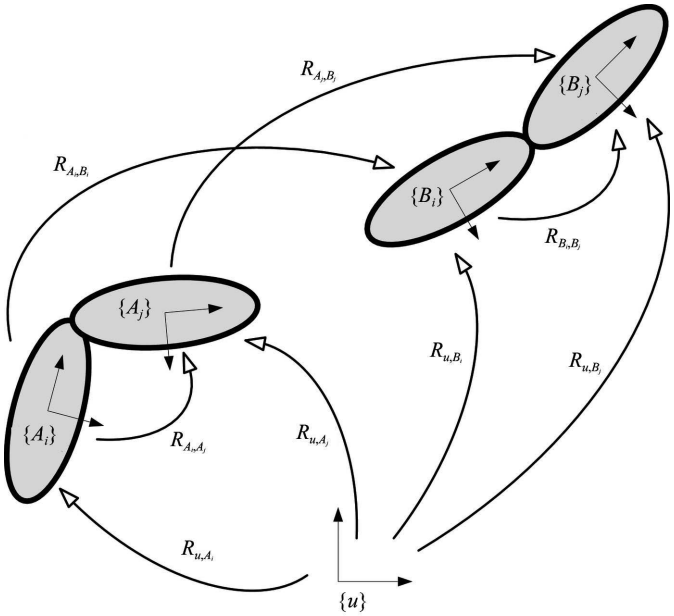

Figure 2
The relationships between relative rotations. $\{u\}$ is the global coordinate frame, which is arbitrarily placed. $\{A_i\}$ and $\{B_i\}$ are the coordinate frames attached to the ith domain in the template and the target, respectively. Our goal is to calculate R_{B_i,B_j}, the relative rotation between domains in the target, by using the rotations R_{u,A_i}, R_{u,A_j}, R_{A_i,B_i} and R_{A_j,B_j}. R_{u,A_i} and R_{u,A_j} can be obtained from the structural details of the template conformation. R_{A_i,B_i} and R_{A_j,B_j} can be calculated by converting the relative rotations H_{A_i,B_i} and H_{A_j,B_j} described in the global frame $\{u\}$ into the relative rotations with respect to the $\{A_i\}$ local frame.

2.3. Keeping geometric constraints using elastic network minimization

The geometric constraints of a protein structure should be retained during the iterative procedures of NMA. When driving the template structure to the target, at least several iterative steps of NMA are required. However, these multiple procedures may disturb the geometric constraints of an intermediate conformation such as preserving virtual bond lengths of the backbone trace and avoiding steric clashes between domains. This arises from the deviation error from linearization of NMA, which can accumulate during the iterations even if it is very small. In addition, 'sticking' between domains is another problem. Along the intermediate pathways two or more domains may pass each other very closely; for example, sliding motions between two domains. However, if two domains are too close and excessive virtual springs are established between these domains, then these will prevent domain motion perpendicular to these surfaces and NMA cannot produce the proper normal mode at the next iteration.

To overcome such a hurdle, we apply the elastic network minimization algorithm to maintain geometric constraints during the iterations. In (1), we replace the current relative distance $R_{i,j}$ between the ith and the jth residues with $L_{i,j}$, which is a predefined distance between the residues such that

$$c(\boldsymbol{\delta}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_{i,j}(\|\mathbf{x}_i + \boldsymbol{\delta}_i - \mathbf{x}_j - \boldsymbol{\delta}_j\| - L_{i,j})^2, \tag{4}$$

where

$$L_{i,j} = \begin{cases} L_{\min} & \text{if } R_{i,j} < L_{\min} \\ R^0_{i,j} & \text{if } |i-j| = 1 \\ R_{i,j} & \text{otherwise} \end{cases}.$$

$R^0_{i,j}$ is the relative distance of the template structure between the ith and jth residues. The minimum deviation of all residues δ can then be found by differentiating (4) (Kim, Chirikjian *et al.*, 2002; Kim, Jernigan *et al.*, 2002). This procedure should be repeated for the candidates at each iteration of iterative NMA until the abnormal relative distances between C^α atoms no longer exist.

2.4. Cost function using relative rotation between domains

2.4.1. Calculating interdomain rotational relationships of the target. To obtain the relative rotation between domains, the rotation function of each domain should first be calculated separately using molecular replacement. To perform this, we use program packages or web services such as *AMoRe* (Navaza, 1994, 2001) and *MolRep* (Vagin & Teplyakov, 1997) at the CaspR homepage (Claude *et al.*, 2004) and the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). We use the rotation functions for the test proteins after calculating the rigid-body refinement algorithm (Castellano *et al.*, 1992) in *AMoRe* and *MolRep*. That is, the

candidate rotation functions are filtered by the rigid-body refinement algorithm of *AMoRe* and *MolRep*.

As a result of MR, several highest peaks in the rotation function can be obtained which are the most likely to reflect the true orientation of the corresponding domain in the target unit cell. Moreover, additional candidate rotations should be considered, since it cannot be known which copy of the asymmetric unit in the unit cell corresponds to the peaks in the rotation function obtained from MR. Thus, all crystallographically symmetric copies of the rotation-function peak must be checked to calculate the relative rotation between domains. We define the candidate orientation from the rotation function of the ith domain between the template and the target as

$$\mathbf{H}^{(p)}_{A_i,B_i} = \mathbf{\Psi}_p \mathbf{H}^{(0)}_{A_i,B_i}, \tag{5}$$

where $\mathbf{H}^{(0)}_{A_i,B_i}$ is the rotation matrix of a peak calculated by MR, p is the index of crystallographic symmetry of the target unit cell and $\mathbf{\Psi}_p$ is the rotation matrix from crystallographic symmetry.

By using pose-estimation methods (Chirikjian & Kyatkin, 2001), we can derive the relative orientations between domains in the 'target' unit cell. Assigning the template conformation as A and the target as B, our goal is to derive the relative rotation between the ith and jth domains in the target. The relative rotation between these two domains in the target can be derived as

$$\mathbf{R}_{B_i,B_j} = \mathbf{R}^T_{A_i,B_i}\mathbf{R}^T_{u,A_i}\mathbf{R}_{u,A_j}\mathbf{R}_{A_j,B_j}, \tag{6}$$

where u is the global coordinate and $\mathbf{R}_{B_i,B_j}$ is the relative rotation between the ith and the jth domain of the target protein B. We present these relationships between rotations in Fig. 2.

It is then only necessary calculate $\mathbf{R}_{A_i,B_i}$ and $\mathbf{R}_{A_j,B_j}$ from $\mathbf{H}_{A_i,B_i}$ and $\mathbf{H}_{A_j,B_j}$, which can be calculated from MR. Here, $\mathbf{H}_{A_i,B_i}$ is a relative rotation of the ith domain from conformation A to B as viewed in the global coordinate frame $\{u\}$, whereas the relative rotation $\mathbf{R}_{A_i,B_i}$ represents the same relative rotation as described in the $\{A_i\}$ local frame. The relationship between these two relative rotations can thus be written by conjugation with respect to $\mathbf{R}^T_{u,A_i}$,

$$\mathbf{R}_{A_i,B_i} = \mathbf{R}^T_{u,A_i}\mathbf{H}_{A_i,B_i}\mathbf{R}_{u,A_i}. \tag{7}$$

Using (7), (6) can be converted into

$$\mathbf{R}_{B_i,B_j} = \mathbf{R}^T_{u,A_i}\mathbf{H}^T_{A_i,B_i}\mathbf{H}_{A_j,B_j}\mathbf{R}_{u,A_j}. \tag{8}$$

After substituting (5) into (8), the final equation

$$\begin{aligned}\mathbf{R}^{(p)}_{B_i,B_j} &= \mathbf{R}^T_{u,A_i}[\mathbf{\Psi}_m\mathbf{H}^{(0)}_{A_i,B_i}]^T[\mathbf{\Psi}_n\mathbf{H}^{(0)}_{A_j,B_j}]\mathbf{R}_{u,A_j}\\ &= \mathbf{R}^T_{u,A_i}{\mathbf{H}^{(0)}_{A_i,B_i}}^T\mathbf{\Psi}^T_m\mathbf{\Psi}_n\mathbf{H}^{(0)}_{A_j,B_j}\mathbf{R}_{u,A_j}\\ &= \mathbf{R}^T_{u,A_i}{\mathbf{H}^{(0)}_{A_i,B_i}}^T\mathbf{\Psi}_p\mathbf{H}^{(0)}_{A_j,B_j}\mathbf{R}_{u,A_j}\end{aligned} \tag{9}$$

can be obtained, where $\mathbf{\Psi}^T_m\mathbf{\Psi}_n = \mathbf{\Psi}_p$ because the crystallographic symmetry is a space group. As a result of (9), several candidates of relative rotation which have the same number of protein molecules in the target unit cell can be calculated. Thus, a method to select the exact relative rotation is required.

2.4.2. Finding the corresponding pairs of rotation matrices. From the result of the previous section, many candidates can be obtained for relative rotations between each domain pair of the target. These candidate conformations can be checked by using all of the relative orientation candidates to drive iterative NMA. Each candidate conformation is then submitted to an MR program such as *AMoRe*, which produces a structure solution and an R factor and correlation coefficient. The conformation with the lowest R factor is accepted.

Another way to judge the proper relative rotation between domains is to drive iterative NMA using relative rotation

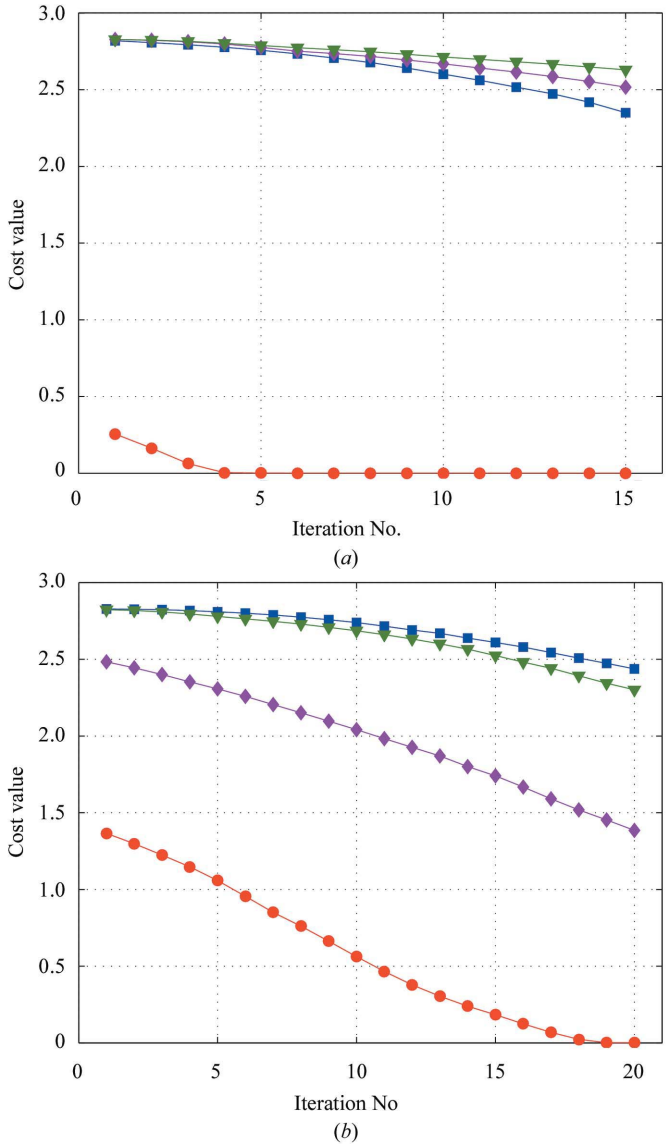

Figure 3
The change in the cost value when each domain pair is tested separately with iterative NMA. The open form of lactoferrin (1lfh) is tested as the template structure and the closed form (1lfg) as the target. We tested four cases of relative rotations for each target rotation. For each domain pair, only one case succeeds in converging. (*a*) Four relative rotations as the target rotation between domain C1 and domain N1; (*b*) four relative rotations as the target rotation between domain C1 and domain N2.

candidates for each domain pair. This method is based on the fact that the conformations of the given protein structure must have physically possible structures in three-dimensional space. When driving the initial conformation into a candidate with an incorrect relative rotation relationship, iterative NMA cannot achieve the final conformation owing to steric clashes between domains and the cost function will not settle to a small value. In Fig. 3, we present the cost values during iterations when trying to drive the initial conformation of lactoferrin (1lfh) into possible candidates with each rotation relationship. This test is separately performed for each domain pair without elastic network minimization: *i.e.* four candidates for each interdomain relative rotation in the target. In the figure, only one case converges to a near-zero value after iteration. In contrast, the candidate conformations in incorrectly matched cases cannot converge by iterative NMA and retained over half their initial cost value. From these results, it can be observed that this provides a tool to weed out such rotations.

Another method that can be considered to judge the proper domain pair is statistical analysis of protein motion. Kreb and Gerstein reported that the maximum relative orientation of conformational change by a hinge motion is 150° by surveying their 'Molecular Movement Database' (Krebs & Gerstein, 2000). This means that hinge motion of greater than 150° will be statistically rare. In this work, we use this maximum angle as a criterion to judge the proper domain pairs. Cases in which the angular difference exceeds 150° can easily be excluded from the set of feasible candidates of relative rotation between domain pairs.

2.4.3. Rotational metric as a cost function. To drive iterative NMA, it is necessary to evaluate a cost value that represents the orientational difference between corresponding domain pairs in the current and target conformations. In this work, the following rotational metric to evaluate the cost values $C_{i,j}$ is used,

$$\begin{aligned} C_{i,j} &= d(\mathbf{R}_{A_i,A_j}, \mathbf{R}_{B_i,B_j}) \\ &= [6 - 2\,\mathrm{trace}(\mathbf{R}^T_{A_i,A_j}\mathbf{R}_{B_i,B_j})]^{1/2} \\ &= 2(1 - \cos\theta_{i,j})^{1/2}, \end{aligned} \tag{10}$$

where the function 'trace(·)' is the trace of a matrix. $\theta_{i,j}$ is the rotation angle between the two relative rotations $\mathbf{R}_{A_i,A_j}$ and $\mathbf{R}_{B_i,B_j}$. This represents the minimum rotational difference between two coordinate frames measured about the unique axis of rotation (Chirikjian & Kyatkin, 2001). Hence, the cost function reduces its value towards zero when an intermediate conformation approaches the target. In the iterative NMA algorithm, we use the cost value from the proposed cost function to decide the normal-mode direction.

Table 1
The properties of the test proteins ribose-binding protein, lactoferrin and calcium ATPase.

We also present the r.m.s.d. values as the result of iterative NMA and the *R* factor and correlation coefficient after applying MR with the final conformations.

Protein type	Ribose-binding protein	Lactoferrin	Calcium ATPase
Template PDB code	1urp	1lfg	1su4
Target PDB code	2dri	1lfh	1t5s
Space group	$P2_12_12_1$	$P2_12_12_1$	$C2$
No. of residues	271	691	994
No. of domains	2	4	4
Initial r.m.s.d. (Å)	4.1	6.4	14.0
Domain range	N (1–103, 236–264), C (104–235, 265–271)	N1 (1–91, 251–339), N2 (92–250), C1 (340–434, 595–691), C2 (435–594)	M (1–124, 240–343, 751–994), A (125–239), P (344–360, 600–750), N (361–599)
R.m.s.d. of the final candidate (Å)	0.7	1.4	5.0
R factor	47.3	48.7	55.2
Correlation coefficient	45.5	50.3	47.7

3. Results

To validate the proposed methodology, we tested three proteins which have open and closed forms: ribose-binding protein, lactoferrin and calcium ATPase. The 15 lowest non-rigid-body normal modes obtained from the elastic network model as described in §2 are used as collective motions in order to morph the template protein structures. The bond-cutoff value (B_c) is set to 3, $R_{\rm intra}$ to 10 Å and $R_{\rm inter}$ to 5 Å to build the elastic network for all test proteins. We obtained the domain information for all test proteins from 3*Dee* and *SCOP* (Dengler *et al.*, 2001; Murzin *et al.*, 1995). Four points inside each domain which tends to remain rigid during conformational change were selected to calculate the relative rotations between domains in the template structure. We present the PDB codes, the residue numbers defining each domain and the initial r.m.s.d. values between the template and target in Table 1.

3.1. Lactoferrin

As the first test protein, we tested lactoferrin, which has 691 residues and four domains labeled C1, C2, N1 and N2. Using a hinge motion, it grabs and transports ligands such as Fe^{3+}, Mg^{2+} and Mn^{2+} ions. We set the open form (1lfh) as the target conformation and the closed form (1lfg) as the template. The sequences of both structures are identical and the r.m.s.d. value between the two conformations is 6.43 Å (Norris *et al.*, 1991).

Firstly, we execute MR using *AMoRe* to find the rotation function of each domain in the target unit cell. We present the results of MR in Table 2. In all MR procedures with each separate domain, the best candidate rotation of each domain can easily be determined, since the correlation factors of the highest peak are much higher than the other candidates. The four candidate rotations for each domain can then be calculated using the crystallographic symmetry of the target. With these first candidates for the rotation matrix, one can attempt to drive the template conformation.

Table 2
The resulting rotation-function peaks when each separate domain of the open form of lactoferrin 1lfh is applied to MR.

Each separate domain of the closed form 1lfg is used as a template conformation for MR. The *AMoRe* program is used to calculate the value of the rotation function. α, β and γ are *ZYZ* Euler angles and *a*, *b* and *c* are fractions of the target unit-cell edges.

Domain		α (°)	β (°)	γ (°)	*a*	*b*	*c*	CC_F	RF_F	CC_I
N1	1	107.73	88.93	248.20	0.4626	0.1298	0.1018	31.7	51.8	30.4
	2	107.69	89.86	247.96	0.4625	0.1303	0.1015	31.7	51.8	30.5
	3	72.33	90.10	67.86	0.0374	0.1302	0.3985	31.6	51.8	30.5
	4	70.50	60.33	71.48	0.4255	0.1327	0.0076	26.9	53.2	26.9
N2	1	20.14	82.04	73.19	0.3916	0.4216	0.1502	31.8	2.1	32.4
	2	19.79	83.30	73.52	0.3914	0.4215	0.1503	31.8	52.0	32.4
	3	9.98	6.83	9.85	0.2899	0.1956	0.3893	28.0	52.8	26.9
	4	69.61	79.63	71.07	0.3361	0.3840	0.3209	28.0	52.9	27.7
C1	1	108.85	81.19	246.58	0.2194	0.2237	0.0211	41.4	49.0	41.1
	2	71.15	92.05	67.55	0.2803	0.2236	0.4786	32.2	52.4	33.6
	3	110.48	48.34	26.99	0.2459	0.1645	0.4763	29.2	53.1	27.7
	4	89.47	32.67	97.60	0.2528	0.2905	0.1375	29.6	53.3	30.6
C2	1	109.58	82.15	246.88	0.2298	0.3925	0.3390	37.0	49.2	34.5
	2	64.29	61.84	83.03	0.2037	0.0357	0.1553	27.4	52.9	27.5
	3	70.81	91.42	66.33	0.2723	0.3788	0.1702	27.9	52.8	27.0
	4	1.01	23.53	201.21	0.2745	0.2325	0.4318	27.1	52.6	26.7

Since the crystallographic symmetry of the target protein is $P2_12_12_1$, there are four copies of the protein structure in the unit cell. Moreover, at least three domain pairs should be considered for lactoferrin because it has four domains. Hence, it is necessary to check the proper relative rotation for each domain among the four candidates. Thus, 12 test runs should be required with each relative rotation when applying iterative NMA for one domain pair separately and checking the possibility of the domain orientation. The C1 domain is selected as the baseline from which the relative rotation to the other domains is considered in the cost function because its highest peak has the highest correlation factor.

Alternatively, the proper relative rotations can be selected without driving iterative NMA, by observing the initial relative rotation for two of three domain pairs. That is, the proper rotation can be found if one of the initial difference angles is much smaller than the others. In Table 3, one of the initial difference angles is calculated as 1.3°, with which iterative NMA should drive the C2 domain based on the C1 domain, whereas the angle differences for the other cases are around 180°. This means that a candidate with a relative rotation of 1.3° between the C1 and C2 domains is already placed near its goal (*i.e.* near 0°), but the other candidates are not. Hence, if the conformational changes between the corresponding domains are very small in the target, one does not need to drive iterative NMA. This procedure can be applied to the relative rotation between the C1 and N1 domains. The smallest relative rotation between the domains is 8.0°, whereas the initial angular differences of the other candidates are near 180°.

Thus, it is only necessary to check the four relative rotations between the C1 and N2 domains with the two relative rotations determined above. Two of the four cases in this domain pair can be excluded according to the maximum angle difference between the template and the target. That is, the initial relative rotation is greater than 150° when *p* is 1 or 4. We therefore need to check the final cost value for two candidates of relative rotations $R^{(2)}_{C1,N2}$ and $R^{(3)}_{C1,N2}$ when driving iterative NMA with the relative rotations of $R^{(1)}_{C1,C2}$ and $R^{(1)}_{C1,N1}$. Consequently, the cost value can converge only with $R^{(3)}_{C1,N2}$. When we use $R^{(2)}_{C1,N2}$ with relative rotations $R^{(1)}_{C1,C2}$ and $R^{(1)}_{C1,N1}$, the final cost value does not converge but retains its value of 2.45. In Fig. 4, the conformational changes of the final candidate are presented when the proper relative rotation pair is chosen. The opening hinge motion of the second domain can be observed and the final r.m.s.d. value for this case is 1.38 Å.

With this final candidate, we tried the MR procedure with the diffraction pattern of 1lfh. Since we need to reconstruct a full-atom model from an α-carbon trace from iterative NMA, we use the *Maxsprout* program (Holm & Sander, 1991) and *Deepview*/*SwissPDBViewer* (Guex & Peitsch, 1997) to relocate the side chains of the candidate. As a result of MR, the final correlation coefficient (CC_F) is 50.3 and the final *R* factor (RF_F) is 48.7%.

3.2. Ribose-binding protein

As the second case, we applied iterative NMA to the ribose-binding protein (RBP). RBP consists of 271 residues grouped into two domains. We used 1urp as the open form and 2dri as the closed form (Bjorkman & Mowbray, 1998; Bjorkman *et al.*, 1994); the two structures change by a hinge-bending motion (Echols *et al.*, 2003) and the r.m.s.d. value between them is 4.06 Å. We set the open form (1urp) as the template and the closed form (2dri) as the target.

Firstly, we calculated the rotation function with the separated N and C domains using *AMoRe*. As a result, the peaks which have higher correlation coefficient values CC_F than the others can be picked out. In the case of the N domain, the highest four peaks are around 35.8; the others are less than 26.3. Similar results can also be observed in the case of the C domain. The highest four peaks are 42.8, with the others being less than 26.3. We set the highest peaks for both domains as the rotation function to calculate the relative rotation.

Since the crystallographic symmetry of the target is $P2_12_12_1$, four relative rotations between these two domains should be

Table 3
The initial angle differences (°) between domain pairs of all test proteins.

The proper relative rotation pair can be selected by checking the initial values. In the case of lactoferrin, the domains pairs can be selected as those with initial angle differences near zero. Domain pairs for which the relative rotation exceeds 150° can be disregarded.

	p	1	2	3	4
Lactoferrin (1lfg to 1lfh)	C1 to N1	8.0	172.1	179.1	179.0
	C1 to N2	173.1	129.6	53.9	163.4
	C1 to C2	1.3	179.0	180.0	179.2
RBP (1urp to 2dri)	N to C	156.2	154.1	41.3	159.5
Calcium ATPase (1su4 to 1t5s)	M to A	41.7	139.5		
	M to P	13.4	178.4		
	M to N	121.8	86.1		

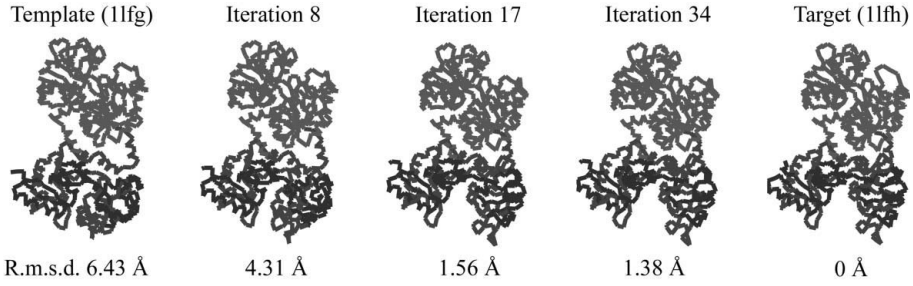

Figure 4
Selected intermediate conformations in the case of lactoferrin when applying iterative NMA from 1lfg to 1lfh. The r.m.s.d. value between each pathway and target conformation is given below each conformation. The final r.m.s.d. is 1.38 Å.

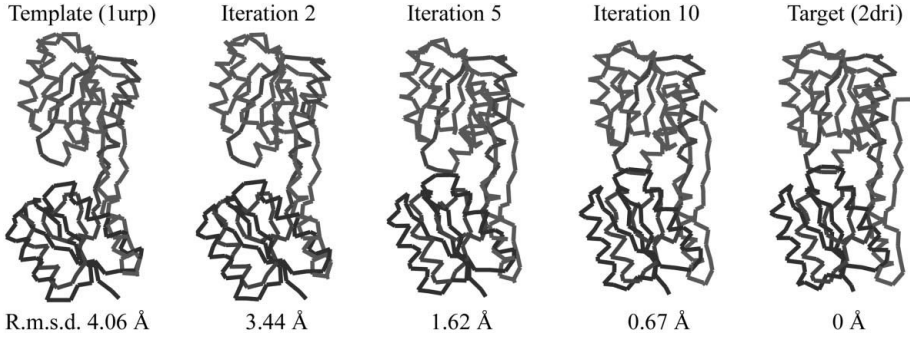

Figure 5
Selected intermediate conformations during iteration of ribose-binding protein from the open form (1urp) to the closed form (2dri). The relative rotation between the N and C domains is used as the cost value of the iterative NMA. The final r.m.s.d. value is 0.67 Å.

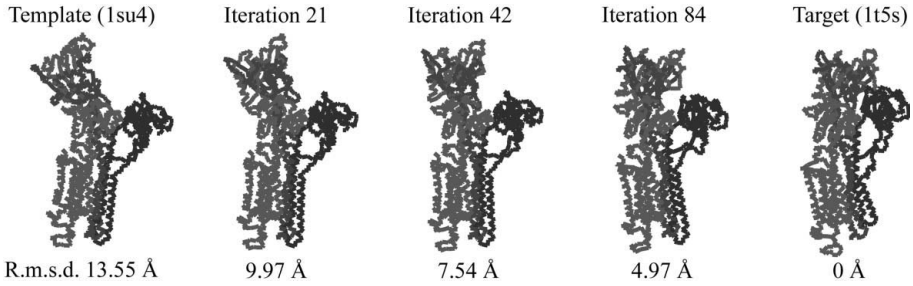

Figure 6
Selected intermediate conformations during iteration of calcium ATPase from the open form (1su4) to the closed form (1t5s). The relative rotations between the M, N, P and A domains are used as the cost function of the iterative NMA. The final r.m.s.d. value is 4.97 Å.

checked to select the proper target cost function. We present the initial angle differences of four candidate relative rotations between these two domains in Table 3. As in the lactoferrin case, only one candidate relative rotation can be obtained when p is 3. The other cases have rotations of over 150°, which has been reported to be a rare case for relative rotation (Krebs & Gerstein, 2000). When p is 3, the candidate conformation converges to the target structure after ten iterations and its final cost value falls to below 0.1 from an initial cost value of 1.0. The r.m.s.d. value between the candidate and the target is 0.67 Å.

We depict several intermediate conformations during the iteration from the template to the target in Fig. 5. The relative hinge motions of the two domains can be observed. We reconstructed the side chains of the final candidate conformation using *Maxsprout* and *Deepview* and executed the MR procedure with *AMoRe*. The final R factor is 47.3 and the correlation coefficient is 45.5.

3.3. Calcium ATPase

As the final test protein, we selected calcium ATPase (1su4 and 1t5s), which is the largest protein structure in this work (Toyoshima *et al.*, 2000; Sorensen *et al.*, 2004). The protein consists of 994 residues and four domains labeled the M, A, P and N domains. The open form 1su4 is set as the template to obtain a candidate for the closed form 1t5s. To drive iterative NMA, we used six domain pairs: $R^{(p)}_{\mathrm{M,P}}$, $R^{(p)}_{\mathrm{M,A}}$, $R^{(p)}_{\mathrm{M,N}}$, $R^{(p)}_{\mathrm{A,N}}$, $R^{(p)}_{\mathrm{A,P}}$ and $R^{(p)}_{\mathrm{P,N}}$.

Firstly, we tried to calculate the rotation function for each domain using *AMoRe*. Each domain was then separated and used for MR without flexible regions such as hinge residues. In the case of the M domain, we only used α-helices M4–M10, which correspond to residues 247–343 and 751–994, to obtain the orientation of the domain (Toyoshima *et al.*, 2000). However, we could not obtain distinct peaks in the cases of the A and P domains and the *AMoRe* program did not yield a rigid-body refinement solution for these two

Table 4
The MR result of calcium ATPase from 1su4 to 1t5s using *MolRep* and *AMoRe*.

Peaks can be distinguished as a result of the rotation function calculated by *MolRep*, while *AMoRe* cannot reveal clear differences between peaks.

	MolRep					*AMoRe*					
	α (°)	β (°)	γ (°)	RotF	RotF/σ	α (°)	β (°)	γ (°)	CC_F	RF_F	CC_I
Domain M											
1	277.05	15.08	12.10	825.8	4.83	281.12	16.59	11.03	31.7	58.7	41.0
2	15.36	10.32	190.35	813.4	4.76	282.33	18.50	11.41	31.3	58.8	40.7
3	299.43	51.58	24.42	799.6	4.68	326.59	41.80	17.25	30.8	58.8	38.8
4	354.56	16.48	207.06	719.0	4.21	309.29	23.15	354.25	30.6	58.9	40.3
5	58.03	17.59	150.54	710.6	4.16	329.52	41.64	16.98	30.4	59.1	38.3
Domain A											
1	208.61	39.70	96.04	3396	4.76	333.03	26.55	97.54	32.5	57.7	38.1
2	68.95	66.51	131.32	2820	3.95	209.51	41.95	94.83	32.5	57.7	37.4
3	12.89	20.32	73.06	2807	3.94	37.18	31.58	359.14	32.2	58.0	39.6
4	175.99	49.55	53.16	2794	3.92	23.74	35.29	336.23	32.2	57.9	39.1
5	335.36	29.50	195.34	2792	3.92	60.33	55.15	312.09	32.2	57.9	38.7
Domain P											
1	268.02	27.86	18.00	2503	4.29	271.50	26.61	14.22	33.7	57.7	40.2
2	269.29	24.28	18.97	2487	4.26	268.75	29.46	14.81	33.6	57.8	40.6
3	83.70	74.86	325.09	2406	4.12	274.62	25.20	13.91	33.6	57.7	40.4
4	233.99	17.51	16.40	2224	3.81	276.50	25.65	13.50	33.5	57.7	40.6
5	15.60	56.84	108.75	2179	3.73	231.00	16.50	21.50	33.5	57.9	41.5
Domain N											
1	312.20	84.36	238.16	2038	9.34	312.15	84.77	237.64	36.7	55.9	41.2
2	228.38	90.00	58.73	953.4	4.37	226.99	90.00	56.23	32.6	57.5	37.5
3	45.05	80.44	188.35	925.2	4.24	226.00	90.00	53.75	32.2	57.6	37.2
4	348.01	65.83	202.04	874.5	4.01	69.66	56.65	258.94	31.7	58.1	38.8
5	168.46	53.70	163.46	857.3	3.93	329.21	71.90	218.50	31.5	57.9	38.2

domains. Thus, we applied the *MolRep* program, which can calculate the real-space rotation function with the separated domains. We present the results of the rotation function using *AMoRe* and *MolRep* in Table 4. From the results, we selected the candidate rotation functions for domains as the highest peak from *MolRep*.

With the selected rotation functions, candidates of the relative rotations with respect to the crystallographic symmetry can be obtained. Since the crystallographic symmetry of the target protein (1su4) is *C*2, only two relative rotations had to be checked per domain pair. The possible relative rotation between domains of the target conformation can then be selected by observing the initial relative rotation between the corresponding domains. The second case between the M and P domains $R^{(2)}_{M,P}$ could be excluded because it was near 180°, whereas $R^{(1)}_{M,P}$ was about 13°. This rule could also be applied to the relative rotation $R_{M,P}$. However, we had to check the relative rotation between the M and N domains. Because $R^{(1)}_{M,N}$ corresponds to 121.8° and $R^{(2)}_{M,N}$ to 83.1°, we have to drive iterative NMA with both relative rotations. When we executed iterative NMA with $R^{(1)}_{M,N}$, however, the cost function did not converge to zero. Finally, we can choose only one candidate which converges to near zero: $R^{(1)}_{M,A}$, $R^{(1)}_{M,P}$ and $R^{(2)}_{M,N}$. The final r.m.s.d. value and the shape of the conformation are presented in Fig. 6. After 84 iterations, the conformation converges and the cost value does not reduce any further. The final r.m.s.d. value of the candidate conformation from the target conformation is 4.97 Å.

With the final candidate for the 1su4 conformation, we applied *Maxsprout* to rebuild the backbone chain from α-carbon traces. The final correlation coefficient and *R* factor from *AMoRe* were 47.7 and 55.2%, respectively. The *R* factor of the final candidate conformation is relatively large in comparison with the results on lactoferrin and ribose-binding protein. We discuss this problem in the next section.

4. Discussion

When driving iterative NMA from the template structure to the target, we do not use the translation function but only use the relative rotation between domains. Six degrees of freedom need to be known to describe the position and orientation of a rigid body in three-dimensional space. However, even though we only use the relative rotation between domains, the final candidate conformations of the two tested globular proteins, *i.e.* lactoferrin and RBP, are very close to their targets: the r.m.s.d. values between the target and the final candidate conformation are only 1.38 and 0.67 Å, respectively. This result means that the conformational changes of multidomain protein structures can be described with less than six degrees of freedom. This is because each domain is connected to the others by a short flexible linker, which acts as a constraint on conformational change.

However, this is not the case in calcium ATPase. The final candidate for the closed form of calcium ATPase 1t5s has a larger r.m.s.d. value (4.97 Å) and *R* factor (55.2%) than lactoferrin and RBP. The r.m.s.d. value and final *R* factor of calcium ATPase arise from the relative translation of domain A as depicted in Fig. 7. Why can domain A of calcium ATPase not be placed into the proper position, while the other domains in RBP, lactoferrin and even the M, N and P domains of calcium ATPase can be? One difference of domain A from

the others is that it connects to the M domain, which is a membrane domain and is relatively more flexible than the globular domain (Toyoshima *et al.*, 2000; Sorensen *et al.*, 2004). Moreover, it connects to domain M by a long flexible region, which can give more than three degrees of freedom to domain A. Thus, even after assigning three angular constraints to domain A, translation is still possible and domain A can remain in one position while it rotates. In contrast, the other domains, which connect to each other *via* short hinge regions and can only move by rotational hinge motion, can have no more than three degrees of freedom for each domain connected. Thus, in the case of membrane proteins such as the calcium pump, translation should be considered to deal with the relative translation between domains.

In most tests, translation functions for separate domains could not be obtained by *AMoRe* or *MolRep*. In the case of finding the rotation function, the highest peaks are distinct from the others and match the proper orientations of the corresponding domain. However, the translation function for each separate domain often places the domain in the wrong position (Rossmann, 1990; DeLano & Brünger, 1995; Giacovazzo *et al.*, 1998). In the case of calcium ATPase, while we can calculate the rotation function for domain P and A of calcium ATPase, we cannot obtain the correct translation function. In contrast, we could find the translation function for an entire molecule of calcium ATPase after driving iterative NMA and finding the final candidate for MR. This result means that the proposed method can be used as a preconditioning procedure for multidomain proteins, which can be followed by successful translation-function computation.

When the rotation function is calculated using programs such as *AMoRe* and *MolRep*, the orientational error might be included in each domain. In the case of lactoferrin, the relative orientation of the N2 domain, which is calculated from MR, is 53.9°. This value is different from the value of 56.1° which is calculated by placing four points inside the C1 and N2 domains and comparing their orientation in the template and the target conformation. Because the relative rotation is used as a cost function between domains, the orientational error can influence the final candidate conformation. To check this influence, we ran the simulation on lactoferrin and ribose-binding protein (data not shown). Even though there is an angular difference of about 2° between the domains, the change in the final r.m.s.d. value is negligible: from 1.37 to 1.38 Å. The difference for ribose-binding protein is also negligible. The angular differences between the N and C domains are calculated to be 41.3 and 43.3° by MR and by calculating the relative rotation with four points inside the domain, respectively. However, the final r.m.s.d.s are little changed: 0.7 and 0.8 Å, respectively.

Next, we checked the geometric constraints of the candidate conformation, which should be retained after driving the template conformation into the final candidate conformation. The main reason for using an iterative procedure is to prevent physically unrealistic conformations during iteration. From observations of the initial and final conformations of our test proteins, we find that the maximum deviation of the relative distance between adjacent backbone residues is of the order of 0.01 Å in the test case from 1su4 to 1t5s and of the order of 0.001 Å in the cases of lactoferrin and RBP. These results arise from the elastic network minimization algorithm, which recovers the geometric constraints after the deviation from NMA is added to the current conformation. Thus, each virtual bond length does not itself change much, but retains its original value during iteration. In addition, the algorithm can also guarantee a candidate conformation without steric clashes between domains. Since the minimum distance between C^{α} atoms in the different domains is set to 4.0 Å and the conformation is updated until it contains no steric clashes, the iteration procedure can always maintain this minimum distance during the iterations.

As a final issue, we discuss possible methods for cases when the search model or the template conformation do not yield distinguishable peaks for the rotation function of the corresponding domain. Firstly, as used in the case of calcium ATPase, an attempt can be made to obtain the rotation function with various modern methodologies such as the real-space rotation function, direct rotation function and Patterson refinement (Brünger, 1990, 1997; DeLano & Brünger, 1995). Secondly, all peaks from the rotation function can be tried as a cost function for use with iterative NMA. The cost value can be checked for each possible conformation, whether it reduces the cost value or not. If there were multiple candidates which satisfy the cost-value threshold, then MR can be applied to these final candidates and the *R* factor checked to find the exact conformation in the target crystal. In the second method, the number of possible candidates could be very large, since the crystallographic copies in the unit cell

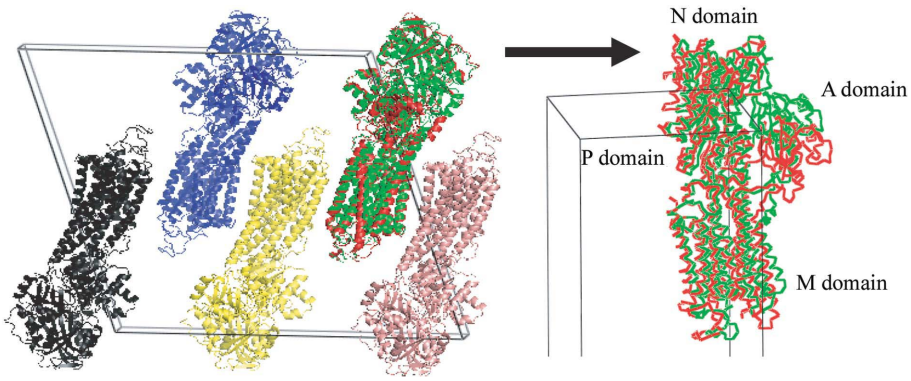

Figure 7
Comparison between crystal packing of the target conformation (1t5s) and the position and orientation of the final candidate conformation morphed from 1su4 after applying MR. Although the N, P and M domains are determined properly, domain A has positional error, while the orientation of the domain is nearly correct. The target conformation is presented in green and the final candidate conformation in red.

also should be considered. For this, great computational power will be needed. However, to quote Jones (2001),

> computer time is now virtually free in most cases

and should no longer be a barrier because of the powerful computational ability of cheap PCs or workstations for MR.

5. Conclusion

In this work, we have presented a method to find candidate conformations of multidomain proteins for use in molecular replacement. We suggested iterative normal-mode analysis to generate candidate structures. We modeled the protein structure as an elastic network model, in which all C^{α} atoms (as representatives of each residue) are treated as point masses with the interactions between them modeled as springs. Using the elastic network model, we calculated the 15 lowest normal modes and applied iterative procedures in order to obtain physically realistic conformations that approximate the final structure closely.

As a cost function to drive iterative NMA, we used the relative rotation between domains of the target protein structure and each candidate. The rotation function of each domain can be calculated by using MR programs such as *AMoRe* and *MolRep*. We then converted the peaks in the rotation function of each domain into candidate relative orientations between domains in the target structure by using pose-estimation techniques from the field of robotics and computer vision.

As a validation of the proposed method, we tested three proteins which have open and closed forms: ribose-binding protein, lactoferrin and calcium ATPase. We showed that the template conformation could be morphed close to the target with the relative rotation cost function. In the cases of ribose-binding protein and lactoferrin, we could obtain the proper candidate conformations with less than 1.5 Å r.m.s.d. from their target structures. In the case of calcium ATPase, we could derive the final candidate conformation with 4.97 Å r.m.s.d. The proper orientation and position of the M, N and P domains of the final candidate conformation could be obtained by iterative NMA. However, domain A has positional error even though the relative rotation between domains could be matched correctly. The source of the error was discussed and denoted as additional degrees of freedom of domain A.

Consequently, the proposed method can extend the application of MR methods of multidomain proteins and can calculate the final candidate conformation without any radius of convergence only if the rotation function for subunits of a multidomain protein can be revealed.

The authors appreciate all the anonymous reviewers who provided valuable comments to improve this work. JJ thanks Moonki, Jinseob, Wooram, Kiju, Adam, Yan, Aris, Georgios, Whitney, Yu Zhou and the many collaborators who have worked with JJ for valuable discussion and with warm support. This work was financially supported by the US Department of Energy (DE-FG02-04ER25626), by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2003-214-D00321), by research program 2006 of Kookmin University in Korea and under NIH grant R01GM075310 'Group-Theoretic Methods in Protein Structure Determination'.

References

Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O. & Bahar, I. (2001). *Biophys. J.* **80**, 505–515.
Bahar, I., Atilgan, A. R. & Erman, B. (1997). *Fold. Des.* **2**, 173–181.
Bahar, I. & Jernigan, R. L. (1997). *J. Mol. Biol.* **266**, 195–214.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. (2000). *Trends Biochem. Sci.* **28**, 235–242.
Bernstein, B. E. & Hol, W. G. J. (1997). *Acta Cryst.* D**53**, 756–764.
Bjorkman, A. J., Binnie, R. A., Zhang, H., Cole, L. B., Hermodson, M. A. & Mowbray, S. L. (1994). *J. Biol. Chem.* **269**, 30206–30211.
Bjorkman, A. J. & Mowbray, S. L. (1998). *J. Mol. Biol.* **279**, 651–664.
Brooks, B. R., Janezic, D. & Karplus, M. (1995). *J. Comput. Chem.* **16**, 1522–1542.
Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.
Brünger, A. T. (1997). *Methods Enzymol.* **276**, 558–580.
Castellano, E. E., Oliva, G. & Navaza, J. (1992). *J. Appl. Cryst.* **25**, 281–284.
Chirikjian, G. S. & Kyatkin, A. (2001). *Engineering Applications of Noncommutative Harmonic Analysis: With Emphasis on Rotation and Motion Groups.* Boca Raton, FL, USA: CRC Press.
Claude, J., Suhre, K., Notredame, C., Claverie, J. & Abergel, C. (2004). *Nucleic Acids Res.* **32**, W606–W609.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cygler, M. & Anderson, W. F. (1988*a*). *Acta Cryst.* A**44**, 38–45.
Cygler, M. & Anderson, W. F. (1988*b*). *Acta Cryst.* A**44**, 300–308.
DeLano, W. L. & Brünger, A. T. (1995). *Acta Cryst.* D**51**, 740–748.
Dengler, U., Siddiqui, A. S. & Barton, G. J. (2001). *Proteins*, **42**, 332–344.
Echols, N., Milburn, D. & Gerstein, M. (2003). *Nucleic Acids Res.* **31**, 478–482.
Giacovazzo, C., Manna, L., Siliqi, D. & Rizzi, M. B. (1998). *Acta Cryst.* A**54**, 617–625.
Guex, N. & Peitsch, M. (1997). *Electrophoresis*, **18**, 2714–2723.
Hinsen, K. (1998). *Proteins*, **33**, 417–429.
Hinsen, K., Reuter, N., Navaza, J., Stokes, D. L. & Lacaperey, J.-J. (2005). *Biophys. J.* **88**, 818–827.
Holm, L. & Sander, C. (1991). *J. Mol. Biol.* **218**, 183–194.
Jeong, J. I., Jang, Y. & Kim, M. K. (2006). *J. Mol. Graph. Model.* **24**, 296–306.
Jones, D. T. (2001). *Acta Cryst.* D**57**, 1428–1434.
Kim, M. K. (2004). PhD thesis. Johns Hopkins University, Baltimore, MA, USA.
Kim, M. K., Chirikjian, G. S. & Jernigan, R. L. (2002). *J. Mol. Graph. Model.* **21**, 151–160.
Kim, M. K., Jernigan, R. L. & Chirikjian, G. S. (2002). *Biophys. J.* **83**, 1620–1630.
Kim, M. K., Jernigan, R. L. & Chirikjian, G. S. (2003). *J. Struct. Biol.* **143**, 107–117.
Kim, M. K., Jernigan, R. L. & Chirikjian, G. S. (2005). *Biophys. J.* **89**, 43–55.
Kim, M. K., Li, W., Shapiro, B. A. & Chirikjian, G. S. (2003). *J. Biomol. Struct. Dyn.* **21**, 395–405.
Krebs, W. G., Alexandrov, V., Wilson, C. A., Echols, N., Yu, H. Y. & Gerstein, M. (2002). *Proteins*, **48**, 682–695.
Krebs, W. G. & Gerstein, M. (2000). *Nucleic Acids Res.* **28**, 1665–1675.

Kurkcuoglu, O., Jernigan, R. L. & Doruker, P. (2004). *Polymer*, **45**, 649–657.
Li, G. H. & Cui, Q. (2002). *Biophys. J.* **83**, 2457–2474.
Li, G. H. & Cui, Q. (2004). *Biophys. J.* **86**, 743–763.
Marques, O. & Sanejouand, Y. H. (1995). *Proteins*, **23**, 557–560.
Moritsugu, K. & Kidera, A. (2004). *J. Phys. Chem. B*, **108**, 3890–3898.
Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.
Norris, G. E., Anderson, B. F. & Baker, E. N. (1991). *Acta Cryst.* B**47**, 998–1004.
Rossmann, M. G. (1972). Editor. *The Molecular Replacement Method.* New York: Gordon & Breach.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Rossmann, M. G. (2001). *Acta Cryst.* D**57**, 1360–1366.
Schuyler, A. D. & Chirikjian, G. S. (2004). *J. Mol. Graph. Model.* **22**, 183–193.
Schuyler, A. D. & Chirikjian, G. S. (2005). *J. Mol. Graph. Model.* **24**, 46–58.
Sorensen, T. L.-M., Moller, J. V. & Nissen, P. (2004). *Science*, **304**, 1672–1675.
Suhre, K. & Sanejouand, Y. (2004*a*). *Nucleic Acids Res.* **32**, 610–614.
Suhre, K. & Sanejouand, Y. (2004*b*). *Acta Cryst.* D**60**, 796–799.
Tama, F., Miyashita, O. & Brooks, C. L. III (2004). *J. Mol. Biol.* **337**, 985–999.
Tama, F. & Sanejouand, Y. (2001). *Protein Eng.* **14**, 1–6.
Tirion, M. M. (1996). *Phys. Rev. Lett.* **77**, 1905–1908.
Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. (2000). *Nature (London)*, **405**, 647–655.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Yan, H., Day, A. R. & Thorpe, M. F. (1988). *Phys. Rev. B*, **38**, 6876–6880.