Prediction and Analysis of Binding Affinities for Chemically Diverse HIV-1 PR Inhibitors by the Modified SAFE_p Approach

MIGUEL ARENAS, M. CARMEN VILLAVERDE, FREDY SUSSMAN
Departamento de Química Orgánica, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain

Received 19 May 2008; Revised 8 August 2008; Accepted 8 September 2008
DOI 10.1002/jcc.21147
Published online 5 November 2008 in Wiley InterScience (www.interscience.wiley.com).

Abstract: One of the biggest challenges in the "*in silico*" screening of enzyme ligands is to have a protocol that could predict the ligand binding free energies. In our group we have developed a very simple screening function (referred to as solvent accessibility free energy of binding predictor, SAFE_p) which we have applied previously to the study of peptidic HIV-1 protease (HIV-1 PR) inhibitors and later to cyclic urea type HIV-1 PR inhibitors. In this work, we have extended the SAFE_p protocol to a chemically diverse set of HIV-1 PR inhibitors with binding constants that differ by several orders of magnitude. The resulting function is able to reproduce the ranking and in many cases the value of the inhibitor binding affinities for the HIV-1 PR, with accuracy comparable with that of costlier protocols. We also demonstrate that the binding pocket SAFE_p analysis can contribute to the understanding of the physical forces that participate in ligand binding. The analysis tools afforded by our protocol have allowed us to identify an induced fit phenomena mediated by the inhibitor and have demonstrated that larger fragments do not necessarily contribute the most to the binding free energy, an outcome partially brought about by the substantial role the desolvation penalty plays in the energetics of binding. Finally, we have revisited the effect of the Asp dyad protonation state on the predicted binding affinities.

 J Comput Chem 30: 1229–1240, 2009

Key words: scoring function; HIV-1 PR inhibitor; SAFE_p; binding free energy; Asp dyad ionization state

Introduction

The single biggest challenge in the "*in silico*" screening of ligands is the development of a binding free energy or scoring function that could predict accurately the actual binding free energies, or at least the binding ranking of a set of compounds of very diverse chemical structure.[1] In this regard, one of the most frequently used testing grounds for ligand binding energy scoring functions is the HIV-1 protease (HIV-1 PR), since it is the most studied enzyme as measured by the number of inhibitor–enzyme structures that have been determined by both X-ray and NMR protocols.[2–4]

Many approaches have been applied for the prediction of the binding affinity to the HIV-1 PR.[5–16] The most basic kind of function is a simple quantitative structure activity relationship based solely on the chemical structure of the inhibitor, like the one performed on cyclic urea (CU)-based HIV-1 PR inhibitors.[5] The next level of complexity is provided by the evaluation of the energy interaction between the ligand and the enzyme.[7,8] Other groups (including us) have elaborated free energy functions that include a solvation contribution, a very important component of the inhibitor binding affinity, specially for the screening of ligands of varying net charge and/or polarity.[10–16] Some of the HIV-1 PR scoring functions have played an useful role in the development of inhibitors that have become lead compounds for antiviral drugs against AIDS.[8,11]

We have developed a very simple function, which we have named solvent accessibility free energy of binding predictor (SAFE_p). Initially we have applied this function to the study of peptidic HIV-1 PR inhibitors[12] and later we extended it to CU HIV-1 PR inhibitors by the addition of an explicit electrostatic interaction contribution.[15] The basic assumption underlying the SAFE_p approach is that the free energy of an inhibitor transfer from water to the binding pockets of an enzyme is analogous to the process of relocating the ligand from a medium of higher

Additional Supporting Information may be found in the online version of this article.

Correspondence to: F. Sussman; e-mail: fsussman@usc.es

Contract/grant sponsor: MEC, Xunta de Galicia

polarity to one of lower polarity, since many residues that line the active site and specificity pockets of this enzyme are hydrophobic in nature.[12,14,15] The resulting free energy function reproduced many of the observed characteristics of HIV-1 PR inhibitor binding. For instance, the resulting free energy function fitted closely the observed free energies of binding for a series of known peptidic inhibitors.[12] The additive nature of this approach enabled us to partition the free energy of binding into the contributions of single fragments. The resulting analysis allowed us to rank the importance of the enzyme's subsites for binding: although all the enzyme's pockets contribute to binding, the ones that bind the P2-P2′ span of the inhibitor are in general the most critical for high inhibitor potency.[12,14] Moreover, perusal of the energy contributions of single side chains has shown a broad specificity for some of the inhibitor fragments located in the central portion of the HIV-1 PR inhibitors. These observations are in agreement with experimental data, providing a validation for the physical relevancy of our method.[12,14,15] Recently, in a separate paper we have shown the usefulness of this approach by successfully applying it to the affinity ranking of ligands whose complexes with the HIV-1 PR are not know experimentally, one of most stringent tests of the quality of any binding free energy scoring function.[16]

The main objective of this work was to develop and apply a binding free energy predictor algorithm that could reproduce the binding affinity of a set of HIV-1 PR inhibitors with a wide variety of chemical motifs and with binding constants differing by several orders of magnitude. We have established the robustness of our free energy prediction results with respect to the size of the training set. The resulting function is able to reproduce the ranking of most inhibitors and in many cases the value of the inhibitor binding affinities for the HIV-1 PR. The analysis tools afforded by our protocol have allowed us to identify (in the case of the piperazine containing compounds) an induced fit phenomena mediated by the inhibitor and have demonstrated that larger fragments do not necessarily contribute the most to the binding free energy, an outcome brought about partly by the substantial role the desolvation penalty plays in the energetics of binding. Finally, with the help of our scoring function, we studied the effect of the HIV-1 PR Asp dyad protonation state on the binding affinities of a subset of peptidic inhibitors.

Methods

Free Energy Predictor's Analytical Form

The original SAFE_p analytical function had two terms, which were originally written as:

$$\frac{\Delta G}{RT} = K_1 \Sigma \Delta g_i + K_2 \Sigma \frac{\mathrm{abs}(Q_i)}{(R_{oi} + R_w)} \Delta g_i \qquad (1)$$

Where Δg_i is a function with a Gaussian analytical form that determines the atom change in solvent accessibility upon binding, Q_i is the partial charge for every atom, R_{oi} is the van der Waals radii, and R_w is radius of a test water molecule. K_1 and K_2 are the weights for the two terms of this equation. The first term in eq. (1) represents the inhibitor–protein hydrophobic interactions, whereas the second takes into account the polar desolvation penalty upon binding. The evaluation of the various components of this equation has been described previously.[12]

In recent publications, we have generated an improved scoring function by modifying the desolvation term and by introducing an explicit electrostatic interaction contribution. The new desolvation penalty replaces the linear form of charge dependence by a quadratic one, based on the idea by Novotny and Sharp that the electrostatic potential field depends on the square of the atomic charges rather than on the charge itself.[17] The explicit term for electrostatic interactions has its origin in a specific contact term for hydrogen bonded interactions in the CHARMM force field.[18] This additional term was shown to be essential for reproducing the binding free energies of CU based HIV-1 PR inhibitors.[15] The resulting expression can be written as:

$$\frac{\Delta G}{RT} = K_1 \Sigma \Delta g_i + K_2 \Sigma \frac{Q_i^2}{(R_{oi} + R_w)} \Delta g_i + K_3 \Delta G_{polar} + K_4 \qquad (2)$$

Where K_3 is the weight of the polar contribution and K_4 is a constant that could be set to 0.

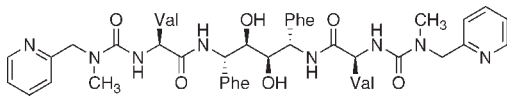
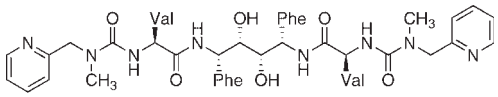
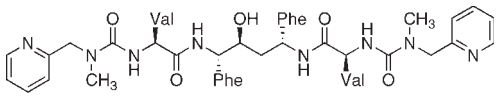
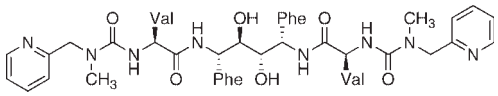
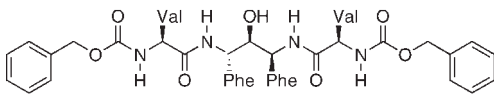
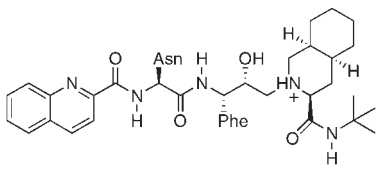
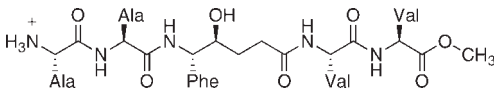
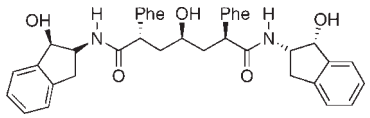
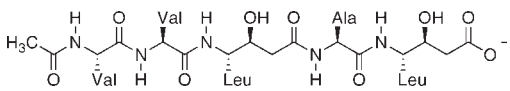
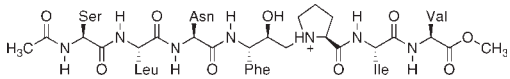
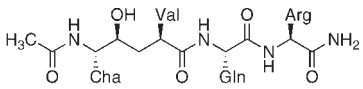
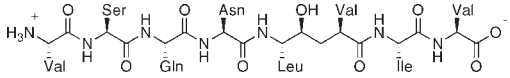
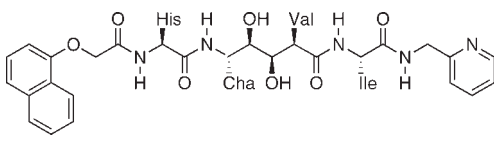
As was pointed out in our first SAFE_p work,[12] our scoring function takes into account implicitly some of the most important entropically related contributions to the free energy of binding. Among them are the ones related to the release of waters from the active site and binding pockets of the enzyme as well as from the ligand surface. This corresponds to the hydrophobic entropic contribution and can be approximated by a term that is proportional to the loss of surface area accessibility and hence, it is included in the first term of our scoring functions [see eqs. (1) and (2)]. A second entropic term relates to the loss of the conformational degrees of freedom associated with the constraints imposed by binding and hence it also should correlate with the change in solvent accessibility. Additional entropic changes include cratic, translational, and rotational entropy. These terms could be considered constant when external factors (e.g., temperature, pressure, etc.) do not change.[12]

Inhibitor Choice and General Procedures

The compounds studied in this article are listed in Tables 1–3. They include a wide variety of ligands, like first generation peptidic inhibitors (Table 1), CU inhibitors (Table 2), and other inhibitors containing piperazine, macrocycle, and sulphonamide moeities (Table 3).

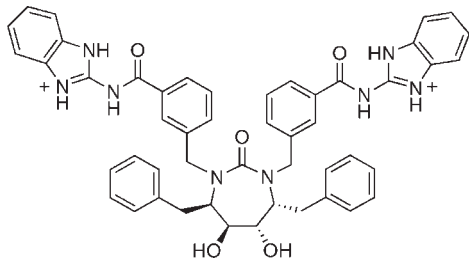
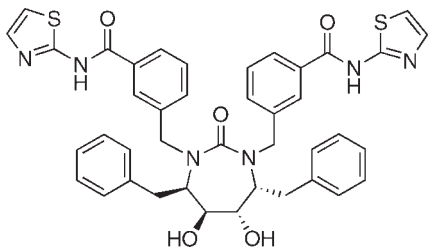
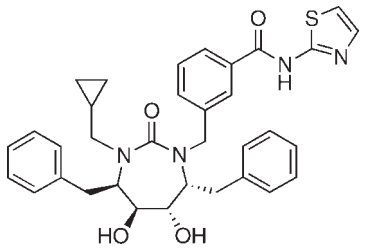
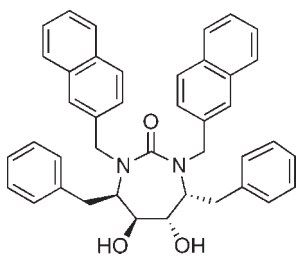
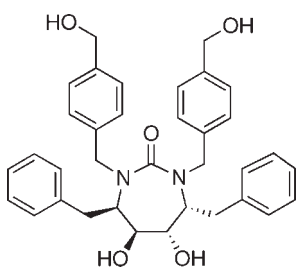
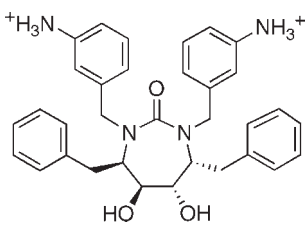
The ligand–protein complexes were downloaded from the PDB web site, all water molecules, but the one that bridges the Ile50/Ile50′ residues in the HIV-1 PR with some polar groups in the inhibitor through a hydrogen bond network were stripped away. Following the work by Fornabaio et al.,[38] we refer to this solvent molecule as water 301, although this water molecule has a different numbering in some crystallographic structures. Hydrogens were added using the Biopolymer module in the InsightII suite of programs[39] at pH = 7.0. The pKa values of

Table 1. Peptidic Inhibitors.

PDB code, inhibitor name	Structure	ΔG exp (kcal/mol)
1HVL, A-76889		-12.4[19a]
1HVK, A-76928		-14.0[19]
1HVJ, A-78791		-14.4[19]
1HVI, A-77003		-13.6[19]
9HVP, A-74704		-11.5[20]
1HXB, Ro31-8959 (Saquinavir)		-13.7[21]
1AAQ, SKB-Va		-10.7[22]
4PHV, L-700,417		-12.7[23]
5HVP, Acetylpepstatin		-10.6[24]
7HVP, JG-365		-13.3[25]
1GNO, U-89360E		-10.6[26]
8HVP, U-85548E		<-12.4[27]
1HIV, U-75875		<-12.4[28]

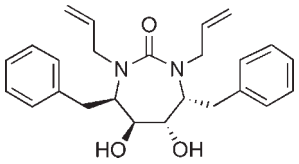
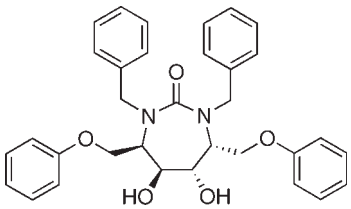
[a]The superscripted number is the reference for the corresponding binding constant.

Table 2. Cyclic Urea Inhibitors.

PDB code, inhibitor name	Structure	ΔG exp (kcal/mol)
1QBT, SD146		−14.8[29a]
1QBR, XV638		−14.5[29]
1QBU, Q8467		−14.1[30]
1HVR, XK263		−13.1[29]
1QBS, DMP323		−13.1[29]
1DMP, DMP450		−13.1[29]

(*continued*)

Table 2. (*Continued*)

PDB code, inhibitor name	Structure	ΔG exp (kcal/mol)
1HWR, XK216		−11.5[29]
1AJX, AH1		−10.9[31]

[a]The superscripted number is the reference for the corresponding binding constant.

some of the inhibitors polar groups (i.e., piperazines) were estimated by the pKa module of SciFinder.[40] The resulting ionization state used in these calculations is shown in Tables 1–3. The N and C terminus of each monomer were kept charged. In earlier studies, we kept the active site Asp dyad ionized. Following experimental and theoretical studies we kept charged only one of the Asp residues of the catalytic machinery.[41–43] The isostere's hydroxyl hydrogen(s) were oriented as to make the best hydrogen bonds with the Asp acid dyad. The hydrogen positions were optimized by a 1000 steps steepest descent energy minimization that kept the heavy atoms fixed, using the CHARMM force field.[18] The final step in our protocol is the SAFE_p calculation, which takes a few seconds per compound in an OCTANE SGI workstation.

Fitting the SAFE_p Equation to Experimental Data

To determine the effect of the training set on the quality of the resulting scoring function, we have used two inhibitor arrays of 9 (set "A") and 16 inhibitors (set "B"), respectively to calculate the eq. (2) constants. The first one contains six peptidic inhibitors (PDB entries 1HVL, 1HVK, 1AAQ, 4PHV, 5HVP, and 7HVP) and three CU inhibitors (PDB entries 1DMP, 1QBS, and 1QBR), whereas the second set holds seven additional inhibitors (PDB entries 2BPX, 1QBT, 1B6J, 2BPZ, 1B6N, 1HPV, and 1B6L). We also have performed the "leave one out" protocol by determining the constants and the fit of a series of training sets that differ in one compound, starting from set "B".

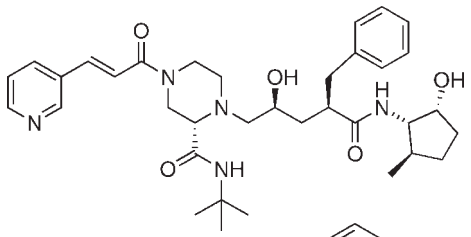
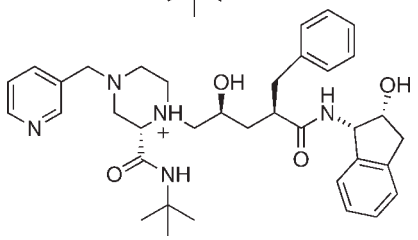
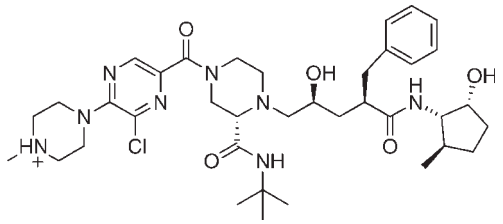
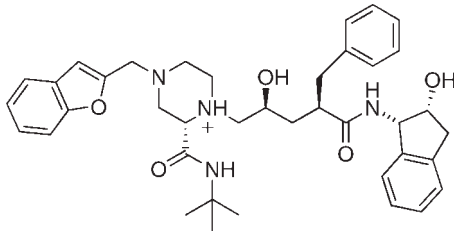
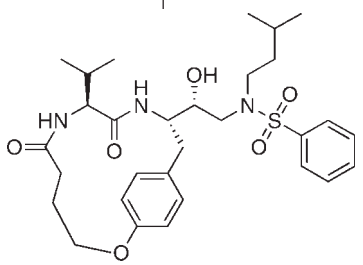
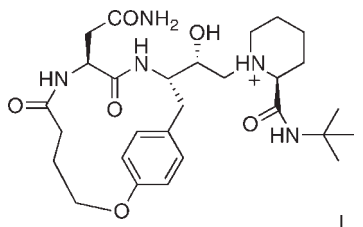
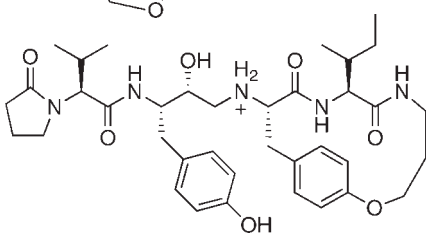
Asp Dyad Protonation State

The issue of the active site Asp dyad charged state was revisited with the help of the SAFE-p scoring function [eq. (2)]. We used as a testing ground an inhibitor subset comprised of 11 peptidic compounds (PDB entries 1HVL, 1HVK, 1HVJ, 1HVI, 9HVP, 1HXB, 1AAQ, 4PHV, 5HVP, 7HVP, and 1GNO). For this study, we selected the two protonation states that have been identified out in the literature as the most probable ones for the HIV-1 PR when bound to a given inhibitor.[41–43] These active site ionization states will be referred to as the "monoprotonated" state in which one of the dyad Asp residues is protonated whereas the other is left ionized, and the "diprotonated" state in which both Asp residues are neutral. The starting structure for both dyad protonation states was built by generating several hydrogen bond network alternatives in which one or two of the Asp residues are protonated on any of the oxygens of the Asp residues in such a way as to produce the best hydrogen bond network, whereas the inhibitor's hydroxyl hydrogen is always directed toward the charged Asp residue. For the case of the diol-based isosteres, we also generated hydrogen bond networks which optimize the number of hydrogen bond interactions. Both Asp dyad ionization states afford several possible initial alternatives, which are screened by an energy minimization that held the heavy atoms fixed (see earlier). To estimate the effect of the structural reorganization due to the change in Asp dyad ionization state, we proceeded to perform an energy minimization on the selected structures, for all atoms located at less than 6 Å around the inhibitor. The first stage of this protocol is a 1500 steps steepest descent minimization with a gradient tolerance of 0.1 followed by a second segment where we perform a conjugate gradient minimization with a 0.05 gradient tolerance. All optimizations were carried out with the CHARMM parameters provided in InsightII. The inhibitor binding affinities in the resulting structures were calculated by our SAFE_p protocol.

Results and Discussion

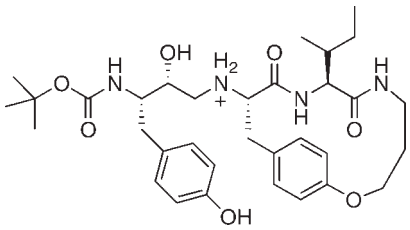
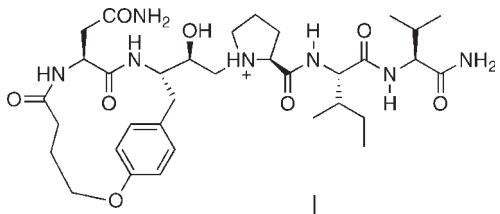
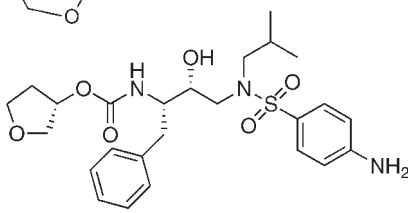
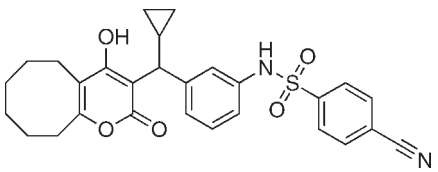
The inhibitors used in this work (collected in Tables 1–3) display very diverse chemical structure, the result of a continuous interest into the development of new generations of HIV-1 PR inhibitors. In many cases (see Table 3) the peptidic nature of

Table 3. Other Inhibitors.

PDB code, inhibitor name	Structure	ΔG exp (kcal/mol)
2BPW, L-738,317		−10.6[32a]
2BPX, L-735,524 (Indinavir)		−13.0[33]
2BPZ, L-739,622		−10.2[32]
1C70, L-756,423		−14.2[34]
1B6N		−11.6[35]
1B6L		−11.5[35]
1B6K		−12.1[35]

(*continued*)

Table 3. (*Continued*)

PDB code, inhibitor name	Structure	ΔG exp (kcal/mol)
1B6M		−11.6[35]
1B6J		−10.9[35]
1HPV, VX-478 (Amprenavir)		−12.7[36]
1HPO, U-103265		−12.6[37]

[a]The superscripted number is the reference for the corresponding binding constant.

these compounds has disappeared almost completely, a welcome trait for drug leads, since this feature favors bioavailability. In some cases a macrocycle has been introduced into the inhibitor's structure. Ligands with this design feature would have the advantage of being structurally constrained and hence will bind with a smaller loss of entropy.[35] Other inhibitors, like the ones containing a piperazine moiety[32–34] are already being used as part of AIDS therapy (i.e., Indinavir). Finally, compounds like the VX-478 (Amprenavir),[36] contain a sulphonamide group that provides a new interacting motif for the active site water molecule.

Scoring Function Prediction and the Size of the Training Set

As pointed out in our methods section we fitted our SAFE_p scoring function [eq. (2)] to two training sets that differ in the number of molecules. Table 4 lists the training sets *R2* values, the error and the standard deviation from the observed binding free energies for the 30 inhibitors studied, as well as the eq. (2) constants for both training sets. Perusal of this table indicates that the quality of the fitting as given by the *R2* parameter is very close for both training sets, irrespective of their size. Moreover, as seen from this table the error and standard deviations from the experimental affinities for the full set of inhibitors are very similar for both training sets. For instance, the average errors are 0.93 and 0.90 kcal/mol, values that are in the range of those obtained with costlier protocols like free energy perturbation or thermodynamic integration.[1] The latter algorithms are only able to calculate changes in the binding affinity for small variations in the inhibitors chemical structure. In this regard our protocol seems capable of predicting the ranking and in many cases the absolute free energies of a chemically diverse set of inhibitors. Furthermore, we have performed the "leave one out protocol" for the set "B". The results of this test are presented as a graph of the prediction of the residuals that leave one compound out against the prediction residual for the whole set. Perusal of this figure indicates that almost all compounds fall close to the 95% confidence region (see Supp. Info., Fig. S1). Moreover, as seen from Figure 1, the rankings predicted by both

Table 4. Results of the Least Square Fit with the Two Training Sets.

Statistical parameter	Training set "A"	Training set "B"
K_1	−0.00634	−0.00663
K_2	0.138	0.148
K_3	0.0006	0.0127
K_4	−4.239	−3.768
$R2$[a]	0.752	0.712
Average error[b]	0.929	0.903
Standard deviation[b]	1.16	1.04

[a]Value for the training set.
[b]Value for the set of 30 inhibitors.

training sets are the same, indicating the robustness of this approach, regardless of the size of the training set. This result is independent of the fact that the electrostatic term weight displays the largest variation between both training sets (see Table 4), and can be explained because the electrostatic contribution is much smaller (usually a fraction of a kcal/mol with any training set) than the other two, and hence, a large change does not have much influence in the actual ranking of the inhibitors.

Some of the main outliers are those inhibitors found in PDB entries 1AJX and 9HVP (see Figure 1). The former inhibitor[31] has different fragments in P1 and P1′ and it was obtained and characterized in a different laboratory than the bulk of the CU inhibitors, indicating once again that different experimental conditions play a role in the discrepancy between the observed and experimental ranking, a perennial problem in building a binding affinity database for prediction.[1] On the other hand, the latter compound was already an outlier in our original SAFE_p prediction performed for peptidic inhibitors.[12]

The presence of mutants in the active site of HIV-1 PR does not seem to pose a hurdle for the SAFE_p-based calculations. Many of the HIV-1 PR structurally known complexes have residue modifications due to viral resistant strains or man-made modifications. Previous SAFE_p-based studies have shown the usefulness of this algorithm for the study of the effect of actual resistance mutations on the binding affinity of some C2-centrosymmetric inhibitors.[14] In this study, we further demonstrate the versatility of our method in the prediction of binding affinity for multiple active site mutant strains of the HIV-1 PR, like in the case of inhibitor U103265 (PDB entry 1HPO) and for macrocycle-based inhibitor complexes (PDB entries 1B6N, 1B6L, 1B6K, 1B6M, and 1B6J). The HIV-1 PR strains bound to these ligands include the Q7K/L33I/L63I or Q7K/L33I variants as part of a strategy to limit the naturally occurring autoproteolysis observed in the HIV-1 PR.[35,37]

In regards to the macrocycle containing inhibitors, their lead optimization have generated smaller compounds due to the replacement of the C-terminal tripeptide (see PDB entry 1B6J in Table 3) by a sulphonamide moiety and by a change in the isostere's OH group configuration similar to the one in VX-478 (Amprenavir),[36] or by a piperidine modified by a terbutyl group at P2′. The resulting ligands (see PDB entries 1B6N and 1B6L in Table 3) have an improved affinity for the protease, a trend that our SAFE_p algorithm is able to predict.

Free Energy Inhibitor Fragment Contributions and Implications for Drug Design

Even in the case where there are differences between the observed and calculated binding free energies, the modified SAFE_p function predicts reasonably well the ranking trends of groups of congeneric compounds, as in the case of the piperazine containing compounds (see Figure 2). In principle, our calculations for these compounds indicate that enlarging the inhibitor by replacement of the cyclopentanol group by an indanol moiety, in the *S'* pockets of the enzyme, (PDB entry 2BPX) improves binding affinity, whereas the replacement of the pyridine fragment by a larger moiety (in the *S* pockets) not always increases binding affinity (e.g., PDB entry 2BPZ). To shed light into these results and to understand their physical origins, we have resorted to a very useful analytical tool afforded by our method that enables us to partition the total binding free energy

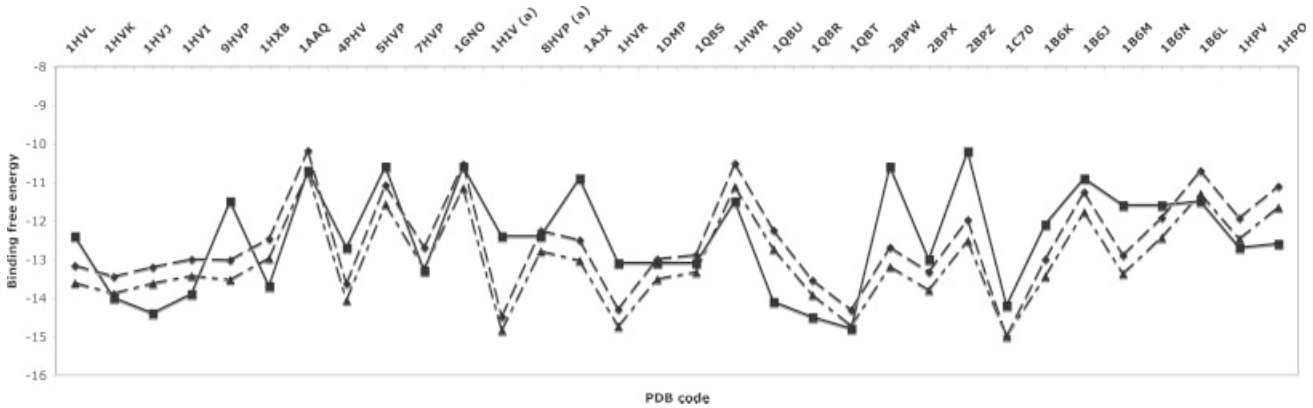


Figure 1. Comparison of the experimental (squares)and predicted binding free energies with training set "A" (triangles) and training set "B" (diamonds) for the whole set of inhibitors studied in this work. The inhibitors marked by an (a) do not have a single well defined affinity (inhibitors in PDB entries 1HIV and 8HVP). Their experimental binding free energies fall below −12.4 kcal/mol, in agreement with our predictions.

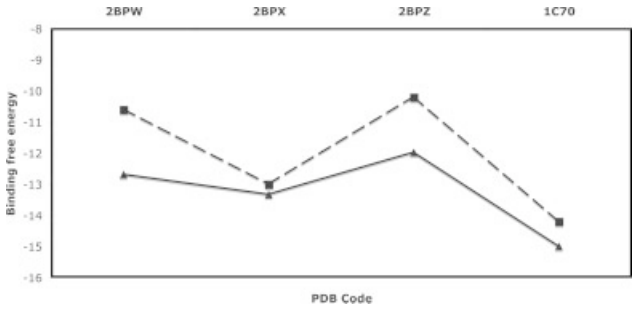


Figure 2. Comparison of the experimental (squares) and calculated (triangles) binding free energies for the piperazine containing inhibitors.

into the fragment contributions, which is made possible by the atom additive nature of eq. (2). The partitioning scheme used in our analysis is shown in Figure 3. The partition results for the piperazine containing compounds are shown in Table 5. We also list in this table the hydrophobic (ΔG_{HC}), desolvation penalty (ΔG_{DP}) and polar contact (ΔG_{HB}) contributions for every pocket, as defined in eq. (2). The direct contribution of water 301 was left out of the picture, as it has nearly the same value for all compounds analyzed. As seen from Table 5, the contributions that determine to the largest extent the ranking binding free energy originate in the P3 fragment. Interestingly enough, some of the largest P3 fragments are not the best binders. This outcome is in agreement with the X-ray studies, which found out that extending the fragments that fit into the *S* pockets leads to an increase in binding affinity only when the resulting groups fit into newly induced *S*1–*S*3 pockets comprised by residues from the 80's loop and the flaps.[34] These residues provide a favorable environment for the hydrophobic pyridine ring of indinavir (PDB entry 2BPX) or the benzofuran moiety (inhibitor L-756,423), a trend that is substantiated by our analysis. As seen from Table 5, the best hydrophobic contacts are produced by the P3 fragment in both 2BPX and 1C70. Moreover, larger P3 fragments are severely penalized by the desolvation term (ΔG_{DP}), when they do not fit into the novel pocket. The largest desolvation term comes from the *N*-methylpiperazinylpyrazine fragment in 2BPZ (1.91 kcal/mol), possibly owing to its charged nature as predicted by our pKa estimates.

As seen from Table 5, enlarging the inhibitor by replacement of the cyclopentanol for an indanol group in the S2′ pocket contributes to a smaller extent to the binding free energy. The indanol generates better hydrophobic contacts (ΔG_{HC}), but those are partially cancelled by a larger desolvation penalty (ΔG_{DP}). Interestingly enough, the results of our partitioning analysis also indicate that there are free energy changes in inhibitor fragments whose chemical structure has not changed from one inhibitor to the other. For instance, the P1′ fragment has a lower binding free energy in 2BPX (−2.41 kcal/mol) than the same fragment in 2BPW (−2.09 kcal/mol) or in 2BPZ (−1.96 kcal/mol). As seen from Table 5, this is the outcome of the better hydrophobic contacts (ΔG_{HC}) made by the indanol P1′ fragment, that may be the result of its interaction with fragments placed in adjacent pockets. Perusal of the inhibitor structure (bound to the enzyme) shows that the indanol moeity generates van der Waals contacts with the Phenyl group in P1′, indicating that the enlargement of the P2′ group improves the contribution of the P1′ to binding by placing it in a conformation with improved enzyme contacts. As a whole, these results demonstrate the plasticity of the HIV-1 PR-binding site and shed some light in ways to improve inhibitor binding affinity.

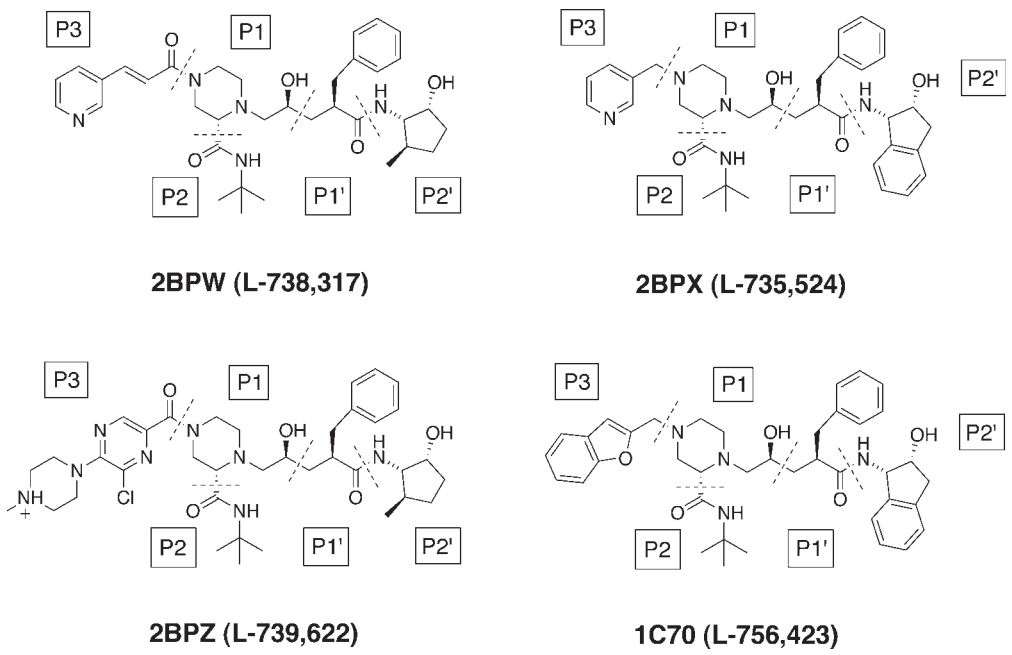


Figure 3. Partitioning scheme for fragment energy contribution in piperazine containing inhibitors.

Table 5. Pocket Binding Analysis for Piperazine Containing Inhibitors.[a]

PDB id	Fragment	ΔG_{HC}[b]	ΔG_{DP}[c]	ΔG_{HB}[d]	ΔG_{Total}
	P3	−1.655	1.231	−0.004	−0.427
	P2	−3.786	1.520	0.000	−2.266
2BPW	P1	−4.330	2.073	−0.013	−2.270
	P1′	−3.746	1.653	0.000	−2.094
	P2′	−4.239	1.974	−0.061	−2.326
	P3	−2.096	0.999	−0.034	−1.130
	P2	−3.727	1.588	−0.001	−2.140
2BPX	P1	−4.405	2.607	−0.023	−1.821
	P1′	−4.071	1.662	−0.003	−2.412
	P2′	−4.695	2.230	−0.057	−2.522
	P3	−1.778	1.910	−0.001	0.130
	P2	−3.954	1.548	−0.001	−2.406
2BPZ	P1	−4.239	2.031	−0.031	−2.239
	P1′	−3.593	1.634	−0.001	−1.960
	P2′	−4.195	2.036	−0.056	−2.215
	P3	−3.667	1.060	0.000	−2.607
	P2	−3.678	1.451	0.000	−2.227
1C70	P1	−4.078	2.348	−0.036	−1.766
	P1′	−4.154	1.756	−0.003	−2.401
	P2′	−4.858	2.237	−0.055	−2.676

[a]All energy data are in kcal/mol.
[b]Hydrophobic contribution, first term in eq. (2).
[c]Desolvation penalty, second term in eq. (2).
[d]Hydrogen bond energy, third term in eq. (2).

Effect of Asp Dyad Charge on Ligand Binding

We have used the SAFE_p approach to calculate the binding affinities of 11 peptidic inhibitors (see Table 1) for the HIV-1 PR in two Asp dyad protonation states that appear to be the most favored according to the literature.[41–43] Table 6 lists the predicted SAFE_p binding free energies for the two protonation states. As seen from this table, the binding free predictions differences from one protonation state to the other are rather small when only the hydrogen placement has been optimized. The binding free energies for seven out of eleven inhibitors are better predicted for the monoprotonated Asp dyad, but the differences are too small to allow us to draw any statistically relevant conclusions. As a whole the results indicate that the SAFE_p predicted binding energies calculated from the X-ray structures are not sensitive to the Asp dyad protonation state, and that any single of these two protonation states could be used across the board of HIV-1 PR inhibitors. On the other hand, a change in the protonation state may induce a reorganization of the residues in the active site, a process that is not taken into account in the calculations discussed above. We attempted to obtain a first order estimate of the active site structural changes brought about by a change in protonation states through an energy minimization protocol that included all atoms located at 6 Å around the inhibitor (see method section). The calculated binding free energies (see Table 6) depart more from the observed affinities than those calculated from the original X-ray structures. This outcome may indicate that the energy minimization protocol may produce spurious effects, probably due to the resulting compaction of the complex upon optimization. Perusal of these results indicates that the average error is larger for the diprotonated Asp dyad complexes than for the monoprotonated ones, indicating that the latter systems are more resilient to this protocol than the former ones.

An important discriminating factor for the Asp dyad ionization assignment in inhibitor-HIV-1 PR complexes is given by the thermodynamics equilibrium for moving a proton from solution to the bound HIV-1 Asp dyad. For this sake, it may be useful to have as a reference state the protonation state of the unbound enzyme. There is a consensus that the unbound protein has a singly protonated Asp dyad.[44] To estimate the feasibility of a diprotonated Asp dyad state it is necessary to determine if the energetic cost of desolvating a proton from water is compensated by the interactions this proton generate in the bound pro-

Table 6. Predicted Affinities for Both Asp Dyad Ionization States.[a]

	Monoprotonated		Diprotonated	
PDB id	ΔG_{Calc}	Error[b]	ΔG_{Calc}	Error[b]
1HVL	−13.17 (−13.56)	0.77 (1.16)	−13.12 (−13.11)	0.72 (0.71)
1HVK	−13.45 (−13.67)	0.55 (0.33)	−13.40 (−13.14)	0.60 (0.86)
1HVJ	−13.20 (−13.36)	1.20 (1.04)	−13.14 (−12.80)	1.26 (1.60)
1HVI	−13.00 (−13.13)	0.60 (0.77)	−12.90 (−11.80)	0.70 (2.10)
9HVP	−13.02 (−12.77)	1.52 (1.27)	−12.96 (−12.93)	1.46 (1.43)
1HXB	−12.47 (−12.16)	1.53 (1.84)	−12.37 (−11.67)	1.63 (2.33)
1AAQ	−10.19 (−10.03)	0.51 (0.66)	−10.12 (−10.29)	0.58 (0.40)
4PHV	−13.64 (−13.68)	0.94 (0.98)	−13.59 (−14.07)	0.89 (1.37)
5HVP	−11.08 (−11.97)	0.48 (1.37)	−10.99 (−11.64)	0.39 (1.04)
7HVP	−12.70 (−12.07)	0.61 (1.23)	−12.65 (−12.75)	0.65 (0.55)
1GNO	−10.54 (−11.08)	0.06 (0.48)	−10.45 (−10.61)	0.15 (0.01)
Average error		0.79 (1.01)		0.82 (1.13)

[a]Energy values are in kcal/mol. The results in parenthesis are from the structures that underwent energy minimization of all atoms at 6 Å around the inhibitor (see "Methods" section).
[b]Error with respect to the experimental binding free energies.

tein. Recent calculations[45] for the unbound HIV-1 PR indicate that this process is very costly energy wise and hence that in many cases the diprotonated state is unlikely, regardless of the small differences in binding free energy between the monoprotonated and diprotonated HIV-1 PR observed in our SAFE_p calculations.

Conclusion

Possibly the main hurdle in the "*in silico*" discovery of new potent protein inhibitors resides in the development of a fast free energy function predictor that could reproduce accurately the binding free energies or at least the binding ranking of compounds of diverse chemical structure and whose binding constants are spread over several orders of magnitude.[1] The aim of this work has been the application of our "in-house" generated SAFE_p protocol to a very heterogeneous set of HIV-1 PR inhibitors that bind to the target enzyme with affinity constants that differ by many orders of magnitude. The resulting scoring function is able to predict the binding free energy ranking for most inhibitors and in many cases the actual binding free energies with error bars that are usually obtained with protocols that are several orders of magnitude more time consuming.

In the case of the piperazine containing inhibitors, we have used the analysis tools afforded by SAFE_p to shed light into the actual fragment contributions to binding and the physical origins behind them. The results indicate that larger fragments do not contribute necessarily the most to the binding free energy, an outcome partly brought about by the substantial role the desolvation penalty plays in the energetics of binding. Our analysis has enabled us to identify an inhibitor induced fit phenomenon between neighboring inhibitor fragments.

Finally, we have revisited the effect of the Asp dyad protonation state on the binding affinities and found that the only factor that could discriminate between the two putative protonation states (in our approach) is the energetic balance of desolvating a proton and placing it in the Asp dyad.

We are in the process of extending our SAFE_p protocol to other members of the aspartic protease family, like the β-secretase, a potential objective for drug leads against Alzheimer's disease. Our main interest is to have a highly accurate inhibitor binding free energy scoring function that could be applied to different enzymes.

Acknowledgments

The authors thank J. L. Martin for sending us the coordinate files of some macrocyclic inhibitors. The authors also thank Dr. Humberto Gonzalez for his advice.

References

1. Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. J Med Chem 2006, 49, 5851.
2. Erickson, J. W.; Burt, S. K. Annu Rev Pharmacol Toxicol 1996, 36, 545.
3. Wlodawer, A.; Vondrasek, J. Annu Rev Biophys Biomol Struct 1998, 27, 249.
4. Leung, D.; Abbenante, G.; Fairlie, D. P. J Med Chem 2000, 43, 305.
5. Debnath, A. K. J Med Chem 1999, 42, 249.
6. Horton, N.; Lewis, M. Protein Sci 1992, 1, 169.
7. Weber, I. T.; Harrison, R. W. Protein Eng 1996, 9, 679.
8. Weber, I. T.; Harrison, R. W. Protein Eng 1999, 12, 469.
9. Lee, C. Y.; Yang, P. K.; Tzou, W. S.; Hwang, M. J. Protein Eng 1998, 11, 429.
10. Huang, D.; Caflisch, A. J Med Chem 2004, 47, 5791.
11. Ohtaka, H.; Muzammil, S.; Schon, A.; Velazquez-Campoy, A.; Vega, S.; Freire, E. Int J Biochem Cell Biol 2004, 36, 1787.
12. Nauchitel, V.; Villaverde, M. C.; Sussman, F. Protein Sci 1995, 4, 1356.
13. Sussman, F.; Villaverde, M. C.; Davis, A. Protein Sci 1997, 6, 1024.
14. Sussman, F.; Martinez, L. A.; Villaverde, M. C. Adv Exp Med Biol 1998, 436, 91.
15. Sussman, F.; Villaverde, M. C.; Martinez, L. Protein Eng 2002, 15, 707.
16. Vilar, S.; Villaverde, M. C.; Sussman, F. J Comput Chem 2007, 28, 2216.
17. Novotny, J.; Sharp, K. Prog Biophys Mol Biol 1992, 58, 203.
18. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
19. Hosur, M. V.; Bhat, T. N.; Kempf, D. J.; Baldwin, E. T.; Liu, B.; Gulnik, S.; Wideburg, N. E.; Norbeck, D. W.; Appelt, K.; Erickson, J. W. J Am Chem Soc 1994, 116, 847.
20. Erickson, J.; Neidhart, D. J.; VanDrie, 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. Science 1990, 249, 527.
21. Krohn, A.; Redshaw, S.; Ritchie, J. C.; Graves, B. J.; Hatada, M. H. J Med Chem 1991, 34, 3340.
22. Dreyer, G. B.; Metcalf, B. W.; Tomaszek, T. A., Jr.; Carr, T. J.; Chandler, A. C., 3rd; Hyland, L.; Fakhoury, S. A.; Magaard, V. W.; Moore, M. L.; Strickler, J. E.; Debouck, C.; Meek, T. D. Proc Natl Acad Sci USA 1989, 86, 9752.
23. Wlodawer, A.; Erickson, J. W. Annu Rev Biochem 1993, 62, 543.
24. Fitzgerald, P. M.; McKeever, B. M.; VanMiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C.; Leu, C. T.; Herber, W. K.; Dixon, R. A.; Darke, P. L. J Biol Chem 1990, 265, 14209.
25. Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S. B.; Wlodawer, A. Proc Natl Acad Sci USA 1990, 87, 8805.
26. Hong, L.; Treharne, A.; Hartsuck, J. A.; Foundling, S.; Tang, J. Biochemistry 1996, 35, 10627.
27. Jaskolski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B.; Wlodawer, A. Biochemistry 1991, 30, 1600.
28. Thanki, N.; Rao, J. K.; Foundling, S. I.; Howe, W. J.; Moon, J. B.; Hui, J. O.; Tomasselli, A. G.; Heinrikson, R. L.; Thaisrivongs, S.; Wlodawer, A. Protein Sci 1992, 1, 1061.
29. Ala, P. J.; DeLoskey, R. J.; Huston, E. E.; Jadhav, P. K.; Lam, P. Y.; Eyermann, C. J.; Hodge, C. N.; Schadt, M. C.; Lewandowski, F. A.; Weber, P. C.; McCabe, D. D.; Duke, J. L.; Chang, C. H. J Biol Chem 1998, 273, 12325.
30. Jadhav, P. K.; Ala, P.; Woerner, F. J.; Chang, C. H.; Garber, S. S.; Anton, E. D.; Bacheler, L. T. J Med Chem 1997, 40, 181.
31. Backbro, K.; Lowgren, S.; Osterlund, K.; Atepo, J.; Unge, T.; Hulten, J.; Bonham, N. M.; Schaal, W.; Karlen, A.; Hallberg, A. J Med Chem 1997, 40, 898.
32. Munshi, S.; Chen, Z.; Li, Y.; Olsen, D. B.; Fraley, M. E.; Hungate, R. W.; Kuo, L. C. Acta Crystallogr D Biol Crystallogr 1998, 54, 1053.

33. Chen, Z.; Li, Y.; Chen, E.; Hall, D. L.; Darke, P. L.; Culberson, C.; Shafer, J. A.; Kuo, L. C. J Biol Chem 1994, 269, 26344.
34. Munshi, S.; Chen, Z.; Yan, Y.; Li, Y.; Olsen, D. B.; Schock, H. B.; Galvin, B. B.; Dorsey, B.; Kuo, L. C. Acta Crystallogr D Biol Crystallogr 2000, 56, 381.
35. Martin, J. L.; Begun, J.; Schindeler, A.; Wickramasinghe, W. A.; Alewood, D.; Alewood, P. F.; Bergman, D. A.; Brinkworth, R. I.; Abbenante, G.; March, D. R.; Reid, R. C.; Fairlie, D. P. Biochemistry 1999, 38, 7978.
36. Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. J Am Chem Soc 1995, 117, 1181.
37. Skulnick, H. I.; Johnson, P. D.; Aristoff, P. A.; Morris, J. K.; Lovasz, K. D.; Howe, W. J.; Watenpaugh, K. D.; Janakiraman, M. N.; Anderson, D. J.; Reischer, R. J.; Schwartz, T. M.; Banitt, L. S.; Tomich, P. K.; Lynn, J. C.; Horng, M. M.; Chong, K. T.; Hinshaw, R. R.; Dolak, L. A.; Seest, E. P.; Schwende, F. J.; Rush, B. D.; Howard, G. M.; Toth, L. N.; Wilkinson, K. R.; Kakuk, T. J.; Johnson, C. W.; Cole, S. L.; Zaya, R. M.; Zipp, G. L.; Possert, P. L.; Dalga, R. J.; Zhong, W.; Williams, M. G.; Romines, K. R. J Med Chem 1997, 40, 1149.
38. Fornabaio, M.; Spyrakis, F.; Mozzarelli, A.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E. J Med Chem 2004, 47, 4507.
39. InsightII. Accelrys Inc.: San Diego, CA, 2000.
40. ACD/pKa DB, for SciFinder; Advanced Chemistry Development Inc: Ontario, Canada.
41. Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. J Am Chem Soc 2001, 123, 8730.
42. Beveridge, A. J.; Heywood, G. C. Biochemistry 1993, 32, 3325.
43. Trylska, J.; Antosiewicz, J.; Geller, M.; Hodge, C. N.; Klabe, R. M.; Head, M. S.; Gilson, M. K. Protein Sci 1999, 8, 180.
44. Piana, S.; Carloni, P. Proteins 2000, 39, 26.
45. Wittayanarakul, K.; Hannongbua, S.; Feig, M. J Comput Chem 2008, 29, 673.