Molecular motions and conformational changes of HPPK

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · NOVEMBER 2002

Impact Factor: 2.63 · DOI: 10.1002/prot.10205 · Source: PubMed

4 AUTHORS, INCLUDING:

Ozlem Keskin
Koc University
120 PUBLICATIONS **4,445** CITATIONS

PROTEINS: Structure, Function, and Genetics 49:191–205 (2002)

Molecular Motions and Conformational Changes of HPPK

O. Keskin,[1,2] X. Ji,[3] J. Blaszcyk,[3] D.G. Covell[1†]
[1]*Computational Technologies Laboratory, Screening Technologies Branch, Developmental Therapeutics Program, National Cancer Institute-Frederick, NIH, Frederick, Maryland*
[2]*College of Arts and Sciences, Department of Chemistry, Koc University, Istanbul, Turkey*
[3]*Macromolecular Crystallography Laboratory, National Cancer Institute-Frederick, NIH, Frederick, Maryland*

ABSTRACT **6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) belongs to a class of catalytic enzymes involved in phosphoryl transfer and is a new target for the development of novel antimicrobial agents. In the present study, the fundamental consideration is to view the overall structure of HPPK as a network of interacting residues and to extract the most cooperative collective motions that define its global dynamics. A coarse-grained model, harmonically constrained according to HPPK's crystal structure is used. Four crystal structures of HPPK (one apo and three holo forms with different nucleotide and pterin analogs) are studied with the goal of providing insights about the function-dynamic correlation and ligand induced conformational changes. The dynamic differences are examined between HPPK's apo- and holo-forms, because they are involved in the catalytic reaction steps. Our results indicate that the palm-like structure of HPPK is nearly rigid, whereas the two flexible loops: L2 (residues 43–53) and L3 (residues 82–92) exhibit the most concerted motions for ligand recognition and presumably, catalysis. These two flexible loops are involved in the recognition of HPPKs nucleotide and pterin ligands, whereas the rigid palm region is associated with binding of these cognate ligands. Six domains of collective motions are identified, comprised of structurally close but not necessarily sequential residues. Two of these domains correspond to the flexible loops (L2 and L3), whereas the remaining domains correspond to the rigid part of the molecule. Proteins 2002;49:191–205.**

Key words: conformational changes; coarse grained network model; 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK); normal mode analysis

INTRODUCTION

The study of motions in large molecular systems is a problem of tremendous importance, current interest and high difficulty.[1–3] From the perspective of molecular biophysics, computational molecular dynamics and thermodynamic studies shed light on numerous biochemical processes such as protein folding,[4,5] ligand diffusion,[6] molecular transitions and ligand binding.[7–10] Motivations for these studies are grounded in the principle that intrinsic motions of proteins play an important role in their function[2,11]; an observation postulated before the first protein structure was elucidated.[12] Among the currently used methods for studying molecular motions, the more common approaches include the use of NMR and related forms of spectroscopy,[13–14] molecular dynamics (MD), Langevin dynamics (LD), Monte Carlo (MC) simulations, and normal mode analysis.[2,3,11] Despite these efforts, the prevailing belief is that motions in proteins are still only poorly understood.

An additional complication in the study of molecular motions involves the high computational requirements for most of these studies. Atomic approaches using all-atom empirical potentials appear to be limited to explorations of protein motions near the neighborhood of their native states[15]; restricted mostly to high-frequency fluctuations of only a few Angstroms. Efforts to improve these computational limitations involve time-scale manipulations or the use of "targeted" or "steered" simulations.[8,16,17] Progress in the study of large-scale molecular motions that occur during transitions or conditions where a protein may have more than one equilibrium state are not yet within the scope of most current methods.

Because of its modest computer requirements and absence of errors, normal mode analysis based on harmonic potentials has emerged as an important complement to traditional molecular dynamics simulations for explorations of large motions in proteins. Recent studies show that simplified force fields and coarse grained protein models are appropriate to reveal vibrational dynamics and collective motions of complex biological systems.[18–22] A recently developed method, the Gaussian network model (GNM) and its directionally modified version, the anisotropic network model (ANM)[23–25] are analytical methods that examine the fluctuation dynamics of proteins modeled as coarse-grained elastic networks. In this model, residues are assumed to undergo Gaussian distributed fluctuations about their mean positions, obeying harmonic potentials. Results of GNM and ANM calculations find good agreement with X-ray crystallographic temperature fac-

†Correspondence to: Ozlem Keskin and David Covell, Ft. Detrick Building, 430, Room 215, Frederick, MD 21702. E-mail: covell@mail.ncifcrf.gov

Received 25 September 2001; Revised 30 April 2002; Accepted 16 May 2002

tors,[21,25–27] the H/D exchange free energies of amide protons with solvent[28] and the order parameters from NMR relaxation measurements.[29]

In this article, we demonstrate that our ANM method can be extended to establish relationships between each conformational state of an apo and one or more holo forms of a protein by determining the manifold of accessible transitions from each "reference" state, thus establishing, *a priori,* the most likely transition available to each conformational state. Our approach lends support to the notion that careful examinations of each conformational state may be used in a predictive sense to relate various conformational states. This approach is general and may be extended to study many transitional motions of proteins. Possible examples include studies of (i) the evolution of optimal conformational changes involved in guiding ligands into their binding sites and (ii) intrinsic motions of proteins that involve deformations of stretching and constricting, as are often observed for transport proteins found in the extracellular matrix and nucleus. Here, we will focus on the large-scale motions and conformational changes of the antimicrobial protein, 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (1HPPK)[30] due to ligand binding. HPPK has been recently observed crystallographically to exist in its apo form and three different liganded states. The conformational differences between HPPK's apo and holo forms involve displacements of up to 17 Å and thus represent large-scale motions observed for this protein. Using HPPK as an example, we will show that the ANM method can be extended to find correlations of motions between each conformational state based on the determination of accessible transitions from each reference state to a "new" state.

HPPK Structures

Folate and its cofactors are required for the biosynthesis of purine and pyrimidine nucleotide bases and thus are essential for life. Enzymes involved in the folate biosynthesis pathway represent important targets for the therapeutic treatment of cancer, AIDS, and related diseases.[31,32] All organisms require reduced folate cofactors for the synthesis of a variety of metabolites. Mammals are unable to synthesize folates but have an active transport system for deriving folates from diets, whereas bacteria, fungi, and protozoa have the ability to synthesize folic acid *de novo.*[33] Inhibitors of dihydropteroate synthase and dihydrofolate reductase, both of which are enzymes in the folate biosynthetic pathway, are currently used as antibiotics for treating many infectious diseases. Therefore, the folate biosynthetic pathway is a target for potential antimicrobial agents. HPPK belongs to a class of enzymes including phosphoribosylpyrophosphate synthase, thiamine pyrophosphokinase, nucleotide pyrophosphokinase, and GTP pyrophosphokinase,[30] that catalyze the transfer of pyrophophate from ATP to 6-hydrohymethyl-7-8-dihydropterin (HPP), an intermediate in the pathway for folic acid biosynthesis (See Fig. 1 in Xiao et al.[30]).

The crystal structure of unliganded *Escherichia coli* HPPK (PDB code 1hka) was determined at 1.5 Å resolution by Xiao et al.[30] The structure contains 158 residues and has an α/β/α fold. HPPK is made of six β-strands forming an antiparallel β-sheet and four α-helices. Figure 1 shows the ribbon diagrams of the apo and three holo forms of the molecule. Figure 1(a) displays the fold of the apo-HPPK molecule consisting of a valley starting from a V-shaped cleft between β6, continuing through loop residues 110–116, and running along helix α2, across the face of the four-stranded β-sheet, ending at the C-terminal loop. Three flexible loops: L1 (8–15), L2 (43–53), and L3 (82–92) form one wall of the valley. The other wall of the valley is relatively rigid and is constructed by the structural motif β6-loop-α3. This valley is likely to form the active center where the two substrates—nucleotide (ATP) and pterin (HP)—bind. The crystal structure of HPPK complexed with ATP and HP is shown in Figure 1(b). The root mean square deviation (rmsd) between the two structures is 1.87 Å, based on the C^{α} atoms, whereas it increases to around 7 Å when all heavy backbone atoms are considered. Figure 1(c) displays an intermediate structure, along the catalytic reaction pathway, complexed with ADP and PO_4, Figure 1(d) shows the structure with its two reaction products; AMP and HPP. The main differences amongst these structures are the conformations of the loops L2 and L3 that form the edges of the ligand binding valley. HP and ACP (an ATP analog) molecules are colored pink in these figures. Table I lists the pairwise rmsds between these four HPPK structures. In this table the apo and three holo forms of HPPK are referred to as lig, bin, and pro. These forms appear in Figure 1(b–d), respectively. The greatest conformational change occurs after the HP is converted to HPP upon phosphoryl transfer of ATP to HP.

The structural changes observed in this system are qualitatively consistent with other HPPK structures available in the protein databank. A search in the protein database for structural analogs of HPPK using the NCBI's VAST utility[34] and the DALI structure searching program[35] finds two additional structures, one complexed with 87Y (pterin analog) and ATP and the other complexed with ROI (pterin analog; PDB codes 1dy3 and 1cbk, respectively). A consistent observation of all these systems is that as pterin derivatives bind to these kinases, their L2 loop closes over the ligand. Subsequent binding of nucleotide analogs results in further closure/opening of the L3 loop around the nucleotide. The exact nature of these structural changes cannot be determined from their crystal structures alone. In the following analysis, we will demonstrate that an apparent synchrony exists between these loop closures (and openings) that reveals a precise ordering of these events.

METHODS

It has been shown that for native proteins, the very low frequency normal modes make major contributions to the conformational fluctuations at thermal equilibrium.[2] Such motions can change the interactions of proteins with other molecules and its environment. One of the standard techniques for studying protein dynamics, and in particular

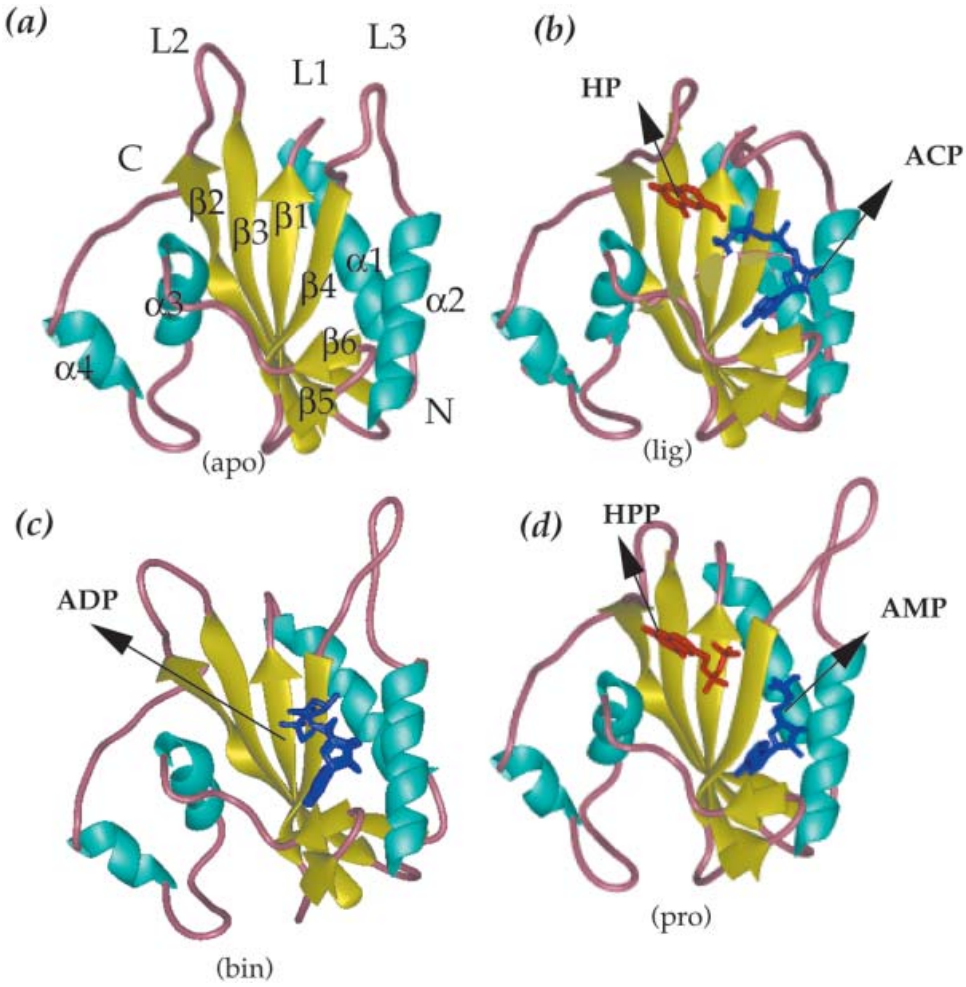

Fig. 1. Ribbon diagrams of (a) apo- and (b–d) holo-forms of HPPK. The helices are colored cyan, strands yellow, and loops pink. (b) The complex with two substrates (ACP and HP). Two loops L2 and L3 close to cover the ligands. (c) The intermediate structure with ADP and PO_4. L3 moves further away from the catalytic center. (d) The enzyme with two products (AMP and HPP) bound. The nucleotide and pterin analogs in (b–d) are shown in blue and red, respectively.

TABLE I. Comparison of apo and holo Forms of HPPK Based on C^{α} rmsd Values†

	Ligands	1hka (apo)	Lig	Bin	Pro
1hka (apo)	—	0	1.87	3.10	3.13
Lig	ATP, HP		0	4.02	3.91
Bin	ADP, PO_4			0	0.68
Pro	AMP, HPP				0

†Values are expressed in Angstroms.

low frequency domain motions, is normal mode analysis.[36] Comparisons of low-frequency normal modes and the directions of large amplitude fluctuations in molecular dynamics simulations indicate clear similarities.[18,37–39]

In Figure 2, a schematic representation of the model used in this study is given. Here, C^{α} atoms are taken as the interaction sites for residues that are connected by virtual bonds (linear springs between the beads) to form the protein backbone. This model assumes that the protein in the folded state is equivalent to a three-dimensional elastic network. Therefore, the molecule is modeled as a linear chain of N beads (residues) that undergo Gaussian distributed fluctuations and $N - 1$ bonded springs connecting them. The beads are subject to harmonic potentials from all neighboring beads, regardless of backbone connections. Only interactions between pairwise beads closer than a cutoff distance are considered to be in contact, thus reducing the spring number from its maximum of $N \times (N - 1)/2$. Previous studies[23–25] accounted for the anisotropic effects in three-dimensional space by decomposing the motions into a series of vibrational modes. In this analysis backbone fluctuations were based on the Hessian matrix as determined from the second derivative of the equilibrium harmonic potential energy. Using the crystal coordinates of C^{α} atoms as the equilibrium state, this analytical method efficiently finds the intrinsic fluctuations and the collective individual motions with their corresponding amplitudes.

The details of our method have appeared in prior publications.[23–25] In this section we briefly review the salient features of our formulation. The interaction potential energy between all pairs of residues is assumed to obey a harmonic potential that establishes a Gaussian distribution of interatomic distances about their equilibrium values, assumed to be the crystal coordinates. The force constant, k, of the harmonic potential neglects differences in amino acid types such that the potential energy governing fluctuations, $V_{(i,j)}$, can be determined from their equilibrium fluctuations, $\Delta R_{(i,j)}$, as $V_{(i,j)} = \frac{1}{2} k \Delta R_{(i,j)}$.[2] For small displacements (fluctuations), the potential energy can be expressed as a power series in the displacements ΔR_i:

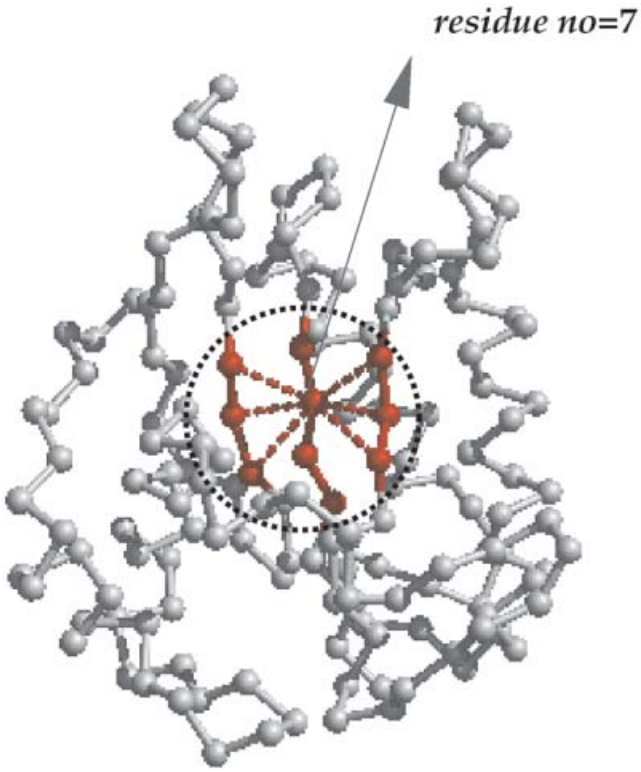

Fig. 2. Schematic representation of the coarse-grained network model used in this study. The protein is represented by a virtual backbone, each interaction site corresponding to an α-carbon. Residue pairs are subject to harmonic forces provided that they are within a cutoff distance around the residue *i*. In this figure, a set of residues within a cutoff distance of 10 Å for the residue $i = 7$ is colored red. The dashed lines represent the interactions between nonbonded sites.

$$2V = 2V_0 + 2\sum_{i=1}^{N}\left(\frac{\partial V}{\partial \Delta R_i}\right)_0 \Delta R_i + \sum_{j=1}^{N}\sum_{i=1}^{N}\left(\frac{\partial^2 V}{\partial \Delta R_i \partial \Delta R_j}\right)_0 \Delta R_i \Delta R_j + \text{higher order terms} \quad (1)$$

In the equilibrium state, the protein assumes a conformation that is energy minimized with respect to all residue fluctuations, eliminating the first term in the series. Furthermore, the first derivative of the potential energy with respect to fluctuations is zero. This is equivalent to the assumption that the net force on the molecule is zero.

$$\left(\frac{\partial V}{\partial \Delta R_i}\right)_0 = 0 \quad (2)$$

For sufficiently small amplitudes of fluctuation, the higher terms (cubic, quadratic, etc. in the ΔR_i's) can be neglected, so that the power series reduces to

$$2V = \sum_{j=1}^{N}\sum_{i=1}^{N}\left(\frac{\partial^2 V}{\partial \Delta R_i \partial \Delta R_j}\right)_0 \Delta R_i \Delta R_j \quad (3)$$

This potential is equivalent to the harmonic potential energy of springs and indicates that the potential energy of

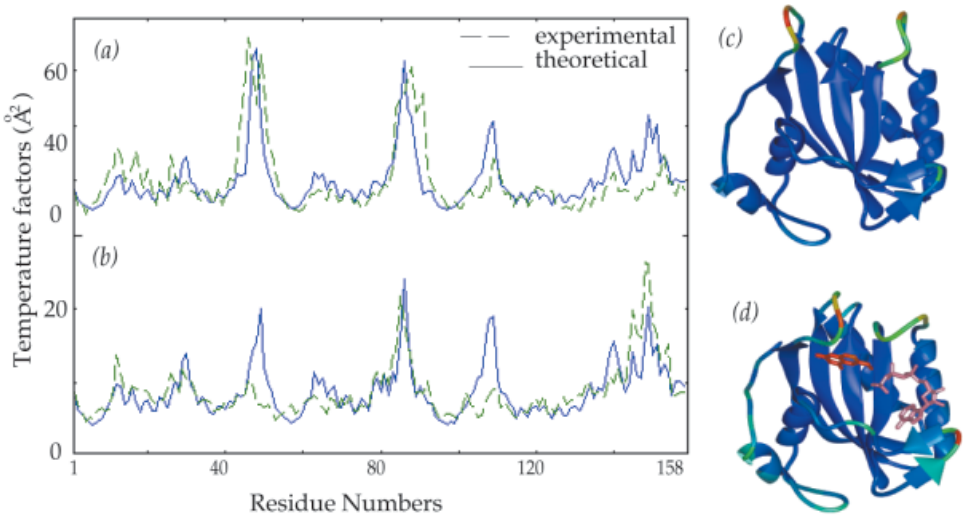

Fig. 3. Temperature factors of (a) apo-HPPK and (b) lig-HPPK complexed with HP and ACP. The solid lines represent the results of our calculations, dashed lines are the experimental temperature factors. (c,d) The ribbon diagrams of the apo- and lig-HPPK color coded on the basis of the relative amplitudes of fluctuations of the individual residues. A color spectrum of red-to-blue is used to represent different levels of motion, where the smallest displacement level is blue and the highest is red. The ligands are colored pink and red in (d).

this model can be approximated as a sum of quadratic terms in fluctuations. The potential energy can be expressed in matrix form in terms of the $1 \times 3N$ fluctuation vector (ΔR) and the $3N \times 3N$ Hessian matrix (H) of interresidue force constants as

$$V = \frac{1}{2}[\Delta R]^T H[\Delta R] \tag{4}$$

where the *Hessian* matrix (H) is defined as

$$H_{ij} = \frac{\partial^2 V}{\partial \Delta R_i \partial \Delta R_j} \tag{5}$$

The ijth super-element ($i \neq j$) of H_{ij} is

$$H_{ij} = \begin{bmatrix} \partial^2V/\partial X_i\partial X_j & \partial^2V/\partial X_i\partial Y_j & \partial^2V/\partial X_i\partial Z_j \\ \partial^2V/\partial Y_i\partial X_j & \partial^2V/\partial Y_i\partial Y_j & \partial^2V/\partial Y_i\partial Z_j \\ \partial^2V/\partial Z_i\partial X_j & \partial^2V/\partial Z_i\partial Y_j & \partial^2V/\partial Z_i\partial Z_j \end{bmatrix}$$

The equilibrium cross-correlations between residue fluctuations is given by[40]

$$\langle \Delta R_i \cdot \Delta R_j \rangle = \int \Delta R_i \cdot \Delta R_j e^{(-V/k_BT)} d\{\Delta R\} / \int e^{(-V/k_BT)} d\{\Delta R\}$$
$$= (k_BT)[H^{-1}]_{ij} \tag{6}$$

where k_B is the Boltzmann constant and T is the absolute temperature. The angular brackets designate the fluctuations of C^α atoms over all vibrational modes of motion, $[H^{-1}]_{ij}$ is the trace of the 3×3 matrix enclosed in the inverse of the $3N \times 3N$ Hessian matrix for residue types i and j, and $d\{\Delta R\}$ is the integration all over the residue fluctuations. In parallel with normal mode analysis, the last equality in the above equation follows from the fact that the mean square displacements in mode space are inversely proportional to their eigenvalues[26,28,41]

$$\langle (\Delta R_i)^2 \rangle = (k_BT)[H^{-1}]_{ii} \tag{7}$$

It should be noted that the determinant of H is zero. Therefore, H cannot be inverted directly. H^{-1} is an approximation calculated from $3N - 6$ nonzero eigenvalues, λ_m, and their corresponding eigenvectors, u_m, of matrix H.

Modal decomposition

Information on global dynamics is acquired by decomposing the motions into a series of modes and concentrating on the modes at the slowest (largest amplitude) end of the spectrum. The latter are the most cooperative modes in the folded state, in that they dominate the coupled collective motions of large size structural blocks. To elucidate the mechanism of these motions, as a first step, H can be written as the product of the matrix U, the diagonal matrix $\mathbf{\Lambda}$, and the transpose of U as

$$H = U\Lambda U^T$$
$$= [u_1 u_2 u_3 \ldots u_{3N}]\mathrm{diag}(\lambda_1\lambda_2\lambda_3 \ldots \lambda_{3N})[u_1 u_2 u_3 \ldots u_{3N}]^T \tag{8}$$

The diagonal elements of the $\mathbf{\Lambda}$ are the eigenvalues (λ_m) of H. We note that because U is an orthonormal matrix whose columns are eigenvectors (u_m) of H, $U^T = U^{-1}$, then the inverse of H is found from $H^{-1} = U\Lambda^{-1}U^T$ or in other words, as the sum of the contributions from each mode:

$$H^{-1} = \sum_m \lambda_m^{-1}[u_m u_m^T] \quad 1 \leq m \leq 3N - 6 \tag{9}$$

The above equation is carried out over the range of modes $1 \leq m \leq 3N - 6$, associated with the normal or fundamental modes of vibration. The last six modes with zero eigenvalues correspond to purely translational and purely rotational motion modes. Vibrational modes give the directions and relative amplitudes of the atomic displacements in each mode. $[u_m u_m^T]$ is a $3N \times 3N$ matrix, representing the contribution of each of the mth eigenvectors to H. This equation provides a simple means of decomposing the dynamics into a series of modes. In other words, the mean square fluctuation of residue type i associated with the mth mode of motion is found from

$$\langle (\Delta R_i)^2 \rangle_m = \lambda_m^{-1}[u_m u_m^T]_{ii}$$
$$\langle (\Delta R_i)^2 \rangle_m = \lambda_m^{-1}[u_m]_i[u_m^T]_i \tag{10}$$

$[u_m u_m^T]_{ii}$ is simply found by the product of $[u_m]_i$ with $[u_m^T]_i$, and the result is a scalar giving the mean square fluctuation of residue type i.

The cross correlations $CC(i, j)$, between fluctuations of C^α atoms are normalized as

$$CC(i,j) = \frac{\langle \Delta R_i \cdot \Delta R_j \rangle}{\{\langle \Delta R_i \cdot \Delta R_i \rangle \langle \Delta R_j \cdot \Delta R_j \rangle\}^{1/2}} \tag{11}$$

The positive and negative limits of $CC(i, j)$ correspond to pairs of residues exhibiting fully correlated (same direction, same sense) and fully anticorrelated (same direction, opposite sense) motions, respectively. Zero correlation refers to totally uncorrelated, or orthogonal motions. These correlations can also be examined for individual modes to provide information about their coupling. In particular, the lowest nonzero eigenvalue (λ_1) refers to the frequency of the most cooperative (slowest) motion, and the corresponding eigenvector, u_1, reflects the shape of this so called global, or dominant, collective mode in three-dimensional Cartesian space. This mode, together with the following modes, is often implicated in biological function.[21–26] For purposes of our analysis, an examination of a few eigenvectors associated with the dominant modes of motions provides information about the identity of structural elements acting as hinges and levers in the cooperative structural changes around the native state. Regions acting as hinges in the global motion are further distinguished in our analysis by having severely hindered fluctuations in this fundamental mode. The slowest motions usually correspond to rigid elements moving about a hinge or rigid structural elements undergoing coherent, coupled movements, and these exhibit motions increasing in amplitude at increased distances from the hinge center.

RESULTS AND DISCUSSION

Temperature Factors

Figure 3 displays the temperature factors of (a) apoenzyme HPPK and (b) lig-HPPK complexed with two substrates: HP and ACP. Curves for the other two conformations of the enzyme are not shown in this figure. The solid lines represent the theoretical mean square fluctuations of amplitudes $\langle \Delta R_i^2 \rangle$, evaluated from Equation 7, and the dashed curves are the X-ray temperature factors for the individual α-carbons, reported in the PDB files of these structures. The theoretical values are scaled to make the areas under the two curves the same. The agreement between theory and experiments is quite satisfactory. The correlation coefficients in all cases are around 0.8. In part (a) the major difference between the theory and experiment arises in the peptide segment (145–152) corresponding to the α-helix (α4)—found to be much flexible in our calculations—and in part (b) the loop 45–52 and the loop-β5-loop motif (residues 102–111) are also calculated to be more flexible than the experimental results. Two of the three flexible loops: L2 (43–53) and L3 (82–92) are identified as the maxima in the curves, the third flexible loop L1 (8–15), which exists in two different conformations in the native state, does not exhibit large fluctuations as in the two other loops. The rest of the molecule is almost rigid.

A superimposition of the normalized overall residue fluctuation curves of the four structures provides insights about the dynamic differences observed for different liganded states. The main flexibility changes are observed in the loops. L2 (43–53) exhibits enhanced flexibility in the bin form (as bound by the nucleotide: ADP) and a suppressed flexibility in the lig (complexed ATP and HP) and prod (complexed with AMP and HPP) forms. L3 (82–92) shows higher flexibility in both bin and pro forms and suppressed flexibility in the lig form. As a summary, the bin structure bound to only ADP exhibits the highest flexibility. This high flexibility when bound by ADP might be necessary to initiate the catalytic process as a prelude to the proper conformation necessary for the subsequent binding or release of pterin. On the other hand, the most rigid structure is the lig form (complexed with ACP and HP). This rigidity might serve to isolate the catalytic center and substrates from the outside effects in order to enhance the efficiency of phosphoryl transfer.

Figure 3(c,d) shows the ribbon diagrams of the apo- and lig-HPPK, color-coded according to the amplitudes of motions displayed in Figure 3(a,b), respectively. The least flexible regions (colored blue) correspond to ligand binding region in its holo- and apo-forms. Spectral colors are used to represent the different levels of flexibilities, with the smallest displacement shown in blue and the highest in red (peaks in parts a and b). The ACP and HP molecules are shown in pink and red, respectively. The loops are the most flexible parts of the HPPK, and it is easily observed that the lig form is much more rigid than the apo form, although the C terminal helix becomes more flexible in the lig form. In its apo form, the L2 and L3 loops are very mobile, with a high capacity to recognize ligands or adjust their positions for the capture of the ligands, whereas in the liganded form the C-terminal helix becomes more mobile as the motions of the core region and loops are constrained by their ligands. The high mobility of the C terminal helix may suggest that this helix plays a role in recognizing other proteins in the folate synthesis pathway, subsequent to phosphoryl transfer.

Dominant Modes of Motions

Figure 4 displays the global collective mode shapes for HPPKs in the (a) first, (b) second, and (c) third slowest modes for individual residues indicated by C^{α} positions. The dashed lines represent the mode shapes for HPPK (apo form), and the solid lines give the mode shapes for HPPK complexed with ACP and HP (lig form). The loops L2 (43–53), L3 (82–92), 102–111 (including strand β5), 132–142, 143–149 (helix α4), 150–153 are observed to exhibit high mobility. Upon ligand binding the loss of flexibility in loops L2 (43–52) and L3 is accompanied by enhanced flexibility in the C-terminal residues (~140–158) for their second and third modes (see Fig. 4). Comparisons of the available four crystal structures of HPPK find that the palm like region of the enzyme is almost rigid; the only flexible parts are the two loops L2 and L3 that change their conformations upon substrate binding and conversion of substrates to products.[42] Thus our calculated results are in good agreement with these observations. The amplitude of motion for loop L3 reaches up to 17 Å when comparing the apo and bin structures and 15 Å, as calculated from the weighted $(1/\lambda i)$ contribution of the fluctuation vectors driven by the first three global modes. The other loop (L2) also moves 7 Å in the crystal structures and exhibits a movement of ~15 Å, as estimated by the first three modes. We also note that the other parts of the molecule are predicted to undergo fluctuations less than 1.5 Å.

The minima of the slowest collective modes would be predicted to correspond to hinge points, active regions or regions where the ligands are bound according to our model.[21,26] HP molecule is sandwiched between the aromatic rings Tyr53 and Phe123. Asn55 is a conserved residue and it also forms hydrogen bonds with the HP molecule, which is in the vicinity of Asp95 and ATP. Mg^{2+} ion is surrounded by Glu77, Asp95, Asp97, and ACP in the liganded structure. Two conserved residues, His115 and Arg121 form H bonds, and Leu111 is the other residue in contact.[30] All the mentioned residues here are indeed observed as minima of the slowest mode shape of HPPK, with ACP and HP substrates. The active center is proposed to stay in the valley formed by one rigid and one flexible wall.[30] The relatively rigid wall includes residues Met118, Met124, Leu125, Trp126, Pro127, Leu128, and Phe129. All of these residues also correspond to the minima of the curves in the figures. As a result, among the highly conserved residues, Gly8, Asn10, Gln50, Asn55, Glu77, Arg82, Arg92, Asp95, Asp97, Pro114, His115, and Arg121, all are located in the active region, and appear as minima of the slowest mode shapes. Nine residues are involved in HP binding: Gly8, The42, Phe43, Leu45, Tyr53, Asn55,

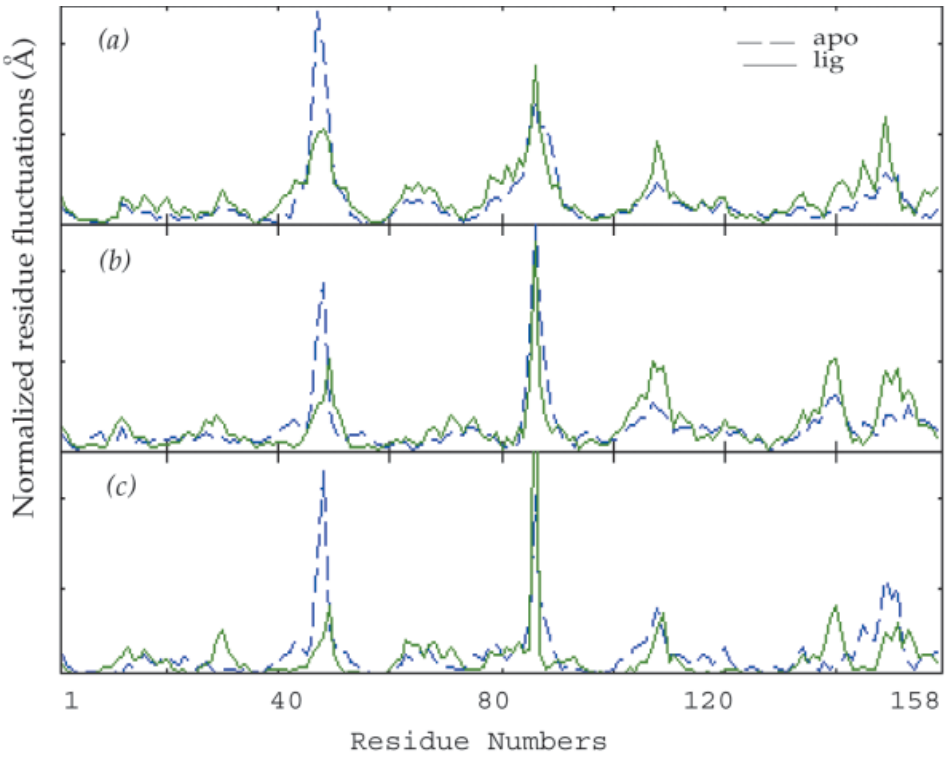

Fig. 4. Most global mode curves for HPPK in the (a) slowest, (b) second slowest, and (c) third slowest modes. The dashed curves represent the mode shapes for HPPK (apo form), and the solid curves give the mode shape curves for lig-HPPK complexed with HP and ACP (lig form).

Trp89, Asp95, and Phe123. Twelve residues (Gln74, Glu77, Arg84, Arg88, Trp89, Arg92, Ile98, Arg110, Thr112, His115, Tyr116, and Arg121) are involved in nucleotide binding. Indeed, these residues also correspond to minima in the slowest mode curves.

Mechanisms of Global Motions

Figure 5 shows the most global motions for HPPK. Figure 5(a–c) displays the first three slowest modes for the apo enzyme. The yellow and magenta structures represent the two conformations that correspond to the extremes of these mode motions (i.e., the native structure is believed to fluctuate between these two conformations). L2 and L3 make negatively correlated motions [Fig. 5(a)]. These two loops move far away from each other and then return as close neighbors (i.e., the loops open and close the active site in the valley). L3 also moves up and down. In Figure 5(b) the two loops L2 and L3 undergo concerted motions; they elongate and shorten simultaneously. Note that as a visual aid, the amplitudes of motions in these figures are exaggerated. In Figure 5(c) these two loops again move in opposite directions, as one shortens the other elongates or vice versa.

Figure 5(d–f) displays the global modes of motions for lig-HPPK with substrates ACP and HP. The yellow and magenta structures represent the two extreme conformations according to our calculations. Loops L2 and L3 are observed to loose their flexibility upon ligand binding. The most discernable motions are observed in the loop-helix-loop motif at the C terminal with loops L2, L3, and helix β5 together with the loops before and after (peptide segment 105–110). Ligand binding appears to introduce an enhanced mobility in these regions.

The coupled motions of these two loops suggest that these structural elements coordinate the ligand binding and catalytic reaction. Opening/closing of the ligand-binding valley isolates the catalytic center from the surrounding environment. Interestingly, two aromatic residues, Arg82 and Arg92, appear at the beginning and end of L3 and are assumed to serve as a gate to enable opening and closing of the loop over the binding site.[42] The calculated global motion for the two residues, Arg82 and Arg92, in the holo enzyme, make positively correlated motions. These two residues are in contact with guanine and thus would be expected to move collectively to coordinate loop closure and opening.

Cross-Correlations of Motions

Figure 6 shows the cross-correlations $\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle$ for residue fluctuations in HPPK for the (a) most, (b) second and (c) third slowest modes, and (d) weighted sum of all modes. The two axes refer to residue indices. Blue and red regions refer to negatively and positively correlated regions. Yellow indicates either there is no correlation between residues or they undergo perpendicular motions. Residues with the same color are positively correlated (i.e. they undergo same directional fluctuations), whereas those in two different colors are anticorrelated (i.e. they fluctu-

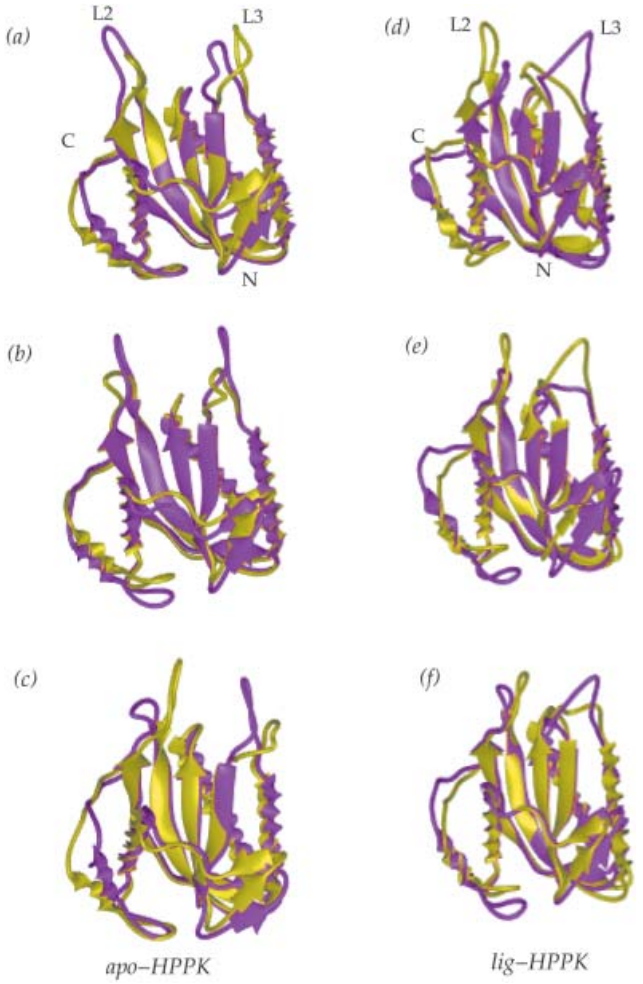

Fig. 5. Mechanisms of the dominant modes of motion. The fluctuating conformations of apo enzyme are shown by the ribbon diagrams in (a–c). The yellow and magenta structures are the two extreme structures with motions for the corresponding mode motions (i.e., the native structure fluctuates between these two structures). (d–f) The most global motions for lig-HPPK liganded with ACP and HP.

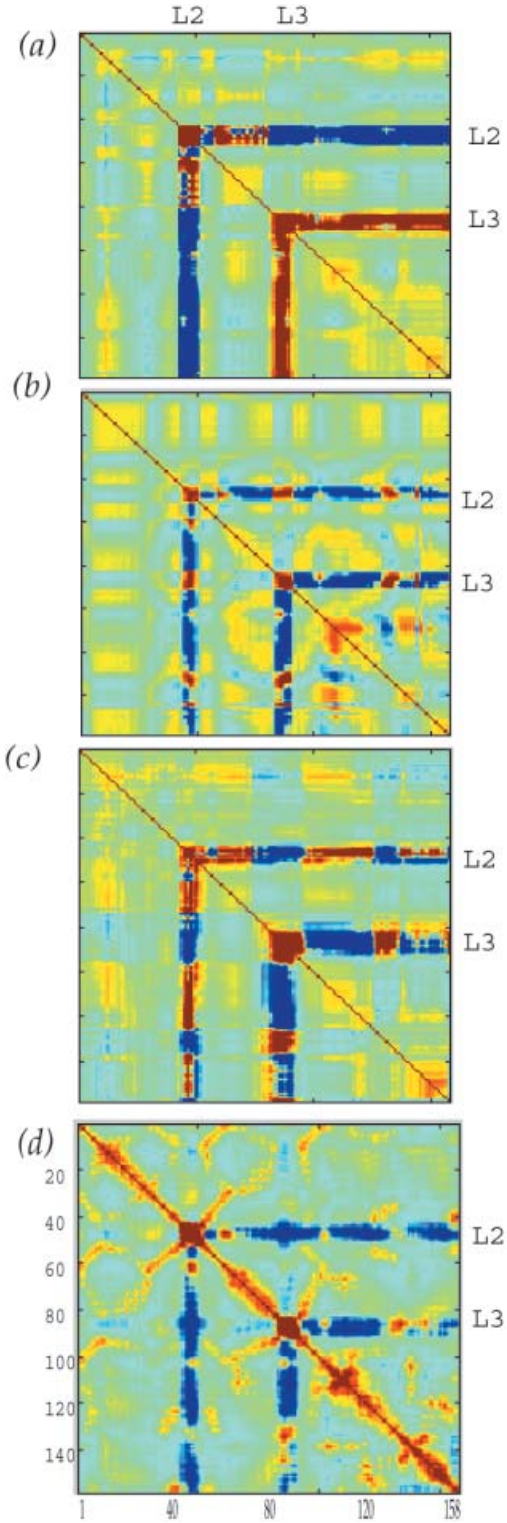

Fig. 6. Correlation maps: cross-correlations $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle$ for residue fluctuations in HPPK for the (a) first, (b) second, (c) third, and (d) overall modes. The two axes refer to residue indices. Blue and red represent the anticorrelated and correlated motions of residue pairs.

ate in opposite directions). In the first mode [Fig. 6(a)], L2 (43–53) undergoes anticorrelated motions with L3 (82–93) and makes positively correlated motions with the helix α3 and the loop following that helix. L3 is positively correlated with the proceeding of the molecule (i.e., residues 94–158). In Figure 6(b) L2 and L3 have positively correlated motions, together with strand β3, the end of helix α3, and the beginning of the loop after that helix. Peptide segment (100–120) undergoes highly correlated motions with the loop and helix α4 (133–149). In the third mode [Fig. 6(c)], L2 and L3 are again anticorrelated. The overall correlation map displays the combined effects of all these three modes of motion to show that L2 and L3 make negatively correlated motions, whereas the remaining parts are almost rigid. A comparison of the correlation maps of the four structures shows that the correlation maps of the bin and pro structures are similar, as their 3D structures are similar. The main difference between their maps and apo form's map is that there is a highly correlated motion of L3 with peptides 100–110 and 130–145 in the bin and pro forms (not shown).

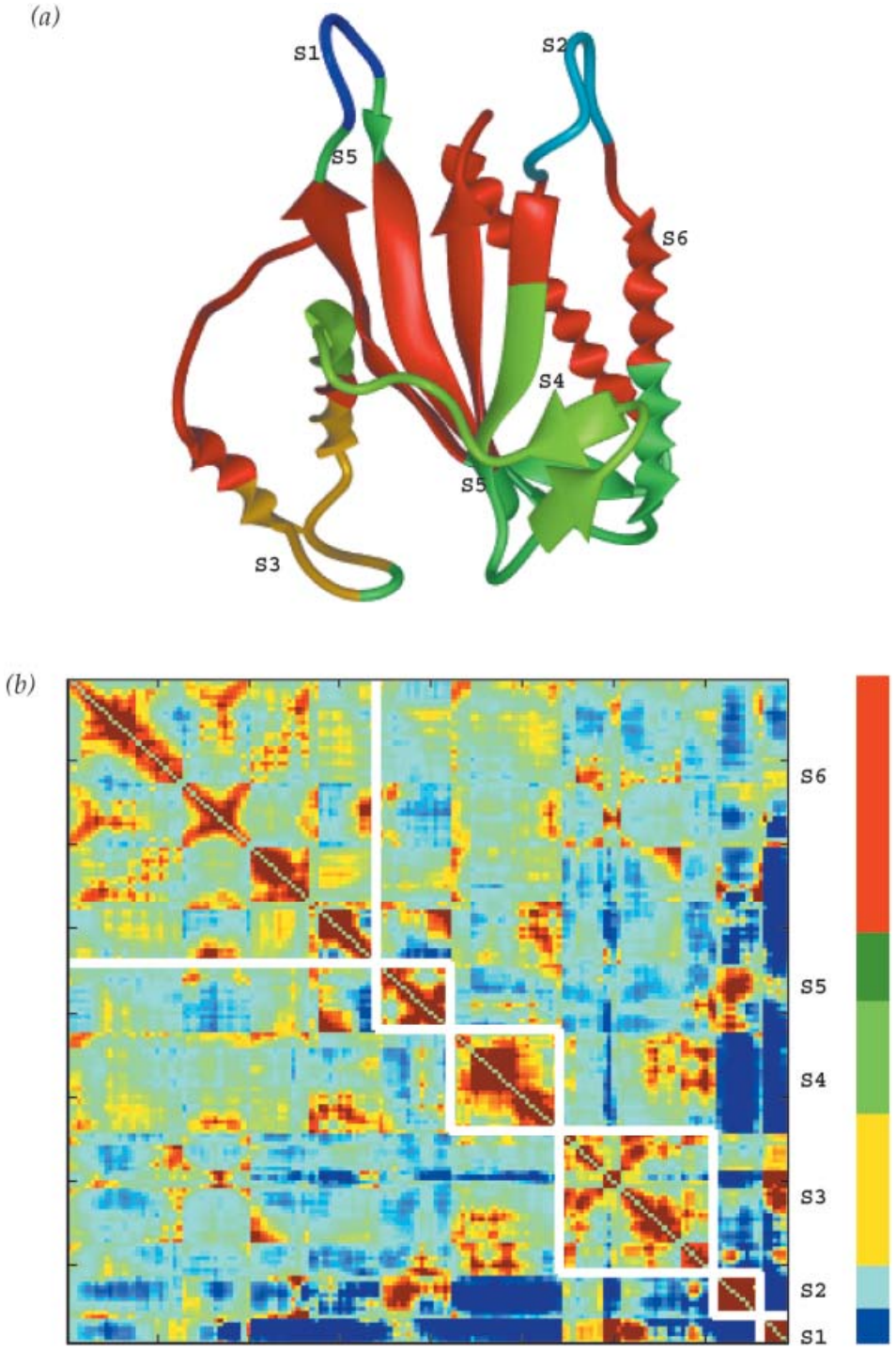

Fig. 7. (a) Ribbon diagram illustrating the correlated motions within the apo enzyme. Six domains of motion are colored differently. The core of the enzyme is almost rigid (red). The six domains were obtained by hierarchical clustering of the residue correlation map shown in Figure 6(d). The reordered correlation map is shown in (b). The domain boundaries are separated by white lines. The bar on the right hand side of the map identifies the coloring of the HPPK ribbon shown in (a) and shows the domain labels S1 to S6.

We have clustered the correlation map in Figure 6(d) in order to identify the residues with the most similar pattern of motion. Residues with similar patterns (although they might be sequentially apart) exhibit collective motions. Average linkage hierarchical cluster analysis was performed on our correlation matrix, and the results are presented in Figure 7. The ribbon diagram in Figure 7(a) identifies by color residues within the six different

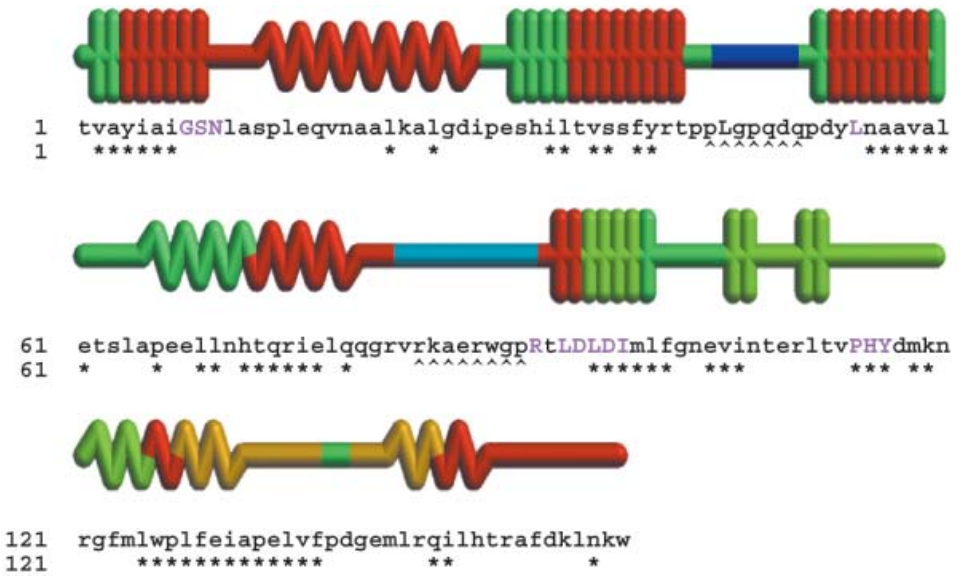

Fig. 8. Secondary structure of HPPK sequence colored according to its six domains of motion. The symbols (*) and (^), below the residue names, represent the mostly constrained and flexible regions of the sequence, respectively, displayed with the mean square fluctuations as shown in the curves of Figure 3(a). Highly conserved residues within a family of 30 HPPK sequences are colored magenta in capital letters. The domain colors are the same as in Figure 7.

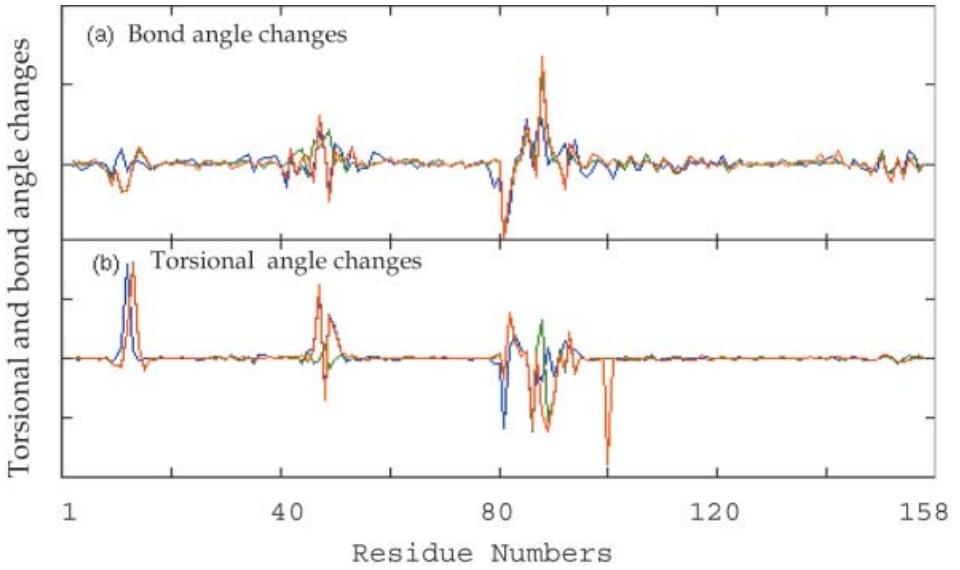

Fig. 9. (a,b) Change in virtual bond angles and virtual torsional angles between the apo and holo structures. Virtual bond and torsional angles are defined by the successive C^{α} atoms along the backbone. Bond angles range between 0° and 180°, and torsional angles range between 0° and 360°.

groups (motion domains), while preserving the HPPK view displayed in Figure 1. Figure 7(b) displays the correlation map shown in Figure 6(d) but reordered to reflect the cluster memberships for the six domains. These domain descriptions divide the HPPK and the companion correlation map into distinct regions according to their collective motions. The bar at the right of the correlation map is colored according to the domains identified in the ribbon structure. The two smallest domains, S1 (dark blue) and S2 (blue) consist of L2 and L3, respectively, and appear at the lower left corner of the re-organized correlation map. These two domains exhibit strongly negatively correlated motions between themselves. The largest domain, S6 (red), consists of residues at the core, and appears at the upper left of the correlation map. Because this domain is a rigid domain, it shows no correlation with other domains or within itself (as indicated by the yellow color in the map). This domain consists of the binding sites for the both ligands.

The secondary structures together with their associated amino acid sequence are shown in Figure 8. The secondary structures are colored in accordance to the domains (groups)

in the previous figure. The location of secondary structure elements along the sequence finds a very close correspondence between domain boundaries, as indicated by a change in color along the sequence. Further inspection of Figure 8 reveals the previously mentioned close relationship between secondary structure and regions of minimal residue fluctuation. The symbols (*) and (^) below the sequence in this figure represent the least and most fluctuating regions, respectively, based on our analysis [results presented in Fig. 3(a)]. All secondary structures have a portion of their secondary structure associated with fluctuation minima. Conversely, loops L2 and L3 correspond to regions of high fluctuation.

To explore any apparent relationships between HPPK sequence and either its secondary structure elements or our domain boundaries, sequence analysis was performed on the HPPK family. A FASTA query of the nonredundant database of protein sequences (Wisconsin Package) identified 30 different sequences. Multiple alignments of these sequences were conducted using the GCG package.[43,44] The most highly conserved sequence positions are identified as magenta colored capital letters along the sequences shown in the figure. The majority of conserved regions are found at nucleotide and pterin binding sites, all of which appear at or near fluctuation minima. Only one residue Arg92 does not correspond to a minimal fluctuating region, and that residue plays a pivotal role in loop closure.

Comparison of the binding sites in folate biosynthesis pathway enzymes

Two enzymes upstream to HPPK in the folate biosynthesis pathway are dihydropteroate synthase (DHPS) and dihydrofolate reductase (DHFR; see Fig. 1 of Xiao et al.[30]). Examinations of these enzymes reveal that folate analogs are surrounded by hydrophobic residues. Although the substrates of these enzymes, namely pterin analogs, share similar core structures (the double hexane rings; 6-hydroxymethyl-7,8-dihydropterin), the phosphate binding sites of the enzymes are significantly different in HPPK, DHPS, and DHFR. This fact may suggest that the side chains and phosphates attached to the pterin group may induce significant changes in these enzymes' substrate binding sites or they might require different conformations for proper binding. On the other hand, active sites in HPPKs, DHPSs and DHFRs are conserved within their own enzyme family (i.e., the active site residues of HPPKs are well conserved), consistent with the results presented above.

Comparison of the global mode motions with the conformational changes upon ligand binding

Normal mode analysis has been used to study the collective motions of biological macromolecules. These studies find that normal modes of several proteins are strongly correlated with the large amplitude conformational changes with ligand binding.[45–47] Here, the root mean square of the atomic fluctuations and motions, obtained by our method, are compared with the experimentally known conformational changes. C^{α} atomic coordinates for the apo and pro crystallographic structures are available in the protein data bank. The two structures are superimposed to calculate the individual residue displacements ($\Delta\mathbf{R} = \mathbf{R}_A - \mathbf{R}_H$), where $\mathbf{R}_A - \mathbf{R}_H$ are the crystallographic coordinates of the apo and holo structures, respectively and $\Delta\mathbf{R}$ is the $3 \times N$ array of coordinate changes in x, y, and z directions for N residues. The theoretical atomic fluctuations are obtained with Equation 7 and the normal mode eigenvectors in the U^T matrix of Equation 8 and used in the analysis of the apo structure of HPPK, as explained in the Methods sections. Similarly, instead of comparing the Cartesian coordinate changes, we also identify the changes in internal bond and torsional angles between the virtual bonds connecting sequential C^{α}s.

Figure 9(a,b) plots the differences in bond and torsional angles for apo and holo HPPK structures upon substrate binding. The three curves in these panels correspond to the differences between the apo structure and the lig, bin, and pro structures, respectively. It is clearly seen that the greatest conformational changes appear for loops L1, L2, and L3, whereas the rest of the molecule shares almost the same internal torsional and bond angles.

The results of the observed and calculated Cartesian displacements are presented in Figure 10(a–c), where the dashed curves represent the experimental relative displacements of C^{α} atoms of the apo and holo forms: lig, bin, and pro HPPKs. The solid curves are obtained from the weighted average of the mean square atomic fluctuations of $3N - 6$ modes of ANM calculations used in this study [see Fig. 3(a)]. Note that both the observed and calculated displacements are normalized in these figures. The correlation coefficients between the observed displacements and theoretical fluctuations are given for each case on the top-right corner of these plots. In three of the cases the correlation coefficients range between 0.71 and 0.86. These high correlations suggest that the information obtained in our calculations may be used to generate hypotheses about conformational motions. Further studies will be required to determine whether individual mode motions derived from selected eigenvectors might be representative of actual displacements upon ligand binding.

We also compared the individual mode motions with the observed displacements upon ligand binding. A summary of these comparisons is presented in Table II. Here, the first column provides the mode numbers, and only results for the first three modes are presented. These three slowest modes are account for nearly 85% of the total calculated motion. In Table II, the second to fourth columns are the correlation coefficients between the directional preferences—simply the corresponding eigenvectors of the corresponding global mode—calculated with our method for the apo HPPK structure and those of the directional displacement changes upon AMP and HPP binding calculated as distances between crystallographic coordinates. The numbers in the parentheses give the correlation coefficients between the mean square residue fluctuations (from Eq. 10), for the first three global modes with those of the observed displacements upon ligand binding. These results find a very high correspondence

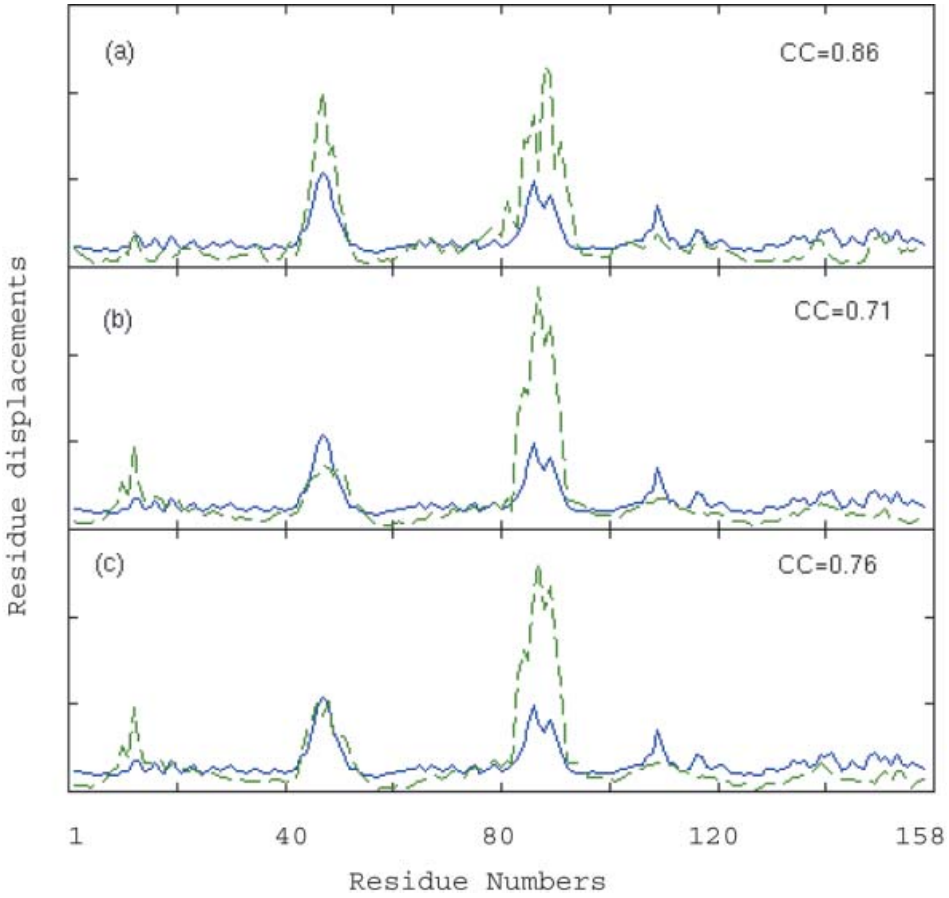

Fig. 10. (a–c) Relative displacements of the C^{α} atoms of the HPPK as observed upon different substrate bindings (solid curve). These are simply the difference of superimposed crystallographic coordinates between the apo-HPPK and lig-, bin-, and pro-HPPK structures, respectively. The dashed curve is the theoretical mean square fluctuations obtained from Equation 7 for all vibrational modes. The areas under the curves are normalized.

TABLE II. Correlation Coefficients between Experimental Displacement Vectors and Theoretical Eigenvectors and between Experimental Conformational Changes and Theoretical Root Mean Square Fluctuations

Mode #	Apo-lig	Apo-bin	Apo-prod
1	0.23 (0.86)	0.45 (0.66)	0.47 (0.72)
2	0.56 (0.86)	0.37 (0.83)	0.33 (0.85)
3	0.09 (0.91)	0.48 (0.83)	0.49 (0.87)

between the experimental X-ray conformational transitions and our theoretical values. These results suggest that our method may be used to predict the directions of motions in apo HPPK when bound by different ligands. Note that transitions from apo to then lig, bin, and pro conformations are highly represented by the first, second, and third eigenvectors, respectively ($CC = 0.56$, 0.48, and 0.49). The second case—the comparison of simply the mean square fluctuations with the distance traveled—introduces higher correlations (the worst case is 0.66). Speculation from these results suggests that the holo conformations of the enzyme are determined by the specific ligands bound to the apo structure, and this holo form is among the possible calculated conformations that the apo structure can assume. Apparently the ligand interactions drive the enzyme's structure to a new, stable, conformation among these possible conformations.

Pathways from apo to holo forms

It is useful to compare the amplitudes of motions as determined from the normal modes of vibration with the conformational changes upon ligand binding. To do this, a sample of snapshots is created that depicts the molecule's motion from its starting (i.e., apo) configuration. For the Cartesian basis, the snapshots may be obtained by the formula:

$$R(s) = R_A + \sum \alpha_m u_m / \lambda_m \cos(\lambda_m^{1/2} t + \delta_m) \qquad (12)$$

where R_A is the starting configuration, u_m and λ_m are the mth eigenvector and the associated eigenvalue of the Hessian matrix, respectively. The summation is carried over the collection of global modes, $R(s)$ is the coordinate array of the snapshots, and α_m is the desired amplitude of motion. The amplitude of motion is selected by simply normalizing the sum of displacements, $|\mathbf{R}_A - \mathbf{R}_H|$, from

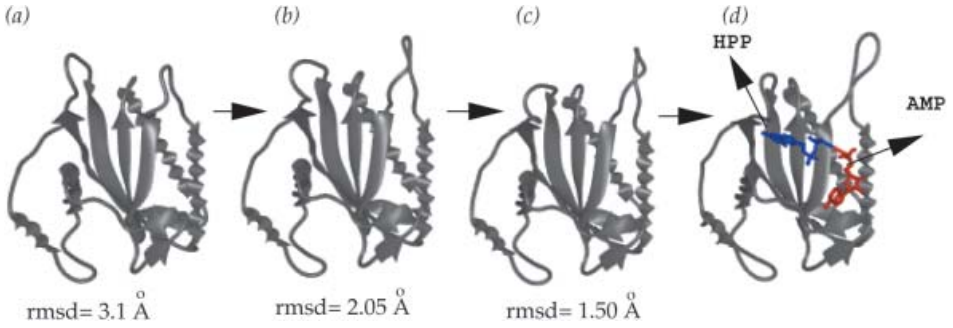

Fig. 11. The ribbon diagrams of (a) apo structure (starting structure), (b,c) two intermediate snapshots, and (d) the binary structure (targeted structure). The intermediate structures are the most similar structures to the pro-HPPK among the first and second set of snapshots. L3 first opens up to provide probably an available space for AMP binding and this introduces a conformation change on L2.

apo to holo forms by the root mean square fluctuation, $\langle \Delta R^2 \rangle^{1/2}$, (i.e., $\alpha = \Sigma_i(|\mathbf{R}_A - \mathbf{R}_H|)_i / \langle \Delta R^2 \rangle_i^{1/2}$). The denominator of this ratio is obtained from Equation 7, and the summation is carried over all residues $1 \leq i \leq N$. δ_m is an arbitrary phase between modes, and we select phases of 0°, 90, and 180° to determine different combinations of modes. Assuming $\lambda_m^{1/2} t$ is very close to zero, this phase difference introduces a sign of ± 1 in the summation.

We tested our method for predicting the conformation of the pro form starting from the apo structure. The rmsd between the apo HPPK and prod was initially 3.13 Å. Using the first three eigenvectors and all combinations of first three eigenvectors, 27 snapshots were generated using Equation 12. The rmsd between the snapshots and the actual pro structure varied between 2.05 and 4.48 Å. Selecting the 2.05-Å snapshot as the next step in the procedure, a new set of snapshots were generated, the best of which had a rmsd of 1.50 Å from the correct structure. The ribbon diagrams of these two intermediate snapshots are shown in Figure 11. These snapshots identify a probable pathway between the starting apo structure and the final bound structures. As a proof of concept, these results suggest that the eigenvectors can be used to drive the apo structure to its holo form, with the caveat that this two-step iteration reflects a strong bias by selecting the "best" rmsd intermediate structure to begin the second iteration. Repeats of this same calculation, based on random eigen inputs, found no conformations similar to the above endpoint.

Our analysis was also repeated on a coarser model based only on the six functional domains as identified above. Here we assume that these domains move as blocks and use the block average eigenvector, assigned to its geometric center, to determine motion. Averages were also determined for these six domains from the crystallographic translations between the apo and holo forms; that is, the residues in the S1 domain of Figure 7 are assigned a center of mass, an average eigenvector, and an average conformational change vector. Figure 12 gives the schematic representations of these blocks and associated eigenvectors and conformational change vectors. In this figure the apo enzyme is colored according to six domains (see Fig. 7). The gray structure is the target structure with AMP and HPP.

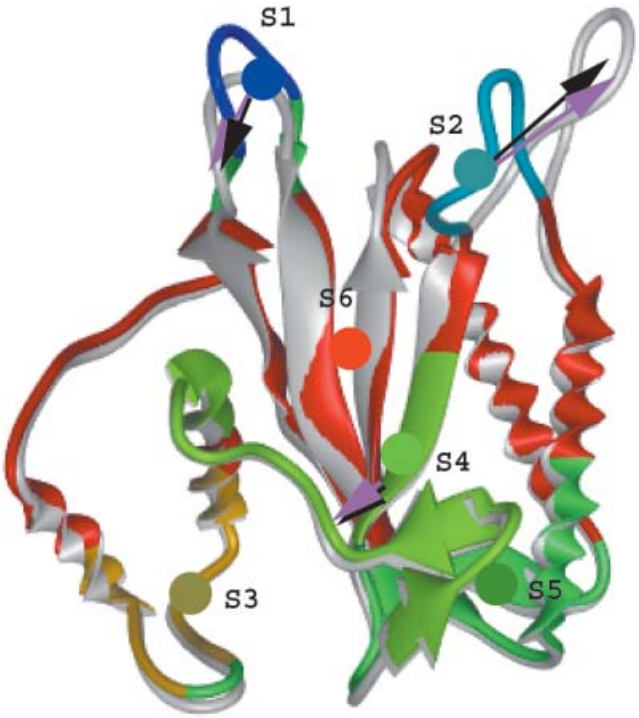

Fig. 12. Schematic representations of the blocks and associated eigenvectors and conformational change vectors. In this figure the apo enzyme is colored according to six domains (see Fig. 7). The gray structure is the target structure with AMP and HPP. The center of mass of each block (domain) is displayed with a circular dot colored according to its associated domain color. The black arrows show the experimental direction of the change in the center of masses from apo to holo form. The magenta arrows give the theoretical weighted average of eigenvectors.

The center of mass of each block (domain) is displayed with a circular dot colored according to its associated domain color. The black arrows show the direction of change in the center of mass from the apo to holo forms as determined from the crystallographic coordinates. The magenta arrows give the weighted average of eigenvectors calculated as $v_{\text{avg,block}} = \Sigma u_{m,\text{block}} / \lambda_{m,\text{block}}$, where m is the first three eigenmodes ($1 \leq m \leq 3$) and block means that the average is taken over the residues in a specific block ($1 \leq$ block ≤ 6). The correlation coefficients between these two

TABLE III. Correlation Coefficients between Experimental and Theoretical Values of Distance upon Conformational Change by Using Six Functional Domains

Mode	Apo-lig	Apo-bin	Apo-prod
1	0.58	0.48	0.55
2	0.38	0.29	0.53
3	0.38	0.75	0.76

sets of vectors did not worsen but improved when compared to the results of the full residue calculations.

Table III gives the correlation coefficients for the six domain coarse model, using the same table format as was used in Table II. We can now solve the analytical problem of finding the coefficients of eigenvectors that give the exact coordinates of these six domains in the holo (pro) form starting from the apo form. The problem can be formulated as ($\Delta\mathbf{R}_{18\times 1} = \mathbf{W}_{18\times 3} \times \mathbf{C}_{3\times 1}$), where $\Delta\mathbf{R}_{18\times 1} = \mathbf{R}_A - \mathbf{R}_H$ is the change in displacement between the centers of mass of the six functional domains of apo and holo structures ($18 = 3 \times 6$) of S1–S6, **W** is the matrix consisting of three slowest modes' eigenvectors for these domains, and **C** is the column of coefficients for these eigenvectors. The solution of this equation gives the elements of the **C** array as $c_1 = -17$, $c_2 = 18$, and $c_3 = -25$. Multiplying these coefficients by the first three eigenvectors yields the exact displacements needed to reach the pro state starting from apo structure. In Table III, note that the third eigenvector is 0.76, correlated with the correct case, and has the highest contribution for this transition. Using these same coefficients for N residues yields a conformation only 1.81 Å from the final structure. Taken together these results suggest that substantial information is contained within the slowest global modes for identifying a subset of protein motions.

CONCLUSIONS

We draw the following conclusions from our results. A good agreement is obtained between the mean square fluctuations and the experimental crystallographic temperature factors for the four conformations of HPPK (one apo and three holo forms with different ligands). The most flexible regions correspond to loops L2 and L3, which indeed exhibit the largest conformational changes upon ligand binding. Different mechanisms of motion for these loops provides a mechanism for conformational changes to the unbound to bound transition. The extension/stretching and opening/closing motions of these loops drive the conformational changes. It is known that ATP is the energy source for activating the kinases for their enzymatic functions. The binary complex of HPPK with ADP displays the largest conformational change for the apo form. Binding of the nucleotide changes the structural properties of the HPPK to provide a more open conformation for the other substrate HP to bind. When the HP binds the loop closes and the kinase becomes fairly rigid: probably to provide an isolated region for the transfer of phosphor in the catalytic reaction. It is also observed that the palm-like region of the enzyme is almost rigid in four of the conformations. This is an interesting observation because usually global modes yield movements of large domains and hinge motions. The equilibrium cross-correlations between residue pairs also identify the units that are correlated. The two loops exhibit anticorrelated motions with respect to each other in apo form, and the rigid part does not show any correlation with the rest of the molecule. Examination of other enzymes in the folate pathway gives no significant signature for the folate binding regions. This finding may support the hypothesis that the same functionality does not necessarily need the same motif in different proteins.

Comparison of the global mode motions with the conformational changes upon ligand binding suggests a high correspondence between the normal mode directions derived from ANM calculations and the actual conformational changes. These high correlations implicate our calculational method for further study of modal motions. Indeed, our results confirm that the individual eigenvectors might be useful to drive the apo structure toward holo structures without resorting to MD calculations. Thus, the apo structure can assume a set of conformations driven by the slowest modes, and ligand binding appears to introduce a new stable structure from this set of accessible conformations. Further investigation of the motions of HPPK's six functional domains, is consistent with our observation that these domains move as blocks.

REFERENCES

1. Berne BJ, Straub JE. Novel methods of sampling phase space in the simulation of biological systems. Curr Opin Struct Biol 1997;7:181–189.
2. Brooks BR, Janezic D, Karplus M. Harmonic analysis of large systems: 1. Methodology. J Comput Chem 1995;16:1522–1542.
3. Elber R. Novel methods for molecular dynamics simulations. Curr Opin Struct Biol 1996;6:232–235.
4. Kazmirski SL, Daggett V. Non-native interactions in protein folding intermediates: molecular dynamics simulations of hen lysozyme. J Mol Biol 1998;284:793–806.
5. Kazmirski SL, Wong KB, Freund SM, Tan YJ, Fersht AR, Daggett V. Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. Proc Natl Acad Sci USA 2001;98:4349–4354.
6. Gibson QH, Regan R, Elber R, Olson JS, Carver TE. Distal pocket residues affect picosecond ligand recombination in myoglobin. An experimental and molecular dynamics study of position 29 mutants. J Biol Chem 1992;267:22022–22034.
7. Ma J, Karplus M. Ligand-induced conformational changes in ras p21: a normal mode and energy minimization analysis. J Mol Biol 1997;274:114–131.
8. Ma J, Sigler PB, Xu Z, Karplus M. A dynamic model for the allosteric mechanism of GroEL. J Mol Biol 2000;302:303–313.
9. Paci E, Karplus M. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. J Mol Biol 1999;288:441–459.
10. Rick SW, Erickson JW, Burt SK. Reaction path and free energy calculations of the transition between alternate conformations of HIV-1 protease. Proteins 1998;32:7–16.
11. Karplus M, McCammon JA. Dynamics of proteins: elements and function. Annu Rev Biochem 1983;52:263–300.
12. Pauling L. Nature of forces between large molecules of biological interest. Nature 1948;161:707–709.
13. Hage W, Kim M, Frei H, Mathies RA. Protein dynamics in the Bacteriorhodopsin photocycle: a nanosecond step-scan FT-IR investigation of the KL to L transition. J Phys Chem 1996;100:16026–16033.
14. Nicholson LK, Yamazaki T, Torchia DA, Grzesiek S, Bax A, Stahl SJ, Kaufman JD, Wingfield PT, Lam PY, Jadhav PK, Hodge CN,

Domaille PJ, Chang CH. Flexibility and function in HIV-1 protease. Nat Struct Biol 1995;2:274–280.
15. Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. Science 1991;254:1598–1603.
16. Olender R, Elber R. Calculation of classical trajectories with a very large time step: formalism and numerical examples. J Chem Phys 1996;105:9299–9315.
17. Schlitter J, Engles M, Kruger P, Jacoby EU, Wollmer A. Targeted molecular dynamics simulation of conformational change: application to the T-R transition in insulin. Mol Sim 1983;10:291–308.
18. Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. Proteins 1993;17:412–425.
19. Bahar I, Jernigan RL. Vibrational dynamics of transfer RNAs: comparison of the free and synthetase-bound forms. J Mol Biol 1998;281:871–884.
20. Hinsen K, Thomas A, Field MJ. Analysis of domain motions in large proteins. Proteins 1999;34:369–382.
21. Keskin O, Jernigan RL, Bahar I. Proteins with similar architecture exhibit similar large-scale dynamic behavior. Biophys J 2000;78:2093–2106.
22. Bahar I, Jernigan RL. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. J Mol Biol 1997;266:195–214.
23. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515.
24. Keskin O, Durell SR, Bahar I, Jernigan RL, Covell DG. Relating molecular flexibility to function: a case study of tubulin. Biophys J 2002;83:663–680.
25. Keskin O, Bahar I, Flatow D, Covell DG, Jernigan RL. Molecular mechanisms of chaperonin GroEL-GroES function. Biochemistry 2002;41:491–501.
26. 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.
27. Haliloglu T, Bahar I. Coarse-grained simulations of conformational dynamics of proteins: application to apomyoglobin. Proteins 1998;31:271–281.
28. Bahar I, Wallqvist A, Covell DG, Jernigan RL. Correlation between native-state hydrogen exchange and cooperative residue fluctuations from a simple model. Biochemistry 1998;37:1067–1075.
29. Haliloglu T, Bahar I. Structure-based analysis of protein dynamics: comparison of theoretical results for hen lysozyme with X-ray diffraction and NMR relaxation data. Proteins 1999;37:654–667.
30. Xiao B, Shi G, Chen X, Yan H, Ji X. Crystal structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, a potential target for the development of novel antimicrobial agents. Struct Fold Des 1999;7:489–496.
31. Kovacs JA, Powell F, Voeller D, Allegra CJ. Inhibition of *Pneumocystis carinii* dihydropteroate synthetase by para-acetamidobenzoic acid: possible mechanism of action of isoprinosine in human immunodeficiency virus infection. Antimicrob Agents Chemother 1993;37:1227–1231.
32. Chabner BA, Longo DL, editors. Cancer chemotherapy and biotherapy: principles and practice. Philadelphia, PA: Lippincot-Raven; 1996.
33. Shi G, Gao J, Yan H. 1H, 13C and 15N resonance assignments of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase and its complex with MgAMPPCP. J BIomol NMR 1999;14:189–190.
34. Hogue CW, Ohkawa H, Bryant SH. A dynamic look at structures: WWW-Entrez and the Molecular Modeling Database. Trends Biochem Sci 1996;21:226–229.
35. Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.
36. Case DA. Normal mode analysis of protein dynamics. Curr Opin Struct Biol 1994;4:285–290.
37. Doruker P, Atilgan AR, Bahar I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. Proteins 2000;40:512–524.
38. Hayward S, Kitao A, Berendsen HJ. Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. Proteins 1997;27:425–437.
39. Kitao A, Go N. Investigating protein dynamics in collective coordinate space. Curr Opin Struct Biol 1999;9:164–169.
40. Flory PJ. Statistical thermodynamics of random networks. Proc R Soc London, Ser A 1976;351:351–380.
41. Haliloglu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. Phys Rev Lett 1997;79:3090–3093.
42. Blaszczyk J, Shi G, Yan H, Ji X. Catalytic center assembly of HPPK as revealed by the crystal structure of a ternary complex at 1.25 Å resolution. Structure Fold Des 2000;8:1049–1058.
43. Wisconsin Package, Genetic Computer Group (GCG). [Version 9.0]. 1998. Madison, Wisconsin.
44. Senger M, Flores T, Glatting K, Ernst P, Hotz-Wagenblatt A, Suhai S. W2H: WWW interface to the GCG sequence analysis package. Bioinformatics 1998;14:452–457.
45. 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.
46. Marques O, Sanejouand YH. Hinge-bending motion in citrate synthase arising from normal mode calculations. Proteins 1995;23:557–560.
47. Tama F, Gadea FX, Marques O, Sanejouand YH. Building-block approach for determining low-frequency normal modes of macromolecules. Proteins 2000;41:1–7.