Effect of Protein Backbone Folding on the Stability of Protein−Ligand Complexes

Ernesto Estrada,* Eugenio Uriarte, and Santiago Vilar†

Complex Systems Research Group, X-ray Unit, RIAIDT, Edificio CACTUS, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain, and Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain

Received September 20, 2005

The role played by the degree of folding of protein backbones in explaining the binding energetics of protein−ligand interactions has been studied. We analyzed the protein/peptide interactions in the RNase-S system in which amino acids at two positions of the peptide S have been mutated. The global degree of folding of the protein S correlates in a significant way with the free energy and enthalpy of the protein-peptide interactions. A much better correlation is found with the local contribution to the degree of folding of one amino acid residue: Thr36. This residue is shown to have a destabilizing interaction with Lys41, which interacts directly with peptide S. Another system, consisting of the interactions of small organic molecules with HIV-1 protease was also studied. In this case, the global change in the degree of folding of the protease backbone does not explain the binding energetics of protein−ligand interactions. However, a significant correlation is observed between the free energy of binding and the contribution of two amino acid residues in the HIV-1 protease: Gly49 and Ile66. In general, it was observed that the changes in the degree of folding are not restricted to the binding site of the protein chain but are distributed along the whole protein backbone. This study provides a basis for further consideration of the degree of folding as a parameter for empirical structural parametrizations of the binding energetics of protein folding and binding.

Keywords: protein degree of folding • protein−peptide interactions • protein-inhibitors interactions • RNase-S • HIV-1 protease • binding energetics • graph theory methods

I. Introduction

The role played by the three-dimensional structure in the binding energetics of protein−ligand complexes is of fundamental importance from both a theoretical and a practical point of view.1,2 On one hand, it permits an understanding of the biophysical and chemical mechanisms involved in these interactions,3 and on the other hand, it is a key element in structure-based drug design.4 General insights on protein−ligand interactions can be obtained by studying how the molecular structure influences the thermodynamic parameters of the binding. Such insights can be of great value in the design of new probes in chemical proteomics, which seeks to define protein function and mechanism at the level of directly observable protein−ligand interactions.5 A number of methods for the estimation of thermodynamic parameters of protein−ligand interactions are based on accessible area parametrizations, which commonly characterize the hydrophobic contribution to the binding free energy.1,2,6,7 widely recognized as playing an important role in protein stability.8,9 For instance, it has been shown that the binding of inhibitors of HIV-1 protease is dominated by the hydrophobic effect according to the structural parametrization of the binding and folding developed by Freire et al.10 These authors have shown that upon binding, not only the inhibitor but also those protease residues located at the binding pocket bury a significant proportion of nonpolar surface, restricting solvent access to it. However, recent findings point to the fact that the loss of packing interactions rather than the hydrophobic effect can be a very significant contributor to the binding thermodynamics in protein−ligand interactions.11,12 It was observed that the change in enthalpy and free energy associated with the peptide/protein complex ribonuclease S, correlate better with changes in packing parameters such as residue depth and occluded surface area than with the change in accessible surface area upon folding.11

In general, not only hydrophobic and packing interactions but also differences in hydrogen bonding, van der Waals and electrostatic interactions as well as conformational changes determine the binding thermodynamics of protein−ligand interactions.13 Consequently, the investigation of other structural characteristics accounting for some of these types of interactions is necessary for a better understanding of the binding of ligands to proteins.

One of the structural parameters that could have a direct role in protein−ligand interactions is the change produced in
II. Methods

The protein chain degree of folding index used here is based on the torsion angles of the protein backbone chain, i.e., the so-called $\phi$, $\psi$, and $\omega$ torsion angles. $\phi_i$ describes the rotation about $N_i-C\alpha_i$ bond, $\psi_i$ describes the rotation about the $C\alpha_i-C_i$ bond and $\omega_i$ defines the rotation about the $C_i-N_{i+1}$ peptide bond. Figure 1 shows a portion of a protein backbone indicating these torsion angles.

The index $I_3$ measuring the degree of folding of a protein backbone expresses the qualitative criterion of backbone-folding on a quantitative scale, in which two proteins can be differentiated by measuring the degree of folding of their backbone chains. It is defined as follows. Let $A$ be the adjacency matrix of this chain representing the adjacency of torsion angles in the protein backbone and let $T$ be a diagonal matrix of the cosines of $\phi_i$, $\psi_i$, and $\omega_i$ angles. Define $B = A + T$ as a matrix representing the protein backbone, which has the spectrum $\sigma = (\lambda_1, \lambda_2, \ldots, \lambda_N)$, where $t = N - 3$ is the number of $\phi$, $\psi$, and $\omega$ torsion angles and $N$ is the number of atoms in the protein backbone. Then the degree of folding index $I_3$ is defined as (see ref 17 for details)

$$I_3 = \frac{1}{N-3} \sum_{j=1}^{N-3} \phi_j$$

(1)

The index $I_3$ represents a global characterization of protein backbone degree of folding through the sum of contributions coming from individual torsion angles and combinations of contiguous torsion angles.

The local contribution of amino acids to the global degree of folding of the protein is defined as follows. Let $i$ denote an amino acid of the protein with torsion angles $\phi_i$, $\psi_i$, and $\omega_i$. Then the contribution of this amino acid to the global protein chain degree of folding can be expressed as follows (see ref 16 and Appendix of ref 19 for details)

$$I_{ij}(i) = \sum_{j=1}^{N} v_j^i \phi_j$$

(2)

where $v_1, v_2, \ldots, v_N$ is an orthonormal basis of $R^N$ composed by eigenvectors of $B$ associated with the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$. On the other hand, $v_j^i$ and $v_j^\phi$ denote the component of $v_j$ corresponding to the torsion angles $\psi$ and $\phi$ of the $i$th amino acid. Larger values of $I_{ij}$ or $I_{ij}(i)$ indicate that the chain is more folded. More than 150 proteins of different sizes and types have been studied and the influence of protein size, crystal packing, temperature, pH, and secondary structure on their values have been analyzed.

The strategy that we use here to find the contribution of the degree of folding of a specific amino acid to the binding energetics of protein-peptide or protein-drug interaction is described as follows. First, we calculate the contribution of all amino acids in the protein to the global degree of folding by using expression 2. Then we use stepwise linear regression analysis to find a regression model that describes the binding energetics parameter in terms of the folding of some specific amino acids. In this way only those variables, i.e., $I_{ij}(i)$, which significantly influence the binding energetic of the complexes studied are included into the linear regression model. These models are of the following form: $P = a_0 + a_1I_{ij}(1) + a_2I_{ij}(2) + \ldots + a_nI_{ij}(n)$, where the coefficients $a_i$ are obtained from the linear regression analysis.

Estrada et al. (2006) Journal of Proteome Research

II. Results and Discussion

Protein–Peptide Interactions in RNase-S. The protein-peptide complex RNase-S is obtained by cleavage of bovine pancreatic ribonuclease A (RNase A) with subtilisin to give the $S$ protein and the $S$ peptide. These two fragments can be differentiated by measuring the degree of folding of their backbone chains. It is defined as follows. Let $A$ be the adjacency matrix of this chain representing the adjacency of torsion angles in the protein backbone and let $T$ be a diagonal matrix of the cosines of $\phi, \psi$, and $\omega$ angles. Define $B = A + T$ as a matrix representing the protein backbone, which has the spectrum $\sigma = (\lambda_1, \lambda_2, \ldots, \lambda_N)$, where $t = N - 3$ is the number of $\phi, \psi$, and $\omega$ torsion angles and $N$ is the number of atoms in the protein backbone. Then the degree of folding index $I_3$ is defined as (see ref 17 for details)

$$I_3 = \frac{1}{N-3} \sum_{j=1}^{N-3} \phi_j$$

(1)

The index $I_3$ represents a global characterization of protein backbone degree of folding through the sum of contributions coming from individual torsion angles and combinations of contiguous torsion angles.

The local contribution of amino acids to the global degree of folding of the protein is defined as follows. Let $i$ denote an amino acid of the protein with torsion angles $\phi_i$, $\psi_i$, and $\omega_i$. Then the contribution of this amino acid to the global protein chain degree of folding can be expressed as follows (see ref 16 and Appendix of ref 19 for details)

$$I_{ij}(i) = \sum_{j=1}^{N} v_j^i \phi_j$$

(2)

where $v_1, v_2, \ldots, v_N$ is an orthonormal basis of $R^N$ composed by eigenvectors of $B$ associated with the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$. On the other hand, $v_j^i$ and $v_j^\phi$ denote the component of $v_j$ corresponding to the torsion angles $\psi$ and $\phi$ of the $i$th amino acid. Larger values of $I_{ij}$ or $I_{ij}(i)$ indicate that the chain is more folded. More than 150 proteins of different sizes and types have been studied and the influence of protein size, crystal packing, temperature, pH, and secondary structure on their values have been analyzed.

The strategy that we use here to find the contribution of the degree of folding of a specific amino acid to the binding energetics of protein-peptide or protein-drug interaction is described as follows. First, we calculate the contribution of all amino acids in the protein to the global degree of folding by using expression 2. Then we use stepwise linear regression analysis to find a regression model that describes the binding energetics parameter in terms of the folding of some specific amino acids. In this way only those variables, i.e., $I_{ij}(i)$, which significantly influence the binding energetic of the complexes studied are included into the linear regression model. These models are of the following form: $P = a_0 + a_1I_{ij}(1) + a_2I_{ij}(2) + \ldots + a_nI_{ij}(n)$, where the coefficients $a_i$ are obtained from the linear regression analysis.

II. Results and Discussion

Protein–Peptide Interactions in RNase-S. The protein-peptide complex RNase-S is obtained by cleavage of bovine pancreatic ribonuclease A (RNase A) with subtilisin to give the $S$ protein and the $S$ peptide. These two fragments can be reconstituted to give rise to RNase-S, which is catalytically active and has a structure very similar to that of RNase-A. The $S$ peptide consists of the first 20 amino acids of RNase-A, but it has been shown that a truncated version formed by residues 1–15 formed a complex with protein S which is structurally identical with RNase-S.11 There are two hydrophobic residues in the $S$ peptide that contribute significantly to the stability of RNase-S. These residues, methionine 13 (M13) and phenylalanine 8 (F8), are buried inside the RNase-S core.27–28 Mutation experiments have led to their replacement by several other
smaller hydrophobic amino acids of different sizes resulting in mutated S peptides of the type F8X and M13X, where X represents alanine (A), methionine (M), norleucine (Nle), α-aminobutyric acid (ANB), valine (V), leucine (L) and isoleucine (I).11 The structures of the RNase-S complexes with S peptide mutants were determined using X-ray crystallography and the free energies (ΔG°) and enthalpies (ΔH°) of the S-peptide–protein binding were determined using titration calorimetry (see ref 11 for details). In Table 1 we show the values of the differences in the thermodynamic parameters upon mutation: ΔΔG° and ΔΔH°, e.g., ΔΔG° = ΔG° (mutant) – ΔG° (wild type).

Using the method of Ratnaparkhi and Varadarajan11 who correlated ΔΔG° and ΔΔH° with the global structural characteristics of protein S we calculated the change in the degree of folding of protein S upon binding: ΔI3(Thr36) = I3(mutant) – I3(wild type). The change in free energy ΔΔG° upon binding shows a good correlation (r = 0.87) with the global degree of folding index of protein S, whereas ΔΔH° shows a slightly poorer correlation (r = 0.82). These correlations observed for thermodynamic properties using the degree of folding index of protein S are significantly better than those obtained using other structural parameters such as change in cavity volume and change in accessible surface area. However, the reported correlations with other global packing parameters, such as the occluded surface area29 and the protein depth,30 show correlation coefficients in excess of 0.9.11

To explore whether local contributions of amino acids correlate with the thermodynamic parameters of binding, we calculated the degree of folding of all residues in protein S. Using stepwise linear regression analysis, we obtained the best linear fits between ΔΔG° and ΔΔH°, which resulted in correlations with the contribution of threonine-36 to the degree of folding: ΔΔG° = −1.058 −11.203ΔI3(Thr36) and ΔΔH° = −5.856 −45.884ΔI3(Thr36). As can be seen in Table 1, when using local contributions to the degree of folding the correlation coefficients improve significantly, to >0.94 for both thermodynamic parameters.

Due to the great heterogeneity in the values of the resolution reported in the crystallographic structures of these complexes it is necessary to analyze the significance of these correlations from the statistical point of view. The wild type (wt) structure (1mrv) has been solved at 1.6 Å, whereas there are structures, such as 1d5d, 1d5e, and 1d5h, which have been solved at 2.25 Å resolution. We have previously shown that the resolution and refinement protocol used to solve the protein structures have little influence on the values of I3 and are unable to explain the variance in the values of this parameter for a series of 150 proteins.17 However, we have not tested whether the resolution influences the values of I3(Thr36) which are used here to describe the binding energetics of protein-peptide complexes. We have used a protocol to determine in which extension the models obtained here can be obtained by chance or are influenced by the resolution at which the crystals were solved. First, we select the value of I3(Thr36) for the wt as the basis for random generation of the values for the other proteins, which are obtained as: I3(Thr36) = I3(Thr36)wt + Rnd, where Rnd is a number, between 0 and 1, generated at random. This procedure simulates a completely random generation of the values of I3(Thr36) to see whether the correlations found here are obtained by chance. The second procedure goes further and analyzes the influence of resolution by considering again the values of I3(Thr36) generated at random by I3(Thr36) = I3(Thr36)wt + Rnd, but taking identical values for those proteins which were solved at the same resolution. In this way, we generate four different random values of I3(Thr36), which correspond to the four groups of proteins solved at identical resolution: 1brg-1bri (1.8 Å), 1brh-1brd (1.7 Å), 2ln (1.85 Å) and 1d5d-1d5e-1d5f (2.25 Å). Using these two strategies, we calculate the correlation coefficient for the linear regression models between the values of the binding thermodynamic parameters and the random I3(Thr36) values. After 104 random realizations we obtained average correlation coefficients of 0.315 and 0.312 for both kind of approaches, well below the values obtained in this work, which are over 0.9. Thus, the probability that the models obtained here can be obtained by chance or are influenced by the resolution at which the crystals were solved is very remote.

Thr36 is located in a loop region below the binding site of the protein and does not interact directly with peptide S. A close inspection of the region around Thr36 reveals that it has a hydrophobic–hydrophilic destabilizing interaction with Lys41, which is located in the same loop at a distance of 3.9 Å from Thr36. Lys41 is located at the bottom of the binding site of protein S, which permits it to interact directly with Gln11 and His12 located in the peptide S. The interaction between Lys41 and Gln11 produced a destabilizing effect on the protein/peptide RNase-S complex due to its hydrophilic–hydrophobic nature. In Figure 2A, we illustrate the interaction involving Gln11 of the peptide S with Lys41 in protein S, residues separated by distances ranging from 3.0 Å (Nle13M) to 3.8 Å (Nle13I). The interaction between Lys41 and Thr36 is also illustrated in this Figure.

To gain insight into the effect of Lys41 on the conformation of Thr36 we have investigated the structural characteristics resulting from the interaction between these two amino acids. The distance between these two residues in the different RNase complexes studied is almost constant and does not offer any explanation for the change in the degree of folding of Thr36. However, the contact surface area (CSA)31 between these two residues shows a relationship with the values of the degree of folding of Thr36. A small value of CSA, well below the values observed for complexes with mutated peptides, is observed for the wild type RNase-S (Table 1). Hence, the increment in CSA in the complexes with mutated peptides can only be produced by the new interactions that arise from the large to small substitutions in the peptide S. In Figure 3, we can see that the de-folding produced in Thr36 is a consequence of the incre-
ment of the CSA between this amino acid and Lys41. Consequently, the relationship between the binding thermodynamics and the change in the degree of folding of Thr36 is a consequence of the increment in destabilizing interaction of this amino acid with Lys41. This increment in the interaction between Thr36 and Lys41 is produced by the direct interaction of Lys41 with peptide S.

Finally, there are some important points that should be noted. The first is that the models obtained here by using the degree of folding illustrate that this parameter is an important factor in explaining the binding thermodynamics of RNase-S complexes. The model explaining the change in binding enthalpy is one of the few examples of correlations between this thermodynamic parameter and structural characteristics. The first clear correlation was the one obtained by Ratnaparkhi and Varadarajan using the occluded surface area.\textsuperscript{11} Further, it is remarkable that the binding free energy and enthalpy are distributed through the protein S structure extending beyond the binding site in a correlated way such that the change in some residues in the binding site induce changes in the folding of others outside the binding site.

**Binding of Synthetic Inhibitors to HIV-1 Protease.** The HIV-1 protease plays a key role in viral maturation and is an important target for development of drugs that can protect against HIV-1 infection. The crystallographic structure of the HIV-1 protease, determined at high resolution, shows that it consists of two identical subunits of 99 residues each.\textsuperscript{32} Since 1995, seven inhibitors of HIV-1 protease have been approved by the FDA. In combination with nucleoside inhibitors or in mixtures of different protease inhibitors\textsuperscript{33} these inhibitors have shown promising results in AIDS patients. They bind to the active site of the viral protease preventing polyprotein processing and maturation to infectious virions. The binding thermodynamics of 13 HIV-1 protease/inhibitors complexes, for which high-resolution X-ray crystallographic structures have been reported, have been used for structural parametrization of the binding and folding energetics of the protein-drug interactions\textsuperscript{10} as well as in a symmetry-based structure—activity relationship study.\textsuperscript{34} In Figure 4 we illustrate the molecular structures for the 12 inhibitors involved in these 13 complexes with the HIV-1 protease (there are two complexes with the same inhibitor but in one a mutated protease is used, see further).

In Table 2 we give the values of the indices of the degree of folding of chains A and B in HIV protease. The values of global degree of folding of either chain of HIV-1 protease do not correlate with the binding affinities of inhibitors. In both cases the correlation coefficient is lower than 0.5, but a somewhat better trend is seen in the degree of folding of chain A ($r = 0.49$) than in that of chain B ($r = 0.32$). These results indicate that the binding of these 13 inhibitors to the HIV-1 protease do not affect the degree of folding of the global protein in such a way as to account for the binding energetics: the free energy of binding is not affected by the changes produced in the folding of the chains as a whole. However, this binding energy could be determined by the change in folding of some specific amino acid residues within these chains. The local contributions to the degree of folding of the different residues in the protease chains were used to correlate with the free energy of
Table 2. Degree of Folding Indices of Chains A and B of the HIV-1 Protease and Local Contributions to Folding of the Two Amino Acid Residues that Play a Role in Explaining the Free Energy of Binding ($\Delta G_{\text{calcd}}$) of 13 Protein/Inhibitor Complexes

<table>
<thead>
<tr>
<th>PDB entry</th>
<th>inhibitor</th>
<th>$I_1$(A)</th>
<th>$I_1$(B)</th>
<th>$I_1$(Gly49)</th>
<th>$I_1$(Ile66)</th>
<th>$\Delta G_{\text{calcd}}$ (cal mol$^{-1}$)</th>
<th>$\Delta G_{\text{calcd}}$ (cal mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1pro</td>
<td>A98881</td>
<td>2.5713</td>
<td>2.5575</td>
<td>4.5648</td>
<td>4.5670</td>
<td>−15417</td>
<td>−14021</td>
</tr>
<tr>
<td>1hvj</td>
<td>A78791</td>
<td>2.5801</td>
<td>2.5908</td>
<td>4.5466</td>
<td>4.5678</td>
<td>−14299</td>
<td>−13334</td>
</tr>
<tr>
<td>1hvk</td>
<td>A76928</td>
<td>2.5716</td>
<td>2.5885</td>
<td>4.5649</td>
<td>4.5674</td>
<td>−13796</td>
<td>−14039</td>
</tr>
<tr>
<td>1hvi</td>
<td>A77003</td>
<td>2.5822</td>
<td>2.5880</td>
<td>4.5646</td>
<td>4.5673</td>
<td>−13699</td>
<td>−13538</td>
</tr>
<tr>
<td>1hvs</td>
<td>A77003</td>
<td>2.5510</td>
<td>2.5497</td>
<td>4.5648</td>
<td>4.5666</td>
<td>−12914</td>
<td>−14184</td>
</tr>
<tr>
<td>1hps</td>
<td>VX478</td>
<td>2.5853</td>
<td>2.5735</td>
<td>4.5640</td>
<td>4.5676</td>
<td>−12580</td>
<td>−12332</td>
</tr>
<tr>
<td>1hpb</td>
<td>SB206343</td>
<td>2.5567</td>
<td>2.5519</td>
<td>4.5626</td>
<td>4.5641</td>
<td>−12580</td>
<td>−11227</td>
</tr>
<tr>
<td>1hvl</td>
<td>A76889</td>
<td>2.5723</td>
<td>2.5848</td>
<td>4.5642</td>
<td>4.5672</td>
<td>−12278</td>
<td>−12856</td>
</tr>
<tr>
<td>1hhb</td>
<td>CGP53820</td>
<td>2.5562</td>
<td>2.5351</td>
<td>4.5633</td>
<td>4.5660</td>
<td>−10976</td>
<td>−11719</td>
</tr>
<tr>
<td>1hsb</td>
<td>SB203386</td>
<td>2.5605</td>
<td>2.5467</td>
<td>4.5636</td>
<td>4.5671</td>
<td>−10565</td>
<td>−11813</td>
</tr>
<tr>
<td>1u7p</td>
<td>U100313</td>
<td>2.5529</td>
<td>2.5522</td>
<td>4.5631</td>
<td>4.5686</td>
<td>−10077</td>
<td>−9893</td>
</tr>
<tr>
<td>1igno</td>
<td>U89360</td>
<td>2.5618</td>
<td>2.5926</td>
<td>4.5622</td>
<td>4.5658</td>
<td>−9441</td>
<td>−9813</td>
</tr>
<tr>
<td>1hsv</td>
<td>SB203238</td>
<td>2.5600</td>
<td>2.5478</td>
<td>4.5635</td>
<td>4.5747</td>
<td>−8685</td>
<td>−8540</td>
</tr>
</tbody>
</table>

*The values of the free energy of binding calculated ($\Delta G_{\text{calcd}}$) from local contributions to the degree of folding are also reported.

Figure 4. Molecular structures of the 13 inhibitors of HIV-1 protease studied here.

The model reported by Baldi et al.\textsuperscript{10} based on a structural parametrization of protein folding and binding shows a correlation coefficient of 0.88 and a mean square error of 310 cal/mol for the last set of complexes, i.e., excluding 1hvs. An improved empirical binding energy function which uses the boundary element algorithm to evaluate the electrostatic solvation energy showed a poor correlation with the free energy of binding for these inhibitors due to the very poor calculated energy for the complex 1hhs.\textsuperscript{35} The correlation coefficient obtained by this method after the elimination of this complex is only 0.63 with an error of 2500 cal/mol, which is 10 times larger than that obtained here.\textsuperscript{35} The plot of observed versus calculated free energy of binding for these inhibitors is given in Figure 5.

According to our model, there are two amino acids that play a crucial role in the binding thermodynamics of these inhibitors to HIV-1 protease. The first, Gly49 which, together with Met46, Ile47, Gly48 and Ile50 defines the top of the binding site of the HIV-1 protease, is located at the tip of the flap (see Figure 3B). Gly49 is at the middle of the flap formed by residues 40–60, which is known to be important for the energetics of protein-inhibitor interactions.\textsuperscript{36} Although the flap region is characterized in the native state by very low stability constants per residue, it is stabilized by its interactions with inhibitors. In particular residues Met46, Ile47, Gly48, Gly49 and Ile50 are largely unstructured before binding and are forced into a unique conformation when interacting with the inhibitor.\textsuperscript{36}

On the other hand, the degree of folding contribution of the second crucial residue, Ile66 to the free energy of interaction of the protease with inhibitors is more subtle. This amino acid is not close to the binding site and is not expected to interact directly with the inhibitors. It has been shown however to have hydrophobic interactions with Phe99 of chain B of the protease. They are separated by a distance ranging from 4.5 to 6.0 Å in the 13 complexes studied here. Phe99 is located at the carboxy terminus of chain B and has been observed to contribute significantly to the total dimerization energy (see Figure 6B). This residue also shows a stabilizing hydrophobic-hydrophobic interaction with Cys67, which is adjacent to Ile66. The critical role played by dimerization in the stabilization of the HIV-1 protease has been experimentally observed by the absence of folded monomers and confirmed theoretically by the structure-based thermodynamic analysis of the protein.\textsuperscript{37,38} However, it is impossible using only this evidence to infer whether inhibitors would have the potential to modify the oligomeric status of the HIV protease. More studies, experimental or molecular simulation, are necessary to clarify this point.
Our findings agree with the previous observations that the energy of stabilization of the protease molecule is not uniformly distributed through the protein structure because it originates at the binding site but extends to many other residues in the molecule, which may be distant from the active site. In this sense, the study of the changes introduced in the degree of folding distribution along the chain by a point mutation in the HIV-1 protease can shed some light. We have analyzed the degree of folding of the complexes formed by inhibitor A77003 with HIV-1 protease (1hvi) and with the mutated protease (1hvs) in which Val82 has been replaced by Ala. The differences in the degree of folding of all amino acids in the protein upon mutation are illustrated in Figure 6A. The mutation is at residue 82 but the principal changes in the degree of folding affect residues Thr80 and Ile84, which are close to Val82 but also others such as Trp6, Pro9, Gly17, and Pro39, which are distant from this site. A similar situation is observed when the change in the degree of folding is analyzed for all the complexes studied here. In Figure 6B, we plot the values of $\Delta I_3$ averaged for the 13 complexes. Our results show that the interaction between the inhibitors and the HIV-1 protease affects not only the residues located close to the binding site but also others such as Ile66, which, as has been seen, plays a role in the stabilization of the dimeric structure of the protease.

IV. Conclusions

Folding and packing are two characteristics of proteins that account for different aspects of their molecular structures. While packing parameters account for the space occupied for the whole protein, the folding of protein chain accounts for the structural organization of the protein backbone only. We have illustrated here that the degree of folding of protein chains can be affected significantly by protein–ligand interactions. The binding energetics of such interactions correlate with the global and local changes in the degree of folding. Such changes are not limited to the binding site of the protein but are distributed.

Figure 5. Linear plot of the observed versus calculated values of the free energy of binding between HIV-1 protease and inhibitors.

Figure 6. Change of the degree of folding of the HIV-1 protease interacting with A77003 upon mutation at Val82 (A) and the averaged change of the degree of folding after the interaction with the 13 inhibitors (B).
along the whole protein chain. In some cases, e.g., RNase mutants, these global changes correlate with the thermodynamic parameters of binding but in others, such as the HIV-1 protease inhibitor complexes, these correlations are only observed for local contributions from specific amino acid residues.

The use of the degree of folding as a parameter for empirical structural parametrization of the energetics of protein folding and binding depends on the generality of the effects observed here for all types of protein–ligand interactions. However, the use of this parameter for correlation-based studies of particular cases of protein–ligand interactions is appropriate and straightforward as we have illustrated. The current findings should encourage other studies toward the generalization of the use of the degree of folding as a structural parameter for studying protein–ligand interactions.

Acknowledgment. E.E. thanks “Ramón y Cajal” program, Spain for partial financial support. Constructive comments made by two anonymous referees are also acknowledged.

References