ProMode: a database of normal mode analyses on protein molecules with a full-atom model

Hiroshi Wako¹,*, Masaki Kato² and Shigeru Endo³

¹School of Social Sciences, Waseda University, Tokyo 169-8050, Japan, ²Department of Science of Biological Supramolecular Systems, Graduate School of Integrated Science, Yokohama City University, Yokohama 222-0045, Japan and ³Department of Physics, School of Science, Kitasato University, Sagamihara 228-8555, Japan

ABSTRACT

Motivation: Although information from protein dynamics simulation is important to understand principles of architecture of a protein structure and its function, simulations such as molecular dynamics and Monte Carlo are very CPU-intensive. Although the ability of normal mode analysis (NMA) is limited because of the need for a harmonic approximation on which NMA is based, NMA is adequate to carry out routine analyses on many proteins to compute aspects of the collective motions essential to protein dynamics and function. Furthermore, it is hoped that realistic animations of the protein dynamics can be observed easily without expensive software and hardware, and that the dynamic properties for various proteins can be compared with each other.

Results: ProMode, a database collecting NMA results on protein molecules, was constructed. The NMA calculations are performed with a full-atom model, using dihedral angles as independent variables, faster and more efficiently than the calculations using Cartesian coordinates. In ProMode, an animation of the normal mode vibration is played with a free plug-in, Chime (MDL Information Systems, Inc.). With the full-atom model, the realistic three-dimensional motions at an atomic level are displayed with Chime. The dynamic domains and their mutual screw motions defined from the NMA results are also displayed. Properties for each normal mode vibration and their time averages, e.g. fluctuations of atom positions, fluctuations of dihedral angles and correlations between the atomic motions, are also presented graphically for characterizing the collective motions in more detail.

Availability: http://promode.socs.waseda.ac.jp

Contact: wako@waseda.jp

INTRODUCTION

Protein dynamics simulation is important to understand principles of architecture of a protein structure and its function. We have several conventional methods to simulate the protein dynamics by computer, e.g. normal mode analysis (NMA), Monte Carlo simulation (MC), and molecular dynamics simulation (MD) (Brooks et al., 1988). MC and MD are more reliable than NMA, because NMA is based on the harmonic approximation (i.e. MC and MD can cover much wider area of the conformational space of the protein than NMA). However, it has been shown by many studies that the results from NMA are not only qualitatively reasonable in most cases, compared to those from MC and MD, but also can provide a proper description of the functionally important motions of the protein (Go et al., 1983; Brooks and Karplus, 1983; Levitt et al., 1985; Gibrat and Go, 1990; Horiuchi and Go, 1991; Marques and Sanejouand, 1995; Hayward and Go, 1995; Wako et al., 1996; Thomas et al., 1996a,b; Hinsen, 1998; Kitao and Go, 1999; Miller and Agard, 1999; Kikuchi et al., 2000; Berendsen and Hayward, 2000; Chacon et al., 2003). Fortunately, NMA is much less time consuming in computation and can be performed in more systematic ways than MC and MD. Consequently, NMA is more adequate to carry out dynamics calculations routinely for many proteins and then, to describe the protein dynamics concisely.

ProMode is a database collecting the results from NMA carried out for many proteins. Although we should be careful in interpreting the results from NMA, owing to the approximation described above, the data presented in ProMode provide fruitful information about protein dynamics, especially for large-scale collective motions. The collection of the NMA results for various proteins is useful for comparative and statistical studies of the dynamic structures of proteins.

Another distinctive feature of ProMode is an animation of a vibrating protein played with a free plug-in, Chime (MDL Information Systems, Inc. http://www.mdl.com/). It is possible for a user with Chime not only to observe a concrete three-dimensional (3D) image in atomic detail about protein dynamics, but also to manipulate the animated protein molecule. The several kinds of time-averaged properties to characterize the motions are also provided graphically.

*To whom correspondence should be addressed.
Recently, a couple of databases for NMA of proteins were developed (Krebs et al., 2002; Chacon et al., 2003; Echols et al., 2003). They used a simplified molecular model and an energy potential function [e.g. an elastic protein model in which atoms representative of residues are connected with harmonic springs (Hinsen, 2000)]. In ProMode, however, a full-atom model is used. We used dihedral angles rather than Cartesian coordinates of atoms as independent variables. Since the number of variables can be reduced to one-eighth of the number of Cartesian coordinates, it is possible to perform NMA with the full-atom model even for relatively larger proteins. Consequently, ProMode can provide realistic molecular motions in full-atomic detail.

**METHODS**

**Regularization**

The most time-consuming part of NMA is regularization of the Protein Data Bank (PDB; Berman et al., 2000) data. We used a program, FEDER, developed by Go’s group for this purpose (Noguti and Go, 1983a,b; Wako and Go, 1987; Wako et al., 1995).

FEDER carries out a conformational energy minimization of a protein rapidly and efficiently with dihedral angles as independent variables (i.e. bond lengths and angles are fixed to their standard values). The number of the dihedral angles is about one-eighth of the number of atomic coordinates, and about 6N for a protein with N residues (compare to a simplified model, in which each residue is represented by one atom; in such a model, the number of variables is 3N). Using the algorithm for fast calculation of the first and second derivatives of a conformational energy function with respect to the dihedral angles, FEDER can precisely obtain a minimum point in the conformational energy space. It is strictly required in NMA that the second derivative matrix of the conformational energy function at the minimum should be positive definite. Since our calculations always satisfied this requirement, all normal modes could be obtained for any proteins.

FEDER generates various conformations with dihedral angles as independent variables. It is necessary to obtain a proper set of dihedral angles that generates a conformation not only at an energy minimum, but also very close to the PDB data whose bond lengths and bond angles are slightly deviated, in general, from their standard values used in FEDER, before NMA is performed. This procedure is referred to as regularization. For this requirement, we carried out the restrained energy-minimization with the following objective function.

\[
F = w \sum_{i,j} (r_{ij} - r_{ij}^X)^2 + pE,
\]

where \(r_{ij}\) and \(r_{ij}^X\) are the distances between atoms \(i\) and \(j\) in a calculated conformation and in the PDB conformation, respectively. \(E\) is a conformational energy function containing van der Waals, electrostatic, hydrogen bonding and torsional potentials (Wako et al., 1995).

If the objective function \(F\) is minimized, the first term works as a geometrical constraint to hold the molecular conformation close to the PDB data and the second term to lower the conformational energy. \(w\) and \(p\) are weight factors for the geometrical constraint and conformational energy terms, respectively. The minimization of the objective function is carried out, as \(p\) is set to zero at first. After the minimization converges, \(p\) and \(w\) are set to one and a large value, respectively, and then the minimization is resumed. The minimization is repeated, as \(w\) is decreased, step by step. Finally, by setting \(w = 0\), we can obtain the conformation located at a local minimum and very close to the PDB conformation. NMA is applied to such an energy-minimum conformation.

**Normal mode analysis**

Normal mode analysis is based on the assumption that a conformational energy surface can be characterized by a parabolic approximation at a minimum. One of the merits of NMA is that the derivation of an analytical solution to the equation of motion subject to this assumption is well established (Levitt et al., 1985). NMA divides the atomic motion into vibrational components called normal modes. A variety of atomic motions of a hundred normal modes (the number of normal modes is around six times the number of residues) in a protein, characterizes the dynamic structure of the protein. A time-averaged property of the motion, which is a superposition of all the normal modes, also provides important information on the protein dynamics (Go et al., 1983; Brooks and Karplus, 1983). The time-averaged properties of the motion depend only on the amplitudes of the normal mode vibrations, and the amplitude depends on the temperature. Calculations are carried out at 310 K (37°C) in ProMode.

The properties provided in ProMode are described below.

**RESULTS**

**Proteins**

We have collected NMA data mainly for proteins with less than 200 residues and with relatively low sequence similarity (<50%) to each other at present. The data for more than 1000 proteins have been collected already. The most time-consuming part of the calculations was the regularization. In Table 1, an example of the regularization schedule for human lysozyme (the number of residues is 130) is shown. The CPU time depends not only on protein size, but also on protein conformation. Especially, when the regularized conformation was largely deviated from the PDB one (see below for the reason why such a deviation occurred), the CPU time was much longer than the average one. However, once the regularized conformation was obtained for a given protein, the NMA calculation could be performed in short time [remember
that the number of variables (dihedral angles) is about 6N for a protein with N residues.

Owing to the computational difficulty in FEDER, any ligands, DNA and water molecules are omitted in our computations. For the proteins comprising two or more chains, each chain is treated independently. The calculation for a multimer system will be a future work.

The conformations of some chains were largely deviated from the PDB data after the regularization, mainly because the hetero molecules or other accompanying chains were ignored. The root-mean-square (RMS) deviations of about 60, 25 and 15% of the calculated proteins from the PDB conformations are less than 2, 2–4 Å and greater than 4 Å, respectively. In ProMode, the data of the proteins with the RMS deviation >2.0 Å are given with a warning message. It is interesting, however, to study which parts of such proteins were changed by the regularization. We are planning to study this problem in the near future. (For the problems in the regularization, see also Brunger 1993).

We will collect data on as many proteins as possible, including larger sized ones. Since a comparative study among homologous proteins is interesting (Wako et al., 1996), we will also add NMA results for such proteins to ProMode.

**Properties presented**

ProMode provides several properties characterizing the dynamics of the proteins. The properties that can be calculated in NMA, however, have a variety, and the total sizes of the resulting files are too large to archive in general. Consequently, the data provided in ProMode have to be limited. What follows are the properties provided in ProMode at present.

(i) **Time-averaged properties.**

(a) **Positional fluctuations of atoms:** The absolute values of the normal mode displacement vectors are averaged over all atoms or over backbone atoms in each residue, respectively, and plotted against residue numbers (Fig. 1a).

(b) **Fluctuations of dihedral angles:** The data on the dihedral angles, \( \phi \), \( \varphi \) and \( \chi \), are plotted against residue numbers (Fig. 1b).

---

### Table 1. An example of regularization schedule

<table>
<thead>
<tr>
<th>Range</th>
<th>Weighting factors(^c) ((\text{kcal/mol/Å}^2\text{, for } w))</th>
<th>Convergence(^d) ((\text{kcal/mol}))</th>
<th>CPU(^e) ((s))</th>
<th>No. of iterations(^f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>12.4 ± 2.9</td>
<td>25 ± 0.7</td>
</tr>
<tr>
<td>2</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>18.0 ± 4.5</td>
<td>29 ± 2.4</td>
</tr>
<tr>
<td>3</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>16.1 ± 1.4</td>
<td>21 ± 1.0</td>
</tr>
<tr>
<td>4</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>11.6 ± 2.6</td>
<td>20 ± 1.1</td>
</tr>
<tr>
<td>5</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>9.4 ± 0.5</td>
<td>18 ± 0.9</td>
</tr>
<tr>
<td>6</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>9.0 ± 0.9</td>
<td>18 ± 1.6</td>
</tr>
<tr>
<td>10</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>9.1 ± 0.3</td>
<td>17 ± 0.6</td>
</tr>
<tr>
<td>14</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>9.6 ± 1.2</td>
<td>17 ± 0.6</td>
</tr>
<tr>
<td>18</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>8.8 ± 0.8</td>
<td>15 ± 1.3</td>
</tr>
<tr>
<td>30</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>9.8 ± 0.8</td>
<td>15 ± 1.3</td>
</tr>
<tr>
<td>60</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>12.1 ± 0.9</td>
<td>15 ± 1.1</td>
</tr>
<tr>
<td>90</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>13.6 ± 1.2</td>
<td>14 ± 1.1</td>
</tr>
<tr>
<td>120</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>16.5 ± 6.5</td>
<td>16 ± 4.6</td>
</tr>
<tr>
<td>150</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>17.5 ± 7.4</td>
<td>14 ± 4.7</td>
</tr>
<tr>
<td>200</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>18.7 ± 4.5</td>
<td>13 ± 0.7</td>
</tr>
<tr>
<td>300</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>27.0 ± 8.5</td>
<td>17 ± 2.8</td>
</tr>
<tr>
<td>All</td>
<td>( p = 0 )</td>
<td>(10^{-3})</td>
<td>21.5 ± 5.7</td>
<td>13 ± 1.5</td>
</tr>
<tr>
<td>All</td>
<td>( p = 1, w = 100 )</td>
<td>(10^{-10})</td>
<td>482.2 ± 181.2</td>
<td>119 ± 40.7</td>
</tr>
<tr>
<td>All</td>
<td>( p = 1, w = 1 )</td>
<td>(10^{-10})</td>
<td>189.2 ± 89.8</td>
<td>48 ± 19.4</td>
</tr>
<tr>
<td>All</td>
<td>( p = 1, w = 0.1 )</td>
<td>(10^{-10})</td>
<td>254.3 ± 91.3</td>
<td>66 ± 23.0</td>
</tr>
<tr>
<td>All</td>
<td>( p = 1, w = 0.01 )</td>
<td>(10^{-10})</td>
<td>419.4 ± 115.2</td>
<td>109 ± 15.3</td>
</tr>
<tr>
<td>All</td>
<td>( p = 1, w = 0 )</td>
<td>(10^{-12})</td>
<td>439.9 ± 185.9</td>
<td>114 ± 38.2</td>
</tr>
<tr>
<td>All</td>
<td>( p = 1, w = 0 )</td>
<td>(10^{-12})</td>
<td>141.1 ± 84.0</td>
<td>57 ± 27.5</td>
</tr>
</tbody>
</table>

Mean total CPU time: 2166 s

---

\(^a\) The data of the regularization for wild type and 19 mutants of human lysozyme are averaged.

\(^b\) The range within which two atoms are possible to interact with each other in the calculation is shown. The distance between the two dihedral angles along the chain (defined in the paper by Wako et al. (1995)) is used to specify the range. The range is gradually expanded to avoid being trapped at a local minimum.

\(^c\) The convergence criterion is given with an energy difference between the two conformations subsequently generated in the regularization.

\(^d\) The calculation was performed on the Pentium 4 processor (2.13 GHz) with 512 MB memory. The mean and SD over 20 human lysozymes are given.

\(^e\) The calculation for a multimer chain is treated independently. The calculation for a multimer system will be a future work.

\(^f\) The mean and SD of iteration times in the optimization of the objective function (1) over 20 human lysozymes are given.
Fig. 1. Examples of time-averaged properties represented graphically in ProMode (these figures are not a screenshot of ProMode, but reproduced with some modifications). (a) Positional fluctuations of atoms, which are absolute values of normal mode displacement vectors averaged over all atoms and averaged over backbone atoms in each residue, respectively. (b) Fluctuations of dihedral angles, $\phi$ and $\varphi$. (c) Correlations between the positional fluctuations of atoms, which are defined as normalized dot products of the normal mode displacement vectors of C$_\alpha$ atoms. The closed and open squares in the triangle map indicate that the correlations of these pair of residues are greater than 0.2 and less than $-0.2$, respectively. The protein shown here is BPTI (PDB ID = 5pti).

(c) **Correlations between the positional fluctuations of atoms:** The normalized dot products of the normal mode displacement vectors of C$_\alpha$ atoms are presented in a triangle map (Fig. 1c).

(ii) Properties of each normal mode vibration (only for the 20 lowest-frequency modes).

(a) **Positional fluctuations of atoms:** The absolute values of the normal mode displacement vectors are averaged over all atoms or over backbone atoms in each residue, respectively, and plotted against residue numbers.

(b) **Fluctuation of dihedral angles:** The data on the dihedral angles, $\phi$, $\varphi$ and $\chi_1$, are plotted against residue numbers.

(c) **Correlations between the atomic motions:** The normalized dot products of the normal mode displacement vectors of C$_\alpha$ atoms are presented in a triangle map.

(d) **Results from DynDom.**

(e) **Animation of normal mode vibration at 310 K.** (The amplitude is too small to observe it in most cases. To exaggerate the atomic movements, the animation in which the amplitude at 310 K is doubled can be also played.)

In ProMode, the provided data are limited to the 20 lowest-frequency normal modes, because it is well known that atomic fluctuations within the subspace spanned by a small number of the lowest-frequency normal modes dominantly contribute to total fluctuation (Go et al., 1983; Levy et al., 1984; Thomas et al., 1996a,b; Kidera and Go, 1992; Krebs et al., 2002). Although some particular higher frequency modes may contribute to the local fluctuations significant for a protein...
Database of normal mode analyses on proteins

Fig. 2. Screenshot of ProMode. A protein is displayed in the two Chime windows: (left) 3D animation; (right) static 3D image; the energy-minimum conformation and the two most fluctuated ones superimposed onto it. The mouse event for translation, rotation and zoom on one window affects the other window simultaneously. The dynamic domains defined by DynDom are distinguished by different colors. The axes of screw motions between the domains are shown in the right window. The manipulation dialog (superimposed in the upper left) can be used for selecting residues and/or atoms and executing commands not available in the pop-up menu of Chime.

function, they are not presented in ProMode. If we could detect such modes from a hundred modes automatically, we would present the data for such modes in addition to the lowest ones. We are also thinking about how to provide the data for all the modes.

The program DynDom, developed by Hayward et al. (1997), is used in ProMode. It is a program to determine dynamic domains, hinge axes and amino acid residues involved in the hinge bending motion, when the two different conformations for a protein are given. In ProMode, the energy minimum conformation and the most fluctuated one in each normal mode motion are used as input data for DynDom. The results characterize the normal mode vibrations well.

To see the animation, DynDom results, and the energy-minimum conformation superimposed onto the PDB one, a plug-in, Chime, is necessary to be installed. Chime is free, and can play an animation. Since the input data file for the animation has no information about amino acid residues in Chime (which requires only atomic types and their Cartesian coordinates), it is difficult for a user to manipulate the protein displayed in the Chime animation beyond the original pop-up menu. For this problem, we created a manipulation dialog to execute scripts and commands with which a user can specify particular atoms and/or residues (e.g. an active site, a hydrophobic core and a sequence/structural motif) to highlight by changing color and atomic representation (Fig. 2).

Example for a comparative study

With the collection of the results from the NMA for various proteins in ProMode, e.g. it is possible to make a comparative study among a wild type protein and its mutants, or among homologous proteins (Wako et al., 1996; Kikuchi et al., 2000). Here, we will illustrate an example for such a study with bovine pancreatic trypsin inhibitor (BPTI). The 3D structures
H.Wako et al.

Fig. 3. Stereoscopic view of the 3D structure of BPTI (5pti) drawn by RASMOL (Sayle, 1995). The chain is shaded according to the three dynamic domains defined for the second lowest frequency normal mode by DynDom (Fig. 4). The residue numbers and side chains are shown for the replaced residues and the residues forming a disulfide bond.

of a wild type and several mutant BPTIs have been determined. The following data are used in this paper: the wild type (PDB ID=5pti; Wlodawer et al. 1984; Fig. 3), Y23A mutant (1bpt; Danishefsky et al. 1993), F22A mutant (1bti; Danishefsky et al. 1993), F45A mutant (1fan; Danishefsky et al. 1993), N43G mutant (1nag; Danishefsky et al. 1993), C30A and C51A mutant (7pti; Eigenbrot et al. 1990) and Y35G mutant (8pti; Housset et al. 1991). The number of residues of BPTI is 58. In the data of 1bpt and 1nag, however, the two C-terminal residues, Gly-57 and Ala-58, are not visible.

Figure 1 shows the positional fluctuations of atoms and dihedral angles, $\phi$ and $\psi$ for the wild type BPTI. The positional fluctuations in the secondary structures (the two $\alpha$-helices, 3–6 and 48–55 and the two $\beta$-strands 18–24 and 29–35) are relatively small, whereas the larger fluctuations are observed in the loop regions, 15–17, 25–27 and 39–42. Exceptionally, the residues 50–53 in the $\alpha$-helix fluctuate largely. According to the animations in ProMode, it seems that the large fluctuations of the loop region 25–27 affect the C-terminal $\alpha$-helix mainly through the disulfide bond 30–51. The fluctuations of dihedral angles are larger not only in the loop regions, but also in some residues in the secondary structures. It is notable that the fluctuations of dihedral angles of the residues 25–27 are not very large, whereas their positional fluctuations are large. This region is a hairpin loop located between the two $\beta$-strands forming a $\beta$-sheet. It seems that the larger fluctuations of dihedral angles in the two $\beta$-strands are responsible for the larger positional fluctuations of the loop.

Figure 4 shows the dynamic domains defined by DynDom for the four lowest frequency normal modes of the seven BPTIs, 5pti, 1bpt, 1bti, 1fan, 1nag, 7pti and 8pti. In each mode, the region with the same color indicates that the residues in the region belong to the same dynamic domain. The white regions are not defined as dynamic domains. For some modes, no dynamic domains are defined. The short double bars indicate hinge regions.

modes, two dynamic domains (three dynamic domains for two of the modes) are defined. For some modes, however, no dynamic domains are defined. One dynamic domain consists of the half of the $\beta$-sheet containing the hairpin loop (around 21–32) and the N- and C-terminals connected to each other with the disulfide bond 5–55. In two of the modes, the C-terminal region is distinguished as a different domain. The other dynamic domain consists of the other half of the $\beta$-sheet and the two long loops connected to each other with the disulfide bond 14–38. Each domain consists of more than one region of consecutive residues. In Figure 3, we can see that these regions, which are separated along the chain, are put together into the domains reasonably in the 3D structure of BPTI (more detail views related to Figs 3 and 4 are available in ProMode).

Figure 5 shows the differences in positional fluctuations of atoms and dihedral angle $\psi$ of the six mutant BPTIs from the wild type. We notice at first that the changes in the fluctuations
Database of normal mode analyses on proteins

Fig. 5. Differences in positional fluctuations of atoms (upper panel) and dihedral angle ϕ (lower panel). The differences of the six mutant BPTIs (1bpt, 1bti, 1fan, 1nag, 7pti and 8pti) from the wild type (5pti) are plotted. The replaced residues are indicated in the parentheses in the legend above the upper panel. The secondary structures (the two α-helices and the two β-sheets), and the dynamic domains and hinge residues defined for the second-lowest normal mode of 5pti by DynDom (Fig. 4), are also indicated.

do not necessarily occur around the replaced residues. It is expected that the substituted residues can fluctuate much more, because the amino acids with larger sidechains in the wild type were replaced by those with smaller sidechains in any mutations. In addition, comparing the fluctuations of atoms and dihedral angles for each residue in Figure 5, we found that the residues with larger fluctuations of the dihedral angles are not necessarily the residues with larger fluctuations of the atoms. Incidentally, the replaced residues 22, 23, 35, 43 and 45 are located near hinge regions (Fig. 4), and most of them are hydrophobic residues forming a hydrophobic core. As a result, the changes in these residues seem to affect the movements of the dynamic domains such as regions 9–17, 25–28 and 40–44.

Fig. 6. Correlations between the positional fluctuations of atoms, which are defined as normalized dot products of the normal mode displacement vectors of Cα atoms for the mutant BPTIs, 1nag and 7pti. See also Figure 1c.

Finally, the suppression of the fluctuations in both atomic positions and dihedral angles at residues 51–52 is remarkable in any mutants (Fig. 5). In other words, these fluctuations are larger only in the wild type BPTI. According to Figure 1c, the two regions, 26–34 and 52–58, move in a negatively correlated manner on average. Such correlative motions of these residues, however, become weaker in the mutant BPTIs (Fig. 6). In general, the number of pairs of residues with relatively larger positive or negative correlations decreases in the mutant BPTIs, compared with the wild type BPTI. Several
factors, such as the changes in the hinge region, the hydrophobic core and the disulfide bond, are considered responsible for the changes in the correlative motions. Further investigation is necessary to reveal to what extent and how these factors affect the changes.

In this kind of study, we observe the responses to the perturbations of the amino acid substitutions. By collecting several cases, not only the characteristic responses in the individual cases, but also common features of the dynamic properties over the mutant proteins are revealed. Since our aim here is to show the analyses of BPTIs and to give a demonstration on how to use the collection of the NMA data in ProMode, we will leave discussion on the detail of the individual mutant BPTIs for another occasion.

DISCUSSION

ProMode provides dynamic properties of various proteins obtained from NMA. By using the dihedral angles as independent variables, we can perform NMA with a full-atom model. Consequently, the animation of the vibrational motions of the proteins can be displayed with a full-atom model. However, one of the problems in ProMode is computational expense. Especially, the regularization performed before the NMA calculation requires much CPU time. On the other hand, in the simplified models (Krebs et al., 2002; Chacon et al., 2003; Echols et al., 2003), the regularization is not necessary, because a PDB conformation is already at an energy minimum by their definition. As a result, it is possible to execute the NMA calculations for the entire PDB and to immediately provide a result on demand by the simplified models, while it is hard in ProMode. Alternatively, information on NMA with the full-atom model is obtained in ProMode. The simplified models and ProMode are complementary to each other in the NMA databases.

A comparison with the MD calculations is important. As for the CPU time, since the PDB conformation is not at a minimum of a potential energy function one uses usually (except for the simplified models), the energy minimization must be performed even in MD. In addition, equilibration is required in advance of the simulation. These procedures are comparable to the regularization in ProMode. The NMA calculation after the regularization can be performed very fast, whereas the MD simulation for statistical analysis after the equilibration requires much more CPU time. As for the memory requirement, it is proportional only to the number of variables, N, in MD, but to N² in NMA. Even though N² in ProMode is reduced to about 1/64 of the NMA using Cartesian coordinates as independent variables, it is still much larger than N in MD. Since MD can cover much wider area of the conformational energy space, MD may replace NMA in the databases of protein dynamics in future, when the CPU time required for MD will become much less than at present.

Visual expression of the NMA results is another characteristic point of ProMode. Not only the global picture of the protein dynamics such as domain motions, but also the fluctuation of the local structures and side chain motions, can be observed realistically through the animation of the full-atom model. By highlighting specific atoms and/or residues (e.g. an active site, a hydrophobic core and sequence/structural motif, with the manipulation dialog we made), it is possible to obtain more detailed information about the dynamics of the protein. The atomic motions and dihedral angle fluctuations plotted against residue numbers and the triangle map for the correlations between the positional fluctuations of atoms are useful to characterize the dynamics of the protein at residue and atom levels. The dynamic domains defined by DynDom and their mutual motions are also helpful to understand the dynamics of the proteins. With the collection of the NMA results for the various proteins, we can compare those results with each other and perform statistical analyses. We expect that such a study will provide new aspects of dynamics of protein structures.

ACKNOWLEDGEMENTS

We gratefully acknowledge Professor Haruki Nakamura, Osaka University, who is a supervisor of one of the BIRD projects, PDBj (Protein Data Bank Japan: http://www.pdbj.org). We also gratefully acknowledge Professor Nobuhiro Go, Japan Atomic Energy Research Institute, who was a supervisor in creating the program FEDER and encouraged us to construct ProMode.

This work was financially supported by BIRD-JST (Institute for Bioinformatics Research and Development of Japan Science and Technology Corporation). This work is running as a subgroup of the PDBj project. ProMode is designed and maintained in collaboration with NEC Soft, Ltd. and Science House, Inc.

REFERENCES


