A preference-based free-energy parameterization of enzyme–inhibitor binding. Applications to HIV-1-protease inhibitor design

A. WALLQVIST,¹ R.L. JERNIGAN,² AND D.G. COVELL¹

¹Frederick Cancer Research and Development Center, National Cancer Institute, Science Applications International Corporation, Frederick, Maryland 21702

²Laboratory of Mathematical Biology, National Cancer Institute, Building 12B, Room B116, MSC 5677 Bethesda, Maryland 20892-5677

(RECEIVED MARCH 29, 1995; ACCEPTED JUNE 28, 1995)

Abstract

The interface between protein receptor–ligand complexes has been studied with respect to their binary interatomic interactions. Crystal structure data have been used to catalogue surfaces buried by atoms from each member of a bound complex and determine a statistical preference for pairs of amino-acid atoms. A simple free energy model of the receptor–ligand system is constructed from these atom–atom preferences and used to assess the energetic importance of interfacial interactions. The free energy approximation of binding strength in this model has a reliability of about ± 1.5 kcal/mol, despite limited knowledge of the unbound states. The main utility of such a scheme lies in the identification of important stabilizing atomic interactions across the receptor–ligand interface. Thus, apart from an overall hydrophobic attraction (Young L, Jernigan RL, Covell DG, 1994, *Protein Sci* 3:717–729), a rich variety of specific interactions is observed. An analysis of 10 HIV-1 protease inhibitor complexes is presented that reveals a common binding motif comprised of energetically important contacts with a rather limited set of atoms. Design improvements to existing HIV-1 protease inhibitors are explored based on a detailed analysis of this binding motif.

Keywords: atom–atom interactions; binding; molecular surface; protein; recognition

The formation of two or more biomolecules into a stable complex is a complicated phenomenon (Perutz, 1990), generally characterized by a delicate balance between entropic and enthalpic differences between the bound and unbound states (Frank & Evans, 1945; Kauzmann, 1959; Hopfinger, 1977; Hansch & Leo, 1979; Tanford, 1980). The net difference of these terms is the free energy of binding, ΔG_{bind} , which is, of course, central to investigations of biomolecular interactions. One component of the free energy results from nonbonded associations between interfacial atoms. For enzymes and their substrates a lock-and-key paradigm has been accepted in which suitable residues form contacts at the interface mostly due to their geometric complementarity and in part also due to their favorable energetic complementarity. A strict mechanical view of two rigid sterically matched surfaces is incomplete (Jorgensen, 1991) but has often served as a starting point for detailed investigations of binding. The analysis proposed here focuses on the details of such rigid

molecular interactions across the interface. A better comprehension of these features should enable a more rational design of potent bioeffector molecules (Hansch & Leo, 1979; Kuntz et al., 1994). In the more complicated induced-fit case, it would also be necessary to comprehend and accommodate changes in intramolecular structure, introducing the corresponding changes to inter- and intramolecular energy terms. In principle, the complete understanding of ΔG_{bind} must include a treatment of both enthalpic effects with respect to the unbound solvated state as well as entropic considerations of both solute and solvent.

The nature of noncovalent atomic interactions within a binding interface has been the subject of many studies aimed, in part, at understanding the determinants of binding strength and specificity (Stryer, 1988; Janin & Chothia, 1990). Empirical attempts to identify preferred nonbonded interactions have emphasized the importance of electric as well as hydrophobic interactions between atoms, functional groups, and residues within the binding interface (Wodak & Janin, 1978; Burley & Petsko, 1985; Miyazawa & Jernigan, 1985, 1993; Miller et al., 1987; Bohacek & McMartin, 1992; Casari & Sippl, 1992; Sippl, 1993; Wodak & Rومان, 1993). These approaches have attempted to establish the preference between contacting atomic pairs by statisti-

Reprint requests to: David Covell, Frederick Cancer Research and Development Center, National Cancer Institute, Science Applications International Corporation, Frederick, Maryland 21702; e-mail: covell@faxpdc.ncifcrf.gov.

cal analyses of distances between interacting entities. Alternative approaches have featured the importance of shape complementarity within the binding interface (Nicholls et al., 1991; Walls & Sternberg, 1992; Lawrence & Colman, 1993; Srivastava & Crippen, 1993; Norel et al., 1994; Fischer et al., 1995). These studies identify patterns in the interfacial geometry for positional preferences of atoms, functional groups, or residues around a central entity (Singh & Thornton, 1992). Such tables provide data about preferred compositions and orientations for various types of interactions possible within a specified interface (Chakrabarti & Dunitz, 1982; Cody & Murray-Rust, 1984; Carrell et al., 1988; Roe & Teeter, 1993; Walshaw & Goodfellow, 1993). All of the above approaches have attempted to identify the important details of those interactions that result in each molecule's mutual recognition of and binding to its target.

In this paper a new free energy description of binding is developed that is based on a systematic analysis of atomic interactions between adjacent buried surfaces within a binding interface. Contact frequencies between adjacent atomic surfaces from each member of a complex are used to determine parameters for a model of binding strength. Motivation for this surface-based approach is based on the thermodynamic aspects describing the transfer of a solvent-accessible surface to an inaccessible location within the binding interface (Ben-Naim, 1990). Previous treatments of this solvophobic effect have considered the energetic cost of removing an accessible surface from water and replacing it with a generic "protein" environment, e.g., *n*-octanol (Eisenberg & McLachlan, 1986). In the context of protein binding, i.e., transfer of surfaces from an aqueous solution to a protein interface, a different mix of energy costs for hydrophobic and hydrophilic surface elements has been considered (Horton & Lewis, 1992). The work presented here is an extension of the importance of hydrophobic interactions in protein-protein recognition (Miller et al., 1987; Covell, 1994) to include a pairwise accounting of preferred atomic surfaces that are adjacent when bound.

The present analysis examines crystallographic protein complexes found in the Brookhaven Protein Data Bank (Bernstein et al., 1977) and retains the atomic identity of atoms with adjacent buried surfaces. The underlying assumption is that only close atoms interact and that there are no long-range interactions. The procedure yields a set of binary atomic preferences that can be used to estimate the free energy cost of burying surfaces of various atomic types depending on the identity of atom pairs and the extent of mutually buried surface. These preferences are then useful for addressing questions on the importance of hydrophobic and hydrophilic contributions to binding strength and specificity. The striking feature of these results is the diversity in the types of atomic interactions between protein surfaces. Even though the largest contributors to the free energy of binding arise from hydrophobic atom pairs, other hydrophilic as well as mixed interactions form a non-negligible portion. Thus, a purely hydrophobic picture is too crude to account for the complete range of interactions between proteins. The results of this present analysis find, for example, that the strongest interactions occur between ionic pairs that form salt bridges. Mutually attractive preferences are still observed among the hydrophobic atoms, with aromatic-aromatic surface contacts being stronger than aliphatic-aliphatic contacts. Strong attractions are observed for acidic atoms to basic atoms in arginine and lysine, with arginine being the stronger. Likewise, a large attrac-

tion is noted between basic atoms to atoms in aspartic and glutamic acids, with aspartic acid being more attractive than glutamic acid. Atoms of serine and threonine are found to have a strong preference for atoms in histidine and aspartic acid.

The results of this analysis can be applied directly to the selection and rational design of inhibitors against target proteins. These atom-atom preferences are used to probe general aspects of ligand binding, such as the contributions of individual residues to binding and the type of atomic interactions at the binding interface. Individual atom positions can be probed to deduce whether these interactions, i.e., atom type and atom position, are optimal for binding free energy. Thus, the procedure lends itself rather directly to proposing improvements for strengthening the binding of known inhibitors to an enzyme.

As a demonstration, an analysis of inhibitors to the HIV-1 protease dimer (HIV-1-PR) is presented; this examination systematically evaluates the atomic surface-surface interactions involved in inhibitor binding and determines the type of atoms that contribute most to binding stability. This approach identifies a common binding motif for the set of inhibitors examined that is comprised of hydrophobic interactions between four aliphatic and two aromatic atom surfaces combined with a set of hydrogen bonding sites. These features characterize all reported protease inhibitors. Researchers have unjustifiably emphasized the importance of the hydrogen bonding pattern at the binding site for an effective HIV-1-PR inhibitor. Our analysis shows that, even though these interactions are important, the bulk of the free energy of binding is gained from the specific hydrophobic interactions between ligand and target. In essence, all the investigated compounds conform to this hitherto undefined and overlooked motif.

Estimating free energies of binding, ΔG_{bind}

The relevant biological process is the association of a ligand and its receptor, i.e.,

where *A* and *B* represent biomolecules that, upon binding, form a stable complex. Examples of this interaction for proteins include the formation of enzyme-inhibitor complexes, as well as the subunit-subunit associations in their oligomeric form. The binding process is assumed to occur in a "generic" aqueous environment at room temperature. The problem of calculating the free energy associated with Equation 1 on the molecular level is truly daunting. In principle, molecular simulations and free energy perturbation techniques offer a direct approach to the problem without the need for mean field approximations of molecular interactions. However, because of the large number of degrees of freedom associated with macromolecular interactions, computer simulations of such systems are not yet capable of providing a viable route for estimating binding strengths (Yun-Yu et al., 1993). Also, it is not clear whether there exists an appropriate and reliable force field that can account for molecular interactions at the required level of accuracy (Rullmann & van Duijnen, 1990).

The free energy change of binding a ligand to a receptor has many contributions (Kauzmann, 1959; Nemethy, 1967), including hydrogen bonds (Shirley et al., 1992), hydrophobic and hy-

drophilic interactions, van der Waals attractions and repulsions, as well as entropic contributions from side- and main-chain motion and the surrounding solvent. It is not possible to invoke any general statements about the relative importance of each of these terms because diverse arrangements occur that can contradict any attempted generalizations. These terms also may not always act independently and thus might not permit a general dissection of the free energy change into pairwise additive components (Mark & van Gunsteren, 1994). Thus, no reduction of the full atomic problem can be exact for the general ligand–receptor binding phenomena.

However, through approximate approaches, progress can follow. In Equation 1, the unassociated molecules interact with the solvent, whereas in their associated state they interact with each other as well as with solvent. In this latter state, they may have changed their average nonassociated conformation, either globally or locally. For the rigid targets considered here, the problem simplifies to identifying the bound state and estimating the difference in free energy upon desolvation of the buried surface area and the concomitant associative free energy between the buried surface elements of each molecule. Yet, any simplification of the free and bound state still entails an estimation of the free energy, a quantity that can be quite different from the potential energy calculated with a force field. In particular, the additional entropic contribution to the free energy of association depends on the available states of the entire system and, in general, would not be proportional simply to the potential energy.

Instead of pursuing the extremely difficult task of calculating directly *a priori* free energies, one can instead utilize the empirical observation that thermodynamic quantities such as enthalpies, entropies, and heat capacities of solvation for biological macromolecules exhibit a strong correlation with solvent-exposed surfaces (Makhatadze & Privalov, 1990; Privalov & Makhatadze, 1990, 1992). The same observation also holds true, in most cases, for small organic solutes (Makhatadze & Privalov, 1993; Privalov & Makhatadze, 1993). Computer simulations of alkanes in water also support the correlation between surface area and thermodynamic properties (Jorgensen et al., 1985). Furthermore, the hydrophobicity of each amino acid can also be related to the degree of solvent exposure in globular proteins (Rose et al., 1985). Lee and Richards calculated the solvent-accessible surface of a fixed molecular conformation by using a rolling probe, with dimensions corresponding to the solvent, to sweep over the surface (Richards, 1977; Richmond, 1984). The close correspondence between solvent-accessible surfaces and thermodynamic quantities has prompted many investigators to explore the connection with models based on solvent exposure. One should emphasize that this connection is empirical and attempts to reconcile these observations with microscopic theories of cavity formation (Nicholls et al., 1991; Sharp et al., 1991; Tunon et al., 1992) have been criticized (Ben-Naim & Mazo, 1993; Ben-Naim, 1994; Giesen et al., 1994; Lee, 1994). However, as an empirical concept, solvent exposure has proven to be a highly effective way of developing an understanding of experimental observations. Both hydrophilic and hydrophobic terms have certain physical aspects that lend themselves to a surface area treatment. For both charge interactions and hydrogen bonds, single or multiple pairings are possible about a single acceptor or donor atom. On the surface of a protein, formation of single or multiple interactions with an atom depends substantially on its accessibility, which can be described by surface ex-

posure. Likewise, the extent of unfavorable water interactions avoided upon burial or by forming hydrophobic interactions depends on the surface area available for interactions. All of these processes for atoms actually involve discrete choices from a few possible numbers of interactions. The geometries of approach in these interactions are complex, however, and serve to smooth over the discrete nature of interactions. The direct expedient use of surface areas to approximate the extent of all these surface interactions avoids these geometric complications and their complicated energetic implications.

Eisenberg and coworkers (Eisenberg & McLachlan, 1986; Eisenberg et al., 1989; Wesson & Eisenberg, 1992) have introduced a set of atomic solvation parameters for the treatment of protein solvation. The free energy of solvation is given as:

$$\Delta G_{\text{solv}} = \sum_i \alpha A_i^{\text{exp}} \sigma_i \quad (2)$$

where αA_i^{exp} is the solvent-exposed surface area of atom type i , and σ_i is the corresponding solvation parameter. Likewise, Ooi et al. (1987) studied the contribution of exposed surface area as an additional energy term to a usual molecular mechanics formulation. Similar concepts have been used to propose models for enzyme–substrate binding. Novotny et al. (1989) combined the effects of hydrophobic surface area burial, explicit electrostatic calculations, and entropy considerations to study antigen–antibody binding. Wilson et al. (1991) coupled a direct potential energy evaluation of enzyme–inhibitor interactions with a solvation term derived from the change in surface area burial upon binding; the relative weights of the potential energy and the solvation contribution were then varied to account for the observed data. Horton and Lewis (1992) focused on the changes in hydrophobic and hydrophilic surface area burial upon binding and were able to fit successfully the experimental free energy data for several protease enzyme–inhibitor complexes, giving perhaps the most compelling reason to use surface area as the physical model of interactions between two rigid bodies upon binding.

Model construction

The method for calculating free energies of binding between proteins is based on the observed atom–atom preferences obtained from the database. This section describes how atom–atom preferences are determined and used as approximate free energies of binding between protein molecules.

Calculation of atom–atom surface preferences

Preferences between molecular surfaces (Lee & Richards, 1971; Richards, 1977; Connolly, 1983) are developed from an analysis of 38 high-resolution protein crystal complexes taken from the Protein Data Bank (PDB) (Bernstein et al., 1977). A list of these structures is given in Table 1, together with their stated resolution and buried surface area. The procedure for developing atom-based surface preferences is comprised of two steps. First, the molecular surfaces for each molecule in its uncomplexed and in its complexed forms are generated using the Connolly algorithm (Connolly, 1983), with a probe radius of 1.4 Å, a surface density of two dots per square Ångström, and atomic radii taken from Bondi (1964). The amount of interface area for each atom type is catalogued by comparison of the complexed and uncom-

Table 1. Crystal structures used for interfacial preferences^a

PDB code	Molecule description	Interface	Area (Å ²)	Resolution (Å)
1AAR	Di-ubiquitin	Dimer	1,235	2.3
1AAZ	Glutaredoxin	Dimer	512	2.0
1AKE	Adenylate kinase	Dimer	470	1.9
1BAR	Fibroblast growth factor	Dimer	1,817	2.7
1CDT	Cardiotoxin	Dimer	895	2.5
1CHO	Chymotrypsin w/inhibitor	Complex	1,420	1.8
1COL	Colicin A	Dimer	1,145	2.4
1DFN	Defensin	Dimer	893	1.9
1DHF	Dihydrofolate reductase	Dimer	1,152	2.3
1GGI	Immunoglobulin w/16-residue peptide	Complex	532	2.8
1HNE	Neutrophil elastase complex	Complex	724	1.8
1MCP	Immunoglobulin FAB fragment	Complex	3,423	2.7
1MEE	Mesentericopeptidase w/Eglin C	Complex	1,737	2.0
1PP2	Phospholipase	Dimer	2,451	2.5
1REI	Immunoglobulin	Dimer	1,397	2.0
1RVE	Endonuclease	Dimer	2,846	2.5
1TEC	Thermitase w/Eglin C	Complex	1,496	2.2
1TGS	Trypsinogen w/inhibitor	Complex	1,705	1.8
1TLP	Thermolysin w/phosphoramidon	Complex	838	2.3
1TPA	Anhydro-trypsin w/inhibitor	Complex	1,420	1.9
2ABX	Neurotoxin	Dimer	705	2.5
2AZA	Azurin	Dimer	994	1.8
2ER9	Endothiapepsin w/inhibitor	Complex	1,471	2.2
2KAI	Kallikrein A w/trypsin inhibitor	Complex	1,406	2.5
2PTC	Beta-trypsin w/inhibitor	Complex	1,400	1.9
2SEC	Subtilisin Carlsberg w/Eglin C	Complex	1,451	1.8
2SNI	Subtilisin Novo w/chymotrypsin inhibitor	Complex	1,601	2.1
2TGP	Trypsinogen w/inhibitor	Complex	1,400	1.9
2UTG	Uteroglobin	Dimer	2,953	1.6
2ZTA	Leucine zipper	Dimer	1,633	1.8
3GAP	Catabolite gene activator protein	Dimer	2,934	2.5
3HLA	Histocomplex antigen	Complex	2,512	2.6
4SGB	Serine protease w/PCI-1	Complex	1,285	2.1
4TPI	Trypsinogen w/inhibitor	Complex	1,454	2.2
5APR	Acid proteinase w/pepstatin	Complex	1,399	2.1
5HVP	HIV-1 protease complex	Dimer	2,880	2.0
6EBX	Erabutoxin	Dimer	545	1.7
6TMN	Thermolysin w/inhibitor	Complex	852	1.6

^a In the calculation of the surface-exposed areas for either enzyme or inhibitor, there is a small discrepancy between the amount of area buried by each member of the complex. This deviation arises from the difference in local surface complementarity as well as from edge effects around the interface. The derived preferences are taken as a mean between enzyme-inhibitor and inhibitor-enzyme preferences.

plexed areas. This results in the characterization of an average buried interfacial area of around 1,500 Å² in terms of 3,000 surface elements. The next step in the calculation is to determine the atom pairs at adjacent surfaces within the interface. A score, S^{ab} , reflecting the interaction geometry of any surface element, a , with any surface element, b , on the adjacent surface is given as:

$$S^{ab} = \frac{1}{d_{ab} |\Theta(\mathbf{r}_a, \mathbf{r}_b) - \pi|}, \quad (3)$$

where d_{ab} is the distance between surface elements a and b and $\Theta(\mathbf{r}_a, \mathbf{r}_b)$ is the angle formed by the surface normal vectors, $\mathbf{r}_a$ and $\mathbf{r}_b$, at these points. Ideally, surfaces of perfectly packed

atoms in van der Waals contact would have $d_{ab} = 0$ Å, with $\Theta(\mathbf{r}_a, \mathbf{r}_b) = 180^\circ$. A small cutoff is introduced in Equation 3 to avoid singularities that might occur at $d_{ab} = 0$ and $\Theta(\mathbf{r}_a, \mathbf{r}_b) = \pi$. Practically, adjacent surfaces are less well packed and have a range of values for d_{ab} and $\Theta(\mathbf{r}_a, \mathbf{r}_b)$. To identify the atom-atom class of each interacting surface element, S^{ab} is calculated for all ab pairs with $d_{ab} \leq 2.8$ Å. An upper limit of 2.8 Å is chosen for d_{ab} to allow consideration of only close surfaces that exclude an intervening water molecule of radius 1.4 Å. The pair of surface elements yielding the highest score, S^{ab} , identifies the atom-atom class of each interacting surface element. Figure 1 shows the observed frequencies of distance, d_{ab} , and $\cos \Theta$ pairs selected by the scoring function S^{ab} in Equation 3. The clear trend demonstrates that the most frequent surface elements

Fig. 1. Frequency distribution for distance between two surface elements and the corresponding cosine of the angle between their respective surface normals for the 38 crystal structures listed in Table 1. A cosine value of -1.0 would indicate that the surfaces are parallel to each other.

selected using Equation 3 have $d_{ab} \sim 2.8$ Å and $\cos \Theta \sim -1.0$. At cutoff distances beyond 2.8 Å, the correspondence between tangential, close surfaces is nearly identical to that shown in Figure 1, that is, better scoring ab pairs are not found at longer distances. For $d_{ab} < 2.8$ Å, an increasingly random frequency of occurrence is found for d_{ab} and $\Theta(\mathbf{r}_a, \mathbf{r}_b)$ pairs as indicated by an increasing frequency of pairs with values of $\Theta(\mathbf{r}_a, \mathbf{r}_b) < 180^\circ$. Based on these considerations, S^{ab} with a cutoff distance of 2.8 Å is used to select pairs of surface elements.

The area of each surface element is accumulated for each atom pair and used to determine an atom–atom preference score, P_{ij} . Each preference is calculated as the ratio of the fraction of the total interfacial area contributed by each atom pair, $F_{ij} = A_{ij}/A_{tot}$, normalized by the product of the fractional contribution by each atom in the pair, $F_i = A_i/A_{tot}$, $F_j = A_j/A_{tot}$:

$$P_{ij} = \frac{F_{ij}}{F_i F_j}. \quad (4)$$

The highest and lowest scoring preferences, P_{ij} , identify the most- and least-observed adjacent atomic surfaces, respectively, in the data set. Atom groupings considered in this analysis are described in the next section.

Ideally, a random distribution of interfacial atom identities yields a preference score, P_{ij} , of one. In support of this claim, an extensive set of surface area calculations has been completed for each of the 38 complexes analyzed. The goal of this calculation was to sample randomly the set of all pairwise atomic surface interactions possible within each interface. This random distribution defines the likelihood of a chance occurrence of an adjacent atom–atom pair within the interface and provides a reference state for identification of nonrandom, or specific, interactions. To achieve this, the atomic identities for all atom positions having any portion of their surface buried within the interface of the native state are shuffled randomly. The surface preferences, Equation 4, are then derived for this data set. Forty random shufflings of interfacial atom identities for each complex analyzed in this study finds an average preference score of 1.05 ± 0.16 .

Selection of reduced atom types

In the calculation of surface preferences, the amino acid atoms are divided somewhat arbitrarily into a large set of nearly unique atom types. Only heavy atoms, C, N, O, and S are considered, avoiding the problem of uniquely assigning hydrogen atoms. The assignment of atoms to these groups is based on chemical similarities between atoms. Table 2 lists these atomic

Table 2. Atom type classification

Element	Reduced atom type	Corresponding amino acid atoms
Carbon	C_{pep}	C_{pep}
	C_α	C^α
	C_β	C^β
	C_{aliph}	$C_V^1, C_V^2, C_L^1, C_L^2, C_I^1, C_I^2, C^\delta, C_T^\gamma, C_K^\gamma$ $C_K^\delta, C_E^\gamma, C_Q^\gamma, C_R^\gamma, C_R^\delta, C_M^\gamma, C_P^\gamma, C_P^\delta$
	C_{arom}	$C_H^{\delta 2}, C_H^1, C_F^\gamma, C_F^{\delta 1}, C_F^{\delta 2}, C_F^1, C_F^2, C_F^\gamma, C_W^{\delta 1}$ $C_W^{\delta 3}, C_W^{\delta 3}, C_W^{\delta 2}, C_Y^{\delta 1}, C_Y^{\delta 2}, C_Y^1, C_Y^2, C_Y^\gamma$
	C_{amid}	C_N^γ, C_Q^δ
	C_{carb}	C_D^γ, C_E^δ
	C_{CNC}	$C_W^\gamma, C_W^{\delta 2}, C_W^2, C_H^\gamma$
	C_N^1	C_R^γ, C_K^δ
	C_Y	C_Y^γ
Nitrogen	N_{pep}	N_{pep}
	N_K	N_K^γ
	N_{arom}	$N_W^1, N_H^{\delta 1}, N_H^{\epsilon 2}$
	N_{amid}	$N_N^{\delta 2}, N_Q^2$
	N_R	$N_R^\epsilon, N_R^{\eta 1}, N_R^{\eta 2}$
Oxygen	O_{pep}	O_{pep}
	O_{alc}	O_S^γ, O_T^γ
	O_{amid}	$O_N^{\delta 1}, O_Q^1$
	O_{carb}	O_D, O_E
	O_Y	O_Y^η
Sulphur	S	S_C^γ, S_M^δ

groupings and their constituent amino acid atom types. All atoms of the peptide group -CONH- have been identified separately in the three classes, C_{pep} , N_{pep} , and O_{pep} . The division among carbon atoms takes into account the positional difference among C^α , C^β and other aliphatic methylene groups. Further groupings of carbon atoms distinguish aromatic atoms as well as carbons directly bonded to oxygen and nitrogen atoms. Similarly, oxygens and nitrogens are grouped together depending on their charge as well as their immediate connecting bonds. In this analysis, a total of 21 different atom types are identified separately. The selection of these atom types is made arbitrarily to facilitate an understanding of enzyme-inhibitor binding in terms of separate atomic entities. Further alternative classification schemes could be formulated easily, but the present groups are considered to be quite sufficiently detailed. The correlations among atom types are discussed in terms of binding parameters in the Appendix.

Free energy approximation

The atom-atom surface burial preferences calculated from a subset of all associated proteins represent a knowledge-based data set. It cannot directly be used to predict binding strengths because it does not contain any information about the unbound states in Equation 1, information necessary to obtain the free energy change between the complex and the free molecules. The probability of observing certain atomic pairs at the interface must be balanced with the probability of those atoms being exposed to solvent. Unfortunately, a precise solution conformation of the isolated molecules may not be derivable from the crystal structure of the complex. Although such structures could be determined experimentally (Otting et al., 1991), relatively few are available. Thus, in order to proceed, it is necessary to make assumptions about the unbound molecules. For both enzyme and inhibitor molecules, their exposed surface areas are calculated for the unbound molecules taken from the co-crystal structure. Structural changes upon complex formation are not considered in the present model. Although most enzymes or proteins of substantial size have a relatively rigid tertiary structure, smaller peptides are not as likely to remain rigid. Thus, the free energy change of desolvating the appropriate region of the bound enzyme-inhibitor complex and the free energy change for association of those surfaces is taken to be proportional to the jointly buried surface area. The free energy of association for the two molecules A and B is then given as,

$$\Delta G_{\text{pred}} = - \sum_{i \in A} \sum_{j \in B} \alpha_{ij} a_{ij} + \beta \quad (5)$$

where the summation runs over all atom types i and j of molecules A and B , respectively, α_{ij} an effective binding parameter dependent only on atom types i and j , a_{ij} is the mutual surface jointly buried by the atom types i and j in the complex, and β is a positive constant. It is not possible to explicitly fit all binding parameters directly because of the paucity of crystal data and accurate binding constants for each complex. Instead, α_{ij} is assumed to have two physical components, one that is directly proportional to the surface area and is independent of the surface atom type, and one that is specific to the buried atom types.

The effective binding parameter, α_{ij} , is then calculated as:

$$\alpha_{ij} = \gamma + \delta \ln P_{ij}, \quad (6)$$

where P_{ij} is the preference of the interaction between atoms i and j , γ and δ are constants to be determined by fitting. Extremely low preference scores indicate that few such atom pairs are observed in the database. Because some pairs are unobserved, a maximum repulsion energy is selected for these cases. Thus, for preferences lower than 0.1, a maximum repulsive value of α_{ij} was set to α_{min} . Implicit in this assignment of free energy contributions from observations on pairwise area burial is that the complexes used to construct the preferences form a complete set. Thus, preferences lower than the random case, $P_{ij} = 1.0$, indicate a relative avoidance between surface elements i and j , i.e., they are associated with a higher free energy. For atom pairs with a low frequency of occurrence, incompleteness of the data set may yield less reliable values for the effective repulsive α_{ij} coefficients. This would be most pronounced for pairs involving infrequently solvent-exposed carbon atoms like, C_γ , C_{amid} , C_{CNC} , C_{carb} , and C_{N^+} . The γ parameter in Equation 6 represents an average surface contribution that is independent of atom type. In Equation 5, a constant term, $\beta = 2.39$ kcal/mol, has been included that can be viewed as including a generic entropy penalty for the binary association of two biomolecules (Erickson, 1989; Finkelstein & Janin, 1989; Horton & Lewis, 1992; Murphy et al., 1994). This value corresponds to the cratic entropy term of Murphy et al. (1994). All other entropic and enthalpic contributions are subsumed into the effective binding parameter α_{ij} by the implicit averaging procedure in the fitting of the binding data.

The unknown parameters, γ and δ , are obtained by minimizing $|\Delta G_{\text{pred}} - \Delta G_{\text{exp}}|$ for a set of enzyme-inhibitor complexes whose free energy of binding is known. The experimental free energy values are derived from measured values of the association constant, K_a , of Equation 1 in the standard state, because $\Delta G_{\text{exp}} = RT \ln K_a$. Although some reliable measurements of the association constant for receptor/ligand formation are available, this is generally not the case for subunit association of protein monomers. Indeed, most association constants have not been reported for the receptor-ligand complexes in Table 1. In addition to the problem of the accuracy of binding data and its relationship to the conditions in the crystal structure, a relatively small set of enzyme-inhibitor structures and free energies is used in order to establish the parameters in the free energy model of Equation 5. In order to minimize the effect of a grossly altered unbound state of the inhibitor, only complexes that are expected to maintain their overall structure, whether bound or free, have been included. The crystal structures used in the estimation of the free energy model together with their estimated and measured free energy values, are given in Table 3. These enzyme complexes have similar buried surface areas, yet their binding energy ranges between -12 and -18 kcal/mol, a factor of about 10^5 in binding affinity. The RMS deviation of the linear fit is 1.5 kcal/mol ($R = 0.74$). The fitted model parameters are given in Table 4A, and the resultant binding coefficients α_{ij} for all designated atom types are in Table 4B. Even though the fitted data set is small, it encompasses variation in inhibitor size as well as representing interactions among many enzymes within the hydrolase enzyme family. Although the fit is not perfect, it establishes the important connection between surface atom-atom

Table 3. Comparison of calculated and experimental free energies of binding

PDB code	Buried area $\text{\AA}^2$	$N_{\text{Enz}}^{\text{aa}}$	$N_{\text{Inh}}^{\text{aa}}$	ΔG_{pred} (kcal/mol)	ΔG_{exp} (kcal/mol)	Reference
1CHO	1,420	245	56	-12.2	-14.6	Empie and Laskowski (1982)
1TPA	1,420	223	58	-17.3	-17.8	Vincent et al. (1974)
2KAI	1,406	232	58	-14.8	-12.4	Chen and Bode (1983)
2PTC	1,400	223	58	-17.5	-18.1	Vincent and Lazdunski (1972)
2SEC	1,451	274	58	-13.7	-13.1	Ascenzi et al. (1988)
2TGP	1,400	229	58	-16.9	-18.2	Bode (1979)
2SNI	1,601	275	83	-16.1	-15	S. Dixon (pers. comm.)
4TPI	1,454	229	58	-15.9	-17.7	S. Dixon (pers. comm.)

preferences and protein-inhibitor free energy interactions. The deviations between predicted and experimental values of binding energy can be attributed in large part to inherent limitations in the crystal data set used to determine atom-atom preferences. In spite of the limitations, as will be seen later, the values defined are interesting and highly informative as to the most important atomic surface interactions.

Interpretation of atom-atom binding coefficients

The physical meaning of the atom-atom preferences can be investigated. The process described in Equation 1 can be favorable if the expulsion of water and the subsequent association between the two atom types are attractive ($\alpha_{ij} > 0$) or unfavorable ($\alpha_{ij} < 0$). The complete set of these interaction energies appears in Table 4B. In the discussion to follow, portions of this table are reproduced in the text for easier reference.

Hydrophobic interactions

The binding parameter for aliphatic carbon-carbon self-interactions, $C_{\text{aliph}}C_{\text{aliph}}$ in Table 4B, is 23 cal/mol/ $\text{\AA}^2$, i.e., the removal of such a hydrophobic surface from solvent exposure yields a gain in free energy of 23 cal/mol/ $\text{\AA}^2$ (Janin & Chothia, 1990). An average of all interactions among nonpolar carbons, C_{α} , C_{β} , C_{aliph} , and C_{arom} types gives a hydrophobic binding parameter of 12 cal/mol/ $\text{\AA}^2$.

α_{ij}	C_{α}	C_{β}	C_{aliph}	C_{arom}
C_{α}	5	4	10	0
C_{β}		13	10	18
C_{aliph}			23	14
C_{arom}				22

It is thus clear that within this relatively homogeneous group of carbon-carbon interactions a large dispersion is observed in their individual interaction strengths, ranging from 0 to 23 cal/mol/ $\text{\AA}^2$. Notably favorable are $C_{\text{aliph}}C_{\text{aliph}}$ and $C_{\text{arom}}C_{\text{arom}}$ pairs. The strong segregation between the two categories is note-

worthy, i.e., the entry for $C_{\text{aliph}}C_{\text{arom}}$ is about half as large as the self-interacting pairs.

Because of the bonds of certain atom types, such as the carbonyl-carbon in the peptide group, C_{pep} , an effective attraction between their surface elements is not generally expected. The hydrogen bond considerations between peptide groups result in specific alignments of atoms, effectively shielding the C_{pep} atoms from one another; hence, their unfavorable α_{ij} value of -23 cal/mol/ $\text{\AA}^2$. Such interdependence in the behavior of atom types is also evident among interactions between C_{aliph} on one hand and C_{amid} and C_{carb} on the other hand. The carbon atom in the carboxyl group is part of a partially charged group and, as such, it is less likely to be placed near a hydrophobic group. Consequently, the binding coefficient for the $C_{\text{aliph}}C_{\text{amid}}$ pair, where such delocalization is slightly weaker, is larger than for the $C_{\text{aliph}}C_{\text{carb}}$ pair. Further, the strong repulsion between C_{carb} atom types can be rationalized by the possible strong hydrogen bonded complex that can be formed between carboxyl groups, thus orienting the carbon atoms away from each other and leaving little exposed surface area between them, i.e., an effective repulsion.

A check of these results comes from estimating the hydrophobic free energy for the association of two methane molecules in liquid water. The removal of the solvent-accessible surface associated with the binding of two methane molecules to form a dimer at a separation of 4 $\text{\AA}$ is about 65 $\text{\AA}^2$ (Jorgensen et al., 1988). Nominally, the methane-methane molecule interaction should correspond to two aliphatic atom types, i.e., C_{aliph} , which would yield an attractive component of $6.5 \cdot 23$ cal/mol ~ -1.5 kcal/mol. Taken together with the generic entropic penalty, β , a net repulsive free energy of association of $\sim +0.9$ kcal/mol is calculated. This result is consistent with molecular simulations of the hydrophobic association of methane molecules (Blokzijl & Engberts, 1993) that also find a favorable dimer association energy, but with an entropy term that results in a net repulsive free energy (Arakawa et al., 1979).

Another system of interest among hydrophobic attractions is the aromatic self-interaction typified by the phenyl-phenyl interaction between two benzene rings. Crystallographic data find that stacked and perpendicular modes of interaction are found for pairs of phenyl rings (Burley & Petsko, 1985). Molecular simulations of the potential of mean force between two benzene molecules in water (Ravishanker & Beveridge, 1985; Jorgensen & Severance, 1990; Linse, 1992, 1993) indicate a lower free energy for the perpendicular dimer. Naturally, in a formu-

[illegible]

direct binding between $O_{\text{pep}}N_{\text{pep}}$ pairs contributes 22 cal/mol/Å², a value almost equal to that for hydrophobic interactions. However, considerations of all possible pairwise combinations of C_{pep} , N_{pep} , and O_{pep} pairs for the six possible interactions among these atom types gives an "average" interaction of only -2 cal/mol/Å², i.e., there is a strong additional orientation constraint that must be taken into account. The individual peptide hydrogen bond parameters within the (C_{pep} , N_{pep} , O_{pep}) group,

α_{ij}	C_{pep}	N_{pep}	O_{pep}
C_{pep}	-23	-8	-7
N_{pep}		3	22
O_{pep}			4

strongly reflect the geometry of the functional group, e.g., the alignment of the amide hydrogen and carbonyl oxygen to form a hydrogen bond, together with the other attached backbone atoms, effectively prevents interaction between carbonyl carbons ($C_{\text{pep}}C_{\text{pep}} = -23$ cal/mol/Å²). For amide-amide interactions among the atom types C_{amid} , N_{amid} , and O_{amid} ,

α_{ij}	C_{amid}	N_{amid}	O_{amid}
C_{amid}	15	22	13
N_{amid}		22	50
O_{amid}			28

a consistent strong attraction is found. These hydrogen bond interactions are found to be at least as important as hydrophobic interactions.

Charge interactions

The fully charged atom groups in our scheme correspond to the N_K , N_R , and O_{carb} types in Table 2. Comparing the lysine and arginine terminal group interactions finds that for their interactions with hydrophobic atom types (C_{aliph} and C_{arom}), arginine exhibits a relatively more favorable binding contribution in its interactions with aromatic carbons.

α_{ij}	N_K	N_R
C_{aliph}	0	-2
C_{arom}	-12	18

Mrabet et al. (1992) showed that Lys → Arg mutations in thermostable D-xylase isomerase (IX) from *Actinoplanes missouriensis* imparted a significant stability under increasing heat conditions. Quaternary ammonium compounds also exhibit a preference for aromatic rings (Gao et al., 1993; Ripoll et al., 1993). The occur-

rence of aromatic phenyl interactions with the $R-N^+(CH_3)_3$ group has also been noted in crystal structures of small molecules from the Cambridge Structural Database (Verdonk et al., 1993).

The self-interaction between lysine N_K -atoms are strongly repulsive, whereas interactions between the N_K and the arginine N_R -atom group are actually favorable.

α_{ij}	N_K	N_R
N_K	α_{min}	12
N_R		-19

The interaction between the like-charged N_R 's is, as expected, also repulsive. The N_K repulsion stems partly from electrostatic repulsion by like charges, but an alternative viewpoint proposes that the main effective repulsion arises from the difference in N_K hydrogen bonding with water in the unbound versus the bound state. The charge on an N_K atom is more localized compared to the position of the charge in an arginine residue. Thus, the lysine can develop strong charge-dipole interactions with the surrounding waters in the free state, possibly stronger relative to a water-water hydrogen bond. This would mean that there is a strong driving force to maintain the favorable solvation of the lysine residue, to give an effective intermolecular repulsion between lysine residues. For arginine residues, the electrostatic repulsive term is similar, but the interaction capabilities of the N_R atoms to solvent waters are weaker than for the N_K atom. Consequently, the unbound state may be stabilized less well and the result is the less repulsive N_RN_R interaction. Arginine-arginine short-range interactions have also been observed and thought to contribute a stabilizing effect in proteins (Magalhaes et al., 1994). The apparent attraction between N_K and N_R atom types cannot be easily explained. Water bridging or solvent-mediated interactions in the bound state may be phenomena important in this case (Ben-Naim et al., 1989a, 1989b, 1990; Ben-Naim, 1990; Durell et al., 1994; Jungwirth & Zahradnik, 1994; Wallqvist & Covell, 1995). Recall that the distance used for data collection is large enough to admit this possibility. Likewise, the same charge pair, $O_{\text{carb}}O_{\text{carb}}$ has a favorable interaction of 21 cal/mol/Å². Stabilization by oppositely charged free ions can also contribute to this favorable interaction (Roseman, 1988). The attraction between aliphatic groups on the flexible carbon chains of lysine and arginine contributes to an apparent effective attraction between the N_R and N_K atom surfaces.

The preference between oppositely charged groups is always attractive and quite strong,

α_{ij}	C_{carb}	O_{carb}
N_K	61	71
N_R	22	36

with N_K pair interactions stronger than the corresponding N_R values. The more localized charge on the lysine residue presum-

ably interacts more favorably with the carboxyl group, but also the possibility for orienting the side-chain terminus on the arginine residue to maximize the N_R carboxyl contact may exist.

The interactions of the terminal groups of arginine and lysine side chains with an amide (C_{amid} , N_{amid} , O_{amid}) exhibit strong amide orientational effects for lysine–amide interactions, with a consistently larger overall attraction shown for arginine–amide atom pairs.

α_{ij}	C_{amid}	N_{amid}	O_{amid}
N_K	–21	–21	11
N_R	1	29	29

As in the previous case, a direct electrostatic effect as well as the effect of differences in hydrogen bonding strength between the N_K and the N_R groups in the unbound state may contribute to these results. The direct interaction between the side-chain termini and the amide nitrogen is mainly electrostatic in character and ought to be similar for the positively charged arginine and lysine side chains, contrary to the result in the table above. The lysine side chain with its single positively charged atom may be capable of making stronger charge–water pairs with the surrounding waters than water–water interactions. Under these conditions, the solvated unbound state is favored and the resultant effect is a repulsive contribution for the free energy of binding between the N_K atom and an amide group.

Alcohols and thiols

The alcohol-containing groups on serine, threonine, and tyrosine as well as the thiol groups on cysteine and methionine, are less polar than the ionized side chains. Yet, they can form hydrogen bonds and contribute significantly to the stability of a receptor–ligand complex. Cysteine residues do not normally form covalent bonds between the binding species, and here they are only considered in the context of noncovalent intermolecular binding.

Connelly et al. (1994) have noted that the mutation Y82F in the FK506 binding protein exhibits a slight destabilizing effect on the protein–ligand complex. The ligand binds through an interaction between a carbonyl group and the tyrosine hydroxyl group. From Table 4B, we see that the $O_{amid}O_Y$ binding parameter is 16 cal/mol/Å², whereas the substitution of a phenyl group would correspond to the $O_{amid}C_{arom}$ interaction of 7 cal/mol/Å². Thus, the observed destabilization is accounted for qualitatively by the preference-based free energy. The self-interactions of O_{alc} , O_Y , and S all show attractive interactions

	$O_{alc}O_{alc}$	O_YO_Y	SS
α_{ij}	22	38	22

that are comparable in strength to a hydrophobic $C_{aliph}C_{aliph}$ contact. The direct interaction, i.e., the potential energy contribution, between these weakly polar hydrogen bonding groups will only contribute a small amount to the free energy of bind-

ing. The bulkier tyrosine side chain is likely to be an even less efficient hydrogen bond former with water and the removal of an O_Y atom gains 38 cal/mol/Å². Obviously, the attachment of the –OH group to a specific side chain contributes to its interaction strength. The $O_{alc}O_Y$ binding parameter has a repulsive value of –13 cal/mol/Å², reflecting the influence of the rest of the side chain on proximate atom types.

The contact between alcohol groups and the peptide group is slightly attractive,

α_{ij}	C_{pep}	N_{pep}	O_{pep}
O_{alc}	1	16	2
O_Y	2	–8	11

whereas the amide contact is strongly favored.

α_{ij}	C_{amid}	N_{amid}	O_{amid}
O_{alc}	24	29	27
O_Y	16	44	16

The positively charged atom types N_K and N_R exhibit an effective attraction with the O_{alc} and O_Y , respectively, but the other cases of N_KO_Y and N_RO_{alc} appear to be neutral.

α_{ij}	N_K	N_R
O_{alc}	46	–7
O_Y	4	25
S	–20	–16

The interactions with the thiols are both repulsive. For the alcohols, these trends in interactions are followed closely by the atom types attached to the hydroxyl group. Aromatic groups in the tyrosine ring repel the lysine nitrogen by –12 cal/mol/Å², whereas they attract the arginine nitrogens at a strength of 18 cal/mol/Å².

“Hydrophobicity” scale for interfacial burial

The values in Table 4B can be used to construct an average interfacial binding parameter for each atom type i , $\bar{\alpha}_i$. This statistic is useful for probing the surface of a target molecule in the absence of a designed ligand, to locate candidate binding sites. Its value can be calculated for all the binding parameters of type i with all other atom types to form a weighted average of atoms present at protein/protein interfaces:

$$\bar{\alpha}_i = \frac{\sum_{j=1}^{N_{atom}} \alpha_{ij} \alpha_{ij}}{\sum_{j=1}^{N_{atom}} \alpha_j} \quad (7)$$

where N_{atom} is the number of atom types defined above (in *Selection of reduced atom types*), and $\mathcal{A}_j$ is the total area for each atom type present among the interfaces specified in Table 1. Values are given in Table 5. The four most favorable interacting atom types are all carbons, C_{aliph} , C_β , C_{N^+} , and C_{arom} . Although the interfacial binding surfaces studied have a preponderance of hydrophobic surface (Covell, 1994), the hydrogen bonding and charged groups contribute significantly to the free energy of binding.

The free energy of transfer between water and an interface can be obtained by using the interfacial-binding parameter together with the calculated solvent-exposed surface area of the Gly-X-Gly tripeptide (Eisenberg et al., 1989) for each amino acid X:

$$\Delta G_{interface/water}^{pref(\bar{\alpha}_i)} = \sum_{k=1}^{N_i} \bar{\alpha}_k \mathcal{A}_k^i \quad (8)$$

where the summation runs over all N_i atom types on the i th amino acid, and $\mathcal{A}_k^i$ is the solvent-exposed surface area of atom type k on the same amino acid. These free energies are compared in Table 6 and Figure 2 with the values for amino acid transfer in the octanol-water system (Fauchere & Pliska, 1983).

Although the two processes are very different physically, and removal of an amino acid from water into octanol may not be considered a particularly good model system for protein interfaces, there is a weak resemblance to the transfer free energies for hydrophobic residues. On the other hand, the correlation with the hydrophilic residues shows no similarity, and indeed most hydrophilic groups have a propensity for being at the interfacial contact surface instead of in solution. The qualitative

Table 5. Averaged "interfacial" binding coefficients

Type	$\mathcal{A}_i$ ($\text{\AA}^2$)	α_i (cal/mol/ $\text{\AA}^2$)
C_{pep}	1,556	2.5
C_α	3,763	5.3
C_β	7,404	9.4
C_{aliph}	14,193	11.5
C_{arom}	6,178	9.1
C_{amid}	223	4.7
C_{carb}	254	-1.2
C_{CNC}	234	-3.0
C_{N^+}	598	9.4
C_Y	150	-3.3
N_{pep}	3,262	2.5
N_K	733	6.1
N_{arom}	779	2.4
N_{amid}	1,116	9.0
N_R	1,678	8.7
O_{pep}	9,069	7.2
O_{alc}	1,770	6.4
O_{amid}	848	7.8
O_{carb}	1,991	2.3
O_Y	990	6.7
S	1,550	8.7

Table 6. "Hydrophobicity"-scale of interfacial burial relative to Gly

Amino acid	Type	$\Delta G_{interface/water}^{pref(\bar{\alpha}_i)}$ (kcal/mol)	$\Delta G_{octanol/water}^{exp}$ (kcal/mol)
Gly	—	(0)	(0)
Ala	Hydrophobic	0.3	0.4
His	Hydrophobic	0.9	0.2
Ile	Hydrophobic	1.4	2.5
Leu	Hydrophobic	1.4	2.3
Met	Hydrophobic	1.5	1.7
Phe	Hydrophobic	1.4	2.4
Pro	Hydrophobic	0.9	1.0
Thr	Hydrophobic	0.8	0.4
Trp	Hydrophobic	1.5	3.1
Tyr	Hydrophobic	1.3	1.3
Val	Hydrophobic	1.1	1.7
Cys ^a	—	0.2	1.3
Arg	Hydrophilic	1.7	-1.4
Asn	Hydrophilic	0.7	-0.8
Asp	Hydrophilic	0.1	-1.1
Gln	Hydrophilic	1.0	-0.3
Glu	Hydrophilic	0.5	-0.9
Lys	Hydrophilic	1.4	-1.4
Ser	Hydrophilic	0.4	-0.5

^a In the calculated, preference-based, transfer free energy, only non-covalently paired cysteines are included.

agreement between the hydrophobic residues for the transfer from octanol or the interfacial region to water indicates that the hydrophobic interaction is a very general effect in many diverse environments. Hydrophobic residues that lie below the line of slope 1 in Figure 2 indicate amino acids that have a preference for the interior of a protein (taken to be equal to the *n*-octanol

Fig. 2. Transfer free energy of the amino acids derived from the tripeptide Gly-X-Gly for the water/*n*-octanol (Fauchere & Pliska, 1983) versus the water/interface system. Values are given as differences with respect to the Gly values. Amino acids in the region marked "INTERIOR" have a stronger preference for being completely buried inside a protein than residing at a protein-protein interface.

environment), whereas those above it have a preference for the interface.

The role of the hydrophilic residues is more complex, as they can both be utilized by the enzyme-receptor complex for mutual attraction as well as forcing the possibility of a highly specific pairing. This suggests that the "hydrophobicity" scale in Table 6 may be more relevant to problems involving the association of protein complexes, or protein subunits, than to values measured for amino acids in the octanol-water system.

HIV-1 protease inhibitors

Antiviral strategies aimed at interfering with the activity of human immunodeficiency virus type 1 protease enzyme (HIV-1-PR) have generally focused on the design of small, tightly bound inhibitors (Huff, 1991). Crystallographic studies have provided the bulk of information about the nature of inhibitor interactions with HIV-1-PR (Fitzgerald et al., 1990; Wlodawer & Erickson, 1993). These data find that most inhibitors bind HIV-1-PR in an extended conformation forming nearly symmetric interactions with both monomers of the active dimer molecule. The approximate twofold crystallographic symmetry of the dimer contributes to these interactions by forming nearly symmetric targets involving the active site and flap regions of HIV-1-PR. These interactions result in an almost complete burial of most HIV-1-PR inhibitors. Conclusions about the energetic contributions to the total binding strength of these inhibitors, based largely on applications of conventional modeling tools (Hutchins & Greer, 1991; Appelt, 1993; Guida, 1994; Chen & Tropsha, 1995; Holloway et al., 1995), find that hydrogen bonding patterns and hydrophobic interactions are frequent components of inhibitor binding (Appelt, 1993; Wlodawer & Erickson, 1993; Gustchina et al., 1994). The extent and importance of these interactions have, however, been difficult to characterize. Furthermore, the relative occurrences of hydrogen bonds and hydrophobic interactions vary depending on the type of inhibitor (Marshall, 1992; Lam et al., 1994).

The atom-atom surface area preferences presented here contain a range of different types of interactions and thus can be applied directly to determine the relative contributions of each type of interaction to total binding strength. The results pre-

sented here are based on an analysis of the 10 inhibitor-protease complexes given in Table 7. These inhibitors are members of two distinct classes; one consists of transition-state inhibitors (TSI) (Dreyer et al., 1992; Oprea et al., 1994) and the other of a class of pseudo C_2 -symmetrical inhibitors. Examples of the former class of inhibitors include a reduced amide- (MVT-101 [Miller, 1989]), and statine (acetylpepstatin [Fitzgerald et al., 1990]), hydroxyethylene (U85548e, Ro8558), dihydroxyethylene (U75875 [Thanki et al., 1992]) or hydroxyethylamine- (JG-365 [Swain et al., 1990], Ro8959-R) containing inhibitors. The compounds A74704 (Erickson et al., 1990) and BMS182,193 are examples of C_2 -symmetric inhibitors, whereas PD134922 is based on the design of existing renin inhibitors.³

Buried surfaces

Table 7 summarizes the results for surface area burial. An average of $1,389 \pm 110 \text{ \AA}^2$ of total surface is buried for this set of inhibitors. The range of buried surface areas is narrow, extending from a low of $1,197 \text{ \AA}^2$ for BMS 182,193 to a maximum of $1,543 \text{ \AA}^2$ for JG-365. Total buried surface per HIV-1-PR residue for all inhibitors is shown in Figure 3 where a histogram of buried areas for all inhibitors is plotted, with the results for monomer A and monomer B appearing below and above the zero axis, respectively.⁴ Only 20 residues on each monomer chain account for all the surface buried by this set of inhibitors:

R8 L23 D25 G27 A28 D29 D30 V32 K45 M46
I47 G48 G49 I50 F53 L76 T80 P81 V82 I84

³ Coordinates of Ro85548e, Ro8959-R, BMS182,193, and PD134922 were kindly supplied by Dr. Alex Wlodawer (NCI-FCRDC). Coordinates for the other complexes are available in the Brookhaven Protein Data Bank (Bernstein et al., 1977). In the following analysis, reference to the protease dimer will be made using the convention of numbering the residues of monomer A from 1 to 9 and monomer B from 1' to 99' (Hutchins & Greer, 1991).

⁴ The results for each histogram are arbitrarily ordered from left to right as follows: MVT-101, acetyl-pepstatin, JG-365, U85548e, A74704, PD 134922, BMS 182193, Ro-8558, Ro-8959, U75875.

Table 7. Characteristics of HIV-1 protease inhibitors

Inhibitor	Heavy atoms	TSI ^a	Buried $\text{\AA}^2$	Protease atoms in contact	Reference ^b
MVT-101	53	No	1,416	222	Miller (1989) (4HVP)
Acetyl-pepstatin	45	Yes	1,366	222	Fitzgerald et al. (1990) (5HVP)
JG-365	61	Yes	1,543	239	Swain et al. (1990) (7HVP)
U-85548E	61	Yes	1,486	236	Jaskolski et al. (1991) (8HVP)
A-74704	54	No	1,440	196	Erickson et al. (1990) (9HVP)
PD 134922	50	No	1,283	187	Wlodawer and Erickson (1993)
BMS 182,193	39	No	1,197	135	Wlodawer and Erickson (1993)
Ro-8558	42	No	1,319	168	Wlodawer and Erickson (1993)
Ro-8959	49	Yes	1,325	153	Wlodawer and Erickson (1993)
U75875	59	Yes	1,513	207	Thanki et al. (1992) (1HIV)

^a TSI, transition state isostere.

^b PDB identifier in parentheses.

Fig. 3. HIV-1-PR surface buried by the complexed inhibitor. Abscissa identifies HIV-1-PR residues having any portion of their surfaces buried. At each residue, a histogram of the buried surface for all 10 complexes is listed. Histograms are ordered arbitrarily from left to right as follows: MVT-101, acetyl-pepstatin, JG-365, U85548e, A74704, PD 134922, BMS 182193, Ro-8558, Ro-8959, U75875. Vertical dotted lines at each residue position indicate the first member of the list. Values for the A and B chains of HIV-1-PR appear below and above the zero axis, respectively.

The buried surfaces of these residues vary from zero to nearly $50 \text{ \AA}^2$, with all but five of these residues (K45, M46, F53, L76, and T80) being buried consistently by this group of inhibitors. The amount of buried surface area falls into several categories. Among the consistently buried residues,⁵ the greatest burial (area $> 20 \text{ \AA}^2$) is accounted for by

R8/8' G27/27' D29/29' D30/30' G48/48' I50/50'

with intermediate burial (area $> 10 \text{ \AA}^2$) by

D25/25' A28/28' I47/47' G49/49' P81/81' V82/82' I84/84',

and the least burial (area $\leq 5 \text{ \AA}^2$) by

L23/23' V32/32'.

Within the group of five residues not consistently buried by these inhibitors, variations in their buried surfaces exceed $15 \text{ \AA}^2$, depending on the inhibitor bound in the complex.

Free energy decomposition

The contribution of each residue's buried surface to the calculated total binding strength (ΔG) based on the present surface interaction parameterization is summarized in Figure 4 for monomer A (below the zero axis) and monomer B (above the zero axis). Fifteen of the 20 HIV-1-PR residues making inhibitor contact are found to interact consistently for the complete set of inhibitors. Isoleucine at position 50 in each monomer (I50/50') contributes the most to the total binding strength, with

values of ΔG per residue ranging from -0.6 to -0.2 kcal/mol . Weaker contributions to ΔG , in the range of -0.2 to $-0.1 \text{ kcal/mol/residue}$, are also found for

R8/8' G27/27' I47/47' G48/48' P81/81' V82/82' I84/84'

with still weaker (ΔG from -0.1 to $-0.05 \text{ kcal/mol/residue}$) contributions observed for

L23/23' D25/25' A28/28' D29/29' D30/30' V32/32' G49/49'.

Within this set of inhibitors, variable but significant interaction strengths are found at five residue positions,

K45/45' M46/46' F53/53' L76/76' T80/80',

residues with atoms located from 8 to $20 \text{ \AA}$ away from the active site aspartic acid (D25/25'). Inhibitor interactions with these latter five residues usually occur with only one monomer of the HIV-1-PR dimer and thus break the symmetry of interactions found for other HIV-1-PR contacts. Contributions of these interactions to the total binding strength are, however, significant and, in some instances, approach or even exceed the interaction strength found for I50/50'. A limited subset of atoms in these residues contributes the greatest to the total binding strength.

Contacts between inhibitors and HIV-1-PR can be further analyzed by identifying the individual atoms within each HIV-1-PR residue that make the largest contribution to the total interaction energy. Only 59 HIV-1-PR atoms (23 backbone and 36 side-chain) have favorable inhibitor interactions. Eleven of the 23 backbone atoms on each monomer consistently participate either in the formation of hydrogen bonds:

G27/27':O D29/29':N D30/30':N, O G48/48':N, O I50/50':N

⁵ The designation of Appelt (1993) (e.g., I50/50') will be used throughout to indicate equivalent monomer residues in the HIV-1-PR dimer.

Fig. 4. Calculated binding energy per residue for all HIV-1-PR residues making inhibitor contact. At each residue a histogram of binding energy per residue for all 10 complexes is listed. Histograms are ordered as in Figure 3.

or hydrophobic interactions:

Inhibitor interactions with these backbone atoms are generally symmetric to both monomer chains. The greatest interaction energies in this set are due to hydrogen bonding by the carbonyl oxygens of G27/27' and G48/48' and the amide nitrogen of I50/50', with consistent but weaker interactions found with G27/27':C, A28/28':C $^{\alpha}$, and A49/49':C, C $^{\alpha}$. Contacts with the remaining 12 HIV-1-PR backbone atoms depend on the specific inhibitor bound; some inhibitors (U85548e, A74704, and BMS 182193) make limited contact with G27/27':C $^{\alpha}$, D30/30':C, and V82/82':O, whereas most inhibitors make contact with a majority of the backbone atoms among the dozen atoms in this set. Nearly all of these latter interactions are asymmetric in that they are limited to backbone atoms on only one monomer of the HIV-1-PR dimer.

Within the set of 36 side chain atoms, 19 interaction sites are consistently found for this set of inhibitors. These consist of

where the underlined entries identify the subset of atoms that consistently provide the strongest interactions for all inhibitors. The remaining side-chain atoms making favorable inhibitor contacts involve 10 residues on each monomer:

These results suggest that inhibitor interactions are characterized by a small set of HIV-1-PR contacts that make large contributions to the total binding energy and a set of more variable interactions that can contribute significant additional stabilization to the complex. In some instances, the strongest contact is found for only one inhibitor in the set. Examples of this exist for R8:C $^{\gamma}$ in Ro-8558, D25:C $^{\gamma}$ in A74704, K45:C $^{\gamma}$ in PD134922, and F53:C $^{\delta}$ in U75875. Variable interactions of this type contribute to the range of total binding strengths for these inhibitors and have been the focus of efforts to optimize inhibitor interactions to many of these atoms.

Table 8A presents the interaction energies for each inhibitor as a function of the location of inhibitor subsites. These subsites are designated P5', P4'...P1', P1...P5, according to the convention of Schecter and Burger (1967), where the boundary between primed and unprimed inhibitor subsites is defined by the location of the nonscissile peptide bond found in the central portion of most of these inhibitors (indicated in this table as the column labeled "Linkage"). Within each inhibitor a relatively uniform distribution of binding strengths is observed for all subsites. This suggests that each inhibitor makes efficient use of all of its constituent atoms to bind the target. In some instances, the shorter inhibitors lacking termini subsites (e.g., P3/3', P4/4', P5/5') have a general tendency for the subsites flanking the nonscissile bond to make a greater contribution to the total binding energy when compared to other inhibitors. The correspondence between measured and predicted binding strengths for these inhibitors is generally good, with the deviations from their measured values on average less than 1.5 kcal/mol (cf. Table 8B). The largest deviations are found for MVT-101 and JG-365. Flexible molecules of partly hydrophobic character, such as these compounds with their several alkane side chains, may very well adopt, if not definite conformations in the solvent, some folded characteristics. If such were the case, one would need to contend with intramolecular binding as well as rotational degrees of freedom in order to characterize prop-

Table 8A. Free energy details for HIV-1 protease inhibitor binding

Inhibitor	P5 ^a	P4	P3	P2	P1	Linkage	P1'	P2'	P3'
MVT-101			Ac-Thr -1.1	Ile -1.6	Nle -1.7	CH ₂ NH -0.1	Nle -1.3	Gln -1.8	Arg NH ₂ -1.0
Ac-Pepst			Ac-Val -1.2	Val -2.0	Sta -1.8	CH(OH)CH ₂ -0.1	CO-NH -1.3	Ala -0.8	Sta -2.4
JG-365		Ac-Ser -0.2	Leu -0.8	Asn -1.1	Phe -1.3	CH(OH)CH ₂ NH -0.1	Pro -1.1	Ile -1.9	Val-OMe -1.2
U-85548e	Val -0.2	Ser -0.4	Gln -1.3	Asn -1.3	Leu -1.5	CH(OH)CH ₂ 0.0	Val -1.6	Ile -1.5	Val -0.9
A-74704			Cbz -0.8	Val -1.9	Phe -1.4	CH(OH) 0.0	Phe -1.6	Val -2.1	Cbz -0.6
PD 134922		Mor -0.8	Phe -0.8	Bac -1.1	Hch -2.1	CH(OH)CH ₂ 0.0	Mva -2.0	Nle -2.1	
BMS 182193				Bms -2.6	Phe -1.7	CH(OH)CH ₂ -0.1	Phe -1.7	Bms -2.1	
Ro-8558				Boc -2.5	Cha -2.0	CH(OH)CH ₂ 0.0	Val -1.3	Ile -2.3	Epy -1.3
Ro-8959			Qun -1.7	Asn -1.2	Hph -1.7	CH(OH) 0.0	Diq -2.8	Ntb -2.1	
U75875			Noa -1.3	His -2.2	Cha -2.1	CH(OH)CH(OH) -0.1	Val -1.3	Ile -2.2	Amp -1.7

^a The convention of Schechter and Burger has been used by others to describe enzyme-specific subsites (S1, S1', S2, S2', etc.) and the corresponding side chains on inhibitors (P1, P1', P2, P2', etc.). These enzyme subsites are bounded by residue pairs from both monomers of HIV-1-PR and have been used to identify important HIV-1-PR targets for the design of more tightly binding inhibitors. Boundaries of the S1/S1' subsites are comprised of atoms from D25/25', L23/23', I50/50', I84/84', T80/80', and P81/81'. The S2/S2' subsites are bounded by atoms from V32/32', I50/50', I47/47', L76/76', A28/28', and D30/30'. The S3/S3' subsites are bounded by L23/23', V82/82', P81/81', and R8/8'. Abbreviations for unusual amino acids follow those of Wlodawer and Erickson (1993).

erly the unbound state. Such considerations are beyond the scope of the present work.

The relative location of the most favorable interactions can be visualized by mapping the binding energy of each inhibitor's atoms onto the joint surface for all inhibitors (cf. Fig. 5). In this figure the position of each inhibitor is determined by matching only the HIV-1-PR backbone coordinates of both monomer chains to those of the acetyl-pepstatin complex.⁶ This view allows visualization of the energetically important contact sites on the complete set of inhibitors, referenced to the target protease dimer. Surfaces are colored spectrally to indicate binding strength: red, strongest; green, intermediate; blue, weakest. HIV-1-PR residues forming various subsites are labeled at the perimeter of these inhibitors. The borders of the S1/S1' subsites are formed by I23/23', D25/25', G27/27', I50/50', P81/81', and I84/84'. The S2/S2' subsite boundary is determined by A28/28', D29/29', D30/30', V32/32', and I47/47'. The S3/S3' subsites are bordered by R8/8' and V82/82'. Four strong interactions with the HIV-1-PR subsites are located symmetrically about the central portion of these inhibitors. These interactions involve subsites S1/S1' and S2/S2' and correspond to hydrogen bonding as well as hydrophobic interactions with the HIV-1-PR dimer. Intermediate strength interactions are found with portions of the S2/S2' and S3/S3' subsites. These interactions include contacts with D29/29' and D30/30' as well as interactions with R8/8':N^ε,

N^γ at the top and bottom of these views. The weakest HIV-1-PR interactions are not confined to a specific location, as is generally true for the stronger interactions, but are scattered throughout the length of these inhibitors. These weaker sites identify inhibitor regions where design improvements to existing inhibitors might be considered.

Inhibitor binding modes

The modes of inhibitor binding to HIV-1-PR have been the subject of many published reports (Wlodawer & Erickson, 1993).

Table 8B. Calculated and experimental free energies of HIV-1 protease inhibitor binding

Inhibitor	ΔG_{pred} (kcal/mol)	K_i (nM)	ΔG_{exp} (kcal/mol)	K_i (nM)
MVT-101	-12.4	0.9	-8.4	780
Ac-P'stn	-11.2	7	-10.6	20
JG-365	-10.8	14	-12.6	0.7
U-85548e	-11.8	3	<12.4	<1
A-74704	-11.2	7	-11.5	4.5
PD 134922	-12.5	0.8	-10.7	15
BMS 182193	-10.6	19	-9.6	100
Ro-8558	-11.7	3	-13.1	0.3
Ro-8959	-12.7	0.6	—	—
U75875	-15.4	0.006	<12.4	<1

⁶ The maximum RMS coordinate deviations for matched C^α positions within this set of HIV-1-PR coordinates are less than 0.8 Å (Appelt, 1993).

Fig. 5. Composite view of atom surfaces for all 10 inhibitors. Inhibitor structures have been positioned by matching their HIV-1-PR backbone coordinates (see description in text). Inhibitor surfaces are colored spectrally from red to blue to indicate most and least favorable interactions with HIV-1-PR. Inhibitors are viewed from (A) the active site and (B) flap residues of HIV-1-PR. HIV-1-PR residues making close contact to these inhibitors appear as labels. Residue labels included in subsites S1/S1', S2/S2', and S3/S3' are colored white, purple, and black, respectively.

Inspection of the strongest interactions between inhibitor and HIV-1-PR atom pairs in this set of structures reveals various shared and unshared binding modes. Table 9 lists the contacts between each pair of inhibitor and HIV-1-PR atoms that contribute more than -0.1 kcal/mol to the total calculated binding energy. A mixture of hydrophobic and hydrophilic atom types appears in this table, depending on inhibitor type. Strong hydrophobic interactions are observed for all of these inhibitors with HIV-1-PR atoms:

Although the positions of inhibitor atoms making contact with each of these HIV-1-PR atoms are naturally clustered close to each other, the inhibitor residue types involved in these interactions generally involve a diverse set of hydrophobic atoms derived from branched, cyclic, and bicyclic side chains in natural and nonnatural amino acids.

Strong hydrophilic interactions between HIV-1-PR and these inhibitors involve hydrogen bonding groups in the active site region

and the flap regions

of HIV-1-PR. In most cases, inhibitor backbone acceptor atoms are involved in hydrogen bonding. Ro-8959 is an exception because hydrogen bonding to HIV-1-PR G48:O is provided by the N^δ atom of quinoline. An example of strong contact between hydrophilic and hydrophobic atoms appears (see Table 9) between R8/8':N^ε on HIV-1-PR and aromatic C^ε atoms supplied by inhibitors with phenol side chains. In some instances, alternative binding interactions are supplied from very different locations in an inhibitor. A specific example of one such dual binding mode involves the contacts of Ro-8959 and JG-365 with V32:C^γ. These two inhibitors share a transition state dipeptide ASN-PHEψ{CH(OH)CH₂N} and are expected to bind this portion of each inhibitor in a similar fashion. Because these inhibitors do not share chirality (JG-365 is an *S*-enantiomer and Ro-5989 is an *R*-enantiomer), it was speculated (Rich et al., 1991) that a binding mode allowing superposition of their hydroxyl groups would require local conformational rearrangements at another position in the inhibitor. These local rearrangements conserve the interaction with V32:C^γ through inhibitor contact with V:9:C^γ in the case of JG-365 and a tertiary-butyl group in the case of Ro-8959.

Considerations for rational inhibitor design

The optimal type for each inhibitor atom can be evaluated with respect to whether an alternative atom type at each position would improve interactions with HIV-1-PR atoms. This analysis simply scans the interaction energy of each inhibitor atom's

Table 9. Favorable inhibitor interactions to HIV-1-PR atoms^a

	MVT-101	Ac-Pepstatin	JG-365	U-85548e	A-74704	PD134922	BMS182193	Ro-8558	Ro-8959	U75875
R8:N ⁷		STA ₆ C ^β	VAL ₆ C	VAL ₁₀ O GLN ₃ O ^δ	*CBZ ₆ C ^ε COR ₄ C ^ε	*PHE ₂ C ^ε	PHE ₂ C ^ε PHE ₃ C ^ε	EPY ₆ C ^ε PHE ₂ C ^ε	*QUN ₁ C ^ε	AMP ₆ C ^ε *NOA ₁ C ^ε
L23:C ⁷	*NLE ₆ C ⁷	*STA ₆ C ⁷		*LOV ₅ C ⁷	COR ₃ C ⁷	*MVA ₆ C ⁷ *HCH ₄ C ⁷		EPY ₆ C ^ε	*DIQ ₄ C ⁷	*VAM ₄ C ⁷ *HCH ₃ C ⁷
G27:O	NLE ₄ N ARG ₇ N ⁷	STA ₄ N		ILE ₇ N		NLE ₆ N	BMS ₁ N		HPH ₃ N	
D29:N		VAL ₂ O ALA ₅ O						ILE ₅ O		NOA ₁ O ILE ₅ O
D29:O ^δ	ARG ₇ N ⁷					MOR ₁ N ⁷				
V32:C ⁷	*ILE ₃ C ⁷			ILE ₈ C ⁷ *ILE ₇ C ⁷	*VAL ₂ C ⁷ *VAL ₅ C ⁷	*NLE ₆ C ⁷	*BMS ₁ C ⁷	*ILE ₅ C ⁷	NTB ₆ C ⁷	*ILE ₅ C ⁷
I47:C ⁷	*ILE ₃ C ⁷			*ILEL ₇ C ⁷	*CBZ ₁ C ^ε	MOR ₁ C ^ε	*BMS ₄ C ⁷ *BMS ₁ C ⁷	*BOC ₁ C ⁷	NTB ₆ C	
G48:N	AC ^ε ₁ O	*STA ₆ O								
G48:O	ILE ₃ N ARG ₇ N		VAL ₉ N		VAL ₅ N	BAC ₃ N		EPY ₆ N	*ASN ₂ N *QUN ₁ ND	*AMP ₆ N
I50:C ⁷	*NLE ₄ C ⁷ *ILE ₃ C ⁷	*VAL ₃ C ⁷	*ILE ₈ C ⁷	*ILE ₇ C ⁷ *LOV ₅ C ⁷	*VAL ₅ C ⁷ COR ₄ C ^ε VAL ₂ C ⁷	*NLE ₆ C ⁷ *HCH ₄ C ⁷ *MVA ₅ C ⁷	*BMS ₁ C ⁷ *BMS ₄ C ⁷	*ILE ₅ C ⁷ *BOC ₁ C ⁷ *VAL ₄ C ⁷	*NTB ₆ C ⁷ *DIQ ₅ C ⁷	*ILE ₅ C ⁷ *HCH ₃ C ⁷
I50:N			PRO ₇ O		VAL ₂ O					
F53:C ^ε		*STA ₄ O	LEU ₃ C ⁷		*CBZ ₁ C ^ε	*PHE ₂ C ^ε			*QUN ₁ C ^ε	*NAO ₁ C ^ε
P81:C ^β				VAL ₈ C ⁷						AMP ₆ C ^ε
P81:C ⁷	NLE ₅ C *NLE ₄ C ⁷	*STA ₆ C ⁷	PRO ₇ C ⁷ PHE ₅ C ^ε	VAL ₈ C ⁷	COR ₄ C ^ε COR ₃ C	*MVA ₅ C ⁷ *HCH ₄ C ⁷	PHE ₂ C PHE ₃ C ^ε	*VAL ₄ C ⁷ PHE ₂ C ^ε	*DIQ ₅ C ⁷ *QUN ₁ C ^ε	AMP ₆ C ^ε *NAO ₁ C ^ε *HCH ₃ C ⁷
V82:C ⁷	*NLE ₅ C ⁷	*STA ₆ C ⁷		*LOV ₆ C ⁷	COR ₄ C ^ε	*MVA ₅ C ⁷	PHE ₂ C	*VAL ₄ C ⁷ EPY ₆ C ^ε	*DIQ ₄ C ⁷	*VAM ₄ C ⁷ AMP ₆ C ^ε
I84:C ⁷	*NLE ₅ C ⁷ *NLE ₄ C ⁷	*VAL ₃ C ⁷	PRO ₇ C ⁷ *ILE ₈ C ⁷	*LOV ₆ C ⁷ *LOV ₅ C ⁷	*VAL ₂ C ⁷ *VAL ₅ C ⁷	*NLE ₆ C ⁷ *HCH ₄ C ⁷	*BMS ₁ C ⁷	*BOC ₁ C ⁷ *VAL ₄ C ⁷ *ILE ₅ C ⁷	*DIQ ₅ C ⁷ *DIQ ₄ C ⁷ HPH ₃ C ^ε *NTB ₆ C ⁷	*ILE ₅ C ⁷ *HCH ₃ C ⁷

^a Cell contents specify inhibitor atom(s) with the most favorable interactions to the HIV-1-PR atoms listed in the leftmost vertical column. Inhibitor names are listed horizontally across the top of the table. HIV-1-PR atoms are listed as pairs to indicate interactions with monomer A (top row of each pair) and B (bottom row of each pair). Entries prefixed with an asterisk identify the most optimal atom for a given inhibitor position. Abbreviations of amino acids follow those of Wlodawer and Erickson (1993).

position as if it were "occupied" by all possible atom types. Atom positions where a more favorable binding energy cannot be found, compared to the native atom, identify optimally occupied sites. Alternatively, atom positions where large improvements in binding strength are found identify candidate positions for placing alternative atom types.

This analysis finds that the optimal atom positions are restricted to six regions of interaction with HIV-1-PR that are nearly symmetrically located above and below the central portion of each inhibitor. The locations of these regions are displayed in Figure 6, where the combined surfaces for all 10 inhibitors are displayed in orientations identical to those used previously in Figure 5. Red colored surfaces identify optimal inhibitor atom positions that have interactions with atoms in HIV-1-PR subsites S1/S1' and S2/S2'. The majority of optimal atoms in these regions correspond to hydrophobic atom types. Those atoms contributing the strongest possible interactions appear with asterisks in Table 9. The white colored surfaces, which can be seen only in Figure 6B, identify inhibitor atoms involved in interactions with only I50/50'. Both hydrophilic and hydrophobic optimal atom types are found in this region (cf. Table 9). Taken together, these optimal interaction sites identify a common design feature of these inhibitors, which attempts to enhance interactions to key atoms in HIV-1-PR subsites S1/S1' and S2/S2'.

Nonoptimal atom positions, shown in yellow, are consistently found at the inhibitor termini. These positions are occupied by hydrophobic and hydrophilic atom types that are predicted to have more favorable interactions if replaced by atom types hav-

ing stronger hydrophilic or hydrophobic features. Such replacements are intended to enhance interactions with D29:N and D30:O^δ, as well as backbone atoms in G27, A28, and G48 in both monomer chains. These interactions are confined to residues at the inhibitor termini, which suggests that inhibitor interactions in these regions have been less well optimized.

Discussion

The observation that thermodynamic binding properties exhibit a correlation with exposed surface area has been extended to interactions within protein-protein interfaces. Measures of interatomic surfaces buried within a binding interface are used in this model of binding free energy to quantify binding strength. Model parameters are based on a statistical analysis of interatomic surface preferences from a set of known enzyme-protein complexes. The parameterization and construction of the model presuppose rigid molecular units and utilize, necessarily, only atom types found among the naturally occurring amino acids. The results of this analysis indicate that, although the greater part of protein-protein interfaces are covered by hydrophobic groups, the corresponding relative importance of hydrophilic groups as highly specific interactions is established. Thus, a general picture of binding does not lead to dominance of either hydrophobic or hydrogen bond interactions. As noted earlier (Young et al., 1994), it is usually possible to locate putative target sites via a surface scan of hydrophobic sites, or patches, in order to determine the general position and shape of binding sites. The present development permits assessment of contribu-

Fig. 6. Composite view of atom surfaces for all 10 inhibitors. Inhibitor positions and views are identical to those of Figure 5. Surfaces of optimal inhibitor atoms are indicated in red and white. Yellow surfaces identify atoms where replacement might yield a higher binding affinity for that inhibitor.

tions from both hydrophobic and hydrophilic amino acid atom burial toward the specificity of target sites, as well as their contribution to the free energy of binding. The observed atomic surface preferences for natural amino acids support the importance of hydrophobic pairing, hydrogen bonding, and the ion pairings among charged or partially charged residues. These atom-atom preferences are also consistent with experimental observations regarding protein stability upon point mutation for a number of different systems. Analysis of the binding coefficients (α_{ij}) reveals the range and details of specific interactions that are important at protein interfaces. Although these binding coefficients have the same units as a surface tension coefficient, there is no apparent theoretical connection between the two quantities. The evaluation of the binding coefficients is intimately connected with the type of surface, i.e., a molecular surface, used in the evaluation of surface preferences and bears no resemblance to the macroscopic interfaces used to evaluate surface tensions. The concept of a "molecular surface" is introduced simply to account empirically for the binding strengths and preferences among enzyme-inhibitor complexes. Although the present data set only considered intermolecular interfaces, the intramolecular surface burials ought to bear a strong resemblance to external surface burial (Bennett et al., 1994) and could be analyzed in a similar fashion. From this viewpoint, the intramolecular energy of assembling surface elements would need to be considered during protein folding, a free energy consideration that is not necessary for treatment of the association of rigid intermolecular enzyme-inhibitor surfaces.

There has been an unprecedented need for the development of an effective inhibitor of HIV-1-PR. The theoretical part of rational drug design (Greer et al., 1994; Kuntz et al., 1994; Verlinde & Hol, 1994) utilizes conventional methods in the modeling phase of developing a specific drug. Such efforts can benefit greatly from the use of a general free energy model of protein-protein binding. Methods such as docking or placing individual side chains at the active binding site rely on the potential energy function of atomic interactions (Moon & Howe, 1991; Caffisch et al., 1993; Gustchina et al., 1994) rather than on free energy considerations to discriminate among possible compounds. Such methods can be quite successful in identifying lead compounds (Lam et al., 1994). More sophisticated and expensive calculations have applied the free energy perturbation method in order to evaluate changes in binding energy upon substitution of small molecular fragments (Reddy et al., 1994; Chen & Tropsha, 1995). The current state of the art in molecular simulations of HIV-1-PR inhibitor binding was reviewed by McCarrick and Kollman (1994). Correlations of potential energy results with IC_{50} -values have been applied to a selected set of HIV-1-PR dimers (Oprea et al., 1994; Holloway et al., 1995) in order to investigate an empirical structure-activity relationship.

The present identification of a common binding motif utilized by a set of inhibitor compounds to the HIV-1-PR molecule permits a retrospective understanding of the currently designed HIV-1-PR inhibitors (Sawyer et al., 1992; Dorsey et al., 1994; Ghosh et al., 1994; Greer et al., 1994; Lam et al., 1994; Lunney et al., 1994; Varney et al., 1994). The importance of hydrophobic residues flanking the HIV-1-PR active site residues is emphasized heavily in this motif. Accordingly, the observation that inhibitor interactions with hydrophobic HIV-1-PR residues enhances the binding affinity is in accord with this picture. One should bear in mind that many of these sites are distant from

the active site and thus might be less critical for the function of the protease. The existence of a common binding motif for the known HIV-1-PR inhibitors suggests a design strategy that exchanges entire sections between known inhibitors to achieve a more effective compound. Thus, the P_1 - P_3 section of one molecule might be replaced by the same section of another molecule (Thompson et al., 1994), not necessarily conserving the C_2 symmetry of the binding pocket.

The property of the HIV-1-PR, and closely related Rous sarcoma virus (RSV) protease, to produce successfully in vivo escape mutants severely limits the usefulness of current protease inhibitors (Grinde et al., 1992a, 1992b; Otto et al., 1993; Ho et al., 1994). The specificity of the substrate to be processed by the protease enzyme is relatively low, the conserved amino acids are only located in the P_2 , P_1 , and P'_2 sites (Poorman et al., 1991), close to the active site. Table 8A indicates that the decomposition of the free energy of binding at any P site is essentially the same. Thus, the inhibitor binding strength is derived from an equal energy arrangement over these sites, but, given the specificity of binding, any loss of attractive binding energy from sites other than the conserved sites will eliminate an essential portion of the binding energy. The nonconserved sites appear to be essential to give the drug an inhibition potential in the nanomolar range. An unfortunate consequence of the shared binding motif is that in vivo resistance developed through the exposure of one inhibitor will automatically decrease inhibition by any other drug utilizing the same binding pattern. The method developed here to understand enzyme-inhibitor binding through an empirical surface burial model is an effort to elucidate the general character of binding epitopes at an intermolecular interface. It can identify specific atom-atom and residue-residue preferences found among naturally occurring protein-protein complexes. The results are weighed by the parameters derived here in order to estimate relative importances and overall contributions to the free energy of binding. Currently, the implementation of the model is at the descriptive level; predictions based on a complementary surface are easily generated, but the chemical construction of such a surface needs to be separately considered. We have shown here the utility of these new surface parameters to compare the relative binding strengths of a series of inhibitors. Their principal utility may lie in suggesting improvements to existing inhibitors.

Acknowledgments

Special thanks are extended to Dr. A. Wlodawer for generously providing us with the X-ray coordinates of the HIV-1 protease inhibitor complexes not in the PDB, and to Dr. Andrew Karplus for valuable comments regarding this manuscript. The staff of the Biomedical Supercomputing Center, FCRDC, Frederick, Maryland is thanked for their assistance and for access to their computer facilities. We thank the NIH Intramural Targeted Anti-viral AIDS Program for support. This research is also sponsored by the National Cancer Institute, DHHS, under Contract with Program Resources, Inc. and Science Applications International Corporation. The contents of this publication do not necessarily reflect the views or policies of the DHHS, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

References

- Appelt K. 1993. Crystal structures of HIV-1 protease-inhibitor complexes. *Perspect Drug Discovery Des* 1:23-48.

- Arakawa K, Tokiwano K, Ohtomo N, Uedaira H. 1979. A note on the nature of ionic hydration and hydrophobic interactions in aqueous solutions. *Bull Chem Soc Jpn* 52:2483–2488.
- Ascenzi P, Amiconi G, Menegatti E, Guarneri M, Bolognesi M, Schnebli HP. 1988. Binding of the recombinant proteinase inhibitor Eglin C from leech *Hirudo medicinalis* to human leukocyte elastase, bovine α -chymotrypsin and subtilisin Carlsberg: Thermodynamic study. *J Enzyme Inhibition* 2:167–172.
- Ben-Naim A. 1990. Solvent effects on protein association and protein folding. *Biopolymers* 29:567–596.
- Ben-Naim A. 1994. Solvation: From small to macro molecules. *Curr Opin Struct Biol* 4:264–268.
- Ben-Naim A, Mazo RM. 1993. Size dependence of the solvation free energies of large solutes. *J Phys Chem* 97:10829–10834.
- Ben-Naim A, Ting KL, Jernigan RL. 1989a. Solvation thermodynamics of biopolymers, I. Separation of the volume and surface interactions with estimates for proteins. *Biopolymers* 28:1309–1325.
- Ben-Naim A, Ting KL, Jernigan RL. 1989b. Solvation thermodynamics of biopolymers, II. Correlation between functional groups. *Biopolymers* 28:1327–1337.
- Ben-Naim A, Ting KL, Jernigan RL. 1990. Solvent effect on binding thermodynamics of biopolymers. *Biopolymers* 29:901–919.
- Bennett MJ, Choe S, Eisenberg D. 1994. Domain swapping: Entangling alliances between proteins. *Proc Natl Acad Sci USA* 91:3127–3131.
- 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.
- Blokzijl W, Engberts JBFN. 1993. Hydrophobic effects. Opinions and facts. *Angew Chem Int Ed Engl* 32:1545–1579.
- Bode W. 1979. The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. II. The binding of the pancreatic trypsin inhibitor and of isoleucine-valine and of sequentially related peptides to trypsinogen and to *p*-guanidinobenzoate-trypsinogen. *J Mol Biol* 127:357–374.
- Bohacek RS, McMartin C. 1992. Definition and display of steric, hydrophobic, and hydrogen-bonding properties of ligand binding sites in proteins using Lee and Richards accessible surface: Validation of a high-resolution graphical tool for drug design. *J Med Chem* 35:1671–1684.
- Bondi A. 1964. Van der Waals volumes and radii. *J Phys Chem* 68:441–449.
- Burley S, Petsko G. 1985. Aromatic-aromatic interactions: A mechanism of protein structure stabilization. *Science* 229:23–28.
- Caffisch A, Miranker A, Karplus M. 1993. Multiple copy simultaneous search and construction of ligands in binding sites: Application to inhibitor of HIV-1 aspartic proteinase. *J Med Chem* 36:2142–2167.
- Carrell C, Carrell H, Erlebach J, Glusker JP. 1988. Structural aspects of metal ion-carboxylate interactions. *J Am Chem Soc* 110:8651–8656.
- Casari G, Sippl MJ. 1992. Structure-derived hydrophobic potential. Hydrophobic potential derived from X-ray structure of globular proteins is able to identify native folds. *J Mol Biol* 224:725–732.
- Chakrabarti P, Dunitz J. 1982. Directional preferences of ether O-atoms toward alkali and alkaline earth cations. *Helv Chim Acta* 65:1482–1488.
- Chen X, Tropsha A. 1995. Relative binding free energies of peptide inhibitors of HIV-1 protease: The influence of the active site protonation state. *J Med Chem* 38:42–48.
- Chen Z, Bode W. 1983. Refined 2.5 Å X-ray crystal structure of the complex formed by porcine kallikrein A and the bovine pancreatic trypsin inhibitor. *J Mol Biol* 164:283–311.
- Cody V, Murray-Rust P. 1984. Iodine-X(O,N,S) intermolecular contacts: Models of thyroid hormone protein binding interactions using information from Cambridge crystallographic data files. *J Mol Struct* 112:189–199.
- Connelly PR, Aldape RA, Bruzzese FJ, Chambers SP, Fitzgibbon MJ, Fleming MA, Itoh S, Livingstone DJ, Navia MA, Thompson JA, Wilson KP. 1994. Enthalpy of hydrogen bond formation in a protein-ligand binding reaction. *Proc Natl Acad Sci USA* 91:1964–1968.
- Connolly ML. 1983. Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221:709–713.
- Covell DG. 1994. Lattice model simulations of polypeptide chain folding. *J Mol Biol* 235:1032–1043.
- Dill KA. 1990. Dominant forces in protein folding. *Biochemistry* 29:7133–7155.
- Dorsey BD, Levin RB, McDaniel SL, Guare JPVJP, Darke PL, Zugay JA, Emini EA, Schleif WA, Quintero JC, Linn JH, Chen IW, Holloway MK, Fitzgerald PMD, Axel MG, Ostovic D, Andersson PS, Huff JR. 1994. L-735,524: The design of a potent and orally bioavailable HIV protease inhibitor. *J Med Chem* 37:3443–3451.
- Dreyer GB, Lambert DM, Meek TD, Carr TJ, Tomaszek TA Jr, Fernandez AV, Bartus H, Cacciavillani E, Hassel AM, Minnich M, Petteway S Jr, Metcalf BW. 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.
- Durell SR, Brooks BR, Ben-Naim A. 1994. Solvent-induced forces between two hydrophilic groups. *J Phys Chem* 98:2198–2202.
- Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. *Nature* 319:199–203.
- Eisenberg D, Wesson M, Yamashita M. 1989. Interpretation of protein folding and binding with atomic solvation parameters. *Chem Scr* 29A:217–221.
- Empie MW, Laskowski M. 1982. Thermodynamics and kinetics of single residue replacements in avian ovomucoid third domains: Effects on inhibitor interactions with serine proteinases. *Biochemistry* 21:2274–2284.
- Erickson HP. 1989. Cooperativity in protein-protein association. The structure and stability of the actin filament. *J Mol Biol* 206:465–474.
- Erickson J, Neidhart DJ, VanDrie J, Kempf DJ, Wang XC, Norbeck DW, Plattner JU, Rittenhouse JW, Turon M, Wideburg N, Kohlbrenner WE, Simmer B, Helfrich R, Paul DA, Knigge M. 1990. Design, activity and 2.8 Å crystal structure of a C-2 symmetric inhibitor complexed to HIV-1 protease. *Science* 249:527–533.
- Fauchere JL, Pliska V. 1983. Hydrophobic parameters π of amino acid side chains from the partitioning of *N*-acetyl-amino-acid amides. *Eur J Med Chem* 18:369–375.
- Finkelstein AV, Janin J. 1989. The price of lost freedom: Entropy of bimolecular complex formation. *Protein Eng* 3:1–3.
- Fischer D, Lin SL, Wolfson HL, Nussinov R. 1995. A geometry-based suite of molecular docking processes. *J Mol Biol* 248:459–477.
- 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 I protease and acetyl-pepstatin at 2.0 Å resolution. *J Biol Chem* 265:14209–14219.
- Frank H, Evans M. 1945. Entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *J Chem Phys* 13:507–532.
- Gao J, Chou LW, Auerbach A. 1993. The nature of cation- π binding: Interactions between tetramethylammonium ion and benzene in aqueous solution. *Biophys J* 65:43–47.
- Ghosh AK, Thompson WJ, Fitzgerald PMD, Culberson JC, Axel MG, McKee SP, Huff JR, Anderson PS. 1994. Structure-based design of HIV-1 protease inhibitors: Replacement of two amides and a 10π -aromatic system by a fused *bis*-tetrahydrofuran. *J Med Chem* 37:2506–2508.
- Giesen DJ, Cramer CJ, Truhlar DG. 1994. Entropic contributions to free energies of solvation. *J Phys Chem* 98:4141–4147.
- Greer J, Erickson JW, Baldwin JJ, Varney MD. 1994. Applications of the three-dimensional structures of protein target molecules in structure-based drug design. *J Med Chem* 37:1035–1054.
- Grinde B, Cameron CE, Leis J, Weber IT, Wlodawer A, Burstein H, Bizub D, Skalka AM. 1992a. Mutations that alter the activity of the Rous sarcoma virus protease. *J Biol Chem* 267:9481–9490.
- Grinde B, Cameron CE, Leis J, Weber IT, Wlodawer A, Burstein H, Bizub D, Skalka AM. 1992b. Analysis of substrate interactions of the Rous sarcoma virus wild type and mutant proteases and human immunodeficiency virus-1 protease using a set of systematically altered peptide substrates. *J Biol Chem* 267:9491–9498.
- Guida WC. 1994. Software for structure-based drug design. *Curr Opin Struct Biol* 4:777–781.
- Gustchina A, Sansom C, Prevost M, Richelle J, Wodak SY, Wlodawer A, Weber IT. 1994. Energy calculations and analysis of HIV-1 protease-inhibitor crystal structures. *Protein Eng* 7:309–317.
- Hansch C, Leo A. 1979. *Substituent constants for correlation analysis in chemistry and biology*. New York: Wiley.
- Ho DD, Toyoshima 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.
- Holloway MK, Wai JM, Halgren TA, Fitzgerald PMD, Vacca JP, Dorsey BD, Levin RB, Thompson WJ, Chen LJ, deSolms J, Gaffin N, Ghosh AK, Giuliani EA, Graham SL, Guare JP, Hungate RW, Lyle TA, Sanders WM, Tucker TJ, Wiggins M, Wiscount CM, Woltersdorf OW, Young SD, Darke PL, Zugay JA. 1995. A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. *J Med Chem* 38:305–317.
- Hopfinger A. 1977. *Intermolecular interactions and biomolecular organization*. New York: Wiley Press.
- Horton N, Lewis M. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci* 1:169–181.

- Huff JR. 1991. HIV protease: A novel chemotherapeutic target for aids. *J Med Chem* 34:2305-2314.
- Hutchins C, Greer J. 1991. Comparative modeling of proteins in the design of novel renin inhibitors. *Crit Rev Biochem Mol Biol* 26:77-127.
- Janin J, Chothia C. 1990. The structure of protein-protein recognition sites. *J Biol Chem* 265:16027-16030.
- Jaskolski 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.
- Jorgensen WL. 1991. Rusting of the lock and key model for protein-ligand binding. *Science* 254:954-955.
- Jorgensen WL, Buckner JK, Boudon S, Tirado-Rives J. 1988. Efficient computation of absolute free energies of binding by computer simulation. Application to the methane dimer in water. *J Chem Phys* 89:3742-3746.
- Jorgensen WL, Gao J, Ravimohan C. 1985. Monte Carlo simulations of alkanes in water: Hydration numbers and the hydrophobic effect. *J Phys Chem* 89:3470-3473.
- Jorgensen WL, Severance DL. 1990. Aromatic-aromatic interactions: Free energy profiles for the benzene dimer in water, chloroform, and liquid benzene. *J Am Chem Soc* 112:4768-4774.
- Jungwirth P, Zahradnik R. 1994. Association and dissociation of nonpolar and polar van der Waals pairs in water. Manifestation of the hydrophobic and hydrophilic effect. *J Phys Chem* 98:1328-1332.
- Kauzmann W. 1959. Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14:1-63.
- Kuntz ID, Meng EC, Shoichet BK. 1994. Structure-based molecular design. *Acc Chem Res* 27:117-123.
- Lam PY, Jadhav PK, Eyermann CJ, Hodge CN, Ru Y, Bacheler LT, Meek JL, Otto MJ, Rayner MM, Wong YN, Chang CH, Weber PC, Jackson DA, Sharpe TR, Erickson-Viitanen S. 1994. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* 263:380-384.
- Lawrence MC, Colman PM. 1993. Shape complementarity at protein/protein interfaces. *J Mol Biol* 234:946-950.
- Lee B. 1994. Relation between volume correction and the standard state. *Biophys Chem* 51:263-269.
- Lee B, Richards FM. 1971. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol* 55:379-400.
- Linse P. 1992. Stacked or T-shaped benzene dimers in aqueous solution? A molecular dynamics study. *J Am Chem Soc* 114:4366-4373.
- Linse P. 1993. Orientation-averaged benzene-benzene potential of mean force in aqueous solution. *J Am Chem Soc* 115:8793-8797.
- Lunney EA, Hagen SE, Domagala JM, Humblet C, Kosinski J, Tait BD, Varmus JS, Wilson M, Ferguson D, Hupe D, Tummino PJ, Baldwin ET, Bhat TN, Liu B, Erickson JW. 1994. A novel nonpeptide HIV-1 protease inhibitor: Elucidation of the binding mode and its application in the design of related analogs. *J Med Chem* 37:2664-2677.
- Magalhaes A, Maigret B, Hoflack J, Gomes JNF, Scheraga HA. 1994. Contribution of unusual arginine-arginine short-range interactions to stabilization and recognition in proteins. *J Protein Chem* 13:195-215.
- Makhatadze GI, Privalov PL. 1990. Heat capacities of proteins. I. Partial molar heat capacities of individual amino acid residues in aqueous solution: Hydration effect. *J Mol Biol* 213:375-384.
- Makhatadze GI, Privalov PL. 1993. Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J Mol Biol* 232:639-659.
- Mark AE, van Gunsteren WF. 1994. Decomposition of the free energy of a system in terms of specific interaction. Implications for theoretical and experimental studies. *J Mol Biol* 240:167-176.
- Marshall GR. 1992. Three-dimensional structure of peptide-protein complexes: Implications for recognition. *Curr Opin Struct Biol* 2:904-919.
- McCarriek MA, Kollman P. 1994. Use of molecular dynamics and free energy perturbation calculations in anti-human immunodeficient virus drug design. *Methods Enzymol* 241:370-384.
- Miller S. 1989. The structure of interfaces between subunits of dimeric and tetrameric proteins. *Protein Eng* 3:77-83.
- Miller S, Lesk A, Janin J, Chothia C. 1987. The accessible surface area and stability of oligomeric proteins. *Science* 238:834-836.
- Miyazawa S, Jernigan RL. 1985. Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules* 18:534-552.
- Miyazawa S, Jernigan RL. 1993. A new substitution matrix for protein sequence searches based on contact frequencies in protein structures. *Protein Eng* 6:267-278.
- Moon JB, Howe WJ. 1991. Computer design of bioactive molecules: A method for receptor-based de novo ligand design. *Proteins Struct Funct Genet* 11:314-328.
- Mrabet NT, van der Broeck A, van der Brande I, Stanssens P, Laroche Y, Lambeir AM, Matthijssens G, Jenkins J, Chiadmi M, van Tilbeurgh H, Rey F, Janin J, Quax WJ, Lasters I, de Maeyer M, Wodak SJ. 1992. Arginine residues as stabilizing elements in proteins. *Biochemistry* 31:2239-2253.
- Murphy KP, Xie D, Thompson KS, Amzel LM, Freire E. 1994. Entropy in biological binding processes: Estimation of translational entropy loss. *Proteins Struct Funct Genet* 18:63-67.
- Nemethy G. 1967. Hydrophobic interactions. *Ang Chem Int Ed Engl* 6:195-206.
- Nicholls A, Sharp KA, Honig B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct Funct Genet* 11:281-296.
- Norel R, Lin SL, Wolfson H, Nussinov R. 1994. Shape complementarity at protein-protein interfaces. *Biopolymers* 34:933-940.
- Novotny J, Brucoleri RE, Saul FA. 1989. On the attribution of binding energy in antigen-antibody complexes McPC 603, D1.3 and HyHEL-5. *Biochemistry* 28:4735-4749.
- Ooi T, Oobatake M, Nemethy G, Scheraga HA. 1987. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc Natl Acad Sci USA* 84:3086-3090.
- Oprea TI, Waller CL, Marshall G. 1994. Three-dimensional quantitative structure-activity relationship of human immunodeficiency virus (I) protease inhibitors. 2. Predictive power using limited exploration of alternate binding modes. *J Med Chem* 37:2206-2215.
- Otting G, Liepinsh E, Wüthrich K. 1991. Protein hydration in aqueous solution. *Science* 254:974-980.
- Otto MJ, Garber S, Winslow DL, Reid CD, Aldrich P, Jadhav PK, Patterson CE, Hodge CN, Cheng YSE. 1993. In vitro isolation and identification of human immunodeficiency virus HIV variants with reduced sensitivity to C-2 symmetrical inhibitors of HIV type 1 protease. *Proc Natl Acad Sci USA* 90:7543-7547.
- Perutz M. 1990. *Mechanisms of cooperativity and allosteric regulation in proteins*. New York: Cambridge University Press.
- Poorman RA, Tomasselli AG, Heinrichson RL, Kezdy FJ. 1991. A cumulative specificity model for proteases from human immunodeficiency virus types 1 and 2, inferred from statistical analysis of an extended substrate data base. *J Biol Chem* 266:14554-14561.
- Privalov PL, Makhatadze GI. 1990. Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptides chain of proteins: Protein unfolding effects. *J Mol Biol* 213:385-391.
- Privalov PL, Makhatadze GI. 1992. Contribution of hydration and noncovalent interactions to the heat capacity effect on protein unfolding. *J Mol Biol* 224:715-723.
- Privalov PL, Makhatadze GI. 1993. Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. *J Mol Biol* 232:660-679.
- Ravishanker G, Beveridge DL. 1985. Potential of mean force for the stacking of phenyl rings in aqueous solution. *J Am Chem Soc* 107:2565-2566.
- Reddy MR, Varney MD, Kalish V, Viswanadhan VN, Appelt K. 1994. Calculation of relative differences in the binding free energies of HIV1 protease inhibitors: A thermodynamic cycle perturbation approach. *J Med Chem* 37:1145-1152.
- Rich DH, Sun CQ, Prasad JNVN, Pathiaseril A, Toth MV, Marshall GR, Clare M, Mueller RA, Houseman K. 1991. Effect of hydroxyl group configuration in hydroxyethylamine dipeptide isosteres on HIV protease inhibition. Evidence for multiple binding modes. *J Med Chem* 34:1222-1225.
- Richards FM. 1977. Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng* 6:151-176.
- Richmond TJ. 1984. Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J Mol Biol* 178:63-89.
- Ripoll DR, Faerman CH, Axelsen PH, Silman I, Sussman JL. 1993. An electrostatic mechanism for substrate guidance down the aromatic gorge of acetylcholinesterase. *Proc Natl Acad Sci USA* 90:5128-5132.
- Roe SM, Teeter MM. 1993. Patterns for prediction of hydration around polar residues in proteins. *J Mol Biol* 229:419-427.
- Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. 1985. Hydrophobicity of amino acid residues in globular proteins. *Science* 229:834-838.
- Roseman MA. 1988. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J Mol Biol* 200:513-522.
- Rullmann JAC, van Duijnen PT. 1990. Potential energy models of biological macromolecules: A case for ab initio quantum chemistry. *Rep Mol Theory* 1:1-21.

- Sawyer TK, Staples DJ, Liu L, Tomasselli AG, Hui JO, O'Connell K, Schostarez H, Hester JB, Moon J, Howe WJ, Smith CW, Decamp DL, Craik CS, Dunn BM, Lowther WT, Harris J, Poorman RA, Wlodawer A, Jaskolski M, Heinrikson RL. 1992. HIV protease (HIV PR) inhibitor structure-activity-selectivity, and active site molecular modeling of high affinity Leu [CH(OH)CH₂]Val modified viral and nonviral substrate analog. *Int J Pept Protein Res* 40:274-281.
- Schecter I, Burger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27:157-162.
- Sharp KA, Nicholls A, Fine RF, Honig B. 1991. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science* 252:106-109.
- Shirley BA, Stanssens P, Hahn U, Pace CN. 1992. Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry* 31:725-732.
- Singh J, Thornton JM. 1992. *Atlas of protein side-chain interactions, vol 1 and 2*. Oxford: IRL Press.
- Sippl MJ. 1993. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J Comput-Aided Mol Des* 7:473-501.
- Srivastava S, Crippen G. 1993. Analysis of cocaine receptor site ligand binding by the three-dimensional Voronoi site modeling approach. *J Med Chem* 36:3572-3579.
- Stryer L. 1988. *Biochemistry*. New York: W.H. Freeman and Company.
- Swain AL, Miller MM, Green J, Rich DH, Schneider J, Kent SBH, Wlodawer A. 1990. X-ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor. *Proc Natl Acad Sci USA* 87:8805-8809.
- Tanford C. 1980. *The hydrophobic effect*. New York: Wiley.
- Thanki N, Rao J, Foundling S, Howe W, Tomasselli A, Heinrikson R, Thaisrivongs S, Wlodawer A. 1992. Crystal structure of a complex of HIV-1 protease with a dihydroethylene-containing inhibitor: Comparisons with molecular modeling. *Protein Sci* 1:1061-1072.
- Thompson SK, Murthy KHM, Zhao B, Winborne E, Green DW, Fisher SM, DesJarlais RL, Tomaszek TA, Meek TD, Gleason JG, Abdel-Meguid SS. 1994. Rational design, synthesis, and crystallographic analysis of a hydroxyethylene-based HIV-1 protease inhibitor containing a heterocyclic P1'-P2' amide bond isostere. *J Med Chem* 37:3100-3107.
- Tunon I, Silla E, Pascual-Ahuir JL. 1992. Molecular surface area and the hydrophobic effect. *Protein Eng* 5:715-716.
- Varney MD, Appelt K, Kalish V, Reddy MR, Tatlock J, Palmer CL, Romines WH, Wu BW, Musick L. 1994. Crystal-structure-based design and synthesis of novel C-terminal inhibitors of HIV protease. *J Med Chem* 37:2274-2284.
- Verdonk ML, Boks GJ, Kooijman H, Kanters JA, Kroon J. 1993. Stereochemistry of charged nitrogen-aromatic interactions and its involvement in ligand-receptor binding. *J Comput-Aided Mol Des* 7:173-182.
- Verlinde CLMJ, Hol WGJ. 1994. Structure-based drug design: Progress, results, and challenges. *Structure* 2:577-587.
- Vincent JP, Lazdunski M. 1972. Trypsin-pancreatic trypsin inhibitor association. Dynamics of the interaction and role of disulfides bridges. *Biochemistry* 11:2967-2977.
- Vincent JP, Peron-Renner M, Pudles J, Lazdunski M. 1974. The association of anhydrotrypsin with the pancreatic trypsin inhibitor. *Biochemistry* 13:4205-4211.
- Wallqvist A, Covell DG. 1995. Cooperativity of water-solute interactions at a hydrophilic surface. *J Phys Chem* 99:5705-5712.
- Walls PH, Sternberg MJE. 1992. New algorithm to model protein-protein recognition based on surface complementarity. Application to antibody-antigen docking. *J Mol Biol* 228:277-297.
- Walshaw J, Goodfellow JM. 1993. Distribution of solvent molecules around apolar side-chains in protein crystals. *J Mol Biol* 231:392-414.
- Wesson L, Eisenberg D. 1992. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci* 1:227-235.
- Wilson C, Mace JE, Agard DA. 1991. Computational method for the design of enzymes with altered substrate specificity. *J Mol Biol* 220:495-506.
- Wlodawer A, Erickson JW. 1993. Structure-based inhibitors of HIV-1 protease. *Annu Rev Biochem* 62:543-583.
- Wodak SJ, Janin J. 1978. Computer analysis of protein-protein interactions. *J Mol Biol* 124:323-342.
- Wodak SJ, Rooman MJ. 1993. Generating and testing protein folds. *Curr Opin Struct Biol* 3:247-259.
- Young L, Jernigan RL, Covell DG. 1994. A role for surface hydrophobicity in protein-protein recognition. *Protein Sci* 3:717-729.
- Yun-Yu S, Mark AE, Cun-Xin W, Fuhua H, Berendsen HJC, van Gunsteren WF. 1993. Can the stability of protein mutants be predicted by free energy calculations? *Protein Eng* 6:289-295.

Fig. 7. Distribution of correlation coefficients c_{ij} among the derived atom-binding parameters. The desired value of 0 indicates that there is no correlation among the selected atom type, whereas a value of 1 corresponds to having a redundant classification between two atom types.

Appendix: Correlation between atom-atom binding coefficients

We can investigate the correlation among binding coefficients by forming the normalized dot product between the matrix arrays in Table 4B. Thus, the correlation coefficient c_{ij} is defined as,

$$c_{ij} = \sum_k \frac{\alpha_{ik} \alpha_{jk}}{|\alpha_{ik}| |\alpha_{jk}|}, \quad (9)$$

where the summation over k spans all reduced atom types. If the atom types were truly independent, c_{ij} would be zero. This ideal state is, how-

Table 10. Similarity of binding coefficients, α_{ij} , expressed as a correlation, c_{ij}

i	j	c_{ij}
O _{alc}	C _N ⁺	0.74
C _β	C _{aliph}	0.67
O _{pep}	O _{alc}	0.66
O _{carb}	C _{carb}	0.65
C _{CNC}	C _γ	0.64
O _{pep}	C _{pep}	0.63
C _γ	O _γ	0.60
N _{amid}	C _{amid}	0.59
N _{amid}	O _γ	0.56
O _{carb}	C _α	0.56
O _{carb}	C _{pep}	0.55
N _{amid}	O _{amid}	0.52
N _R	C _{aliph}	0.52
O _{pep}	O _{amid}	0.52
C _α	C _{carb}	0.52
C _{carb}	C _{CNC}	0.52
N _{arom}	O _{alc}	0.51
O _{pep}	O _{carb}	0.51
C _{aliph}	C _{arom}	0.51

ever, not achieved due to the natural proximity of certain atom types with other atom types in the amino acids. In Figure 7, we give the distribution of binding coefficients for values of the correlation coefficient, c_{ij} . Our choice of atom types successfully avoids the highest correlations, but there is still a minor shift of the entire distribution toward positive correlation. Table 10 lists the highest scoring correlation coefficients. The reason for a high correlation among the interactions can arise for many different reasons. Like element types such as the C_β and C_{aliph} possess a similar chemical nature and can be expected to be correlated among their interactions with other atom types. A somewhat smaller correlation, 0.51, is associated with aliphatic and aromatic carbon pairs. In Table 4B, we can also see directly that the hydrophobic carbon pairs, with either an aliphatic or an aromatic carbon, are similar. There is, how-

ever, a marked difference in the nitrogen-carbon pairs. The behavior of aliphatic carbons is almost the opposite of aromatic carbon-nitrogen interactions.

The positions of heavy atoms in an amide or carboxylate group are strongly dependent on each other and the surface exposure of an N_{amid} and C_{amid} will be similar to each other. This is partly the reason why the $(N_{amid}C_{amid})$, $(O_{carb}C_{carb})$, and $(O_{pep}C_{pep})$ pairs exhibit enhanced correlations. This correlation is also connected to the preferential association of these groups in the enzyme-inhibitor complex through complementary hydrogen bonding. Atom types may also exhibit a correlation through their effective interactions. Thus, the oxygen atoms of serine and threonine, O_{alc} , exhibit a similarity to the interactions of a ζ - and ϵ -carbon on arginine and lysine.