Cooperative fluctuations of unliganded and substrate-bound HIV-1 protease: A structure-based analysis on a variety of conformations from crystallography and molecular dynamics simu…

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · MAY 2003

Impact Factor: 2.63 · DOI: 10.1002/prot.10350 · Source: PubMed

4 AUTHORS, INCLUDING:

Nese Kurt Yilmaz
University of Massachusetts Medical School
31 PUBLICATIONS **251** CITATIONS
SEE PROFILE

Celia A Schiffer
University of Massachusetts Medical School
148 PUBLICATIONS **3,863** CITATIONS
SEE PROFILE

Turkan Haliloglu
Bogazici University
107 PUBLICATIONS **1,829** CITATIONS
SEE PROFILE

PROTEINS: Structure, Function, and Genetics 51:409–422 (2003)

Cooperative Fluctuations of Unliganded and Substrate-Bound HIV-1 Protease: A Structure-Based Analysis on a Variety of Conformations From Crystallography and Molecular Dynamics Simulations

Nese Kurt,[1] Walter R.P. Scott,[2] Celia A. Schiffer,[3*] and Turkan Haliloglu[1]
[1]*Polymer Research Center and Chemical Engineering Department, Bogazici University, Istanbul, Turkey*
[2]*Oxford, United Kingdom*
[3]*Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts*

ABSTRACT **The dynamics of HIV-1 protease, both in unliganded and substrate-bound forms have been analyzed by using an analytical method, Gaussian network model (GNM). The method is applied to different conformations accessible to the protein backbone in the native state, observed in crystal structures and snapshots from fully atomistic molecular dynamics (MD) simulation trajectories. The modes of motion obtained from GNM on different conformations of HIV-1 protease are conserved throughout the MD simulations. The flaps and 40's loop of the unliganded HIV-1 protease structure are identified as the most mobile regions. However, in the liganded structure these flaps lose mobility, and terminal regions of the monomers become more flexible. Analysis of the fast modes shows that residues important for stability are in the same regions of all the structures examined. Among these, Gly86 appears to be a key residue for stability. The contribution of residues in the active site region and flaps to the stability is more pronounced in the substrate-bound form than in the unliganded form. The convergence of modes in GNM to similar regions of HIV-1 protease, regardless of the conformation of the protein, supports the robustness of GNM as a potentially useful and predictive tool. Proteins 2003;51:409–422.**

Key words: Gaussian network model; molecular dynamics; principal modes; global motions; correlations; residue fluctuations; mobility; functional motions

INTRODUCTION

HIV-1 protease is an essential enzyme for the replication of HIV, the virus that causes AIDS.[1] It is a homodimeric aspartyl protease made up of two identical subunits of 99 residues each. The crystal structure of unliganded HIV-1 protease is completely symmetric, with its flap extended.[2] As with most proteins, HIV-1 protease can adopt many different conformations[3,4]; the extended conformation of flaps in the crystal structure[2] is likely due to crystal packing effects, as the tips of the flaps hydrogen bond to a symmetry-related molecule in the crystal lattice.[5] NMR relaxation studies have shown the flap tips to be highly mobile in solution[6] adopting a variety of conformations in solution from fully closed to open.[7] They also likely adopt different conformations in an asymmetric fashion as suggested from solvated fully atomistic MD simulations.[3] To bind asymmetric substrates, the symmetric HIV-1 protease adopts an asymmetric conformation.[8,9] Therefore, the analysis of structural fluctuations in HIV-1 protease can reveal valuable information on the dynamics of motion related to the functioning of this enzyme, which is critical in designing the next generation of drugs against AIDS.

Recently, a structure-based analytical model was developed to describe the dynamics of proteins, referred to as the Gaussian Network Model (GNM). GNM has been applied to a number of different biomolecular systems including RNA complexes,[10] enzyme complexes,[11] substrate-binding proteins,[12] and monomeric proteins,[13] and the predictions compare favorably with crystallographic temperature factors,[14] H/D exchange behavior,[15] and order parameters from NMR relaxation measurements.[16] It has also been shown to be useful in identifying functional motions of proteins.[17,18] GNM uses the known topology of protein-protein contacts to model the protein as an elastic network with uniform single-parameter harmonic potentials between the α-carbons of residue pairs in contact. Adopting harmonic interaction potentials means assum-

Grant sponsor: BU Research Funds; Grant numbers: 00HA502D-00HA503; Grant sponsor: DPT Project; Grant number: 01K120280; Grant sponsor: Turkish Academy of Sciences.

N. Kurt's present address is Department of Chemistry, University of Wisconsin-Madison, Madison, WI 53706.

W. R. P. Scott's present address is Department of Chemistry, University of British Columbia, Vancouver, B. C., Canada V6T 1Z1.

C.A.S. and T.H. contributed equally and are joint last authors.

*Correspondence to: Celia Schiffer, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation St. LRB, Worcester, MA 01605. E-mail: Celia.Schiffer@umassmed.edu

ing that the residues are undergoing Gaussian fluctuations about their mean positions.

With use of GNM, the dynamics of a biomolecular system can be decomposed into a collection of internal modes at different frequencies as in normal mode analysis (NMA),[19,20] yet without the need of energy minimization or inclusion of specific energy parameters. Instead, an analytical solution that is deterministic for a given molecular architecture is quickly calculated, but with the expense of losing some atomic details.[18] The slowest modes, those with the lowest frequencies, represent the most cooperative motions involving the overall structure. These dominant modes of motion give information about the molecular dynamics relevant to biological function occurring on a global scale.[5,21–24] The fastest modes reflect localized motions involving high-frequency fluctuations of individual residues. These residues are generally in dense regions of the structure with high coordination numbers, and they are potentially important for the protein's structural stability.

Details of molecular motion can be obtained from conventional molecular dynamics (MD) simulations using empirical potentials. In MD simulations, an overwhelming amount of information is contained in the atomic coordinates trajectories, and the significant motions need to be extracted from the data to relate them to function. In determining the dominant modes from the fluctuations seen in MD simulations, Principal Component Analysis (PCA) is an established computational means of studying protein dynamics.[25] PCA can be carried out over the MD trajectory for this purpose if the conformational space is sampled efficiently. However, if the sampling is inefficient; PCA falls short because of the finite simulation times.[26–28] The set of modes that capture most of the fluctuations depends on the width of the particular sampling time window. The sampling problem becomes more difficult as the size of protein increases, and the computational time is longer and the particular conformational transitions that operate at the slowest mode regimen are not necessarily reproduced in independent runs of MD.[26,29]

The advantages and limitations of the GNM model in comparison with PCA were discussed in a recent study.[29] The correlation between GNM modes and those determined by PCA of MD trajectories were studied by analyzing the same protein at the same level of resolution (a single site per residue). The major criticism of the GNM is inadequacy to describe anharmonic fluctuations or multimeric transitions driven by the slowest collective modes observed in MD. However, despite neglecting the anharmonic motions, the average properties (mean-square fluctuations and cross correlations) controlled by the low-frequency motions were reproduced by both PCA and GNM, which demonstrated the robustness of these properties and utility of GNM, which is computationally much more efficient.

In this article, we describe how the GNM and MD methods were combined to extensively analyze the behavior of HIV-1 protease in unliganded and substrate-bound forms. In a previous study, the mean-square fluctuations predicted by GNM for the unliganded crystal structure was given by Bahar et al.[24] These theoretical predictions were in reasonable agreement with the crystallographic temperature factors, which supports the applicability of GNM for further analysis of protease dynamics. In all previous GNM studies of different biomolecular systems, the method was applied to only one conformation seen in the crystal structure. This limited application may cause an incomplete or even flawed dynamic analysis when the conformation seen in the crystal structure is affected by crystal packing and the protein samples different conformations in solution. In this study, we applied GNM to several different conformations accessible to HIV-1 protease, as obtained from MD simulations. Similar overall dynamic behaviors were obtained for all conformations examined. This finding supports the reliability of GNM analysis, as well as elucidating the dynamic consistency of configurations in the MD trajectories.

MODELS AND METHODS

Gaussian Network Model (GNM)

GNM theory finds its roots in the elasticity theory of random polymer networks.[30] In this theory, it is assumed that the native-state protein is equivalent to a three-dimensional elastic network. The junctions in this network are the α-carbon atoms, and the interactions between the neighboring residues are represented by harmonic potentials with a uniform spring constant. The residues i and j in the folded protein are assumed to undergo Gaussian fluctuations ΔR_{ij} about their mean positions in the separation $R_{ij} = |R_j - R_i|$.

According to the GNM theory,[14] the equilibrium correlation between the fluctuations $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ of residues i and j is given by

$$< \Delta\mathbf{R_i} \cdot \Delta\mathbf{R_j} > = (3\ k_B\ T/\gamma)\ [\Gamma^{-1}]_{ij} \qquad (1)$$

where Γ is a symmetric matrix known as the Kirchoff or connectivity matrix, $\mathbf{R}_i$ is the position vector of the i^{th} α-carbon, k_B is the Boltzmann constant, T is the absolute temperature. γ is the normalization constant, which is the counterpart of the single parameter of the Hookean pairwise potential originally proposed by Tirion[31] and represents the interresidue interactions in the native state. The mean-square fluctuations of individual residues can be readily found from Eq. 1, taking i = j.

The elements of the Kirchoff matrix are defined as

$$\Gamma_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } R > r_c \\ -\sum_{i=j} \Gamma_{ij} & \text{if } i = j \end{cases} \qquad (2)$$

Here r_c is the cutoff separation defining the range of interaction of nonbonded α-carbons. A reasonable cutoff distance including all residue pairs within a first interaction shell is 7.0 Å.[32,33] The i^{th} diagonal element of **Γ** characterizes the local packing density or the coordination number of residue i.

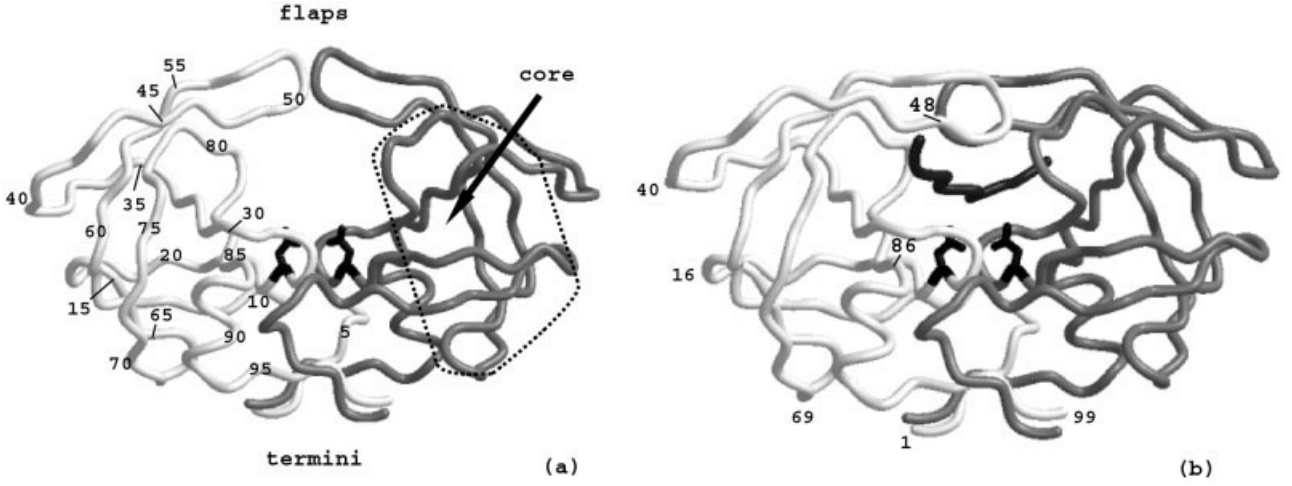

Fig. 1. Ribbon diagrams of (**a**) unliganded and (**b**) liganded HIV-1 protease crystal structure. The two monomers are in white and gray; the substrate in (b) is in dark gray. The active site residue D25 is shown explicitly in black. On one of the monomers, (a) every fifth residue is labeled and (b) residues at the tips of loops and residues 86 and 48 (which are discussed later in the text) are labeled.

The inverse of $\mathbf{\Gamma}$ may be written as

$$\mathbf{\Gamma^{-1}} = \mathbf{U}\,(\mathbf{\Lambda}^{-1})\,\mathbf{U}^{T} \tag{3}$$

where $\mathbf{U}$ is an orthogonal matrix whose columns $\mathbf{u}_i$ are the eigenvectors of $\mathbf{\Gamma}$, and $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues (λ_i) of $\mathbf{\Gamma}$, usually organized in ascending order. Mean-square fluctuations of the α-carbon atoms and the cross correlations between them are given by the respective diagonal and off-diagonal elements of Γ^{-1}.

It is possible to decompose $\mathbf{\Gamma}^{-1}$ into the sum of contributions from individual modes as

$$\Gamma^{-1} = \sum_{k=2}^{N} \lambda_k^{-1}\mathbf{u}_k\mathbf{u}_k^{T} = \sum_{k=2}^{N} \mathbf{A}^{(k)} \tag{4}$$

Here $\mathbf{A}^{(k)}$ is the N × N matrix (for a protein of N residues) describing the contribution of the k^{th} vibrational mode to atomic fluctuations. The first eigenvalue of $\mathbf{\Gamma}$, identically equal to zero, is not included in the above summation. The i^{th} eigenvalue is representative of the frequency of the i^{th} mode of motion (which is $(\gamma\lambda_i)^{1/2}$), and the i^{th} eigenvector gives the shape of this mode as a function of residue index. GNM has verified to provide information on the magnitudes of the fluctuations. All fluctuations are implicitly assumed to be isotropic; no directional preferences are accounted for. Thus, the molecule is viewed as a collection of N sites, one for each residue, resulting in an ensemble of N-1 independent modes, instead of 3N-6 modes, which would be obtained in 3D description.

The factor λ^{-1} scales the contributions of the individual modes. For instance, the correlations conveyed by several modes $k_1 \le k \le k_2$ of interest are given by

$$C_{ij}\,(k_1 \le k \le k_2) \equiv \frac{\sum_{k} \lambda_k^{-1}[\mathbf{u}_k\mathbf{u}_k^{T}]_{ij}}{\left(\sum_{k} \lambda_k^{-1}[\mathbf{u}_k\mathbf{u}_k^{T}]_{ii}\ \sum_{k}\lambda_k^{-1}[\mathbf{u}_k\mathbf{u}_k^{T}]_{jj}\right)^{1/2}} \tag{5}$$

where the summations are over only the modes of interest.

Previous manuscripts where GNM was introduced include additional details of the theory.[13,14,24]

MD Simulations

Solvated MD simulations of HIV-1 protease in unliganded and substrate-bound forms were conducted and reported before with a detailed description of the simulation setup.[3] Here, we used only the α-carbon backbone coordinates of HIV-1 protease from particular time points during the trajectories.

RESULTS

GNM was applied to the crystal structures of the unliganded HIV-1 protease and one of its substrate complexes with codes 1hhp and 1f7a, respectively.[2,8,34] Ribbon diagrams of these crystal structures are given in Figure 1. As representatives of different conformations accessible to the protease in the native state, four snapshots at 1, 2, 5, and 10 ns for the unliganded protease and three snapshots at 1, 2, and 5 ns for the liganded complex were taken from the solvated fully atomistic MD simulation trajectories.[3]

Mean-Square Fluctuations in Atomic Coordinates

The mean-square fluctuations predicted by GNM for the unliganded crystal structure, together with the crystallographic temperature factors, were given by Bahar et al.[24] The force constant γ in Eq. 1 is calculated here as 0.38 (kcal/mol)/A^2 for both unliganded[2] and liganded[8] crystal structures, from the comparison of the respective temperature factors. Here, we present the fluctuations predicted for different conformations of the structure examined [Fig. 2(a)]. General characteristics of fluctuations are conserved such as minima and maxima occurring at the same region although the symmetry of protease is lost. The only significant change is the higher amplitude fluctuations of flaps, with either one monomer or the other being more mobile at different snapshots. The structure of the liganded protease dimer is not symmetric because it adapts to

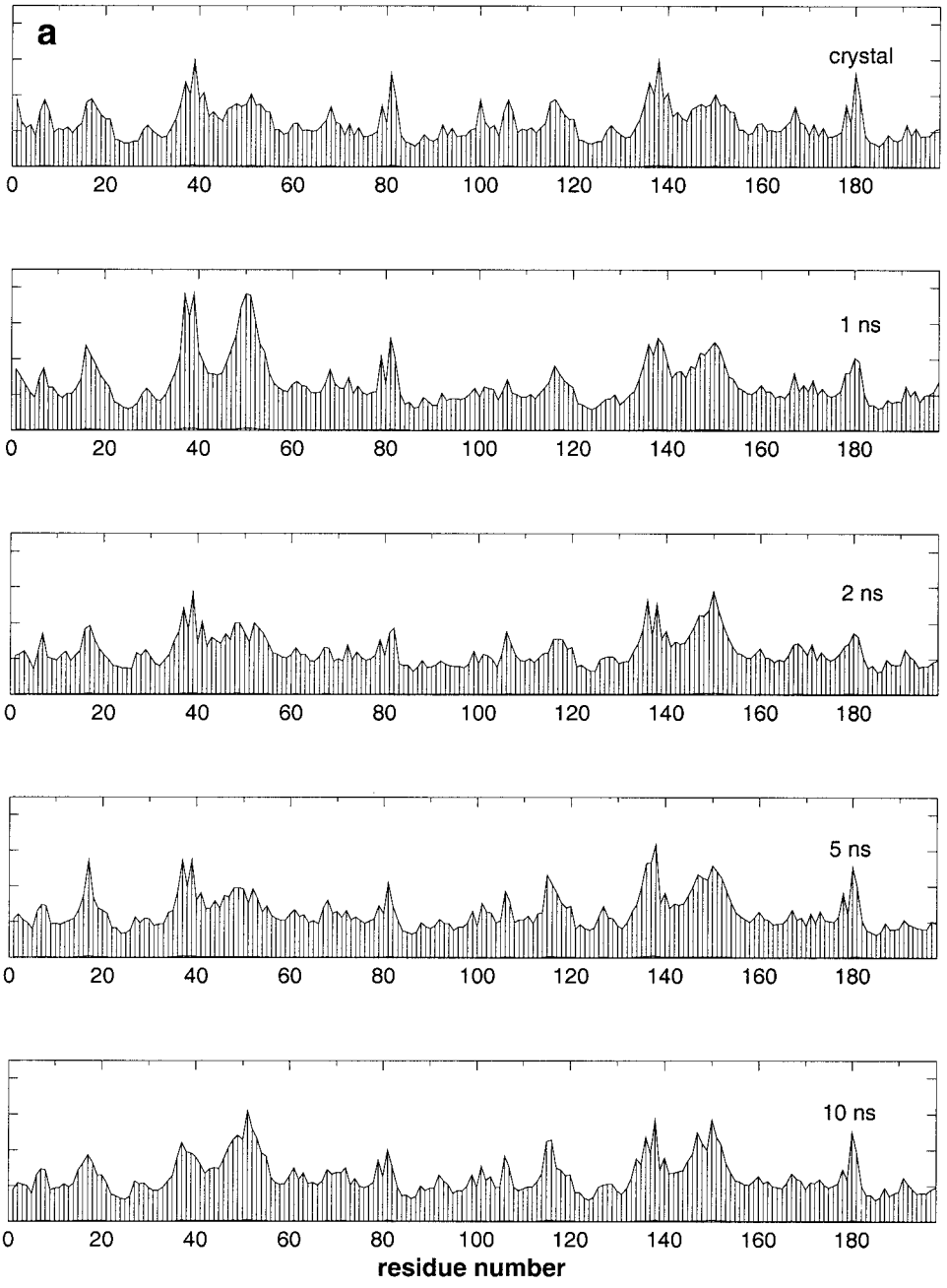

Fig. 2. Mean-square fluctuations of α-carbon atoms for (**a**) unliganded and (**b**) liganded HIV-1 protease predicted by GNM. Each row is for a different conformation of the structure. The results are normalized and in arbitrary units.

bind an asymmetric peptide. However, the fluctuations in the crystal structure are almost the same for the two monomers [Fig. 2(b)]. This symmetric behavior changes for other conformations sampled during the MD simulation; still the general characteristics are conserved. The highest amplitude fluctuations are displayed by residues in the 40's loop, which is the least restricted region of the protease once the substrate is bound. These flexible loops (Fig. 1) are connected to the flaps and, therefore, may have a role in the movement of flaps.[4]

Global Motion Induced by the Principal Modes of Actions

Unliganded protease

The dynamic behavior of a macromolecular system can be described by the superposition of several different modes. The slowest modes are associated with cooperative long wavelength motions occurring on a global scale. They indicate the most probable mechanism of action as a unique function of the overall molecular architecture. The

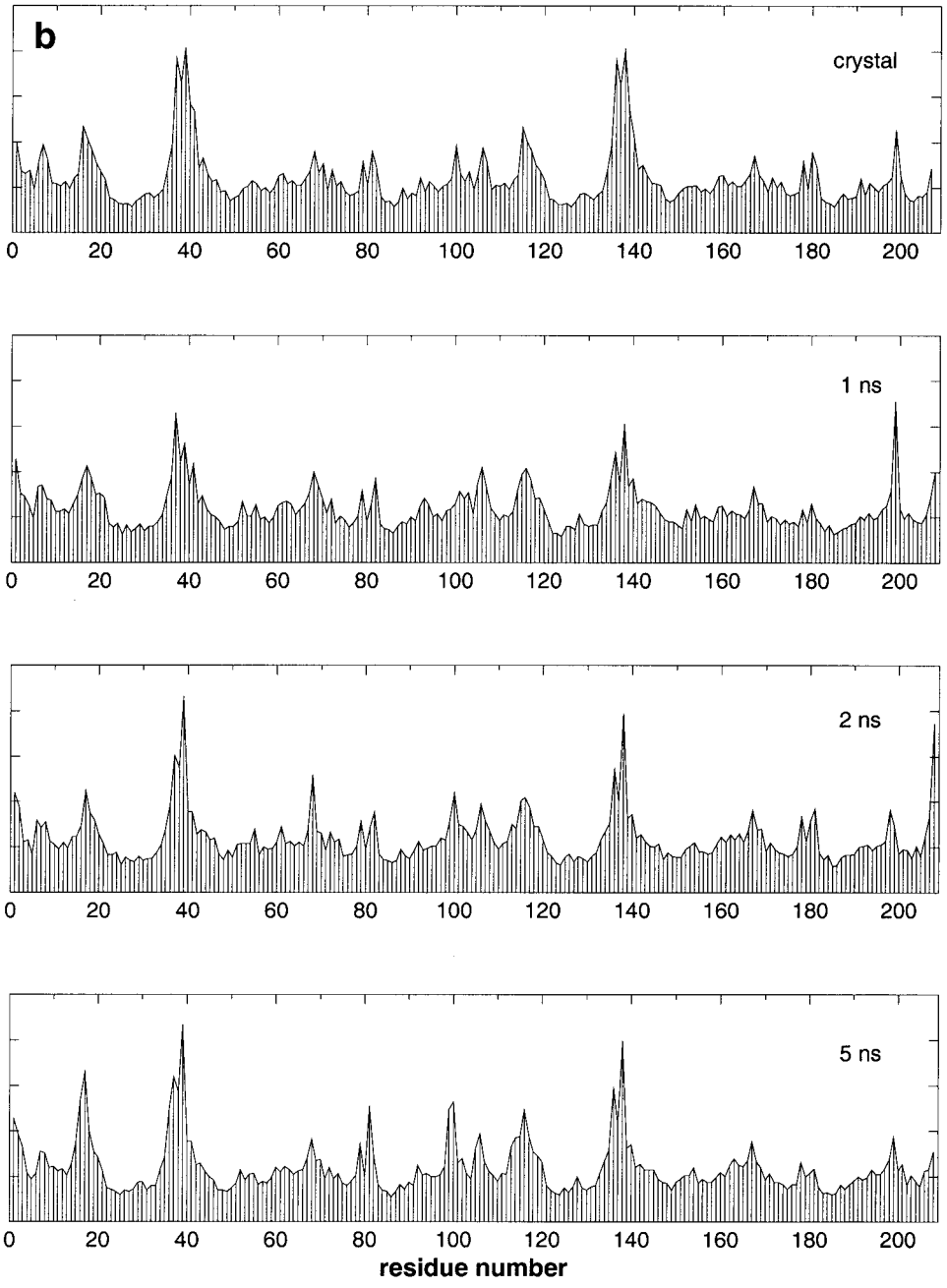

Figure 2. (Continued.)

maxima in the shape of these principal modes describe regions of highest mobility/flexibility, whereas minima correspond to rigid parts with restricted motion that may act like hinges or pivots in the global motion.

In general, the orientational correlations between residue motions vary in the range $-1 \leq C_{ij} \leq +1$; where the upper and lower limits refer to fully correlated and fully anticorrelated motions, respectively. A clearer picture is obtained by examining isolated modes where the values $+1$, -1, or 0 are assigned to each residue pair as calculated in Eq. 5. Thus, the residues move either in the same direction $(+1)$, in the opposite direction (-1), or have no component in this mode and are stationary (0). Hence, in a single mode, one expects to identify two regions of the structure, each correlated within itself but anticorrelated with the other.

The mean-square fluctuations in the first, second, third, and fourth, and the average of the first 10 slowest modes for different conformations of the unliganded protease are shown in Figure 3(a). The linear correlation coefficients between each mode shape of the different conformations are given in Table I. Figure 4(a) shows both the amplitude

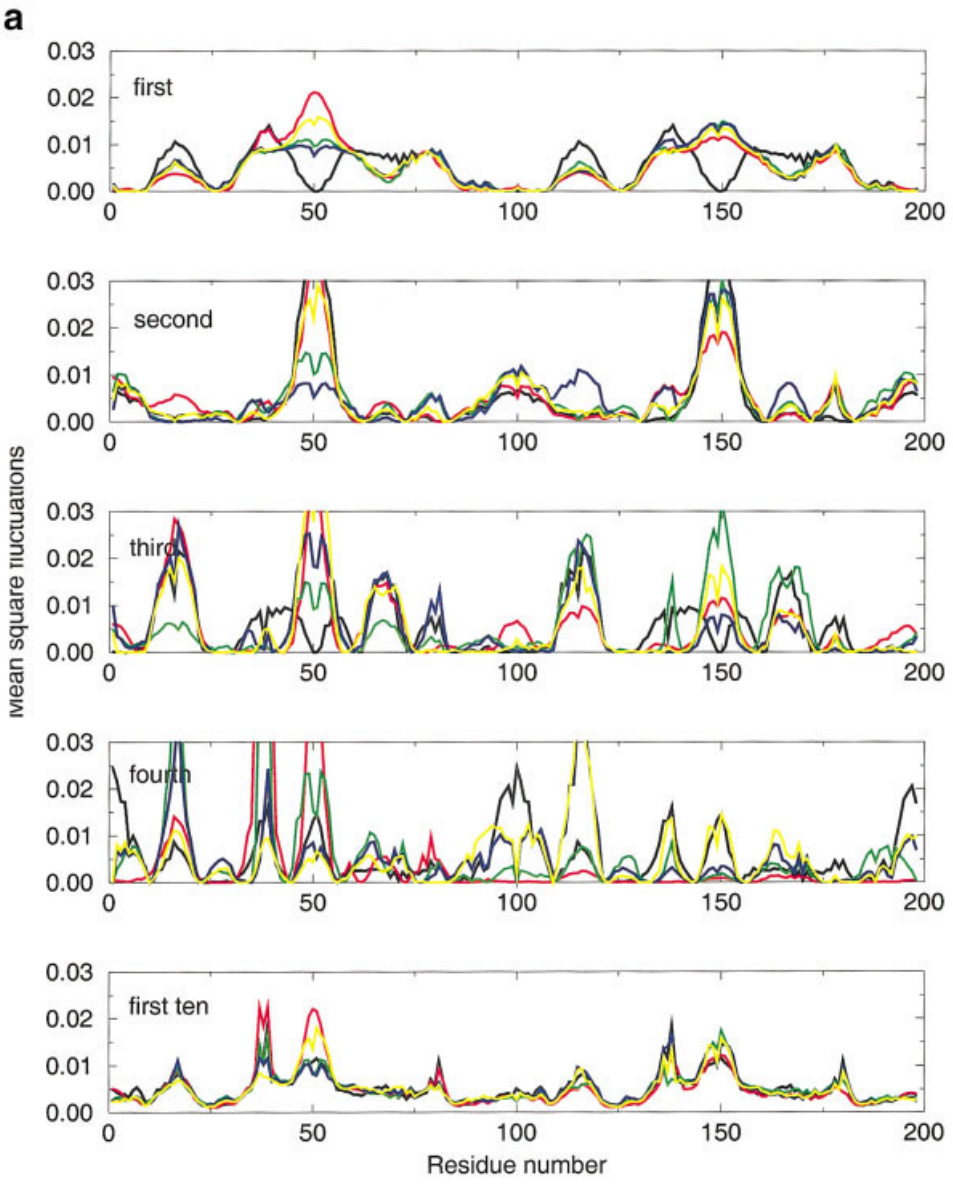

Fig. 3. Mean-square fluctuations in the first, second, third, fourth, and first 10 slowest modes of motion at different conformations of (**a**) unliganded and (**b**) liganded proteases. The colors refer to the following conformations: black (crystal), red (1 ns), green (2 ns), blue (5 ns), and yellow (10 ns). The results are normalized and in arbitrary units.

and its correlation for the first four slowest modes of the unliganded protease structure. The two regions identified by cross correlations are in different colors. The lighter shade indicates more restricted mobility as revealed by smaller amplitude fluctuations in the mode shape, and the more mobile parts of the protein are shown in darker shades.

The first rows in Figures 3 and 4 give the dynamic behavior in the first principal mode of motion, whose shape was previously calculated for the crystal structure by Bahar et al.[24] This mode corresponds to the most cooperative motion of the structure. The motion becomes more localized with the involvement of fewer residues in the next modes and with correspondingly higher frequencies. These first principal modes represent relatively global motions of the structure, and their superposition determines the overall apparent dynamic behavior.

In the slowest mode of the crystal structure, the motion at the dimer interface, including the tips of extended flaps in close proximity, is predicted to be restricted, whereas the 40's loop is the most mobile region. In the MD simulation, the extended conformation of the flaps is lost on a very short timescale, and they exhibit a variety of different conformations in the native state. In the confor-

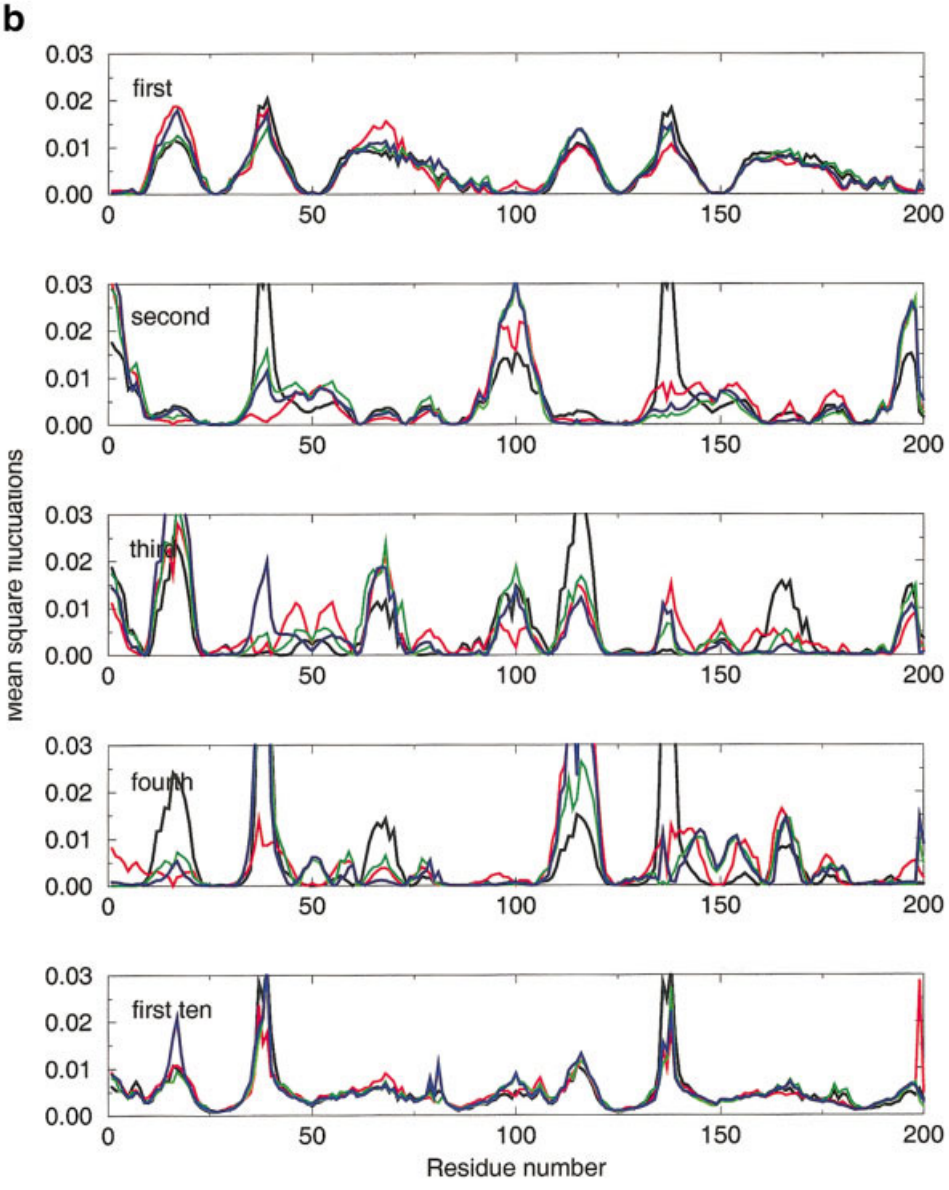

Figure 3. (Continued.)

mations adopted, the flaps are the most mobile regions of the structure—as expected from NMR results[6]—and either one or the other of the two flaps is the most mobile in the different snapshots. The 40's loop does not show significant mobility once the flaps open up.

In the second principal mode of motion [Figs. 3(a) and 4(a), second row], the flaps are the most mobile regions of the protease crystal structure. The flaps continue to be the most mobile for the second mode of the other conformations examined. This agreement includes predicting which flap is the more mobile in that conformation. In this mode, the mobility of a band of residues on an axis perpendicular to the dimer interface are restricted, and the cross-correlation analysis suggests that the (upper and lower) regions on either side of this axis make anticorrelated fluctuations.

The third and fourth slowest modes [Figs. 3(a) and 4(a), third and fourth rows] are also conserved in all the conformations examined, except in the unliganded crystal structure where the extended conformation of flaps affects the behavior; this is reflected in the correlation coefficients (Table I). The set of residues that cluster in particular modes are the same in different conformations, but the third mode for the crystal structure is the fourth mode in the other conformations from MD. This third mode shows the motion of the dimer interface versus the monomer

TABLE I. The Linear Correlation Coefficient Between the Mode Shapes of Different Conformations of Liganded and Unliganded Proteases

Unliganded protease										
Mode number	x-1	x-2	x-3	x-4	1–2	1–3	1–4	2–3	2–4	3–4
First	0.36	0.51	0.54	0.43	0.88	0.86	0.96	0.98	0.96	0.94
Second	0.91	0.85	0.69	0.95	0.74	0.56	0.93	0.88	0.90	0.94
Third	0.25	0.25	0.31	0.42	0.47	0.82	0.90	0.57	0.64	0.87
Fourth	0.43	0.44	0.54	0.43	0.76	0.23	0.07	0.54	0.21	0.81
Slowest 2	0.92	0.9	0.86	0.95	0.85	0.77	0.96	0.97	0.94	0.88
Slowest 10	0.87	0.84	0.91	0.84	0.84	0.78	0.85	0.90	0.88	0.89
Liganded protease										
Mode number	x-1	x-2	x-3	1–2	1–3	2–3				
First	0.86	0.94	0.94	0.88	0.94	0.96				
Second	0.52	0.59	0.57	0.88	0.93	0.97				
Third	0.31	0.64	0.53	0.83	0.80	0.85				
Fourth	0.61	0.77	0.59	0.52	0.75	0.89				
Slowest two	0.70	0.77	0.74	0.89	0.92	0.95				
Slowest ten	0.78	0.89	0.87	0.75	0.78	0.91				

x, 1, 2, 3, and 4 refer to the crystal structure, and conformations at 1, 2, 5, and 10 ns of MD simulations, for the respective states of the protease.

cores. When the flaps open up, their mobility is enhanced along with the tips of the 15′s loop. In the fourth slowest mode of the crystal structure, the third for others, the regions defined by cross correlations is diagonal to what is seen in the third mode.

To understand the overall molecular motion better, we looked at the combined effect of the first 10 slowest modes [Fig. 3(a), fifth row; Fig. 5, first two rows]. This combination includes the major modes while excluding the relatively more localized motions. The contribution (see Models and Methods) of the first 10 slowest modes is ~40% of the entire spectrum in the present analysis. The high correlation coefficient values (Table I) between the shape of first 10 slowest modes of different conformations depict the consistency of the mode shapes throughout the MD simulation. Nevertheless, the first and second slowest modes still dominate the overall global motion of the molecule because they alone contribute ~20% of the spectrum. Despite the extended conformation in crystal structure, the high mobility of flaps can be predicted in the global motion of the protease, and the residues at the interface and especially at the active site loop have the most restricted motion.

Substrate-bound protease

Upon binding a peptide or peptidomimetic, the flaps of HIV-1 protease undergo a large conformational change to lie over the substrate and make a parallel β-sheet with one side of the substrate and an antiparallel β-sheet on the other. To understand the effect of substrate binding on the cooperative fluctuations of HIV-1 protease, we analyzed a complex of an inactive variant of HIV-1 protease (D25N) with one of its natural substrate sites, capsid-p2.[8] Both the conformation of it found in the crystal structure and from several time points during MD simulation[3] were used.

The mean-square fluctuations along the first, second, third, and fourth, and the average of the first 10 slowest modes are shown in Figure 3(b) for the liganded protease at different conformations and the linear correlation coefficients between the mode shapes are given in Table I. In the liganded structure [Figs. 3(b) and 4(b)], the regions defined by cross correlations between fluctuations at different modes are essentially the same as for the unliganded case, except for differences in the amplitude of the motions. It is not surprising that the binding of the substrate severely restricts the mobility of the flap regions.

In the first principal mode of motion, the mobility at the dimer interface (including the flaps and substrate) are restricted in all conformations adopted by the substrate-bound protease. The farther the residues are from the interface, the greater their amplitude of fluctuations. The two regions identified by cross correlations are the two monomers [Fig. 4(a), different colors], whereas a few residues at the dimer interface belong to the region defined by the other monomer, similar to the unliganded case.

In the second slowest mode, the 40's loop is the most mobile region in the crystal conformation, followed by the terminal region. The flexibility of the terminal region is enhanced for other conformations obtained from MD. Compared to the unliganded case where the flaps have an exceptional flexibility as indicated by the high amplitudes in the second mode [Fig. 4(a)], the restriction induced by substrate binding is remarkable in all the conformations examined. The regions defined by cross-correlation analysis are the upper and lower parts of the structure (as depicted in the figures) as in the unliganded case, except that the active site loop and some of the 80's loop do not belong to the upper region.

Fig. 4. Variation of the first four principal modes of motion for (**a**) unliganded (**b**) liganded HIV-1 protease. The row numbers (from top to bottom) correspond to mode numbers, and each column is for a different conformation of the structure. The two regions each correlated within itself but anticorrelated with each other as identified by the cross-correlation analysis at the particular mode in blue and red. The shade of the color is an indication of high-amplitude fluctuations of the particular mode; darker shades correspond to higher amplitudes.

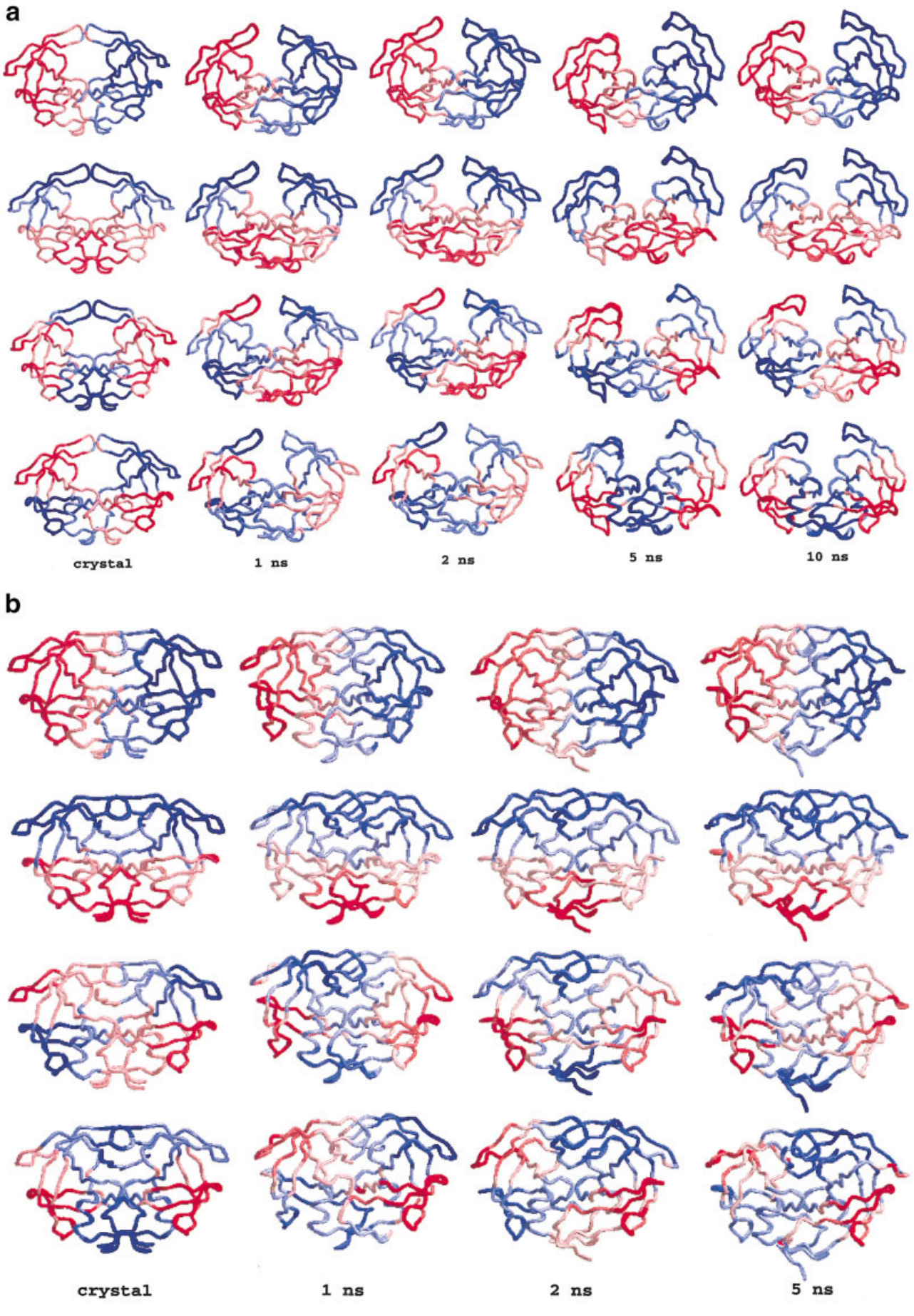

Figure 4.

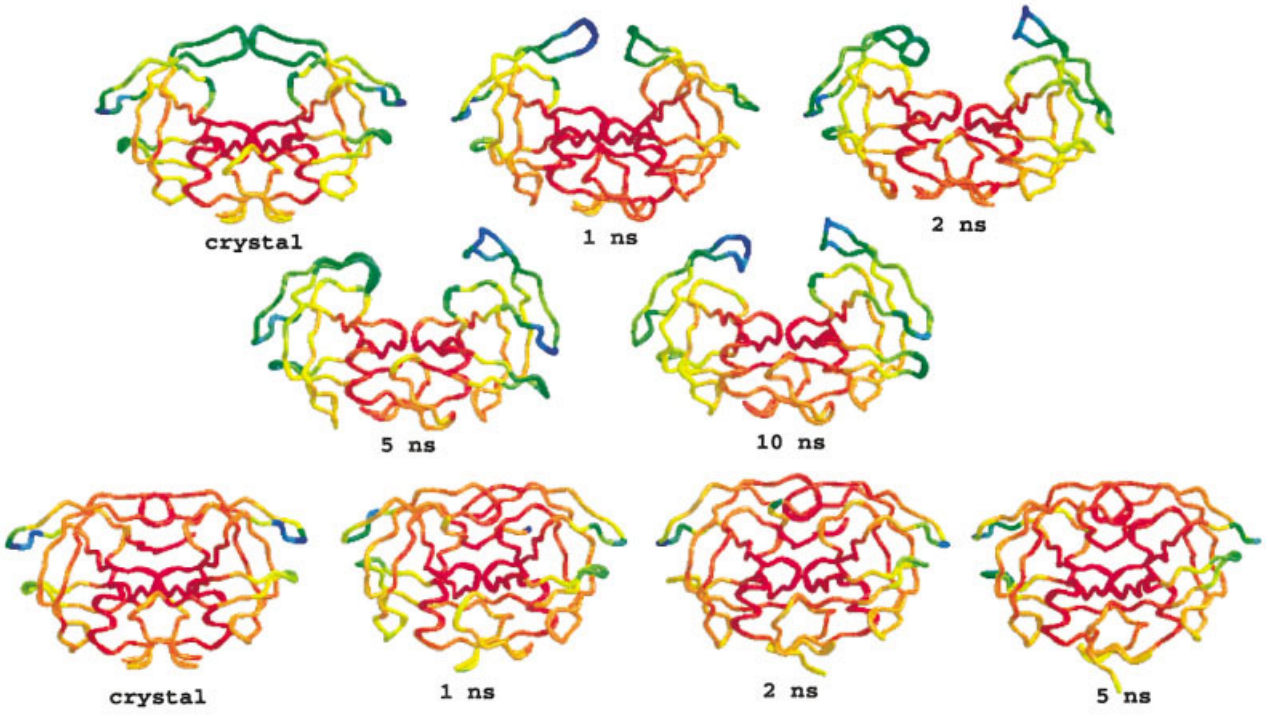

Fig. 5. Rainbow diagram showing the flexibility in the combined first 10 slowest modes of motion. The first two rows are for the unliganded HIV-1 protease, and the last row is for its liganded complex structure. The color spectrum from red to blue corresponds to increasing mobility.

The switch between the crystal structure and conformations from MD seen in the third and fourth slowest modes is seen also for the liganded case. These modes are also switched when comparing the liganded and unliganded cases; the regions of cross correlation in the third and fourth slowest modes for the liganded crystal structure are the same as those in the fourth and third modes of unliganded crystal structure, respectively. In these modes, the flexibilities of the tips of 40's loop and 15′s loop are most pronounced.

The combined effect of the first 10 slowest modes (Fig. 3, last row; Fig. 5, last row) supports the finding[6,35] that the binding of substrate causes a severe restriction of the mobility of HIV-1 protease, and only some loop tips can be considered to be mobile in the global structure. The respective correlation coefficients in Table I depict the consistency of the mode shapes throughout the simulations.

Fast Modes

At the other end of the vibrational spectrum lie the highest frequency fluctuations of individual residues. As opposed to the smooth mode shapes in the low-frequency modes from the collective motions of many residues, only a single or a few residues contribute to the sharp peaks of the fastest modes. These residues are supposed to be centers of localized energy because their amplitudes are restricted due to the high number of neighboring residues (i.e., high contact number). Hence, their motions are tightly restricted to high-frequency motions and may play a key role in the stability of the protein structure.

The residues having highest frequency motions in the fastest five modes of motion are labeled in Figure 6(a) for different conformations of the unliganded structure. The results are identical for the two monomers in the crystal structure because the conformation is completely symmetric. Almost all the residues of highest frequency motion in each monomer of the crystal structure are located on an axis extending from the dimer interface near the active site toward the base of the flaps. In other conformations of HIV-1 protease sampled in MD, the same residues in this region of the protein structure had fast modes, with a few exceptions. In all conformations of the unliganded protease examined, Gly86 makes the highest-frequency fluctuations without exception.

In the liganded case, the highest frequency residues are located around the substrate [Fig. 6(b)]. More residues around the active site and in the flaps become important for stability in the substrate-bound form. Gly48 in the flap competes with Gly86 in making the highest frequency fluctuations. Upon substrate binding, residues 58–59 and 76–77 lose their importance for stability.

DISCUSSION

In examining several conformations of the adaptable HIV-1 protease structure we found that except for the mobile flaps that display large conformational changes, the overall shapes of the principal modes, as calculated from GNM, are conserved in all the conformations examined. The similarity in the results obtained supports the application of GNM to any conformation accessible in the native state, including the crystal structure, as a reason-

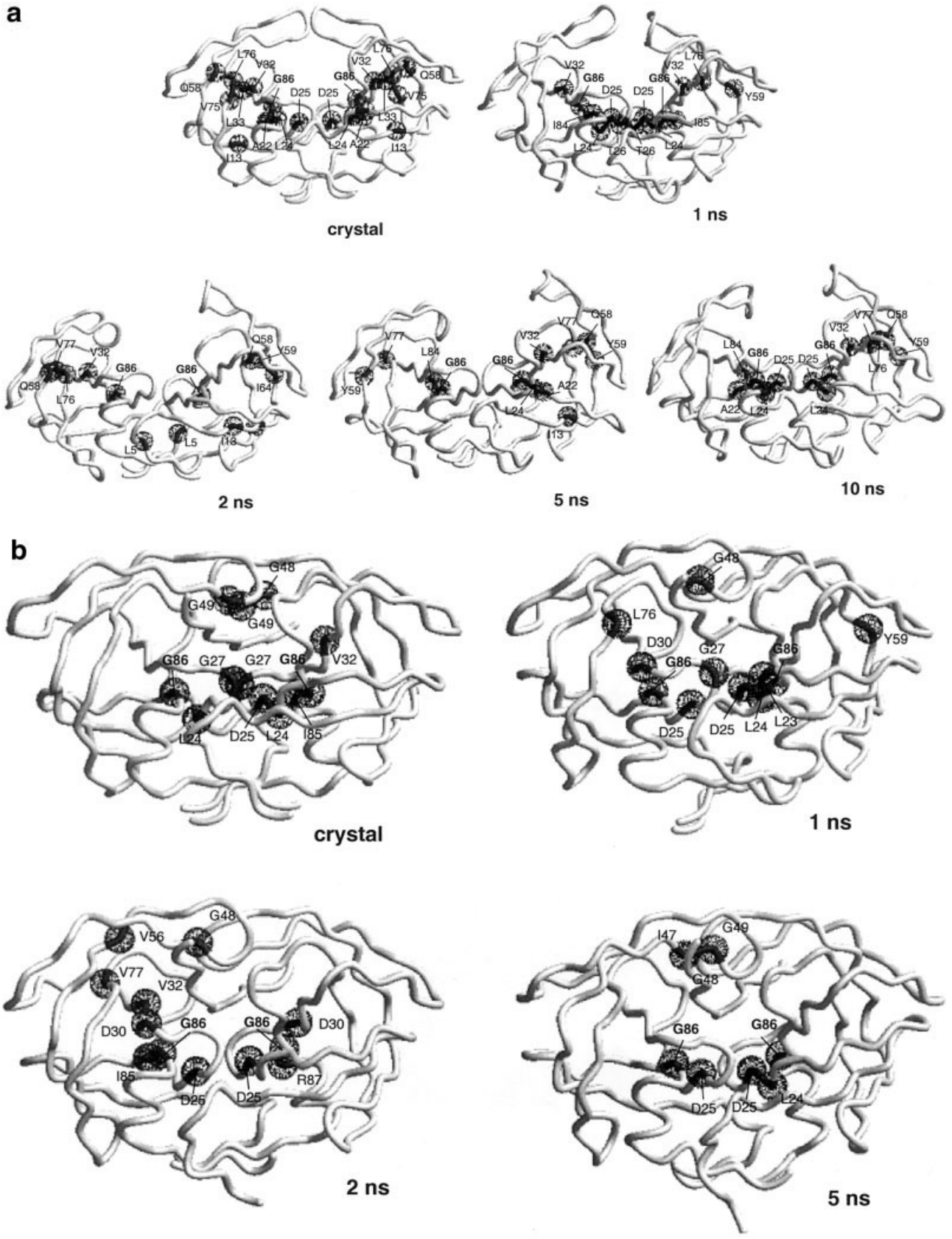

Fig. 6. High-frequency modes of (**a**) unliganded and (**b**) liganded HIV-1 protease at different conformations where the residues vibrating with high frequencies in the fastest five modes are indicated as spheres on their α-carbons and labeled.

able starting point for the analysis of the fluctuations of the overall molecular architecture. However, if the system undergoes large conformational changes in solution or if the conformation in the crystal structure has artifacts due to crystal packing or other effects, examination of only a single structure may not be sufficient for obtaining reliable results on the details of the dynamics. Still, the results show that most global and high-frequency motions are conserved for all conformations studied here and are, therefore, likely to be biologically relevant.

The slowest mode in all the conformations, and therefore the most global motion of both the unliganded and substrate-bound protease structure, is the independent motion of two monomers on a principal axis that is parallel to the dimer interface (Fig. 4). This could be because the two monomers have to form this dimeric structure to recognize and bind a substrate. The dimerization process is the most global phenomenon associated with this structure and is vital for its biological functioning. The mobility of residues around the dimer interface is constrained, whereas the amplitude of the fluctuations is higher at the core domains of the monomers. There is a broad band with low mobility around the dimer interface, rather than an axis to define one hinge. This observation suggests that the motion of monomers occurs around two hinges at either side of this stable and rigid band. The fluctuations of residues in the two monomers are anticorrelated, except for a few residues at the interface that group with the other monomer. These few residues, located in the quite rigid interface and fluctuating within the other monomer, might act as anchors to keep the two monomers together in a dimeric structure. In this slowest mode of the unliganded structure, the motion of flaps is restricted when they are in an extended conformation, but they gain mobility once the tips start to curl in. This is also reflected in the low-correlation coefficients for the slowest mode of the crystal structure with other conformations (Table I). Hence, looking only at the crystal structure and only at the slowest mode of motion could lead to misinterpretations about the flap mobility.

The correlation coefficients between the mode shapes of the different conformations in both liganded and unliganded states of the protease are the highest for the first 2 slowest and first 10 slowest modes (Table I). This shows that the first two modes are conserved and dominate; however, switches of modes with changes in the conformations are observed for the higher frequency modes in the spectrum. This, together with the differences in the amplitude of more localized motions, leads to relatively poor linear correlations between these higher frequency modes. However, it is clear from Figure 4 that the general behavior reflected by cross correlations display the consistency of the dynamics between different conformations in the trajectories.

The behavior of the protease in the second global mode of motion is approximately perpendicular to the first one, with residues in upper and lower parts (Fig. 4) forming the two anticorrelated regions. There is mobility of the flaps in this mode, whereas residues with restricted mobility at the interface act like hinges between the two flexible flaps. Therefore, this second dominant mode is associated with the large amplitude movement of flaps and hence functioning of the protease.

The third and fourth slowest modes of motion describe the motion of dimer interface with respect to the monomer cores and a diagonal motion. In these modes, the flaps and tips of loops are the most flexible regions of the structure. Although the order of their frequencies (third vs forth) is switched for the crystal structures and there are differences in the amplitude of the motions, they consistently describe the same behavior for different conformations examined.

Examination of the collective fluctuations of the 10 slowest, most dominant modes (Fig. 5) indicates a high mobility of the flaps for the unliganded protease. Upon substrate binding, a severe restriction of mobility is observed except for the tips of 40's and 15′s loops. The most restricted region is around the active site both in the liganded and unliganded cases, in agreement with their relatively large order parameters seen in NMR measurements[35] and by the root-mean-square deviation analysis of 73 crystal structures.[36]

Upon substrate binding, the motion of the flaps becomes restricted through new contacts, whereas the terminal region gains mobility. The high mobility of the terminal region, especially in the second slowest mode of motion, is consistent with NMR results[6] and may be related to the autoprocessing of HIV-1 protease.[37] In the liganded case, the tips of 40's loop are the most flexible regions of the structure. This finding is also in agreement with the results of a recently developed technique for predicting protein flexibility according to the hydrogen-bonding network using graph theory.[38] The 40's loops, also referred to as the "flap elbows," are predicted to be highly mobile by both MD simulations and NMR studies of HIV-1 protease/KNI-272 complex.[35,39] Residues 38–45 only mutate rarely and then only tolerate conservative substitutions in patients treated with protease inhibitors,[40] indicating that the high mobility of the 40's loop may be important for function. In our analysis of different conformations, we find that the 15′s loop may also become mobile as seen by the third and fourth principal modes of action. Actually, the 15′s loop is reported to have smaller than average (but larger than those in 40's loop) order parameters.[35]

Binding affects the dynamics of not only the residues involved in interactions with the substrates, but of the overall structure. Todd and Freire[41] proposed that the effect of binding propagates to regions not in direct contact with the substrate because of a network of cooperative interactions distributed within the protease structure. Although substrate binding induces restriction of mobility on the overall structure, the mobility of some regions are enhanced as indicated above. It has been proposed[41] that the enthalpic interactions on binding of small molecules are likely to be small and not enough to overcome the entropic loss due to restrictions in the binding region. Hence, some other regions become more mobile to provide a compensating entropic energy component necessary to

drive binding.[42] Along this line, a recent study on CI2 demonstrates how the decrease in entropy introduced by binding is compensated by changes in other parts of the structure.[43]

The fastest modes at the other end of the vibrational spectrum are also conserved to a great extent for the various conformations of HIV-1 protease examined. Although the exact order of the residues may not be the same in the five fastest modes for different conformations, their identities are conserved within the 10 fastest modes. The residue making the highest frequency fluctuations is Gly86 for the unliganded structure and either Gly48 or Gly86 for the substrate-bound structure without any exceptions. It is of interest that the residues in the fastest modes of the liganded structure are essentially the glycine residues surrounding the bound substrate and the active site residue.

The GNM analysis suggests that Gly86 is a key residue for the stability of the protease. Gly86 is located at the C-cap of the small helix near the active site and has a high coordination number, contacting 14 residues within a 7 Å sphere. According to Stanford HIV RT and Protease Sequence Database, this residue is almost invariant among patients treated with protease inhibitors.[40] Database analysis of other residues seen in the fastest modes revealed that these residues are either highly conserved or associated with drug-resistant mutations.

CONCLUSIONS

In this study, an analytical tool is combined with the results of a simulation method to explore the cooperative fluctuations of a variety of conformations of HIV-1 protease. As is well known, a protein is not rigid in its folded state, but it adopts different conformations in the course of performing its biological function. The difficulties in efficiently sampling these conformations using fully atomistic simulations to understand global motions leads to the search for more computationally efficient techniques. In another study, normal mode analysis was used to study the low-temperature vibrational inhomogeneity of globular proteins in a similar manner to ours by calculating the modes on a series of conformations minimized from the selected points along the MD trajectory.[44] In this study, we combine simulations with the simple analytical model GNM to extract useful and biologically significant information from the trajectories, such as identifying the rigid and mobile regions of HIV-1 protease in an efficient way.

The conservation of overall dynamic behavior seen in HIV-1 protease in crystal structures and different conformations in the native state obtained from MD simulations supports the application of GNM to available PDB structures for reliable dynamic analysis as well as elucidating the dynamic consistency of MD conformations. This finding has also recently been similarly confirmed by using normal mode analysis on a series of crystal structures.[36] However, the details of dynamic behavior obtained by GNM may change when the protein undergoes large conformational changes in solution, and examining only a single mode for the crystal structure may be misleading in cases where there are important crystal-packing effects. Therefore, applying the method to several conformations of a protein is more reliable, especially for interpreting the details of dynamic behavior. Nevertheless, our results of this study on HIV-1 protease show that the dynamics at either end of the vibrational spectrum are an intrinsic property of the overall molecular architecture of a protein and are highly conserved among its various conformations.

ACKNOWLEDGMENTS

This work was partially supported by BU Research Funds grant numbers 00HA502D-00HA503, DPT Project 01K120280 and by the Turkish Academy of Sciences, in the framework of the Young Scientist Award Program (EA-TUBA-GEBIP/2001-1-1). N.K. thanks TUBITAK for her BDP fellowship.

REFERENCES

1. Kohl NE, Emini EA, Schlieff WA, Davis LJ, Heimbach JC, Dixon RA, Scolnick EM, Sigal IS. Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci USA 1988;85:4686–4690.
2. Spinelli S, Liu QZ, Alzari PM, Hirel PH, Poljak RJ. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. Biochimie 1991;73:1391–1396.
3. Scott WRP, Schiffer CA. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. Structure 2000;8:1259–1265.
4. Rose RB, Craik CS, Stroud RM. Domain flexibility in retroviral proteases: structural implications for drug resistant mutations. Biochemistry 1998;37:2607–2621.
5. Lange-Savage G, Berchtold H, Liesum A, Budt KH, Peyman A, Knolle J, Sedlacek J, Fabry M, Hilgenfield R. Structure of HOE/BAY 793 complexed to human immunodeficiency virus (HIV-1) protease in two different crystal forms—structure/function relationship and influence of crystal packing. Eur J Biochem 1997;248:313–322.
6. Ishima R, Freedberg DI, Wang YX, Louis JM, Torchia DA. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. Structure Fold Des 1999;7:1047–1055.
7. Freedberg DI, Ishima R, Jacobs J, Wang YX, Kustanovich I, Louis JM, Torchia DA. Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with prediction of dynamics calculations. Protein Sci 2002;11:221–232.
8. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. J Mol Biol 2000;301:1207–1220.
9. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA. Substrate shape determines specifisity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. Structure 2002;10:369–381.
10. Bahar I, Jernigan RL. Vibrational dynamics of transfer RNAs: comparison of the free and synthetase-bound forms. J Mol Biol 1998;281:871–884.
11. Bahar I, Jernigan RL. Cooperative fluctuations and subunit communication in tryptophan synthase. Biochemistry 1999;38:3478–3490.
12. Keskin O, Jernigan RL, Bahar I. Proteins with similar architecture exhibit similar large-scale dynamic behavior. Biophys J 2000;78:2093–2196.
13. Haliloglu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. Phys Rev Lett 1997;79:3090–3093.
14. 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.
15. Bahar I, Wallquist A, Covell DG, Jernigan RL. Correlation between native state hydrogen exchange and cooperative residue fluctuations from a simple model. Biochemistry 1998;37:1067–1075.

16. 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.
17. Bahar I, Erman B, Jernigan RL, Atilgan AR, Covell DG. Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function. J Mol Biol 1999;285:1023–1037.
18. Isin B, Doruker P, Bahar I. Functional motions of influenza virus hemagglutinin: a structure-based analytical approach. Biophys J 2002;82:569–581.
19. Go N, Noguti T, Nishikawa T. Dynamics of a small protein in terms of low-frequency vibrational modes. Proc Natl Acad Sci USA 1983;80:3693–3700.
20. Levy RM, Karplus M. Vibrational approach to dynamics of an α-helix. Biopolymers 1979;18:2465–2495.
21. Amadei A, Linssen AB, Berendsen HJC. Essential dynamics of proteins. Proteins 1993;17:412–425.
22. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins 1998;33:417–429.
23. de Groot BL, Hayward S, van Aalten DMF, Amadei A, Berendsen HJC. Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data. Proteins 1998;31:116–127.
24. Bahar I, Atilgan AR, Demirel MC, Erman B. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. Phys Rev Lett 1998;80:2733–2736.
25. Kitao A, Go N. Investigating protein dynamics in collective coordinate space. Curr Opin Struct Biol 1999;9:164–169.
26. Caves LSD, Evanseck JD, Karplus M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. Protein Sci 1998;7:649–666.
27. Clarage JB, Romo T, Andrews BK, Pettitt BM, Phillips JGN. A sampling problem in molecular dynamics simulations of macromolecules. Proc Natl Acad Sci USA 1995;92:3288–3292.
28. Balsera MA, Wriggers W, Oono Y, Schulten K. Principal component analysis and long time protein dynamics. J Phys Chem 1996;100:2567–2672.
29. 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.
30. Flory PJ. Statistical thermodynamics of random networks. Proc R Soc Lond A 1976;351:351–380.
31. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 1996;77:1905–1908.
32. Miyazawa S, Jernigan RL. Energy functions that discriminate x-ray and near-native folds from well-constructed decoys. Macromolecules 1985;18:534–552.
33. 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.
34. Bernstein FC, Tasumi M. The protein data bank: a computer-based archival file for macromolecular structures. J Mol Biol 1997;112:535–542.
35. Freedberg DI, Yun-Xing W, Stahl SJ, Kaufman JD, Wingfield PT, Kiso Y, Torchia DA. Flexibility and function in HIV protease: dynamics of the HIV-1 protease bound to the asymmetric inhibitor kynostatin 272 (KNI-272). J Am Chem Soc 1998;120:7916–7923.
36. Zoete V, Michielin O, Karplus M. Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for analysis of protein flexibility. J Mol Biol 2002;315:21–51.
37. Louis JM, Clore GM, Gronenborn AM. Autoprocessing of HIV-1 protease is tightly coupled to protein folding. Nat Struct Biol 1999;6:868–875.
38. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein flexibility predictions using graph theory. Proteins 2001;44:150–165.
39. Luo XC, Kato RH, Collins JR. Dynamic flexibility of protein-inhibitor complexes: a study of the HIV-1 protease KNI-272 complex. J Am Chem Soc 1998;120:12410–12418.
40. Kantor R, Machekano R, Gonzales MJ, Dupnik K, Schapiro JM, Shafer RW. Human immunodeficiency virus reverse transcriptase and protease sequence database: an expanded data model integrating natural language text and sequence analysis programs. Nucleic Acids Res 2001;29:296–299.
41. Todd MJ, Freire E.The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. Proteins 1999;36: 147–156.
42. Forman-Kay JJ. The "dynamics" in the thermodynamics of binding. Nat Struct Biol 1999;6:1086–1087.
43. Baysal C, Atilgan AR. Coordination topology and stability for the native and binding conformers of chymotrypsin inhibitor 2. Proteins 2001;45:62–70.
44. Lamy AV, Souaille M, Smith JC. Simulation evidence for experimentally detectable low temperature vibrational inhomogeneity in a globular protein. Biopolymers 1996;39:471–478.