Solvent accessibility as a predictive tool for the free energy of inhibitor binding to the HIV-1 protease

VLADIMIR NAUCHITEL,^{1,2} M. CARMEN VILLAVERDE,^{1,2,3} AND FREDY SUSSMAN^{1,2}

¹ Research and Education Computing, Oklahoma University Health Sciences Center, Oklahoma City, Oklahoma 73104

² Protein Studies Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104

(RECEIVED February 23, 1995; ACCEPTED April 11, 1995)

Abstract

We have developed a simple approach for the evaluation of the free energies of inhibitor binding to the protease of the human immunodeficiency virus (HIV-1 PR). Our algorithm is based on the observation that most groups that line the binding pockets of this enzyme are hydrophobic in nature. Based on this fact, we have likened the binding of an inhibitor to this enzyme to its transfer from water to a medium of lower polarity. The resulting expression produced values for the free energy of binding of inhibitors to the HIV-1 PR that are in good agreement with experimental values. The additive nature of this approach has enabled us to partition the free energy of binding into the contributions of single fragments. The resulting analysis clearly indicates the existence of a ranking in the participation of the enzyme's subsites in 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. Moreover, our method has allowed us to determine the nature of the functional groups that fit into given enzyme binding pockets. Perusal of the energy contributions of single side chains has shown that a large number of hydrophobic and aromatic groups located in the central portion of the HIV-1 PR inhibitors present optimal binding. All of these observations are in agreement with experimental evidence, providing a validation for the physical relevancy of our model.

Keywords: binding analysis; HIV-1 protease; inhibitor binding; solvent accessibility

The HIV virus encodes a protease (HIV-1 PR) whose activity is vital for the processing of mature virions (Kramer et al., 1986). This enzyme together with the reverse transcriptase are among the most promising targets for rational drug design against AIDS (Mitsuya et al., 1990; Roberts et al., 1990; Huff, 1991; Debouck, 1992). For this reason HIV-1 PR has become one of the most studied proteins to date. Efforts in the search for high-affinity HIV-1 PR inhibitors were based initially on peptide substrate analogs in which the scissile peptide bond was replaced by a non-hydrolyzable isostere with tetrahedral geometry (Dreyer et al., 1989, 1992; Wlodawer & Erickson, 1993). Most inhibitors have at least one hydroxy group at the non-scissile junction, which is positioned to make H-bonds with at least one of the two active site aspartic acids of the enzyme.

A wealth of information has been gained from the X-ray crystallographic studies performed on a large number of HIV-1 PR–

inhibitor complexes (Davies, 1990; Fitzgerald & Springer, 1991; Wlodawer & Erickson, 1993). Analysis of the structural data has shown that all known peptidomimetic inhibitors bind to the HIV-1 protease in an almost extended conformation. Most of the binding pockets of the HIV-1 PR are lined with hydrophobic side chains, with the exception of the active site aspartic dyad. This has led us to test the idea that inhibitor binding to the HIV-1 PR can be likened to the transfer of the ligand from a polar environment (water) to a less polar one. To investigate this hypothesis we have generated a simple approach that allowed close fitting of the observed free energies of binding for a series of known inhibitors. This approach has also allowed us to determine the free energy of binding for an inhibitor, whose three-dimensional structure of its complex with the HIV-1 PR is not known. The result is in good agreement with experimental evidence, suggesting a way of scanning the affinity of a large number of related inhibitors prior to their synthesis and characterization.

Partition of the free energies of binding into the contributions of the inhibitor P fragments interacting with the HIV-1 PR binding pockets has allowed us to gauge the binding capabilities of single binding pockets. The outcome of this partitioning indi-

Reprint requests to: Fredy Sussman, Protein Studies Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104; e-mail: fredy@rev.uokhsc.edu.

³ Permanent address: Departamento de Química Orgánica, Facultad de Química, 15706-Santiago de Compostela, Spain.

cates a clear pattern of molecular recognition between the HIV-1 PR and its ligands, a piece of information that should be useful in the design of novel inhibitors.

Methodological approach

The approach we present in this paper is ideologically close to the one developed earlier (Eisenberg & McLachlan, 1986). The basic assumption is that the free energy of an inhibitor transfer from water to the binding pockets of HIV-1 PR is analogous to the process of transfer from a medium of higher polarity to one of lower polarity. In this approach the free energy of binding is assumed to be an additive function of all atoms involved. Each atom in the HIV-1 PR-inhibitor complex contributes to the free energy an amount proportional to its change in solvent accessibility that occurs upon binding. For any given system, the free energy is expressed as:

$$\Delta G/RT = k_1 \sum \Delta g_i + k_2 \sum \frac{abs(Q_i)}{(R_{oi} + R_w)} \Delta g_i, \quad (1)$$

where the first term characterizes the transfer free energy of each atom regardless of its partial charge, and the second term provides a correction to the free energy that depends on the atom's charge to radius ratio; Δg_i characterizes the change of accessibility per atom, Q_i is the partial charge of atom i , R_{oi} is its van der Waals radius, and R_w is the radius of the water molecule, and the sums are over all atoms of the complex. For a given evaluation of ΔG , the temperature was set at 300 K. The empirical parameters k_1 and k_2 control the transfer free energy per unit of accessibility. Our choice of the form for the second term of the equation is based on the assumption that the interaction of an atom with the surrounding water molecules is proportional to the electrostatic field a given atom produces around itself. To calculate the change in solvent accessibility of the atoms (i.e., Δg_i), we use the Gaussian neighborhood approach (Nauchitel & Somorjai, 1993). This method has been shown to be about two orders of magnitude faster than those that are based on the calculation of surface areas (Lee & Richards, 1971). To estimate the degree of exposure of an atom to water we use a discriminating function that filters surrounding atoms according to their ability to screen from water the atom of interest. As the discriminating function we use a two-parameter Gaussian, $\exp[-(r - r_o)^2/2\sigma^2]$, where σ controls its width, and $r_o = R_{oi} + R_w$ positions its maximum. The Gaussian neighborhood of atom i is defined by

$$g_i = \sum_{j \neq i} \exp[-(R_{ij} - (R_{oi} + R_w))^2/2\sigma^2], \quad (2)$$

where R_{ij} is the separation between atoms i and j . As the burial of any given atom increases so does its Gaussian neighborhood.

Our empirical expression (Equation 1) takes into account implicitly some of the most important entropically related contributions to the free energy of binding. Among them are the free energies related to the release of water molecules from the active site and binding pockets of the enzyme, as well as from the ligand surface, that occur on binding. This entropic free energy contribution can be approximated by a term that is proportional to the loss of surface accessibility, and hence, it is included in

the first term of Equation 1. A second entropic term relates to the loss of the conformational degrees of freedom associated with the constraints imposed by binding and hence should also correlate with the change of solvent accessibility. Similar treatments for the desolvation entropy and the loss of conformational entropy were employed previously (Horton & Lewis, 1992; Krystek et al., 1993). Additional entropic changes include cratic, translational, and rotational entropy. These terms could be considered constant when external factors (e.g., temperature, pressure) do not change and should be taken into account in the framework of more explicit approaches.

Calculation of the free energy of ligand-enzyme binding

To check the validity of our approach we fitted Equation 1 to the free energies of some known entries for the inhibitor-HIV-1 PR complexes available in the Brookhaven Protein Data Bank (Bernstein et al., 1977) and some that were made available to us by A. Wlodawer and J.W. Erickson. The list of the inhibitors is presented in Figure 1. The differences in solvent accessibility of the enzyme and the inhibitors, complexed and separated, were calculated after adding hydrogens to the molecules. Because the conformations of the inhibitors in solution are unknown we assumed them to have the same structure as when bound to HIV-1 PR. Even though the structure of the apoenzyme is available, we have used, for the sake of consistency, the structures this molecule has in the complexed state. This is a rigid body approximation that has been widely used in calculation of the free energies of binding (Horton & Lewis, 1992; Krystek et al., 1993). The results of fitting the free energies of binding with Equation 1 are presented in Table 1.

Table 1. Free energy of inhibitor binding to HIV-1 protease

Inhibitor	PDB file name ^a	$K_{i,nM}$ ^b	ΔG , exp (kcal/mol)	W^c	ΔG , calc (kcal/mol)
1 SKB-Va	1aaq	18	-10.7	0	-9.7
2 L-700,417	4phv	0.65	-12.7	0	-12.3
3 Ac. pepstatin	5hvp	20.0	-10.6	1	-10.6
4 JG-365	7hvp	0.66	-12.7	1	-12.9
		0.24	-13.3		
5 A-74704	9hvp	4.5	-11.5	0	-13.3
6 A-76889		1	-12.4	1	-13.5
7 A-76928		0.077	-14.0	1	-14.2
8 A-77003		0.15	-13.6	1	-13.5
		0.084	-13.9		
9 A-78791		0.035	-14.5	1	-13.7
10 Ro-31-8558		0.3	-13.1	1	-13.8
11 Ro-31-8959		0.073	-14.0	1	-12.8
12 U-75875	1hiv	<1	<-12.4	0	-14.0
13 U-85548E	8hvp	<1	<-12.4	0	-11.8
14 U-89360E		20	-10.6	0	-10.1

^a Structures of the inhibitors that do not have a PDB entry name in the table were supplied to us by A. Wlodawer and J.W. Erickson, with the exception of U-89360E.

^b K_i data are from Dreyer et al. (1989), Richards et al. (1989), Erickson et al. (1990), Tomasselli et al. (1990), Wlodawer and Erickson (1993), Ho et al. (1994), Hosur et al. (1994), and Lin et al. (1995). In some cases there is more than one published K_i value for a single inhibitor.

^c Only data with $W = 1$ were used for the fitting.

Calculation of the free energy of binding of U-89360E

To further test the predictive power of Equation 1, we have calculated the binding energy for the inhibitor U-89360E, whose complexed structure with HIV-1 PR has not been studied crystallographically. This inhibitor has the chemical structure shown in Figure 2.

The central isostere group conformation was model built based on the structure of the U-85548E inhibitor (Jaskólski et al., 1991), whereas the starting structures of Gln and Arg side chains, as well as the N-terminal acetyl group of this inhibitor, were built from analogously placed side chains in the reduced amide inhibitor MVT-101 (Miller et al., 1989). The initial structure of the Cha side chain was taken from U-75875 (Thanki et al., 1992). The structure of the enzyme from the complex with the U-85548E inhibitor was used as a starting conformation for the modeled enzyme. The whole system was surrounded by an 8-Å solvent layer containing 970 water molecules and energy minimized using the Charmm-22 set of parameters (Brooks et al., 1983). The convergence criterion was 0.01 in the gradient tolerance between successive iterations.

Energy partition

Because the total free energy expression presented in Equation 1 is an additive function of single atoms, it is easy to partition it into contributions from particular fragments of interest. All the inhibitors we have analyzed are peptidic and can (in principle) be divided into their constituent fragments following the definition for substrate molecules (Schechter & Berger, 1967). The only obstacle to this partition is the inhibitor's isostere, which is nonpeptidic and differs in length from one inhibitor to another. In what follows we propose special rules for its partitioning. Those inhibitors with the noncissile fragments that contain six main-chain atoms are partitioned into groups of three atoms, following a similar rule used for the peptidic fragments of the inhibitors. Among the inhibitors of this type are: SKB-Va, U-75875, U-85548E, and Ro-31-8558 (see Fig. 1). We divide the central isostere main-chain atoms and their side-chain atoms between two adjacent fragments (P_1 and P'_1). The three isostere main-chain atoms that are closer to the N-terminus are ascribed to P_1 , and the three that are closer to the C-terminus are ascribed to P'_1 . Some inhibitors have isostere groups with five or seven atoms in their main chains. We divided the isosteres that contain five atoms (i.e., acylpepstatin, A-74704) into two smaller subsets (P_1 and P'_1) each containing two main-chain atoms and their respective side chains plus an equal share of the central -CH(OH)- group. We have followed a similar rule with the isosteres containing seven main-chain atoms. For instance, in the case of inhibitor JG-365 we have divided the central motif into two fragments (P_1 and P'_1), each containing three atoms

Fig. 2. Chemical structure of U-89360E, an inhibitor of HIV-1 protease.

plus an equal share of the central methylene group. Symmetrical peptidic inhibitors have the same type of termini at both ends because they have reversed the N- to C-terminus direction after the isostere group. Here also we draw boundaries between the P fragments through the peptide bonds as we did in the case of the asymmetric inhibitors. (See Fig. 1 for chemical structures of the inhibitors.) The energy partitioning shown in our results and discussion section is based on the structure apportioning described here.

Once the inhibitor's binding fragments have been defined, the residues that form the HIV-1 PR binding pockets (S_i) can be determined based on a free energy criterion. We have identified the residues that are involved in a given pocket S_i as those that contribute to the free energy of binding of a P_i fragment above a threshold set at $|\Delta G/RT| = 0.1$ (0.06 kcal/mol at 300 K).

Results and discussion

The fitting of the free energies of binding to the observed values for the set of inhibitors presented in Figure 1 yields $k_1 = -0.007778$ and $k_2 = 0.05891$ for the empiric parameters of Equation 1. Hence, the first term of this equation always rewards burial of atoms, whereas the second provides a correction that penalizes the burial of charged atoms. Based on the magnitude and sign of k_1 and k_2 , the final outcome only penalizes the transfer of atoms that have large partial charges. The values of the critical charges for which reward and penalty cancel out each other can be readily determined from Equation 1, by equating its two terms. For a given value of k_1 and k_2 we have:

$$\frac{abs(Q_i)}{(R_{oi} + R_w)} = \frac{k_1}{k_2} \quad (3)$$

This relationship establishes exposure/burial preferences for each type of atom based on its charge and van der Waals radius. The absolute values of the critical charges are: 0.36 for H, 0.40 for N and O, and 0.42 for C. These figures are a model-sensitive characteristic, but they represent a real phenomenon of charge-dependent partitioning between media of different polarity.

The list of observed and calculated free energies of binding is presented in Table 1. For this set of inhibitors, the difference in the free energy of binding between the ligand with the lowest and that with the highest affinity for HIV-1 PR exceeds 4 kcal/mol (i.e., three orders of magnitude in the inhibition constants), whereas the average difference between the observed and calculated free energies is only 0.9 kcal/mol. We calculated the average deviation based only on those compounds that have a single K_i value in Table 1. Some of the compounds listed in Table 1 have more than one published K_i value, because different laboratories perform their experiments under different conditions. Some laboratories provide only the upper boundary of possible K_i values (e.g., $K_i < 1$ nM for inhibitors U-75875 [Thanki et al., 1992] and U-85548E [Tomasselli et al., 1990]), which reduces their utility for fitting. In all cases, our method produces reasonable estimates for the free energy of binding.

Our approach has enabled the prediction of the free energy of binding of the U-89360E inhibitor, whose 3D structure complexed with the HIV-1 PR is unknown. Following the protocol described in the methodology section we have model built this complex.

The calculated value for the inhibition constant is 50.6 nM, whereas the observed one is 20 nM (Lin et al., 1995). This corresponds to a difference of 0.54 kcal/mol between the observed and calculated free energies of binding. This result shows that the method described here, when complemented by a model-built structure of a complex, is capable of yielding reasonable estimates of the binding energies, and therefore could be of help in the rational design of HIV-1 PR inhibitors. A few members of a family of inhibitors would have to be synthesized and characterized kinetically and structurally. Then, a large number of variations with different hydrophobic and hydrophilic side chain groups could be scanned, using our approach for the prediction of their inhibition constants prior to their synthesis and binding characterization.

The X-ray studies of HIV-1 PR-inhibitor complexes have revealed the existence of an extensive backbone-to-backbone hydrogen bond network between the ligands and the enzyme (Wlodawer & Erickson, 1993). In the framework of our model we have analyzed the role of hydrogen bonding in inhibitor binding to this enzyme. We have found that burial of the amide H-bond donors is rewarded, whereas that of the carbonyl H-bond acceptors is penalized. Simultaneous internment of these two groups in order to form a hydrogen bond has little effect on the energetics of binding. We have found that an H-bond contributes on average 0.1 kcal/mol to the destabilization of the inhibitor-HIV-1 PR complex. However, if the inhibitor would not form hydrogen bonds with some of the protein's donors or acceptors, the inhibitor's binding constant could be even lower because some of these groups would have to be buried against hydrophobic segments. This result is consistent with the view held by some authors that electrostatic interactions between polar groups in proteins do not completely compensate for the desolvation of these groups (Novotny & Sharp, 1992). They infer that buried polar groups are located inside an enzyme to put in place specific interactions that are essential for a unique complex structure. The outcome from our approach supports this point of view for the case of the HIV-1 PR. The inhibitor-enzyme hydrogen bond network could help indirectly to increase the free energy of binding by directing a single inhibitor fragment P_i into its optimal binding pocket S_i . We intend to critically re-examine the energetic contribution of the inhibitor-enzyme H-bond network with the help of a model that has built in explicit electrostatic interactions.

Analysis of inhibitor binding

The residue composition of the enzyme pockets that bind to a single inhibitor's fragment can be identified using the energy criterion described in the methodology section. The results are shown in Table 2. As seen from this table, single P_i fragments of different inhibitors interact with practically the same residues of the enzyme. These contact residues are located in four regions of the enzyme and are almost invariant for all inhibitors studied here. The first region only includes residue R8 (of either monomer), which contacts the P_3 and P'_3 and in some cases the P_1 and P'_1 fragments. The second region of interaction includes most of the residues in the segment between L23 to V32 of both monomers. In this contact area the central fragments P_1 and P'_1 interact with some of the residues that stretch from L23 to A28, and P_2 and P'_2 interact with some of the residues in the segment 28–32, and fragments P_3 and P'_3 contact residues 28–30. The

third segment comprises residues stretching from K45 to F53. In this contact area, fragments P_1 and P'_1 interact with residues 49–50 of both monomers, and P_2 and P'_2 contact some of the residues of the segment from I47 to I50, and P_3 and P'_3 contact mostly residues from segment 47–49 and F53 of the enzyme. The fourth and final contact region includes residues 76–84. In this segment, residues 80–82 and 84 of both monomers contact fragments P_1 and P'_1 , whereas residue 76 (in some cases) and residue 84 (in all cases) contact fragments P_2 and P'_2 . Finally residues 81–82 contact fragments P_3 and P'_3 . These contact segments play a very important role in the function of the protein. Multiple site-directed mutagenesis experiments of all the HIV-1 PR residues have identified those whose change had a detrimental influence on the proteolytic activity of the enzyme (Loeb et al., 1989). They are located in three segments that coincide to a large extent with the last three contact regions between inhibitor and the HIV-1 PR enzyme shown in Table 2.

The contributions of the P fragments to the free energy of binding of symmetric and asymmetric inhibitors are shown in Figure 3A and B, respectively. This figure also depicts the free energy of binding of a given fragment to each of the enzyme's monomers. As seen from Figure 3, the calculated free energy of binding to each monomer of this homodimer enzyme alternates for consecutive P fragments. The free energies of binding have a "saw-tooth" pattern with A-monomer maxima complemented by B-monomer minima, and vice versa. Both monomers of the HIV-1 PR contribute to the interaction with each fragment P_i , but one of the monomers (either A or B) makes a larger stabilizing contribution than the other. This result seems to correlate with the residue composition, at least for the binding pockets S_2 , S_1 , S'_1 , and S'_2 , as shown in Table 2. The enzyme pocket for a given inhibitor fragment is built of a mixture of residues from both monomers although the above-mentioned pockets have a larger contribution from one of the monomers (either A or B). The enzyme's binding pockets that can be defined more readily with one monomer are S_2 and S'_2 . In this case the only residue from the second monomer that participates in binding is I50, a residue located on the flaps of the enzyme. This translates into the largest calculated difference in the free energy of binding (for a large number of HIV-1 PR-inhibitor complexes) at P_2 or P'_2 (see Fig. 3), indicating that these fragments interact noticeably stronger with one of the monomers than with the other. For the most part the two monomers of the HIV-1 PR do not contribute the same residues to the constitution of a binding pocket. The only exception are binding pockets S_1 and S'_1 , which contain in some cases residues A28, I50, and I84 from both monomers.

Analysis of free energy profiles (see Fig. 3) has allowed us to study the binding capabilities of the enzyme pockets on one hand, and the roles of the inhibitor's fragments in binding on the other. The free energy profiles presented in Figure 3 show a clear pattern of binding capacity for the enzyme pockets. As seen from this figure the largest binding energies are concentrated almost equally at binding pockets S_1 and S'_1 for *all* symmetric inhibitors, and at S_1 or S'_1 for *most* asymmetric inhibitors. This conclusion is consistent with the findings of other investigators. A similar pattern has been obtained from QSAR type studies using the hypothetical active-site lattice methodology (HASL). This method generated a pharmacophore for the HIV-1 PR inhibitors, containing 11 distinctive points or areas (Doweyko, 1994). The segment that contributed the most to the free energy of binding was found to be the part of the peptidic

Table 2. Residue composition of the enzyme pockets^a

		R 08	L 23	D 25	G 27	A 28	D 29	D 30	V 32	K 45	I 47	G 48	G 49	I 50	F 53	L 76	T 80	P 81	V 82	I 84
P ₁	1	B	B	B	A	A							A	A				B	B	B
	2	B	B	B	A	AB							A	A				B	B	B
	3		B	B	A	A							A	AB				B	B	B
	4		B	B	A	AB							A	A				B	B	B
	5		B	B	A	AB							A	A				B	B	B
	6		B	B	A	AB							A	A				B	B	B
	7		B	B	A	A							A	AB				B	B	BA
	8		B	B	A	A							A	A				B	B	B
	9		B	B	A	AB							A	A				B	B	B
	10	B	B	B	A	AB							A	A			B	B	B	B
	11		B	B	A	AB							A	A			B	B	B	B
	12		B	B	A	AB							A	A			B	B	B	B
	13		B	B	A	AB							A	A			B	B	B	B
P ₁ '	1		A			B								B						A
	2	A	A	A	B	BA							B	B				A	A	A
	3					B														B
	4		A	A	B	A							B	B			A	A	A	A
	5	A	A	A	B	B							B	B				A	A	A
	6		A	A	B	BA							B	B				A	A	AB
	7		A	A	B	B							B	B				A	A	A
	8		A	A	B	BA							B	B				A	A	A
	9		A	A	B	BA							B	B				A	A	A
	10		A	A	B	B							B	BA				A	A	A
	11		A	A	B	B							B	BA			A	A	A	A
	12		A	A	B	BA							B	BA			A	A	A	A
	13		A	A	B	BA							B	B				A	A	A
P ₂	1					A		A	A		A			B						A
	2					A	A	A	A		A	A	A	B		A				A
	3					A	A	A	A		A	A	A	B						A
	4					A	A	A	A		A	A	A	B						A
	5					A	A	A	A		A	A	A	B						A
	6					A	A	A	A		A	A	A	B						A
	7					A	A	A	A		A	A	A	B						A
	8					A	A	A	A		A	A	A	B						A
	9					A	A	A	A		A	A	A	B						A
	10					A	A	A	A		A	A	A	B						A
	11					A	A	A	A		A	A	A	B						A
	12					A	A	A	A		A	A	A	B		A				A
	13					A	A	A	A		A	A	A	B						A
P ₂ '	1					B	B	B	B		B		B	A						B
	2					B	B	B	B		B	B	B	A		B				B
	3					B			B		B		B	A						B
	4					B	B	B	B		B		B	A						B
	5					B	B	B	B		B		B	A						B
	6					B	B	B	B		B		B	A						B
	7					B	B	B	B		B		B	A						B
	8					B	B	B	B		B		B	A						B
	9					B	B	B	B		B		B	A						B
	10					B	B	B	B		B	B	B	A						B
	11					B	B	B	B		B	B	B	A						B
	12					B	B	B	B		B	B	B	A						B
	13					B	B	B	B		B	B	B	A		B				B
P ₃	1						A				A	A						B	B	B
	2																			
	3	B				A					A	A	A					B		
	4	B				A					A	A	A		A			B		
	5					A				A									B	
	6	B	B			A							A						B	
	7	B	B			A							A						B	
	8	B	B			A							A						B	
	9	B	B			A							A						B	
	10																			
	11	B					A					A	A		A			B	B	
	12	B					A					A	A					B		
	13	B					A					A	A					B	B	
P ₃ '	1	A					B				B	B	B					A	A	
	2																			
	3	A	A			B	B	B		B	B	B	B					A	A	
	4	A					B				B	B	B		B			A		
	5	A									B	B	B						A	
	6	A	A			B	B					B	B						A	
	7	A	A			B	B					B	B						A	
	8	A	A			B	B					B	B						A	
	9	A	A			B	B					B	B						A	
	10	A	A				B					B						A	A	
	11																			
	12	A	A				B				B	B	B		B			A	A	
	13	A					B				B	B	B			B		A		

^a The first column (leftmost) of the table identifies the fragments, and the second one lists the inhibitors according to their serial numbers in Table 1. The uppermost row identifies amino acid residues of HIV-1 protease that participate in binding. The second row identifies sequential numbers of corresponding amino acid residues in the monomer enzyme. Letters A and B indicate which of the two monomers of the homodimer enzyme is involved in binding a particular P fragment. P₁ and P₁' fragments of some of the inhibitors have interactions with the same amino acid residues of both monomers, and those residues are denoted as AB or BA.

inhibitors fragment that binds to the S_1 - S_1' pockets of the enzyme, in agreement with our study. The same role of the central fragments has been found from the study of a series of related inhibitors (Clare, 1993). There are two exceptions to this general pattern of binding. They correspond to inhibitors U-85548E and JG-365. In these complexes the binding of fragment P_2' is more favored than that of P_1 or P_1' . The observed displacement in the location of the strongest binding site could be due to the functional groups located at P_1 . For instance, two of the inhibitors that differ on their placement of the strongest binding site are U-85548E, which has its maximum binding located at P_2' , and Ro-31-8558, which has it maximum at P_1 . The central triads of these inhibitors differ only in the residue at position P_1 . In the former inhibitor this fragment is occupied by a Leu, whereas in the latter it is filled by a Cha, a difference that could explain this shift.

Although some binding sites provide more binding capacity than others, all fragments provide a stabilizing contribution to binding. This result is consistent with the experimental finding that longer homologous peptidic inhibitors bind more strongly (Reddy et al., 1991). The larger molecules take advantage of the binding affinity of a greater number of binding sites (Kempf et al., 1990). Beyond the range of the segment stretching from P_2 through P_2' , the binding energies of particular groups have a broader range of variability and contribute, in general, less to

binding. These groups present weaker binding because they are part of terminal portions of the inhibitors, which include polar groups and have a higher degree of exposure to solvent.

In the early stages of the search for potent HIV-1 PR inhibitors it was argued that the best inhibitor would be the one that could take full advantage of the twofold symmetry of the unbound enzyme. It seemed that the biggest advantage of this type of inhibitors could be their ability to provide almost symmetrically distributed strong binding for the stretch of P and P' fragments (Erickson et al., 1990). The binding patterns of symmetric inhibitors (Fig. 3A) show that these ligands do not bind in a totally symmetric manner. The symmetry breakage already occurs in a subtle manner for the segment P_1 - P_1' and is more pronounced for fragments lying closer to the inhibitor's termini. The deviations from symmetry originate from small differences in the interaction energy of the inhibitors with the A and B monomers (Fig. 3A) and are caused by the impossibility of structurally satisfying some of the strongest interactions with both monomers simultaneously (Hosur et al., 1994).

The upper four panels of Figure 3A include three diastereomeric C_2 symmetry-based diol inhibitors and a deshydroxy analogue. The three-dimensional structures of these compounds were recently solved (Hosur et al., 1994). The predicted free energy of binding shows that the R,R diol compound (i.e., A-76889) has one of the poorest affinities for the HIV-1 PR in this group,

in agreement with experiment (see Table 1). Inspection of the superimposed structures show that the difference in free energy between the R,R diol and the rest of the diastereoisomers may reside at least in part on the isostere fragment (Hosur et al., 1994). In this subset of inhibitors, A-78791 binds as strongly as some of the inhibitors that contain two OH groups in its isostere (see Table 1). This suggests that at least one of the OH groups of the series studied above does not contribute significantly to binding. This conclusion is in line with the idea presented above that the transfer of polar groups from water to the interior of the HIV-1 PR is not always compensated by the electrostatic interactions generated between the enzyme and the inhibitor.

Asymmetric inhibitors give pronounced differences in the free energy contributions for their P and P' fragments. All the asymmetric inhibitors have one or two P fragments that provide the strongest contribution to their binding, whereas the symmetric ones distribute their binding capacity more equally between the four central P fragments (see Fig. 3). Nevertheless, asymmetric inhibitors can reach as high an affinity as their symmetric counterparts because single P positions of the asymmetric inhibitors contribute sometimes more to binding than equivalent P positions of the symmetric inhibitors. For instance the P₁' position of inhibitor Ro-31-8959 contributes more than 4.2 kcal/mol, whereas the strongest binder among the symmetric inhibitors at this position (L-700,417) contributes less than 3.6 kcal/mol.

Our approach has enabled us to identify the contributions of single functional groups to the free energy of binding. The profiles in Figure 3 show that identical hydrophobic side chains of different inhibitors provide similar contributions to the binding at any position in the range from P₂ through P₂'. For instance the contribution for Val is around 2.4 kcal/mol; for Ile it is close to 3 kcal/mol. The largest spread in the contribution of single functional groups is provided by the Phe groups located in the P₁-P₁' segment of symmetric inhibitors. This group contributes from 2.7 kcal/mol (in inhibitors like A-76889, A-76928, A-77003, and A-78791) to 3.3 kcal/mol (in L-700,417). The stronger affinity of the central fragment of the latter inhibitor seems to be related to the size of the S₁-S₁' binding pockets shown in Table 2. As seen in this table, the P₁-P₁' fragment of L-700,417 has a bigger binding pocket than the rest of the symmetric inhibitors because it includes the side chain of Arg-8 of both monomers, absent in the diol diastereoisomers and the des-hydroxy analogue (see Table 2).

The known complexes of HIV-1 PR with the symmetric inhibitors only include those compounds that have Phe groups in their P₁-P₁' segments. It remains to be seen if other functional groups at this position can contribute as much to binding. In the case of asymmetric inhibitors many groups may bind with similar strength. For instance S₁ can accommodate a large variety of hydrophobic groups ranging from Phe in SKB-Va, Cha in Ro-31-8558 and U-75875, and Leu in acetylpepstatin. All these residues bind to the S₁ pockets with binding energies ranging from 2.8 to 3.6 kcal/mol. This extended specificity is also indicated by our earlier observation that the strongest binding may occur at any of the four central positions (i.e., from P₂ to P₂'). It seems that the binding at these positions is based on hydrophobic interactions that lack the specificity of electrostatic interactions. This broader specificity has been experimentally observed in the substrate recognition by this enzyme. It is known that the HIV-1 PR hydrolyzes the precursor molecule in at least seven sites that differ in their amino acid sequences (Darke et al., 1988).

Some of the X-ray structures reveal two alternative orientations (roughly twofold related) of the inhibitor inside HIV-1 PR (Fitzgerald et al., 1990). We have calculated free energy of binding for both observed orientations of acetylpepstatin. The resulting contributions for the P fragments are very close for both orientations, but P positions for the second orientation have a reversed pattern of their free energy of binding to A and B monomers.

Concluding remarks

In this paper we have introduced a simple model for the prediction of the binding constants of a series of inhibitors to the HIV-1 PR. Our method produced free energies of binding that are in good agreement with experiment. Moreover, analysis of the results has allowed us to rank the importance of the binding pockets of the enzyme as well as to determine their specificity for given functional groups. These results are in good agreement with experimental results, providing physical validation for our approach, and should be useful for designing novel high-affinity inhibitors.

A better fit for the free energies of binding (Table 1) could be achieved only at the cost of using a more elaborate approach that would require substantially greater computer time. Some of the differences between the observed and calculated free energies could be ascribed to the rigid body approximation. In this approximation the structure of the inhibitor in solution is assumed to be the same as the one in the complex (Horton & Lewis, 1992; Krystek et al., 1993). In a few cases it has been found that this approximation has led to overestimates in the binding affinities. For instance, the calculated energy of binding between trypsinogen and BPTI was found to be 10 kcal/mol stronger than the observed one (Horton & Lewis, 1992), the difference being ascribed to the conformational changes that accompany binding.

The problem of the inhibitor conformation (or a set of probable conformations) in water requires special consideration. As seen in Table 1, the biggest difference between the observed and the calculated free energies of binding is for the complex with A-74704; our approach overestimates the binding of this compound by 1.8 kcal/mol. We have carried out simulated annealing calculations on this inhibitor in solution in order to check the role of conformational changes that could occur on binding. Preliminary analysis of the results show that the A-74704 inhibitor may have a conformation in water with two aromatic rings of its central part screening each other (L. Martínez & F. Sussman, unpubl. results). In its complex with the enzyme the inhibitor acquires a more extended conformation with the rings buried among residues of the enzyme. Our simulated annealing results are in line with some recent NMR studies of small peptides with more than one aromatic group in their sequences, which have shown that the aqueous solution structure of such molecules is highly populated with folded conformers in which the aromatic rings are stacked against other residues (Yao et al., 1994). When we estimated the free energy of binding of A-74704, using some of the more compact structures obtained by simulating annealing, the difference between the observed and calculated free energies of binding of this inhibitor dropped from 1.8 kcal/mol to 1.3 kcal/mol. This demonstrates that for some inhibitors, the inclusion of the structure of the inhibitor

in solution helps to bring the free energy of binding close to experimental values.

Our approach works satisfactorily for most of the compounds even though Equation 1 does not have an explicit electrostatic term. It is possible to improve the method by adding an explicit term to account for the difference in electrostatic interactions of these molecules in the bound and unbound states. A more realistic presentation of the distribution of charges for the molecules may be necessary as well. This should include the determination of the state of protonation of the two active-site aspartates, Asp-25 and Asp-25', that participate directly in the enzymatic reaction. Work on these open questions is in progress.

Acknowledgments

This work was supported by NIH grant R01-GM49553 and by OCAST grant HS3-011. We are thankful to Dr. J. Tang for providing binding constants of several HIV-1 PR-inhibitor complexes, including the one with the U-89360E inhibitor, as well as Drs. A. Wlodawer and J.W. Erickson for sharing with us their unpublished data on crystal structures.

References

- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:535-542.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187-217.
- Clare M. 1993. HIV protease: Structure-based design. *Perspect Drug Discovery Des* 1:49-68.
- Darke PL, Nutt RF, Brady SF, Garsky VM, Ciccarone TM, Leu CT, Lumma PK, Freidinger RM, Veber DF, Sigal IS. 1988. HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins. *Biochem Biophys Res Commun* 156:297-303.
- Davies DR. 1990. The structure and function of the aspartic proteinases. *Annu Rev Biophys Biophys Chem* 19:189-215.
- Debouck C. 1992. The HIV-1 protease as a therapeutic target for AIDS. *Aids Res Hum Retroviruses* 8:153-164.
- Doweyko AM. 1994. Three-dimensional pharmacophores from binding data. *J Med Chem* 37:1769-1778.
- Dreyer GB, Lambert DM, Meek TD, Carr TJ, Tomaszek TA Jr, Fernandez AV, Bartus H, Cacciavillani E, Hassell AM, Minnich M, Petteway SR Jr, Metcalf BW, Lewis M. 1992. Hydroxyethylene isostere inhibitors of human immunodeficiency virus-1 protease: Structure-activity analysis using enzyme kinetics, X-ray crystallography, and infected T-cell assays. *Biochemistry* 31:6646-6659.
- Dreyer GB, Metcalf BW, Tomaszek TA Jr, Carr TJ, Chandler AC III, Hyland L, Fakhoury SA, Magaard VW, Moore ML, Strickler JE, Debouck C, Meek TD. 1989. Inhibition of human immunodeficiency virus 1 protease in vitro: Rational design of substrate analogue inhibitors. *Proc Natl Acad Sci USA* 86:9752-9756.
- Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. *Nature* 319:199-203.
- Erickson J, Neidhart DJ, VanDrie J, Kempf DJ, Wang XC, Norbeck DW, Plattner JJ, Rittenhouse JW, Turon M, Wideburg N, Kohlbrenner WE, Simmer R, Helfrich R, Paul DA, Knigge M. 1990. Design, activity, and 2.8 Å crystal structure of a C₂ symmetric inhibitor complexed to HIV-1 protease. *Science* 249:527-533.
- Fitzgerald PMD, McKeever BM, Van Middlesworth JF, Springer JP, Heimbach JC, Leu CT, Herber WK, Dixon RAF, Darke PL. 1990. Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0-Å resolution. *J Biol Chem* 265:14209-14219.
- Fitzgerald PMD, Springer JP. 1991. Structure and function of retroviral proteases. *Annu Rev Biophys Biophys Chem* 20:299-320.
- Ho DD, Tayoshima T, Mo H, Kempf DJ, Norbeck D, Chen CM, Wideburg NE, Burt SK, Erickson JW, Singh MK. 1994. Characterization of human immunodeficiency virus type 1 variants with increased resistance to a C₂-symmetric protease inhibitor. *J Virol* 68:2016-2020.
- Horton N, Lewis M. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci* 1:169-181.
- Hosur MV, Bhat TN, Kempf DJ, Baldwin ET, Liu B, Gulnik S, Wideburg NE, Norbeck DW, Appelt K, Erickson JW. 1994. Influence of stereochemistry on activity and binding modes for C₂ symmetry-based diol inhibitors of HIV-1 protease. *J Am Chem Soc* 116:847-855.
- Huff JR. 1991. HIV protease: A novel chemotherapeutic target for AIDS. *J Med Chem* 34:2305-2314.
- Jaskólski M, Tomasselli AG, Sawyer TK, Staples DG, Heinrichson RL, Schneider J, Kent SBH, Wlodawer A. 1991. Structure at 2.5-Å resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry* 30:1600-1609.
- Kempf DJ, Norbeck DW, Codacovi L, Wang XC, Kohlbrenner WE, Wideburg NE, Paul DA, Knigge MF, Vasavanonda S, Craig-Kennard A, Saldívar A, Rosenbrook W Jr, Clement JJ, Plattner JJ, Erickson J. 1990. Structure-based, C₂ symmetric inhibitors of HIV protease. *J Med Chem* 33:2687-2689.
- Kramer RA, Schaber MD, Skalka AM, Ganguly K, Wong-Staal F, Reddy EP. 1986. HTLV-III gag protein is processed in yeast cells by the virus pol-protease. *Science* 231:1580-1584.
- Krystek S, Stouch T, Novotny J. 1993. Affinity and specificity of serine endopeptidase-protein inhibitor interactions. *J Mol Biol* 234:661-679.
- Lee B, Richards FM. 1971. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol* 55:379-400.
- Lin Y, Lin X, Hong L, Foundling S, Heinrichson RL, Thaisrivongs S, Leelamanit W, Raterman D, Shah M, Dunn BM, Tang J. 1995. Effect of point mutations on the kinetics and the inhibition of human immunodeficiency virus type 1 protease: Relationship to drug resistance. *Biochemistry* 34:1143-1152.
- Loeb DD, Swanstrom R, Everitt L, Manchester M, Stamper SE, Hutchison CA III. 1989. Complete mutagenesis of the HIV-1 protease. *Nature* 340:397-400.
- Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SBH, Wlodawer A. 1989. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science* 246:1149-1152.
- Mitsuya H, Yarchoan R, Broder S. 1990. Molecular targets for AIDS therapy. *Science* 249:1533-1544.
- Nauchitel VV, Somorjai RL. 1993. Gaussian neighborhood: A new measure of accessibility for residues of protein molecules. *Proteins Struct Funct Genet* 15:50-61.
- Novotny J, Sharp K. 1992. Electrostatic fields in antibodies and antibody/antigen complexes. *Prog Biophys Mol Biol* 58:203-224.
- Reddy MR, Viswanadhan VN, Weinstein JN. 1991. Relative differences in the binding free energies of human immunodeficiency virus 1 protease inhibitors: A thermodynamic cycle-perturbation approach. *Proc Natl Acad Sci USA* 88:10287-10291.
- Richards AD, Roberts R, Dunn BM, Graves MC, Kay J. 1989. Effective blocking of HIV-1 proteinase activity by characteristic inhibitors of aspartic proteinases. *FEBS Lett* 247:113-117.
- Roberts NA, Martin JA, Kington D, Broadhurst AV, Craig JC, Duncan IB, Galpin SA, Handa BK, Kay J, Kröhn A, Lambert RW, Merrett JH, Mills JS, Parkes KEB, Redshaw S, Ritchie AJ, Taylor DL, Thomas GJ, Machin PJ. 1990. Rational design of peptide-based HIV proteinase inhibitors. *Science* 248:358-361.
- Schechter I, Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27:157-162.
- Thanki N, Rao JKM, Foundling SI, Howe WJ, Moon JB, Hui JO, Tomasselli AG, Heinrichson RL, Thaisrivongs S, Wlodawer A. 1992. Crystal structure of a complex of HIV-1 protease with a dihydroxyethylene-containing inhibitor: Comparisons with molecular modeling. *Protein Sci* 1:1061-1072.
- Tomasselli AG, Hui JO, Sawyer TK, Staples DJ, Bannow C, Reardon IM, Howe WJ, DeCamp DL, Craik CS, Heinrichson RL. 1990. Specificity and inhibition of proteases from human immunodeficiency viruses 1 and 2. *J Biol Chem* 265:14675-14683.
- Wlodawer A, Erickson JW. 1993. Structure-based inhibitors of HIV-1 protease. *Annu Rev Biochem* 62:543-585.
- Yao J, Dyson HJ, Wright PE. 1994. Three-dimensional structure of a type VI turn in a linear peptide in water solution. *J Mol Biol* 243:754-766.