Coordination topology and stability for the native and binding conformers of chymotrypsin inhibitor 2

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · OCTOBER 2001

Impact Factor: 2.63 · DOI: 10.1002/prot.1124 · Source: PubMed

2 AUTHORS:

Canan ATILGAN
Sabanci University
81 PUBLICATIONS **836** CITATIONS
SEE PROFILE

Ali Rana Atilgan
Sabanci University
83 PUBLICATIONS **3,362** CITATIONS
SEE PROFILE

PROTEINS: Structure, Function, and Genetics 45:62–70 (2001)

Coordination Topology and Stability for the Native and Binding Conformers of Chymotrypsin Inhibitor 2

Canan Baysal[1]* and Ali Rana Atilgan[2]
[1]*Laboratory of Computational Biology, Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli, Tuzla, Istanbul, Turkey*
[2]*School of Engineering and Polymer Research Center, Bogazici University, Bebek, Istanbul, Turkey*

ABSTRACT **We demonstrate that the stabilization of the binding region is accomplished at the expense of a loss in the stability of the rest of the protein. A novel molecular mechanics (MM) approach is introduced to distinguish residue stabilities of proteins in a given conformation. As an example, the relative stabilities of folded chymotrypsin inhibitor 2 (CI2) in unbound form, and CI2 in complex with subtilisin novo is investigated. The conformation of the molecule in the two states is almost identical, with an ~0.6-Å root-mean-square deviation (RMSD) of the Cα atoms. On binding, the packing density changes only at the binding loop. However, residue fluctuations in the rest of the protein are greatly altered solely due to those contacts, indicating the effective propagation of perturbation and the presence of remotely controlling residues. To quantify the interplay between packing density, packing order, residue fluctuations, and residue stability, we adopt an MM approach whereby small displacements are inserted at selected residues, followed by energy minimization; the displacement of each residue in response to such perturbations are organized in a perturbation-response matrix L. We define residue stability $\lambda_i = \sum_j L_{ij} / \sum_j L_{ji}$ as the ratio of the amount of change to which the residue is amenable, to the ability of a given residue to induce change. We then define the free energy associated with residue stability, $\Delta G_\lambda = -RT \ln \lambda$. ΔG_λ intrinsically selects the residues that are in the folding core. Upon complexation, the binding loop becomes more resistant to perturbation, in contrast to the α-helix that favors change. Although the two forms of CI2 are structurally similar, residue fluctuations differ vastly, and the stability of many residues is altered upon binding. The decrease in entropy introduced by binding is thus compensated by these changes. Proteins 2001;45:62–70.**

Key words: protein conformers; stability; chymotrypsin inhibitor 2

INTRODUCTION

The native contact topology of proteins appear to govern folding rates and mechanisms. The interplay between the packing density and packing order is thus proposed to offer a simplistic view of protein folding.[1] Kinetically, a correlation between topological complexity of proteins and folding rates have been established in single domain proteins exhibiting two-state behavior, independent of protein length.[2] In a more recent work, stability was shown to play an important role in determining folding rates, along with contact order.[3] Protein topology was also shown to be detrimental in its functioning.[4–10] Fersht has proposed the existence of an extended nucleus which is composed of reminiscents of the native secondary structures stabilized by local and long-range tertiary interactions.[11] This nucleus appears to explain the observations of folding rate depending on both topology and stability. The relationship between topology and residue stability has been investigated in some theoretical studies, e.g., Baldwin and Rose,[12,13] and Rumbley et al.[14]

Recent understanding of protein structure related events address a probabilistic view where the relevant event is not governed by the static three-dimensional structure. Rather a distribution of populations of conformers are operational both in folding[15] and binding.[16,17] In this respect, proteins may be thought of exhibiting the kind of behavior observed in small peptides showing conformational multiplicity with a limited number of microstates accessible in the "folded" state.[18,19]

It has been further shown both experimentally[20,21] and computationally[10,22] that the effects of perturbations (such as those induced by binding) are transmitted to remote regions. Freire and coworkers[23] have introduced a method for the structure-based thermodynamic analysis of proteins, which they have applied to a number of cases to analyze the relation between function, binding, and stability.[24–26]

In this study we develop a perturbation-response analysis using a molecular mechanics (MM) methodology. This is a simple approach whereby the energy minimized native coordinates of the protein is altered by inserting a displacement at a selected residue followed by energy minimization. We exploit the fact that when a small perturbation is introduced in a region of the molecule, relaxation to equilibrium does not occur by directly restoring the initial

*Correspondence to: Canan Baysal, Laboratory of Computational Biology, Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli 81474, Tuzla, Istanbul, Turkey. E-mail: canan@sabanciuniv.edu

Received 13 February 2001; Accepted 8 June 2001

conditions.[27,28] For example, if a bond is lengthened, this bond will not directly shrink to its original coordinates during energy minimization. Rather, relaxation at softer degrees of freedom will take precedence, and rearrangements, e.g., in the torsional degrees of freedom, will also be observed. We assume that the perturbation is applied quasi-statically. Such a disturbance is related to mutations, i.e. alterations of packing topology. We show that residue stability arises as a tool reflecting the character of the response.

Recently, a new methodology was introduced that is a low-resolution structural model that aims to identify the control architecture governing the communication between the flexible and rigid parts of a folded protein by computing the adaptive response of the protein to inserted perturbations.[9] This and other recently developed low-resolution approaches have implicated contact topology as a major factor influencing protein dynamics.[7,10,29–31] Directionality of the displacements was also shown to be important.[32,33] Apart from being a high-resolution model, the current MM approach differs from the analytical method introduced by Yilmaz and Atilgan in that the natural response of the protein is followed here as opposed to the adaptive response; i.e., the residues that control the response to the perturbation will emerge in the latter analysis.

The MM methodology is applied to chymotrpsin inhibitor 2 (CI2) which is a commonly adapted protein for studying folding/unfolding events, transition-state structures, and function/stability issues. It is a good structural model because of its relatively small size (64 residues), lack of disulfide bridges, and existence of both short- and long-range stabilization. The methodology allows us to study the perturbation-response behavior of the protein, residue correlations that are comparable to X-ray crystallographic temperature factors, and residue stability. It is applied to CI2 in unbound form as well as a form bound to subtilisin novo. Although the two "conformers" are very similar with only ~0.6 Å RMSD of the C_α atoms, major differences are observed in residue fluctuations and residue stability. These findings are directly related to the change in contact topology upon binding.

MATERIALS, MODELS, AND METHODS

Structures

CI2 is a 83-residue protein with the first 19 residues being unstructured; however, their removal does not change the unfolding properties and truncated CI2 structures are reported in experimental work. In this study, the original residues 20–83 are renumbered 1–64, consistent with other work on CI2.[34,35] For the native coordinates of CI2, we use the X-ray crystallographic structure determined at 1.7-Å resolution[36] [Protein Data Bank (PDB)[37,38] code 1ypc]. We also analyze CI2 in complex with subtilisin novo (pdb code 2sni).[39] The X-ray resolution of the 339 residue complex is 2.1 Å. The overall structure of the unbound and bound forms are very similar with a root mean square deviation (RMSD) of only 0.59 Å between the C_α atoms. The structure (Fig. 1) may be viewed as consisting of two substructures, the first one consisting of the N-terminus and the α-helix and the second one made up of the β-sheet and the reactive site loop containing residues.[35–44] The hydrophobic core is formed by a delicate association of the β-sheet and the α-helix. It consists of 10 hydrophobic residues: W5, L8, A16, I20, I29, V47, L49, V51, I57, and P61.[40] Residues A16, L49, and I57 are also identified to be kinetically hot residues, i.e. those that are directly involved in the formation and stabilization of folding nuclei.[41] One side of the β-sheet that packs against the α-helix accommodates five hydrogen bonds, which are between $^{24}H^Z$ and 44O, 63H and 3O, 5H and 61O, 10H and 57O, and 57H and 11O, while the other side comprises a network of hydrogen bonds and electrostatic interactions with the loop. The hydrogen bonds in the hydrophobic core provide further stabilization.

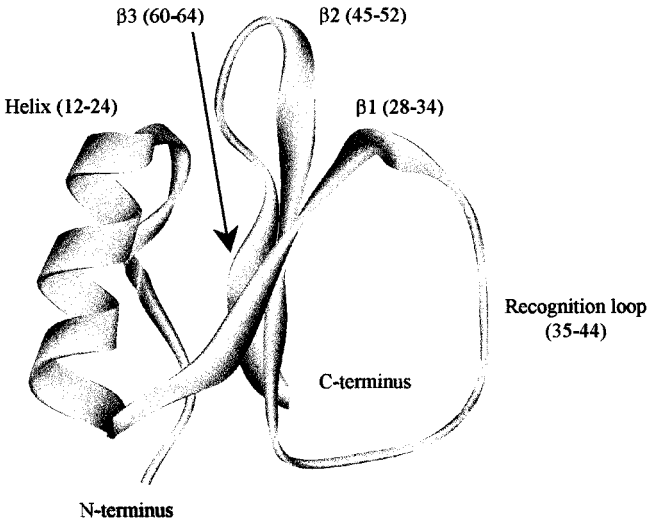

Fig. 1. Ribbon representation of the native coordinates of CI2 mutant (E33A, E34A) from barley seeds. We use the X-ray crystallographic structure determined at 1.7-Å resolution (pdb code 1ypc).[36]

It should be noted that in the X-ray structure 1ypc, the side-chain atoms C_γ, S_δ, and C_ϵ are missing. These atoms are added to the initial structure, assuming a *trans*-conformation for the χ^1 and χ^2 dihedral angles of the side-chain, and equilibrium bond lengths and angles. Also, the side-chains of Glu23, Lys37, and Arg81 are modeled in two conformations in 1ypc, and we take the first of these two conformers in our treatment.

Molecular Mechanics Model

The CVFF forcefield implemented within the Molecular Simulations Inc. InsightII 98.0 software package was used in the MM calculations.[42] Initially, the structures of CI2 in free state (1ypc) and in complex with subtilisin novo (2sni) are each energy minimized to 0.001 kcal/mol/Å of the derivative using conjugate gradients method. Group-based cutoffs are employed with a 9.5-Å cutoff distance. A

switching function is used with the spline and buffer widths set to 1.0 and 0.5 Å, respectively. The RMSD between the original PDB and the energy minimized C_α coordinates are 1.5 and 2.0 Å for the 1ypc and 2sni structures, respectively. These coordinates of the molecule are treated as the main coordinates of the equilibrium structure. The final energies of the minimized structures are E_o(1ypc) = 332.985 kcal/mol and E_o(2sni) = 2702.123 kcal/mol.

In the following discussion, we are interested in the relative displacements of the atoms, and the criterion for terminating the minimization must be selected accordingly. Extensive testing on various proteins have shown us that the usage of a more stringent criterion (e.g., 10^{-4} kcal/mol/Å) affects the absolute value of the energy attained by ~0.2 kcal/mol, whereas the final atomic locations are within ~0.01 Å of their absolute minima at a derivative of 0.1 kcal/mol/Å. This is especially true when the minimization is started from a slightly distorted structure obtained from a very well minimized one as we do in this work. We therefore employ a termination criterion of 0.1 kcal/mol/Å of the derivative in all the subsequent energy minimizations.

Next we use a procedure of systematically perturbing each of the protein residues according to the following scheme: (1) The coordinates of the C_α of the selected residue are displaced by an amount d_i, which leads to a distorted local structure; (2) the new coordinates of the displaced C_α atom is fixed in space, whereas the rest of the protein atoms are free to move; (3) the energy of the protein is minimized to 0.1 kcal/mol/Å of the derivative; (4) the rearranged atomic coordinates in response to the perturbation are recorded for further analysis; (5) the coordinates are reset to those of the equilibrium structure and the process (1–4) is repeated over all the C_α atoms. Note that the perturbation of residue i mentioned in item (1) may be carried out in several different ways. We elaborate on this issue under Results.

The N sets of protein coordinates obtained at the end of the MM calculations are organized into the perturbation-response matrix of order $3N \times 3N$, where the displacement vector of residue i in response to a directed perturbation placed at residue k is denoted by $\Delta\mathbf{R}_{ik}$:

$$\Delta\mathbf{R} = \begin{bmatrix} \Delta\mathbf{R}_{11} & \Delta\mathbf{R}_{12} & \cdots & \Delta\mathbf{R}_{1N} \\ \Delta\mathbf{R}_{21} & \Delta\mathbf{R}_{22} & \cdots & \Delta\mathbf{R}_{2N} \\ \vdots & \vdots & & \vdots \\ \Delta\mathbf{R}_{N1} & \Delta\mathbf{R}_{N2} & \cdots & \Delta\mathbf{R}_{NN} \end{bmatrix} \quad (1)$$

We then compute the $N\times N$ displacement matrix, $\Delta\mathbf{L}$, each of whose elements, L_{ik} (= ΔR_{ik}), are the magnitudes of $\Delta\mathbf{R}_{ik}$ and hence is a measure of the amount of displacement experienced by i in response to the perturbation placed at k:

$$\mathbf{L} = \begin{bmatrix} \Delta R_{11} & \Delta R_{12} & \cdots & \Delta R_{1N} \\ \Delta R_{21} & \Delta R_{22} & \cdots & \Delta R_{2N} \\ \vdots & \vdots & & \vdots \\ \Delta R_{N1} & \Delta R_{N2} & \cdots & \Delta R_{NN} \end{bmatrix} \quad (2)$$

Residue Cross-Correlations

Cross-correlations between residue pairs i and j may be calculated from the MM methodology by the following scheme: $\Delta\mathbf{R}$ (eq 1) multiplied by its transpose yields the $3N\times 3N$ second moment matrix $\mathbf{A} = \Delta\mathbf{R}\,\Delta\mathbf{R}^T$, which carries the average effect of all the perturbations. Alternatively, $\mathbf{A}$ may be viewed as an $N\times N$ matrix, whose ijth element is the 3×3 second moment matrix of correlations between the x-, y-, and z-components of the fluctuations $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ of residues i and j:

$$\mathbf{A}_{ij} = \begin{bmatrix} \langle \Delta X_i \Delta X_j \rangle & \langle \Delta X_i \Delta Y_j \rangle & \langle \Delta X_i \Delta Z_j \rangle \\ \langle \Delta Y_i \Delta X_j \rangle & \langle \Delta Y_i \Delta Y_j \rangle & \langle \Delta Y_i \Delta Z_j \rangle \\ \langle \Delta Z_i \Delta X_j \rangle & \langle \Delta Z_i \Delta Y_j \rangle & \langle \Delta Z_i \Delta Z_j \rangle \end{bmatrix} \quad (3)$$

Finally, the cross-correlations between residues i and j in response to the inserted perturbations is given by

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = \frac{1}{N} tr(\mathbf{A}_{ij}) \quad (4)$$

RESULTS AND DISCUSSION

Nature of the Perturbation

The MM scanning methodology used in this work relies on introducing a perturbation at a selected residue in the form of a displacement, and extracting the response of the protein. The size and the directionality of the perturbation is expected to be consequential on the response pattern of the structure. We therefore first investigate the effect of the size of the displacement. We have carried out the MM scanning with moves of sizes 0.17, 0.35, 0.87, and 1.7 Å, and we have reproduced perturbation-response maps from the $\Delta\mathbf{L}$ matrix of equation 2. The pattern of the response is the same in all runs [e.g., as in Fig. 2(a)]. However, when the size of the displacements is increased to 2.5 Å, a different perturbation-response pattern begins to emerge. We thus conclude that we are in the limits of linear response for perturbations up to about one bond length. We choose moves of size 0.87 Å in all the following calculations.

We next investigate the effect of the direction of the perturbation to the response pattern. We have compared the perturbation-response maps from three different scans: (1) a displacement of 0.5 Å in each of the x-, y-, and z-coordinates is introduced, resulting in an overall displacement of 0.87 Å; (2) same as (1), but with displacements of −0.5 Å; and (3) uniformly distributed randomly selected moves in each of the x-, y-, and z-coordinates that result in a displacement of 0.87 Å. Again, the pattern of Figure 2(a) is captured in all cases, implying that moves are readily localized, and conducted to the rest of the molecule through the same mechanisms. This localization is coordinated by the packing geometry and packing density so that the system responds to an external perturbation with its topology. That the system responds with its intrinsic properties is another indication of linear response. All the calculations presented in the following are obtained with the random moves of case (3) averaged over five different scans. A scan of the 64-residue 1ypc takes an average of 4.5 h of CPU time on a Silicon Graphics O2

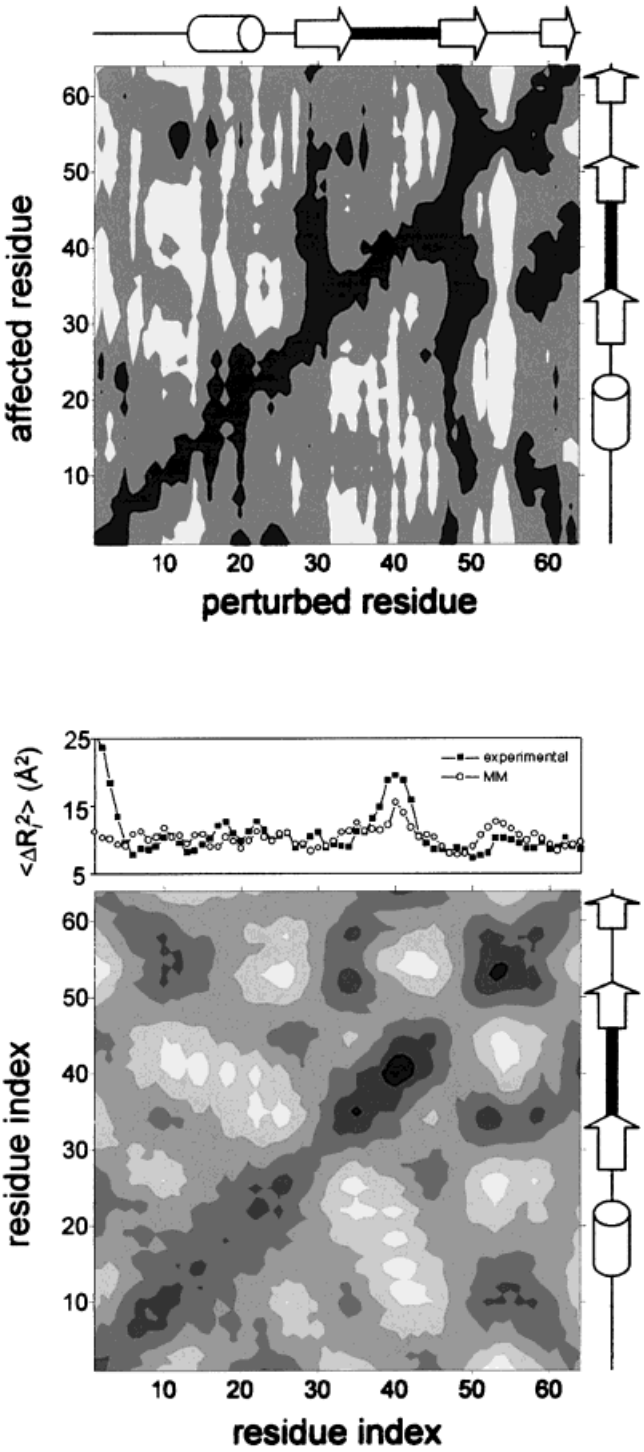

Fig. 2. **a:** Perturbation-response map of unbound CI2. The positions of the secondary structural units are shown along the sides; cylinder indicates α-helix, arrow indicates β-sheet, and a bold line indicates the recognition loop. **b:** Residue–residue correlations of unbound CI2. Autocorrelations (diagonal of the correlation map, ○) are placed along the top and are compared with the X-ray crystallographic data (■). The positions of the secondary structural units are shown along the side.

workstation with a 200-MHz R5000 CPU and 128 Mb RAM; that of the 339-residue 2sni scanned over the 64 CI2 residues is 31.5 h.

Residue Correlations

The experimental and calculated residue autocorrelations of CI2 (eq 4 for $i = j$) are presented at the top of Figure 2(b). The former correspond to the crystallographic temperature factors (filled squares), while the latter are those averaged over the five independent MM scans and scaled to reproduce the average of the experimental data (open circles). The average standard error on the autocorrelation data obtained from the five runs is 2.2%. The peaks and sinks in the experimental data are well reproduced by the MM method except for the first three residues. In fact, since the MM methodology utilizes small displacements followed by energy minimization, we do not expect to capture the large fluctuations of the free N-terminus.

The 2-d correlation map of Figure 2(b), on the other hand, represents the cross-correlations between residue pairs averaged over the five runs; there is a mean error of 2.6% on the data points. Here, white and black correspond to the least and the most correlated regions, respectively. Most of the large correlations arise between regions that are located within the first coordination shell (7 Å) of each other. Correlations of residues 1–12 and 50–64 due to the close proximity of the N- and C-termini, as well as those between the β1–β2, and β2–β3-strands are examples of this. In fact, CI2 is known to expand around this region during the unfolding pathway to the transition states.[40] More interesting is the observed correlations between regions that are located far from each other. By examining Figure 2(b), we see a large correlation between the N-terminus and the turn connecting the α-helix and the β1-strand. Also, the initial part of the recognition loop (residues 33–35) correlates with the turn preceding the α-helix (residues 10–12) as well as the loop adjoining the β2- and β3-strands. Similar short- and long-range cross-correlations were also observed by coarse-grained dynamic Monte Carlo simulations of CI2.[43]

The observation that residues remotely positioned have correlated fluctuations implies a transitivity of interactions: The contact matrix reveals relations between residues such that, for instance, the ith residue may be in contact with the jth that is in turn contacting the kth, but the ith residue is not necessarily in contact with the kth. By contrast, the fluctuations of the ith residue may be correlated with the kth,[10] and this information may be visualized by the correlation maps. The roots of this transitivity, however, lie in the perturbation-response behavior of individual residues.

Overall Response of CI2 to Inserted Perturbations

We now revert to Figure 2(a) and investigate the response of the protein to perturbations inserted at each residue. Hence, L_{ij} is the displacement experienced by residue i after a perturbation at residue j. Figure 2(a) displays the perturbation-response behavior of the molecule as an average over five independent MM runs. Compared with the size of the perturbation of 0.87 Å, white corresponds to a small response $L_{ij} < 0.35$ Å, gray corresponds to a medium-size response within the range

$0.35 \leq L_{ij} < 0.70$ Å, and black corresponds to a large response $L_{ij} \geq 0.70$ Å. The largest response observed is 1.15 Å. By examining this figure, it is possible to visualize the regions where the residues that manipulate the actions of the rest of the molecule are located.

Residues within the β2 region emerge as the center of the structural core among the four conformationally stable regions, the α-helix, and the three β-strands forming the β-sheet. This feature readily arises, since perturbations inserted at residues within this region are able to induce large displacements in essentially all the residues of the protein, identified by the black stripe running along the perturbed residue axis (vertical). In contrast, residues of the α-helix can only affect its close neighbors residing within the helix and the N-terminus packed onto itself. Similarly, perturbations in the β1- and β3-strands give rise to large changes only in their vicinity: themselves and loops 35–45 and 53–58. Additionally, in the case of the β3-strand, the N-terminus is also highly affected due to two stabilizing backbone hydrogen bonds between residues 3–63 and 5–61.[35] Thus, we identify the residues that have the ability to induce large changes at distant positions. Implicitly, this implies that these residues lead to a transitivity of interactions and conduct their effect to remote locations through the directly contacting residues.

We also observe from Figure 2(a) that different loops have different effects on the response. For example, residues 35–45 that reside in the recognition loop are able to induce some medium sized changes in many of the protein residues, although they are located far from much of the protein. In fact, for any given protein structure, we may classify protein residues into four broad groups: Some residues will be located in the protein core, and will have a marked effect on residue stability and function, whereas others at the hydrophobic core with high packing will have less defined effects on stability and/or function. Similarly, some loop regions will be associated with binding sites whereas others will be idle loops. These ideas will be investigated further in the next section.

Packing Density and Residue Correlations

Recently, Nussinov and coworkers have put forth a new view of binding where the binding does not induce a conformational change, but merely induces a redistribution of the equilibrium states.[16,17] We now examine the X-ray structures of CI2 in unbound form and in complex with subtilisin novo to understand how these population shifts occur. Note that "conformationally" the structures attained by CI2 are almost identical in the two states, with an RMSD of C_α carbons of only 0.59 Å.

We first compare the contact number of each residue; this is a measure of the density around each residue, and is computed here as the number of residues residing within a sphere of 7 Å of a given residue. As shown in Figure 3(a), upon binding, the packing density of most of the protein residues remain essentially unchanged, except in the recognition loop which now resides within a region of subtilisin novo. We note that the solvent accessible surface areas (SASA) of the residues behave in exactly the same way; i.e. a change in the SASA is observed only at the recognition loop residues. Here the SASA is defined as the area swept by the center of a spherical solvent probe with radius 1.4 Å as it is rolled over the van der Waals surface of the protein. We use the *naccess* program, which is an implementation of the method of Lee and Richards.[44,45]

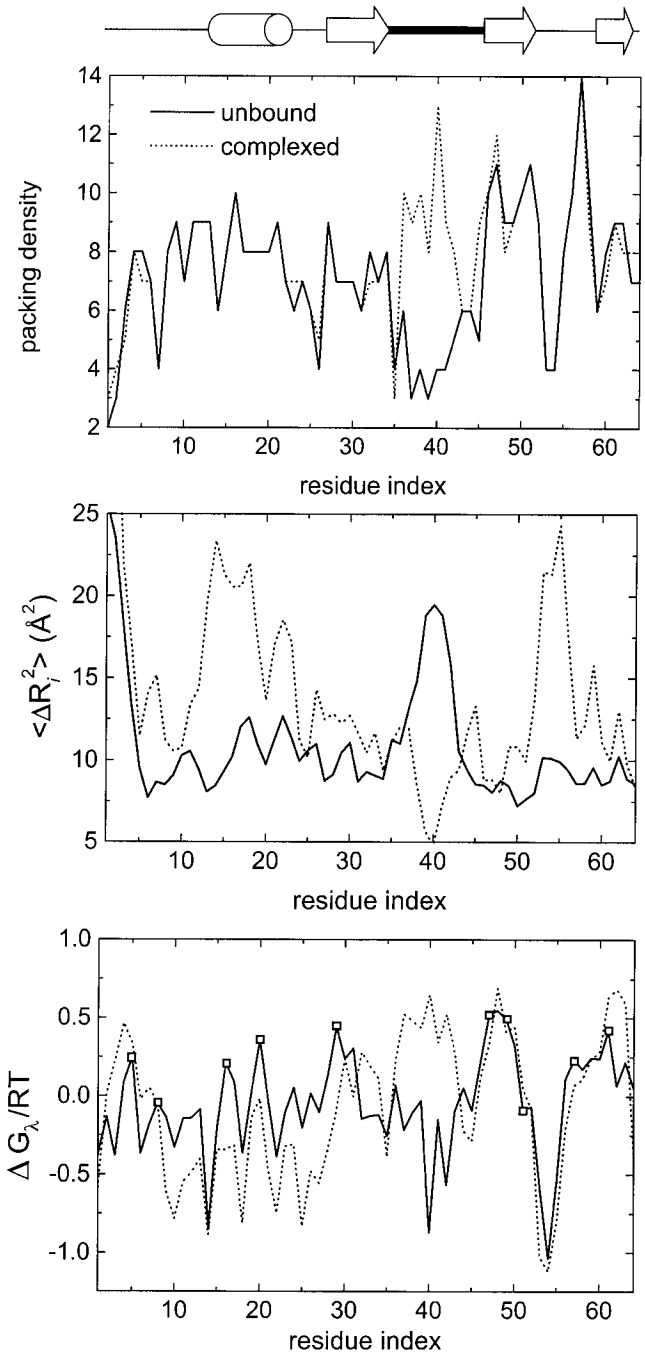

Fig. 3. Comparison of various properties of unbound (solid line) and complexed (dotted line) CI2. **a:** Packing density. **b:** Crystallographic temperature factors. **c:** $\Delta G_\lambda/RT$. The positions of the secondary structural units are shown along the top: cylinder indicates α-helix, arrow indicates β-sheet, a bold line indicates the recognition loop.

The contact matrix carries information on residues that are directly associated with each other. These data, com-

bined with the packing density of each residue, yield most of the indispensable information on residue cross-correlations as follows: Formally, the structural organization of a protein may be arranged in the form of the Kirchhoff matrix:

$$\Gamma_{ij} = \begin{cases} -H(r_c - r_{ij}) & i \neq j \\ -\sum_{i(\neq j)}^{N} \Gamma_{ij} & i = j \end{cases} \quad (5)$$

where the diagonal entries ($i = j$) represent the packing density of the α-carbons and the off-diagonal elements depict the packing geometry. Here, r_{ij} is the distance between the ith and jth α-carbons, H(x) is the Heavyside step function given by H(x) = 1 for $x > 0$ and H(x) = 0 for $x \leq 0$. We take r_c = 7.0 Å in this study. In previous work[29] it was shown that the cross-correlations between residues i and j is related to the Γ matrix by

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle \propto [\Gamma^{-1}]_{ij} \quad (6)$$

Note that the above treatment lacks any information on residue specificity. Yet, the cross-correlation maps such as that of Figure 2(b) of both 1ypc and 2sni obtained in this manner are very similar to those from MM calculations (data not shown).

The positional fluctuations of the residues, or autocorrelations, obtained from X-ray crystallography are displayed in Figure 3(b). These data, whose general trends are also well reproduced by the MM calculations [see also top of Fig. 2(b)] and the methodology outlined in equations 5–6, merely show that the mobility of the recognition loop markedly decreases upon binding, as would be expected. A somewhat less apparent result is that the mobility of most of the other residues increase in the bound form. This increase in mobility is also manifested in the average energy differences between the minimized conformers in the perturbed (E_i for the ith residue) and unperturbed states (E_o), given by $\langle \Delta E \rangle = \Sigma_i (E_i - E_o)/N$. In the unbound form $\langle \Delta E \rangle$ = 1.75 ± 0.28 cal/mol/residue, whereas in the bound form $\langle \Delta E \rangle$ = 2.45 ± 0.22 cal/mol/residue. In other words, perturbations to CI2 in the free state lead to conformers with about 40% less energy difference from the original structure compared with the complex. Such substantial conformational flexibility in a protein complex has also been implicated by Lavigne et al.[46]

The general picture painted in Figure 3(b) indicates that upon binding many of the residues become more flexible, and hence may be more susceptible to structural change. Our MM calculations show that this does not occur in the form of large conformational changes, but rather the roles taken on by certain residues change upon binding. This finding is also corroborated by the work of Freire and coworkers[22,25] and may be the underlying factor of the newly introduced concept by Nussinov and coworkers.[16,17] They state that binding is accompanied by shifts in energy landscapes such that the barrier heights between the sampled substates change upon binding. This is in contrast to the lock-and-key-like mechanisms implied by the induced fit idea, where conformational changes are induced and propagated upon complexation.

It should be noted that, for a protein of N residues described by the inter-residue potential of the coarse grained model given by $\mathcal{H} = \frac{1}{2}\gamma\, tr(\Delta\mathbf{R}^T\, \Gamma\, \Delta\mathbf{R})$, where γ is a single-parameter stiffness constant for representing all pairwise interactions in the folded state, the vibrational entropy of protein residues was shown to be proportional to the inverse of Γ.[30] However, the data in Figure 3(b) are not obtained from the inverse of Figure 3(a), which lacks information on the contact order, i.e., the identity of contacting pairs. In fact, for the vibrational dynamics of folded CI2, Figure 3(a) may be envisaged as conveying information on the potential energy of residues, whereas Figure 3(b) reflects the vibrational entropy. However, the overall stability of a residue depends on the interplay between these two effects.

Elucidating Conformational Shifts Accompanying Binding

To investigate topology/stability/function issues, we define residue stability, λ, as

$$\lambda_i = \frac{\sum_j L_{ij}}{\sum_j L_{ji}} \quad (7)$$

where the denominator represents the overall ability of a given residue to induce a change in the rest of the protein, and the numerator represents the total amount of change that may be induced on the same residue by equal amounts of perturbations inserted in the rest of the protein. L_{ij} are the elements of the perturbation − response matrix obtained from the MM calculations (eq. 2). We then define the free energy associated with residue stability by

$$\Delta G_\lambda = -RT \ln \lambda \quad (8)$$

Thus, idle loops will have $\Delta G_\lambda < 0$, and residues that contribute to stability and function will have $\Delta G_\lambda > 0$. Functional loops will have $\Delta G_\lambda \approx 0$, as they will have relatively large values in both the numerator (they will be easily perturbed as they are located in a region of small packing) and the denominator (they will have some effect on other residues, in connection with their function) of eq. 7. Similarly, hydrophobic core residues that have insignificant contributions to stability or function will have a relatively small value in the denominator (small effect on other residues due to lack of "function") and in the numerator (due to large packing) of eq. 7, and hence, $\Delta G_\lambda \approx 0$.

ΔG_λ are displayed in Figure 3(c) for CI2 in unbound form, and in complex with subtilisin novo. Note that a comparison of the locations of peaks and sinks in Figure 3(c) and residue stabilities measured by H/D exchange[47] reveal a correspondence between the trend followed by the residues. Moreover, in a recent publication on the dynamics of the chemokine family,[10] the residue–residue correlations, $\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle$ obtained by the current MM approach and molecular dynamics simulations have shown remarkable similarities. These results demonstrate that the set of

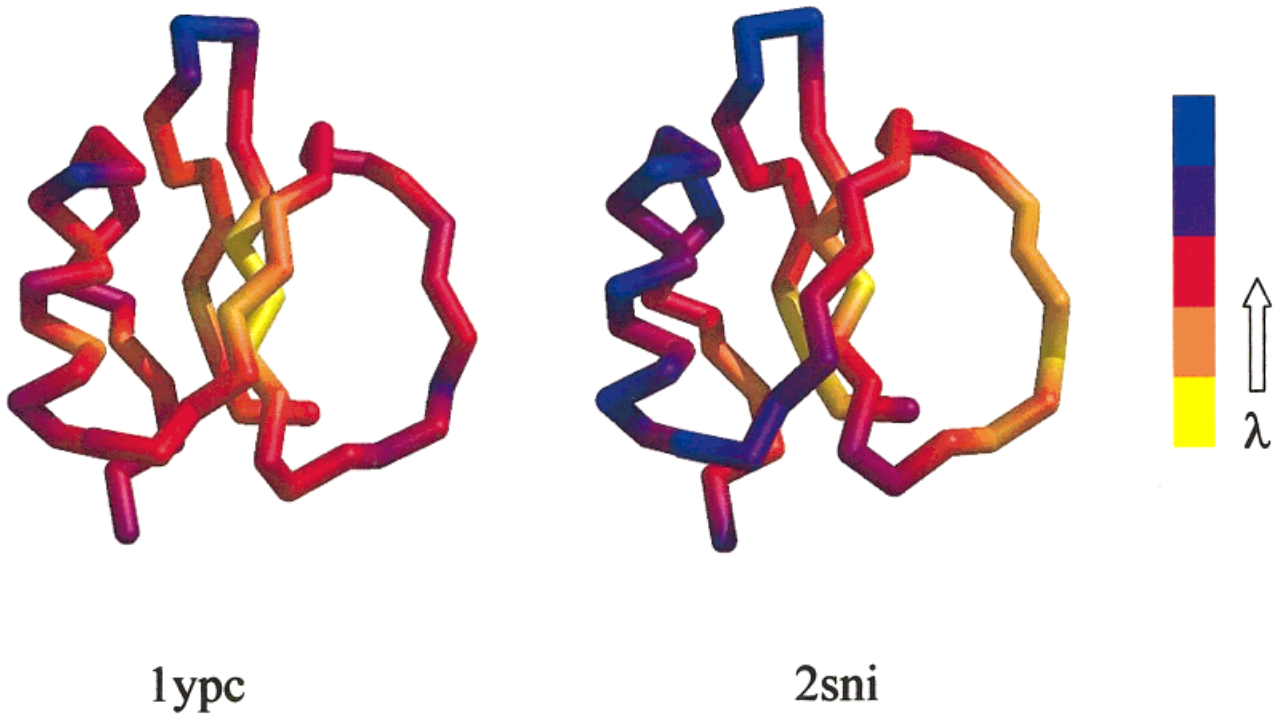

Fig. 4. Backbone of (**a**) 1ypc, and (**b**) 2sni colored according to their λ values. The smallest λ corresponds to residues with the highest ability to induce change and simultaneously resist change, i.e., the functional core residues.

conformers treated in our study are representative of the ensemble under native conditions. The hydrophobic core residues of free CI2 identified by Fersht and coworkers are also marked on the figure by squares.[40] Interestingly, these residues coincide with the peaks of the $\Delta G_\lambda/RT$ curve, except for Leu8 and Val51 that reside at the edge of the hydrophobic core. By contrast, Arg48 has a relatively large ΔG_λ value, but does not reside in the hydrophobic core, implying that it has a major function. In fact, Hilser et al. find Arg 48 together with Leu 49 to energetically link the two subunits of CI2, such that the effects of perturbations on either side are propagated through them.[22]

A comparison of the two curves in Figure 3(c) reveals that upon binding, the stability of the N-terminus, the recognition loop, and part of the β1-strand approaching this loop increases, whereas that of the α-helix and the few residues surrounding it on either side decreases. The region following the β2-strand, on the other hand, does not have a change in its stability. This is contrary to the expectation implied by the substantial mobility gain of this region displayed in Figure 3(b).

To understand the stability/structure relationships, we have colored the CI2 molecule in free and complexed forms according to the λ values (Fig. 4). CI2 in free form has its structural core built around residues 47–49 residing within the β2-strand, Ile29 in the β1-strand, and Pro61 in β3-strand. The α-helix has medium sized stability in this form. The least stable residues, in contrast, are 14 and 54–56. In the complex, residues within the β1- and β3-strands have diminished stability, and the structural core shifts to residues 48, and 61–63 within the β-sheet. In addition, most of the residues in the recognition loop gain stability and controlling residue status. By contrast, residues of the α-helix and the β1-strand become unstable. This is the region, which is known to surround the hydrophobic core. It is also where CI2 is known to expand during the unfolding pathway to the transition states.[40]

CONCLUSIONS

The simple perturbation-response approach used in this work provides a strong tool for following residue interactions in a systematic and well-defined manner. The triad of packing density, residue fluctuations, and residue stability reveal many interesting features of phenomena underlying interactions in proteins. The example case of CI2 is important since it is in the same overall conformation in unbound and complexed forms and allows us to study the topology/stability relationships free of this variable.

In CI2 the only region with a substantial packing density change is that of the recognition loop [Fig. 3(a)], which is in direct contact with the substrate, subtilisin novo. The latter is much larger than CI2 (275 residues versus 64 of CI2), but approaches to CI2 only in the residue 35–44 region. The response of CI2 to this entrapment is not in the form of a conformational change, but rather as substantial elevations in residue fluctuations in the rest of the protein [Fig. 3(b)]. Thus, CI2 shakes itself like a mouse caught in a trap by one of its limbs—the limb is immobile, but the rest of the animal moves frantically. However, as far as stability is concerned, larger fluctuations do not necessarily lead to less stable regions [Fig. 3(c)]. The general trend followed by residues is not largely affected,

except in the binding loop; in particular, consecutive residues residing in the α-helical region displays alternating gain and loss in stability, but the value of ΔG_λ is reduced considerably here. Most interestingly, the C-terminus region following the recognition loop, which gains enormous flexibility in the loop adjoining the β2- and β3-strands does not show any change in its stability. The loss in stability, however, does not necessarily lead to a loss of structure. It merely implies that there is a shift in the residues that are involved in communicating incoming signals in the two forms.

Going back to the caught mouse analogy, the "conformations" sampled by body parts have different weights for the trapped and free mouse. Parts of the mouse not trapped are now moving more than they did when the mouse was freely running, but different body parts may be responding differently. For example, the head which is rather constrained is moving in the directions that it did earlier, but more vigorously this time. The tail, which is rather free to move under both circumstances, is still taking those shapes, but they may be changing from one "conformation" to another more briskly. The spine, on the other hand, may be curving in a manner it did not do so before. This, however, does not mean that the mouse could not curl its spine into that shape earlier, but merely that it did not need to do so. The shapes each body part of the mouse can take is still bounded by the mouse anatomy.

Residue stability may be thought of as the inclination of residues to locally unfold–refold. Thus, the ΔG_λ values provide an idea on the conformations sampled by the molecule around the "global" minima represented by the minimized folded structure in free and bound forms. These ideas are directly related to the dynamic landscapes and population shifts concept put forth by Nussinov and coworkers.[16,17] If binding is accompanied by shifts in energy landscapes such that the barrier heights between the sampled substates change upon binding, the picture of the molecule in Figure 4 shows the propensity of various regions to readjust their barrier heights.

REFERENCES

1. Baker D. A surprising simplicity to protein folding. Nature 2000;405:39–42.
2. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;277:985–994.
3. Dinner AR, Karplus M. The roles of stability and contact order in determining protein folding rates. Nature Struct Biol 2001;8:21–22.
4. Crump MP, Gong J-H, Loetscher P, Rajarathnam K, Amara A, Arenzana-Seisdesos F, Virelizier J-L, Baggiolini M, Sykes BD, Clark-Lewis I. Solution structure and basis for functional activity of stromal cell-derived factor—1; dissociation of CXCR4 activation from binding and inhibition of HIV-1. EMBO J 1997;16:6996–7007.
5. Frauenfelder H, McMahon B. Dynamics and function of proteins: The search for general concepts. Proc Natl Acad Sci USA 1998;95:4795–4797.
6. Zavodszky P, Kardos J, Svingor A, Petsko GA. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. Proc Natl Acad Sci USA 1998;95:7406–7411.
7. Bahar I, Erman B, Jernigan RL, Atilgan AR, Covell DG. Collective dynamics of HIV-1 reverse transcriptase: Examination of flexibility and enzyme function. J Mol Biol 1999;285:1023–1037.
8. Feher V, Cavanagh J. Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F. Nature 1999;400:289–293.
9. Yilmaz LS, Atilgan AR. Identifying the adaptive mechanism in globular proteins: Fluctuations in densely packed regions manipulate flexible parts. J Chem Phys 2000;113:4454–4464.
10. Baysal C, Atilgan AR. Elucidating the structural mechanisms for biological activity of the chemokine family. Proteins 2001;43:150–160.
11. Fersht AR. Transition-state structure as a unifying basis in protein folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism. Proc Natl Acad Sci USA 2000;97:1525–1529.
12. Baldwin R, Rose G. Is protein folding hierarchic? I. Local structure and peptide folding. Trends Biochem Sci 1999;24:26–33.
13. Baldwin R, Rose G. Is protein folding hierarchic? II. Folding intermediates and transition states. Trends Biochem Sci 1999;24:77–83.
14. Rumbley J, Hoang L, Mayne L, Walter S. An amino acid code for protein folding. Proc Natl Acad Sci USA 2001;98:105–112.
15. Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M. Understanding protein folding via free-energy surfaces from theory and experiment. Trends Biochem Sci 2000;25:331–339.
16. Tsai C-J, Ma B, Nussinov R. Folding and binding cascades: Shifts in energy landscapes. Proc Natl Acad Sci USA 1999;96:9970–9972.
17. Kumar S, Ma B, Tsai C-J, Sinha N, Nussinov R. Folding and binding cascades: Dynamic landscapes and population shifts. Protein Sci 2000;9:10–19.
18. Baysal C, Meirovitch H. Determination of the stable microstates of a peptide from NOE distance constraints and optimization of atomic solvation parameters. J Am Chem Soc 1998;120:800–812.
19. Baysal C, Meirovitch H. Ab initio prediction of the solution structures and populations of a cyclic pentapeptide in DMSO based on an implicit solvation model. Biopolymers 2000;53:423–433.
20. Olejniczak EJ, Zhou M-M, Fesik SW. Changes in the NMR derived motional parameters of the insulin receptor substrate 1 phosphotyrosine binding domain upon binding to an interleukin 4 receptor phosphopeptide. Biochemistry 1997;36:4118–4124.
21. Monaco-Malbet S, Berthet-Colominas C, Novelli A, Battai N, Piga N, Cheynet V, Mallet F, Cusack S. Mutual conformational adaptations in antigen and antibody upon complex formation between a Fab and HIV-1 capsid protein p24. Structure 2000;8:1069–1077.
22. Hilser VJ, Dowdy D, Oas TG, Freire E. The structural distribution of cooperative interactions in proteins: Analysis of the native state ensemble. Proc Natl Acad Sci USA 1998;95:9903–9908.
23. Hilser VJ, Freire E. Structure-based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection factors. J Mol Biol 1996;262:756–772.
24. Todd MJ, Semo N, Freire E. The structural stability of the HIV-1 protease. J Mol Biol 1998;283:475–488.
25. 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.
26. Luque I, Freire E. Structural stability of binding sites: consequences for binding affinity and allosteric effects. Proteins 2000; suppl 4:63–71.
27. Baysal C, Meirovitch H. New conformational search method based on local torsional deformations for cyclic molecules, loops in proteins, and dense polymer systems. J Chem Phys 1996;105:7868–7871.
28. Baysal C, Meirovitch H. Efficiency of the local torsional deformations method for identifying the stable structures of cyclic molecules. J Phys Chem A 1997;101:2185–2191.
29. 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.
30. Bahar I, Atilgan AR, Demirel MC, Erman B. Vibrational dynamics of folded proteins: significance of slow and fast modes in relation to function and stability. Phys Rev Lett 1998;80:2733–2736.
31. Demirel MC, Atilgan AR, Jernigan RL, Erman B, Bahar I. Identification of kinetically hot residues in proteins. Protein Sci 1998;7:2522–2532.
32. Doruker P, Atilgan AR, Bahar I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to α-amylase inhibitor. Proteins 2000;40:512–524.
33. 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.
34. Itzhaki LS, Otzen DE, Fersht AR. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. J Mol Biol 1995;254:260–288.
35. Li A, Daggett V. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulation. J Mol Biol 1996;257:412–429.
36. Harpaz Y, Elmasry N, Fersht AR, Henrick K. Direct observation of better hydration at the N-terminus of an α-helix with glycine rather than alanine as the N-cap residue. Proc Natl Acad Sci USA 1994;91:311–315.
37. Abola EE, Bernstein FC, Bryant SH, Koetzle TF, Weng J, Abola EE, Bernstein FC, Bryant SH, Koetzle TF, Weng J, editors. Data Commission of the International Union of Crystallography: Bonn, Cambridge and Chester; 1987. p 107.
38. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanovichi T, Tasumi M. The Protein Data Bank: A computer based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
39. McPhalen CA, Svendsen I, Jonassen I, James MNG. Crystal and molecular structure of chymotrypsin inhibitor 2 from barley seeds in complex with subtilisin Novo. Proc Natl Acad Sci USA 1985;82: 7242–7246.
40. Daggett V, Li A, Itzhaki LS, Otzen DE, Fersht AR. Structure of the transition state for folding of a protein derived from experiment and simulation. J Mol Biol 1996;257:430–440.
41. Gay GD, Ruiz-Sanz J, Neira JL, Corrales FJ, Otzen DE, Ladurner AG, Fersht AR. Conformational pathway of the polypeptide chain of chymotrypsin inhibitor-2 growing from its N terminus in vitro. Parallels with the protein folding pathway. J Mol Biol 1995;254: 968–979.
42. Dauber-Osguthorpe P, Roberts VA, Osguthorpe DJ, Wolff J, Genest M, Hagler AT. Structure and energetics of ligand binding to proteins: *E. coli* dihydrofolate reductase-trimethoprim, a drug-receptor system. Proteins 1988;4:31–47.
43. Kurt N, Haliloglu T. Conformational dynamics of Chymotrypsin inhibitor 2 by coarse-grained simulations. Proteins 1999;37:454–464.
44. Lee B, Richards F. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.
45. Hubbard S, Campbell S, Thornton J. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. J Mol Biol 1991;20:507–530.
46. Lavigne P, Bagu JR, Boyko R, Willard L, Holmes CFB, Sykes BD. Structure-based thermodynamic analysis of the dissociation of protein phosphatase-1 catalytic subunit and microcystin-LR docked complexes. Protein Sci 2000;9:252–264.
47. Itzhaki LS, Neira JL, Fersht AR. Hydrogen exchange in chymotrypsin inhibitor 2 probed by denaturants and temperature. J Mol Biol 1997;270:89–98.