On the relationship between the protein structure and protein dynamics

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · AUGUST 2008

Impact Factor: 2.63 · DOI: 10.1002/prot.21954 · Source: PubMed

8 AUTHORS, INCLUDING:

Chih-Hao Lu
China Medical University (ROC)
11 PUBLICATIONS **411** CITATIONS
SEE PROFILE

Shao-Wei Huang
Tzu Chi University
15 PUBLICATIONS **148** CITATIONS
SEE PROFILE

Tsun-Tsao Huang
National Chiao Tung University
9 PUBLICATIONS **170** CITATIONS
SEE PROFILE

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

On the relationship between the protein structure and protein dynamics

Chih-Hao Lu, Shao-Wei Huang, Yan-Long Lai, Chih-Peng Lin, Chien-Hua Shih, Cuen-Chao Huang, Wei-Lun Hsu, and Jenn-Kang Hwang*

Institute of Bioinformatics, National Chiao Tung University, HsinChu 30050, Taiwan, Republic of China

ABSTRACT

Recently, we have developed a method (Shih et al., Proteins: Structure, Function, and Bioinformatics 2007;68: 34–38) to compute correlation of fluctuations of proteins. This method, referred to as the protein fixed-point (PFP) model, is based on the positional vectors of atoms issuing from the fixed point, which is the point of the least fluctuations in proteins. One corollary from this model is that atoms lying on the same shell centered at the fixed point will have the same thermal fluctuations. In practice, this model provides a convenient way to compute the average dynamical properties of proteins directly from the geometrical shapes of proteins without the need of any mechanical models, and hence no trajectory integration or sophisticated matrix operations are needed. As a result, it is more efficient than molecular dynamics simulation or normal mode analysis. Though in the previous study the PFP model has been successfully applied to a number of proteins of various folds, it is not clear to what extent this model will be applied. In this article, we have carried out the comprehensive analysis of the PFP model for a dataset comprising 972 high-resolution X-ray structures with pairwise sequence identity $\leq$25%. We found that in most cases the PFP model works well. However, in case of proteins comprising multiple domains, each domain should be treated separately as an independent dynamical module with its own fixed point; and in case of the protein complex comprising a number of subunits, if functioning as a biological unit, the whole complex should be considered as one single dynamical module with one fixed point. Under such considerations, the resultant correlation coefficient between the computed and the X-ray structural B-factors for the data set is 0.59 and 75% (727/972) of proteins with a correlation coefficient $\geq$0.5. Our result shows that the fixed-point model is indeed quite general and will be a useful tool for high throughput analysis of dynamical properties of proteins.

Proteins 2008; 72:625–634.

Key words: protein dynamics; thermal fluctuations; molecular dynamics; normal mode analysis; B-factors.

INTRODUCTION

Protein dynamics is closely related to protein function. The ability to compute dynamical properties of proteins may to help shed new light on protein function. With the progress in experimental structural biology, the number of protein structures deposited in Protein Data Bank grows rapidly, and it becomes increasingly important to develop more efficient methods to extract dynamical properties from protein structures in a high throughput manner. Molecular dynamics method[1–4] is a powerful method to compute the dynamical properties of proteins. However, molecular dynamics simulation is computationally expensive. For example, to obtain the average thermal fluctuations (or the *B*-factors), one will need to integrate a long time trajectory in order to compute the root means square fluctuations of the residues. A recent study[5] of massive molecular dynamics simulations of 30 proteins using four different force fields in aqueous solution reportedly took computational time equivalent to around 50 years of CPU. The average protein dynamics can also be studied by normal mode analysis (NMA).[6–8] The NMA calculates the 2nd derivative matrix (also called the Hessian matrix) of the total potential function of the energy-optimized structure, and then calculates the normal mode eigenvectors and eigenvalues of the matrix. The Elastic Network Model (ENM) or Gaussian Network Model (GNM)[9–11] provides a coarse-grained version of NMA to compute average dynamical properties of proteins. In the ENM or GNM, all neighboring resides are connected to each other by a uniform harmonic potential function, and, just like NMA, a Hessian matrix is calculated based on that simple force field, and the modes of motion are calculated through the diagonalization of this matrix.

We recently reported a study[12] that the atoms lying on the same shell centered at the fixed point of the protein tend to

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: National Science Council; Grant number: NSC 95-3114-P-002-005-Y; Grant sponsor: The MoE ATU program.

Chih-Hao Lu and Shao-Wei Huang contributed equally to this work.

*Correspondence to: Jenn-Kang Hwang, Institute of Bioinformatics, National Chiao Tung University, HsinChu 30050, Taiwan, Republic of China. E-mail: jkhwang@faculty.nctu.edu.tw

Received 27 April 2007; Revised 27 October 2007; Accepted 19 November 2007
Published online 4 February 2008 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21954

have similar thermal fluctuations. The fixed point is the position in protein that has the least fluctuations, which was identified in the previous study with the centroid (or the center of mass) of the protein. We refer to this as the protein fixed-point (PFP) model. When computing only the average thermal fluctuations, that is, the *B*-factors, the PFP model requires only simple vector operations like vector addition and vector inner product. Unlike NMA, this method does not require matrix operations like matrix diagonalization and, therefore, can handle quite large proteins. For example, the PFP model can compute the thermal fluctuations of a protein as large as 50S ribosomal subunit[13] comprising more than 3700 residues on a common desktop computer, while none of the aforementioned methods can do it due to the hardware memory problem. On the other hand, the PFP method can also calculate modes of motion, as will be shown in later sections.

The PFP model has been successfully applied to a small sample set comprising 14 single-chain structures,[12] but it is still not clear whether the PFP model can be applied to proteins containing more than one domain or protein complexes comprising two or more chains. It is also not clear how one can define the fixed point for a multidomain protein, since there may be more than one fixed point for a multidomain protein. One conspicuous example is the calcium-binding calmodulin (PDB ID: 1CLL), which appears to have 2 fixed points. We will try to address these issues in this report.

METHODS

The protein fixed-point model

In the PFP model, the protein of N residues is characterized by the r profile $\mathbf{R}$

$$\mathbf{R} = (\mathbf{r}_1 - \mathbf{r}_0, \mathbf{r}_2 - \mathbf{r}_0, \ldots, \mathbf{r}_N - \mathbf{r}_0) \quad (1)$$

where $\mathbf{r}_i$ is the coordinate of Cα atom of the ith residue, and $\mathbf{r}_0$ is the fixed point. The fixed point $\mathbf{r}_0$ is identified with the centroid (or the center of mass) of the protein chain, that is, $\mathbf{r}_0 = \sum_i \mathbf{r}_i/N$. The matrix element $\boldsymbol{C}_{ij}$, that is, the correlation between fluctuations of atom i and j, of the normalized correlation matrix $\boldsymbol{C}$ is formally given by

$$C_{ij} = \frac{\left\langle \Delta\mathbf{r}_i \, \Delta\mathbf{r}_j \right\rangle}{\sqrt{\left\langle \Delta\mathbf{r}_i \, \Delta\mathbf{r}_i \right\rangle \left\langle \Delta\mathbf{r}_j \, \Delta\mathbf{r}_j \right\rangle}}, \quad (2)$$

where $\Delta\mathbf{r}_i$ and $\Delta\mathbf{r}_j$ are the fluctuations of the atom i and j, respectively, around their equilibrium positions.

In the PFP model, the correlation matrix element is approximated by $C_{ij} \sim (\mathbf{R}^T\mathbf{R})_{ij}$, where $\mathbf{R}^T$ is the transpose of $\mathbf{R}$. The PFP correlation matrix $\mathbf{M}$ is given as

$$\mathbf{M} = \mathbf{R}^T\mathbf{R} \quad (3)$$

The above equation computes not only the auto-correlation of atomic motions, but also the cross-correlation of atomic motions. Note that the PFP model provides a straightforward way to compute the correlation matrix directly from protein shapes without matrix diagonalization. In other words, the PFP model computes the complete correlation matrix without the use of the Hessian matrix, whereas other method like the NMA or the ENM needs to first diagonalize the Hessian matrix in order to obtain the correlation matrix.

The PFP model also provides a way to compute the normal modes of protein motion. Since the correlation matrix is the inverse of the Hessian matrix, we can easily compute all modes of protein motion by first diagonalizing $\mathbf{M}$,

$$\mathbf{M} = \mathbf{U}\Lambda\mathbf{U}^{-1} \quad (4)$$

where $\mathbf{U}$ is the matrix whose column vectors are eigenvectors of $\mathbf{M}$, and Λ is the diagonal matrix of eigenvalues. Using $\mathbf{M}^{-1} = \mathbf{U}\Lambda^{-1}\mathbf{U}^{-1}$, we obtain the normal mode frequencies from the diagonal elements of Λ^{-1} and the normal vectors from $\mathbf{U}$.

The *B*-factor is given by

$$B = (8\pi^2/3)\langle \Delta\mathbf{r}_i \, \Delta\mathbf{r}_i \rangle, \quad (5)$$

From this, one can see that the *B*-factors are in fact proportional to the diagonal terms of the correlation matrix. We can identify the *B*-factor profile with the following r^2 profile $\mathbf{R}_2$ within a scaling factor,

$$\mathbf{R}_2 = \big((\mathbf{r}_1 - \mathbf{r}_0)\,(\mathbf{r}_1 - \mathbf{r}_0), (\mathbf{r}_2 - \mathbf{r}_0)\,(\mathbf{r}_2 - \mathbf{r}_0), \ldots, (\mathbf{r}_N - \mathbf{r}_0)\,(\mathbf{r}_N - \mathbf{r}_0)\big) \quad (6)$$

Since $(\mathbf{r}_i - \mathbf{r}_0)\,(\mathbf{r}_i - \mathbf{r}_0) \geq 0$, the fixed point $\mathbf{r}_0$ actually defines the point of the smallest fluctuations in proteins—hence the origin of the name—the fixed point.

For illustration, the PFP model is schematically shown in Figure 1. We use a simple hypothetical 4-atom protein as an example. Each atom i is associated with a vector $\mathbf{v}_i$ issuing from the fixed point to the atom with a magnitude of the distance from the fixed point to the atom. The correlation of motion between any two atoms is computed by taking the inner product of the vectors associated with the atoms. The *B*-factor (or the autocorrelation of motion) of the atom is proportional to the square of the magnitude of its associated vector, that is, the inner product of the vector with itself. In this particular example, atom 1 has the smallest *B*-factor of all, whereas atom 2 has the largest *B*-factor. Atoms 3 and 4 have identical *B*-factors because both lie on the same

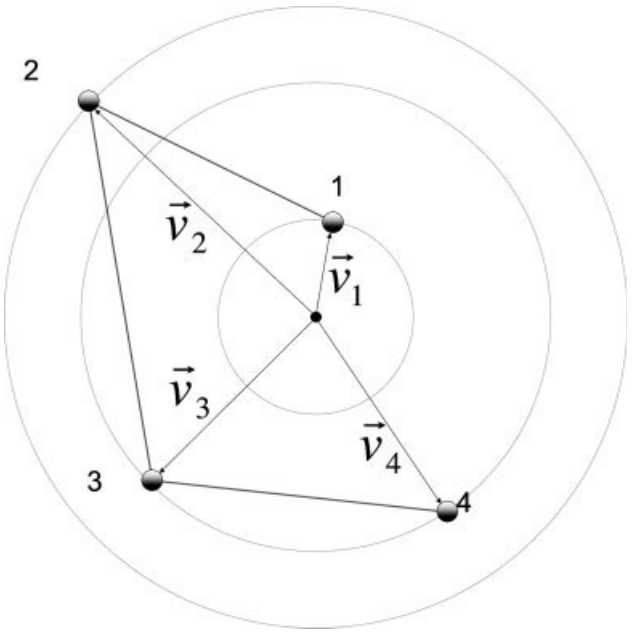

Figure 1

The schematic illustration of the PFP model of a hypothetical 4-atom protein. The hypothetical protein is represented by four shaded circles connected by the thick gray lines. The dotted circle represents the spherical shell whose center is at the fixed point (the black dot) and whose radius is the distance from the fixed point to the atom on the shell. Four atoms are shown in shaded circles and are labeled sequentially from 1 to 4. The vector associated with the atom i *is labeled* $\vec{v}_i$*, where* $i = 1,\ldots,4$*. The atoms lying on the same shell will have the same fluctuations or the same B-factors (for example, atom 3 and 4). The correlation of fluctuations between atoms is computed by taking the inner product of their associated vectors. The motions of atom 1 and 2 are positively correlated, while the motions of atom 1 and 3 (or 4) are negatively correlated.*

shell from the fixed point. Atom 1 and atom 2 are positively correlated because of the acute angle between them, while atoms 1 and 3 (or 4) are negatively correlated because of the obtuse angle between them.

Data set

We selected from PDB-REPRDB[14] 972 protein chains with length $\geq$60 residues. All structures are solved by X-ray crystallography with resolution $\leq$2.0 Å and *R*-factors $\leq$0.2. All chains are of pair-wise sequence identity $\leq$25%. The chains of the data set are listed in Table S1 in the supplementary material.

RESULTS

The r^2 profiles of the nonhomologous data set

For easy comparison, both the *B*-factors and the r^2 profiles are normalized to the corresponding *Z*-scores, which is defined as $Z = (u - \overline{u})/\sigma_u$, where $\overline{u}$ and σ_u are the mean and standard deviation of u. Here u designates either the *B*-factor or the r^2 value. The *B*-factor and the r^2 profiles are denoted by the vectors $\mathbf{Z_B}$ and $\mathbf{Z}_{r^2}$, respectively. The correlation coefficient between them is computed by $\mathbf{Z_B}\cdot\mathbf{Z}_{r^2}$.

Here we consider, as in the previous study,[12] every chain in the dataset as a single dynamical module, each with its own fixed point or centroid, and compute its r^2 profile. The mean correlation coefficient is 0.530. There are 61% (596/972) of the proteins in the data set with a correlation coefficient $\geq$0.5. For the sake of illustration, we will present the examples that the PFP performs well [Fig. 2(A)] and those that the PFP performs poorly [Figs. 3(A)–5(A)]. Figure 2(A) compares the computed and the X-ray *B*-factor profiles of two proteins: 1DQZ:A,[16] the chain A of antigen 85-C from mycobacterium tuberculosis; and 1O6V:A,[17] the internalin functional domain. The correlation coefficients between the calculated and experimental *B*-factors are 0.851 and 0.873, respectively. The atomic *B*-factor, which describes the magnitude of atomic thermal fluctuation, is in fact the auto-correlation function of atomic motion. The more general way to provide the dynamical overview of a protein is the correlation matrix [Eq. (2)], which covers both auto-correlation and cross-correlation functions of atomic motion. There are no cross-correlation data analogous to *B*-factors in the PDB file. The correlation map is usually computed by the NMA,[6–8] which describes the protein dynamics through full molecular mechanics force fields. In the NMA, the protein conformation is first optimized by energy minimization. And then a matrix of the 2nd derivatives of the potential function (referred to as the Hessian matrix) is computed. By diagonalizing the Hessian matrix, we can compute the correlation map from its eigenvectors and eigenvalues. In Figure 2(B), we compare the correlation maps computed by Eq. (2) with those computed by NMA. The correlation maps computed by both methods are quite similar. The computed *B*-factors by NMA, (i.e., the diagonal terms of the NMA correlation map) of both proteins (i.e. 1DQZ:A and 1O6V:A) correlate with the X-ray *B*-factors with the correlation coefficients 0.693 and 0.860, respectively. Note that the PFP model and the NMA are based on completely different principles—the former uses only the geometrical shape of a protein while the latter requires the use a force field, minimization, and diagonalization a matrix. The similarity of these computed correlation maps suggests that much information about protein dynamics can extracted from only the geometrical shape of a protein. In later sections, we will present examples that are of relatively poor correlation coefficients.

Multidomain proteins

The N-terminal DH/PH domain of Trio

Figure 3(A) compares the computed and X-ray *B*-factor profiles of the N-terminal DH/PH domain of Trio

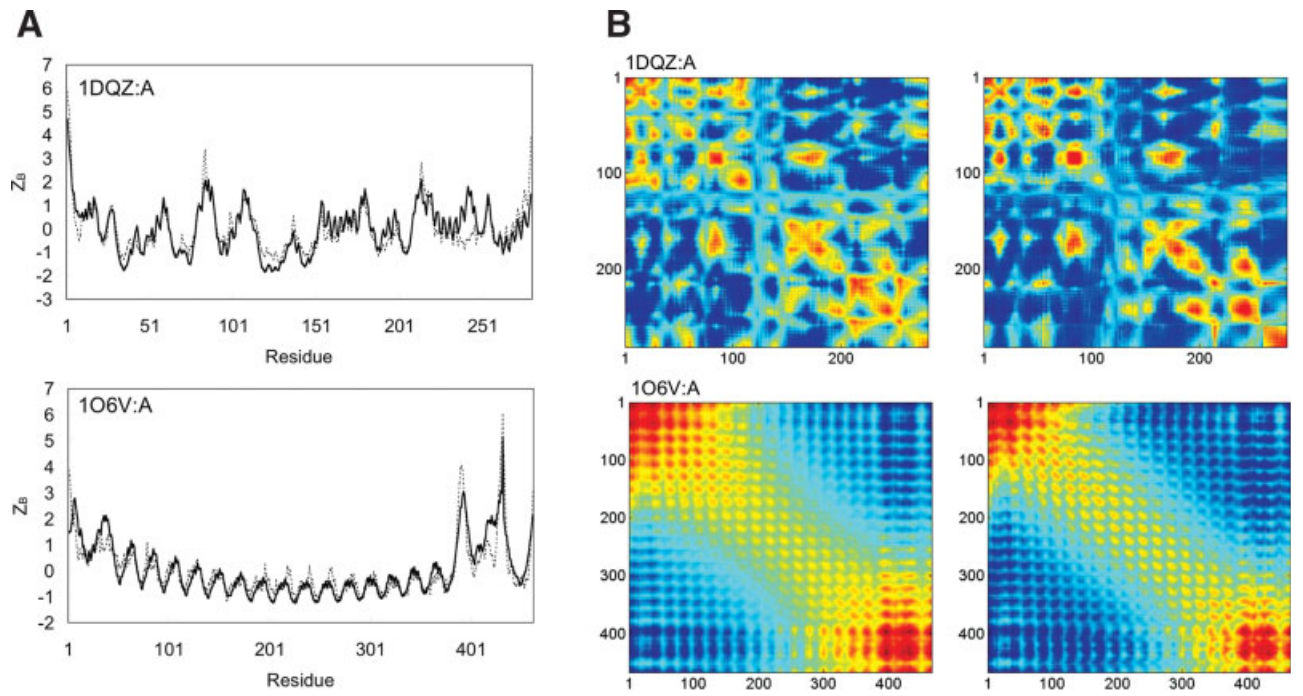

Figure 2

*(**A**) Comparison of the computed (solid line) and the X-ray B-factors (dotted line) of 1DQZ:A (top) and 1O6V:A (bottom). (**B**) The correlation maps of the 1DQZ:A (top) and 1O6V:A (bottom). For each protein, the map on the left is computed using the PFP model and the map on the right by the NMA. The ENZYMIX[15] force-field is used for the NMA coupled with energy minimization. The colors of the map ramp from red (positive correlation) to blue (negative correlation).*

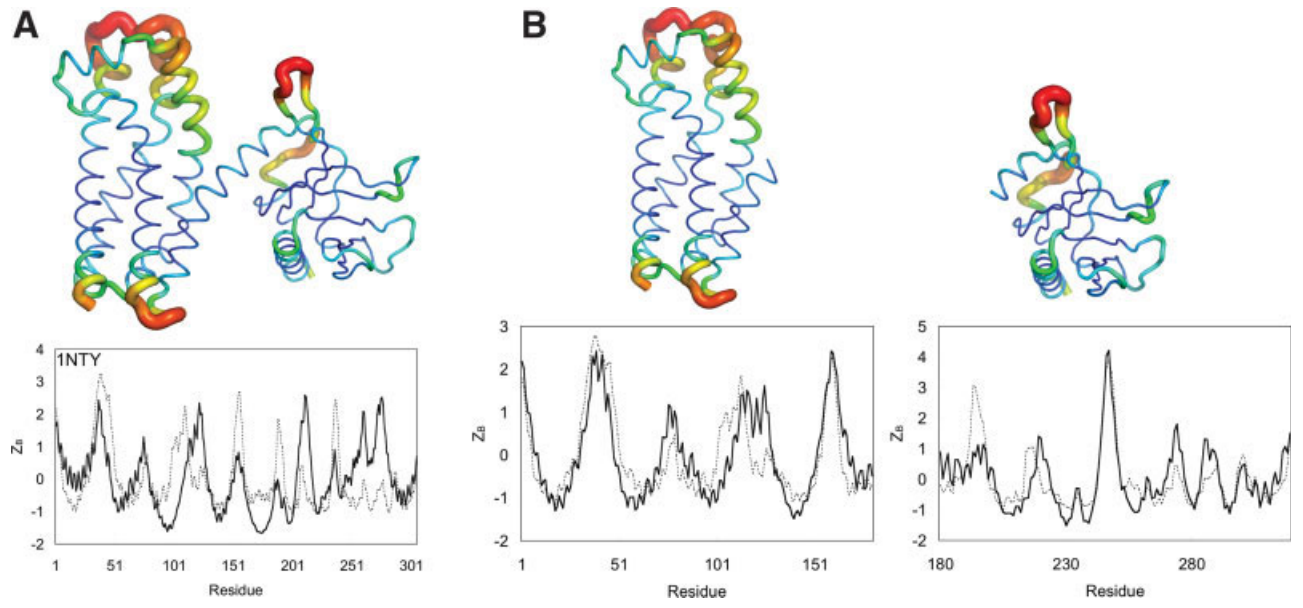

Figure 3

*(**A**) The structure of the N-terminal DH/PH domain of Trio (1NTY) is represented in the cartoon putty representation, where the color is ramped by residue from blue at the lowest B-factor value to red at the highest B-factor value; in addition, the size of the tube reflects the value of B factor—the larger the B-factor value, the thicker the tube. The lower part shows the computed (solid line) and X-ray (dotted line) B factors. (**B**) 1NTY can be divided into two domains, one comprising primarily α helices (upper left) and another one β sheets (upper right). Their computed (solid line) and experimental (dotted line) B-factor profiles based on the centroids of the domains are shown in lower left and lower right, respectively.*

(1NTY). The correlation coefficient between the computed and X-ray *B*-factors is 0.348. We note that the N-terminal DH/PH domain of Trio comprises two domains,[18,19] one consisting of primarily α helices and the other β sheets [the upper part of Fig. 3(B)]. If each domain is treated as a separately unit with its own fixed point, the computed *B*-factor profile is much improved [the lower part of Fig. 3(B)]. The correlation coefficients for two domains become 0.781 and 0.800, respectively. These results seem to suggest that these domains behave like independent dynamical modules, since the PFP model is able to accurately compute the thermal fluctuations of one domain without reference to the other.

Glutamine phosphoribosylpyrophosphate amidotransferase

Another example is glutamine phosphoribosylpyrophosphate amidotransferase (1ECF). The calculated and experimental *B*-factor profiles of its chain B are shown in Figure 4(A). The correlation coefficient between the computed and X-ray *B*-factors is 0.401. However, we note that 1ECF:B comprises two domains[18,19]: a 4-layer sandwich and a 3-layer sandwich [the upper part of Fig. 4(B)]. If each domain is treated as an independent dynamical module, the correlation coefficient becomes 0.813 and 0.789, respectively, almost twice as high as before.

Protein complexes

Formate dehydrogenase-N

The second example is formate dehydrogenase-N (1KQF). Figure 5(A) compares the computed and X-ray *B*-factor profiles of 1KQF:C. The agreement is poor and the correlation coefficient is only 0.120. The asymmetric unit 1KQF is in fact a heterotrimer,[20] as shown in Figure 5(B) (the upper part). The chain C makes extensive contacts with other chains of the complex; for example, the long helix of the chain A, which extends along the long axis of the chain B, makes extensive contacts with the chain C. Hence, it is reasonable to expect that the dynamics of the chain will be affected by other chains. We compute the *B*-factors of the chain C using the centroid of the trimeric complex (instead of that of the sole chain C), and the correlation coefficient is improved to 0.920. On the other hand, the more interesting PDB entry is the biological unit (instead of the asymmetrical unit), since it is supposed to be the functional form of the macromolecule. At present, one can obtain the coordinates of biological units from either PDB or PQS.[21] The PDB and PQS biological units agree on 82% of entries.[22] In this work, we use the PDB biological units. The PDB biological units are built from the crystallographic space group using symmetry operation. The PDB biological unit of 1KQF comprises three asymmetric units, that is, a nonamer. We also do the computation for 1KQF:C using the centroid of the biological unit. The correlation is improved to 0.948. Our results suggest that the dynamics of the chains of the complex are in fact closely coupled to each other, and we cannot accurately compute the thermal fluctuations of one chain without taking into account the presence of the other chains.

The nitrogenase MoFe-protein

Figure 6(A) shows the computed and X-ray *B*-factor profiles of the chain B of nitrogenase MoFe-protein (2MIN:B). The correlation coefficient is 0.416. However, the biological unit of the nitrogenase MoFe-protein is in fact a homotetramer[23] [Fig. 6(B), the upper part]. The biological unit is built from the crystallographic space group using symmetry operations. The biological unit is the macromolecule that has been shown or is believed to be functional. The coordinates of biological unit are taken from the PDB database. If we use the centroid of the tetramer as the fixed point, instead of using the centroid of a single chain, to compute for *B*-factors of the chain, the correlation coefficient improves to 0.801.

Inorganic pyrophosphatase

The third example is inorganic pyrophosphatase (1I6T). Figure 7(A) compares the computed and X-ray *B*-factor profiles of 1I6T. The structure of 1I6T is shown on the upper part of Figure 7(A). The correlation coefficient is 0.502. However, the biological unit of inorganic pyrophosphatase is a homohexamer, comprising six pyrophosphatase molecules.[24] Using the centroid of the hexamer, we computed the *B*-factor profiles of 1I6T [Fig. 7(B)]. The correlation coefficient improves to 0.795.

Comparison of the r^2 profile with the *B*-factor profile

We compute the *B*-factor profiles for the nonhomologous dataset using the following procedures: each protein chain in the data set is checked through the Protein Domain Parser (PDP).[25] If the PDP domains of the chain are not defined, we will resort to the SCOP server; however, if no SCOP entry of the protein chain is defined, we will proceed to the CATH server. If the chain is a multidomain chain, we will compute the r^2 profile for each domain based on the centroid of each domain. The reported correlation of the chain is evaluated as the mean correlation of the domains however, if the chain is not a multidomain chain, we will check whether it is a part of a protein complex or a biological unit—in this work, we use the coordinates of biological units provided by PDB. If the chain is part of a larger protein complex,

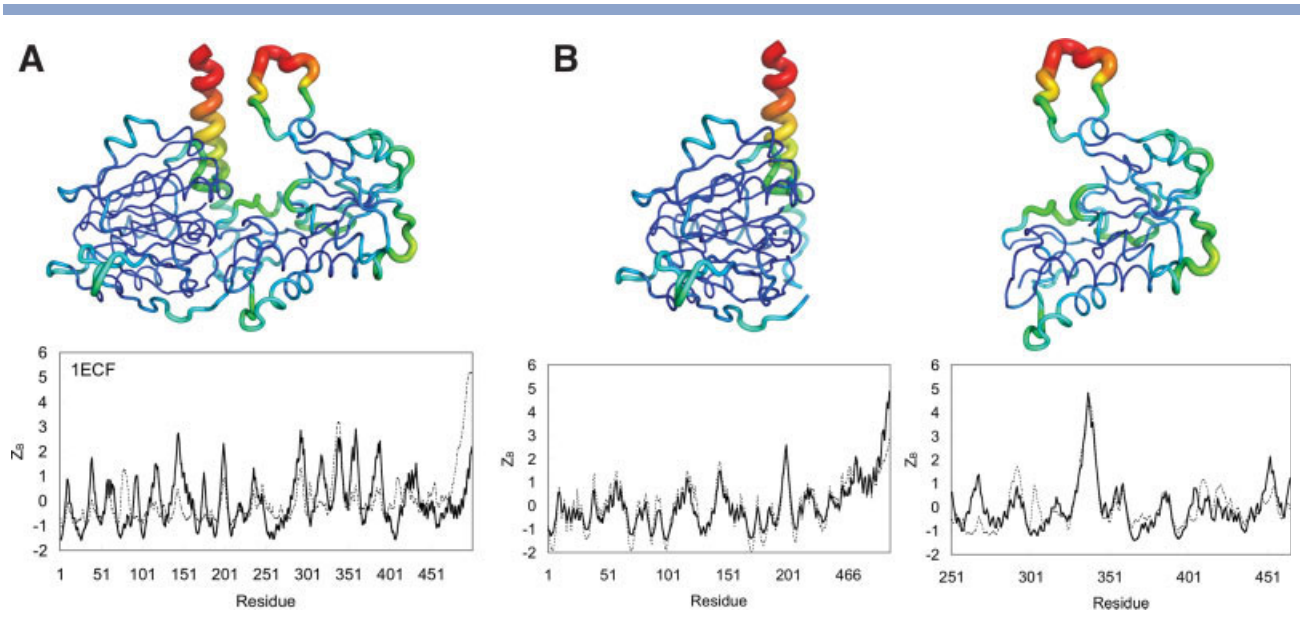

Figure 4

*(**A**) The structure of glutamine phosphoribosylpyrophosphate amidotransferase (1ECF) is shown in the upper part. The computed (solid line) and X-ray (dotted line) B factors are shown below in the lower part. (**B**) 1ECF can be divided into two domains, one comprising 4-layer sandwich (upper left) and the other a 3-layer sandwich. The computed (solid line) and experimental (dotted line) B-factor profiles of each domain based on its respective centroid are shown in lower left and lower right, respectively.*

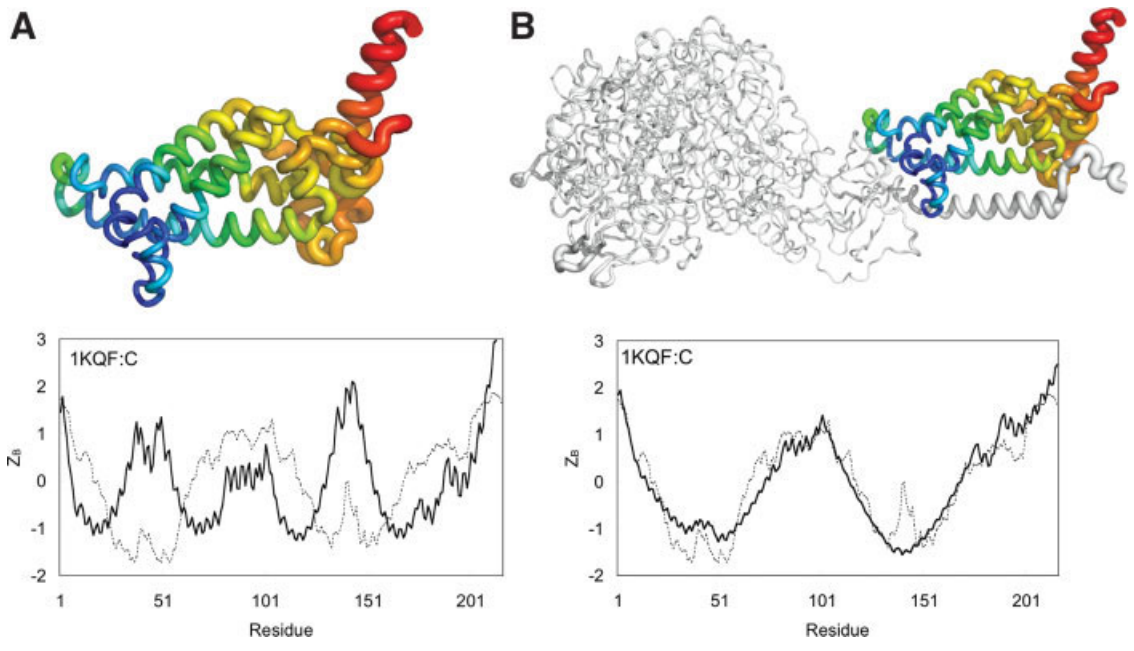

Figure 5

*(**A**) The structure of the chain C of formate dehydrogenase-N (1KQF) is shown in the upper part. The computed (solid line) and X-ray (dotted line) B factors are shown below in the lower part. (**B**) The asymmetric unit (i.e., the trimeric form) of formate dehydrogenase-N is shown in the upper part (the chain C is rainbow-colored, while the others are colored in white). The computed (solid line) and experimental (dotted line) B-factor profiles of the chain B using the centroid of the trimer are shown in lower part.*

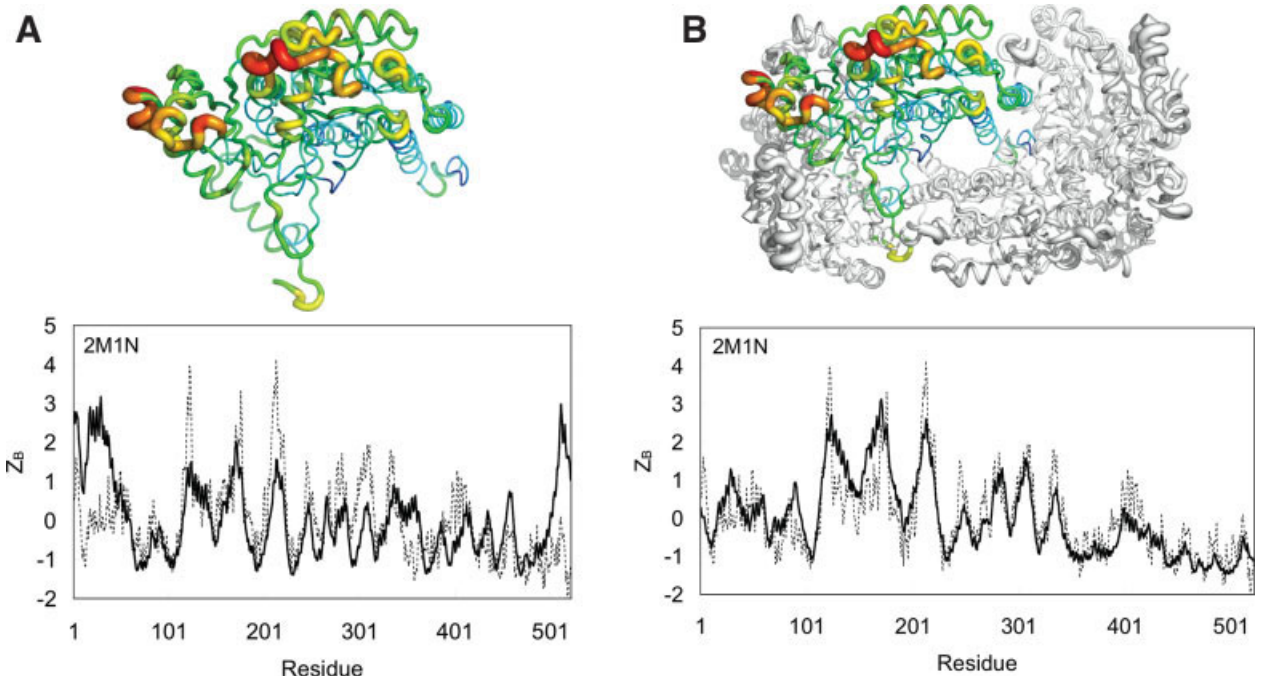

Figure 6

*(**A**) The structure of the nitrogenase MoFe-protein (2MIN) is shown in the upper part. The computed (solid line) and X-ray (dotted line) B factors are shown below in the lower part. (**B**) The tetrameric form of the nitrogenase MoFe-protein is shown in the upper part (one of the monomer is rainbow-colored, while the others are colored in white). The computed (solid line) and experimental (dotted line) B-factor profiles of the colored monomer using the centroid of the tetramer are shown in lower part.*

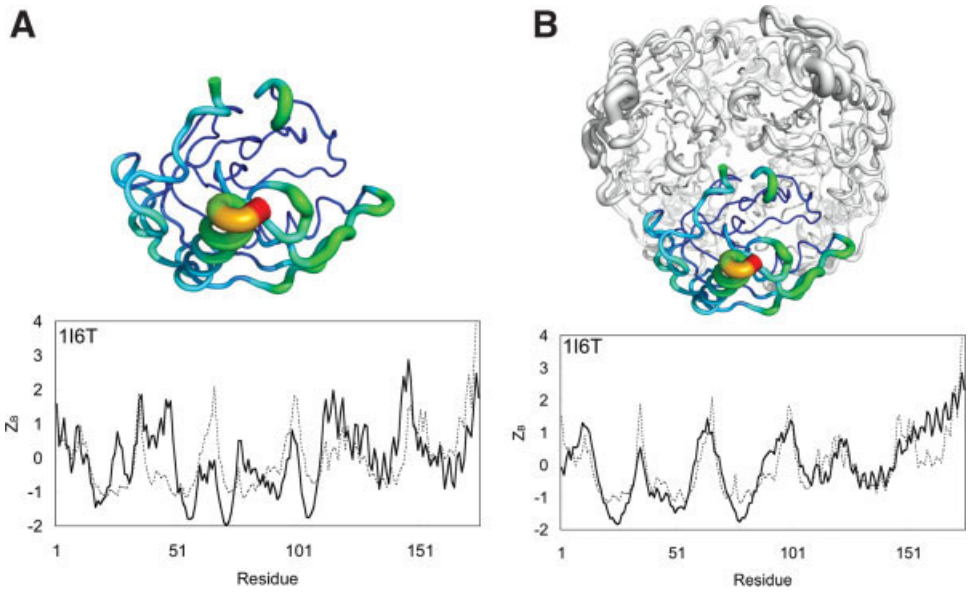

Figure 7

*(**A**) The structure of the chain B of inorganic pyrophosphatase (1I6T) is shown in the upper part. The computed (solid line) and X-ray (dotted line) B factors are shown below in the lower part. (**B**) The biological form (i.e., the hexameric form) of inorganic pyrophosphatase is shown in the upper part (the chain B is rainbow-colored, while the others are colored in white). The computed (solid line) and experimental (dotted line) B-factor profiles of the chain B using the centroid of the hexamer are shown in lower part.*

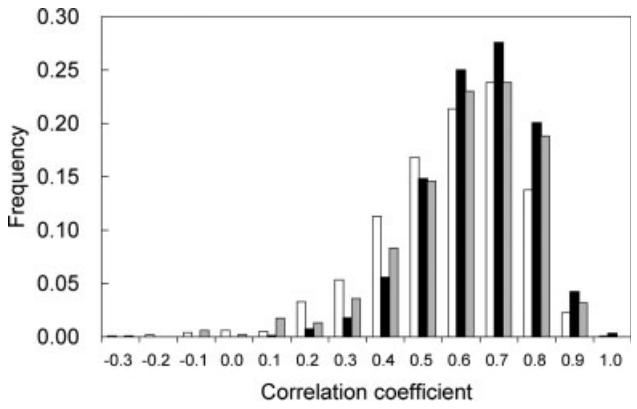

Figure 8

The distribution of correlation coefficients between the X-ray and the computed B-factor profiles using the original PFP (empty bars), the revised PFP (black bars), and the GNM (the grey bars).

we will compute the r^2 profile of the chain based on the centroid of the protein complex, or we will compute the r^2 profile based on the centroid of the original protein chain. The new FPF model improves the mean correlation coefficient to 0.590 from 0.530 and the fraction of proteins with a correlation coefficient $\geq$0.5 increase to 75% from 61%. In addition, we compute the *B*-factor profiles using the GNM for the same set of data set. In the GNM, the mean correlation is 0.558 and the fraction of proteins with a correlation coefficient $\geq$0.5 is 69%. In Figure 8, we compare the distributions of the correlation coefficients between the X-ray and the computed *B*-values using these methods.

Protein dynamics

Recent studies[26–29] showed that the coarse-grained NMA (such as the ENM or GNM) can provide greater insight into the structure-dynamics-function relationship

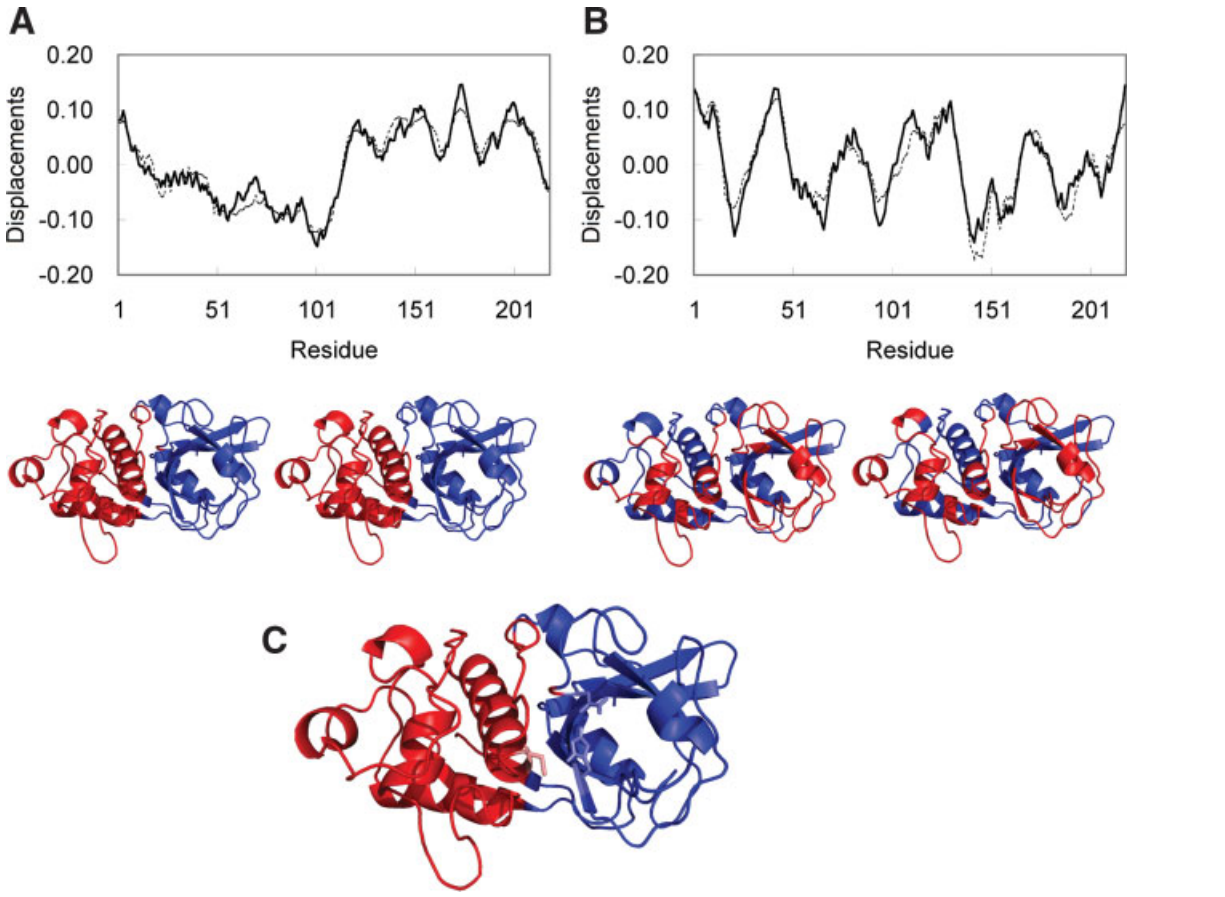

Figure 9

*The normal mode displacements of actinidin (PDBID: 1AEC) along (**A**) the 1st mode and (**B**) the 2nd mode as a function of the residues. The solid line shows the result of the PFP model, while the solid line the GNM. The corresponding ribbon diagrams are shown below with the PFP model on the left and the GNM on the right. The blue and red regions indicate the opposite directions of the displacements of the modes. (**C**) The 1st normal mode computed by the PFP with the catalytic triads (Cys-25, His-162, and Asn-182) shown in the stick model.*

of proteins than the conventional MD simulations because of its ability of sampling a wider range of collective motions. This method uses a simple potential function to describe the protein shape: each Cα atom is connected through a single-parameter harmonic potential to its neighboring atoms that are within a certain cut-off distance, from which a connectivity matrix (i.e., the Hessian matrix) is constructed. This matrix is then diagonalized to obtain the normal mode vectors and frequencies. The correlation matrix, which contains information about correlation between the motions of residues, is calculated by inverting Hessian matrix. On the other hand, the PFP calculates the correlation matrix directly from the geometrical shape of a protein [Eqs. (3) and (4)]. To illlustrate this, we use the PFP model to compute the normal norm motions of the protein actinidin.[30] Figure 9 compares the first two normal modes computed by the PFP model and those computed by the GNM. The correlation coefficients of the 1st and 2nd normal mode eigenvectors between these two models are 0.97 and 0.94, respectively. The agreement is excellent. Previous study[29] shows that the catalytic residues are usually immobilized in order to maintain the delicate arrangement of functional groups. This is evident in Figure 9(B), the catalytic residues (Cys-25, His-162, and Asn-182) of actinidin, which are shown in the stick model, are in the crossover region between the two substructures undergoing oppositely correlated motions.

DISCUSSION

Molecular dynamics method computes the dynamical properties of proteins using a sophisticated molecular force field,[1–3] which describes bond stretching and bond bending in terms of spring-and-ball models, and nonbonded interactions in terms of steric contacts (van der Waals interactions) and the point charge model (the Coulombs law). On the other hand, the PFP model, utilizing only the geometrical shape of a protein, computes the dynamical properties of proteins. In the PFP model, each atom is associated with a vector issuing from the fixed point to the atom. This model states that the atoms lying on the same spherical shell centered at the fixed point will have the same thermal fluctuations, and that correlation of the fluctuations is related to the inner product of the associated vectors of the atom pair. In this article, we show that proteins should be treated in terms of dynamical modules—each module has its own fixed point and its dynamics is essentially independent of others. Most proteins can be considered as a dynamical module with one fixed point; however, for a multidomain protein, each domain may be considered as an independent dynamical module with its own fixed point identified with the centroid of the domain. On the other hand, a biological unit comprising many chains should be considered as a single dynamical module with one fixed point.

Recent studies[31,32] show that the catalytic residues are usually located near to the centroid regions of the enzymes. There are also observations[29,33] that the catalytic residues usually have smaller *B*-factors than others. According to the PFP model, the observation that catalytic residues tend to lie near to the centroid regions is equivalent to the observation that the catalytic residues usually have smaller *B*-factors or less flexible. This is consistent with the computational works of Warshel and coworkers[34–37] who have shown that the major effect to enzyme catalysis comes from smaller reorganization of the protein. The lack of mobility of catalytical residues allows them to maintain similar conformations in both the reactant state and the transition state. In this way, the enzyme is able to reduce the reorganization energy by partially pre-organizing the catalytical residues to stabilize the transition state. The enzyme plays a role of super solvent with smaller reorganization energy than the corresponding reactions in aqueous solutions.

The PFP model as well as other studies[9–12] suggest that the dynamical properties of the folded protein can be to a large extent determined by its folded architecture and seem to be independent of its protein sequence. Hence, it is possible to infer approximate positions of the functional group directly from the folded architecture of a protein without sequence information.

It is not clear whether the PFP model can be generally applied to any "random" structures. Further study is needed to understand what types of structural features of proteins give rise to the PFP model, or what types of evolutional pressure on the protein structure result in the structure-dynamics relationship as described by the PFP model.

ACKNOWLEDGMENTS

We are also grateful to both hardware and software supports of the NRPGM Structural Bioinformatics Core at National Chiao Tung University.

REFERENCES

1. Levitt M, Warshel A. Computer simulation of protein folding. Nature 1975;253:694–698.
2. Warshel A. Bicycle-pedal model for the first step in the vision process. Nature 1976;260:679–683.
3. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature 1977;267:585–590.
4. Warshel A. Molecular dynamics simulations of biological reactions. Acc Chem Res 2002;35:385–395.
5. Rueda M, Ferrer-Costa C, Meyer T, Perez A, Camps J, Hospital A, Gelpi JL, Orozco M. A consensus view of protein dynamics. Proc Natl Acad Sci USA 2007;104:796–801.
6. Brooks B, Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci USA 1983;80:6571–6575.

7. Levitt M, Sander C, Stern PS. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J Mol Biol 1985;181:423–447.
8. Kidera A, Go N. Normal mode refinement: crystallographic refinement of protein dynamic structure. I. Theory and test by simulated diffraction data. J Mol Biol 1992;225:457–475.
9. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter. Atom Anal Phys Rev Lett 1996;77:1905–1908.
10. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 1997;2:173–181.
11. Ming D, Kong Y, Lambert MA, Huang Z, Ma J. How to describe protein motion without amino acid sequence and atomic coordinates. Proc Natl Acad Sci USA 2002;99:8620–8625.
12. Shih CH, Huang SW, Yen SC, Lai YL, Yu SH, Hwang JK. A simple way to compute protein dynamics without a mechanical model. Proteins 2007;68:34–38.
13. Tu D, Blaha G, Moore PB, Steitz TA. Structures of MLSBK antibiotics bound to mutated large ribosomal subunits provide a structural explanation for resistance. Cell 2005;121:257–270.
14. Noguchi T, Akiyama Y. PDB-REPRDB: a database of representative protein chains from the Protein Data Bank (PDB) in 2003. Nucleic Acids Res 2003;31:492–493.
15. Lee FS, Chu ZT, Warshel A. Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIX programs. J Comput Chem 1993;14:161–185.
16. Ronning DR, Klabunde T, Besra GS, Vissa VD, Belisle JT, Sacchettini JC. Crystal structure of the secreted form of antigen 85C reveals potential targets for mycobacterial drugs and vaccines. Nat Struct Biol 2000;7:141–146.
17. Schubert WD, Urbanke C, Ziehm T, Beier V, Machner MP, Domann E, Wehland J, Chakraborty T, Heinz DW. Structure of internalin, a major invasion protein of Listeria monocytogenes, in complex with its human receptor E-cadherin. Cell 2002;111:825–836.
18. Pearl FM, Martin N, Bray JE, Buchan DW, Harrison AP, Lee D, Reeves GA, Shepherd AJ, Sillitoe I, Todd AE, Thornton JM, Orengo CA. A rapid classification protocol for the CATH domain database to support structural genomics. Nucleic Acid Res 2001; 29:223–227.
19. Andreeva A, Howorth D, Brenner SE, Hubbard TJ, Chothia C, Murzin AG. SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Res 2004;32:D226–D229.
20. Jormakka M, Tornroth S, Byrne B, Iwata S. Molecular basis of proton motive force generation: structure of formate dehydrogenase-N. Science 2002;295:1863–1868.
21. Henrick K, Thornton JM. PQS: a protein quaternary structure file server. Trends Biochem Sci 1998;23:358–361.
22. Xu Q, Canutescu A, Obradovic Z, Dunbrack RL, Jr. ProtBuD: a database of biological unit structures of protein families and superfamilies. Bioinformatics 2006;22:2876–2882.
23. Peters JW, Stowell MH, Soltis SM, Finnegan MG, Johnson MK, Rees DC. Redox-dependent structural changes in the nitrogenase P-cluster. Biochemistry 1997;36:1181–1187.
24. Samygina VR, Popov AN, Rodina EV, Vorobyeva NN, Lamzin VS, Polyakov KM, Kurilova SA, Nazarova TI, Avaeva SM. The structures of *Escherichia coli* inorganic pyrophosphatase complexed with Ca(2+) or CaPP(i) at atomic resolution and their mechanistic implications. J Mol Biol 2001;314:633–645.
25. Alexandrov N, Shindyalov I. PDP: protein domain parser. Bioinformatics 2003;19:429–430.
26. Ming D, Kong Y, Wakil SJ, Brink J, Ma J. Domain movements in human fatty acid synthase by quantized elastic deformational model. Proc Natl Acad Sci USA 2002;99:7895–7899.
27. Bahar I, Rader AJ. Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 2005;15:586–592.
28. Ma J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure 2005;13:373–380.
29. Yang LW, Bahar I. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. Structure 2005;13:893–904.
30. Varughese KI, Su Y, Cromwell D, Hasnain S, Xuong NH. Crystal structure of an actinidin-E-64 complex. Biochemistry 1992;31:5172–5176.
31. Ben-Shimon A, Eisenstein M. Looking at enzymes from the inside out: the proximity of catalytic residues to the molecular centroid can be used for detection of active sites and enzyme-ligand interfaces. J Mol Biol 2005;351:309–326.
32. del Sol A, Fujihashi H, Amoros D, Nussinov R. Residue centrality, functionally important residues, and active site shape: analysis of enzyme and non-enzyme families. Protein Sci 2006;15:2120–2128.
33. Yuan Z, Zhao J, Wang ZX. Flexibility analysis of enzyme active sites by crystallographic temperature factors. Protein Eng 2003;16:109–114.
34. Warshel A. Energetics of enzyme catalysis. Proc Natl Acad Sci USA 1978;75:5250–5254.
35. Warshel A. Calculations of enzymatic reactions: calculations of pKa, proton transfer reactions, and general acid catalysis reactions in enzymes. Biochemistry 1981;20:3167–3177.
36. Warshel A, Naray-Szabo G, Sussman F, Hwang JK. How do serine proteases really work? Biochemistry 1989;28:3629–3637.
37. Shurki A, Strajbl M, Villa J, Warshel A. How much do enzymes really gain by restraining their reacting fragments? J Am Chem Soc 2002;124:4097–4107.