Biophys Rev (2017) 9:877–893
https://doi.org/10.1007/s12551-017-0330-2

REVIEW

Normal mode analysis as a method to derive protein dynamics information from the Protein Data Bank

Hiroshi Wako[1] · **Shigeru Endo**[2]

Received: 14 June 2017 / Accepted: 4 October 2017 / Published online: 4 November 2017

Abstract Normal mode analysis (NMA) can facilitate quick and systematic investigation of protein dynamics using data from the Protein Data Bank (PDB). We developed an elastic network model-based NMA program using dihedral angles as independent variables. Compared to the NMA programs that use Cartesian coordinates as independent variables, key attributes of the proposed program are as follows: (1) chain connectivity related to the folding pattern of a polypeptide chain is naturally embedded in the model; (2) the full-atom system is acceptable, and owing to a considerably smaller number of independent variables, the PDB data can be used without further manipulation; (3) the number of variables can be easily reduced by some of the rotatable dihedral angles; (4) the PDB data for any molecule besides proteins can be considered without coarse-graining; and (5) individual motions of constituent subunits and ligand molecules can be easily decomposed into external and internal motions to examine their mutual and intrinsic motions. Its performance is illustrated with an example of a DNA-binding allosteric protein, a catabolite activator protein. In particular, the focus is on the conformational change upon cAMP and DNA binding, and on the communication between their binding sites remotely located from each other. In this illustration, NMA creates a vivid picture of the protein dynamics at various levels of the structures, i.e., atoms, residues, secondary structures, domains, subunits, and the complete system, including DNA and cAMP. Comparative studies of the specific protein in different states, e.g., apo- and holo-conformations, and free and complexed configurations, provide useful information for studying structurally and functionally important aspects of the protein.

Keywords Elastic network model · Protein structure network · Catabolite activator protein · Full-atom system · Decomposition into internal and external motions

Introduction

Protein folding and structure–function relationships are major challenges in molecular biology, biophysics, and other related fields. The three-dimensional structural data deposited in the Protein Data Bank (PDB) (Berman et al. 2003) have played an important role in investigating these problems. However, the information extracted from the PDB contains mainly static structural features. It is generally recognized that both the dynamic and static aspects of the protein structures are necessary to fully understand these problems. The only dynamics-related data provided in the PDB is a crystallographic temperature factor. This provides some information on the fluctuations of individual atoms, but also reflects various problems specific to a crystal, such as crystalline disorder and crystal contacts with neighboring molecules. Furthermore, as most of the PDB data are obtained based on fluctuations as isotropic motions, such data lack directional information on the fluctuations. As the temperature factor is not sufficient for resolving most of the protein dynamics problems, it is necessary to perform computer simulations, such as molecular dynamics

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12551-017-0330-2) contains supplementary material, which is available to authorized users.

✉ Hiroshi Wako
wako@waseda.jp

[1] School of Social Sciences, Waseda University, Tokyo 169-8050, Japan

[2] Department of Physics, School of Science, Kitasato University, Sagamihara 252-0373, Japan

(MD) and normal mode analysis (NMA), to acquire more information about protein dynamics from the PDB data.

MD and NMA are useful methods for characterizing various dynamic aspects of biological macromolecules. MD is more accurate than NMA. Although NMA is based on the normal mode vibrations that are defined as simple harmonic oscillations around an energy minimum, MD can encompass a significantly greater volume of the conformational space. Consequently, the performance of NMA is limited to investigating fluctuations around a specific conformation, and cannot reveal a significant conformational change that may occur during MD simulations. Nonetheless, many investigations have demonstrated that the results obtained from NMA are capable of characterizing the dynamic aspects of biomolecules well. Notably, it has been shown that some of the low-frequency normal modes are strongly correlated with the large-amplitude conformational changes in proteins that have been observed upon ligand binding (Tama and Sanejouand 2001; Krebs et al. 2002; Reuter et al. 2003; Alexandrov et al. 2005; Bahar and Rader 2005; Tobi and Bahar 2005; Dobbins et al. 2008; Wako and Endo 2011). Furthermore, time-averaged properties, in particular the root mean square (RMS) fluctuation of an atom averaged over all normal modes and over time, have been shown to be well correlated with the temperature factor provided in the PDB data (Gō et al. 1983; Brooks and Karplus 1983; Levitt et al. 1985).

NMA has some advantages compared to MD. NMA can be carried out considerably more quickly than MD, and the properties calculated in NMA are analytically well defined. This makes it possible to systematically perform NMA for many PDB structures. Using this advantage, several databases and online servers delivering NMA have been developed to reveal the dynamic features of proteins (Echols et al. 2003; Suhre and Sanejouand 2004; Wako et al. 2004; Yang et al. 2005, 2006; Skjærven et al. 2014; Tiwari et al. 2014; Eyal et al. 2015). ProMode, ProMode-Oligomer, and ProMode-Elastic are examples of such databases developed by the authors (Wako et al. 2004; Wako and Endo 2012, 2013). ProMode and ProMode-Oligomer provided our results obtained by orthodox NMA, but they have now been closed. ProMode-Elastic is available at PDBj (https://pdbj.org; Kinjo et al. 2017). Although it is typically difficult to derive necessary and sufficient information about protein dynamics using only NMA, NMA can be a powerful method for preliminary surveys before more accurate investigations such as MD simulations are undertaken.

In this review, we will describe NMA using dihedral angles as independent variables, which are characteristic of our NMA calculations, and then illustrate what kind of information can be derived from the PDB data using a specific example.

Theory and methods

General theory of normal modes

The theory of normal modes gives a complete analytical solution to the equation of motion for a molecular system subject to the assumption that a given conformation is at the conformational energy minimum and the energy surface in the vicinity of the minimum is a multidimensional parabola. In this assumption, the conformational energy E can be expressed as a quadratic function of the variables to define a conformation of molecules, q_i:

$$E = \frac{1}{2}\sum_{i,j} F_{ij}\Delta q_i \Delta q_j, \tag{1}$$

where Δq_i is the deviation of the ith variable from the minimum point and F_{ij} is the second derivative of E with respect to q_i and q_j at the minimum, $\partial^2 E/\partial q_i \partial q_j$. The variables $\boldsymbol{q} = \{q_i\}$ are generalized, and can be atomic Cartesian coordinates, translation and rotation variables, dihedral angles, or any other variables to define a molecular system.

The kinetic energy T is also approximated as a quadratic function of the velocities $\dot{q}_i$ with the coefficient matrix $\mathbf{H}$:

$$T = \frac{1}{2}\sum_{i,j} H_{ij}\dot{q}_i\dot{q}_j, \tag{2}$$

Then, the normal modes of vibration and their frequency are obtained by solving the generalized eigenvalue equation:

$$\mathbf{HA\Lambda} = \mathbf{FA}, \tag{3}$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal element $\Lambda_{ii} = \omega_i^2$ is the frequency of the ith normal mode and $\mathbf{A}$ provides the variations of variables for individual normal mode vibrations. Various time-averaged properties, such as atomic fluctuations, dihedral angle fluctuations, and correlations between these properties, can be analytically formulated with $\mathbf{\Lambda}$ and $\mathbf{A}$, and, thus, numerically calculated easily.

As for variables to define a conformation of a molecule, most researchers on NMA adopt the Cartesian coordinate system (CCS). In this review, however, we will focus on the dihedral angle system (DAS) because it is not familiar to the researchers. We will clarify the differences between the two systems.

Elastic network model-based NMA

Conformational energy analysis of a protein molecule using dihedral angles as independent variables was developed by Scheraga's group at Cornell University in the 1970s (Momany et al. 1975). The developed program was named ECEPP. Based on ECEPP, the authors developed the program

to perform conformational energy minimization and NMA, termed as FEDER (Wako and Gō 1987; Wako et al. 1995). However, we experienced problems in extending FEDER to a system that included molecules other than a polypeptide. The all-atom force fields were provided mainly for the proteins in ECEPP. They were not available for other molecules such as DNA or various ligands in the PDB data.

In such a situation, the elastic network model-based NMA (ENM-NMA) opened a new era (Bahar et al. 1997, 2010; Hinsen 1998) in the field of NMA including both CCS and DAS. It replaced standard potentials that were used in an orthodox NMA by a simple pairwise Hookean potential:

$$E = \sum_{i,j} k\left(d_{ij} - d_{ij}^{\mathrm{PDB}}\right)^2, \tag{4}$$

where d_{ij} and d_{ij}^{PDB} are distances between the atoms i and j in the calculated and PDB conformations, respectively, and k is a spring constant. There are some options for defining k. We adopted the form:

$$k = c\exp\left\{-\left(\frac{d_{ij}^{\mathrm{PDB}}}{a}\right)^2\right\}, \tag{5}$$

where a and c are constants determined independently of the atomic types of i and j. In this formulation, the spring constant k decreases with the increasing distance between the atoms. It is possible to set a cutoff distance to neglect the interaction between remote atoms. We usually used $c = 1$ and $a = 5$ Å. The ENM-NMA results do not strongly depend on the functional form or the parameter values.

A coarse-grained model is used in most ENM-NMA calculations, where each residue is represented by a Cα atom only. In contrast, we adopted a full-atom system in the ENM-NMA computational program named PDBETA (Wako and Endo 2013). In practice, the full-atom system can be adopted, because a much smaller number of independent variables are required in DAS than in CCS (discussed below). Consequently, PDBETA is applicable to a system involving DNA, RNA, and ligand molecules without transforming them into a coarse-grained model.

Furthermore, ENM-NMA frees the conformational energy minimization that is necessary to perform the orthodox NMA in advance. This is because the PDB structure is considered as the energy-minimum conformation in ENM-NMA. In fact, very precise energy minimization of a PDB structure required in orthodox NMA is the most time-consuming process. In addition, the problem that the energy-minimum conformation sometimes deviates significantly from the PDB structure is resolved in ENM-NMA. In spite of such a sweeping approximation, many studies have shown that ENM-NMA can predict global conformational changes associated with the biological functions of the proteins (Tama and Sanejouand 2001; Krebs et al. 2002; Alexandrov et al. 2005; Tobi and Bahar 2005; Dobbins et al. 2008; Bahar et al. 2010; Skjærven et al. 2011; Wako and Endo 2011).

Dihedral angle system

The NMA method used in ProMode, ProMode-Oligomer, and ProMode-Elastic developed by the authors is characterized by the use of dihedral angles as independent variables, and its application to a full-atom system of any molecule including not only proteins but also DNA, RNA, and ligand molecules in the PDB data. Whereas an amino acid residue is represented by a Cα atom in most NMA calculations, we have been focusing on the use of the full-atom system for properly accommodating any molecules in the PDB data without coarse-graining them.

We have used a molecular structure model with fixed bond-length and bond-angle geometry. In this system, an independent variable is a dihedral angle of a rotatable covalent bond rather than the Cartesian coordinates used in most NMA applications. If more than one molecule exists in a system, however, six additional variables (i.e., translational and rotational variables) are necessary for each additional molecule. Therefore, it is possible to involve water molecules in a system. However, we have removed them to keep the number of variables small.

DAS has an advantage with respect to the number of variables to define the molecular conformations, compared to CCS. In the case of a protein molecule, the number of variables in DAS is approximately one-eighth of that in CCS for a full-atom system. Because memory size requirements, and, thus, computing time, are proportional to the square or to the cube of the number of variables, this is critical, particularly for huge systems. Another advantage is that DAS naturally involves the chain connectivity that plays an important role in determining the folding pattern of a polypeptide chain, whereas CCS is required to take measures to maintain the chain connectivity in NMA calculations. However, DAS is not as popular as CCS because NMA in DAS, in particular matrices **F** and **H** defined in Eqs. 1–3, is formulated in a mathematically complicated manner than compared to that of CCS (Noguti and Gō 1983a, b).

In the NMA calculation in DAS, it is required to represent a molecular system as a tree in the sense of graph theory. This is essential for rapid calculation of the second derivatives of the potential energy functions with respect to the dihedral angles (Noguti and Gō 1983b; Abe et al. 1984; Wako and Gō 1987). It is possible to extend this into a system consisting of two or more molecules in the same framework as the formulation for a one-molecule system. Such a system can be also represented as a tree, including translational and rotational variables (Braun et al. 1984; Higo et al. 1985; Wako et al. 1995). Although the NMA calculation includes mathematically

complicated formulas, the computation program can be coded in a comprehensive manner so that it may work for any molecular systems represented as a tree. Therefore, the tree representation of a system is essential in the NMA calculation in DAS.

In the tree representation, a molecule comprises of a rigid-body portion that contains one or more atoms and a rotatable covalent bond connecting the rigid-body portions. The rigid-body portion and the rotatable covalent bond correspond to a vertex and an edge of a tree graph, respectively. In our previous works, the vertices and edges are simply referred to as a *unit* and *bond*, respectively (Wako and Gō 1987; Wako et al. 1995).

A critical feature of the tree representation of a molecule is that the tree cannot have any loops. In particular, a loop containing any rotatable covalent bond is not allowed in the NMA algorithm for a rapid calculation of the second derivatives of the potential functions. One of the challenging points is associated with various ring structures in the PDB data including an aromatic ring and a sugar ring. These are not problematic as long as they can be regarded as rigid bodies that do not contain any rotatable bonds. However, if, for example, a disulfide bond is formed in a protein, a loop is generated containing rotatable covalent bonds within it. In this case, it is difficult to acknowledge that the rigid-body assumption of the loop is appropriate. The same result is possible if a ligand molecule binds to a protein with more than one covalent bond. In such cases, we have avoided the loop problem by assuming that one of the covalent bonds in the loop is non-bonded in order to break the loop. Instead, a strong force is imposed at some distance between the atoms to maintain the original bond length and bond angles by incorporating a pseudo-potential energy for loop closing.

Inversely, using this strategy, we can introduce flexibility in the sugar ring of DNA and RNA. For example, an O4'-C1' bond was regarded as non-bonded, and the loop-closing potential was applied to the four atom pairs (O4', C1'), (C4', C1'), (O4', C2'), and (O4', N1) to maintain the bond length and bond angles. The introduction of flexibility to the sugar ring in this manner provided better quantitative results than those calculated with the rigid-body sugar rings (Wako and Endo 2013).

The process to construct a tree from PDB data is given in our previous paper (Wako and Endo 2013). Essentially, any molecules in the PDB data can be represented as a tree. Because a disulfide bond and a ring structure can be algorithmically detected, the tree representation of the PDB data can be performed automatically except for few peculiar data. Once the system is presented as a tree, NMA can be performed by the same computer program. In fact, NMA has been demonstrated for various systems composed of not only proteins but also DNA, RNA, and various ligand molecules.

Making full use of the characteristics of NMA in DAS, we have tried to introduce another possibility. We fixed some of the dihedral angles in the NMA calculations. By this operation, the number of independent variables can be reduced while the full-atom system is being considered. For example, it is possible to fix the side chain dihedral angles and one of the main chain dihedral angles, ω, which keeps a flat peptide plane. In this system, the conformations of the system can be treated with only the main chain dihedral angles, ϕ and ψ, as independent variables. That is, each residue has only two variables, in contrast to three variables (x, y, and z coordinates) of the Cα atom used in CCS. Figure 1 shows how the units and bonds are re-defined in fixing the dihedral angles.

NMA for the system with only ϕ and ψ as independent variables (referred to as system B) provides fluctuation profiles of atoms similar to the system in which any rotatable dihedral angles are taken into computation (system A), as shown in Fig. 2 and Fig. S1 in the supplementary material. In Fig. 2, the fluctuation profiles for residues of the two systems are compared for the various PDB data with resolution less than 1 Å. The closed circles in Fig. 2 show the RMS difference of the calculated fluctuations from those estimated by temperature factors for system B plotted against that for system A. In this comparison, the temperature in the NMA calculation was adjusted so that the mean fluctuation may be coincident with the fluctuation estimated from a temperature factor. The diagonal line indicates that the RMS differences

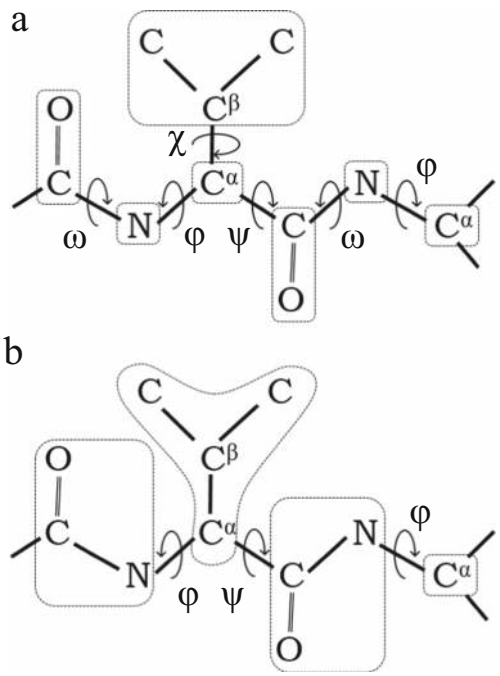

Fig. 1 Tree representation of a molecule. The tree is composed of a *unit* and a *bond*. The atoms enclosed with a dashed line constitute a rigid-body component referred to as a *unit*. A chemical bond with a rotation arrow is rotatable and referred to as a *bond*. **a** All dihedral angles are considered rotatable (system A). **b** Main-chain dihedral angle ω and side-chain dihedral angles are fixed, and, thus, only main-chain dihedral angles, φ and ψ, are rotatable (system B)

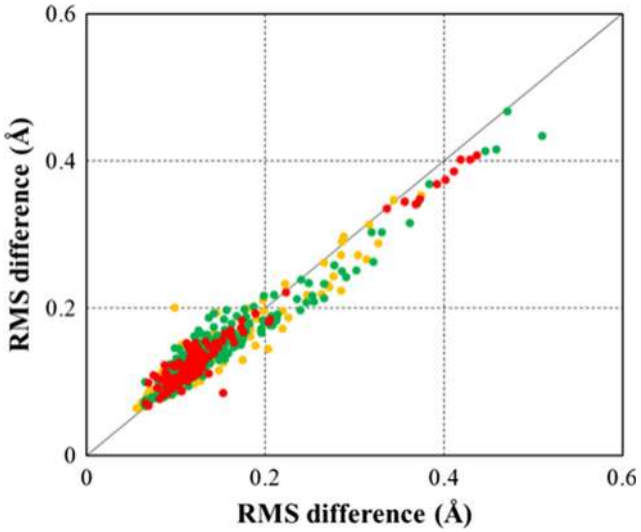

Fig. 2 Root mean square (RMS) differences of the two systems, A and B, defined in Fig. 1. The protein structure data in the Protein Data Bank (PDB) with high resolution (less than 1 Å) were examined. A closed circle denotes RMS differences between calculated and experimentally observed atomic fluctuations for systems A and B (horizontal and vertical axes, respectively). The diagonal line indicates that they are identical. The proteins are classified into three groups by their sizes: yellow, green, and red indicate that the protein size is less than 100, 100 to 300, and greater than 300, respectively

for the two systems are identical. Figure 2 suggests that the deviations between the RMS differences for the two systems do not differ significantly. In Fig. 2, the proteins are classified according to their sizes. Clear dependence on protein size is not found. Figure 2 also indicates that most of the RMS deviations of the calculated fluctuation profiles from the temperature factors are less than 0.2 Å.

In Fig. S1, the fluctuations in the system where only ϕ and ψ are changeable were compared with those calculated from temperature factors for a specific example: a spliceosomal protein, Prp8, which is a heterodimer of two subunits of 334 and 1407 amino acid residues (PDB ID: 4i43) (Query and Konarska 2013). The correspondence between the calculated and experimentally observed fluctuations was good except for residues on the protein surface. In the NMA calculation, the side chains of residues on the protein surface were allowed to fluctuate rather freely, because there were no molecules surrounding it. In contrast, such side chains are restrained by other proteins surrounding them in the crystal.

We also examined RMS differences between the fluctuation profiles of the two systems, A and B (not shown here). The small differences of less than 0.1 Å also suggest that fixing the side chain and ω dihedral angles is a good approximation.

Furthermore, it is possible to set up various situations in NMA in DAS. For example, the fluctuations of protein in a crystalline environment have been studied in CCS (Hinsen 2008; Riccardi et al. 2009; Hafner and Zheng 2010; Lu and Ma 2013). In contrast to these studies, we explored a simple model where the atoms surrounding the relevant molecules in the crystal are considered as a rigid body. These atoms are considered as one unit in a tree representation. Because it is not necessary to consider the connections between the atoms within the rigid-body part, it is possible to consider only the atoms relatively near to the relevant molecules. By this model, it was confirmed that the large fluctuations of the side chains on the protein surface, a problem mentioned above, are controlled so that they may coincide with the fluctuations calculated from the temperature factors (unpublished results). For another possibility, NMA can be applied to the system of virtual bonds (Cα-Cα bonds) with fixed virtual bond angles. In this case, the number of independent variables is one per residue. Interestingly, the profiles of atomic fluctuations are reproduced well in such a roughly constructed model (unpublished results).

NMA in DAS can provide various systems in which some parts of molecules are fixed and, thus, they are treated as a rigid body. The preliminary results shown above indicate that such rigid-body approximation is of little influence to NMA. Although it is important to clarify the reason why such approximations are effective, we cannot show it at this point. It requires further study.

As mentioned before, for any model, it is required to represent the system as a tree, and the NMA calculation program need not be modified any further.

Illustration of normal mode analysis

Protein for an illustration

In this section, we will demonstrate various dynamic aspects of molecules that can be derived by NMA for the PDB data. For this purpose, we used the PDB data for catabolite activator protein (CAP). CAP is a transcriptional activator known to regulate hundreds of transcription units by responding to fluctuations in the cellular concentration of cAMP. CAP undergoes cAMP-mediated allosteric transition to modulate DNA-binding activity. Allostery, in which ligand binding to a protein alters an activity at a distant site, was interesting from the perspective of protein dynamics. Although Rodgers et al. (2013) and Townsend et al. (2015) have already discussed the allostery of CAP based on NMA, we will discuss this textbook example of allostery based on our ENM-NMA calculations.

The three-dimensional structures of CAP have been determined in the three states: (a) in the absence of cAMP and free from DNA (PDB ID: 2wc2) (Popovych et al. 2009), (b) in the cAMP-bound state and free from DNA (1g6n) (Passner et al. 2000), and (c) in the cAMP-bound state and in complex with DNA (1j59) (Parkinson et al. 1996).

CAP is a homodimer (Fig. 3a). Although it is essentially symmetric, the 3D structures of the two subunits are not exactly identical in the PDB data. Minor differences between the NMA results for the two subunits are given below. Each subunit is composed of two domains: an N-terminal cAMP-binding domain (CBD; residues 1–136) and a C-terminal DNA-binding domain (DBD; residues 139–209). The two domains are linked by a short hinge region (residues 137–138). Of seven helices termed A to F, C-helix in CBD (residues 111–136 in 1g6n and 1j59 and 112–126 in 2wc2) forms an interface between the two subunits; F-helix in DBD (180–192 in 1j59) are inserted into the DNA double-helix major groove. These two helices are important to describe the cAMP-mediated allosteric transition of CAP. The cAMP binding sites are residues, 30, 49, 61, 71–73, 82–83, 124, and 127–128 in CBD. According to the PDB data, it was found that the F-helices in DBD are rotated by ~60° and translated ~7 Å to insert into the DNA major groove upon cAMP binding to apo CAP (Popovych et al. 2009).

Fluctuation of atoms

One of the major properties calculated in NMA involves atomic fluctuations. Two types of atomic fluctuations are calculated: displacement vectors of atoms of individual normal modes and their average over all the normal modes and time (RMS atomic fluctuations). As the RMS atomic fluctuations can be estimated from temperature factors provided in the PDB data, the calculated and experimentally observed atomic fluctuations are compared for assessing NMA results. As temperature factors can only provide information regarding the magnitude of the atomic fluctuations in most of the PDB data, it is generally problematic to compare the calculated and experimentally observed atomic fluctuations with respect to their directional properties. Such a comparison is possible for the PDB data with anisotropic temperature factors, and there is some literature that discusses this method (Atilgan et al. 2001; Eyal et al. 2007; Yang et al. 2009; Hafner and Zheng 2011). However, we have not discussed this in our review.

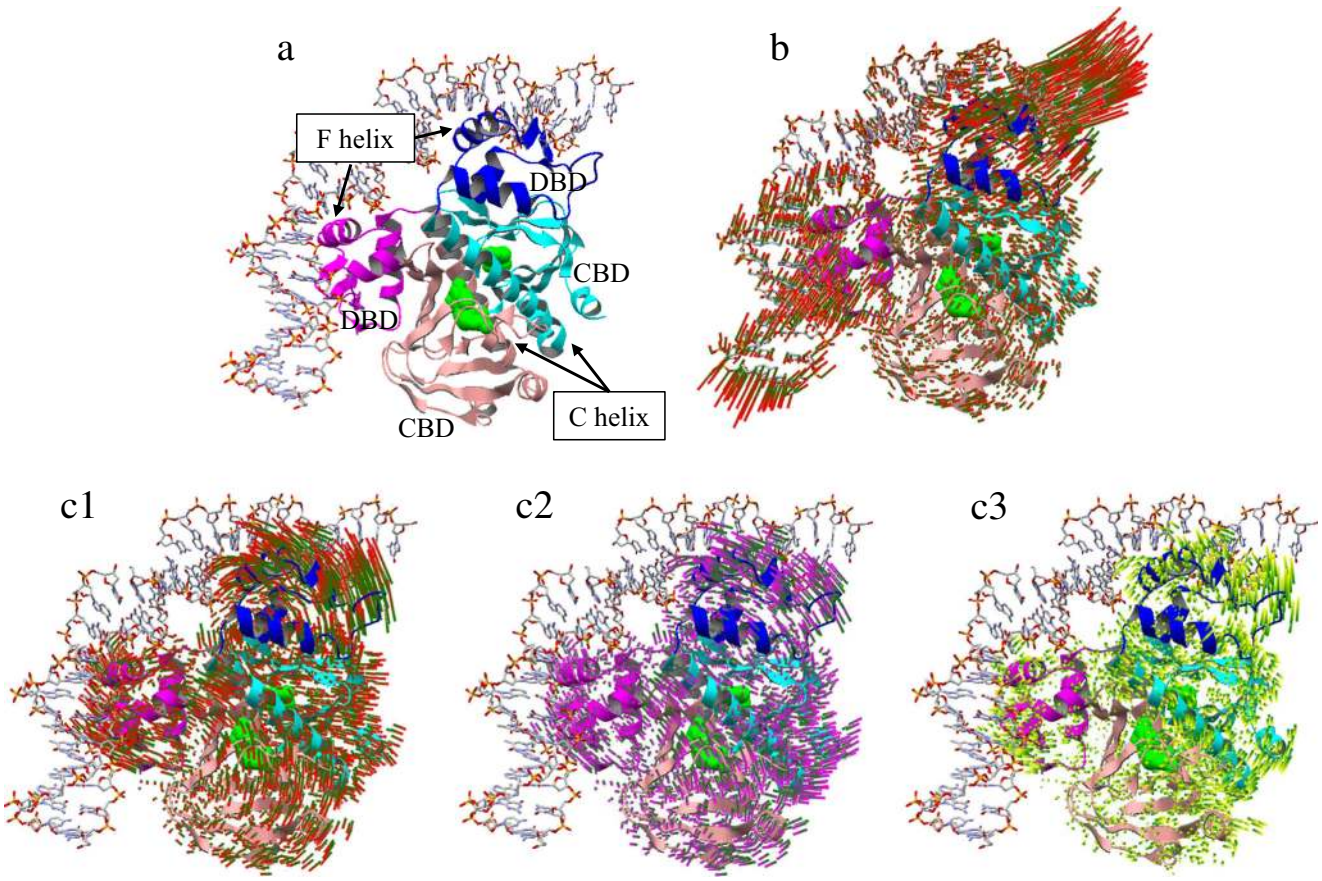

Fig. 3 Normal mode displacement vectors of atoms of 1j59. The magnitude of the displacement vector is exaggerated for clarity but is proportional to its actual value. It is presented as different colors for tail and head halves. The lowest frequency normal mode is shown for the two systems. **a** PDB conformation with protein, DNA, and cAMP. N-terminal cAMP-binding domain (CBD) and C-terminal DNA-binding domain (DBD) are in pink and magenta in chain A, and in cyan and blue in chain B, respectively. **b** The displacement vectors of atoms for NMA of the CAP-DNA-cAMP complex. **c1**–**c3** The displacement vectors of atoms for NMA of CAP only. Although the DNA is shown, it was not taken into account in the NMA calculations. The displacement vectors shown in **c1** are decomposed into the external and internal motions in **c2** and **c3**, respectively

In contrast, the displacement vectors of atoms for individual normal modes provide significant information on protein dynamics. The individual normal modes show characteristic motions that differ from mode to mode, and it is suggested that some of them are closely related to the functional movements of proteins, as mentioned below. In particular, the low-frequency normal modes show concerted motions of amino acid residues, such as opening-and-closing motions and sliding motion around the interfaces of domains and subunits. Such information cannot be obtained by a simple visual inspection of the PDB structure without calculating the dynamic properties by NMA (and MD).

Figure 3 presents the displacement vectors of atoms in the lowest-frequency normal mode of CAP with and without DNA and cAMP. The concerted motions of several groups of atoms, i.e., CBD and DBD, are clearly observed in Figs. 3b and c1. In the CAP-DNA-cAMP complex (Fig. 3b), the DBD domain and DNA are integrated and move together.

Figure 4 shows the residue-by-residue fluctuation profiles of atoms of apo CAP and CAP-DNA-cAMP complex, together with those calculated from the temperature factors in the PDB data of the CAP-DNA-cAMP complex. As the amplitudes of normal modes cannot be determined in ENM-NMA, the calculated fluctuation profiles were adjusted so that their mean values would coincide with those of the experimentally observed profiles. Figure 4 illustrates that the experimentally observed fluctuations are well reproduced, including DNA, by ENM-NMA.

In Fig. 4, large fluctuations are found around DNA and F-helices (180–192) bound to the DNA. The irregularly structured region in DBD (153–162), which is remote from the interface between the two subunits and is not involved in DNA binding, and B-helices in CBD, which is the most remote from the DNA, also fluctuate on a large scale. However, the fluctuations of the residues involved in cAMP binding and C-helices (111–136) forming the interface of the two subunits are relatively small. Generally speaking, the residues on the exterior fluctuate considerably, whereas those in the interior fluctuate to a much lesser extent.

It should be noted that the low-frequency normal modes dominantly contribute to the time-averaged properties (Gō 1990). In addition, the movements of some low-frequency normal modes are observed in the conformational change from apo and holo forms of an enzyme protein, as described later.

Correlation between atomic movements

The coordinated movements of atoms are one of the most important characteristics to be examined in protein dynamics studies. The correlation of motions is calculated as an inner product of displacement vectors of two atoms, and usually

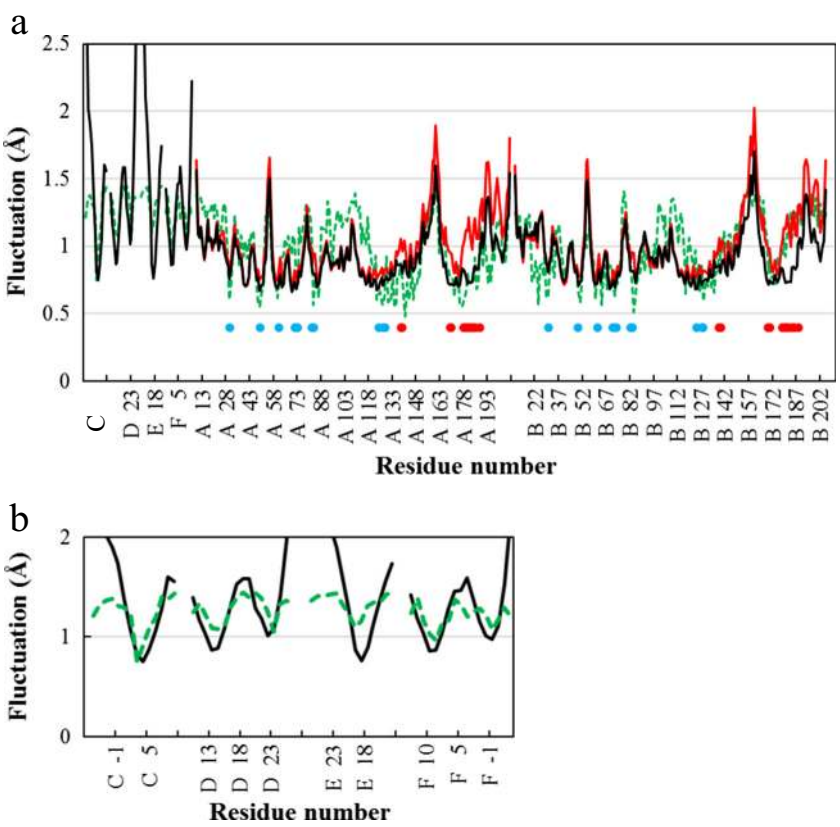

Fig. 4 Fluctuations of atoms with and without DNA. The black and red solid lines represent fluctuations of the CAP-DNA-cAMP complex and apo CAP, respectively. The cyan dashed line indicates fluctuations calculated from the temperature factors in the PDB data (1j59). The A and B chains are subunits of the protein dimer, the C-F and D-E chains are DNA double helices. Panel **b** is an enlargement of the DNA regions in panel **a**

normalized by their magnitudes. These values are calculated for individual normal modes, and then averaged over all normal modes and time. The averaged property is referred to as the correlation coefficient (CC). The CC has values between − 1 and 1.

The large positive CC indicates that the atoms move in the same direction. If an atom pair is spatially close, this suggests that the atoms form strong interactions. In the protein structure network analysis, the clusters of residues with large positive CCs are considered to play structurally and functionally important roles. The communication of distant residues through such residues is discussed in subsequent sections.

Conversely, a large negative CC indicates that the two residues move in the opposite direction. Because such residue pairs are found around active sites, and interface areas between domains and between subunits, such information is important for relating protein dynamics to protein function.

Figure 5 shows a correlation map for the CAP-DNA-cAMP complex, which reveals the CC values of residue pairs. Each residue is represented by one specific atom: Cα and C4' atoms for protein and DNA, respectively. The correlation map highlights clusters of residues with large positive CCs between them, which correspond to secondary structures and domains. The large negative CCs between the clusters suggest their repulsive movements. In Fig. 5, the two domains, CBD and DBD, appear as clusters of residues with large positive CCs. Positively correlated residue pairs are also found in the following pairs of regions: DNA strands forming double helices (region a in Fig. 5), DNA and F-helix (regions b and c), the C-helices of the A and B chains (region z), and the C-helix and CBD (regions x and y). Negatively correlated residue pairs are found in the following pairs of regions: CBD and DBD in the same chain (regions p and q) and CBDs in the A and B chains (region s). The correlation coefficients between DBDs in the A and B chains (region t) are mixed; some residue pairs are positively correlated, but others are negatively correlated. These observations help create a global image of the dynamic structure of the CAP-DNA-cAMP complex.

Correlation maps can be created for individual normal modes. These are significantly different from the correlation map for the average CC (Fig. 5) and between themselves, as shown in Fig. S2 in the supplementary material. This is another representation of the individual normal modes, different from the displacement vectors shown in Fig. 3.

In a comparative study (Wako et al. 1996) of homologous proteins that show essentially the same folding characteristics but have a slight difference in amino acid sequences and local structures, the correlations calculated by the full-atom NMA are found to be useful to detect their differences in static and dynamic structures. Differences can be difficult to detect by simple visual inspection of static structures, whereas these can be revealed because a change in the interactions is definitely reflected in the NMA correlations between atoms.

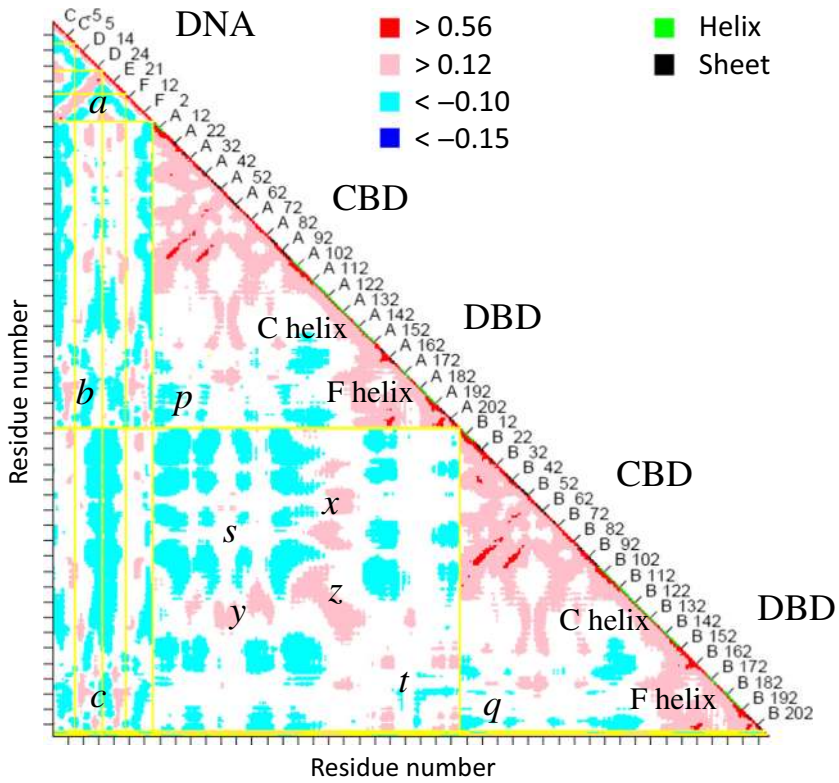

Fig. 5 Correlation map for the CAP-DNA-cAMP complex. Mean correlation coefficients over all normal modes and time for the representative atoms, Cα for the protein and C4' for the DNA, are plotted. The residue pairs are ranked according to their correlation coefficients and shown by the different colors indicated in the legend

Internal and external motions of individual subunits in an oligomeric protein

NMA has been applied to oligomeric proteins such as hemoglobin (Seno and Gō 1990a, b), G-protein coupled receptors (Niv and Filizola 2008), membrane proteins (Bahar et al. 2010), and immunoglobulins (Wako and Endo 2012). It has also been applied to multimeric structures such as subtilisin-eglin c complexes (Ishida et al. 1998), protein-DNA complexes (Yang et al. 2006), and protein-RNA complexes (Wako and Endo 2013) to reveal the structure–function relationships, with particular attention to the influence of oligomerization and complex formation with other molecules.

In these studies, Ishida et al. (1998) proposed an approach to investigate the mutual motions between the constituent subunits and molecules. In an NMA of oligomeric and complexed proteins, translational and rotational motions of the whole protein are eliminated by the Eckart condition (Eckart 1935). Accordingly, all motions considered are attributed to the internal degrees of freedom of the whole system. However, when these internal motions are further broken down into motions of the constituent subunits and molecules, we have two motions characteristic to the oligomer and complex, i.e., internal and external motions of individual subunits and molecules. The internal motion is a deformation motion, and the external one is a rigid-body motion of each constituent changing its mutual disposition. They may be referred to as tertiary and quaternary movements, respectively. The external motion is considered key to understanding the dynamics of oligomeric and complexed proteins.

A simple formula for the decomposition of motions into the internal and external components is possible in NMA in DAS, and given in Ishida et al. (1998). Eventually, the mean-square fluctuation of atom α, $<(\Delta r_\alpha)^2>$, is decomposed into three terms: internal motion, $<(\Delta r_\alpha{}^{\mathrm{int}})^2>$, external motion, $<(\Delta r_\alpha{}^{\mathrm{ext}})^2>$, and their cross-correlation, $<\Delta r_\alpha{}^{\mathrm{int}}\Delta r_\alpha{}^{\mathrm{ext}}>$:

$$< (\Delta r_\alpha)^2 >=< \left(\Delta r_\alpha^{\mathrm{int}}\right)^2 > + < \left(\Delta r_\alpha^{\mathrm{ext}}\right)^2 > +2 < \left(\Delta r_\alpha^{\mathrm{int}}\Delta r_\alpha^{\mathrm{ext}}\right) >, \quad (6)$$

where <…> denotes an average over all the normal modes and time.

These values for CAP of apo and complex forms are plotted against the residue numbers in Fig. 6. The internal motions are considerably larger than the external motions for both forms. However, this depends on the specific system considered. We have other examples in which both motions are comparable to each other (Wako and Endo 2012). Figure 6 indicates that the internal motions of individual components, i.e., subunits and DNA, determine the sum of motions of CAP. However, as far as the low-frequency normal modes are concerned, the internal and external motions are comparable to each other, as shown in Fig. 7. Because the motions of low-frequency normal modes are presumably related to biological function, this fact should be noted. In other words, the external motions have the potential to play an important role in protein dynamics related to function.

In the internal motions in Fig. 6a, the differences between apo and complex forms are distinctive around DBD; the internal motions in the apo form are much larger than in the complex form because of the absence of the interactions with DNA.

The third term on the right-hand side of Eq. 6, i.e., the cross-correlation between internal and external fluctuations, is plotted in Fig. 6b. For the apo-form CAP, large positive values around DBD suggest that the deformative fluctuations and the rigid-body ones of DBD are concerted and they amplify the total fluctuations (see large fluctuations of DBD in the apo-form CAP in Fig. 6a). In contrast, large negative values around C-helices of the two subunits, which form the interface between them, suggest that the internal motions have an almost inverse phase with the external motions to attenuate the total motions. These types of behavior in the interface region are found in other examples (Ishida et al. 1998; Wako and Endo 2012). Similarly, in CAP complexed with DNA, negative values of DNA and F-helices, i.e., interface of the DNA and protein, are found, suggesting the same situation. In addition, much larger negative values of the cross-correlation term of DNA are remarkable. This reflects the flexibility of DNA. When DNA moves away from the protein, DNA easily changes its conformation to resist such movement.

When we turn our attention to the individual low-frequency normal modes, it is interesting to characterize the external motions of the individual subunits and molecules. For this purpose, we defined translational and rotational vectors for the external motions of the individual subunits in the λth normal mode as follows (Wako and Endo 2012):

$$T_\lambda = \left(\sum_\alpha m_\alpha \Delta r_{\alpha,\lambda}^{\mathrm{ext}}\right) / \sum_\alpha m_\alpha \quad (7)$$

$$R_\lambda = \Gamma^{-1} \sum_\alpha m_\alpha \left(r_a - r^{\mathrm{G}}\right) \times \Delta r_{\alpha,\lambda}^{\mathrm{ext}} \quad (8)$$

where r_α and r^{G} are coordinate vectors of atom α in the minimum-energy conformation and the center of mass of the subunit, respectively, and m_α is a mass of atom α. The summation is taken over the atoms in the subunit. The inertia tensor Γ is given as:

$$\Gamma = \sum_\alpha m_\alpha \begin{pmatrix} y_\alpha^2 + z_\alpha^2 & -x_\alpha y_\alpha & -x_\alpha z_\alpha \\ -x_\alpha y_\alpha & z_\alpha^2 + x_\alpha^2 & -y_\alpha z_\alpha \\ -x_\alpha z_\alpha & -y_\alpha z_\alpha & x_\alpha^2 + y_\alpha^2 \end{pmatrix}, \quad (9)$$

where $(x_\alpha, y_\alpha, z_\alpha) = (r_\alpha - r^{\mathrm{G}})$. The dimensions of the two vectors, T_λ and R_λ, are Å and radians, respectively. The two vectors are useful for characterizing mutual movements of

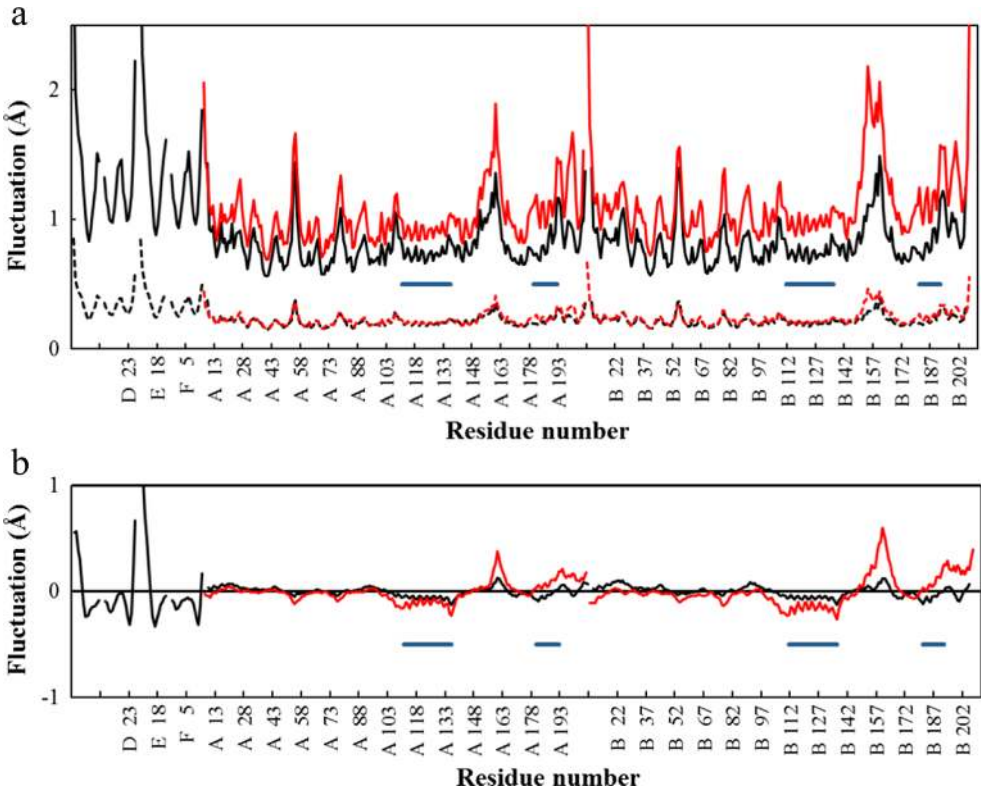

Fig. 6 Decomposition of fluctuations of atoms into internal and external motions for the CAP-DNA-cAMP complex and apo CAP. **a** The internal and external motions are plotted by solid and dashed lines, respectively. **b** The correlation term of the internal and external motions. The black and red lines are for the CAP-DNA-cAMP complex and apo CAP, respectively

subunits.

The translational and rotational vectors of the three lowest-frequency normal modes defined for individual subunits for CAP are shown in Fig. 8. The lengths of vectors are exaggerated, but proportional to their absolute values. Although the sizes of the translational vectors cannot be compared with the rotational vectors directly, the rotational motions of atoms distant from the center of mass are usually more dominant than the translational motions. Interestingly, these vectors of the three normal modes are perpendicular to each other in general, although it is not necessarily true between the translational vectors of modes 1 and 3 in this illustration. This suggests that various motions are easily generated by a combination of them including opening-and-closing and sliding motions. As the external motions are comparable with the internal motions in the low-frequency normal modes as mentioned above, it is plausible that the external motions may

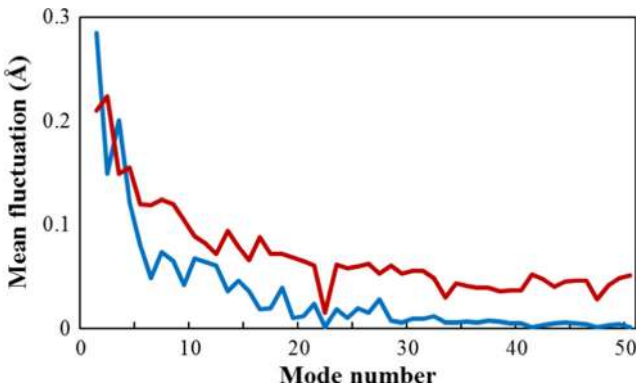

Fig. 7 The mean magnitudes of displacement vectors of atoms of internal and external motions for the 50 lowest-frequency normal modes. The red and blue lines show internal and external motions, respectively

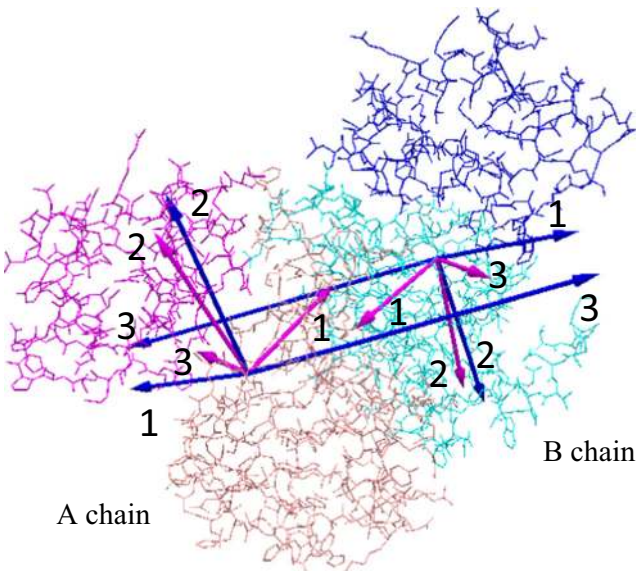

Fig. 8 Translational and rotational vectors of the three lowest-frequency normal modes for the subunits A and B. These are shown by the blue and red arrows with the normal mode numbers, respectively. The original points of the vectors are centers of mass of the individual subunits. The vector lengths are exaggerated, but are proportional to their absolute values

contribute significantly to a correlative motion of distant residues, such as allosteric movement.

Ligand-induced conformational change

Ligand binding to a protein is an essential process in many biochemical phenomena. A conformational change to either a larger or a smaller size plays an important role in the ligand-binding mechanism. In fact, conformational changes of the side chains of a binding site (Najmanovich et al. 2000; Zavodszky and Kuhn 2005), small movements and/or deformation of loops (Malkowski et al. 1997; Karthikeyan et al. 2003), large-scale domain motions (Kraft et al. 2002; Magnusson et al. 2004; Amemiya et al. 2012), and quaternary movements of subunits (Xiong et al. 2002; Benison et al. 2008) have been observed during the binding of a ligand to proteins. The intrinsic thermal motion of proteins is considered responsible for the conformational changes upon ligand binding. It has been found that the observed conformational changes upon ligand binding correlate well with the conformational change of some low-frequency normal mode vibrations (Tama and Sanejouand 2001; Krebs et al. 2002; Alexandrov et al. 2005; Tobi and Bahar 2005; Dobbins et al. 2008; Wako and Endo 2011).

Following Wako and Endo (2011), we will illustrate how the conformational change vectors of atoms from the apo to holo forms of a protein can be represented by a linear combination of the displacement vectors of atoms in the apo form calculated for the lowest-frequency m normal modes ($m = 50$ in this paper).

First, two conformations of the apo and holo forms of a protein, which are represented by atomic coordinates sets $\{x_\alpha\}$ and $\{y_\alpha\}$, respectively, were best-fitted. The NMA of the apo form provides a displacement vector of atom α in the jth normal mode, $b_{\alpha,j}$ ($j = 1, 2,\ldots, m$). As temperature cannot be defined in the ENM-NMA, $b_{\alpha,j}$ is normalized such that $\Sigma {b_{\alpha,j}}^2 = 1$ for any normal modes, j. It is assumed that the conformational changes of the apo form can be expressed by a linear combination of the displacement vectors with a proper set of weighting factors for the jth normal mode, $\{w_j\}$ ($j = 1, 2,\ldots, m$), as follows:

$$q_\alpha = x_\alpha + \sum_{j=1}^{m} w_j b_{\alpha,j} \tag{10}$$

Then, we estimated the set of weighting factors to best-fit $\{q_\alpha\}$ to $\{y_\alpha\}$ by minimizing the following objective function by the least-squares method:

$$F = \sum_{\alpha} (y_\alpha - q_\alpha)^2 \tag{11}$$

The conformation generated with the estimated weighting factors is referred to as a least-squares fitting (LSF) conformation.

For CAP, three cases were examined: (a) from the apo form to cAMP binding form (2wc2 → 1g6n), (b) from the apo form to cAMP and DNA binding form (2wc2 → 1j59), and (c) from cAMP binding form to cAMP and DNA binding form (1g6n → 1j59). The notation, X → Y, in parentheses, means that X is changed based on its displacement vectors of atoms obtained from its NMA calculation to best-fit to Y by the least-squares method with respect to weighting factors. Figure 9 shows superimposed conformations for the cases (a) and (c). The RMS differences between the superimposed conformations are shown in Table 1. The regions with large deviations in the LSF conformations from the target are shown in Fig. S3 in the supplementary material. Figure 10 shows the sets of weighting factors obtained using the least-squares method for the three cases.

These results show that a few low-frequency normal modes mainly contribute to their conformational changes: (a) the third, (b) the first and third, and (c) the first and second lowest-frequency normal modes according to Fig. 10. In the case of (a), however, the LSF conformation still differs from the targets. Significant deviations are found in the C-terminal regions of C-helices (Fig. S2). The comparison between the PDB conformations indicates that the key conformational change induced by cAMP is a coil-to-helix transition of the segment of residues 126–136 in the C-helices. This kind of large conformational change cannot be predicted by NMA.

Nonetheless, the LSF and target conformations are globally very similar. The molecular motions of the first, second, and third lowest-frequency normal modes can be observed in the figures of atomic displacement vectors and GIF animations presented in the ProMode-Elastic database at PDBj. In these motions, four domains, i.e., two CBDs and two DBDs in two subunits, A and B, vibrate in a repulsive manner to one another but in various directions, depending on the modes. The basic conformations of the individual domains are retained in these modes; the two subunits exhibit the same behavior. The above results show that apo-to-holo conformational changes are undergone in combination of these motions. Consequently, even if two residues are distant from each other, they can move in a coherent manner.

This fact suggests that it might be incorrect to attribute the conformational change upon ligand binding entirely to the coil-to-helix transition of the C-terminal segment of the C-helices. It is emphasized above that the external motions are dominant over, or comparable to, the internal motions in the low-frequency normal modes. Therefore, a possible explanation of the real-world scenarios in CAP is a combination of global and local, or quaternary and tertiary, conformational changes.

Twenty-two paired proteins in the holo and apo forms, which undergo relatively large conformational changes, were examined previously, and similar results were

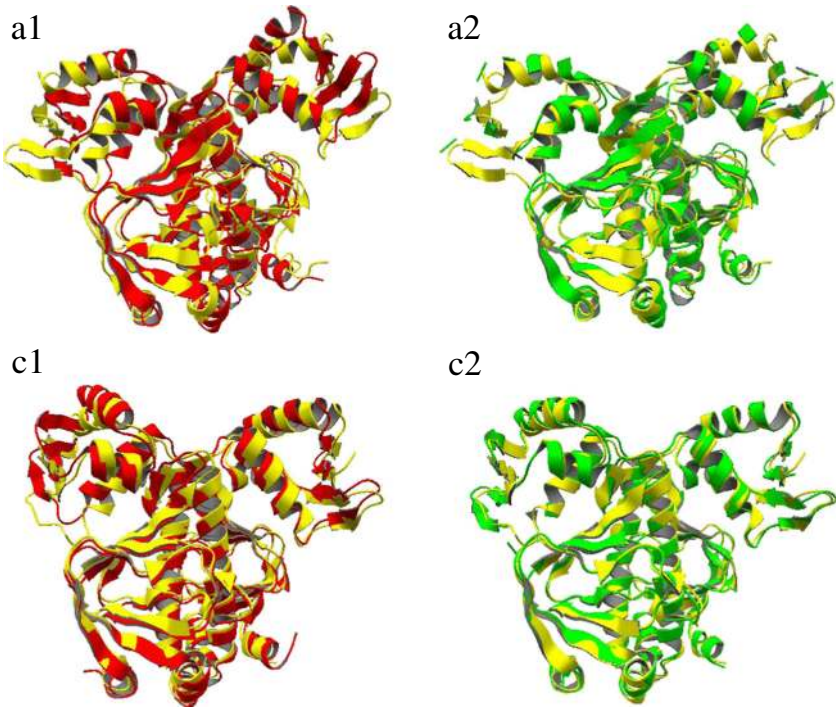

Fig. 9 Superimposed conformations before and after applying the least-squares method. **a1** apo form (2wc2; red) and cAMP binding form (1g6n; yellow) conformations, and **a2** LSF (green) and cAMP binding form conformations are superimposed. **c1** cAMP binding form (1g6n; red) and cAMP and DNA binding form (1j59; yellow) conformations, and **c2** LSF (green) and cAMP and DNA binding form conformations are superimposed

obtained (Wako and Endo 2011). The results showed that, in most cases, the conformational change was reproduced well by a linear combination of the displacement vectors of a small number of low-frequency normal modes. Although the conformational change around an active site was reproduced as well as the entire conformational change in these cases, it was not successful for some proteins that underwent significant conformational changes around active sites.

In the above study, we supposed a situation that conformations around the apo form of a protein are generated by the linear combination of the displacement vectors of low-frequency normal modes obtained by ENM-NMA by varying the weighting factors in Eq. 10. This may help to solve the flexible-docking problem of a protein with another molecule because the results presented here suggest that they have a relatively high probability of being involved in an actual conformational change (Meireles et al. 2011).

Table 1 Root mean square (RMS) differences between conformations

X → Y	X and Y (Å)	LSF and Y (Å)
(a) 2wc2 → 1g6n	5.15 (4.53)	3.16 (2.11)
(b) 2wc2 → 1j59	4.79 (4.01)	3.35 (2.19)
(c) 1g6n → 1j59	2.73 (2.34)	1.63 (0.90)

RMS difference between two conformations is calculated for all atoms and only for main-chain atoms. The latter value is shown in parentheses

Network analysis

One of the important dynamic aspects obtained from NMA is coordinated movements of atoms and residues. The concerted motions of a cluster of residues described above are one of such phenomena to be noted. Another interesting aspect is a network of residues where the remote residues communicate with each other via other residues. Recently, a network analysis based on graph theory, which is popular in social sciences and in complex-system studies, has been applied to a protein structure to elucidate structurally and functionally important residues, intra- and inter-protein communication, and allostery (Vendruscolo et al. 2001; Amitai et al. 2004; Böde et al. 2007; Tang et al. 2007; Chennubhotla and Bahar 2007; Vishveshwara et al. 2009; Papaleo et al. 2012; Raimondi et al. 2013). This kind of network is referred to as a protein structure network (PSN). PSN is analyzed for not only static but also dynamic structures of proteins. In the PSN analysis, various properties are calculated to characterize the network: shortest communication pathways, various centrality measures to elucidate important residues such as degree, closeness, and betweenness, a cluster of residues, and so on. In this review, we tentatively applied a network analysis method to the NMA results of CAP, particularly with respect to the betweenness centrality, and demonstrated its application.

The PSN of CAP was individually defined for the ten lowest-frequency normal modes, in addition to the static structure (i.e., the PDB structure). The network is composed of vertices and edges. A Cβ atom is assigned to a vertex for

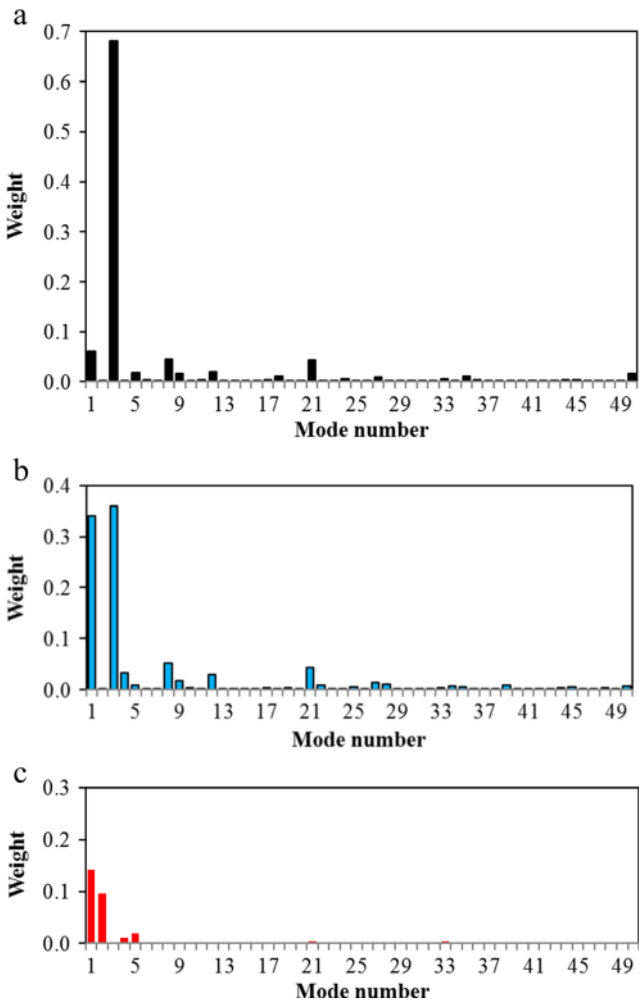

Fig. 10 Contributions of the 50 lowest-frequency normal modes in the least-squares fitting (LSF). The normalized weighting factor, w_i^2/W (i = 1, 2,…, 50), where $W = \Sigma w_i^2$, is shown. **a** 2wc2 → 1g6n, **b** 2wc2 → 1j59, and **c** 1g6n → 1j59

representing a residue (a Cα atom for a Gly residue). For the PDB structure, spatially neighboring residues were connected by an edge; i.e., if at least one atom pair between two residues was located within a given cutoff distance (5.0 Å in this study), they were connected by an edge. For the network based on the normal mode, two residues were connected by an edge, if they were spatially neighboring and moved in a coherent manner; i.e., one more condition that the inner product of displacement vectors of two Cβ atoms is greater than a given cutoff value (0.8 in this study) was added for the edge definition. This condition means that the two connected Cβ atoms move in the similar direction. The networks defined for the individual normal mode vibrations are 'subgraphs' of the network defined for the PDB structure.

We calculated the betweenness centrality for every residue (i.e., vertex) in these networks. For every pair of residues in a network, there exists a shortest path between them. The betweenness for a residue is defined as the number of these shortest paths that pass through the residue. A residue with higher betweenness would have more control over the network because more residues communicate with each other passing through that residue.

Figure 11a shows the betweenness values of individual residues for the PDB conformation and ten lowest-frequency normal modes. The betweenness is normalized by dividing by the total number of residue pairs. The betweenness profiles for the ten normal modes differ from mode to mode, and from that for the PDB structure. In the profiles for the normal modes, compared to that for the PDB structure, there are residues whose betweenness values are significantly larger than those of others. In Table S1 in the supplementary material, the residues with higher betweenness (> 0.12) are shown with the mode numbers. These residues are indicated in the space-filling model in Fig. 11b. Interestingly, these residues continue from the cAMP binding site (residues in green in Fig. 11b) into the DNA binding site, F-helix, via C-terminal of C-helices and the interface region between CBD and DBD, particularly E-helices.

Moreover, we analyzed certain enzyme proteins with this method (results not shown here). The residues with higher betweenness in some normal modes were found at ligand binding sites. These results suggested that the ligand binding sites are located in the central part of the residue–residue communications. This also implies that network analysis is useful in characterizing the individual normal modes.

Discussion

NMA is useful to analyze both local fluctuations and global motions. In local fluctuations, not only the mean fluctuations average over all the normal modes and time, but also the displacement vectors of atoms in the individual normal modes are examined (Figs. 3 and 4). It is also important to notice that the structure of a specific region is flexible or rigid-body-like. For this analysis, fluctuations of dihedral angles and a decomposition of the fluctuation of the specified region into internal and external motions can provide useful information (Hirose et al. 2010). For global motions, the concerted motions appearing in the low-frequency normal modes play a significant role. To inspect such motions, the correlation map shown in Fig. 5 and translational and rotational motions of the individual subunits shown in Fig. 8 are helpful for visualizing the global behaviors of a protein. Thus, NMA can create vivid pictures of mutual motions between the secondary structures, between the domains, between the subunits, and between the proteins and other molecules.

In this review, we have demonstrated NMA by applying it to an allosteric protein, CAP. The mechanism of allostery is still controversial. Several theoretical possibilities have been proposed (Laskowski et al. 2009; Townsend et al. 2015).

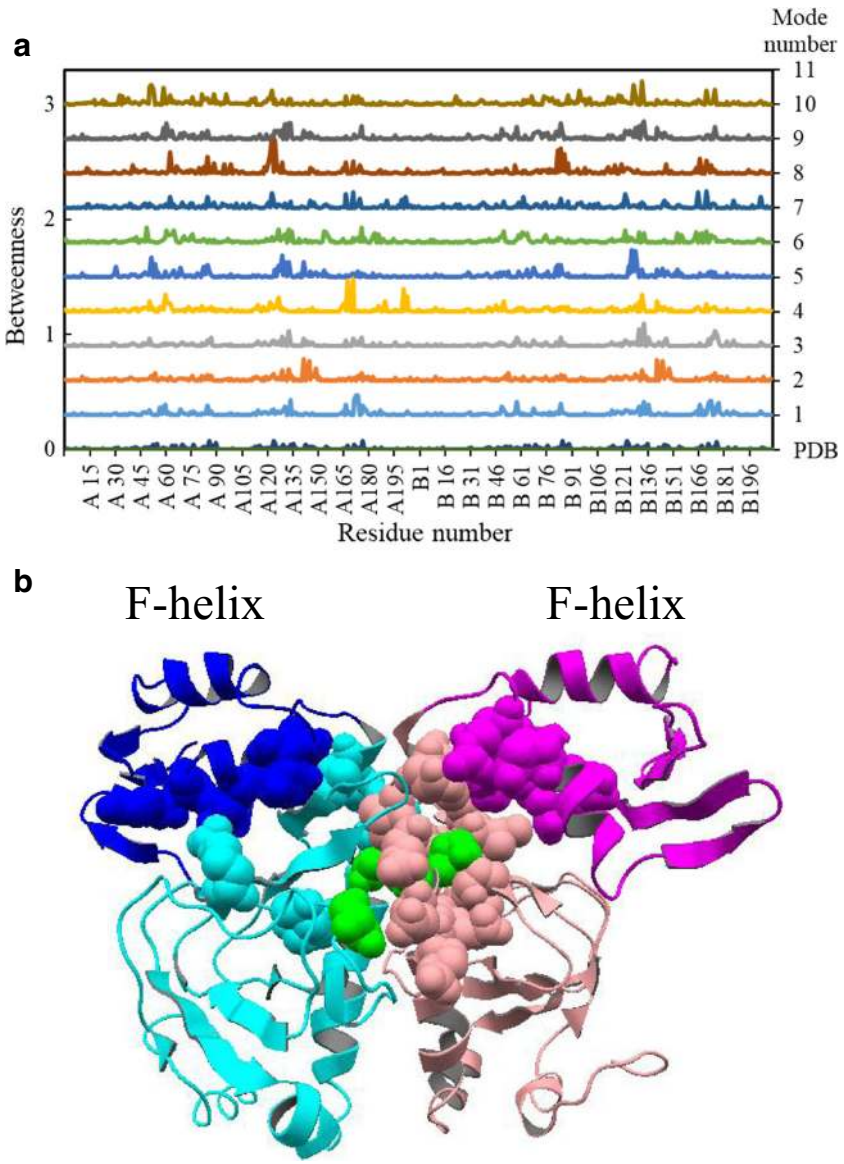

Fig. 11 Betweenness of residues for the PDB structure and ten lowest-frequency normal modes of CAP. **a** Betweenness values are plotted and are shifted upward by 0.3× (normal mode number) for clarity. The bottom line is the betweenness of the PDB conformation. **b** Residues with a high betweenness value are represented by a space-filling model on a cartoon model of CAP (see also Table S1 in the supplementary material). Of the residues with a high betweenness value, those that are bound to cAMP are shown by a green space-filling model. CBD and DBD in the A chain are colored pink and magenta, respectively, and those in the B chain cyan and blue, respectively

One possibility is that a change of a local structure upon an effector molecule binding induces a change in conformation or orientation of the remote active site to fit a ligand, irrespective of the fluctuation patterns of the low-frequency normal modes. This type of description is typically provided in a comparative study of static structures. In fact, the comparative study of the PDB structures of apo and complexed CAP revealed differences in the C-termini of C-helices, and, thus, they are attributed to allosteric conformational changes (Popovych et al. 2009).

Another possibility is that conformational changes related to the low-frequency normal modes make it possible for an active site to respond to the effector molecule binding at the remote position. In the CAP study, it was shown that the ligand-induced conformational change can be reproduced with a combination of several low-frequency normal modes, at least to some extent. The low-frequency normal modes are characterized by the long-range correlation distance. In oligomeric proteins, the rigid-body-like motions of individual subunits are also possible candidates to induce correlative motions of remote regions.

The third possibility is that the communications among positively correlated residues carry an allosteric signal and then cause some response of the active site on effector molecule binding. In the CAP, the residues located from the cAMP binding site to the DNA binding site have relatively higher betweenness centrality values. However, at this point, this explanation is interesting, but plausible in a conceptual sense. Nothing is revealed about what kind of signal might be carried through these residues to trigger an allosteric response on effector molecule binding.

The above study suggests that the actual events occurring in CAP seem to be a mixture of these possibilities. We have demonstrated that NMA can be utilized to overview protein

dynamics at various levels of structures, i.e., atoms, residues, secondary structures, domains, and subunits. The comparison of the same protein in different states, e.g., apo and holo configurations, and free and complexed states, as well as comparisons between homologous proteins provide useful information.

In general, it is difficult to arrive at rigorous theoretical explanations for protein dynamics only employing NMA due to the approximations that are introduced. Additional measures such as MD is necessary to complement NMA. Nevertheless, many researchers have shown that NMA can provide appropriate results consistent with MD simulations. Rapid computation and calculation of analytically well-defined properties are key advantages under NMA, which allows systematic analyses of many proteins. NMA should be a useful first step toward the study of protein dynamics and can be utilized to facilitate the understanding of protein dynamics.

Acknowledgements This work was supported by a JSPS Grant-in-Aid for Scientific Research (C) (grant no. 16K00407).

Compliance with ethical standards

Conflict of interest Hiroshi Wako declares that he has no conflict of interest. Shigeru Endo declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abe H, Braun W, Noguti T, Gō N (1984) Rapid calculation of first and second derivatives of conformational energy with respect to dihedral angles for proteins. General recurrent equations. Comput Chem 8: 239–247. https://doi.org/10.1016/0097-8485(84)85015-9

Alexandrov V, Lehnert U, Echols N, Milburn D, Engelman D, Gerstein M (2005) Normal modes for predicting protein motions: a comprehensive database assessment and associated Web tool. Protein Sci 14: 633–643. https://doi.org/10.1110/ps.04882105

Amemiya T, Koike R, Kidera A, Ota M (2012) PSCDB: a database for protein structural change upon ligand binding. Nucleic Acids Res 40:D554–D558. https://doi.org/10.1093/nar/gkr966

Amitai G, Shemesh A, Sitbon E, Shklar M, Netanely D, Venger I, Pietrokovski S (2004) Network analysis of protein structures identifies functional residues. J Mol Biol 344:1135–1146. https://doi.org/10.1016/j.jmb.2004.10.055

Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 80:505–515. https://doi.org/10.1016/s0006-3495(01)76033-x

Bahar I, Rader AJ (2005) Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 15:586–592. https://doi.org/10.1016/j.sbi.2005.08.007

Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 2:173–181. https://doi.org/10.1016/S1359-0278(97)00024-2

Bahar I, Lezon TR, Bakan A, Shrivastava IH (2010) Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. Chem Rev 110:1463–1497. https://doi.org/10.1021/cr900095e

Benison G, Karplus PA, Barbar E (2008) The interplay of ligand binding and quaternary structure in the diverse interactions of dynein light chain LC8. J Mol Biol 384:954–966. https://doi.org/10.1016/j.jmb.2008.09.083

Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide Protein Data Bank. Nat Struct Mol Biol 10:980. https://doi.org/10.1038/nsb1203-980

Böde C, Kovács IA, Szalay MS, Palotai R, Korcsmáros T, Csermely P (2007) Network analysis of protein dynamics. FEBS Lett 581:2776–2782. https://doi.org/10.1016/j.febslet.2007.05.021

Braun W, Yoshioki S, Gō N (1984) Formulation of static and dynamic conformational energy analysis of biopolymer systems consisting of two or more molecules. J Phys Soc Jpn 53:3269–3275. https://doi.org/10.1143/JPSJ.53.3269

Brooks B, Karplus M (1983) Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci U S A 80:6571–6575

Chennubhotla C, Bahar I (2007) Signal propagation in proteins and relation to equilibrium fluctuations. PLoS Comput Biol 3:1716–1726. https://doi.org/10.1371/journal.pcbi.0030172

Dobbins SE, Lesk VI, Sternberg MJ (2008) Insights into protein flexibility: the relationship between normal modes and conformational change upon protein–protein docking. Proc Natl Acad Sci U S A 105:10390–10395. https://doi.org/10.1073/pnas.0802496105

Echols N, Milburn D, Gerstein M (2003) MolMovDB: analysis and visualization of conformational change and structural flexibility. Nucleic Acids Res 31:478–482. https://doi.org/10.1093/nar/gkg104

Eckart C (1935) Some studies concerning rotating axes and polyatomic molecules. Phys Rev 47:552–558. https://doi.org/10.1103/PhysRev.47.552

Eyal E, Chennubhotla C, Yang L-W, Bahar I (2007) Anisotropic fluctuations of amino acids in protein structures: insights from X-ray crystallography and elastic network models. Bioinformatics 23:i175–i184. https://doi.org/10.1093/bioinformatics/btm186

Eyal E, Lum G, Bahar I (2015) The anisotropic network model web server at 2015 (ANM 2.0). Bioinformatics 31:1487–1489. https://doi.org/10.1093/bioinformatics/btu847

Gō N (1990) A theorem on amplitudes of thermal atomic fluctuations in large molecules assuming specific conformations calculated by normal mode analysis. Biophys Chem 35:105–112. https://doi.org/10.1016/0301-4622(90)80065-F

Gō N, Noguti T, Nishikawa T (1983) Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc Natl Acad Sci U S A 80:3696–3700

Hafner J, Zheng W (2010) Optimal modeling of atomic fluctuations in protein crystal structures for weak crystal contact interactions. J Chem Phys 132:014111. https://doi.org/10.1063/1.3288503

Hafner J, Zheng W (2011) All-atom modeling of anisotropic atomic fluctuations in protein crystal structures. J Chem Phys 135:144114. https://doi.org/10.1063/1.3646312

Higo J, Seno Y, Gō N (1985) Formulation of static and dynamic conformational energy analysis of biopolymer systems consisting of two or more molecules—avoiding a singularity in the previous method. J Phys Soc Jpn 54:4053–4058. https://doi.org/10.1143/JPSJ.54.4053

Hinsen K (1998) Analysis of domain motions by approximate normal mode calculations. Proteins 33:417–429. https://doi.org/10.1002/(SICI)1097-0134(19981115)33:3<417::AID-PROT10>3.0.CO;2-8

Hinsen K (2008) Structural flexibility in proteins: impact of the crystal environment. Bioinformatics 24:521–528. https://doi.org/10.1093/bioinformatics/btm625

Hirose S, Yokota K, Kuroda Y, Wako H, Endo S, Kanai S, Noguchi T (2010) Prediction of protein motions from amino acid sequence and

its application to protein–protein interaction. BMC Struct Biol 10: 20. https://doi.org/10.1186/1472-6807-10-20

Ishida H, Jochi Y, Kidera A (1998) Dynamic structure of subtilisin-eglin c complex studied by normal mode analysis. Proteins 32:324–333. https://doi.org/10.1002/(SICI)1097-0134(19980815)32:3<324::AID-PROT8>3.0.CO;2-H

Karthikeyan S, Zhou Q, Osterman AL, Zhang H (2003) Ligand binding-induced conformational changes in riboflavin kinase: structural basis for the ordered mechanism. Biochemistry 42:12532–12538. https://doi.org/10.1021/bi035450t

Kinjo AR, Bekker G-J, Suzuki H, Tsuchiya Y, Kawabata T, Ikegawa Y, Nakamura H (2017) Protein Data Bank Japan (PDBj): updated user interfaces, resource description framework, analysis tools for large structures. Nucleic Acids Res 45:D282–D288. https://doi.org/10.1093/nar/gkw962

Kraft L, Sprenger GA, Lindqvist Y (2002) Conformational changes during the catalytic cycle of gluconate kinase as revealed by X-ray crystallography. J Mol Biol 318:1057–1069. https://doi.org/10.1016/s0022-2836(02)00215-2

Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M (2002) Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. Proteins 48:682–695. https://doi.org/10.1002/prot.10168

Laskowski RA, Gerick F, Thornton JM (2009) The structural basis of allosteric regulation in proteins. FEBS Lett 583:1692–1698. https://doi.org/10.1016/j.febslet.2009.03.019

Levitt M, Sander C, Stern PS (1985) Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J Mol Biol 181:423–447. https://doi.org/10.1016/0022-2836(85)90230-X

Lu M, Ma J (2013) PIM: phase integrated method for normal mode analysis of biomolecules in a crystalline environment. J Mol Biol 425:1082–1098. https://doi.org/10.1016/j.jmb.2012.12.026

Magnusson U, Salopek-Sondi B, Luck LA, Mowbray SL (2004) X-ray structures of the leucine-binding protein illustrate conformational changes and the basis of ligand specificity. J Biol Chem 279: 8747–8752. https://doi.org/10.1074/jbc.M311890200

Malkowski MG, Martin PD, Guzik JC, Edwards BF (1997) The co-crystal structure of unliganded bovine α-thrombin and prethrombin-2: movement of the Tyr-Pro-Pro-Trp segment and active site residues upon ligand binding. Protein Sci 6:1438–1448. https://doi.org/10.1002/pro.5560060708

Meireles L, Gur M, Bakan A, Bahar I (2011) Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins. Protein Sci 20:1645–1658. https://doi.org/10.1002/pro.711

Momany FA, McGuire RF, Burgess AW, Scheraga HA (1975) Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J Phys Chem 79:2361–2381. https://doi.org/10.1021/j100589a006

Najmanovich R, Kuttner J, Sobolev V, Edelman M (2000) Side-chain flexibility in proteins upon ligand binding. Proteins 39:261–268. https://doi.org/10.1002/(SICI)1097-0134(20000515)39:3<261::AID-PROT90>3.0.CO;2-4

Niv MY, Filizola M (2008) Influence of oligomerization on the dynamics of G-protein coupled receptors as assessed by normal mode analysis. Proteins 71:575–586. https://doi.org/10.1002/prot.21787

Noguti T, Gō N (1983a) Dynamics of native globular proteins in terms of dihedral angles. J Phys Soc Jpn 52:3283–3288. https://doi.org/10.1143/JPSJ.52.3283

Noguti T, Gō N (1983b) A method of rapid calculation of a second derivative matrix of conformational energy for large molecules. J Phys Soc Jpn 52:3685–3690. https://doi.org/10.1143/JPSJ.52.3685

Papaleo E, Lindorff-Larsen K, De Gioia L (2012) Paths of long-range communication in the E2 enzymes of family 3: a molecular dynamics investigation. Phys Chem Chem Phys 14:12515–12525. https://doi.org/10.1039/c2cp41224a

Parkinson G, Wilson C, Gunasekera A, Ebright YW, Ebright RE, Berman HM (1996) Structure of the CAP-DNA complex at 2.5 Å resolution: a complete picture of the protein–DNA interface. J Mol Biol 260: 395–408. https://doi.org/10.1006/jmbi.1996.0409

Passner JM, Schultz SC, Steitz TA (2000) Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 Å resolution. J Mol Biol 304:847–859. https://doi.org/10.1006/jmbi.2000.4231

Popovych N, Tzeng SR, Tonelli M, Ebright RH, Kalodimos CG (2009) Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. Proc Natl Acad Sci U S A 106:6927–6932. https://doi.org/10.1073/pnas.0900595106

Query CC, Konarska MM (2013) Structural biology: Spliceosome's core exposed. Nature 493:615–616. https://doi.org/10.1038/nature11857

Raimondi F, Felline A, Seeber M, Mariani S, Fanelli F (2013) A mixed protein structure network and elastic network model approach to predict the structural communication in biomolecular systems: the PDZ2 domain from tyrosine phosphatase 1E as a case study. J Chem Theory Comput 9:2504–2518. https://doi.org/10.1021/ct400096f

Reuter N, Hinsen K, Lacapère J-J (2003) Transconformations of the SERCA1 ca-ATPase: a normal mode study. Biophys J 85:2186–2197. https://doi.org/10.1016/S0006-3495(03)74644-X

Riccardi D, Cui Q, Phillips GN Jr (2009) Application of elastic network models to proteins in the crystalline state. Biophys J 96:464–475. https://doi.org/10.1016/j.bpj.2008.10.010

Rodgers TL, Townsend PD, Burnell D, Jones ML, Richards SA, McLeish TC, Pohl E, Wilson MR, Cann MJ (2013) Modulation of global low-frequency motions underlies allosteric regulation: demonstration in CRP/FNR family transcription factors. PLoS Biol 11:e1001651. https://doi.org/10.1371/journal.pbio.1001651

Seno Y, Gō N (1990a) Deoxymyoglobin studied by the conformational normal mode analysis: I. Dynamics of globin and the heme-globin interaction. J Mol Biol 216:95–109. https://doi.org/10.1016/s0022-2836(05)80063-4

Seno Y, Gō N (1990b) Deoxymyoglobin studied by the conformational normal mode analysis: II. The conformational change upon oxygenation. J Mol Biol 216:111–126. https://doi.org/10.1016/s0022-2836(05)80064-6

Skjærven L, Reuter N, Martinez A (2011) Dynamics, flexibility and ligand-induced conformational changes in biological macromolecules: a computational approach. Future Med Chem 3:2079–2100. https://doi.org/10.4155/fmc.11.159

Skjærven L, Yao XQ, Scarabelli G, Grant BJ (2014) Integrating protein structural dynamics and evolutionary analysis with Bio3D. BMC Bioinformatics 15:399. https://doi.org/10.1186/s12859-014-0399-6

Suhre K, Sanejouand YH (2004) ElNémo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res 32:W610–W614. https://doi.org/10.1093/nar/gkh368

Tama F, Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. Protein Eng 14:1–6. https://doi.org/10.1093/protein/14.1.1

Tang S, Liao JC, Dunn AR, Altman RB, Spudich JA, Schmidt JP (2007) Predicting allosteric communication in myosin via a pathway of conserved residues. J Mol Biol 373:1361–1373. https://doi.org/10.1016/j.jmb.2007.08.059

Tiwari SP, Fuglebakk E, Hollup SM, Skjærven L, Cragnolini T, Grindhaug SH, Tekle KM, Reuter N (2014) WEBnm@ v2.0: web server and services for comparing protein flexibility. BMC Bioinformatics 15:427. https://doi.org/10.1186/s12859-014-0427-6

Tobi D, Bahar I (2005) Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state.

Proc Natl Acad Sci U S A 102:18908–18913. https://doi.org/10.1073/pnas.0507603102

Townsend PD, Rodgers TL, Pohl E, Wilson MR, McLeish TCB, Cann MJ (2015) Global low-frequency motions in protein allostery: CAP as a model system. Biophys Rev 7:175–182. https://doi.org/10.1007/s12551-015-0163-9

Vendruscolo M, Paci E, Dobson CM, Karplus M (2001) Three key residues form a critical contact network in a protein folding transition state. Nature 409:641–645. https://doi.org/10.1038/35054591

Vishveshwara S, Ghosh A, Hansia P (2009) Intra and inter-molecular communications through protein structure network. Curr Protein Pept Sci 10:146–160. https://doi.org/10.2174/138920309787847590

Wako H, Endo S (2011) Ligand-induced conformational change of a protein reproduced by a linear combination of displacement vectors obtained from normal mode analysis. Biophys Chem 159:257–266. https://doi.org/10.1016/j.bpc.2011.07.004

Wako H, Endo S (2012) ProMode-Oligomer: database of normal mode analysis in dihedral angle space for a full-atom system of oligomeric proteins. Open Bioinformatics J 6:9–19. https://doi.org/10.2174/1875036201206010009

Wako H, Endo S (2013) Normal mode analysis based on an elastic network model for biomolecules in the Protein Data Bank, which uses dihedral angles as independent variables. Comput Biol Chem 44:22–30. https://doi.org/10.1016/j.compbiolchem.2013.02.006

Wako H, Gō N (1987) Algorithm for rapid calculation of hessian of conformational energy function of proteins by supercomputer. J Comp Chem 8:625–635. https://doi.org/10.1002/jcc.540080507

Wako H, Endo S, Nagayama K, Gō N (1995) FEDER/2: program for static and dynamic conformational energy analysis of macromolecules in dihedral angle space. Comp Phys Comm 91:233–251. https://doi.org/10.1016/0010-4655(95)00050-P

Wako H, Tachikawa M, Ogawa A (1996) A comparative study of dynamic structures between phage 434 Cro and repressor proteins by normal mode analysis. Proteins 26:72–80. https://doi.org/10.1002/(SICI)1097-0134(199609)26:1<72::AID-PROT7>3.0.CO;2-I

Wako H, Kato M, Endo S (2004) ProMode: a database of normal mode analyses on protein molecules with a full-atom model. Bioinformatics 20:2035–2043. https://doi.org/10.1093/bioinformatics/bth197

Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, Goodman SL, Arnaout MA (2002) Crystal structure of the extracellular segment of integrin αVβ3 in complex with an Arg-Gly-Asp ligand. Science 296:151–155. https://doi.org/10.1126/science.1069040

Yang LW, Liu X, Jursa CJ, Holliman M, Rader AJ, Karimi HA, Bahar I (2005) iGNM: a database of protein functional motions based on Gaussian network model. Bioinformatics 21:2978–2987. https://doi.org/10.1093/bioinformatics/bti469

Yang LW, Rader AJ, Liu X, Jursa CJ, Chen SC, Karimi HA, Bahar I (2006) oGNM: online computation of structural dynamics using the Gaussian network model. Nucleic Acids Res 34:W24–W31. https://doi.org/10.1093/nar/gkl084

Yang L, Song G, Jernigan RL (2009) Comparisons of experimental and computed protein anisotropic temperature factors. Proteins 76:164–175. https://doi.org/10.1002/prot.22328

Zavodszky MI, Kuhn LA (2005) Side-chain flexibility in protein–ligand binding: the minimal rotation hypothesis. Protein Sci 14:1104–1114. https://doi.org/10.1110/ps.041153605