Structure-Based Thermodynamic Analysis of HIV-1 Protease Inhibitors†

Jason S. Bardi, Irene Luque, and Ernesto Freire*

Department of Biology and Biocalorimetry Center, The Johns Hopkins University, Baltimore, Maryland 21218

Received January 24, 1997; Revised Manuscript Received March 17, 1997⊗

ABSTRACT: A structural parametrization of the binding and folding energetics previously developed in this laboratory accounts quantitatively for the binding of 13 HIV-1 protease inhibitors for which high-resolution structures are available (A77003, A78791, A76928, A74704, A76889, VX478, SB203386, SB203238, SB206343, U100313, U89360, A98881, CGP53820). The binding free energies for the inhibitors are predicted with a standard deviation of ±1.1 kcal/mol or ±10%. Furthermore, the formalism correctly predicts the observed change in inhibition constant for the complex of A77003 and the resistant protease mutant V82A, for which the high-resolution structure is also available. The analysis presented here provides a structural mapping of the different contributions to the binding energetics. Comparison of the binding map with the residue stability map indicates that the binding pocket in the protease molecule has a *dual character*: half of the binding site is defined by the most stable region of the protein, while the other half is unstructured prior to inhibitor or substrate binding. This characteristic of the binding site accentuates cooperative effects that permit mutations in distal residues to have a significant effect on binding affinity. These results permit an initial assessment of the effects of mutations on the activity of protease inhibitors.

The development of successful strategies for structure-based molecular design requires the ability to accurately predict binding affinities from structural considerations. Previously, we have reported a structural parametrization of the folding and binding energetics of proteins and peptides (D'Aquino et al., 1996; Gomez & Freire, 1995; Gomez et al., 1995; Hilser et al., 1996; Luque et al., 1996). This parametrization has been shown to be accurate enough to predict the helical propensities of amino acids with an accuracy of better than 0.2 kcal/mol (Luque et al., 1996) and to correctly predict the global stability of proteins and the stability constants per residue as reflected in the pattern of NMR-detected hydrogen exchange protection factors (Hilser & Freire, 1996a−c). Here, we report the application of that methodology to the association of 13 different inhibitors of the HIV-1 protease for which high-resolution crystallographic structures are available. Inhibition constants for these inhibitors, some of which are in clinical trials or clinical use, are available. The 13 inhibitors are A77003, A78791, A76928, A74704, A76889, VX478, SB203386, SB203238, SB206343, U100313, U89360, A98881, and CGP53820 (Abdel-Meguid et al., 1994; Erickson et al., 1990; Fassler et al., 1993; Hoog et al., 1995; Kim et al., 1995; Lin et al., 1993; Madhusoodan et al., 1994; Thaisrivongs et al., 1995; Thompson et al., 1994). Their structures are shown in Figure 1. The analysis was also performed on the complex of A77003 with the inhibitor resistant protease mutant V82A for which the high resolution is available (Baldwin et al., 1995).

The HIV-1 protease has been an important target for drug development against HIV-1 infection due to its key role in viral maturation. Several HIV-1 protease inhibitors are already in clinical use and have shown significant promise in combination therapies that include nucleoside inhibitors or several protease inhibitors. A significant clinical outcome has been the emergence of viral strains that exhibit resistance to multiple HIV-1 protease inhibitors (Condra et al., 1995; Ho et al., 1994; Kaplan et al., 1994; Roberts, 1995; Tisdale, 1996). Loss of sensitivity to protease inhibitors occurs because the resistant viral strains encode for protease molecules containing specific amino acid mutations that lower the affinity for the inhibitors, yet maintain sufficient affinity for the substrate. The origins of the resistance are still unclear. While some of the observed mutations are located directly in the binding site, other mutations are far away from the binding pocket. It is also apparent that different inhibitors elicit different mutational patterns and that the patterns of cross resistance are not the same, despite the fact that all inhibitors target the same binding site.

The development of a new generation of protease inhibitors that effectively addresses the issue of resistance requires a better understanding of the interactions, both between protease and inhibitors and between protease and substrates. The sequencing of viral isolates from patients undergoing therapy with protease inhibitors has allowed identification of the location and character of the mutations but has provided no molecular description of the origin of resistance. The analysis presented here provides a detailed structural mapping of the binding energetics for 13 different protease inhibitors and a quantitative account of the effects of V82A mutation on the affinity for the inhibitor A77003. As such, these studies should help in the development of a new generation of inhibitors.

METHODS

I. Structural Parametrization of Binding Energetics. In all cases presented here the Gibbs energy of binding, ΔG, was calculated from the published crystallographic structures using procedures previously described (D'Aquino et al.,

† Supported by grants from the National Institutes of Health (RR04328 and GM51362). I.L. is a visiting student from the Universidad de Granada, Granada, Spain, partially supported by a fellowship from the Ministerio de Educación y Ciencia (PB93-1163).

* To whom correspondence should be addressed. Phone: (410) 516-7743. Fax: (410) 516-6469. E-mail: bcc@biocal2.bio.jhu.edu.

⊗ Abstract published in *Advance ACS Abstracts,* May 1, 1997.

S0006-2960(97)00174-8 CCC: $14.00

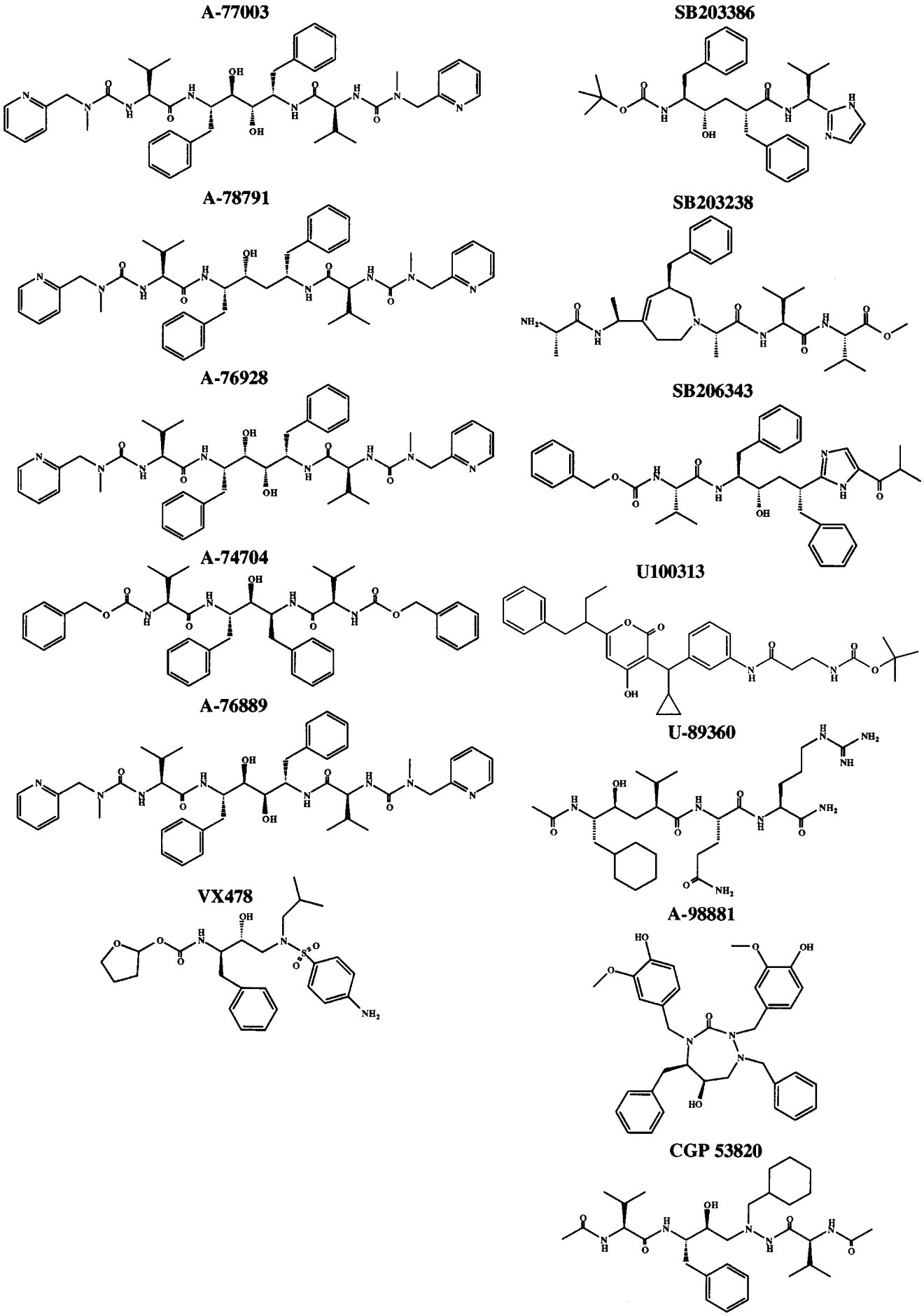

FIGURE 1: Chemical structures of the 13 HIV-1 protease inhibitors considered in this paper. The original references for each inhibitor are given in the text.

1996; Gomez & Freire, 1995; Gomez et al., 1995; Hilser et al., 1996; Luque et al., 1996). These calculations require the structures of the complex as well as the structures of the unligated protein and unligated inhibitor. In this approach, the generic portion of the Gibbs energy, ΔG_{gen}, is calculated from a separate computation of its enthalpy and entropy components. This portion of the Gibbs energy contains those contributions typically associated with the formation of secondary and tertiary structure (van der Waals interactions, hydrogen bonding, hydration, and conformational entropy). Additional contributions to the Gibbs energy of binding are not separated into enthalpic and entropic components. They include electrostatic and ionization effects, ΔG_{ion}, and the contribution of the change in translational degrees of freedom, ΔG_{tr}:

$$\Delta G = \Delta G_{gen} + \Delta G_{ion} + \Delta G_{tr} \quad (1)$$

Generic Contributions to Gibbs Energy. The most significant structural/solvation contributions to the total free energy of binding are contained in the term $\Delta G_{gen} = \Delta H_{gen} - T\Delta S_{gen}$ which is calculated by estimating separately its enthalpy and entropy components. The important structural changes for these calculations are the changes in apolar and polar solvent-accessible surface areas (ΔASA_{ap} and ΔASA_{pol}) and the distribution of interatomic distances between different atom types which determines the packing density.

The changes in accessible surface areas were calculated by implementing the algorithm of Lee and Richards (1971). In all calculations a solvent radius of 1.4 Å and a slice width of 0.25 Å were used. In order to better define small differences in solvent accessibility between inhibitors or mutants, 64 different protein/inhibitor orientations with respect to the slicing plane were considered in the accessible surface area calculations. These orientations were generated by rotating the molecule around the *x*, *y*, and *z* axis every 90°. In this way, the resulting solvent accessibility for each atom is the numerical average of the values obtained for all molecular orientations. When the solvent accessibilities for unfolded conformations are needed, a set of free energy optimized values is used (Luque et al., 1996).

(a) The Enthalpy Change. In binding or folding, the bulk of the enthalpy change originates from the formation of internal interactions (van der Waals, hydrogen bonds, etc.) and the parallel desolvation of the interacting groups. Not surprisingly, the bulk of the enthalpy change scales both in terms of ΔASA changes and the interatomic distances between the interacting groups. At the reference temperature of 60 °C it can be written as (Hilser et al., 1996)

$$\Delta H_{gen}(60) = (\alpha_{ap} + \beta_{ap}U_{ap}^{6})\Delta ASA_{ap} + (\alpha_{pol} + \beta_{pol}U_{pol}^{6})\Delta ASA_{pol} + \beta_{mix}U_{mix}^{6}\Delta ASA_{total} \quad (2)$$

where the empirical coefficients α and β have been estimated from an analysis of the protein thermodynamic database and are equal to $\alpha_{ap} = -12.96$, $\beta_{ap} = 25.34$, $\alpha_{pol} = 24.38$, $\beta_{pol} = 16.57$ and $\beta_{mix} = 16.42$. The terms U_i represent the packing density of apolar, polar, and mixed atoms and are equal to the energy-weighted average of the ratio between the separation distance at the minimum in the potential well and the actual separation between atom types. For the average packing density in proteins, eq 2 is well approximated by

$$\Delta H_{gen}(60) = -8.44\Delta ASA_{ap} + 31.4\Delta ASA_{pol} \quad (3)$$

At any other temperature, $\Delta H_{gen}(T)$ is obtained from the standard thermodynamic equation:

$$\Delta H_{gen}(T) = \Delta H_{gen}(60) + \Delta C_p(T - 333.15) \quad (4)$$

ΔH_{gen} needs not be equal to the experimental enthalpy. For example, it has been shown that for binding processes in which protons are either released or absorbed the measured enthalpy depends on the ionization enthalpy of the buffer (Gomez & Freire, 1995).

(b) The Entropy Change. In the calculation of the entropy change two primary contributions are included, one due to changes in solvation and the other due to changes in conformational degrees of freedom [$\Delta S_{gen}(T) = \Delta S_{solv}(T) + \Delta S_{conf}$]. The entropy of solvation is temperature dependent while the conformational entropy is essentially a constant at different temperatures. The entropy of solvation can be written in terms of the heat capacity if the temperatures at which the apolar and polar hydration entropies are zero ($T^*_{S,ap}$ and $T^*_{S,pol}$) are used as reference temperatures:

$$\Delta S_{solv}(T) = \Delta S_{solv,ap}(T) + \Delta S_{solv,pol}(T) \quad (5)$$

$$\Delta S_{solv}(T) = \Delta C_{p,ap}\ln(T/T^*_{S,ap}) + \Delta C_{p,pol}\ln(T/T^*_{S,pol}) \quad (6)$$

$T^*_{S,ap}$ has been known to be equal to 385.15 K for some time (Baldwin, 1986; Murphy & Freire, 1992) and $T^*_{S,pol}$ has been recently found to be close to 335.15 K (D'Aquino et al., 1996). While the entropy of apolar solvation appears to be additive, the situation for polar solvation is known to depend on the number and proximity of polar functional groups in the molecule (Cabani et al., 1981).

Conformational entropies upon binding or folding are evaluated by explicitly considering the following three contributions for each amino acid: (1) $\Delta S_{bu\rightarrow ex}$, the entropy change associated with the transfer of a side chain that is buried in the interior of the protein to its surface; (2) $\Delta S_{ex\rightarrow u}$, the entropy change gained by a surface-exposed side chain when the peptide backbone changes from a unique conformation to an unfolded conformation; and (3) ΔS_{bb}, the entropy change gained by the backbone itself upon unfolding from a unique conformation. The magnitude of these terms for each amino acid has been estimated by computational analysis of the probability of different conformers as a function of the dihedral and torsional angles (D'Aquino et al., 1996; Lee et al., 1994; Luque et al., 1996).

Since the HIV-1 protease inhibitors considered here are not peptides, a special parametrization was implemented in order to account for the change in conformational degrees of freedom between the bound and free forms of the inhibitors. As shown in Figure 1, the total number of atoms as well as the number of rotatable bonds varies between inhibitors. In general, the conformational entropy of the inhibitor free in solution will be proportional to the number of rotatable bonds. However, for a given number of rotatable bonds, excluded volume effects will increase with the total number of atoms in the molecule. Therefore, as a first approximation, the conformational entropy change of the nonpeptide inhibitors upon binding, ΔS_{np}, was considered to be a linear function of the number of rotatable bonds (N_{rb}) and total number of atoms (N_{atoms}):

$$\Delta S_{np} = k_1 N_{rb} + k_2 N_{atoms} \quad (7)$$

The coefficients k_1 and k_2 were estimated from a larger set of nonpeptide inhibitors that includes also renin inhibitors. The coefficient k_1 was found to be equal to −1.76 cal/(K·mol), which is close to the conformational entropy value observed for C-C bonds in long chain paraffins (Schellman, 1955a,b). The coefficient k_2 was found to be equal to 0.414 cal/(K·mol) and essentially accounts for the conformational entropy restrictions in the free inhibitor due to excluded volume.

(c) The Heat Capacity Change. The heat capacity change is a weak function of temperature and has been parametrized in terms of changes in solvent-accessible surface areas (ΔASA) since it originates mainly from changes in hydration (Gomez & Freire, 1995; Gomez et al., 1995; Murphy et al., 1992):

$$\Delta C_p = \Delta C_{p,ap} + \Delta C_{p,pol} \quad (8)$$

$$\Delta C_p = a_C(T)\Delta ASA_{ap} + b_C(T)\Delta ASA_{pol} + c_C(T)\Delta ASA_{OH} \quad (9)$$

where the coefficients $a_C(T) = (0.45 + 2.63 \times 10^{-4})(T - 25) - (4.2 \times 10^{-5})(T - 25)^2$ and $b_C(T) = (-0.26 + 2.85 \times 10^{-4})(T - 25) + 4.31 \times 10^{-5})(T - 25)^2$. The hydration of the hydroxyl group in aliphatic hydroxyl side chains (Ser and Thr) appears to contribute positively and not negatively to ΔC_p as previously assumed (0.17 cal·K^{-1}·mol^{-1} Å^2 at 25 °C) (Gomez & Freire, 1995). In the equation above, ΔASA changes are in Å^2 and the heat capacity is in cal/(K·mol). In general, for low-temperature calculations (T < 80 °C) the temperature-independent coefficients are sufficient (Gomez & Freire, 1995; Gomez et al., 1995). Specific effects like heat capacity changes associated to changes in protonation, differential binding of ligands or denaturants, etc. need to be considered individually (Gomez & Freire, 1995; Gomez et al., 1995).

Ionic and Electrostatic Contributions to Gibbs Energy. Two types of electrostatic effects need to be considered in this situation: Coulombic contributions due to the interactions between charged sites and the self-energy arising from charging a single site or alternatively the self-energy arising from transferring a charge between environments with different dielectric constants. These electrostatic contributions were computed as described by García-Moreno (1995) using the standard equation:

$$\Delta G_{el} = \sum_i \frac{332 Z_i^2}{2 r_i}\left(\frac{1}{A} - \frac{1}{A_{ref}}\right) + 332 Z_i \sum_j \frac{Z_j}{A r_{ij}} \quad (10)$$

where Z is the charge, r_i is the radius of the charged particle, r_{ij} is the separation between two charges, and A and A_{ref} are the attenuation parameters in the complex and reference. These parameters incorporate dielectric and screening effects as discussed in García-Moreno (1995) and García-Moreno et al. (1997).

Protonation effects are treated as described before (Gomez & Freire, 1995) from a knowledge of the pK_a of the groups that change ionization state upon binding. For the protease molecule, the active site residues Asp 25 in each chain have been observed to have distinct protonation states in the complex (pK_a < 2.5 and pK_a > 6.5, respectively) but similar pK_a's in the free form (pK_a ∼ 4) (Smith et al., 1996; Wang et al., 1996). The standard experimental assay for HIV-1 protease activity is at pH 4.7−5.5 (Abdel-Meguid et al., 1994; Erickson et al., 1990; Fassler et al., 1993; Hoog et al., 1995; Kim et al., 1995; Lin et al., 1993; Madhusoodan et al., 1994; Thaisrivongs et al., 1995; Thompson et al., 1994), which corresponds to the experimental range in which the pH dependence of the activity is minimal (Hyland et al., 1991; Smith et al., 1996; Wang et al., 1996). Under those experimental conditions, the coupling to protonation/deprotonation is expected to be small.

Translational Entropy Contribution to Gibbs Energy. The association of two or more molecules reduces the translational degrees of freedom available to those molecules. There has been considerable discussion regarding the exact magnitude of this term since no precise calculations are available [see, for example, Janin (1995), Kauzmann (1959), and Murphy et al. (1994)]. In our work (Gomez & Freire, 1995; Murphy et al., 1994), we have found that the value that best accounts for experimental results is the cratic entropy proposed by Kauzmann (1959). For the formation of a bimolecular complex the cratic entropy is equal to −8 cal/(K·mol), which amounts to approximately 2.4 kcal/mol at 25 °C.

II. Calculation of Residue Stability Constants of the Protease Molecule. An important set of parameters for mapping the structural stability of different regions of a protein is given by the apparent residue stability constants. For any given residue, the apparent stability constant per residue, $\kappa_{f,j}$, is defined as the ratio of the probabilities of all states in which that residue is folded, $P_{f,j}$, to the probabilities of the states in which that same residue is not folded:

$$\kappa_j = \frac{\sum_{\text{states with residue } j \text{ folded}} P_i}{\sum_{\text{states with residue } j \text{ not folded}} P_i} = \frac{P_{f,j}}{P_{nf,j}} \quad (11)$$

The apparent stability constant per residue, $\kappa_{f,j}$, is the quantity that one will measure if it were possible to experimentally determine the stability of the protein by monitoring each individual residue. Therefore, variations in stability constants per residue permit an evaluation of structural stability differences between regions of the protein. In many cases, hydrogen-exchange protection factors measured by NMR permit an experimental determination of the apparent stability constants per residue (Hilser & Freire, 1996a−c).

The set of residue stability constants for the HIV-1 protease molecule was calculated from the structure according to the COREX algorithm (Hilser & Freire, 1996a−c) using the structure of the free protein (1hhp) or protein structures obtained from the complex coordinates by removing the inhibitor. A total of 126 496 states with degrees of folding ranging from the native to the completely unfolded states were used in these calculations. The states were generated by using a sliding block of windows of 16 amino acids each. The Gibbs energy, partition function, and relative probability of the 126 496 states were calculated using the structural parametrization of the energetics described above.

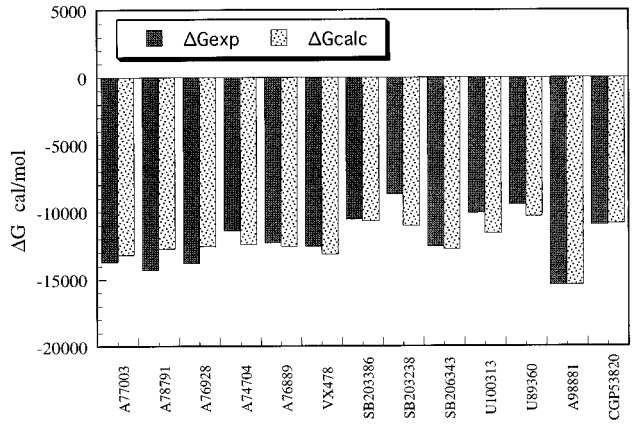

FIGURE 2: Predicted and experimental binding affinities for the 13 HIV-1 protease inhibitors considered here. The calculations were performed as described using the corresponding PDB files for each complex (A77003, 1hvi; A78791, 1hvj; A76928, 1hvk; A74704, 9hvp; A76889, 1hvl; VX478, 1hpv; SB203386, 1sbg; SB203238, 1hbv; SB206343, 1hps; U100313, 2upj; U89360, 1gno; A98881, 1pro; CGP53820, 1hih).

RESULTS AND DISCUSSION

I. Structure-Based Prediction of Inhibitor Affinities. Figure 2 shows the predicted and experimental binding affinities for the 13 HIV-1 protease inhibitors considered here. For those protease/inhibitor complexes for which the structure of the free enzyme is available, the calculations were performed by using both the structure of the free enzyme (Spinelli et al., 1991) and the structure of the enzyme in the complex but without the inhibitor as the unligated protein. The results were equivalent in both cases, the differences in Gibbs energies being smaller than 0.5 kcal/mol on the average. For those cases in which deviations were larger (pdb files 2upj, 1hvi, 1hvj, 1hvk, 9hvp) the deviations were traced to the side-chain conformations of Phe 53B, Lys 55B, Arg 41A, and Arg 41B. These side chains are solvent exposed and far away from the binding site, indicating that the conformational differences are not related to the inhibitor. The statistical analysis of the data reveals that the free energies of binding are predicted with a standard deviation of ±1.1 kcal/mol and a standard error of 0.3 kcal/mol. The standard deviation amounts to a relative uncertainty of ±10%. The correlation analysis between the experimental and predicted ΔG values yields a slope of 0.982 with a correlation coefficient of 0.85. The structural predictions show no systematic deviations and are accurate enough to permit an examination of the different contributions to the binding energetics.

According to the analysis, the binding of the 13 inhibitors to the enzyme is dominated by the hydrophobic effect. Upon binding, not only the inhibitor itself but also protease residues located in the binding pocket bury a significant nonpolar surface from the solvent. In fact, the average fraction of nonpolar area buried from the solvent upon binding is 0.737 ± 0.02, which is much higher than the fraction buried by a typical globular protein upon folding (between 0.55 and 0.60). Not surprisingly, the major contribution to the Gibbs energy of binding is given by the favorable entropy resulting from the release of water molecules associated with the desolvation of those surfaces. On the other hand, the enthalpic contributions due to those generic effects are unfavorable at room temperature since they are dominated by the positive enthalpy of desolvating hydrophobic groups. The breakdown of the energetics is summarized in Table 1.

The heat capacity values listed in Table 1 are of the same magnitude as those measured for other protease inhibitors (Gomez & Freire, 1995). The magnitude of the heat capacity changes is dominated by changes in solvation of polar and nonpolar groups and is not expected to be significantly affected by other interactions (Gomez et al., 1995). The enthalpy values listed in Table 1 include only generic contributions and cannot be compared directly to experimental values since protonation and buffer effects are not included. This generic enthalpy is composed primarily of two opposite effects, a favorable component due to the formation of van der Waals, hydrogen bonds, and other interactions between inhibitor and protease and an unfavorable component due to desolvation. Due to the highly hydrophobic character of the inhibitors the dominant term is the desolvation term. This is, however, not a general phenomenon as demonstrated for the binding of some peptide inhibitors which exhibits a favorable enthalpy under certain conditions (Gomez & Freire, 1995). As shown in Table 1, the structure/solvation terms included in ΔG_{gen} make the largest contribution to the total Gibbs energy of binding. All the inhibitors are highly hydrophobic and lack polar groups. For this reason, electrostatic interactions are predicted to contribute very little to the binding Gibbs energy. The only significant electrostatic contribution described by eq 10 arises from the change in the environment of Asp 25, Asp 29, and Asp 30 which may contribute up to 0.7 kcal/mol depending on the inhibitor. This contribution arises from the transferring of the charge from water to an environment with a somewhat lower dielectric constant. According to the experimental results of García-Moreno et al. (1997) the dielectric constant in the interior of a protein is no lower than ∼15. According to these authors, the dielectric constant of different protein regions ranges between 15 and 78.5 depending on the solvent accessibility.

II. Structural Mapping of Protease Residues Contributing to Binding Affinity. For the entire set of inhibitors, essentially the same set of protease residues, albeit with different strength, was observed to contribute to the binding energetics, reflecting the fact that they have been targeted to the same site on the molecule. Table 2 summarizes those amino acids in the protease molecule that contribute more than 0.1 kcal/mol to the total free energy of binding for at least one of the inhibitors. The values listed in the table do not include the contribution corresponding to the inhibitor (∼55% of the total Gibbs energy of binding) or the translational entropy that cannot be ascribed to a particular amino acid. Figure 3 shows the location of those amino acids in the protease structure.

From an energetic standpoint, the binding pocket is defined by amino acids belonging to four noncontiguous regions in the sequence: amino acids in the region containing the catalytic Asp group (Asp 25, Gly 27, Ala 28, Asp 29, Asp 30), the so-called flap region (Met 46, Ile 47, Gly 48, Gly 49, Ile 50), the strand between residues 80−86 (Pro 81, Val 82, Ile 84), and Arg 8, which contributes significantly to the binding energetics. Due to the chemical structure of the inhibitors, both chains in the protease molecule contribute in a more or less symmetrical fashion to the total Gibbs energy of binding. In all cases the region containing the

Table 1: Structure-Based Thermodynamics of Inhibitor Binding to HIV-1 Protease[a]

inhibitor	ΔC_p	ΔH_{gen}	ΔS_{solv}	ΔS_{conf}	ΔG_{gen}	ΔG_{other}	ΔG_{total}	$\Delta\Delta G$
A77003	−397	11 733	115.2	−23.7	−15 552	2385	−13 167	532
A78791	−400	12 183	115.8	−24.2	−15 127	2385	−12 742	1557
A76928	−392	11 853	113.4	−23.6	−14 932	2385	−12 547	1249
A74704	−379	11 254	110.1	−20.8	−15 378	2953	−12 425	−1037
A76889	−387	11 303	112.7	−23.7	−15 229	2680	−12 549	−271
VX478	−320	8 641	93.9	−11.1	−16 046	2903	−13 143	−563
SB203386	−343	10 410	96.2	−16.8	−13 263	2555	−10 688	−123
SB203238	−320	8 641	94.0	−18.1	−13 959	2918	−11 041	−2356
SB206343	−332	8 109	98.9	−19.8	−15 481	2724	−12 757	−177
U100313	−317	8 471	93.4	−16.6	−14 416	2807	−11 608	−1531
U89360	−236	1 255	76.5	−28.0	−13 211	2877	−10 334	−893
A98881	−293	7 512	86.6	−1.0	−18 018	2615	−15 403	14
CGP53820	−294	6 294	88.7	−22.3	−13 504	2643	−10 861	115

[a] Calculated thermodynamic parameters for inhibitor binding to HIV-1 protease. ΔC_p is in cal/(K·mol); ΔS values are in cal/(K·mol); ΔH and ΔG values are in cal/mol. ΔH_{gen} and ΔG_{gen} include only the structure/solvation contributions to ΔG. Under ΔG_{other} the electrostatic and cratic contributions have been combined.

Table 2: Mapping of HIV-1 Residue Contributions to Gibbs Energy of Inhibitor Binding[a]

residue	A77003	A78791	A76928	A74704	A76889	VX478	SB203386	SB203238	SB206343	U100313	U89360	A98881	CGP53820
ARG−8−A	−351	−358	−397	−602	−385	−132	−311	−468	−228	−106	−140	−363	−425
ASP−25−A	−344	−351	−342	−146	−337	−283	−320	−283	−248	−313	−346	−290	−312
GLY−27−A	−778	−747	−579	−629	−596	−538	−810	−436	−775	−437	−703	−542	−789
ALA−28−A	−189	−178	−207	−192	−209	−146	−165	−207	−159	−181	−135	−195	−177
ASP−29−A	−435	−447	−407	−702	−413	−312	−197	−455	−422	−308	−570	−331	−336
ASP−30−A	−198	−248	−192	−184	−232	−110	−158	−137	−220	−382	−437	−187	−202
MET−46−A	0	0	0	−290	0	0	0	0	0	0	−17	0	0
ILE−47−A	−43	−39	−38	−146	−30	−84	−46	−78	−87	−107	−97	−77	−55
GLY−48−A	−965	−962	−906	−935	−933	−719	−592	−616	−1081	−1021	−1186	−821	−904
GLY−49−A	−351	−378	−265	−173	−241	−125	−155	−128	−140	−192	−282	−107	−253
ILE−50−A	−271	−294	−253	−185	−272	−239	−260	−167	−271	−210	−151	−235	−231
PHE−53−A	0	0	0	−91	0	0	0	0	−65	−69	−161	0	0
PRO−81−A	−121	−129	−152	−166	−160	−78	−131	−181	−123	−42	−97	−98	−148
VAL−82−A	−351	−357	−347	−191	−359	−119	−200	−179	−165	−95	−211	−159	−120
ILE−84−A	−86	−64	−82	−67	−83	−99	−60	−102	−90	−71	−95	−73	−51
ARG−8−B	−505	−513	−376	−230	−462	−245	−214	−283	−764	−624	−418	−352	−486
ASP−25−B	−323	−345	−369	−315	−366	−273	−225	−275	−334	−384	−372	−337	−367
GLY−27−B	−648	−652	−590	−501	−561	−594	−722	−444	−832	−353	−755	−451	−729
ALA−28−B	−184	−189	−196	−237	−195	−146	−203	−243	−183	−198	−167	−180	−183
ASP−29−B	−473	−478	−377	−432	−404	−275	−360	−523	−589	−420	−24	−218	−294
ASP−30−B	−167	−164	−156	−88	−149	−273	−326	−427	−313	−267	0	−270	−188
ILE−47−B	−56	−53	−60	−75	−59	−73	−68	−110	−60	−91	−32	−58	−66
GLY−48−B	−968	−954	−906	−1413	−931	−507	−966	−1089	−846	−701	−433	−591	−958
GLY−49−B	−156	−189	−191	−201	−185	−150	−166	−216	−192	−170	−174	−144	−150
ILE−50−B	−319	−256	−263	−125	−261	−215	−193	−204	−362	−215	−216	−212	−305
PRO−81−B	−247	−233	−154	−133	−179	−153	−126	−76	−219	−362	−348	−122	−160
VAL−82−B	−398	−388	−352	−151	−393	−142	−179	−134	−156	−215	−121	−135	−94

[a] Gibbs energies are in cal/mol.

catalytic Asp group is the major contributor to the binding energetics.

III. Structural Stability of Protease Residues Contributing to Inhibitor Binding. Figure 4 displays the calculated residue stability constants for the HIV-1 protease molecule. These constants map the protein molecule in terms of the structural stability of different regions (Hilser & Freire, 1996a−c). Protein residues with a high probability of being in the native conformation are characterized by high stability constants while residues that are most likely to be unstructured have low stability constants.

Two regions of the HIV-1 protease molecule are predicted to have the highest stability: the region including residues 13−32 and the region including residues 82−92. These two regions are distant in sequence but close in three-dimensional space and define, to a significant extent, the hydrophobic core of the molecule. Portions of these two regions (residues 23−29 in the amino terminus and 86−99 in the carboxy terminus) as well as the first nine residues in the sequence are predicted to contribute significantly to the dimerization interface. This region of the protein is predicted to be folded and well structured in the vast majority of conformational states that are accessible to the protease under native conditions. The active site triad (Asp 25, Thr 26, Gly 27) is located in the most stable part of the molecule as shown in Figure 4. This conclusion agrees with the results of the crystallographic analysis which identify this area of the molecule as quite rigid due to a dense network of hydrogen bonds (Wlodawer & Erickson, 1993). Residues 82−92 comprise most of the well-defined and highly stable h′ helix. Conversely, the region between residues 40−60, which corresponds to the flap, is characterized by very low stability constants per residue and is predicted to be unstructured even under native conditions. In the complexes, the flap is stabilized by its interactions with the inhibitors. Similar results were obtained with the structure of the free protein (pdb file 1hhp) or with protein structures obtained from complexes by removing the inhibitor coordinates. This

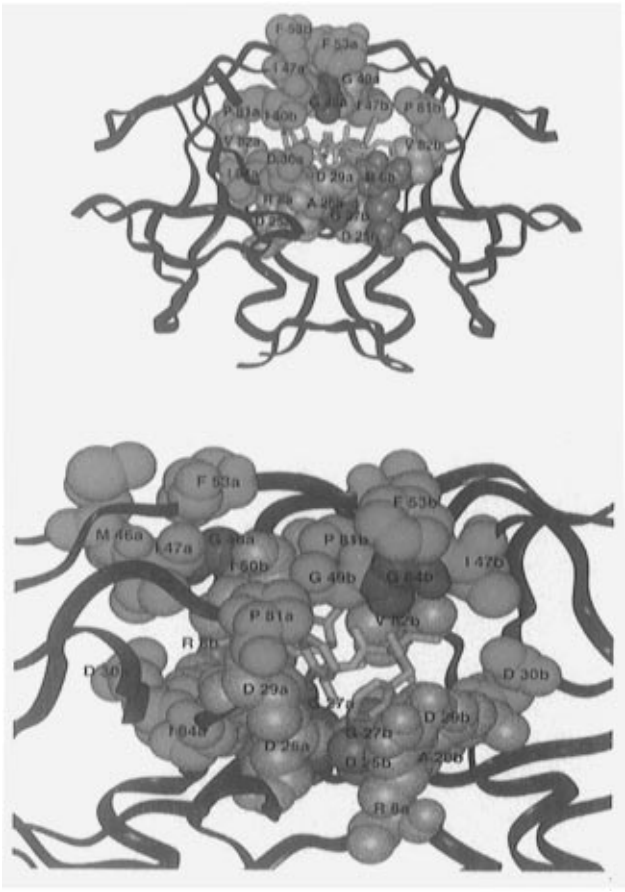

FIGURE 3: Two different views of the amino acid residues in the binding pocket of the HIV-1 protease molecule that contribute more than 0.1 kcal/mol to the Gibbs energy of binding. The residues depicted in red contribute between −0.7 and −0.9 kcal/mol, the residues depicted in orange between −0.5 and −0.7 kcal/mol, the residues depicted in yellow between −0.3 and −0.5 kcal/mol, and the residues depicted in green −0.1 and −0.3 kcal/mol. As a guide to the eye, the inhibitor A77003 is shown using a stick representation.

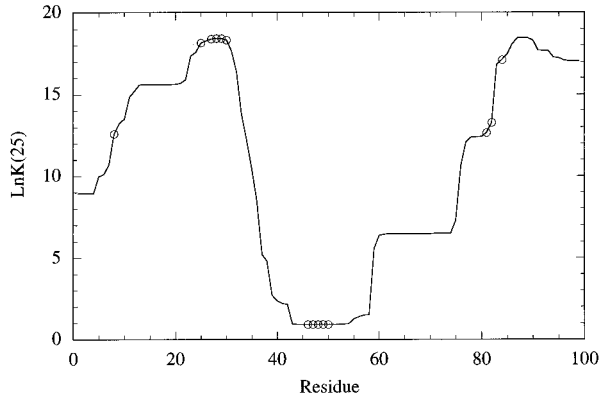

FIGURE 4: Calculated residue stability constants for the HIV-1 protease molecule. These constants were calculated according to the COREX algorithm (Hilser & Freire, 1996a−c) and map the protein molecule in terms of the structural stability of different regions. The circles indicate the location of the residues that contribute more than 0.1 kcal/mol to the Gibbs energy of inhibitor binding (Table 2).

observation suggests that in the protease/inhibitor complex the flap is stabilized by interactions with the inhibitor and not with the protein.

Figure 4 also indicates the location of the residues that contribute significantly to the Gibbs energy of binding. It is clear that the binding site is made up of residues belonging to both the most and the least stable regions of the protease molecule. This *dual character* of the binding pocket defines one of the most fundamental features of inhibitor binding to the protease molecule. Essentially, half of the binding site is preformed while the other half is formed during binding. The most stable region (containing binding site components Asp 25, Gly 27, Ala 28, Asp 29, Asp 30, Pro 81, Val 82, Ile 84) is essentially locked in a binding-competent conformation before binding occurs. The flap region, on the other hand (containing binding site components Met 46, Ile 47, Gly 48, Gly 49, Ile 50), is largely unstructured before binding and is forced into a unique conformation by its interaction with the inhibitor. For this reason, protease residues not in direct contact with the inhibitor but capable of affecting, structurally or energetically, the facility with which the flap adopts its bound conformation will influence the overall binding energetics.

IV. The Molecular Basis of Protease Resistance. The binding energetics described above provide some insight into the nature of the changes in HIV-1 protease mutants that have been observed to elicit *in vivo* resistance to multiple inhibitors. Several mutations have been identified in viral isolates from patients treated with protease inhibitors. For example, treatment with Ro31-8959 (saquinavir) has been shown to consistently induce the double mutant G48V + L90M (Roberts, 1995). In vitro selection of mutants by A77003 include V82I, M46L, M46F, V32I, V32I + V82I, and R8Q (Kaplan et al., 1994). Resistant variants to VX478 that have been identified are M46I and I50V(Schinazi et al., 1996; Tisdale, 1996). A recent study has shown that four mutations (M46I + L63P + V82T + I84V) are sufficient to elicit cross resistance to the inhibitors MK639, XM323, A80987, Ro31-8959, VX478, and SC52151 (Condra et al., 1995). Some of these mutations are located on the binding pocket and are thought to affect the binding affinity by a direct alteration of protease/inhibitor interactions. Other mutations are at distant locations and are expected to affect affinity by cooperative interactions.

In general, mutations in HIV-1 protease may affect the binding energetics by a direct interaction with the inhibitor, by a cooperative effect in which the mutated amino acid does not interact directly with the inhibitor but affects interactions between protease residues that elicit an altered protease/inhibitor interaction, or by some combination of both. For example, some mutations are located either in the flap or the hinge region, and some of them are distal from the binding site (e.g., L63P, A71V). As discussed above, the flap is essentially disordered in the free enzyme. Therefore, if a mutation induces a substantial energy barrier for the flap to adopt its bound conformation, it will affect the binding affinity even if the mutation is distal from the binding pocket. Mutations like L63P or A71V will decrease the conformational degrees of freedom of that region and make some conformations unaccessible or energetically unfavorable (D'Aquino et al., 1996).

The inhibition constants for A77003 to some mutants have been measured (Kaplan et al., 1994), and there is one case for which the crystallographic structure of the complex is also available: the complex of A77003 with the V82A mutant HIV-1 protease (Baldwin et al., 1995). Structure-based thermodynamic calculations were performed on the mutant complex resulting in a binding free energy of −12 394

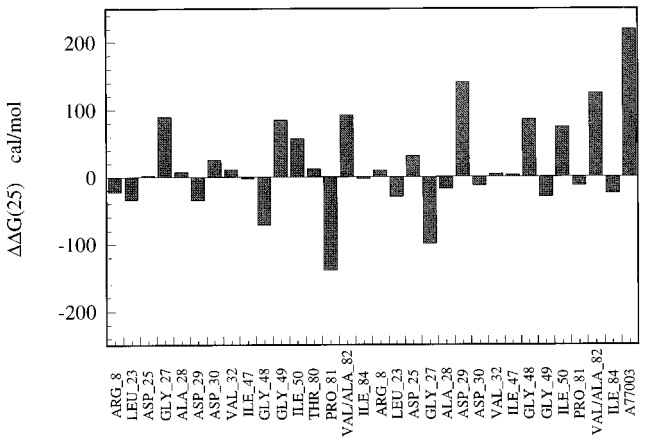

FIGURE 5: Difference in the HIV-1 protease residue contribution to the Gibbs energy of binding of A77003 to the wild-type protease and the resistant mutant V84A. $\Delta\Delta G$ is calculated as ($\Delta G_{\text{mutant}} - \Delta G_{\text{wild type}}$) for all residues that contribute more than 0.1 kcal/mol to the binding free energy.

cal/mol compared with the value of −13 167 cal/mol obtained for the wild type. These free energies translate into a 3.7-fold reduction in the binding affinity for the mutant which compares favorably with the 4-fold reduction measured experimentally (Baldwin et al., 1995). More important than the numerical comparison is the mechanism by which a single mutation affects the binding affinity. The structural mapping of the binding energetics in the wild type and mutant complexes reveals that the effect of the mutation cannot be assigned to a single site and that the Gibbs free energy of binding is redistributed throughout the entire binding pocket. This result is in agreement with the crystallographic analysis of Baldwin et al. (1995), who also concluded that the effect of the V82A mutation cannot be rationalized by the deletion of a single methyl group in each chain and that the overall effect is due to the backbone rearrangement induced by that mutation. Figure 5 shows the calculated $\Delta\Delta G$ values between mutant and wild type for all residues that contribute more than 0.1 kcal/mol to the binding free energy. This case illustrates very clearly that even for the replacement of a single methyl group the Gibbs energy of binding cannot be rationalized by simply adding group contributions. Contributions such as those tabulated in Table 2 depend on the global context of each group. It also indicates that the interpretation of the effect of mutations on the binding affinity of an inhibitor requires a global analysis even if the mutation is located in the binding pocket.

Conclusions. The free energy of inhibitor binding to the protease molecule reflects a delicate balance between mutually compensatory enthalpy and entropy terms. These compensatory terms define the main thermodynamic roadblocks in molecular design. Molecular modifications resulting in a more favorable enthalpy bring about unfavorable entropy contributions, a phenomenon known as enthalpy−entropy compensation. Also, the enthalpy and entropy changes themselves are composed of opposing contributions. The enthalpy of formation of internal interactions is opposed by the enthalpy of desolvation, and the entropy of desolvation is opposed by the conformational entropy. Molecular design requires accurate prediction of these effects, in order to minimize the undesirable consequences of compensatory changes.

The high accuracy with which the binding affinities of a number of HIV-1 protease inhibitors can be predicted from structure suggests that the approach presented here can be used to help in the design of new protease inhibitors. In addition, given that the structural parametrization of the binding energetics accounts well for the observed change in inhibition constant between the wild type and a resistant mutant of the HIV-1 protease for which the high-resolution structure is available, this approach has the potential for addressing the issue of resistance in molecular design.

REFERENCES

Abdel-Meguid, S. S., Metcalf, B. W., Carr, T. J., Demarsh, P., DesJarlais, R. L., Fisher, S., Green, D. W., Ivanoff, L., Lambert, D. M., Murthy, K. H. M., Petteway, S. R., Jr., Pitts, W. J., Tomaszek, T. A., Jr., Winborne, E., Baoguang Zhao, B., Dreyer, G. B., & Meek, T. D. (1994) *Biochemistry 33*, 11671−11677.

Baldwin, E. T., Bhat, T. N., Liu, B., Pattabiraman, N., & Erickson, J. W. (1995) *Nat. Struct. Biol. 2*, 244−249.

Baldwin, R. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8069−8072.

Cabani, S., Gianni, P., Mollica, V., & Lepori, L. (1981) *J. Solution Chem. 10*, 563−595.

Condra, J. H., Schleif, W. A., Blahy, O. M., Gabryelski, L. J., Graham, D. J., Quintero, J. C., Rhodes, A., Robbins, H. L., Roth, E., Shivaprakash, M., Titus, D., Yang, T., Teppler, H., Squires, K. E., Deutsch, P. J., & Emini, E. A. (1995) *Nature 374*, 569−571.

D'Aquino, J. A., Gómez, J., Hilser, V. J., Lee, K. H., Amzel, L. M., & Freire, E. (1996) *Proteins 25*, 143−156.

Erickson, J. W., Neidhart, D. J., Van Drie, J., Kempf, D. J., Wang, X. C., Norbeck, D. W., Plattner, J. J., Rittenhouse, J. W., Turon, M., Wideburg, N., Kohlbrenner, W. E., Simmer, R., Helfrich, R., Paul, D. A., & Knigge, M. (1990) *Science 249*, 527−529.

Fassler, A., Rosel, J., Grutter, M., Tintelnot, M., Alter, E., Bold, G., & Lang, M. (1993) *Bioorg. Med. Chem. Lett. 3*, 2817−2842.

García-Moreno, B. E. (1995) *Methods Enzymol. 259*, 512−538.

García-Moreno, B. E., Dwyer, J. J., Gittis, A. G., Lattman, E. E., Spencer, D. S., & Stites, W. E. (1997) *Biophys. Chem.* (in press).

Gomez, J., & Freire, E. (1995) *J. Mol. Biol. 252*, 337−350.

Gomez, J., Hilser, J. V., Xie, D., & Freire, E. (1995) *Proteins: Struct., Funct., Genet. 22*, 404−412.

Hilser, V. J., & Freire, E. (1996a) *Proteins* (in press).

Hilser, V. J., & Freire, E. (1996b) *J. Mol. Biol. 262*, 756−772.

Hilser, V. J., & Freire, E. (1996c) *Biophys. Chem.* (in press).

Hilser, V. J., Gomez, J., & Freire, E. (1996) *Proteins 26*, 123−133.

Ho, D. D., Toyoshima, T., Mo, H., Kempf, D. J., Norbeck, D., Chen, C., Wideburg, N. E., Burt, S. K., Erickson, J. W., & Singh, M. K. (1994) *J. Virol. 68*, 2016−2020.

Hoog, S. S., Zhao, B., Winbornel, E., Fisher, S., Green, D. W., DesJarlais, R. L., Newlander, K. A., Callahan, J. F., Moore, M. L., Huffman, W. F., & Abdel-Meguid, S. S. (1995) *J. Med. Chem. 38*, 3246−3252.

Hyland, L. J., Tomaszek, J. T. A., & Meek, T. D. (1991) *Biochemistry 30*, 8454−8463.

Janin, J. (1995) *Proteins 21*, 30−39.

Kaplan, A. H., Michael, S. F., Wehbie, R. S., Knigge, M. F., Paul, D. A., Everitt, L., Kempf, D. J., Norbeck, D. W., & Erickson, J. W. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5597−5601.

Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1−63.

Kim, E. E., Baker, C. T., Dwyer, M. D., Murcko, M. A., Rao, B. G., Tung, R. D., & Navia, M. A. (1995) *J. Am. Chem. Soc. 117*, 1181−1182.

Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.

Lee, K. H., Xie, D., Freire, E., & Amzel, L. M. (1994) *Proteins: Struct., Funct., Genet. 20*, 68−84.

Lin, Y., Lin, X., Hong, L., Foundling, S., Heinrikson, R. L., Thaisrivongs, S., Leelamanit, W., Raterman, D., Shah, M., Dunn, B. D., & Tang, J. (1995) *Biochemistry 34*, 1143−1152.

Luque, I., Mayorga, O., & Freire, E. (1996) *Biochemistry 35*, 13681−13688.
Madhusoodan, V. H., Bhat, T. N., Kempf, D. J., Baldwin, E. T., Liu, B., Gulnik, S., Wideburg, N. E., Norbeck, D. W., Appelt, K., & Erickson, J. W. (1994) *J. Am. Chem. Soc. 116*, 847−855.
Murphy, K. P., & Freire, E. (1992) *Adv. Protein Chem. 43*, 313−361.
Murphy, K. P., Bhakuni, V., Xie, D., & Freire, E. (1992) *J. Mol. Biol. 227*, 293−306.
Murphy, K. P., Xie, D., Thompson, K., Amzel, L. M., & Freire, E. (1994) *Proteins: Struct., Funct., Genet. 18*, 63−67.
Roberts, N. A. (1995) *AIDS 9*, S27−S32.
Schellman, J. A. (1955a) *C. R. Trav. Lab. Carlsberg, Ser. Chim. 29*, 230−259.
Schellman, J. A. (1955b) *C. R. Trav. Lab. Carlsberg, Ser. Chim. 29*, 223−229.
Schinazi, R. F., Larder, B. A., & Mellors, J. W. (1996) *Int. Antiviral News 4*, 95−100.
Smith, R., Brereton, I. M., Chai, R. Y., & Kent, S. B. H. (1996) *Nat. Struct. Biol. 3*, 946−950.
Spinelli, S., Liu, Q. Z., Alzari, P. M., Hirel, P. H., & Poljak, R. J. (1991) *Biochimie 73*, 1391−1396.
Thaisrivongs, S., Watenpaugh, K. D., Howe, W. J., Timich, P. K., Dolak, L. A., Chong, K. T., Tomich, C. S., Tomasselli, A. G., Turner, S. R., Strohbach, J. W., Mulichak, A. M., Janakiraman, M. N., Moon, J. B., Lynn, J. C., Horng, M. M., Hinshaw, R. R., Curry, K. A., & Rothrock, D. J. (1995) *J. Med. Chem. 38*, 3624−3637.
Thompson, S. K., Murthy, K. H. M., Zhao, B., Winborne, E., Green, D. W., Fisher, S. M., DesJarlais, R. L., Tomaszek, T. A., Jr., Meek, T. D., Gleason, J. G., & Abdel Meguid, S. S. (1994) *J. Med. Chem. 37*, 3100−3107.
Tisdale, M. (1996) *Int. Antiviral News 4*, 38−40.
Wang, Y. X., Freedberg, D. I., Yamazaki, T., Wingfield, P. T., Stahl, S. J., Kaufman, J. D., Kiso, Y., & Torchia, D. A. (1996) *Biochemistry 35*, 9945−9950.
Wlodawer, A., & Erickson, J. W. (1993) *Annu. Rev. Biochem. 62*, 543−585.

BI9701742