PROTEINS: Structure, Function, and Genetics 49:181–190 (2002)

Structural Parameterization of the Binding Enthalpy of Small Ligands

Irene Luque and Ernesto Freire*
Department of Biology, The Johns Hopkins University, Baltimore, Maryland

ABSTRACT **A major goal in ligand and drug design is the optimization of the binding affinity of selected lead molecules. However, the binding affinity is defined by the free energy of binding, which, in turn, is determined by the enthalpy and entropy changes. Because the binding enthalpy is the term that predominantly reflects the strength of the interactions of the ligand with its target relative to those with the solvent, it is desirable to develop ways of predicting enthalpy changes from structural considerations. The application of structure/enthalpy correlations derived from protein stability data has yielded inconsistent results when applied to small ligands of pharmaceutical interest (MW < 800). Here we present a first attempt at an empirical parameterization of the binding enthalpy for small ligands in terms of structural information. We find that at least three terms need to be considered: (1) the intrinsic enthalpy change that reflects the nature of the interactions between ligand, target, and solvent; (2) the enthalpy associated with any possible conformational change in the protein or ligand upon binding; and, (3) the enthalpy associated with protonation/deprotonation events, if present. As in the case of protein stability, the intrinsic binding enthalpy scales with changes in solvent accessible surface areas. However, an accurate estimation of the intrinsic binding enthalpy requires explicit consideration of long-lived water molecules at the binding interface. The best statistical structure/enthalpy correlation is obtained when buried water molecules within 5–7 Å of the ligand are included in the calculations. For all seven protein systems considered (HIV-1 protease, dihydrodipicolinate reductase, Rnase T1, streptavidin, pp60c-Src SH2 domain, Hsp90 molecular chaperone, and bovine β-trypsin) the binding enthalpy of 25 small molecular weight peptide and nonpeptide ligands can be accounted for with a standard error of ±0.3 kcal · mol^{-1}. Proteins 2002;49: 181–190.** © 2002 Wiley-Liss, Inc.

Key words: binding energetics; binding thermodynamics; isothermal titration calorimetry; structure-based drug design; HIV-1 protease

INTRODUCTION

The completion of the Human Genome Project as well as the genomes of several pathogens has generated new challenges in ligand and drug design. The number of targets for drug development is expected to increase dramatically during the next few years. For each new target, lead compounds will need to be identified and optimized in order to achieve the required binding affinity, specificity, selectivity, bioavailability, and toxicological properties. This new reality accentuates the need for improved design paradigms and efficient ways of predicting binding parameters from structure. One approach to the prediction of binding parameters that does not require extensive computations is the development of accurate empirical parameterizations based on structure/thermodynamic correlations. Previously, such approach has been successfully applied to protein stability,[1,2,7,8] and its predictive ability has been demonstrated in several laboratories.[1–4]

Currently, the binding affinity is used as the main selection criteria in high-throughput screening and computational analysis. Nonetheless, the binding affinity is defined by the Gibbs energy of binding (ΔG), which, in turn, is determined by the enthalpy (ΔH) and the entropy (ΔS) changes ($\Delta G = \Delta H - T\Delta S$). In principle, many combinations of ΔH and ΔS can give rise to the same ΔG value and, therefore, to the same binding affinity. Because the magnitudes of the enthalpy and entropy changes reflect different underlying interactions, ligands that have been enthalpically or entropically optimized exhibit different responses to changes in the target or the environment, even if they have the same affinity under the initial set of conditions.[5,6] For those reasons, additional selection and optimization criteria that includes enthalpic and entropic contributions have been proposed.[5,6]

Here we describe a first attempt at an empirical structural parameterization of the binding enthalpy for small ligands (MW < 800). As in the case of protein stability, the basic requirement for a statistical parameterization is the availability of a set of protein complexes for which high-resolution structures and binding thermodynamic data are available. Because the conformational change in the

Grant sponsor: National Institutes of Health; Grant number: GM 57144. Grant sponsor: National Science Foundation; Grant number: MCB9816661. Grant sponsor: postdoctoral fellowship from the Fundación Ramón Areces, Madrid (Spain) to I. L.

*Correspondence to: Ernesto Freire, Department of Biology, The Johns Hopkins University, Baltimore, MD 21218. E-mail: ef@jhu.edu

Received 27 March 2002; Revised 16 May 2002; Accepted23 May 2002

protein is expected to be an important contributor to the binding enthalpy, the statistical analysis requires that each protein is represented by a minimum of two complexes with ligands that target the same site and induce the same protein conformation. Because of these constraints, the number of datasets was limited to seven protein systems, for which high-resolution structures and thermodynamic data for multiple ligands are available. These proteins are the HIV-1 protease (seven ligands), dihydrodipicolinate reductase (four ligands), Rnase T1 (two ligands), streptavidin (three ligands), pp60c-Src SH2 domain (three ligands), Hsp90 molecular chaperone (two ligands), and bovine β-trypsin (four ligands). Because the HIV-1 protease contains the most extensive dataset, it was used as the primary training set in the analysis. The resulting parameterization accounts for the binding enthalpies of all ligands considered with a standard error of ±0.3 kcal · mol^{-1} at 25°C.

RESULTS AND DISCUSSION

Enthalpy Changes Associated with Protein Denaturation

It is now well established that the enthalpy change for protein denaturation scales in terms of changes in solvent accessible surface areas (ASA) for different atom types associated with the transition between native and denatured states. In the most detailed parametric equation, the scaling coefficients for the changes in solvent accessibility are a function of the atomic packing density (δ) of the native structure.[7]

$$\Delta H(T,\delta) = \sum_i a_i(T,\delta) \times \Delta ASA_i \qquad (1)$$

where ΔASA_i represents the changes in solvent accessible surface area for atoms of type i and $a_i(T,\delta)$ their corresponding scaling coefficients. The $a_i(T,\delta)$ coefficients are a function of temperature and the atomic packing density δ. There are two limits to these coefficients, the zero packing density limit in which the coefficients approximate the solvation enthalpy from the gas phase and the high packing density limit in which the coefficients contain an additional contribution corresponding to the enthalpy of sublimation of the atom type under consideration.[7] This approach was able to predict the denaturation enthalpy of the proteins studied with a standard deviation of 4 kcal · mol^{-1}. Because the interior of proteins is highly homogeneous in terms of both atomic composition and packing density, it has been found that highly simplified forms of Equation 1 are able to account for experimental data within error.[2,7,8] In particular, an equation that contains only polar and nonpolar atom types and assumes a constant packing density has been widely used[2,7,8]:

$$\Delta H(T) = a(T) \times \Delta ASA_{ap} + b(T) \times \Delta ASA_{pol} \qquad (2)$$

where $a(T)$ and $b(T)$ are empirically determined coefficients and ΔASA_{apolar} and ΔASA_{polar} are the changes in solvent accessible surface area for nonpolar and polar atoms, respectively. Analysis of 60 proteins by Robertson and Murphy[2] in terms of Equation 2 yielded $a(60) = -1.9 \pm 2.6$ cal · mol^{-1} · Å^{-2} and $b(60) = 20.6 \pm 4.1$ cal · mol^{-1} · Å^{-2}. The regression coefficient was 0.9 with a zero intercept and a slope of 0.96. The standard deviation between experimental and calculated values was close to 10 kcal mol^{-1} (~10% of the mean denaturation enthalpy value).

Prediction of Binding Enthalpies

When applied to ligand binding, the structural parameterization of the enthalpy change derived from protein denaturation and embodied in Equation 2 has yielded inconsistent results.[9–12] The observed inconsistency can be attributed to inherent errors in the protein stability parameterization and temperature extrapolation errors as well as the presence of other factors that become relevant in the structural analysis of the binding enthalpy.

Reference temperature.

The structural parameterization of the enthalpy change for protein stability was obtained at 60°C, which is the median temperature for protein heat denaturation and also the temperature at which the experimental data are obtained. Ligand binding is usually measured at or close to 25°C, which requires a 35°C extrapolation. So, in the application of the protein stability parameterization to ligand binding, the total error is compounded by both the intrinsic error in the enthalpy and heat capacity parameterizations. In addition, because the magnitude of binding enthalpies is about one order of magnitude smaller than the measured denaturation enthalpies, any error is significantly amplified. A 10% error in the prediction of the denaturation enthalpy corresponds to a 100% error in the binding enthalpy. The parameterization of the binding enthalpy needs to be derived from protein/ligand structure/thermodynamic datasets and refined at 25°C, which is the temperature at which most binding experiments are performed.

Protein conformation.

The binding of small ligands is usually associated with a change in protein conformation. This change in protein conformation is not necessarily of a global character; in fact, most of the time it only involves a local rearrangement or stabilization of unstructured regions near the binding site.[13] In addition, the ligand itself may occupy enthalpically different states in the free and bound states. Because binding enthalpies are usually small, contributions due to conformational changes, especially of the protein, need to be considered explicitly even if they are only of a local nature. In theory, the availability of the free and bound conformations provides the necessary information. In practice, however, the accumulation of small crystallographic differences (e.g., orientation of exposed side chains in regions remote from the binding site) is often larger than the effects due to the binding of low-molecular-weight ligands. One way to address this issue is to use the structure of the complex alone and consider the conformational enthalpy as a fitting parameter in the parameterization equation. This approach, however, requires a minimum of two structure/thermodynamic datasets with

different ligands for each protein and the condition that the ligands induce the same bound conformation of the protein. This approach will be presented here.

Linkage to protonation/deprotonation reactions

Ligand binding is frequently coupled to the protonation or deprotonation of certain ionizable groups either in the protein or the ligand itself.[9,14,15] Depending on the nature of the ionizing groups, ionization enthalpies can be of the same order of magnitude as the intrinsic binding enthalpy itself and therefore need to be considered explicitly. Calculation of the ionization contribution to the binding enthalpy requires knowledge of the p*K*'s and protonation enthalpies of the ionizable groups. In addition, when binding is coupled to a protonation/deprotonation reaction the measured enthalpy is buffer dependent.[9] At the present time, this information can only be obtained experimentally. Explicit methods to dissect proton linkage contributions to the binding enthalpy were developed by Baker and Murphy[14] and have been applied to various systems.[14,15]

Buried water molecules

The desolvation of the ligand and protein interfaces upon binding is not always complete. Frequently, long-lived, buried water molecules are found at the binding interface in many complexes. These buried water molecules may play a critical structural and energetic role in mediating the interactions between a protein and its ligands by serving as adapters that fill nonoccupied volumes, satisfying the hydrogen bonding potential of the ligand and the binding site or assisting in the dissipation of charges. All these terms are expected to contribute favorably to the binding enthalpy. Conversely, the incomplete desolvation of the ligand-protein interface will oppose this effect by a decrease in the solvation entropy (enthalpy/entropy compensation). Thus, the enthalpic effect of buried water can be expected to be larger than the corresponding effect in the Gibbs energy.

Parameterization of the Binding Enthalpy

According to the discussion above, as a first approximation the experimental binding enthalpy of small ligands, ΔH_{exp}, can be considered as the combination of at least three terms:

$$\Delta \mathrm{H}_{\mathrm{exp}} = \Delta \mathrm{H}_{\mathrm{intrinsic}} + \Delta \mathrm{H}_{\mathrm{conformation}} + \Delta \mathrm{H}_{\mathrm{protonation}} \qquad (3)$$

The intrinsic enthalpy, $\Delta H_{\mathrm{intrinsic}}$, is the most important term for ligand optimization because it reflects interactions between ligand and protein (hydrogen bonds, van der Waals interactions, etc.) and solvation changes upon binding. Because the conformation enthalpy is treated separately, $\Delta H_{\mathrm{intrinsic}}$ corresponds to the enthalpy that would be observed if protein and ligand had the same conformation in the free and bound states. These contributions are expected to scale with changes in accessible surface area (ΔASA) and the atomic packing density (δ),[7] and can be parameterized according to Equations 1 and 2.

The enthalpic contributions arising from conformational changes, $\Delta H_{\mathrm{conformation}}$, in the protein cannot be easily parameterized in terms of changes in solvent accessibility upon complexation. First, an accurate estimation of these contributions would require the availability of the high-resolution structure of the free protein obtained under exactly the same conditions as that of the complex. This is rarely the case. Second, because binding sites are usually characterized by the presence of regions with low structural stability,[13] the conformation observed in the crystal structure might not be representative of the native state ensemble. Third, it is common to observe some variability in the conformation of exposed side chains among structures of free and bound proteins. Small changes in solvent accessibilities due to different conformations of side chains not directly involved in binding or in the conformational change can lead to over- or underestimation of ΔASA values, and therefore, to erroneous enthalpy predictions. For example, the difference in solvent accessibility between two of the available structures of the free HIV-1 protease (pdb codes 1hhp and 3phv) is on the order of 500 Å^2, which is of the same magnitude of the area buried by a small ligand.

Contributions to the binding enthalpy arising from protonation/deprotonation processes, $\Delta H_{\mathrm{protonation}}$, if any, need to be dissected experimentally by measuring the enthalpy of binding at different pH values and with buffers characterized by different ionization enthalpies.[14,15] Once corrected for protonation/deprotonation effects, the resulting protonation-independent binding enthalpy will be the combination of intrinsic contributions and the enthalpy associated with any possible conformational change in the protein and/or the ligand upon formation of the complex.

For those situations in which several ligands induce the same bound conformation in the protein and assuming that any enthalpy associated with conformational changes in the ligand is small compared to that of the protein, the protonation-independent binding enthalpy at any given temperature, $\Delta H_{\mathrm{binding}}(T)$, will be the sum of a constant term corresponding to the conformational enthalpy plus a ligand-specific term that accounts for the specific interactions of each ligand with the target:

$$\Delta \mathrm{H}_{\mathrm{binding}}(\mathrm{T}) = \Delta \mathrm{H}_{\mathrm{conformation}}(\mathrm{T}) + \Delta \mathrm{H}_{\mathrm{intrinsic}}(\mathrm{T}) \qquad (4a)$$

$$\Delta \mathrm{H}_{\mathrm{binding}}(\mathrm{T}) = \Delta \mathrm{H}_{\mathrm{conformation}}(\mathrm{T}) + \mathrm{a}(\mathrm{T}) \cdot \Delta \mathrm{ASA}_{\mathrm{apolar}} + \mathrm{b}(\mathrm{T}) \cdot \Delta \mathrm{ASA}_{\mathrm{polar}} \qquad (4b)$$

where the intrinsic enthalpy is represented in terms of changes in solvent accessible surface areas, and $a(T)$ and $b(T)$, as in Equation 2, are the scaling coefficients for nonpolar and polar groups, respectively.

HIV-1 Protease Database

The binding of inhibitors and substrates to the HIV-1 protease is coupled to a conformational change in the protein. The flexible beta-hairpin loops (known as "flaps") located at the top of the binding site (Fig. 1) become stabilized into a unique conformation that closes the binding cavity upon binding. Also, inhibitor binding induces a rigid body rotation of the two subunits that tightens slightly the binding site cavity. The bound confor-

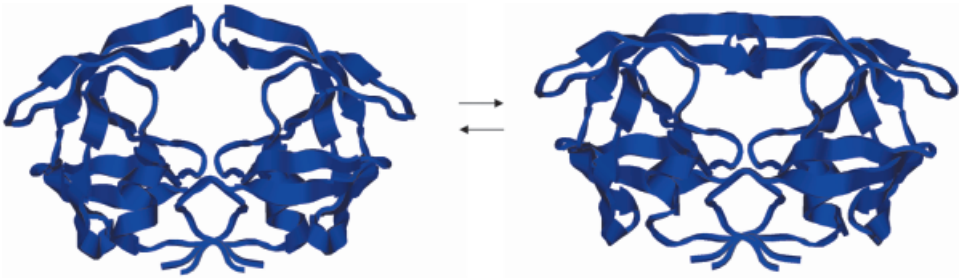

Fig. 1. The conformational change between the free and bound conformations of the HIV-1 protease. Upon substrate or inhibitor binding the flaps are structurally stabilized and adopt a closed conformation. The energetics associated with this conformational change contributes to the binding energetics and consequently to the binding affinity.

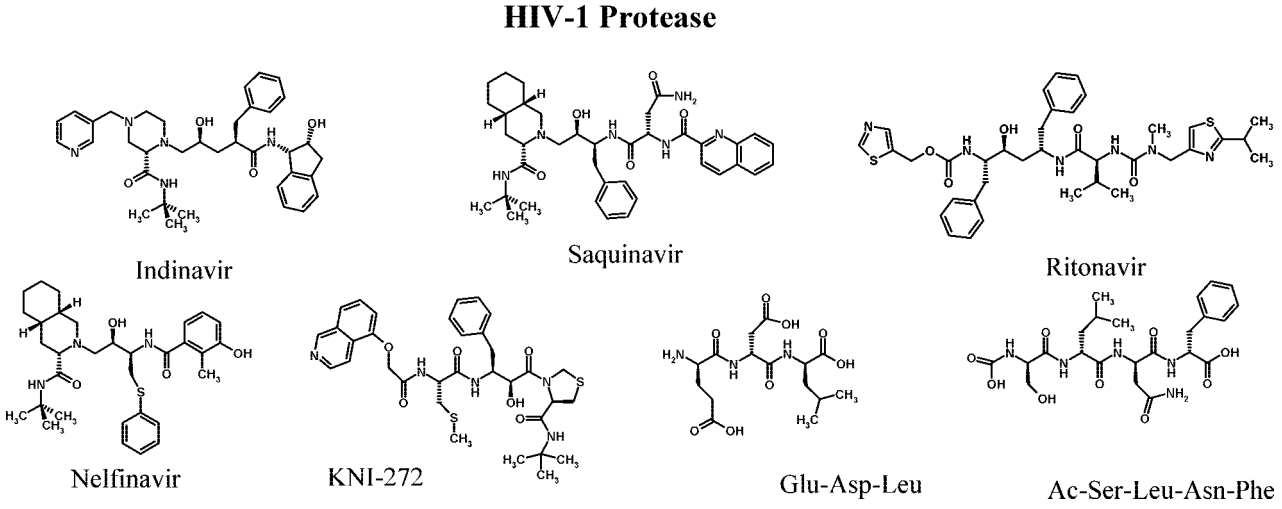

Fig. 2. Chemical structures of the HIV-1 protease inhibitors included in the structure/thermodynamic dataset.

mation of the protease is quite conserved, showing a rms deviation between C_α of $<$0.6 Å among different structures.[16] Accordingly, $\Delta H_{\mathrm{conformation}}(T)$ in Equation 4 can be assumed to be essentially constant for different inhibitors.

Currently a database of seven HIV-1 protease/inhibitor complexes for which high-resolution structures as well as a complete thermodynamic description of the binding process is available.[5,15,17,18] The database includes four inhibitors currently in clinical use (Indinavir, Nelvinavir, Saquinavir, and Ritonavir), the allophenylnorstatine-based inhibitor KNI-272, the tripeptidic inhibitor EDL from the transframe region of Gag-Pol, and the substrate-hydrolysis product SLNF. The chemical structures of these compounds are shown in Figure 2. The binding enthalpy values are summarized in Table I and were measured under identical solvent conditions (10 mM sodium acetate, pH 5.0, 2% DMSO) as reported by Velazquez-Campoy et al.[5,15,18] and Luque et al. (unpublished results). All the thermodynamic values used in the analysis were obtained in the presence of 2% DMSO rather than acetonitrile,[17] because the presence of 2% DMSO does not affect the binding enthalpy by more than $\pm$0.1 kcal $\cdot$ mol^{-1}, judging by measurements of the peptidic protease inhibitor acetyl pepstatin in the absence and presence of 2% DMSO and the clinical inhibitor Indinavir at 1 and 2% DMSO (unpublished data from this laboratory).

ASA Calculations and Water Molecules at HIV-1 Protease Binding Interface

Changes in accessible surface area were calculated according to the Lee and Richards algorithm.[19] 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 differences in solvent accessibility between small ligands, 64 different orientations of the ligands and proteins with respect to the slicing plane were considered in the calculations. The orientations are generated by rotating the molecule around the *x, y,* and *z* axis. The solvent accessibility for each atom is obtained as the numerical average of the values calculated for all molecular orientations.

The PDB files for the complexes with Indinavir, Nelfinavir, Saquinavir, Ritonavir, KNI-272, EDL, and SLNF (pdb codes 1hsg, 1ohr, 1hxb, 1hxw, 1hpx, 1a30, and 1yth, respectively) were used in the analysis. The changes in accessible surface areas used in the estimation of the intrinsic contributions to the binding enthalpy [$\Delta H_{\mathrm{intrinsic}}(T)$] were calculated by considering the structures of the free protein and ligand to be identical to those

TABLE I. Summary of Structure/Thermodynamic Analysis for Seven Proteins in Database

Protein	Ligand	ΔASA Apolar (Å^2)	ΔASA Polar (Å^2)	Number Waters	ΔH_{calc} (25°C) (kcal · mol^{-1})	ΔH_{exp} (25°C) (kcal · mol^{-1})	ΔH_{conf} (25°C) (kcal · mol^{-1})
HIV-1 protease	Indinavir	−1035	−361	2	2.34	2.10	5.9 ± 1.3
	Nelfinavir	−849	−339	2	1.66	2.60	
	Saquinavir	−921	−336	1	2.26	1.90	
	Ritonavir	−947	−450	2	−1.08	−3.70	
	KNI-272	−833	−612	6	−6.96	−6.30	
	EDL	−393	−410	1	−3.92	−5.21	
	SLNF	−502	−378	1	−2.10	−1.91	
DHPR	NADH	−557	−548	0	−7.82	−8.93	5.1 ± 0.9
	NHDH	−554	−600	0	−9.47	−8.36	
	NADPH	−507	−648	0	−11.29	−10.98	
	AcNADH	−578	−556	1	−7.92	−8.23	
Rnase T1	2′ GMP	−183	−391	0	−8.90	−9.00	1.9 ± 0.1
	3′ GMP	−193	−356	0	−7.74	−7.64	
Streptavidin	Haba 2-((4′hydroxyphenyl)-azo)benzoate (HABA)	−425	−195	1	1.60	1.70	4.5 ± 0.1
	3′ Methyl HABA	−469	−189	1	2.10	2.15	
	3′5′ diMethyl HABA	−455	−207	1	1.44	1.28	
pp60^{eSre} SH_2	Ace-phosphoTyr-Glu2 (N,N dipentylamine) (compound 1)	−449	−289	0	−5.60	−4.30	0.07 ± 1.4
	Ace-phosphoTyr-Glu2 (N,Me (−(CH_2)3 cyclopentyl)) (compound 4)	−477	−297	0	−5.65	−7.10	
	Ace-phosphoTyr-Glu2 (3-butylpeperidine) (compound 5)	−456	−289	0	−5.56	−5.40	
Hsp60	Geldanamycin	−655	−556	5	−10.88	−10.70	1.6 ± 0.3
	Radicicol	−480	−334	3	−5.27	−5.45	
Trypsin	n-α-(2-naphthylsulfonyl)-n-(3-amidino-1-phenylalanilyl)-d-pipecolinic acid (1b)	−367	−473	4	−10.80	−10.9	1.2 ± 0.8
	n-α-(2-naphthylsulfonyl)-n(3-amidino-1-phenylalaninyl) isopipecolinic acid methyl ester (1bMe)	−465	−419	3	−8.42	−9.47	
	n-α-(2-naphthylsulfonyl)-3-amidino-1-phenylalanine piperazide (1d)	−432	−409	3	−8.33	−7.87	
	n-α-(2-naphtylsulfonyl)-3-amidino-1-phenylalaninyl-4-acetyl-piperazine (1dAc)	−439	−429	3	−8.93	−8.23	

observed in the complex. It is important to keep in mind that, as defined here, the intrinsic binding enthalpy reflects the interactions between ligand and protein and the changes in solvation in the absence of any conformational change. Any contributions to the binding enthalpy arising from conformational changes of the protein and/or the ligand upon binding are reflected in the $\Delta H_{conformation}$ term, which is left as a fitting parameter. In principle, $\Delta H_{conformation}$ includes changes in conformation of protein and ligand; however, for small ligands, conformationally related differences in accessible surface area will be small compared with those associated with the protein. For example, the maximal differences in solvent accessible surface area between all backbone and side-chain conformations accessible to a phenylalanine residue in a Ala-Phe-Ala tripeptide are 20 and 17 Å of apolar and polar accessible surface area, respectively. These changes will result in a maximum enthalpic contribution to the binding enthalpy of 380 cal/mol, according to Equation 5b. Accordingly, enthalpic contributions associated with conformational changes are expected to predominantly originate from the protein.

The changes in solvent accessible surface area upon binding are calculated considering the contribution of water molecules at the protein/ligand interface. These water molecules are included in the analysis by considering them as part of the ligand (i.e. upon binding the ligand is not completely desolvated and retains some of the solvation waters). This approach assumes that these molecules are not associated with the free protein in solution. In fact, for the HIV-1 protease none of the water molecules summarized in Table II are present in the existing three dimensional structures of the free protein (pdb codes 1hhp, 1hps, 3hpv, 3phv).

A considerable number of water molecules are observed in the crystal structures of the complexes with the HIV-1 protease; some of these water molecules are buried at the binding interface, bridging the inhibitor molecule and protein. The number and position of water molecules at the interface varies from one complex to another, although

TABLE II. Water Molecules at HIV-1 Protease/Inhibitor Interfaces and Their Solvent Accessibilities[†]

Inhibitor					
Indinavir	w5 (0.00 Å^2)[a]	w13 (5.86 Å^2)	w22 (0.00 Å^2)[b]		
Saquinavir	w1 (0.00 Å^2)[b]	w42 (33.77 Å^2)	w46 (19.63 Å^2)	w73 (7.55 Å^2)	
Nelfinavir	w303 (0.00 Å^2)[b]	w307 (0.00 Å^2)[a]	w315 (7.52 Å^2)	w422 (4.48 Å^2)	w510 (31.99 Å^2)
Ritonavir	w1 (0.0 Å^2)[b]	w5 (0.0 Å^2)[a]	w14 (14.34 Å^2)	w20 (22.17 Å^2)	w25 (5.59 Å^2)
	w36 (0.95 Å^2)	w41 (21.47 Å^2)	w44 (18.84 Å^2)	w50 (11.95 Å^2)	w51 (23.66 Å^2)
	w52 (0.30 Å^2)	w70 (33.83 Å^2)	w72 (26.72 Å^2)	w10 (22.48 Å^2)	w15 (43.93 Å^2)
EDL	w1169 (0.00 Å^2)[a]	w1075 (8.92 Å^2)	w1076 (9.71 Å^2)	w1087 (35.65 Å^2)	w1092 (34.43 Å^2)
	w1097 (4.07 Å^2)[b]				
SNLF	w401 (1.06 Å^2)[b]	w424 (0.01 Å^2)[a]	w429 (13.35 Å^2)	w430 (15.08 Å^2)	
KNI-272	W301 (0.00 Å^2)[b]	W607 (0.29 Å^2)[a]	W406 (0.00 Å^2)	W426 (0.00 Å^2)	W566[†] (0.36 Å^2)
	W608 (9.15 Å^2)				

[†]Water molecules are labeled according to their number in the original pdb files. The values in parentheses are the solvent accessibilities in the complex.
[a]The ubiquitous water molecule that bridges the inhibitors and the flaps.
[b]The water molecule at the P2 subsite that interacts with amino acids Arg8, Leu23, Gly27, Ala28, and Asp29.

a tetrahedrally hydrogen-bonded water molecule that connects the flaps of the protease to the inhibitor is observed in nearly every crystal structure of the inhibited protein.

Ideally, it would be desirable to have information about the persistence of water molecules at the interface for each of the complexes considered. However, direct observation of long-lived water molecules in solution has only been reported for one of the inhibitors in the database: the allophenylnorstatine-based inhibitor KNI-272.[20] In this case, a total of six long-lived water molecules have been observed both in the crystal and in solution. Four of these water molecules (waters 301, 566, 607, 608 in the crystal structure) are in direct contact with the inhibitor at distances smaller that 5Å; the other two (waters 406 and 426) are located at longer distances (7Å), but still remain completely buried and tightly bound to the protease/KNI-272 complex.[20] All six water molecules are considered in the analysis. For the other complexes only information from the crystallographic data is available; hence some criteria need to be established to consistently select those water molecules at the interface that are likely to be influential to the binding enthalpy.

It is apparent that only those water molecules at the binding interface that are significantly buried (ASA $\leq$ 10 Å^2) need to be considered (Table II). However, no precise guidelines for possible cutoff values are available. In order to develop a consistent criteria, the analysis of the thermodynamic data in terms of Equation 4 was performed using different cutoff values. For each value, the structure/thermodynamic dataset was analyzed by nonlinear least squares in terms of Equation 4b. The fitting parameters were $\Delta H_{conformation}$, $a(25)$ and $b(25)$. In each case ΔASA_{ap} and ΔASA_{pol} were calculated by including the water molecules within the cutoff limits. The best statistical correlation between calculated and experimental binding enthalpies was obtained when only fully buried water molecules within 5–7 Å of the inhibitor were considered. These conditions resulted in a coefficient of multiple determination (R^2) of 0.9. More relaxed criteria, allowing for the consideration of partially buried water molecules resulted in poorer correlations. More importantly, neglecting buried water molecules yielded a very low coefficient of multiple determination (R^2) of 0.7. These results indicate that buried water molecules are important contributors to a favorable binding enthalpy. Two effects may be responsible for this contribution: first, interfacial water molecules improve the atomic packing density within the binding site, bringing it to values similar to those of the protein interior, and second, water molecules that remain associated to the inhibitor diminish the unfavorable enthalpy associated with the complete desolvation of a hydrophobic molecule.

The location of the fully buried water molecules (waters 5 and 22 for indinavir; water 1 for saquinavir; waters 303 and 307 for nelfinavir; waters 1 and 5 for ritonavir; water 1169 for EDL; and water 424 for SNLF, using the numbering in their corresponding pdb files) is very conserved among all the complexes. In all cases these water molecules correspond either to the ubiquitous water molecules at the tip of the flaps or to a water molecule occupying the P2 subsite, in contact with amino acids Arg 8, Leu23, Gly 27, Ala 28, and Asp29.

Results for HIV-1 Protease Structure/Thermodynamic Data Set

Nonlinear least squares analysis of the HIV-1 protease in terms of Equation 4 yielded the following parameters:

$$\Delta H_{binding}(25) = \Delta H_{conformation}(25) + a(25) \cdot \Delta ASA_{apolar} + b(25) \cdot \Delta ASA_{polar} \quad (5a)$$

$$5900\ (\pm\ 1300) - 7.35(\pm\ 2.55) \cdot \Delta ASA_{apolar} + 31.06(\pm\ 6.32) \cdot \Delta ASA_{polar} \quad (5b)$$

where the enthalpy values are in cal $\cdot$ mol^{-1}, ΔASA_{apolar} and ΔASA_{polar} are in Å^2 and the $a(25)$ and $b(25)$ coefficients in cal $\cdot$ (mol $\cdot$ Å^2)$^{-1}$, respectively.

The accuracy of Equation 5 is illustrated in Figure 3 where the experimental (solid bars) and calculated (hatched bars) enthalpies are compared. The analysis accounts for the experimental enthalpy with a standard error of $\pm$0.4 kcal $\cdot$ mol^{-1}. The conformational enthalpy is small and

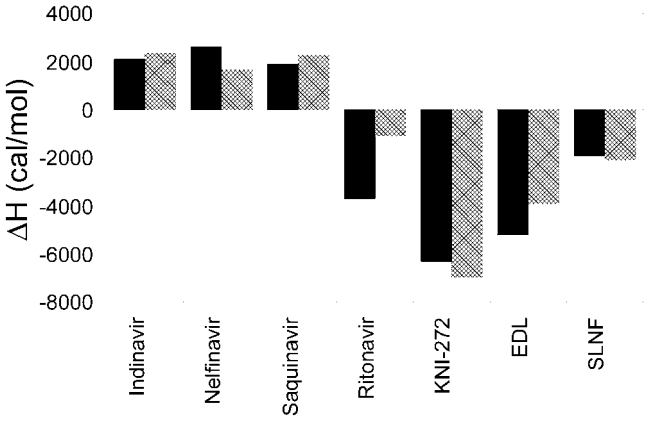

Fig. 3. Experimental (solid bars) and calculated (hatched bars) enthalpies for seven inhibitors of the HIV-1 protease. This graph illustrates the accuracy of a protein-specific refinement of the structural parameterization of the binding enthalpy. In this case, the enthalpies of binding are calculated from structure with a standard error of ± 0.4 kcal $\cdot$ mol^{-1}.

positive, indicating that the free conformation is enthalpically favored over the bound conformation. The values of the scaling coefficients $a(25)$ and $b(25)$ in Equation 5b define an initial parameterization of the intrinsic enthalpy with the limited dataset currently available. As more proteins are added to the dataset, these coefficients will be refined to more robust values. On a relative basis, the standard error for $a(25)$ is larger than that for $b(25)$ (35% versus 20%). Accordingly, it is expected for $a(25)$ to undergo more pronounced variations as the size of the database increases and the coefficients converge to a stable value. In this respect, an earlier attempt at parameterizing the binding enthalpy with a more restricted dataset yielded a value for $a(25)$ of -4.7 ± 2.3 cal $\cdot$ (mol $\cdot$ Å^2)$^{-1}$ and a value of 31.9 ± 5 cal $\cdot$ (mol $\cdot$ Å^2)$^{-1}$ for $b(25)$.[21] Clearly, $a(25)$ has exhibited the largest change, whereas $b(25)$ has essentially the same value reported here. Interestingly, the values in Equation 5b fall within the error brackets of the protein denaturation coefficients derived by Robertson and Murphy[2] when extrapolated to the same temperature.

Extension to Other Systems

The applicability of the scaling coefficients $a(25)$ and $b(25)$ derived from the analysis of the HIV-1 protease dataset to other systems was tested by using literature structure/thermodynamic datasets for other proteins. For this analysis we identified six proteins for which the requirement of structure/thermodynamic data for at least two ligands was satisfied: (1) Binding of the nucleotide substrates NADH (1dru.pdb), NHDH (1drw.pdb), NADPH (1dih.pdb), and AcNADH (1drv.pdb) to dihydrodipicolinate reductase, a component of the biosynthetic pathway leading to L-lysine in bacteria[22]; (2) The interaction of guanidine nucleotides 2′GMP (1bui.pdb) and 3′GMP (6gsp.pdb) with ribonuclease T1, that catalyzes the hydrolysis of single-stranded ribonucleic acid at the 3′ side of guanosine residues in RNA[23]; (3) Binding of HABA (2-[(4′-hydroxyphenyl)-azo]benzoate) derivatives to streptavidin, a biotin-binding protein secreted by *Streptomyces avidinii*; these compounds [HABA (1sre.pdb), 3′-methylHABA (1srg.pdb), and 3′,5′-dimethylHABA (1sri.pdb)] are dyes used to quantitate the concentrations of biotin binding sites in streptavidin solutions[24–26]; (4) Binding of a series of peptide ligands to the pp60c-Src SH2 domain,[27] which are derivatives of the general motif *N*-acetylphosphotyrosine-Glu: Ace-pTyr-Glu2(*N,N* dipentylamine) (1A1B.pdb), Ace-pTyr-Glu2(*N*,Me(-(CH_2)3cyclopentyl)) (1A1C.pdb), and Ace-pTyr-Glu2(3-butylpiperidine) (1A1E.pdb); (5) Binding of the antitumor antibiotics Geldanamycin (1A4H.pdb) and Radicicol (1bgq.pdb) to the Hsp90 molecular chaperone[28]; and (6) Binding of a series of closely related N^{α}-(2-naphthylsulphonyl)-L-3-amidino-phenylalanine derivatives to bovine β-trypsin[29]: N^{α}-(2-naphthylsulphonyl)-L-(3-amidino-phenylalanylyl)-*d*-pipecolinic acid (1k1i.pdb), N^{α}-(2-naphthylsulphonyl)-L-(3-amidino-phenylalanilyl) isopipecolinic acid methyl ester (1k1j.pdb), N^{α}-(2-naphthylsulphonyl)-L-3-amidino-phenylalanine piperazide (1k1m.pdb), and N^{α}-(2-naphthylsulphonyl)-L-3-amidino-phenylalanilyl-4-acetyl-piperazine (1k1n.pdb).

Superimposition of the bound protein structures considered in the analysis confirmed that the bound conformations of each protein can be considered equivalent (backbone RMS values were found to range between 0.19 and 0.46 Å for dihydrodipicolinate reductase, 0.5 Å for ribonuclease T1, 0.25–0.27 Å for streptavidin, 0.18–0.28 Å for pp60$^{c\text{-}Src}$ SH2 domain, 0.37 Å for Hsp 90 molecular chaperone, and 0.11–0.34 for β-trypsin). The structures of a representative complex for each family are shown in Figure 4, whereas the chemical structures of the ligands are summarized in Figure 5. These ligands represent a heterogeneous group of molecules differing in size, structure, and chemical composition that bind target proteins with different structures, active site geometries, and binding modes.

For the systems considered, there were some additional ligands not included in the analysis because they were reported to induce different bound conformations of the protein. This was the case for the binding of biotin to streptavidin. According to the authors, the loop 48–51 of the protein is better ordered in the complex with biotin than in the complexes with the HABA derivatives.[24–26] Also, in the crystal structure, some residues of the peptide ligand Phe-Ser-His-Pro-Gln-Asn-Thr are not well resolved. These two ligands were not included in the analysis. A similar situation was found for the binding of peptide derivatives to pp60c-Src SH2 domain. Of the 13 ligands for which the thermodynamic analysis has been described, only the structures for compounds 1, 4, 5, 8, 11, 12, and 13 were available in the PDB database (pdb codes 1a1b, 1a1c, 1a1e, 1shd, 1ao7, 1ao8, and 1ao9, respectively). Of these, compound 8 was not fully resolved in the crystal structure, compounds 11 and 12 induced a different conformation of the BC loop in the complex, and compound 13 was found to interact covalently with the enzyme.[27]

For each system, the binding thermodynamics were analyzed using the $a(25)$ and $b(25)$ parameters derived from the HIV-1 protease dataset, leaving $\Delta H_{\text{conformation}}$ as

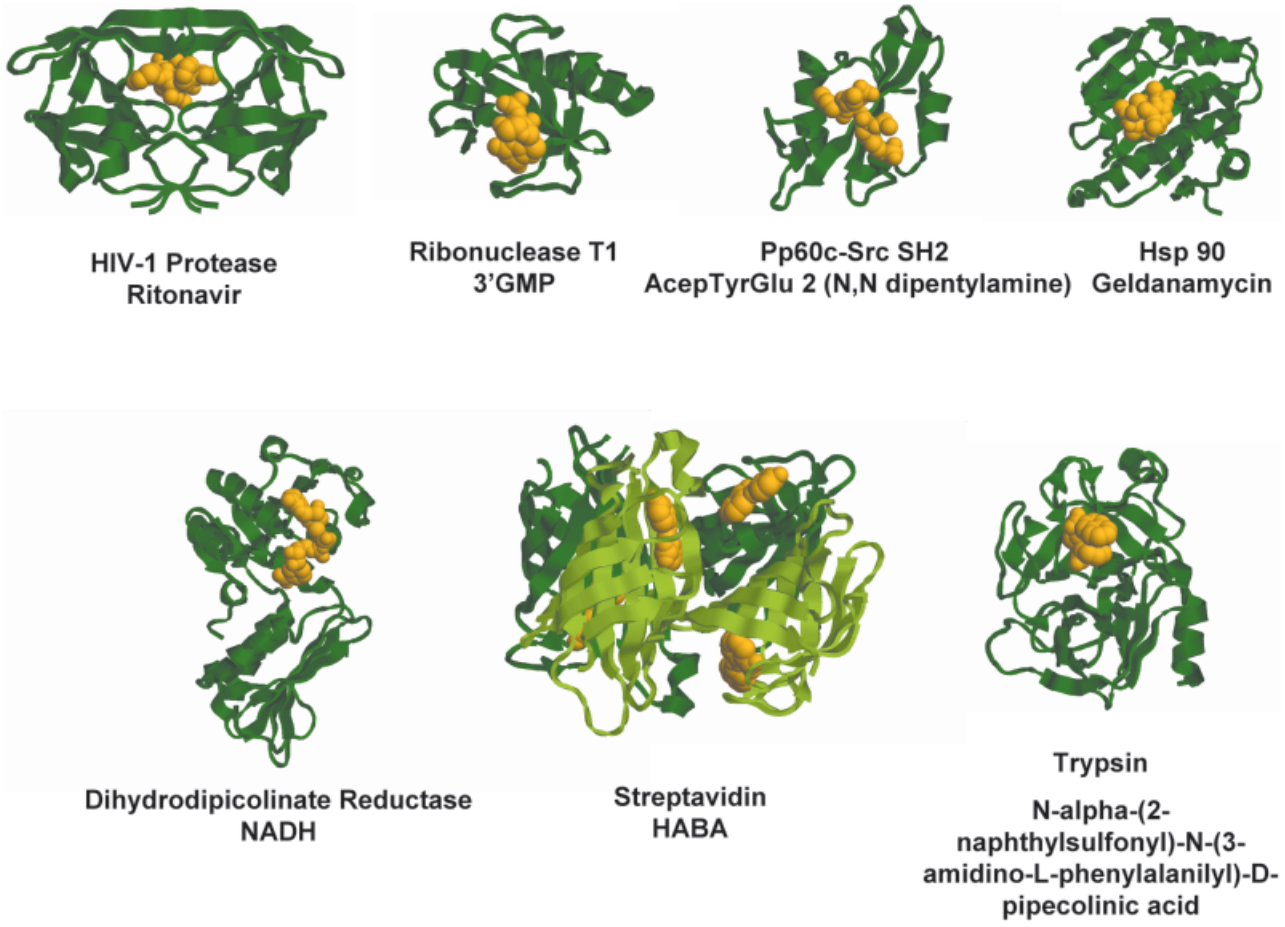

Fig. 4. Crystallographic structures of the proteins included in the structure/thermodynamic database. Each protein (green) is displayed bound to a representative member of each family of ligands. The complete list is in Table I. The inhibitors are shown in yellow using a CPK representation.

the only adjustable parameter for each protein. The complete results from the analysis are summarized in Table I. The experimentally determined (solid bars) and the calculated binding enthalpy values (hatched bars) for all complexes are shown in Figure 6. The differences between calculated and experimental values are in all cases <1.5 kcal $\cdot$ mol^{-1} with an average error of 0.3 kcal $\cdot$ mol^{-1}. The fact that the sum of the differences is very close to zero (-0.006 kcal $\cdot$ mol^{-1}) with a standard deviation of 0.7 kcal $\cdot$ mol^{-1} is indicative of the lack of systematic errors in the calculated values. The values for the conformational enthalpies obtained from these calculations are 5.1 $\pm$ 0.9 kcal $\cdot$ mol^{-1} for dihydrodipicolinate reductase, 1.9 $\pm$ 0.2 kcal $\cdot$ mol^{-1} for ribonuclease T1, 4.5 $\pm$ 0.1 kcal $\cdot$ mol^{-1} for streptavidin, 0.1 $\pm$ 1.4 kcal $\cdot$ mol^{-1} for pp60$^{c\text{-}Src}$ SH2 domain, 1.6 $\pm$ 0.3 kcal $\cdot$ mol^{-1} for Hsp90, and 1.2 $\pm$ 0.8 kcal $\cdot$ mol^{-1} for β-trypsin. The results indicate that for each protein, $\Delta H_{\text{conformation}}$ assumes a single value with a small standard deviation in accordance with the assumptions of the model.

The agreement between calculated and experimental binding enthalpies in the dataset is remarkable, considering that not all the literature data has been corrected for protonation-deprotonation. For example, the binding affinity of compound 1 [Ace-pTyr-Glu2(*N,N* dipentylamine)] to pp60^{cSrc} SH_2 shows a dependence on pH, although no correction of the binding enthalpy has been reported. Fortunately, the observed pH dependence appears to be linked to the protonation state of a phosphate group, which is known to exhibit a small ionization enthalpy.[27] The data for β-trypsin, has been corrected for protonation effects. The enthalpy values for the HIV-1 protease are buffer-independent values corrected for buffer protonation enthalpies. Except for the inhibitor KNI-272,[15] which itself contains an ionizable isoquinoline group that protonates upon binding with an enthalpy of -6.3 kcal/mol, no correction has been made for the enthalpic contributions due to ligand or protein protonation/deprotonation of the other inhibitors. In the absence of ligand ionizable groups that protonate/deprotonate upon binding, binding to the protease is only coupled to a fractional change in the degree of protonation of a carboxylic group, which is expected to contribute <1 kcal/mol[15,17] to the total enthalpy.

CONCLUSIONS

The results presented in this work constitute a first attempt at a structural-based parameterization of the binding enthalpy at 25°C. The results indicate that in the

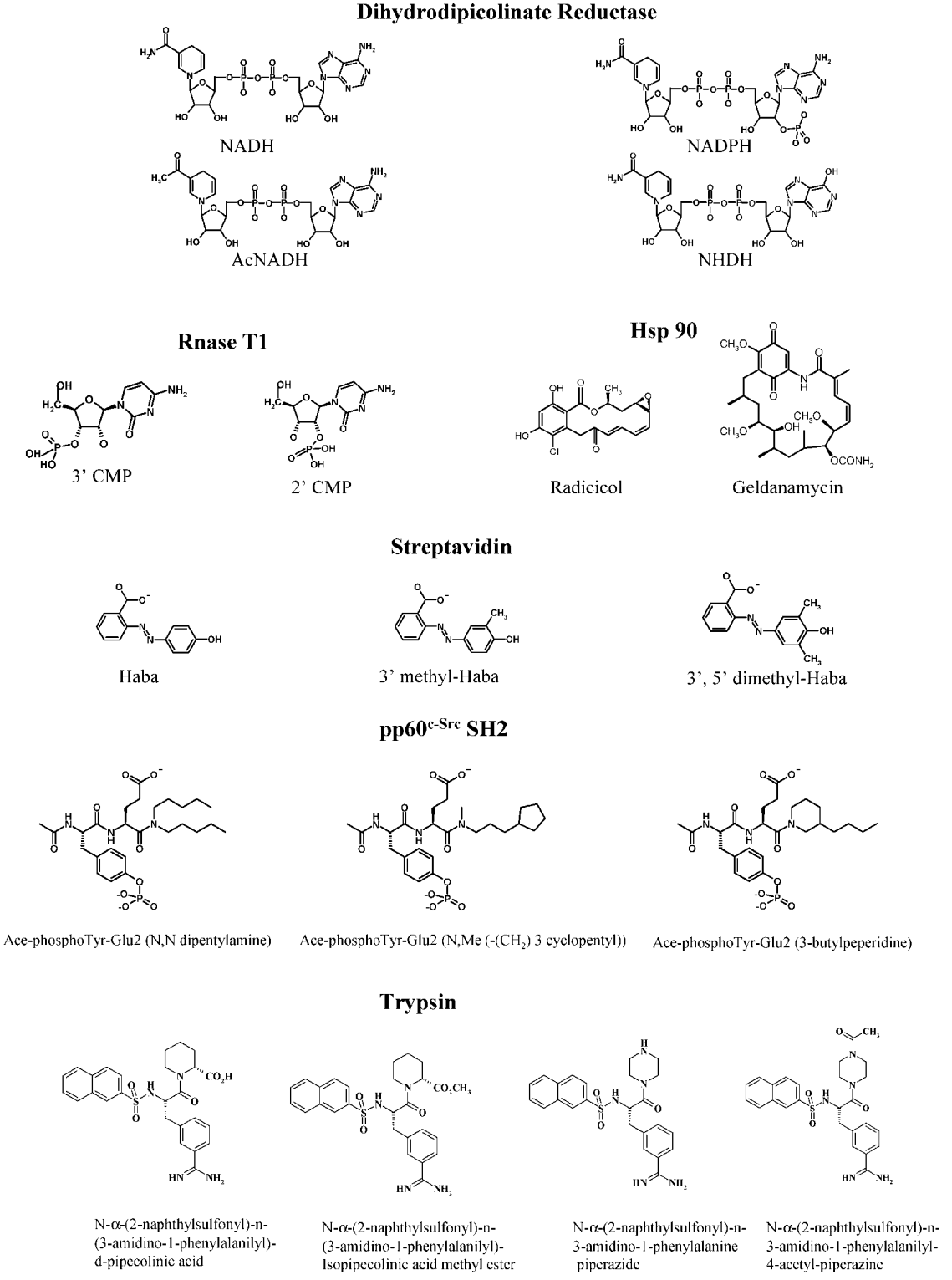

Fig. 5. Chemical structures of the ligands considered in our study.

absence of or after correction for protonation effects, the binding enthalpy can be accounted for in terms of an intrinsic binding enthalpy that scales with changes in solvent accessibilities and a conformational enthalpy change. Changes in solvent accessibility need to be calculated by taking into consideration buried water molecules within 5–7Å of the ligand. The resulting parametric equation accounts for the binding enthalpy of the ligands

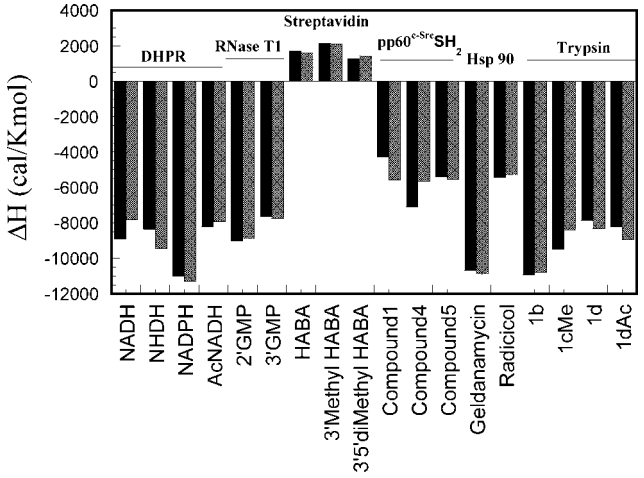

Fig. 6. Binding enthalpies of several inhibitors to each of the proteins considered. Solid bars represent experimental values and hatched bars the binding enthalpies calculated using the parameterization obtained from HIV-1 protease data. The overall standard error is <0.3 kcal $\cdot$ mol^{-1}. In the graph, compound 1 refers to Ace-phosphoTyr-Glu2 (*N,N* dipentylamine); compound 2 to Ace-phosphoTyr-Glu2(*N*, Me (-(CH_2) 3 cyclopentyl)); compound 3 to Ace-phosphoTyr-Glu2(3-butylpeperidine); 1b to *n*-α-(2-naphthylsulfonyl)-*n*-(3-amidino-1-phenylalanilyl)-*d*-pipecolinic acid; 1bMe to *n*-α-(2-naphthylsulfonyl)-*n*-(3-amidino-1-phenylalanilyl)-isopipecolinic acid methyl ester; 1d to *n*-α-(2-naphthylsulfonyl)-*n*-3-amidino-1-phenylalanine piperazide; and 1dAc to n-α-(2-naphthylsulfonyl)-*n*-3-amidino-1-phenylalanilyl-4-acetyl-piperazine.

considered in this article, with an error of 0.3 kcal $\cdot$ mol^{-1}, a significant improvement with respect to previous parameterizations. It is expected that the implementation of this approach in design algorithms will contribute to better predictions of binding energetics and to the development of more rational and efficient ligand design strategies.

The authors thank Adrian Velazquez-Campoy and Stephanie Leavitt for helpful comments on this manuscript.

REFERENCES

1. Pace NC, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. Biophys J 1998;75:422–427.
2. Robertson AD, Murphy KP. Protein structure and the energetics of protein stability. Chem Rev 1997;97:1251–1267.
3. 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. Prot Sci 2000;9:252–264.
4. Sundberg EJ, Urrutia M, Braden BC, Isern J, Tsuchiya D, Fields BA, Malchiodi EL, Tormo J, Schwarz FP, Mariuzza RA. Estimation of the hydrophobic effect in an antigen-antibody protein-protein interface. Biochemistry 2000;39:15375–15387.
5. Velazquez-Campoy A, Todd MJ, Freire E. HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity. Biochemistry 2000;39:2201–2207.
6. Velazquez-Campoy A, Kiso Y, Freire E. The binding energetics of first and second generation HIV-1 protease inhibitors: implications for drug design. Arch Biochim Biophys 2001;390:169–175.
7. Hilser VJ, Gomez J, Freire E. The enthalpy change in protein folding and binding. refinement of parameters for structure based calculations. Proteins 1996;26:123–133.
8. Xie D, Freire E. Molecular basis of cooperativity in protein folding. V. Thermodynamic and structural conditions for the stabilization of compact denatured states. Proteins Struct Func Genet 1994;19:291–301.
9. Gomez J, Freire E. Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin. J Mol Biol 1995;252:337–350.
10. Luque I, Gomez J, Semo N, Freire E. Structure-based thermodynamic design of peptide ligands. Application to peptide inhibitors of the aspartic protease endothiapepsin. Proteins 1998;30:74–85.
11. Edgcomb SP, Murphy KP. Structural energetics of protein folding and binding. Curr Opin Biotechnol 2000;11:62–66.
12. Henriques DA, Ladbury JE, Jackson RM. Comparison of binding energetics of SrcSH2-phosphotyrosyl peptides with structure-based prediction using surface area based empirical parameterization. Prot Sci 2000;9:1975–1985.
13. Luque I, Freire E. The structural stability of binding sites. Consequences for binding affinity and cooperativity. Proteins 2000;4:63–71.
14. Baker BM, Murphy KP. Evaluation of linked protonation effects in protein binding using isothermal titration calorimetry. Biophys J 1996;71:2049–2055.
15. Velazquez-Campoy A, Luque I, Todd MJ, Milutinovich M, Kiso Y, Freire E. Thermodynamic dissection of the binding energetics of KNI-272, a powerful HIV-1 protease inhibitor. Prot Sci 2000;9:1801–1809.
16. Wlodawer A, Wondrasek J. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. Annu Rev Biophys Biomol Struct 1998;27:249–284.
17. Todd MJ, Luque I, Velazquez-Campoy A, Freire E. The thermodynamic basis of resistance to HIV-1 protease inhibition. Calorimetric analysis of the V82F/I84V active site resistant mutant. Biochemistry 2000;39:11876–11883.
18. Velazquez-Campoy A, Vega S, Freire E. Amplification of the effects of drug-resistance mutations by background polymorphisms in HIV-1 protease from African subtypes. Biochemistry 2002;41:8613–8619.
19. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.
20. Wang YX, Freedberg DI, Wingfield PT, Stahl SJ, Kaufman JD, Kiso Y, Bhat TN, Erickson JW, Torchia DA. Bound water molecules at the interface between the HIV-1 protease and a potent inhibitor, KNI-272, determined by NMR. J Am Chem Soc 1996;118:12287–12290.
21. Leavitt S, Freire E. Direct measurement of protein binding energetics by isothermal titration calorimetry. Curr Opin Struct Biol 2001;11:560–566.
22. Reddy SG, Scapin G, Blanchard TS. Interaction of pyridine nucleotide substrates with *Escherichia coli* dyhydrodipicolinate reductase: thermodynamic and structural analysis of binary complexes. Biochemistry 1996;35:13294–13302.
23. Hu C, Sturtevant JM. Thermodynamic of binding of mononucleotides to ribonuclease T1. J Phys Chem 1992;96:4052–4056.
24. Weber PC, Wendoloski JJ, Pantoliano MW, Salemme FR. Crystallographic and thermodynamic comparison of natural and synthetic ligands bound to streptavidin. J Am Chem Soc 1992;114:3197–3200.
25. Weber PC, Pantoliano MW, Simons DM, Salemme FR. Structure-based design of synthetic azobenzene ligands for streptavidin. J Am Chem Soc 1994;116:2717–2724.
26. Weber PC, Pantoliano MW, Salemme FR. Crystallographic and thermodynamic comparison of structurally diverse molecules binding to streptavidin. Acta Cryst. 1995;D52:590–596.
27. Charifson PS, Shewchuk LM, Rocque W, Hummel CW, Jordan SR, Mohr C, Pacofsky GJ, Peel MR, Rodriguez M, Sternbach DD, Consler TG. Peptide ligands of pp60$^{c\text{-}Src}$ domains: a thermodynamic and structural study. Biochemistry 1997;36:6283–6293.
28. Roe SM, Prodromou C, O'Brien R, Ladbury JE, Piper PW, Pearl LH. Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics Radicicol and Geldanamycin. J Med Chem 1999;42:260–266.
29. Dullweber F, Stubbs MT, Musil D, Sturzebecher J, Klebe G. Factorising ligand affinity: a combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. J Mol Biol 2001;313:593–614.