Biophysical Chemistry 159 (2011) 257–266

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Ligand-induced conformational change of a protein reproduced by a linear combination of displacement vectors obtained from normal mode analysis

Hiroshi Wako [a,*], Shigeru Endo [b]

[a] School of Social Sciences, Waseda University, Tokyo 169-8050, Japan
[b] Department of Physics, Faculty of Science, Kitasato University, Sagamihara 252-0373, Japan

ARTICLE INFO

Article history:
Received 22 April 2011
Received in revised form 9 July 2011
Accepted 9 July 2011
Available online 21 July 2011

Keywords:
Ligand binding
Holo-form protein
Apo-form protein
Normal mode analysis
Conformational change
Conformational sampling

ABSTRACT

The conformational change of a protein upon ligand binding was examined by normal mode analysis (NMA) based on an elastic-network model (ENM) for a full-atom system using dihedral angles as independent variables. Specifically, we investigated the extent to which conformational change vectors of atoms from an apo form to a holo form 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 ($m = 1, 2, \ldots, 20$). In this analysis, the latter vectors were best fitted to the former ones by the least-squares method. Twenty-two paired proteins in the holo and apo forms, including three dimer pairs, were examined. The results showed that, in most cases, the conformational change vectors were reproduced well by a linear combination of the displacement vectors of a small number of low-frequency normal modes. The conformational change around an active site was reproduced as well as the entire conformational change, except for some proteins that only undergo significant conformational changes around active sites. The weighting factors for 20 normal modes optimized by the least-squares fitting characterize the conformational changes upon ligand binding for these proteins. The conformational changes sampled around the apo form of a protein by the linear combination of the displacement vectors obtained by ENM-based NMA may help solve the flexible-docking problem of a protein with another molecule because the results presented herein suggest that they have a relatively high probability of being involved in an actual conformational change.

1. Introduction

Ligand binding to a protein is an essential process in many biochemical phenomena. The lock-and-key model by Fisher [1] and the induced-fit model by Koshland [2] are well-known hypotheses for the binding mechanism. The lock-and-key model was proposed to explain the ligand specificity of an enzyme, but does not necessarily support the rigidity of the active site of a protein. As postulated in the induced-fit hypothesis, a conformational change to either a larger or smaller size should play an essential role in the ligand-binding mechanism. In fact, conformational changes of the side chains of a binding site [3,4], small movements and/or deformation of loops [5,6], large-scale domain motions [7,8], and quaternary movements of subunits [9,10] have all been observed during the binding of a ligand to proteins. According to a recent study by Brylinkski and Skolnick [11] on 521 paired protein structures in the ligand-free and ligand-bound forms, most proteins undergo relatively small conformational changes upon ligand binding, i.e., their root-mean-square differences (RMSDs) are less than 1 Å.

The intrinsic thermal motion of proteins is considered to be responsible for the conformational changes upon ligand binding. Normal mode analysis (NMA) is one of the most promising methods to explore such an intrinsic motion around the native conformation. In fact, it has been found that the observed conformational change upon ligand binding correlates well with the conformational change of some low-frequency normal mode vibrations [12–16]. In these studies, most conformational changes with high correlations are large-scale domain motions because the low-frequency normal modes, which are usually examined in NMA, cause a concerted motion of the atoms.

In the ligand and protein docking problem, a simplistic rigid model for the ligand–protein interaction is inadequate and the incorporation of ligand and protein flexibility is required for an accurate docking prediction [17,18]. However, the prediction of a conformational change upon ligand binding remains a formidable challenge because it is difficult to generate adequate perturbed conformations in a reasonable computation time. This problem has been addressed by a variety of algorithms; NMA is one of such methods [19]. However, NMA has only been applied to systems comprising C^{α} atoms regarding conformational changes upon ligand binding. For the flexible docking problem, the modification of NMA for a more detailed molecular model including side-chain atoms is required. In addition, it

* Corresponding author at: 1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050, Japan. Tel.: +81 3 5286 1430; fax: +81 3 5272 0168.
E-mail addresses: wako@waseda.jp (H. Wako), endo@sci.kitasato-u.ac.jp (S. Endo).

0301-4622/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2011.07.004

is necessary to find a methodology to select a reasonably small number of representative conformational changes from the enormous number of conformations generated.

We have modified FEDER/2 [20,21] for NMA based on an elastic-network model (ENM). This program, named PDBETA, uses dihedral angles as independent variables. Therefore, the number of independent variables used to describe a protein structure for a full-atom system is much smaller than that of other NMA programs using Cartesian coordinates for the same full-atom system, but comparable to that for a C^{α}-atom-only system. For example, while 3 variables are required for a residue represented by only C^{α} atoms using the Cartesian-coordinate system, an average of around 4.7 variables, e.g., dihedral angles φ, ψ, ω, χ_1, and χ_2, are sufficient for a residue comprising all of the non-hydrogen atoms using the dihedral-angle system. Accordingly, it is practical to account for all atoms in NMA calculations using PDBETA. The extension of this calculation to a system comprising more than one molecule, such as an oligomer or a complex with other molecules, is also possible [21]. Even when non-protein molecules are incorporated, it is not necessary to select representative atoms to reduce the number of variables. In fact, PDBETA can perform NMA calculations with all of the atoms, including those of non-protein molecules, contained in the Protein Data Bank (PDB) data. Although the ENM-based NMA method sacrifices accuracy, it has been successfully applied to exploring the relationship between function and dynamics for many proteins [22–24]. Furthermore, its rapid computation is a great advantage attained by skipping the conformational-energy minimization required in orthodox NMA.

In this paper, we investigated two points associated with the above-mentioned problems: (*i*) the applicability of ENM-based NMA for a full-atom system to the ligand-binding problem and (*ii*) the possibility of properly representing the conformational changes as an ensemble of various normal modes. With regard to the latter point, although we cannot present a practical methodology for the docking problem in this paper, our research suggests that conformational sampling by a linear combination of displacement vectors generated by ENM-based NMA for a full-atom system can be used to predict conformational changes upon ligand binding.

2. Materials and methods

2.1. Examination of the holo-form and apo-form protein pairs

We examined twenty-two protein structure pairs in the holo and apo forms. Their PDB codes and protein names are as follows: 1usiA–1usgA, L-leucine binding protein; 1akeA–4akeA, adenylate kinase; 1q0qA–1k5hA, 1-deoxy-d-xylulose-5-phosphate reductoisomerase; 1sw2A–1sw5A, osmoprotection protein; 1y3nA–1y3qA, alginate binding protein; 1rpjA–1gudA, D-allose binding protein; 1rf4A–1rf5A, 5-enolpyruvylshikimate-3-phosphate synthase; 1o03A–1zolA, β-phosphoglucomutase; 1tj5A–1s2oA, sucrose phosphatase; 2gkeA–1gqzA, diaminopimelate epimerase; 1vhlA–1viyC, dephospho-CoA kinase; 1q95A–1za1A, aspartate transcarbamylase; 1cg0A–1hooB adenylosuccinate synthetase; 1ftmA–1ftoA, glutamate receptor; 1eejB–1tjdA, disulfide bond isomerase; 1jg6A–1jejA, β-glucosyltransferase; 1rpfA–1aqpA, ribonuclease A; 1cimA–1ca2A, carbonic anhydrase II; 1kpeB–1kpaB, protein kinase C interacting protein; 1hiiAB–1hsiAB, HIV-2 protease; and 2rveAB–1rveAB and 4rveAB–1rveAB, EcoRV endonuclease. The proteins were mainly selected from the study by Brylinski and Skolnick [11] as examples with large-scale domain motions upon ligand binding. Proteins in studies by Najmanovich et al. [3] and Seeliger and Groot [25] were also considered. In addition, some proteins with smaller conformational changes upon ligand binding were included in the examination for comparison. The last three pairs in the above list, i.e., 1hiiAB–1hsiAB, 2rveAB–1rveAB, and 4rveAB–1rveAB, are dimer pairs in the holo and apo forms comprising subunits A and B. Although ligand binding to oligomer proteins was of interest, only these three pairs were examined in this paper.

2.2. Normal mode analysis of the apo form

NMA was carried out by the PDBETA program that was developed by the authors. PDBETA is an ENM-based NMA program that uses dihedral angles as independent variables. It is possible to include all the atoms of any molecule, including both the protein and non-protein molecules, in the PDB data. Except for the potential function, the algorithm for NMA in dihedral angle space is the same as that used in the program FEDER/2 [20,21] and the NMA database *Pro*Mode [26] developed earlier by the authors. In PDBETA, the potential function, $E(r)$, for an atomic pair separated by a distance, r, is as follows [22]:

$$E(r) = c\,\exp\left[-\left(r^{\mathrm{PDB}}/r_0\right)^2\right]\left(r-r^{\mathrm{PDB}}\right)^2, \tag{1}$$

where r^{PDB} is the actual distance of the relevant atom pair in the PDB data. The exponential decay of the coefficient allows for the exclusion of significantly distant atomic pairs. The constant r_0 was set to 5 Å since the results are less sensitive to this value. The value of c is arbitrary because it causes a uniform scaling of all vibrational frequencies.

Owing to the small number of variables required for PDBETA, all of the atoms in the protein were accounted for in the NMA calculation. Notable exceptions are the hydrogen atoms, which are usually ignored to reduce the number of variables even if they are provided in the PDB data. Additionally, some side-chain atoms were discarded if they project from the protein surface into the solvent and have few interactions with other atoms; the atoms in the long side chains of amino acids such as Arg, Lys and Met are usually subject to this exclusion criterion when they are on the protein surface. Since such atoms can move quite freely, their movements are extraordinarily large and could therefore possibly disturb the analysis.

The 20 lowest-frequency normal modes were calculated for the apo-form protein structures. It is well known that atoms fluctuate in a more concerted manner in the low-frequency normal modes. Although there is a possibility that higher-frequency normal modes play a significant role in local conformational changes upon ligand binding, we focused our attention on the low-frequency normal modes in this study.

For the normal mode, i ($i=1, 2, \ldots, 20$), a displacement vector of atom α, $a_{\alpha,i}$, was calculated. Since temperature cannot be defined in the elastic network model, the displacement vector $a_{\alpha,i}$ is normalized such that $\sum_{\alpha\in \mathrm{H}} a_{\alpha,i}^2 = 1\,\text{Å}^2$ in the subsequent analysis. The summation is taken over a specified set of atoms, H; we considered two atom sets, H = {all atoms} and H = {C^{α} only}, as described below.

2.3. Least-squares fitting of a linear combination of normal-mode displacement vectors

Two conformations of the apo form and holo form of a protein, which are represented by atomic coordinate sets $\{x_\alpha\}$ and $\{y_\alpha\}$, respectively, were best-fitted. The difference between the two superimposed conformations is given by a set of difference vectors: $\{b_\alpha = y_\alpha - x_\alpha\}$. Hereafter, b_α is referred to as the conformational change vector. The next question to address is to what extent the linear combination of m normal-mode displacement vectors, $\sum_{i=1}^{m} w_i a_{\alpha,i}$, can be least-squares fitted to the conformational change vector, b_α, by adjusting weighting factors, w_i ($i=1,2,\ldots,m$; $m=1,2,\ldots,20$). Although it is natural that the conformational change be represented by the linear combination of the dihedral angles that are the independent variables in the NMA calculation used here, we did not do that for the following reason. Since the molecular geometries

(bond lengths and bond angles) in the holo-form and the apo-form PDB structures are not exactly the same, even the same set of dihedral angles cannot generate the same structure for the holo- and apo-form proteins. Accordingly, we decided that it was improper to discuss conformational differences using dihedral angles. In addition, we considered that the conformational changes represented by the atomic displacement vectors are more intuitive than the dihedral angle changes, and made it easier to apply the results obtained here to other calculations, such as a docking simulation using Cartesian coordinates as independent variables.

The least-squares method finds the optimum set of weighting factors, w_i, with which the sum of the squared residuals

$$F_m = \frac{1}{n}\sum_{\alpha\in \mathrm{H}}\left(b_\alpha - \sum_{i=1}^{m} w_i a_{\alpha,i}\right)^2 \quad (2)$$

is at a minimum. The summation includes a specified set of atoms, H, as described above, and the normalization factor, n, is the number of atoms in H. The minimum of the sum of the squares is found by setting the gradient to zero with respect to w_k, i.e., $\frac{\partial F_m}{\partial w_k} = 0$. The m gradient equations are then

$$\sum_{i=1}^{m} w_i\left(\sum_{\alpha\in \mathrm{H}} a_{\alpha,i}a_{\alpha,k}\right) = \sum_{\alpha\in \mathrm{H}} b_\alpha a_{\alpha,k} (k = 1,2,\ldots,m). \quad (3)$$

We numerically solved this set of equations using LAPACK (Linear Algebra Package; http://www.netlib.org/lapack/).

Using the weighting factors, w_i, obtained by the least-squares method, we can represent an optimum conformation, q_α, deformed from an apo form, x_α, as

$$q_\alpha = x_\alpha + \sum_{i=1}^{m} w_i a_{\alpha,i}. \quad (4)$$

Hereafter, we refer to this conformation as a least-squares-fitted (LSF) conformation. With this optimum set of weighting factors, Eq. (2) is rewritten as

$$F_m = \frac{1}{n}\sum_{\alpha\in \mathrm{H}} (y_\alpha - q_\alpha)^2. \quad (5)$$

The square root of F_m is the root-mean-square difference (RMSD) between the holo-form and LSF conformations.

3. Results

3.1. General results

For the twenty-two sets of apo-form and holo-form protein pairs, we performed least-squares fitting of the linear combination of the lowest-frequency m normal mode displacement vectors, $\sum_{i=1}^{m} w_i a_{\alpha,i}$ ($m = 1, 2, \ldots, 20$), to the observed conformational change vectors, b_α, from an apo form to a holo form with weighting factors, w_i, as unknown parameters. Two atom sets were considered in the least-squares fitting: (*i*) a set of all atoms and (*ii*) a set of C^α atoms only, referred to as "all" and "C^α," respectively. In both cases, NMA was performed for the same full-atom system. Table 1 shows the RMSDs of the LSF conformation from the holo form calculated using Eq. (5) for $m = 10$ and 20, together with the RMSD between the holo form and the apo form in PDB. They are referred to as $RMSD_{10}$, $RMSD_{20}$, and $RMSD_{PDB}$, respectively.

Out of 22 pairs, 17 pairs have an $RMSD_{10}$(all), $RMSD_{10}(C^\alpha)$, $RMSD_{20}$ (all), and $RMSD_{20}(C^\alpha)$ of less than 1.82, 1.44, 1.70, and 1.18 Å, respectively. However, the following five pairs are exceptions: 1usiA–1usgA, 1akeA–4akeA, 1q0qA–1k5hA, 1sw2A–1sw5A, and 4rveAB–1rveAB. Although these pairs have rather large $RMSD_{PDB}$ values of 7.30, 7.19, 5.28, 5.28, and 4.13 Å, respectively, they are significantly lowered by the results of the least-squares fitting, which are the $RMSD_{10}$ (all) values of 2.11, 2.06, 4.19, 1.93, and 2.24 Å, respectively, and the $RMSD_{20}$(all) values of 2.08, 1.90, 3.56, 1.91, and 2.08 Å, respectively. However, the $RMSD_{10}$ and $RMSD_{20}$ values of the five pairs are still larger than those of the other 17 pairs.

The differences between $RMSD_{10}$ and $RMSD_{20}$ are small, at less than 0.1 Å, for most protein pairs. This fact indicates that the higher-frequency normal modes do not contribute much to deformation by least squares fitting.

Fig. 1 shows how $RMSD_m$ decreases as m increases for the three examples of the holo-apo form pairs of 1akeA–4akeA, 1rpjA–1gudA,

Table 1
Results of least-squares fitting of a linear combination of normal-mode displacement vectors to conformational change vectors.

Holo	Apo	No. of residues	$RMSD_{PDB}$(Å)		$RMSD_{10}$(Å)		$RMSD_{20}$(Å)		No. of residues[b]	RMSD (Å) of an active site	
			All	C^α	All	C^α	All	C^α		Apo–holo	LSF–holo
1usiA	1usgA	346	7.30	7.14	2.11	1.92	2.08	1.88	12	6.36	2.26
1akeA	4akeA	214	7.19	7.13	2.06	1.72	1.90	1.55	33	6.09	2.01
1q0qA	1k5hA	398	5.28	4.93	4.19	3.69	3.56	2.94	34	5.32	5.67
1sw2A	1sw5A	270	5.28	5.04	1.93	1.61	1.91	1.58	9	3.44	1.50
1y3nA	1y3qA	490	4.89	4.79	1.40	1.10	1.32	1.00	14	2.85	1.15
1rpjA	1gudA	288	4.71	4.45	1.60	1.24	1.52	1.18	17	4.87	1.88
1rf4A	1rf5A	427	3.92	3.69	1.69	1.18	1.66	1.17	18	5.32	3.72
1o03A	1zolA	221	3.40	3.11	1.67	1.03	1.58	0.92	17	2.26	1.33
1tj5A	1s2oA	244	3.37	3.30	1.01	0.65	0.97	0.59	16	2.35	0.83
2gkeA	1gqzA	274	2.63	2.45	1.26	0.85	1.13	0.70	26	2.95	2.05
1vhlA	1viyC	208	2.59	2.61	1.18	0.93	1.15	0.91	13	1.11	1.08
1q95A	1za1A	310	2.57	2.25	1.67	1.22	1.61	1.14	15	2.45	2.18
1cg0A	1hooB	431	2.42	2.21	1.82	1.44	1.70	1.17	37	3.35	3.31
1ftmA	1ftoA	257[a]	2.32	2.26	0.77	0.43	0.75	0.39	13	1.93	0.74
1eejB	1tjdA	216	2.31	2.08	1.53	1.18	1.36	0.96	4	0.24	0.30
1jg6A	1jejA	351	2.29	2.10	1.18	0.81	1.16	0.77	15	2.78	2.55
1rpfA	1aqpA	124	0.91	0.51	0.82	0.36	0.77	0.31	9	0.58	0.53
1cimA	1ca2A	256	0.61	0.21	0.60	0.18	0.60	0.17	17	0.63	0.60
1kpeB	1kpaB	113	0.53	0.26	0.50	0.20	0.50	0.19	1	0.05	0.04
1hiiAB	1hsiAB	198	2.18	1.90	1.40	0.90	1.34	0.84	28	1.76	1.17
2rveAB	1rveAB	426[a]	4.76	4.55	1.73	1.33	1.58	1.14	28	5.09	2.79
4rveAB	1rveAB	480[a]	4.13	3.83	2.24	1.77	2.08	1.57	50	4.25	2.87

[a] The numbers of residues of the holo and apo PDB data differ because they have missing residues: The number of residues is 258 for 1ftmA but 257 for 1ftoA, 426 for 2revAB but 488 for 1revAB, and 480 for 4revAB but 488 for 1revAB. Although all of the constituent atoms in the PDB data were considered in the individual NMA calculations, only the common atoms in each pair are considered in the comparative analyses of the least-squares-fitting and RMSD calculations.

[b] The number of residues in an active site. The residues indicated by PDBsum [31] to have contact with a ligand or DNA are considered to make up the active-site.

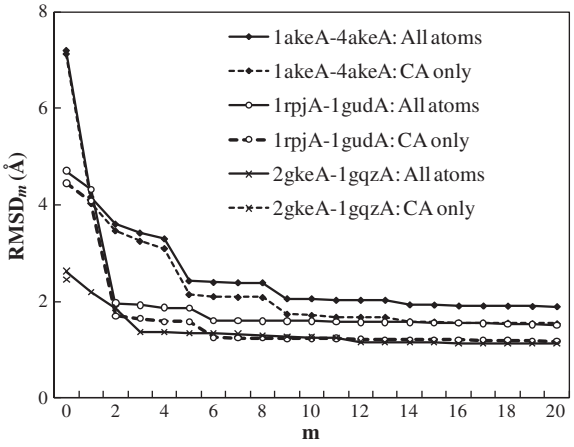

Fig. 1. Dependence of RMSD between the LSF and holo-form conformations on the number of normal modes considered. RMSD$_m$(all) and RMSD$_m$(C^α) ($m=1, 2, \ldots, 20$) are shown for three examples, i.e., 1akeA–4akeA, 1rpjA–1gudA, and 2gkeA–1gpzA. $m=0$ means the RMSD$_{PDB}$ between holo-form and apo-form PDB conformations.

and 2gkeA–1gqzA. The changes in RMSD$_m$ are discussed in association with Fig. 2 below.

In this study, we consider the lowest-frequency m normal modes for $m=1, 2, \ldots, 20$. In order to reveal which normal mode contributes more dominantly significantly to deformation, we examined w_i^2/W for the case of $m=20$, where $W=\sum_{i=1}^{20} w_i^2$ is a normalizing factor. In Table 2, the largest four w_i^2/W values of the 20 normal modes and their sum, S_4, are shown, where $S_k=\sum w_i^2/W$ is defined as a sum over k modes with the largest w_i^2/W. Note that $a_{\alpha,i}$ does not satisfy the orthonormal condition, i.e., $\sum_{\alpha\in H} a_{\alpha,i}a_{\alpha,j}\equiv a_i a_j \neq \delta_{ij}$, where $a_i=\{a_{\alpha,i}\}$. As a result, $\left(\sum_i w_i a_i\right)^2 \neq \sum_i w_i^2$. Although w_i^2/W does not exactly mean a contribution of the ith normal mode in the sense used in multivariable statistics, we use this property to approximately assess the contribution of the ith normal mode for convenience.

Out of the 22 holo-apo form pairs, five pairs, i.e., 1usiA–1usgA, 1sw2A–1sw5A, 1y3nA–1y3qA, 1rf4A–1rf5A, and 1tj5A–1s2oA, have an $S_1>0.8$, and six pairs, i.e., 1akeA–4akeA, 1rpjA–1gudA, 1o03A–1z0lA, 1vhlA–1viyC, 1ftmA–1ftoA, and 1jg6A–1jejA, have an $S_1<0.8$, but an $S_2>0.8$. This indicates that the conformational changes of the proteins in the former group are well reproduced by only one dominant normal mode while those in the latter group are reproduced by a combination of two dominant normal modes. However, even in the latter group, the second largest contribution is significantly smaller than the largest one. It should be also noted that the lowest-frequency three normal modes occupy 80% of the top two largest w_i^2/W values in the individual pairs (see Table 2).

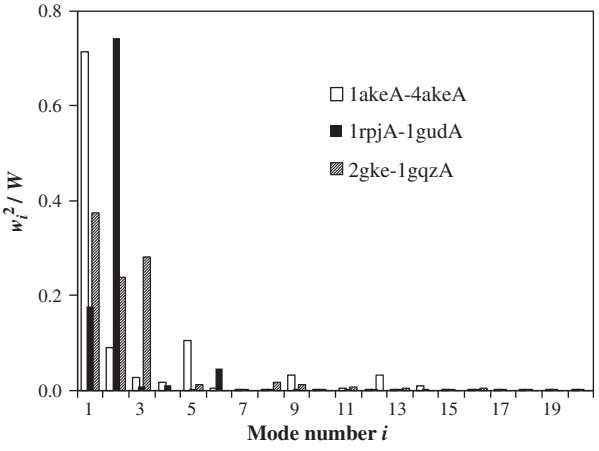

Fig. 2. Contribution of the 20 lowest-frequency normal modes. The normalized weighting factor, w_i^2/W ($i=1, 2, \ldots, 20$), for $m=20$ is shown for the same three examples, 1akeA–4akeA, 1rpjA–1gudA, and 2gkeA–1gpzA, as in Fig. 1.

Table 2
Top four modes with the largest w_i/W values.

Holo	Apo	Mode number i and w_i^2/W				$S_4=\sum w_i^2/W$
1usiA	1usgA	1	5	4	3	
		0.942	0.036	0.006	0.006	0.989
1akeA	4akeA	1	5	2	9	
		0.714	0.103	0.090	0.030	0.938
1q0qA	1k5hA	1	8	6	16	
		0.278	0.216	0.124	0.065	0.683
1sw2A	1sw5A	1	5	6	2	
		0.920	0.031	0.024	0.017	0.994
1y3nA	1y3qA	1	2	6	8	
		0.893	0.056	0.011	0.009	0.969
1rpjA	1gudA	2	1	6	4	
		0.743	0.177	0.046	0.010	0.977
1rf4A	1rf5A	2	8	4	6	
		0.938	0.016	0.010	0.008	0.972
1o03A	1zolA	3	2	1	10	
		0.648	0.153	0.133	0.013	0.946
1tj5A	1s2oA	1	2	3	8	
		0.921	0.036	0.011	0.008	0.975
2gkeA	1gqzA	1	3	2	12	
		0.376	0.281	0.239	0.033	0.930
1vhlA	1viyC	1	3	2	10	
		0.488	0.381	0.100	0.007	0.977
1q95A	1za1A	2	1	4	6	
		0.266	0.167	0.150	0.098	0.681
1cg0A	1hooB	1	2	8	10	
		0.330	0.246	0.127	0.060	0.763
1ftmA	1ftoA	1	2	5	9	
		0.511	0.389	0.034	0.016	0.950
1eejB	1tjdA	2	1	7	3	
		0.300	0.153	0.146	0.143	0.743
1jg6A	1jejA	2	4	3	7	
		0.763	0.117	0.046	0.018	0.944
1rpfA	1aqpA	2	1	12	11	
		0.403	0.110	0.103	0.098	0.714
1cimA	1ca2A	2	1	8	10	
		0.258	0.167	0.152	0.068	0.645
1kpeB	1kpaB	10	8	3	9	
		0.292	0.136	0.094	0.093	0.616
1hiiAB	1hsiAB	3	8	1	5	
		0.383	0.241	0.198	0.096	0.919
2rveAB	1rveAB	2	1	5	3	
		0.396	0.157	0.143	0.113	0.809
4rveAB	1rveAB	1	2	4	7	
		0.449	0.189	0.062	0.062	0.762

On the other hand, $S_4<0.7$ for four holo–apo pairs, i.e., 1q0qA–1k5hA, 1q95A–1za1A, 1cimA–1ca2A, and 1kpeB–1kpaB. In these pairs, there is no mode that dominantly reproduces the conformational change. Simplistically, this means that their conformational changes are reproduced by an ensemble of several normal modes rather than one dominant one. In fact, the significantly small RMSD$_{10}$ and RMSD$_{20}$ values for 1q95A–1za1A indicate successful least-squares fitting of the displacement vectors obtained by NMA to the conformational change vectors. The three other pairs, 1cg0A–1hooB, 1eejA–1tjdA, and 4rveAB–1rveAB, have an S_4 value within the range of 0.7–0.8. For these pairs, it is reasonable to consider that their conformational changes are reproduced by an ensemble of several normal modes rather than one dominant one. However, since the RMSD$_{10}$ and RMSD$_{20}$ values for the 1q0qA–1k5hA pair are slightly larger than those of the other pairs, the linear combination of the lowest-frequency normal modes is insufficient to explain the conformational change from the apo form to the holo form. This holo-apo pair is discussed in more detail below. The holo-apo pairs, 1cimA–1ca2A and 1kpeB–1kpaB, are considered to be different from the above cases, because their RMSD$_{PDB}$(all) values are quite small at 0.61

and 0.53 Å, respectively. It may be reasonable to assume from the results for these pairs that the ENM-based NMA used in this study is not sufficiently accurate to represent such small conformational changes. We will comment again on this point below.

Fig. 2 shows the w_i^2/W values for $i=1, 2,\ldots, 20$ for the same three examples as those given in Fig. 1. The relationship between $RMSD_m$ in Fig. 1 and w_i^2/W in Fig. 2 is clear. For 1akeA–4akeA, the w_i^2/W values for $i=1$, 2, and 5 are relatively large; in response, $RMSD_1$ and $RMSD_5$ lower considerably, and the reduction of $RMSD_2$ is small but significant. For 1rpjA–1gudA, the w_i^2/W value for $i=2$ is dominant and the reduction of $RMSD_2$ is apparent. In contrast, although the w_i^2/W values for $i=1$ and 6 are relatively large, the reductions of $RMSD_1$ and $RMSD_6$ are small. This shows that a large weighting factor does not always cause a large reduction in the RMSD. For 2gke–1gqzA, the w_i^2/W values for $i=1$, 2, and 3 are relatively large and comparable in size. In response, the reductions of $RMSD_1$, $RMSD_2$ and $RMSD_3$ are also comparable, and larger than other $RMSD_m$.

Out of the 22 holo–apo pairs, 19 pairs have $RMSD_{PDB}$(all) of greater than 2.0 Å, and have $RMSD_{10}$ and $RMSD_{20}$ values that are significantly smaller than their $RMSD_{PDB}$ value (Table 1). These facts indicate that the conformational fluctuations represented by the normal mode vibrations are intrinsically involved in the conformational changes from an apo form to a holo form in each protein. In contrast, the $RMSD_{10}$ and $RMSD_{20}$ values for the remaining three holo–apo pairs, 1rpfA–1aqpA, 1cimA–1ca2A, and 1kpeB–1kpaB, are similar to their $RMSD_{PDB}$ values, which are quite small, at 0.91, 0.61, and 0.53 Å, respectively, compared with those of the other pairs. As mentioned above, the ENM-based NMA used in this study does not have enough accuracy to reproduce smaller conformational changes. There are two possible reasons for the low-accuracy reproductions: (*i*) the coarse-grained potential function used in the ENM-based NMA (Eq. (1)) and (*ii*) the restriction of the normal modes to the 20 lowest-frequency ones. As far as the latter point is concerned, although there is a possibility that the inclusion of one specific normal mode out of the enormous number of higher-frequency ones could drastically improve the reproduction, it does not make sense to search for it in this generic study.

In summary, the above results indicate that the conformational differences between a holo form and an apo form can be well reproduced in most cases by a linear combination of displacement vectors calculated using ENM-based NMA for a full-atom system. In particular, the lowest-frequency three normal modes are sufficient to represent the conformational change as long as the conformational change is large enough.

3.2. Examples

Some examples are given in this section to illustrate in detail how the apo-form conformation is deformed by the least-squares fitting (see Eq. (4)). The first example is the L-leucine binding protein (LBP) (1usiA–1usgA) [8]. The $RMSD_{PDB}$(all) and $RMSD_{PDB}(C^{\alpha})$ between the holo-form and apo-form conformations are 7.30 and 7.14 Å, respectively (Table 1). This difference appears large in their superimposed conformations shown in Fig. 3a. LBP comprises two domains [8]. Whereas the individual structures of the two domains in the holo and apo forms are very similar to each other, one domain is rotated with respect to the other when the overall structures of the two forms are compared. Although the $RMSD_{10}$ and $RMSD_{20}$ values are slightly larger than those of other protein pairs, the LSF and holo-form conformations are reasonably close to each other, as seen in Fig. 3b. According to Table 2, the first lowest-frequency normal mode can essentially determine the deformation from the apo to the holo form. The movement in this normal mode is just a rotation of one domain with respect to the other (data are not shown here).

Some residues in the LSF conformation deviate significantly from the holo-form one; they are indicated by the cyan coloring in Fig. 3b. Most of these residues are in the surface loops, while some are in the hairpin β-sheet 311–318 located on the protein surface. The loops are flexible and the β-sheet can easily shift. Consequently, they deform or move beyond the changes expressed by the normal modes. On the other hand, the residues around the ligand-binding site are well fitted to each other, as seen in Fig. 3b (see also the discussion associated with Table 1 in the Discussion section).

The second example is 1-deoxy-d-xylulose-5-phosphate reductoisomerase (DXR) (1q0qA-1k5hA) [27,28]. This protein pair gives the worst results of the study (Fig. 4): the $RMSD_{10}$(all), $RMSD_{10}(C^{\alpha})$, $RMSD_{20}$(all), and $RMSD_{20}(C^{\alpha})$ values are 4.19, 3.69, 3.56, and 2.94 Å,

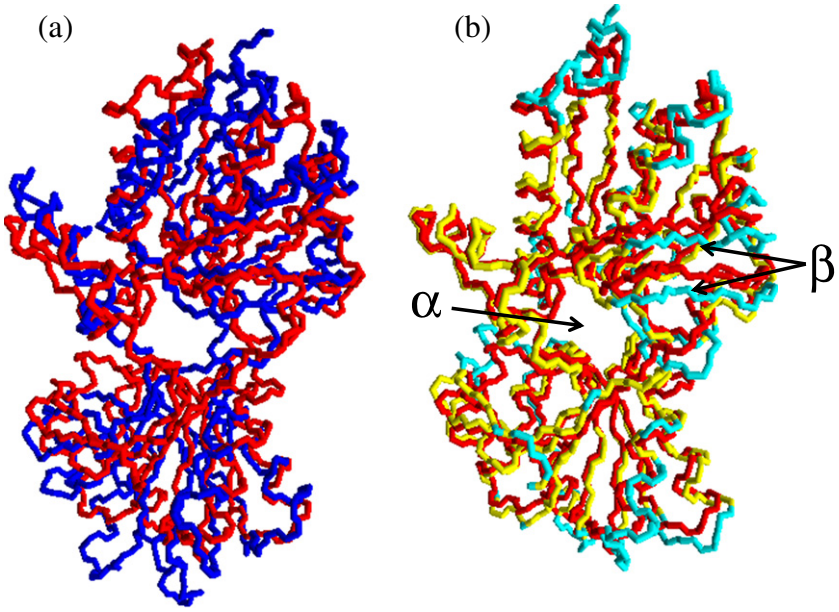

Fig. 3. L-leucine binding protein (1usiA–1usgA). The main-chain conformations of the holo-form, apo-form, and LSF conformations for $m=10$ are superimposed and colored red, blue, and yellow, respectively. (a) Apo-form and holo-form conformations. (b) Holo-form and LSF conformations. The residues in the LSF conformation that deviate by more than 2.0 Å from the holo-form conformation are colored cyan. Arrows α and β indicate a ligand-binding site and a hairpin β-sheet (311–318), respectively.

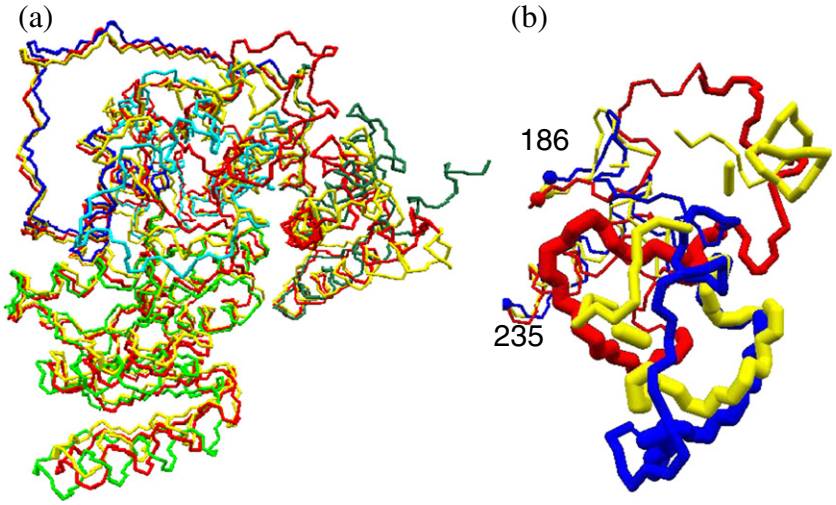

Fig. 4. 1-Deoxy-d-xylulose-5-phosphate reductoisomerase (1q0qA–1k5hA). (a) The main-chain conformations of the holo-form, apo-form, and LSF conformations for $m = 10$ are superimposed. The holo-form and LSF conformations are colored red and yellow, respectively. The residues 1–150, 160–270, and 312–398 of the apo-form conformation are colored green, cyan, and greenblue, respectively, to distinguish the three domains; the remaining residues are colored blue (see text). (b) An expanded view of the main-chain conformations of residues 186–235. The 195–205 and 206–216 regions are represented by thick and thicker tubes, respectively, relative to the other regions to identify the corresponding relationship between the three chains. Some regions of the yellow tubes are segmented because the regions in which molecular geometries (bond lengths and bond angles) deviate significantly from the standard values are not displayed on a molecular viewer.

respectively. The structure of DXR comprises three domains [27]: (*i*) a larger N-terminal NADPH binding domain (1–150), (*ii*) a linker domain (160–270), and (*iii*) a smaller C-terminal α-helical domain (312–398). As seen in Fig. 4a, the overall fitting of the conformations is not too bad; however, the conformations of the linker domain differ considerably (Fig. 4b). Residues 206–216 in the linker domain are known as a flexible lid that is closed upon ligand binding to shield the active site from the solvent. Evidently, this deformation is too large for NMA to be applied.

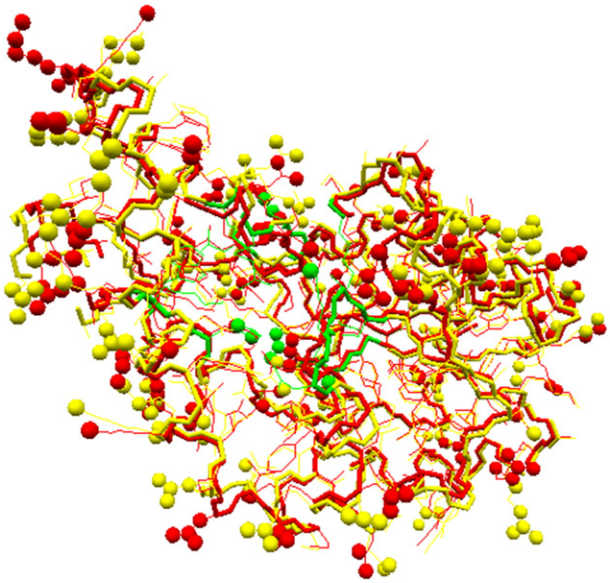

Fig. 5. β-Phosphoglucomutase (1o03A–1zolA). The conformations of the holo-form and LSF conformations for $m = 10$ are superimposed. The main and side chains are shown by tubes and thin lines, respectively. The yellow balls are side-chain atoms in the LSF conformation that deviate by more than 2.0 Å from the holo-form conformation; the same atoms in the holo-form conformation are represented by red balls. There are largely deviated main-chain atoms, but they are not indicated here in order to focus the attention on the side-chain atoms. The ligand-binding residues defined by PDBsum (residues 8–10, 20, 24, 44, 46, 47, 49, 52, 76, 114–118, and 145) [31] in the LSF conformation are colored green. See the Fig. 4 caption and the text for an explanation of why the yellow tube is segmented.

In Fig. 4, it is observed that the LSF conformation (yellow chain) is segmented in spite of the fact that the chain is actually continuous. This phenomenon occurs when a molecular viewer cannot properly determine the chain connectivity in the region where the molecular geometry (bond lengths and bond angles) significantly deviates from the standard values. This is one of the defects of the present method. It arises because the molecular geometry of the conformation is not checked when optimizing the linear combination of displacement vectors obtained from NMA. Therefore, it is necessary to improve the methodology for practical use.

In the third example, we examined side-chain atoms that were accounted for in the computation by ENM-based NMA for a full-atom system. Fig. 5 shows the results for β-phosphoglucomutase (1o03A–1zolA) [29,30], which has the largest difference between the $RMSD_{10}$ (all) and $RMSD_{10}(C^{\alpha})$ values, as shown in Table 1. This large difference suggests that the side-chain conformations in the LSF conformation are quite different from those in the holo-form one. In Fig. 5, the largely deviated side-chain atoms are represented by colored balls and are mainly located on the surface. In the NMA used in this study, the side chain can move rather freely if it is on the surface and has few atoms to interact with; therefore, such a side chain is considered to be intrinsically mobile. The largely deviated side-chain atoms in the binding site defined by PDBsum [31] are represented by green balls in Fig. 5. Only a few side-chain atoms in the binding site deviate significantly (see also the discussion associated with Table 1 in the Discussion section). The small deviations of the atoms at the binding site are probably because they are in the interior of the protein.

The final example is EcoRV endonuclease (2rveAB–1rveAB and 4rveAB–1rveAB) [32]. It is a homo dimer comprising subunits A and B and it makes a complex with DNA. PDB structures 1rve, 2rve, and 4rve refer to a free form, a complex with non-cognate DNA, and a complex with cognate DNA, respectively. The conformational differences between them are found in both the tertiary and quaternary structures. In Fig. 6, the main-chain conformations of the holo-form, apo-form, and LSF conformations are shown superimposed. According to Table 1, the $RMSD_{10}$ and $RMSD_{20}$ values for 2rveAB–1rveAB are much better than those for 4rveAB–1rveAB, even though the

$RMSD_{PDB}$ value for 2rveAB–1rveAB is larger than that of 4rveAB–1rveAB. One of the reasons is that there are many missing residues in the 2rve PDB data; since the missing residues are in mobile regions, the RMSD value is expected to be lower if they are ignored. Another possibility is a difference in the interactions between DNA and the proteins of 2rve and 4rve. It has been shown that the total surface area buried upon complex formation is larger in the case of cognate DNA binding (4rve) than non-cognate DNA binding (2rve) [32]. This indicates that there are many more interactions between the protein and the DNA in 4rve than in 2rve. Consequently, it is reasonable to consider that the conformation of 2rve can fluctuate more freely than that of 4rve, and consequently would be more affected by normal mode fluctuations.

In Fig. 7, the weighting factors, w_i^2/W ($i = 1, 2, \ldots, 20$), are shown for 2rveAB–1rveAB and 4rveAB–1rveAB. The figure shows that the LSF conformations for 2rve and 4rve have different linear combinations of displacement vectors than those obtained for the apo-form conformation of 1rve. While the second lowest-frequency normal mode is dominant in 2rveAB–1rveAB, the first one is dominant in 4rveAB–1rveAB. However, the difference between 2rve and 4rve in the distributions of w_i^2/W is relatively small because the RMSD between the PDB structures of 2rve and 4rve is not very large ($RMSD_{PDB}(all) = 2.28$ Å).

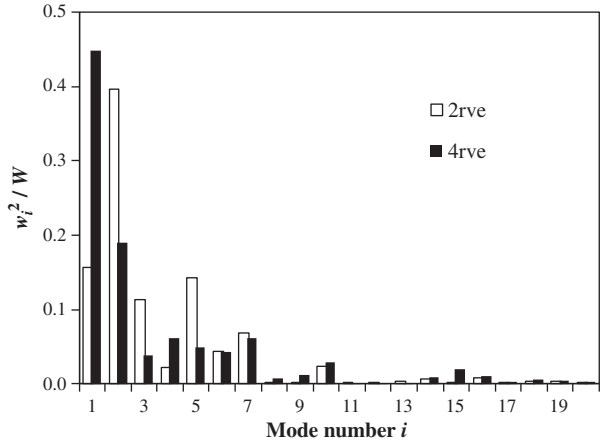

Fig. 7. Contribution of the 20 lowest-frequency normal modes for EcoRV endonuclease (2rveAB–1rveAB and 4rveAB–1rveAB). The normalized weighting factor, w_i^2/W ($i = 1, 2, \ldots, 20$), for $m = 20$ is shown for the two holo–apo pairs.

4. Discussion

In this study, we demonstrated the potential for the use of ENM-based NMA for a full-atom system in a dihedral-angle space to reproduce the conformational changes in proteins upon ligand binding. From a different perspective, these results suggest that the conformational change from an apo to a holo form is within the intrinsic conformational fluctuations of the apo form calculated by NMA in most cases. However, since NMA can only identify intrinsic conformational fluctuations around the minimum-energy conformation (the apo form itself in the ENM-based NMA performed here), the identified variations are not guaranteed to be equivalent to the intrinsic conformational fluctuations of the holo form. Accordingly, it is unclear if the conformational change from a holo to an apo form can be reproduced by a linear combination of displacement vectors obtained from the NMA for the holo form. Fig. 8a shows the $RMSD_{10}$ (all) values obtained by least-squares fitting of a linear combination of the displacement vectors calculated for the holo form to the conformational change vectors from the holo form to the apo form (referred to as $RMSD_{10}(Holo \rightarrow Apo)$). These values are plotted against the $RMSD_{10}(Apo \rightarrow Holo)$ values described above for comparison. In most cases, $RMSD_{10}(Holo \rightarrow Apo)$ was nearly equal to or greater than $RMSD_{10}(Apo \rightarrow Holo)$. The differences were not very large, except for the 1akeA–4akeA protein pair.

Maragakis and Karplus [33] studied the conformational change pathway between the 1ake and 4ake conformations of the 1akeA–

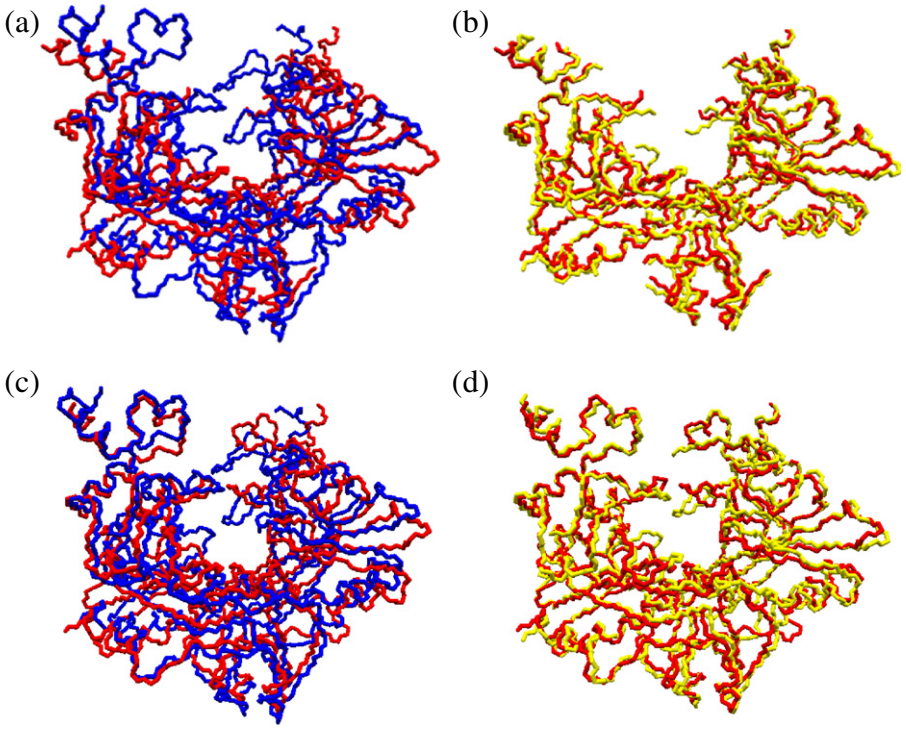

Fig. 6. EcoRV endonuclease (2rveAB–1rveAB and 4rveAB–1rveAB). The main-chain conformations of the holo-form, apo-form, and LSF conformations are superimposed and colored red, blue, and yellow, respectively. (a) and (b) 2rveAB in red and 1rveAB in blue. (c) and (d) 4rveAB in red and 1rveAB in blue. Since the PDB data for these proteins have missing atoms, the chains are not continuous.

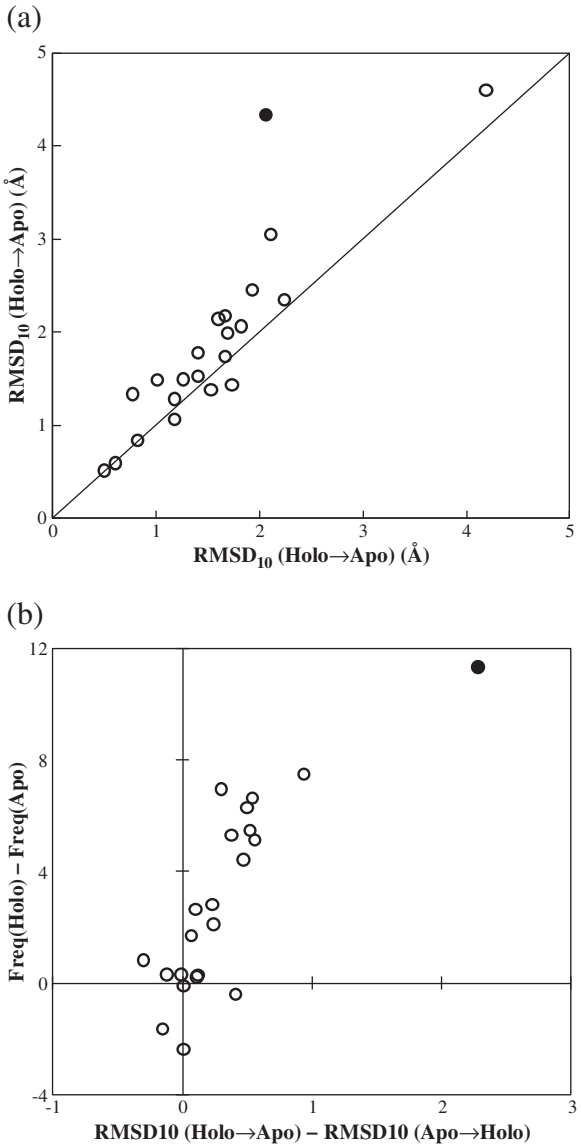

Fig. 8. (a) Comparison between $RMSD_{10}$(Apo → Holo) and $RMSD_{10}$(Holo → Apo). $RMSD_{10}$(Apo → Holo) is the root-mean-square difference between the holo-form and the LSF conformations obtained from NMA of an apo-form conformation. $RMSD_{10}$ (Holo → Apo) is the root-mean-square difference between the apo-form and the LSF conformations obtained from NMA of a holo-form conformation. (b) Correlation between $RMSD_{10}$(Holo → Apo) − $RMSD_{10}$(Apo → Holo) and Freq(Holo)–Freq(Apo). Freq(H olo) and Freq(Apo) denote the lowest frequencies of normal modes of a holo-form and an apo-form conformation, respectively. The closed symbols in both panels indicate the 1akeA–4akeA pair.

4akeA holo–apo pair using the plastic-network model and found a transition state structure rather close to that of 1ake (holo form). Furthermore, they found that only a limited number of normal modes of 4ake (apo form) contribute to most of the conformational changes that occur from the apo form to the transition point, so the energy change is small over a large deformation. Their results explain our finding that the $RMSD_{10}$(Apo → Holo) for 4akeA → 1akeA is much smaller than the $RMSD_{10}$(Holo → Apo) for 1akeA → 4akeA.

Tama and Sanejouand [15] previously discussed the general differences between $RMSD_{10}$(Holo → Apo) and $RMSD_{10}$(Apo → Holo). They showed that the conformational change from an open form to a closed form (corresponding to a change from an apo form to a holo form in this study) is almost always better described than that from the closed form to the open form. The results of the present study confirm these conclusions. According to Tama and Sanejouand, this occurs because the properties captured by NMA are considered to reflect the shape of the protein and the domains, which are structures with characteristic shapes in the protein structure, are more clearly defined in the open form than in the closed form.

The difference between $RMSD_{10}$(Apo → Holo) and $RMSD_{10}$ (Holo → Apo) can be attributed to the difference between conformational energy surfaces around the apo-form and holo-form conformations. Therefore, we compared the lowest frequencies of the normal modes of the two forms for the 22 protein pairs. The lowest frequencies reflect the largest curvatures of the energy surfaces around the two forms, with smaller frequencies indicating a larger curvature and therefore a larger possible conformational change for a given energy change. The difference in the lowest frequency of a holo form from that of an apo form, Freq(Holo)–Freq(Apo), is plotted against the difference, $RMSD_{10}$(Holo → Apo)–$RMSD_{10}$(Apo → Holo), in Fig. 8b. These two properties were well correlated, with a larger $RMSD_{10}$(Holo → Apo)–$RMSD_{10}$(Apo → Holo) value being associated with Freq(Holo) being larger than Freq(Apo). These findings suggest that the changeability of an apo-form conformation, i.e., how easily it can undergo various changes, is greater than that of a holo-form; however, it is necessary to examine more low-frequency normal modes to elucidate additional detailed features of the conformational energy surfaces. This is another expression of the conformational characterization of an open form and a closed form by Tama and Sanejouand [15] discussed above. Specifically, the considerably large value of Freq(Holo)–Freq(Apo) for the 1akeA–4akeA pair is consistent with the results of the study conducted by Maragakis and Karplus [33] described above. Our findings further suggest that features similar to 1akeA–4akeA may be observed in other holo–apo protein pairs if the analysis of Maragakis and Karplus is applied.

The weighting factors for the linear combination of the displacement vectors optimized by least-squares fitting characterize the conformational change. In the above examples, only one normal mode can sufficiently reproduce the conformational change from apo to holo forms for some proteins, and the combination of two or more normal modes is necessary for others. Conversely, some conformational changes cannot be reproduced by the linear combination of the displacement vectors obtained from ENM-based NMA. It may be interesting to characterize conformational differences among multiple conformations generated by methods other than NMA such as molecular dynamics simulation and NMR. Although it is not necessarily possible to detect improbable conformational changes using only ENM-based NMA since there are cases in which the conformational change cannot be reproduced by linear combination of the displacement vectors, this type of information may help exclude improper conformations in collaboration with other methods such as conformational energy calculations and some knowledge-based examinations. Even if the conformational change cannot be reproduced via linear combination of the displacement vectors for some proteins, this information is valuable for further consideration of alternative mechanisms for the conformational change rather than intrinsic fluctuations of the protein.

As in previous studies, most conformational changes with highly correlated NMA results were large-scale domain motions in this study. Additionally, dimer proteins in the holo and apo forms were examined in the present study. With respect to the RMSD, the results for the dimers were similar to those for the monomers. Essentially, the mutual motions between subunits are comparable to the large-scale domain motions. However, the conformational change in the dimers (oligomer, in general) is a more complicated phenomenon. Specifically, these changes comprise the conformational changes in the individual subunits as well as their quaternary movements. The displacement vectors obtained from NMA can be decomposed into two types of motions; an external one (rigid-body motions of individual subunits) and an internal one (deformation motions of individual subunits) [34]. The external motion describes mutual

movements between subunits. Although we did not address the problem from this perspective in the present study, it would be interesting to examine the conformational changes in oligomers upon ligand binding in more detail.

When using a linear combination of the displacement vectors obtained from several low-frequency normal modes, we assumed a flexible-docking problem. In cases in which an intact apo-form conformation given in PDB is used as a first approximation and the results are not satisfactory, it is necessary to generate candidate conformations for a holo-form conformation from the apo-form. Since the generated conformations must be individually assessed with respect to the complementary conformational features between the protein and ligand or to the ligand–protein pairwise interaction energies, the random generation of conformations is not feasible because of the computation time required. Therefore, it is necessary to develop a method to automatically generate reasonably small, but representative, sets of conformational changes. The results of the present study suggest that NMA could be used for this purpose. For example, candidate conformations can be generated by assigning random numbers to a set of weights, w_i, in Eq. (4). Although this significantly limits the conformational space that is searched, it may be necessary to further restrict the number of conformations generated by the NMA. Ikeguchi et al. [35] showed that the conformational change upon ligand binding was explained well by linear response theory. Their results suggest that by properly incorporating protein–ligand interactions, it is possible to select and merge the low-frequency normal modes. It is also necessary to reveal why a few normal modes can dominantly reproduce a conformational change upon ligand binding, even though there are various low-frequency normal modes associated with the large-scale domain motions. In other words, it is necessary to determine the characteristic motion upon ligand binding in various normal modes. Such studies would not only contribute to further development of the flexible-docking problem, but also to understanding the structure–function relationship of proteins.

For a flexible-docking problem, we are concerned with the side-chain conformations around an active site because ENM-based NMA was conducted for a full-atom system. The right-hand columns in Table 1 show the RMSD values between sets of atoms involved in the active site of an apo-form and a holo-form conformation and those between LSF for $m = 10$ and holo-form conformations, respectively. For atoms in the active site, we considered all atoms in the residues that were found to have contact with a ligand or DNA in PDBsum [31]. It is natural that all RMSD values between the LSF and holo-form conformations, except for 1q0qA–1k5hA, are lower than those between the apo-form and holo-form conformations, and we discussed the exceptionally large RMSD values of 1q0qA–1k5hA in Section 3.2. While most RMSD values are comparable to those of the whole conformations, $RMSD_{10}$(all), some, such as those of 1rf4A–1rf5A and 1cg0A–1hooB, are remarkably large. As far as the 1rf4A–1rf5A pair is concerned, while residues 334–342 involved in the active site are in a disordered conformation in 1rf5A, they become ordered (residues 334–338 form an α-helix) upon ligand binding [36]. This area is considered to act as a gate for ligand binding. For the 1cg0A–1hooB pair, the conformations of residues 126–129 and 298–301 involved in the active site in LSF and a holo form deviated greatly from each other. Segments of residues 121–130 and 298–304 were observed for poorly ordered loops in an apo-form (1hooB) [37] and were thus intrinsically mobile. In these proteins, while whole conformational changes were found to be within the intrinsic fluctuations represented by normal modes, the local conformational changes around the active sites were beyond such fluctuations.

Although it is possible that the RMSD values can be improved by considering higher-frequency normal modes, it is practical to refine the generated conformations using other methods such as molecular dynamics (MD) and rotamer sampling. In fact, MD has been conducted from a modeling perspective and the role of individual residues and water molecules has been discussed [38]. Water molecules were not considered in this study because they require too many more degrees of freedom for NMA; however, there is no algorithmic difficulty in introducing water molecules into NMA. Nevertheless, even if water molecules are taken into account, it is not possible to predict the active-site conformation as accurately as with MD using only ENM-based NMA. Rather, ENM-based NMA may be used to generate reasonable sets of conformational changes for further refinements by other methods such as MD and rotamer sampling, thereby taking advantage of its rapid computation.

5. Conclusions

The conformational changes in 22 paired proteins, including three dimer pairs, upon ligand binding were examined by ENM-based NMA. The results revealed that conformational change vectors of atoms from the apo to the holo form of a protein could be reproduced well by a linear combination of the displacement vectors of atoms in the apo form calculated for low-frequency normal modes for most of these proteins. The weighting factors for the linear combination of the displacement vectors optimized by least-squares fitting can characterize the conformational change upon ligand binding. For example, while only one normal mode can sufficiently reproduce the conformational change from the apo to the holo form for some proteins, the combination of two or more normal modes is necessary for others. Conversely, some conformational changes could not be reproduced by linear combination of the displacement vectors obtained from ENM-based NMA. In such cases, it is necessary to consider an alternative mechanism for the conformational change rather than intrinsic fluctuations of the protein, such as protein–ligand interaction, large deformability of the whole and local conformations, and large mobility of a conformational unit.

In this study, ENM-based NMA was employed for full-atom systems using dihedral angles as independent variables. Since the solvent molecules were not considered during computation, some side-chain conformations on the surface of the LSF conformation deviated greatly from the holo-form conformation. However, these side-chains are intrinsically flexible; therefore, this may not be a serious problem. Conversely, for the conformation around the active site, the RMSD values were very small for some proteins, but rather large for others. The large RMSD values can be attributed to the remarkable flexibility of the segment involved in ligand binding. These results were also used to characterize the conformational changes upon ligand binding.

From a modeling perspective, the conformations can be generated by assigning random numbers to weight factors during linear combination of the displacement vectors obtained using ENM-based NMA as a candidate for possible conformational changes around an apo-form conformation. The results of the present study suggest it is highly likely that an actual conformational change is involved in the conformations generated in this manner. The generated conformations for assessment of ligand binding are greatly limited when compared to random generation of conformations, but can be still in huge numbers. However, the assessment of conformational changes upon ligand binding might not be a formidable challenge because of rapid advances in computer technology.

References

[1] E. Fischer, Einfluss der configuration auf die wirkung der enzyme, Ber. Dtsch. Chem. Ges. 27 (1894) 2985–2993.
[2] D.E. Koshland Jr., Application of a theory of enzyme specificity to protein synthesis, Proc. Natl. Acad. Sci. U.S.A. 44 (1958) 98–104.
[3] R. Najmanovich, J. Kuttner, V. Sobolev, M. Edelman, Side-chain flexibility in proteins upon ligand binding, Proteins 39 (2000) 261–268.
[4] M.I. Zavodszky, L.A. Kuhn, Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis, Protein Sci. 14 (2005) 1104–1114.

[5] S. Karthikeyan, Q. Zhou, A.L. Osterman, H. Zhang, Ligand-binding induced conformational changes in riboflavin kinase: structural basis for the ordered mechanism, Biochemistry 42 (2003) 12532–12538.
[6] M.G. Malkowski, P.D. Martin, J.C. Guzik, B.F. Edwards, 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 (1997) 1438–1448.
[7] L. Kraft, G.A. Sprenger, Y. Lindqvist, Conformational changes during the catalytic cycle of gluconate kinase as revealed by X-ray crystallography, J. Mol. Biol. 318 (2002) 1057–1069.
[8] U. Magnusson, B. Salopek-Sondi, L.A. Luck, S.L. Mowbray, X-ray structures of the leucine-binding protein illustrate conformational changes and the basis of ligand specificity, J. Biol. Chem. 279 (2004) 8747–8752.
[9] G. Benison, P.A. Karplus, E. Barbar, The interplay of ligand binding and quaternary structure in the diverse interactions of dynein light chain LC8, J. Mol. Biol. 384 (2008) 954–966.
[10] J.-P. Xiong, T. Stehle, R. Zhang, A. Joachimiak, M. Frech, S.L. Goodman, M.A. Arnaout, Crystal structure of the extracellular segment of integrin αVβ3 in complex with an Arg-Gly-Asp ligand, Science 296 (2002) 151–155.
[11] M. Brylinski, J. Skolnick, What is the relationship between the global structures of apo and holo proteins? Proteins 70 (2008) 363–377.
[12] W.G. Krebs, V. Alexandrov, C.A. Wilson, N. Echols, H. Yu, M. Gerstein, Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic, Proteins 48 (2002) 682–695.
[13] V. Alexandrov, U. Lehnert, N. Echols, D. Milburn, D. Engelman, M. Gerstein, Normal modes for predicting protein motions: a comprehensive database assessment and associated web tool, Protein Sci. 14 (2005) 633–643.
[14] S.E. Dobbins, V.I. Lesk, J.E. Sternberg, Insights into protein flexibility: the relationship between normal modes and conformational change upon protein-protein docking, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 10390–10395.
[15] F. Tama, Y.-H. Sanejouand, Conformational change of proteins arising from normal mode calculations, Protein Eng. 14 (2001) 1–6.
[16] D. Tobi, I. Bahar, Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 18908–18913.
[17] M.L. Teodoro, L.E. Kavraki, Conformational flexibility models for the receptor in structure based drug design, Curr. Pharm. Des. 9 (2003) 1635–1648.
[18] M. Totrov, R. Abagyan, Flexible ligand docking to multiple receptor conformations: a practical alternative, Curr. Opin. Struct. Biol. 18 (2008) 178–184.
[19] C.N. Cavasotto, J. Kovacs, R.A. Abagyan, Representing receptor flexibility in ligand docking through relevant normal modes, J. Am. Chem. Soc. 127 (2005) 9632–9640.
[20] H. Wako, N. Gō, Algorithm for rapid calculation of hessian of conformational energy function of proteins by supercomputer, J. Comp. Chem. 8 (1987) 625–635.
[21] H. Wako, S. Endo, K. Nagayama, N. Gō, FEDER/2: program for static and dynamic conformational energy analysis of macro-molecules in dihedral angle space, Comp. Phys. Comm. 19 (1995) 233–251.
[22] K. Hinsen, Analysis of domain motions by approximate normal mode calculations, Proteins 33 (1998) 417–429.
[23] A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, I. Bahar, Anisotropy of fluctuation dynamics of proteins with an elastic network model, Biophys. J. 80 (2001) 505–515.
[24] I. Bahar, A.J. Rader, Coarse-grained normal mode analysis in structural biology, Curr. Opin. Struct. Biol. 15 (2005) 1–7.
[25] D. Seeliger, B.L. de Groot, Conformational transitions upon ligand binding: holo-structure prediction from apo conformations, PLOS Comp. Biol. 6 (2010) e1000634.
[26] H. Wako, M. Kato, S. Endo, ProMode: a database of normal mode analyses on protein molecules with a full-atom model, Bioinformatics 20 (2004) 2035–2043.
[27] A. Mac Sweeney, R. Lange, R.P.M. Fernandes, H. Schulz, G.E. Dale, A. Douangamath, P.J. Proteau, C. Oefner, The crystal structure of *E. coli* 1-deoxy-d-xylulose-5-phosphate reductoisomerase in a ternary complex with the antimalarial compound fosmidomycin and NADPH reveals a tight-binding closed enzyme conformation, J. Mol. Biol. 345 (2005) 115–127.
[28] K. Reuter, S. Sanderbrand, H. Jomaa, J. Wiesner, I. Steinbrecher, E. Beck, M. Hintz, G. Klebe, M.T. Stubbs, Crystal structure of 1-deoxy-d-xylulose-5-phosphate reductoisomerase, a crucial enzyme in the non-mevalonate pathway of isoprenoid biosynthesis, J. Biol. Chem. 277 (2002) 5378–5384.
[29] S.D. Lahiri, G. Zhang, D. Dunaway-Mariano, K.N. Allen, The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction, Science 299 (2003) 2067–2071.
[30] G. Zhang, J. Dai, L. Wang, D. Dunaway-Mariano, L.W. Tremblay, K.N. Allen, Catalytic cycling in β-phosphoglucomutase: a kinetic and structural analysis, Biochemistry 44 (2005) 9404–9416.
[31] R.A. Laskowski, PDBsum new things, Nucleic Acids Res. 37 (2009) D355–D359.
[32] F.K. Winkler, D.W. Banner, C. Oefner, D. Tsernoglou, R.S. Brown, S.P. Heathman, R.K. Bryan, P.D. Martin, K. Petratos, K.S. Wilson, The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments, EMBO J. 12 (1993) 1781–1795.
[33] P. Maragakis, M. Karplus, Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase, J. Mol. Biol. 352 (2005) 807–822.
[34] H. Ishida, Y. Jochi, A. Kidera, Dynamic structure of subtilisin–eglin c complex studied by normal mode analysis, Proteins 32 (1998) 324–333.
[35] M. Ikeguchi, J. Ueno, M. Sato, A. Kidera, Protein structural change upon ligand binding: linear response theory, Phys. Rev. Lett. 94 (2005) 078102.
[36] H. Park, J.L. Hilsenbeck, H.J. Kim, W.A. Shuttleworth, Y.H. Park, J.N. Evans, C.H. Kang, Structural studies of Streptococcus pneumonia EPSP synthase in unliganded state, tetrahedral intermediate-bound state and S3P-GLP-bound state, Mol. Microbiol. 51 (2004) 963–971.
[37] B.W. Poland, Z. Hou, C. Bruns, H.J. Fromm, R.B. Honzatko, Refined crystal structures of guanine nucleotide complexes of adenylosuccinate synthetase from *Escherichia coli*, J. Biol. Chem. 271 (1996) 15407–15413.
[38] A.L. Milac, N.V. Buchete, T.A. Fritz, G. Hummer, L.A. Tabak, Substrate-induced conformational changes and dynamics of UDP-N-acetylgalactosamine: polypeptide N-acetylgalactosaminyltransferase-2, J. Mol. Biol. 373 (2007) 439–451.