SHORT COMMUNICATION

A Simple Way to Compute Protein Dynamics Without a Mechanical Model

Chien-Hua Shih, Shao-Wei Huang, Shih-Chung Yen, Yan-Long Lai, Sung-Huan Yu, and Jenn-Kang Hwang*
Institute of Bioinformatics, National Chiao Tung University, HsinChu 30050, Taiwan, Republic of China

ABSTRACT We found that in proteins the average atomic fluctuation is linearly related to the square of the atomic distance from the center of mass of the protein. Using this simple relation, we can accurately compute the temperature factors of proteins of a wide range of sizes and folds, and the correlation of the fluctuations in proteins. This simple relation provides a direct link between protein dynamics and the static protein's geometrical shape and offers a simple way to compute protein dynamics without either long time trajectory integration or any matrix operations. Proteins 2007;68:34–38.
© 2007 Wiley-Liss, Inc.

INTRODUCTION

The computation of protein dynamics is usually based on a mechanical model. For example, molecular dynamics computes the protein's trajectory based on a sophisticated molecular mechanical model of bond stretching, bond angle bending, bond twisting, van der Waals and electrostatic interactions.[1–3] The recently developed elastic network model,[4–6] which has been successfully applied to analyzing large-scale protein motion, is based on a much simpler mechanical model that each atom in proteins are connected to its surrounding atoms that are within a certain cut-off distance by a one-parameter harmonic spring. Here we report a simple method to compute protein dynamics directly from the static protein geometrical shape without any mechanical models. Our method is based on the observation that the deeper an atom is buried inside a protein structure, the less it will fluctuate around its equilibrium position. We found that this observation goes beyond a mere qualitative description. We found that in proteins the atomic fluctuation is in fact linearly related to the square of the atomic distance from the center of mass of the protein. Using this simple relation, we can accurately compute the temperature factors and the correlation of the fluctuations in proteins.

METHODS

Center of Mass Distance

Let $\mathbf{X}_0$ be the center of mass of the protein, that is, $\mathbf{X}_0 = \sum_k m_k \mathbf{X}_k / \sum_k m_k$, where m_k and $\mathbf{X}_k$ are the mass and the crystallographic position of atom k, respectively. The distance of atom i from the center of mass of the protein is computed by

$$r_i^2 = (\mathbf{X}_i - \mathbf{X}_0)(\mathbf{X}_i - \mathbf{X}_0), \qquad (1)$$

Each protein of size N will have its distinct distribution given by $(r_1^2, r_2^2, \ldots, r_N^2)$, referred to as the r^2 profile. In this work, we computed the r^2-profiles of the Cα atoms of several proteins of sizes ranging from 54 to 736 and different folds including all-α, all-β, $\alpha+\beta$, and α/β folds. The descriptions of these proteins are listed in Table I. We will show in later sections that the r^2 profile is in fact closely related to the temperature factors.

Correlation Between Atoms in Proteins

We can generalize Eq. (1) to compute the correlation between the center of mass distances of atom i and j, that is,

$$c_{ij} = (\mathbf{X}_i - \mathbf{X}_0)(\mathbf{X}_j - \mathbf{X}_0), \qquad (2)$$

Note that when $i=j$, $c_{ij} = r_i^2$. In other words, the autocorrelation reduces to the square of the distance from the center of mass of the protein. We will show later that the c_{ij} is closely related to the correlation between fluctuations of atom i and j, which is given by

$$C_{ij} = \frac{\langle \delta\mathbf{X}_i \ \delta\mathbf{X}_j \rangle}{\sqrt{\langle \delta\mathbf{X}_i \ \delta\mathbf{X}_i \rangle \langle \delta\mathbf{X}_j \ \delta\mathbf{X}_j \rangle}}$$

TABLE I. Information of Proteins Studied in this Work

PDB ID[a]	Length	Protein name	Fold[b]
1PD3:A	54	Influenza A NEP M1-binding domain	ROP-like (α)
1U0S:A	87	Chemotaxis kinase CheA P2 domain	Ferredoxin-like (α+β)
1VJH:A	118	At1g24000	TBP-like (α+β)
1MIJ:A	148	Homeo-prospero Domain	DNA/RNA-binding 3-helical bundle (α)
1F35:A	157	Olfactory marker protein	Olfactory marker protein (β)
1WUB:A	178	Polyisoprenoid-binding protein	Hypothetical protein (β)
1B12:A	239	Singnal peptidase I	LexA/signal peptidase (β)
1IOW	306	D-Ala-D-Ala ligase	PreATP-grasp domain and ATP grasp (α/β)
1NVM:A	340	Bifunctional aldolase-dehydrogenase	TIM barrel and RuvA C-terminal domain-like-(α/β)
1DKQ:A	410	Phytase	Phosphoglycerate mutase-like (α/β)
1FUO:A	456	Fumarase C	Orthogonal bundle and up-down bundle (α)
1FCE	629	Endocellulase CelF	α/α toroid (α)
1LF9:B	674	Glucoamylase (1AYX)	Six-hairpin glycosidases and galactose mutarotase-like (α)
1J7N:B	736	Anthrax toxin lethal factor	Zincin-like and ADP-ribosylation (α+β)

[a]The first four letter is the PDB ID. The fifth letter (if any) after the colon mark is the chain designator.
[b]The definition of the fold follows that of SCOP.

where $\delta\mathbf{X}_i$ and $\delta\mathbf{X}_j$ are the fluctuations of the atom i and j, respectively, around their equilibrium positions. The correlation between fluctuations can be computed using the normal mode analysis,[7–9] which requires the evaluation of the second derivative matrix of the total potential function and the diagonalization of the matrix. In this work, the Enzymix[10] force field is used for the normal mode analysis.

RESULTS

Temperature Factors

Figure 1 compares the computed r^2 profile and the corresponding crystallographic temperature factors (or thermal B factors) of these proteins. The agreement between them is excellent. The good agreement is surprising since our method [i.e., Eq. (1)] is very simple and does not requires either a long time trajectory integration[11] or any matrix operations.[4–6]

Since the B factor is given as $B_i = (8\pi^2/3)\langle\delta\mathbf{X}_i\,\delta\mathbf{X}_i\rangle$, our results suggest the following interesting relation,

$$\langle\delta\mathbf{X}_i\,\delta\mathbf{X}_i\rangle \sim (\mathbf{X}_i - \mathbf{X}_0)(\mathbf{X}_i - \mathbf{X}_0) \qquad (3)$$

Equation (3) provides a direct link between protein's dynamics properties and its static geometrical shape. On the practical side, Eq. (3) also offers a very simple way to compute the temperature factors of proteins.

Correlation Maps

Figure 2 compares the correlation maps computed by Eq. (2) and those computed using the normal mode analysis. Again, the agreement between c_{ij} and C_{ij} is excellent.

DISCUSSION

The dynamic properties of a protein result from a complex network of various interactions; however, our results indicate that they can be computed directly from the protein's geometrical shape without assuming any mechanical models. Now, the question is why equations as simple as Eqs. (1) and (2) will work for a molecule as complex as a protein. It will be instructive to compare our method with the recently developed elastic network model.[4–6] The elastic network model assumes that the fluctuations of an atom are related to the positions of its surrounding atoms that are within a certain cut-off distance. This is in certain way similar to our center of mass distance—fluctuation relation [i.e., Eq. (3)], since intuitively the deeper an atom is buried inside the protein, the more number of atoms it will be surrounded, and the less it will fluctuate. The elastic network model assumes a mechanical model in which the atom and its surrounding atoms are connected to each other through a single-parameter harmonic potential. To compute the temperature factor, the elastic network model inverts a $N\times N$ connectivity matrix (or the Kirchhoff matrix), where N is the size of the protein. This matrix is constructed based on the positions of the surrounding atoms of each Cα atom in the protein.

Originally, it was rather surprising that a simple method like the elastic network model could describe protein dynamics so well[4–6] when compared with other more sophis-
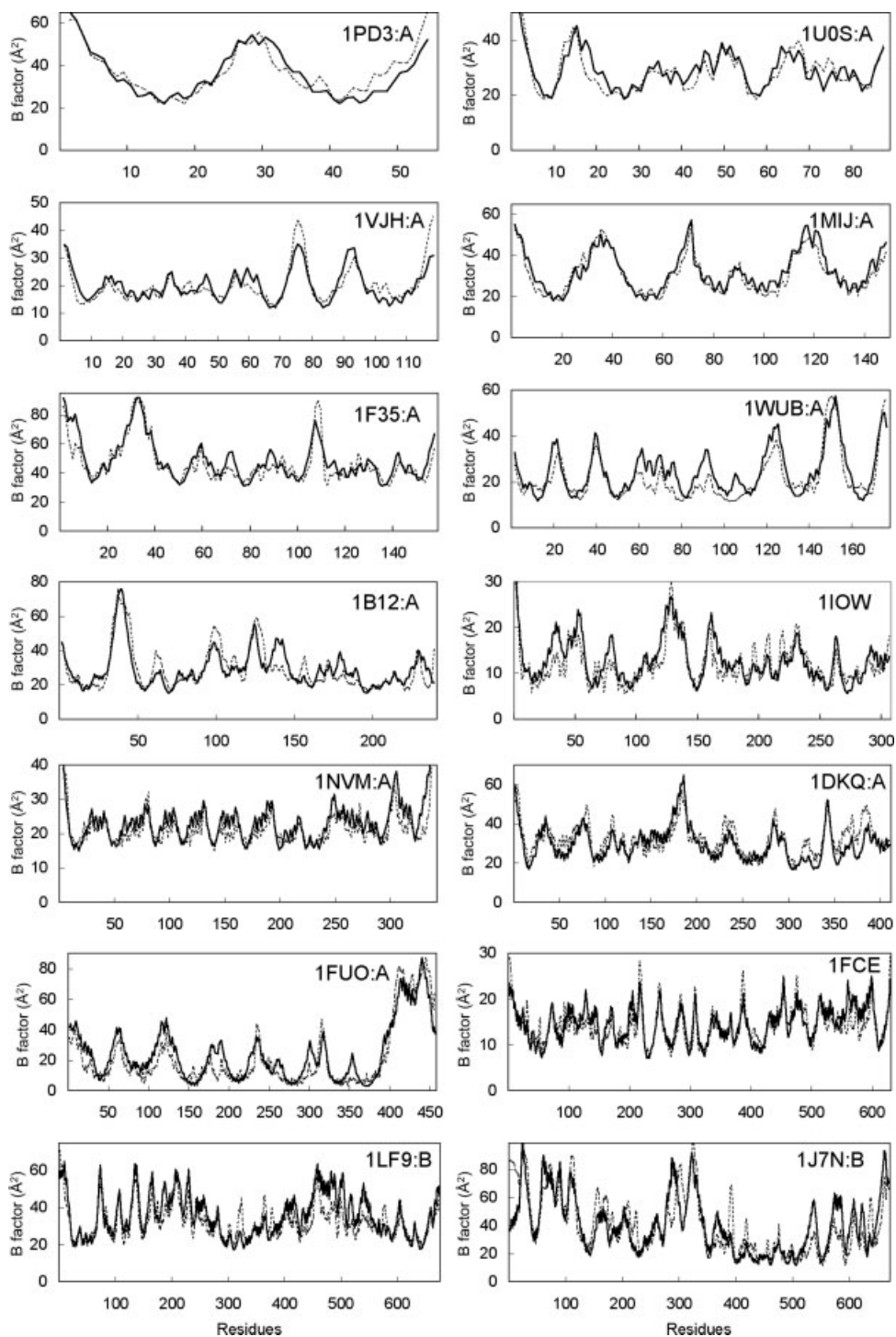
Fig. 1.   The computed (bold line) and the crystallographic (dotted) temperature factors (or *B* factors) of Cα atoms of 1PD3:A, an influenza A NEP M1-binding domain with a ROP-like all-α fold; 1U0S:A, a chemo-taxis kinase CheA P2 domain with a ferrodoxin-like α+β fold; 1VJH:A, a hypothetical protein (At1g24000) with an α+β fold; 1MIJ:A, a homeo-prospero domain with a 3-helical bundle; 1F35:A, an olfactory marker protein with an all-β fold; 1WUB:A, a polyisoprenoid-binding protein with an all-β fold; 1B12:A, singnal pepti-dase I with an all-β fold; 1IOW, ligase with an α/β fold; 1NVM:A, bifunctional aldolase-dehydrogenase with a TIM barrel (α/β) fold; 1DKQ:A, phytase with an α/β fold; 1FUO:A, fumarase C with an all-α fold; 1FCE, endocellulase CelF with an all-α fold; 1LF9:B, glucoamylase with an all-α fold; 1J7N:B, anthrax toxin lethal factor with an α+β fold. The computed *B* factors are normalized to the scale of the crystallographic *B* factors.
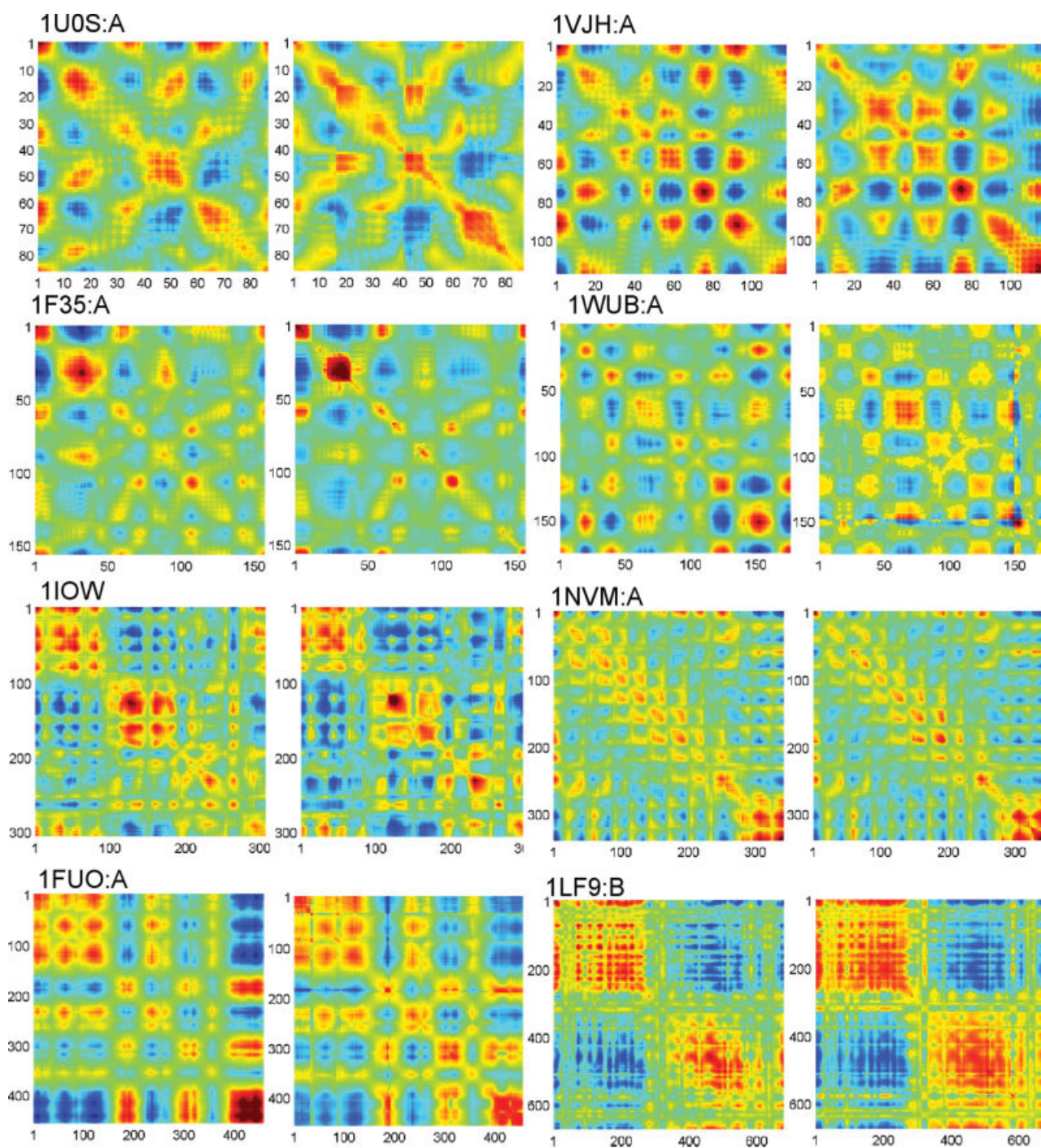
Fig. 2.   The cross-correlation maps for some of the proteins examined in Figure 1. For each protein, the map on the left is computed by Eq. (2) and the map on the right by the normal mode analysis coupled with energy minimization. The colors of the map ramp from red (positive correlation) to blue (negative correlation).

ticated approaches such as molecular dynamics[11] and the normal mode analysis.[7–9] Hence, it is even more surprising that an even simpler model [i.e., Eqs. (2) and (3)] can be developed, which does not assume any mechanical model, does not have any adjustable parameters, and does not require any matrix operations.

Both the elastic network model and our method do not require the knowledge of amino acid sequence. This suggests that protein dynamics of a folded protein are dictated mainly by its folded structure rather than by its chemical properties. Our results further suggest that the folded protein behaves like a rotating sphere
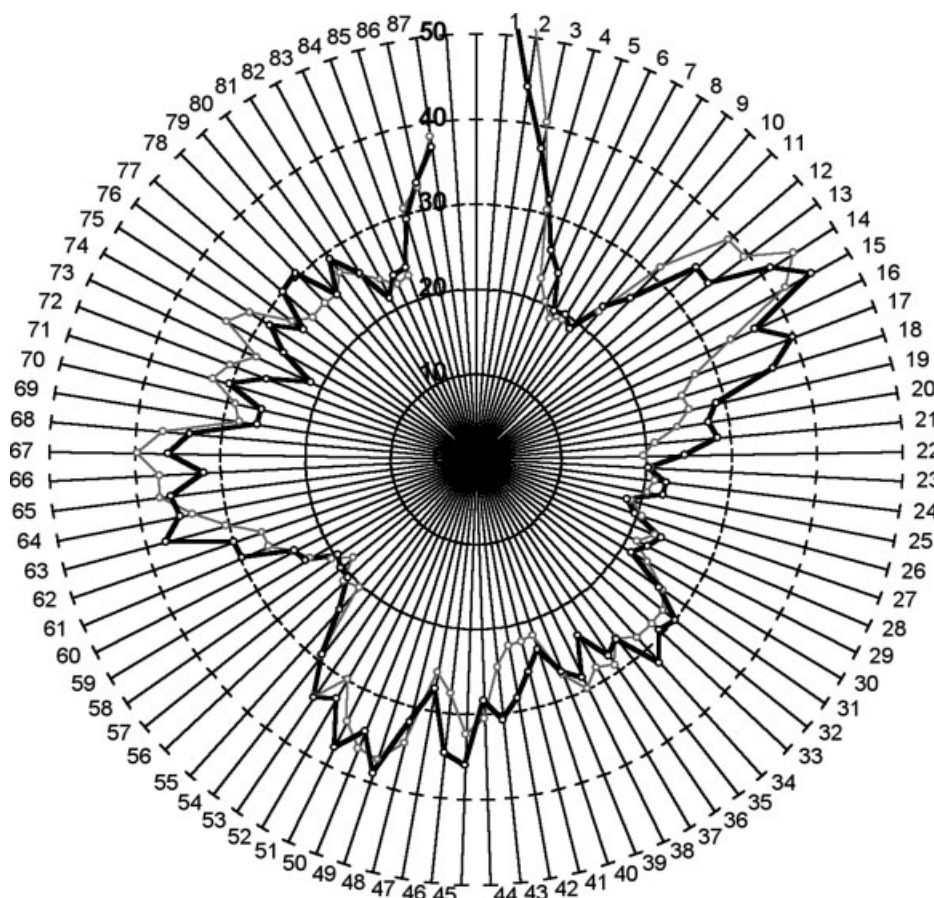
Fig. 3.   The radar plot of the r^2-profile (black line) and B values (gray line) of the Cα atoms of chemotaxis kinase CheA P2 domain (1U0S:A). The residues are numbered on the outer rim of the radar in a clockwise direction and the B values are numbered in the radial axis. The r^2 values are normalized to the scale of the crystallographic B factors for easy comparison.

around its center of mass such that the angular dependent components of its dynamic properties (such as the temperature factor or the correlation of fluctuations) are averaged out. As a result, they can be well approximated by a function depending only on the atomic distances from the center of the protein. To further illustrate this point, Figure 3 shows a radar plot overlapping the r^2 profile and B values of chemotaxis kinase CheA P2 domain with a $\alpha+\beta$ domain. The excellent overlap between these two indicates the atoms with the similar radius from the center of the protein will generally have similar thermal fluctuations (or B factors). In the real three-dimensional space, it is as if that there are many concentric spheres, all of which are centered on the center of mass of the protein but are of different radii, and that the atoms lying on the same surface will tend to have the same thermal fluctuations, regardless of their chemical properties.

With the growing number of protein structures deposited in PDB, our method may provide a very efficient way to systematically characterize protein dynamics in the study of the structure-dynamics relationship of proteins.

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. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 1996;77:1905–1908.
5. 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.
6. 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.
7. 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.
8. Levitt M, Sander C, Stern PS. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J Mol Biol 1985;181: 423–447.
9. 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.
10. 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.
11. Levitt M. Molecular dynamics of native protein. I. Computer simulation of trajectories. J Mol Biol 1983;168:595–617.