Occluded Molecular Surface Analysis of Ligand−Macromolecule Contacts: Application to HIV-1 Protease−Inhibitor Complexes

Nagarajan Pattabiraman*

Advanced Biomedical Computing Center, SAIC−NCI/FCRDC, P.O. Box B, Frederick, Maryland 21701-1201

Received September 4, 1998

Herein is described a method of quantifying and visualizing ligand−macromolecule contacts with the occluded surface algorithm by utilizing Connolly's van der Waals molecular surface dots together with the associated normals, to scoop out surrounding macromolecule atoms within a distance of 6.4 Å from any ligand atom. On the basis of the intersections of surface normals with the van der Waals spheres of surrounding macromolecule atoms, the van der Waals molecular surface area for each atom is divided into occluded and nonoccluded surface areas. Also, we calculate a packing parameter for each occluded surface, measuring the closeness of the occluded surface against the macromolecule atom in contact. From the partial charges of ligand and macromolecule atoms and the occluded and nonoccluded surface areas due to the contact, we were able to identify favorable and unfavorable contacts. From the value of occluded surface constant, nonoccluded surface constant, and solvent-exposed constant for ligands, we qualitatively rank order the binding of ligands to the same target. From the individual parameters, group parameters for groups of atoms in a ligand or for each residue in a ligand-binding pocket of a macromolecule could be calculated. The group and the residue-based parameters could be used in structure-based ligand design and protein engineering experiments. In this paper, we present our analysis of ligand−macromolecule contacts, using five X-ray crystal structures of HIV-1 protease−ligand complexes.

Introduction

The specificity and tight binding of small ligands to their corresponding macromolecule targets are very important in biochemical processes. A number of laboratories developed computational methods to calculate the binding free energies of small molecules to a macromolecule.[1] However, these computational techniques either are computer-intensive or can be applied only to a limited number of closely related analogues. Presently, in structure-based drug design studies, medicinal chemists have to visually evaluate ligand−macromolecule contacts on a computer graphics screen.

In some of the reported crystal structures of ligand−macromolecule complexes, only a few of the ligand−macromolecule contacts (within a distance of 3.5 Å) were shown on a 2D cartoon representation of the ligand and its binding site.[2] The program LIGPROT[3] projects the 3D coordinates of the ligand and its binding site residues onto a plane. To get nonoverlapping atoms in the 2D projection, some of the internal parameters have to be changed. Also, only distances between the hydrogen-bonded atoms and some of the hydrophobic contacts involved between the ligand and the macromolecule were shown.

These 2D representations give a semiqualitative description of the ligand−macromolecule contacts and do not quantify these contacts. Is it possible to quantify each ligand−macromolecule contact? By quantifying these contacts, is it possible to define a "local" binding pocket for each atom and also for the whole ligand? In some of the reported crystal structure complexes,[4] a cutoff distance of 4.1 Å was used to identify ligand−macromolecule contacts. In the case of protein-folding simulation studies, a cutoff distance of 6.4 Å was used.[5] What is the cutoff distance one has to use to identify all the critical ligand−macromolecule contacts?

In a ligand−macromolecule complex, the specificity and the strength are determined first by the shapes of the ligand as well as the shape of the binding site on the macromolecule target and by hydrogen bonds and hydrophobic−hydrophobic interactions. In all these interactions, the ligand−macromolecule complex not only increases intermolecular hydrogen bonds and hydrophobic−hydrophobic interactions but also decreases unfavorable intermolecular hydrophobic−polar interactions. Lee and Richard[6] calculated the solvent-accessible surface of a macromolecule (such as a protein) by rolling a probe sphere with a radius of 1.4 Å (as in a water molecule) over the protein atoms. A number of laboratories have used solvent-accessible surfaces to calculate the atomic solvation parameters (ASP),[7] a simple free energy model based on protein−protein interfaces[8] and change in hydrophobic and hydrophilic surface area buried upon binding.[9] These methods address the chemical nature of the protein−protein interactions.

Gregoret and Cohen,[10] in a study based on high-resolution crystal structures of proteins, developed a sphere-growth method to calculate pairwise potentials for amino acids qualitatively measuring the interior packing of proteins. In this method, each amino acid side chain was represented as one, two, or three points, depending upon the size of the side chain. For this reason, this method could not be applied to identify individual ligand−macromolecule contact. In addition,

* Phone: (301)846-5705. Fax: (301)846-5762. E-mail: pattabir@ncifcrf.gov.

10.1021/jm980512c CCC: $18.00
Published on Web 08/27/1999

this method is not suitable to calculating the packing of nonpeptide ligands in its binding pocket.

Sobolev et al.[11] developed an approach not based on interaction energy but based on contact surface between a ligand and its binding site. In their method, contact surface of atom−atom contact was calculated by the intersection of spheres placed at the center of each atom. The radius of each sphere was set to the van der Waals radius extended by the radius of a water probe sphere (1.4 Å). Each sphere was divided into a number of surface patches. If the distance from the center of a patch to the center of the contact sphere was less than the extended van der Waals radius of the contact sphere, then the surface patch was said to be in contact. If a point-of-contact surface area fell within the intersecting volumes of two or more spheres, the surface point was arbitrarily set to be in contact with the closest atom. This obviously distorted the calculation of the contact surface areas. In addition, the ligand−macromolecule interface occurs at the van der Waals surface, not at the solvent-accessible surface.

Recently, we developed a method of calculating the occluded molecular surface for each residue in a protein structure.[12] This method does not use solvent-accessible surfaces but instead uses Connolly's van der Waals molecular surfaces. We also calculated packing parameters for the occluded surfaces as discussed in the Computational Methods section. By calculating surface areas of occluded and nonoccluded surfaces and the packing parameter for each residue in a protein structure, we were able to identify misfolded protein structures. Dedecker et al.,[13] by using the occluded surface algorithm to calculate the packing parameter of the crystal structure of *P. woesei* TATA-box binding protein, have shown that one of the chemical characteristics for increase in protein thermostability is more compact packing. Also, Fleming et al.[14] found a remarkable agreement between the differences in ΔG_d of the wild type and the mutant of transmembrane helix dimers when measured by ultracentrifuge and the calculated free energy based on the difference in occluded surface areas. Recently, Ratnaparkhi et al.[15] concluded that, on the basis of the occluded surface algorithm, the NMR structure of uncomplexed barstar shows an unusually high packing value, compared to that of the crystal structure, indicating errors in the packing of the NMR structure. Thus, the occluded surface algorithm has been shown to be useful in the analysis of protein structures. In this paper, we describe a method of quantifying and visualizing ligand−macromolecule contacts, using the occluded surface algorithm. From the partial charges of the ligand and macromolecule atoms and the occluded and nonoccluded surface areas, we calculated the occluded surface, nonoccluded surface, and solvent-exposed constants for ligands. With these constants, we were able not only to identify favorable and unfavorable contacts but also to qualitatively rank order the binding of ligands to the same site of a macromolecule. Five HIV-1 protease inhibitor crystal structure complexes were chosen to illustrate our method of analysis.

Computational Methods

Following the method[12] of calculating occluded and nonoccluded surfaces of a residue in a protein structure, we calculated the occluded and nonoccluded surface areas of a ligand bound to a macromolecule. In our method, the Connolly's van der Waals molecular surface (referred to in this paper as molecular surface) dots were calculated using the program *dms* in the MidasPlus molecular modeling package.[16] Since we were calculating the molecular surface for the ligand only, we used a very high density of dots (50 dots/Å^2) to describe the surface. A 2D schematic representation of our method of calculation is shown in Figure 1. Dashed lines represent a hypothetical ligand. A cross section of the molecular surface of the ligand is shown by dark solid line surrounding the ligand. The van der Waals radius of macromolecule atoms that are within a distance of 6.4 Å from the ligand atom L_i are represented as circles and labeled as M_i and M_j. The surface points associated with the ligand atom, L_i, are labeled 1, 2, 3, 4, and 5. The corresponding surface normals (n1 to n5) at these points are represented as arrows. A circle represents the surface area of each dot. The normals n1 and n2 do not intersect with any van der Waals circle of macromolecule atoms M_i and M_j. The surface areas associated with these normals are set to be nonoccluded. The normal n3 intersects with the van der Waals circle of the atom M_i at two points, *a* and *a′*. The surface point 3 is set to be occluded, and the surface area associated with the point is the occluded surface area. Similarly, the points of interaction for the normals n4 and n5 with the van der Waals circle for M_i are marked as *b*, *b′* and *c*, *c′*, respectively. The surface area associated with surface dots 4 and 5 are also set to be occluded by the macromolecule atom M_i. Thus, for the ligand atom L_i, the molecular surface is divided into two parts: the occluded surface (3, 4, and 5 dots) and the nonoccluded surface (1 and 2 dots). The summation of the surface areas associated with the occluded surface dots is the occluded surface area for the particular ligand−macromolecule contact and is represented by $[OS]_{L_i}^{M_i}$ (Å^2). Similarly, the summation of the surface area of the nonoccluded surface dots is the nonoccluded surface area for the particular ligand−macromolecule contact and is represented by $[NOS]_{L_i}^{M_i}$ (Å^2). The distance between the surface dot and the closest point of intersection of normals at the van der Waals sphere of the ligand M_i is represented by d_3, d_4, and d_5. By averaging the distances d_3, d_4, and d_5, we calculated the packing of the occluded surface for this particular ligand−protein contact. For a given ligand−macromolecule contact, the average distances of the occluded surface dots are normal-

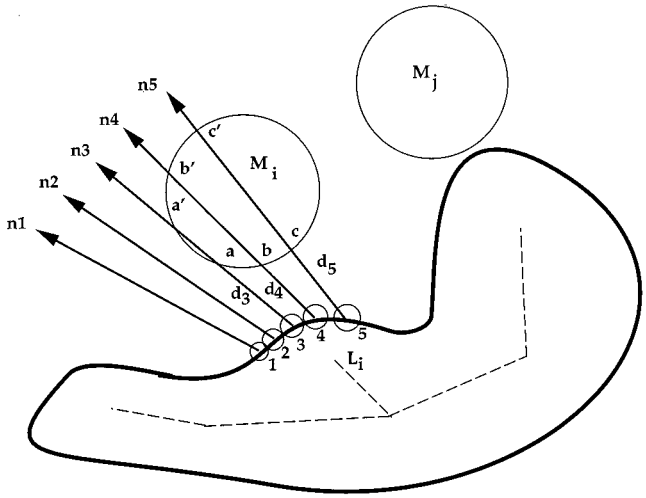

Figure 1. Schematic representation of calculating occluded and nonoccluded surface areas for a ligand atom in a ligand−macromolecule complex. M_i and M_j are the macromolecular atoms that are within 6.4 Å from the ligand atom L_i. The dashed lines represent a ligand molecule. The solid line surrounding the ligand is a cross section of the molecular surface of the ligand. The vectors n1 to n5 are the normals from the molecular surface dots of the ligand atom L_i. The vectors n3, n4, and n5 intersect the van der Waals circle of the ligand atom M_i at *a*,*a′*, *b*,*b′*, and *c*,*c′*, respectively.

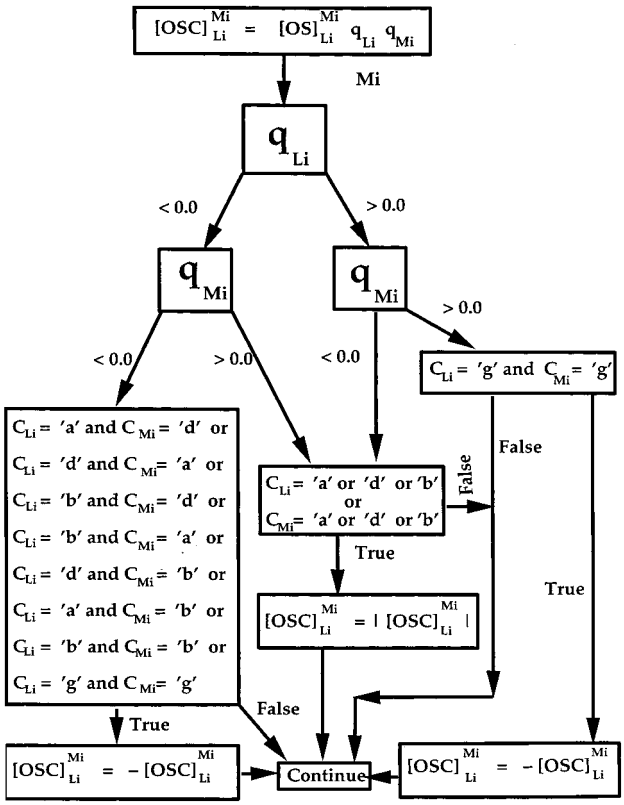

Figure 2. Flowchart for our method of calculating the occluded surface constant, $[OSC]_{L_i}^{M_i}$, using the occluded surface area $[OS]_{L_i}^{M_i}$, the charges on the ligand and the macromolecule, and the chemical nature of the atoms involved in the contact.

ized by the diameter of a water molecule (2.8 Å). This normalized ratio is called the packing parameter, represented by $[PP]_{L_i}^{M_i}$ for the particular occluded patch. If the average distance of the occluded surface dots is greater than 2.8 Å, then a water molecule may be present between the ligand and the macromolecule atoms. Thus, if $[PP]_{L_i}^{M_i}$ is equal to 1, the packing is loose, and if $[PP]_{L_i}^{M_i}$ is equal to 0, then the packing is tight (at the van der Waals contact). For example, the atom M_j is closer to the ligand molecular surface than the atom M_i. We also calculated a weighed occluded surface area (represented by $[WOS]_{L_i}^{M_i}$ in Å^2) for a ligand−macromolecule contact as follows:

$$[WOS]_{L_i}^{M_i} = [OS]_{L_i}^{M_i}(1 - [PP]_{L_i}^{M_i})$$

The weighted occluded surface takes into account not only the amount of surface area occluded but also the packing of the occluded surface against the macromolecule atom.

To quantify favorable and unfavorable ligand−macromolecule contacts, we used the partial atomic charges, the chemical nature of the atoms in contact, and the occluded and nonoccluded surface areas. Figure 2 shows a flowchart of our method of calculation. The charges were calculated using Biosym molecular modeling software of Molecular Simulations Inc.

The occluded surface constant (represented by $[OSC]_{L_i}^{M_i}$) for each ligand−macromolecule contact is calculated as the product of the area of occluded surface, the charge on the ligand atom (q_{L_i}), and the change on the macromolecule atom (q_{M_i}). The $[OSC]_{L_i}^{M_i}$ is modified based on the signs of the charges q_{L_i} and q_{M_i} and also on the chemical nature of the atoms in contact such as donors, acceptors, both acceptor and donor, and hydrophobic. If both q_{L_i} and q_{M_i} are negative and if the ligand atom is an acceptor and the macromolecule atom is a donor, then the $[OSC]_{L_i}^{M_i}$ is set to be favorable (the sign of $[OSC]_{L_i}^{M_i}$ is changed to negative). Similarly, as shown in the box, there are eight such favorable contacts. In Figure 1, the letters a, d, b, and g represent acceptor, donor, both acceptor and donor, and greasy (hydrophobic) atom. If the charges q_{L_i} and q_{M_i} are both positive and if the ligand and the macromolecule atoms are hydrophobic, then the $[OSC]_{L_i}^{M_i}$ is favorable. If either one of the signs of the charges q_{L_i} and q_{M_i} is negative and if the ligand atom or the macromolecule atom is either an acceptor, 'a', or a donor, 'd', or both acceptor and donor, 'b', this contact is unfavorable.

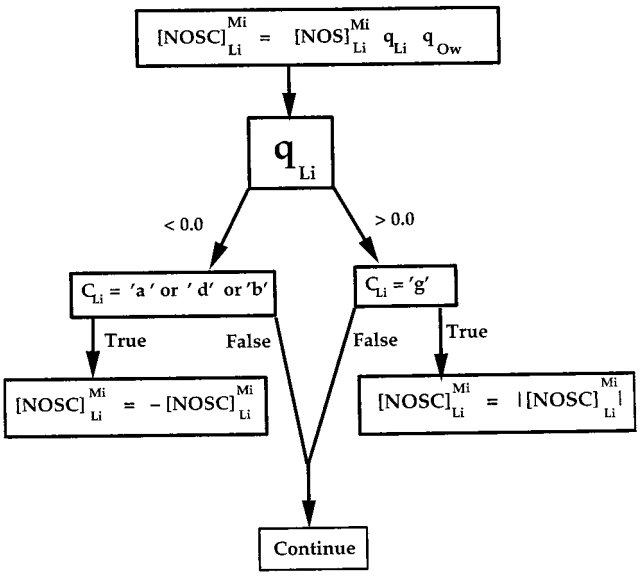

Figure 3. Flowchart for our method of calculating the nonoccluded surface constant, $[NOSC]_{Li}^{M_i}$, using the nonoccluded surface area $[NOS]_{Li}^{M_i}$, the charges on the ligand and the macromolecule, and the chemical nature of the atoms involved in the contact.

For the nonoccluded surface part of a ligand−macromolecule contact, we calculated the nonoccluded surface constant ($[NOSC]_{L_i}^{M_i}$) as shown in Figure 3. We assumed the nonoccluded surface to be exposed to solvent (water). The partial charge on the oxygen atom (q_{O_w}) of the water molecule was assumed to be −0.834. As shown in Figure 3, depending upon the sign of the charge and the nature of the ligand atom, the $[NOSC]_{L_i}^{M_i}$ is set to be favorable or unfavorable. For example, if the ligand charge is positive and if it is a hydrophobic atom, then the $[NOSC]_{L_i}^{M_i}$ is unfavorable.

To measure the strength of the ligand binding to the site, we calculated the solvent-exposed constant ($[SEC]_L$) for the ligand using the method as described in Figure 2. In this calculation, instead of $[NOS]_{L_i}^{M_i}$ we used the total molecular surface area of each atom which will be exposed to water when uncomplexed with the macromolecule.

Following the methodology described in Figure 2, we also calculated the $[OSC]^r$ and $[NOSC]^r$ for each residue of the macromolecule without the ligand. In this case, only the contacts between protein−protein atoms were used in this calculation. By summing $[OSC]^r$ and $[NOSC]^r$ over the entire macromolecule, we calculated the macromolecular stability constant, represented by [MSC].

A MidasPlus display program (version 2.0)[16] running on a Silicon Graphics machine was used to display and analyze the occluded molecular surfaces. In this paper we present our analysis of ligand−macromolecule contacts using five X-ray crystal structures of HIV-1 protease-ligand complexes (the Protein Data Bank codes are 1hiv,[17] 1hpv,[18] 1hpx,[19] 1hvi,[20] and 1hvj[20]). The resolutions of these X-ray structures are less than 2.0 Å, and the *R*-values are less than 20%. In our calculations, we did not include any solvent molecules reported with these structures.

Results and Discussions

Quantifying Ligand−Protease Contacts. In Table 1, the ligand atom name, the protein residue name, the

Table 1. Occluded Surface Constants for Ligand−Protease Contacts of 1hpx Complex[a]

	occluding protein atom						
ligand atom	amino acid	sequence	name	distance (Å)	$[OS]_{L_i}^{M_i}$ (Å^2)	$[PP]_{L_i}^{M_i}$	$[OSC]_{L_i}^{M_i}$
C1	**F**	**153**	**CZ**	**4.10**	**2.41**	**0.24**	**−0.03**
C1	**G**	**148**	**C**	**3.56**	**1.50**	**0.05**	**−0.07**
C1	**G**	**149**	**CA**	**3.89**	**1.46**	**0.09**	**−0.02**
C1	**P**	**81**	**CG**	**4.42**	**1.33**	**0.35**	**−0.02**
C1	**G**	**149**	**C**	**4.22**	**1.20**	**0.26**	**−0.05**
C1	**F**	**153**	**CE1**	**4.43**	**1.18**	**0.39**	**−0.02**
C1	**G**	**148**	**CA**	**4.05**	**1.09**	**0.25**	**−0.01**
C1	G	149	O	3.91	0.84	0.34	0.04
C2	**G**	**149**	**CA**	**3.82**	**4.42**	**0.20**	**−0.06**
C2	**G**	**148**	**C**	**3.36**	**1.87**	**0.04**	**−0.09**
C2	G	148	O	3.16	1.88	0.05	0.09
C4	**P**	**81**	**CB**	**4.15**	**1.52**	**0.32**	**0.00**
C5	**P**	**81**	**CG**	**3.64**	**3.65**	**0.10**	**−0.04**
C5	**P**	**81**	**CB**	**3.76**	**2.31**	**0.10**	**−0.03**
C7	**R**	**8**	**CZ**	**3.87**	**3.02**	**0.17**	**−0.15**
C7	R	8	NH2	3.73	1.38	0.27	0.14
C8	**V**	**82**	**CG1**	**5.12**	**1.82**	**0.55**	**−0.04**
C8	R	8	NH1	4.07	1.10	0.35	0.11
C9	**P**	**81**	**CB**	**4.23**	**1.99**	**0.29**	**−0.03**
C9	**V**	**82**	**CG1**	**5.26**	**1.48**	**0.60**	**−0.03**
C10	**D**	**129**	**CG**	**4.29**	**1.30**	**0.24**	**0.00**
C10	**D**	**129**	**OD2**	**3.96**	**0.95**	**0.37**	**0.00**
C10	**R**	**8**	**CZ**	**4.66**	**0.81**	**0.37**	**0.00**
C11	G	127	O	3.81	1.59	0.24	0.25
C12	G	148	O	3.13	2.61	0.05	0.04
C12	**I**	**50**	**CG2**	**4.58**	**1.80**	**0.38**	**−0.01**
C12	**I**	**147**	**CB**	**4.83**	**1.65**	**0.50**	**0.00**
C12	**G**	**149**	**CA**	**4.71**	**0.83**	**0.38**	**0.00**
C12	**G**	**148**	**C**	**4.09**	**0.83**	**0.27**	**−0.01**
C13	**G**	**149**	**CA**	**4.42**	**2.85**	**0.34**	**−0.12**
C13	**A**	**128**	**CB**	**4.75**	**1.71**	**0.45**	**−0.11**
C15	D	125	OD2	3.11	2.10	0.07	0.05
C15	**A**	**128**	**CB**	**4.73**	**1.95**	**0.47**	**−0.02**
C15	**D**	**125**	**CG**	**3.78**	**0.90**	**0.08**	**−0.02**
C16	D	25	OD2	3.43	1.93	0.11	0.09
C16	**L**	**23**	**CD2**	**4.75**	**1.58**	**0.47**	**−0.03**
C16	**I**	**84**	**CD1**	**4.40**	**1.38**	**0.27**	**−0.02**
C16	G	127	O	3.70	1.07	0.20	0.04
C16	**G**	**127**	**CA**	**4.54**	**1.00**	**0.41**	**−0.01**
C17	**G**	**27**	**CA**	**4.59**	**1.26**	**0.48**	**−0.05**
C17	**G**	**27**	**C**	**4.20**	**1.17**	**0.25**	**−0.18**
C17	D	125	OD2	3.51	0.74	0.22	0.12
C18	**A**	**28**	**CA**	**4.61**	**1.36**	**0.36**	**−0.03**
C18	G	27	O	3.69	1.45	0.24	0.09
C18	**G**	**27**	**C**	**4.38**	**0.85**	**0.28**	**−0.05**
C19	**G**	**49**	**CA**	**4.16**	**1.94**	**0.23**	**−0.01**
C19	G	48	O	4.24	1.61	0.43	0.03
C19	**V**	**182**	**CG1**	**4.66**	**1.61**	**0.45**	**−0.01**
C20	**I**	**184**	**CD1**	**3.95**	**3.83**	**0.22**	**−0.10**
C20	**I**	**50**	**CD1**	**4.69**	**2.24**	**0.42**	**−0.06**
C20	D	125	OD2	3.86	0.94	0.30	0.07
C21	**G**	**49**	**CA**	**4.29**	**1.58**	**0.22**	**−0.07**
C23	**I**	**84**	**CD1**	**3.47**	**4.99**	**0.06**	**−0.13**
C23	**A**	**28**	**CB**	**4.05**	**2.91**	**0.20**	**−0.07**
C23	**I**	**150**	**CG1**	**4.02**	**2.86**	**0.16**	**−0.05**
C23	**V**	**32**	**CG2**	**4.08**	**2.67**	**0.24**	**−0.07**
C23	**I**	**84**	**CG2**	**4.16**	**2.37**	**0.32**	**−0.06**
C24	**I**	**47**	**CD1**	**3.94**	**3.02**	**0.19**	**−0.08**
C24	**I**	**150**	**CG1**	**3.80**	**2.70**	**0.10**	**−0.05**
C24	**G**	**49**	**CA**	**4.30**	**2.29**	**0.28**	**−0.04**
C24	**I**	**150**	**CG2**	**3.83**	**2.37**	**0.13**	**−0.06**
C24	**I**	**47**	**CB**	**4.32**	**1.93**	**0.33**	**−0.02**
C24	**G**	**48**	**C**	**4.09**	**1.91**	**0.21**	**−0.12**
C24	G	48	O	3.79	1.57	0.33	0.10
C24	**I**	**150**	**CB**	**4.07**	**1.07**	**0.27**	**−0.01**
C25	**V**	**32**	**CG2**	**4.06**	**2.98**	**0.23**	**−0.08**
C25	**A**	**28**	**CB**	**4.05**	**2.42**	**0.19**	**−0.06**
C25	**D**	**30**	**CB**	**4.27**	**2.12**	**0.30**	**−0.04**
C25	**I**	**47**	**CD1**	**4.52**	**1.51**	**0.37**	**−0.04**
C25	**I**	**47**	**CG2**	**4.95**	**1.35**	**0.56**	**−0.03**
C25	**A**	**28**	**CA**	**4.45**	**1.15**	**0.34**	**−0.03**

Table 1 (Continued)

ligand atom	occluding protein atom: amino acid	sequence	name	distance (Å)	$[OS]_{L_i}^{M_i}$ (Å^2)	$[PP]_{L_i}^{M_i}$	$[OSC]_{L_i}^{M_i}$
C26	**A**	**128**	**CA**	**4.31**	**1.91**	**0.22**	**−0.05**
C26	**A**	**128**	**CB**	**4.53**	**0.75**	**0.29**	**−0.02**
C27	**A**	**128**	**CB**	**3.77**	**3.57**	**0.14**	**−0.05**
C27	**V**	**132**	**CG2**	**3.73**	**2.90**	**0.11**	**−0.04**
C27	**D**	**130**	**CB**	**4.07**	**1.97**	**0.22**	**−0.02**
C27	**I**	**184**	**CG2**	**4.84**	**1.90**	**0.52**	**−0.03**
C27	D	130	O	3.23	1.83	0.08	0.07
C27	**D**	**130**	**C**	**3.88**	**1.34**	**0.14**	**−0.05**
C27	**I**	**147**	**CD1**	**4.70**	**1.23**	**0.44**	**−0.02**
C27	D	130	N	3.83	1.18	0.25	0.07
C27	**A**	**128**	**CA**	**4.33**	**1.04**	**0.24**	**−0.02**
C27	**D**	**129**	**N**	**4.38**	**0.80**	**0.37**	**0.05**
C28	**I**	**84**	**CD1**	**4.10**	**0.87**	**0.23**	**0.00**
C29	G	127	O	3.82	1.92	0.25	0.10
C30	**V**	**82**	**CG1**	**3.43**	**3.92**	**0.14**	**−0.08**
C31	**P**	**81**	**CG**	**4.11**	**3.20**	**0.23**	**−0.04**
C31	**V**	**82**	**CG1**	**3.57**	**2.22**	**0.12**	**−0.04**
C31	**P**	**81**	**CB**	**4.67**	**1.43**	**0.44**	**−0.02**
C31	**P**	**81**	**CD**	**4.41**	**1.09**	**0.33**	**−0.01**
C32	**G**	**149**	**CA**	**3.85**	**5.66**	**0.20**	**−0.08**
C32	**I**	**150**	**CG1**	**3.74**	**2.32**	**0.09**	**−0.03**
C32	**G**	**149**	**C**	**3.85**	**1.90**	**0.15**	**−0.10**
C32	I	150	N	3.49	1.56	0.14	0.13
C32	**P**	**81**	**CD**	**4.52**	**1.31**	**0.38**	**−0.02**
C32	**I**	**150**	**CD1**	**4.36**	**0.97**	**0.40**	**−0.02**
C32	**I**	**84**	**CD1**	**4.22**	**0.83**	**0.32**	**−0.02**
C33	**I**	**84**	**CD1**	**3.75**	**4.03**	**0.09**	**−0.08**
C33	**I**	**150**	**CG1**	**3.77**	**3.49**	**0.13**	**−0.05**
C33	**G**	**149**	**CA**	**4.29**	**0.79**	**0.35**	**−0.01**
N2	G	149	CA	4.50	1.61	0.40	0.11
N2	**G**	**148**	**O**	**3.16**	**0.88**	**0.11**	**−0.23**
N3	**G**	**127**	**O**	**3.17**	**3.16**	**0.17**	**−0.81**
N3	A	128	CA	4.09	1.68	0.30	0.17
N3	G	127	C	3.85	0.62	0.19	0.16
N5	A	28	CA	4.21	1.72	0.36	0.18
N5	A	28	CB	4.20	1.46	0.28	0.15
O2	G	127	C	3.07	2.90	0.08	0.62
O2	G	127	CA	3.24	2.01	0.12	0.11
O2	D	25	CG	3.48	1.73	0.15	0.37
O2	D	125	CG	3.18	1.67	0.08	0.36
O2	**D**	**125**	**OD1**	**2.62**	**1.30**	**0.05**	**−0.28**
O2	**D**	**25**	**OD1**	**3.12**	**0.92**	**0.21**	**−0.20**
O3	A	128	CA	3.58	2.71	0.20	0.17
O3	D	129	CG	3.37	2.22	0.14	0.35
O3	**D**	**129**	**N**	**2.93**	**2.10**	**0.14**	**−0.54**
O3	G	127	O	3.40	1.10	0.27	0.17
O4	A	28	CB	3.68	2.39	0.23	0.15
O4	G	27	C	3.57	1.93	0.18	0.30
O4	**D**	**25**	**CD2**	**2.55**	**1.80**	**0.03**	**−0.30**
O4	D	25	CG	3.23	1.66	0.06	0.26
O4	A	28	CA	3.77	1.37	0.26	0.09
O5	I	150	CG1	4.12	2.27	0.28	0.10
O5	G	49	CA	3.88	2.07	0.28	0.09
O5	I	150	CB	4.28	1.32	0.47	0.03
O6	G	149	CA	3.78	3.13	0.28	0.13
O6	I	50	CB	3.94	2.46	0.39	0.05
O6	I	50	CD1	4.42	2.04	0.44	0.13
O6	I	50	CG2	4.02	1.22	0.36	0.08
S1	**I**	**50**	**CD1**	**3.82**	**3.83**	**0.22**	**−0.08**
S1	**I**	**184**	**CD1**	**4.26**	**3.01**	**0.32**	**−0.06**
S1	**I**	**147**	**CD1**	**4.35**	**2.74**	**0.36**	**−0.06**
S1	**V**	**132**	**CG2**	**3.83**	**2.09**	**0.15**	**−0.04**
S1	**I**	**50**	**CG1**	**4.23**	**1.58**	**0.34**	**−0.02**
S1	**A**	**128**	**CB**	**4.41**	**1.27**	**0.44**	**−0.03**
S2	**G**	**49**	**CA**	**3.94**	**2.53**	**0.18**	**−0.03**
S2	**P**	**181**	**CG**	**4.05**	**2.56**	**0.25**	**−0.04**
S2	**I**	**50**	**CD1**	**4.40**	**2.13**	**0.37**	**−0.04**
S2	**V**	**182**	**CG1**	**4.45**	**2.12**	**0.38**	**−0.04**
S2	**I**	**50**	**CB**	**4.67**	**1.44**	**0.46**	**−0.01**
S2	**I**	**184**	**CD1**	**4.39**	**1.08**	**0.38**	**−0.02**
S2	**G**	**49**	**C**	**4.17**	**0.94**	**0.25**	**−0.05**
S2	**I**	**50**	**CA**	**4.49**	**0.92**	**0.37**	**−0.02**
S2	I	50	N	3.90	0.76	0.28	0.06

[a] Sequence 1−99 belongs to the A chain. Sequence 101−199 belongs to the B chain.

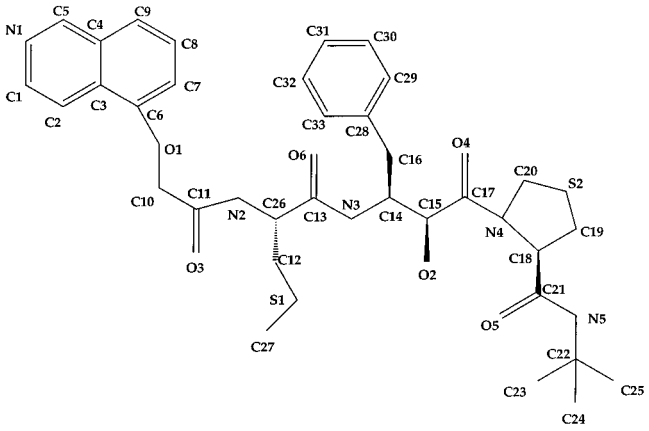

Figure 4. Chemical structure and atom names of the ligand, KNI272, bound to protease in the 1hpx complex.

sequence number, the name of the occluding protease atom, the occluded surface area $[OS]_{L_i}{}^{M_i}$, the packing parameter $[PP]_{L_i}{}^{M_i}$, and the occluded surface constant $[OSC]_{L_i}{}^{M_i}$ for each ligand−protease contact of the crystal structure of the 1hpx complex are listed. The ligand atom names are listed in alphabetical order. In this table, the negative values are considered to be favorable contacts (indicated by bold fonts), where as the positive values are unfavorable contacts. For a few of the contacts, the value of $[OSC]_{L_i}{}^{M_i}$ is 0 (for example, C4 atom) because of small partial charges for the ligand atom. Also, the actual distances in Å (as observed in the crystal structure) are tabulated in the fifth column. For this complex, the total number of ligand−protease contacts is 333. Some of the contacts result in small occluded surface areas, and some of them are very close to the cutoff distance of 6.4 Å. Since we have calculated the weighted occluded surface area ($[WOS]_{L_i}{}^{M_i}$, Å^2) which measures the surface area of occlusion and the packing of these occluded patches, it is possible to eliminate such contacts using a cutoff value for $[WOS]_{L_i}{}^{M_i}$. In Table 1, only contacts that have $[WOS]_{L_i}{}^{M_i}$ greater than 0.5 Å^2 are shown. By using this cutoff, the number of contacts is reduced from 333 to 145.

Figure 4 shows the 2D chemical representation with the atom names of the ligand, KNI-272, as reported by the authors in the PDB entry (1hpx). The ligand atom, C1, which is hydrophobic and is in an aromatic ring, makes eight contacts with the protease molecule at the binding site. Out of the eight contacts, only one contact with the backbone carbonyl oxygen atom of G149 is unfavorable. The other seven contacts are made with hydrophobic atoms. If we add the last column for the atom C1, this atom makes favorable contacts (with $[OSC]_{L_i} = -0.18$) with the binding site. The atom N3, which forms a hydrogen bond when bound, makes a very favorable interaction with the carbonyl oxygen of G127 and makes two unfavorable interactions with the CA atom of A128 and the C atom of G127. For the oxygen atom O2, out of six contacts it makes with the protease, only two contacts are favorable because of two hydrogen-bonding interactions. Overall, in the bound complex, this atom is not making favorable contact with the protease. The summation of the last column of Table 1 gives the total occluded surface constant $[OSC]_{L_i}$ for each ligand atom and is listed in the second column of Table 2. Some of the ligand atoms do not completely bury their surfaces into the binding pocket. So, some of the ligand atoms have nonoccluded surface areas. As discussed in the Computational Methods section, the calculated nonoccluded surface constant, $[NOSC]_{L_i}$, for each ligand atom is listed in column 3 of Table 2.

Table 2. Occluded Surface, Nonoccluded Surface, and Solvent-Exposed Constants for Each Ligand Atom of 1hpx

ligand atom	$[OSC]_{L_i}$	$[NOSC]_{L_i}$	$[SEC]_{L_i}$
C1	−0.18	0.70	2.01
C2	−0.06	0.52	1.50
C4	0.00	0.00	0.00
C5	−0.07	1.33	1.92
C7	−0.01	0.86	1.42
C8	0.07	1.43	1.88
C9	−0.06	1.45	1.92
ClO	0.00	0.05	0.08
C11	0.25	2.15	2.77
C12	0.02	0.22	0.61
C13	−0.23	0.00	1.50
C15	0.01	0.04	0.38
C16	0.07	0.15	1.05
C17	−0.11	0.57	1.80
C18	0.01	0.52	1.07
C19	0.01	0.49	0.71
C20	−0.09	0.87	2.24
C21	−0.07	0.00	0.52
C23	−0.38	0.56	3.09
C24	−0.28	0.75	3.60
C25	−0.28	1.34	3.86
C26	−0.07	0.00	0.37
C27	−0.04	0.49	2.44
C28	0.00	0.00	0.00
C29	0.04	0.58	1.24
C30	−0.08	0.58	1.07
C31	−0.11	0.24	1.23
C32	−0.14	0.42	2.22
C33	−0.14	0.40	1.43
N2	−0.12	0.00	−1.35
N3	−0.48	0.00	−3.33
N5	0.33	−2.74	−4.52
O2	0.98	−0.20	−5.98
O3	0.15	−1.14	−4.37
O4	0.50	−0.12	−4.12
O5	0.22	−1.41	−3.50
O6	0.39	−1.28	−4.56
S1	−0.29	0.44	2.21
S2	−0.19	0.62	2.38

It is clear that the majority of carbon atoms make favorable contacts at the binding site. Two of the nitrogen atoms also make favorable contacts at the binding site. This is due to the formation of hydrogen bonds with a backbone carbonyl oxygen atom of the protease molecule. All the carbonyl oxygen atoms of the ligand do not make favorable contacts even though they make hydrogen bonds with the peptide backbone nitrogen atoms in the site. This is because these carbonyl atoms are surrounded by hydrophobic atoms which make contacts with the carbonyl oxygen atom. The sulfur atoms S1 and S2 have favorable contacts with the protease atoms when bound.

In the uncomplexed state, the entire ligand surface is surrounded by solvent (water) molecules. The solvent-exposed constant $[SEC]_{L_i}$ is listed in the last column of Table 2 for each ligand atom. As expected, all atoms of the ligand are unfavorable except the hydrophilic atoms-nitrogens and oxygens.

Group Parameters from Individual Ligand−Protease Contacts. From the individual occluded surface-based constants, $[OSC]_{L_i}{}^{M_i}$, $[NOSC]_{L_i}{}^{M_i}$ of each contact, and $[SEC]_{L_i}{}^{M_i}$, it is possible to calculate the

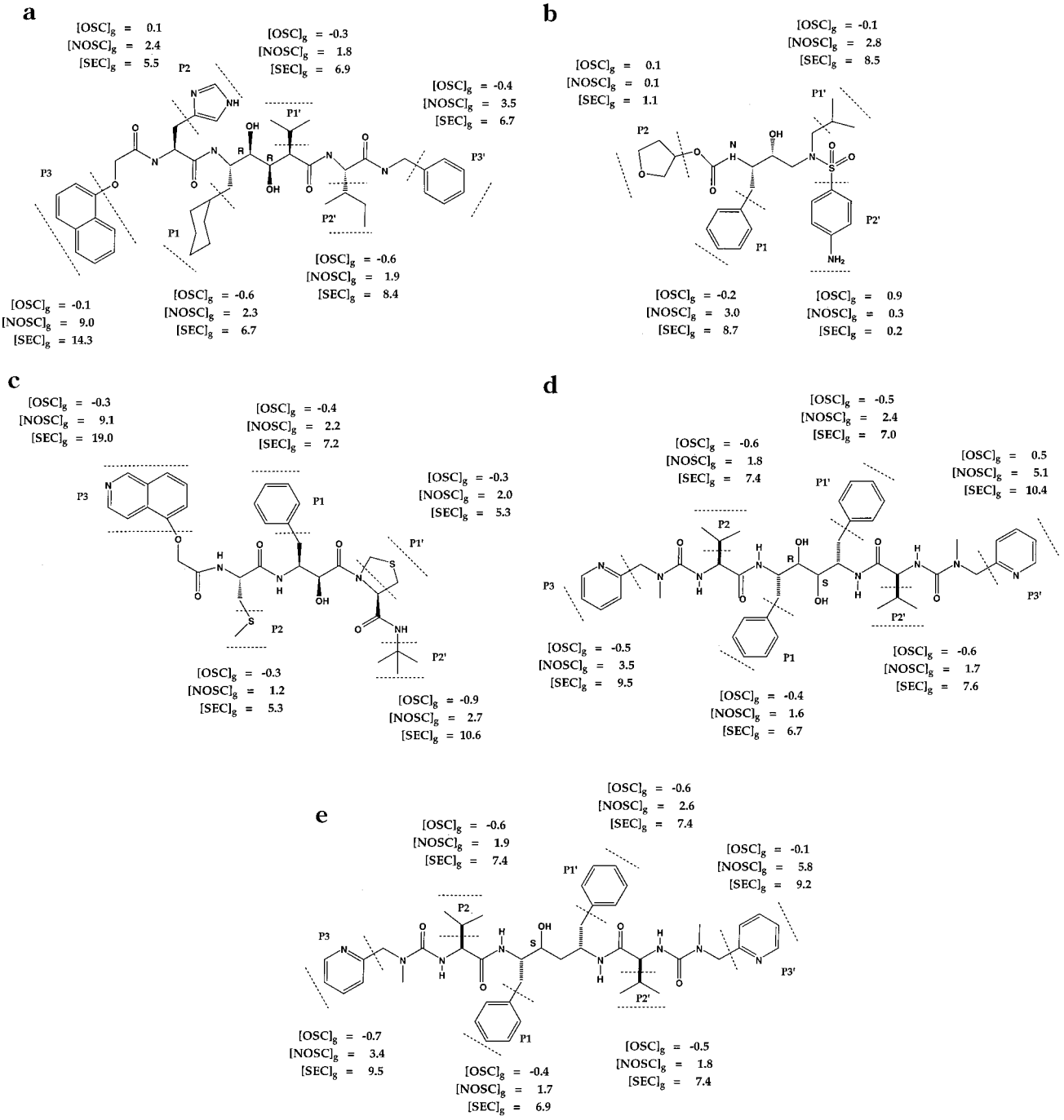

Figure 5. Occluded surface $[OSC]_g$, nonoccluded surface $[NOSC]_g$, and solvent-exposed $[SEC]_g$ group constants for group of atoms (shown between two dashed lines) at the unprimed (P1, P2, and P3) and primed (P1′, P2′, and P3′) sites for the five complexes: (a) 1hiv, (b) 1hpv, (c) 1hpx, (d) 1hvi, and (e) 1hvj.

group parameters for a group of atoms, such as side chains, in a ligand. In Figure 5 the group parameters for the five ligands in the ligand−protease complexes are shown. In these figures, a group of atoms are indicated between the dashed lines and the groups are labeled as P1, P2, P3, P1′, P2′, and P3′ based on the ligand. For each group, we have listed $[OSC]_g$, $[NOSC]_g$, and $[SEC]_g$. It is easy to label the groups for 1hiv, 1hpx, and 1hpv because of the asymmetry in the ligand. For 1hvi and 1hvj complexes, the P1, P2, and P3 groups are respectively the same as that of the P1′, P2′, and P3′ groups. For these compounds, the groups are labeled based on the position of the hydroxyl group with respect to the active site aspartic acid residues. The majority of the groups in these complexes have negative values for $[OSC]_g$, thereby making favorable contacts with the protease. But for the P2 group of the 1hiv complex, the P2 and P2′ groups of the 1hpv complex, and the P3′ group of the 1hvi complex, the $[OSC]_g$ values are positive because some of the atoms in these groups are hydrophilic. As expected, the nonoccluded surface constant ($[NOSC]_g$) and the solvent-exposed constant ($[SEC]_g$) for these groups are all unfavorable. However, the $[SEC]_g$ values for these groups are much more unfavorable when exposed to water. For example, the P1 group for the 1hiv complex has the value of $[SEC]_g$ equal to 6.7. By adding the $[OSC]_g$ and $[NOSC]_g$ for this group, the total group binding constant is 1.7. This group favors

binding to the pocket instead of exposed to water by 5.0 because the group consists of hydrophobic atoms. But, for the P2′ group of the 1hpv complex, it favors exposure to water instead of binding to the pocket. This is due to the hydrophilic nitrogen atom attached to the ring. The P2′ group of the 1hpx complex favors binding to the pocket over exposure to water by 8.8. All the P1 and P1′ groups for these ligand−protease complexes favor binding to the site over exposure to solvent. Even though the P1 group is chemically different from the P1′ group, both groups contribute equally to the binding of the ligand. However, in the case of the 1hpx complex, the P1′ group, thioproline, binds less favorably than the P1 group. It could be possible to improve the binding of the P1′ group by adding hydrophobic groups to the thioproline ring. Except for the P2′ group of the 1hpv complex, all other P2 and P2′ groups have favorable interactions in the complex formation. For the 1hvi complex, even though the P1, P2, and P3 groups are respectively the same as those of the P1′, P2′, and P3′ groups, our group analysis indicates that this ligand does not bind symmetrically. It is interesting to conclude that all the ligands do not bind symmetrically to the homodimer of protease.

We can also regroup ligand−protein contacts into individual atoms of each residue in the protein. In Table 3, the amino acid, the sequence number, the atom name for each occluding residue, and the $[OSC]^{M_i}$ for the five complexes are listed. In the calculation of $[OSC]^{M_i}$, we considered the ligand−protease contacts with $[WOS]_{L_i}{}^{M_i}$ greater than 0.5 Å^2. For these complexes, the results for the two chains are listed separately. This table gives the atom in each residue that is involved in binding of the ligand and also the nature of interaction (favorable, indicated by bold font, or unfavorable). By adding the $[OSC]^{M_i}$ for each residue in each complex, it is possible to identify the residues that make favorable and unfavorable contacts with the ligand. In Table 4, we have listed for each chain the favorably and unfavorably interacting residues with the ligand at this site.

On the basis of the primed and unprimed groups as shown in Figure 5, we renamed the protein chains as A that interacts with the unprimed part of the ligand and B that interacts with the primed part of the ligand. From Table 4, it is obvious that these five ligands do not bind symmetrically to the protease homodimer. For chain A, the binding region on the protease can be grouped into four regions: around residue 8, a stretch of resides 23−32 near the active site, another stretch of residues in the flap region 47−50, and a stretch of residues 80−84. The four regions are the same for chain B except that the flap region has one more residue. For both chains, the majority of the unfavorably interacting residues fall in the first three regions. The majority of the residues listed in Table 4 are shown to be involved in protease drug resistance.

What is the cutoff distance to account for all important contacts between a protein and a ligand in a protein−ligand complex? Table 5 shows the total number of ligand−protease distances whose $[WOS]_{L_i}{}^{M_i}$ greater than 0.0 Å^2, $[WOS]_{L_i}{}^{M_i}$ greater than 0.5 Å^2 and distance greater than 4.1 Å, and $[WOS]_{L_i}{}^{M_i}$ greater than 0. 5 Å^2 are listed for the five protease−ligand complexes. When $[WOS]_{L_i}{}^{M_i}$ is greater than 0.0 Å^2, we have included some of the contacts that are not critical to the binding of the ligand. When $[WOS]_{L_i}{}^{M_i}$ is greater than 0.5 Å^2, the number of contact distances is reduced drastically by about 60%. In the analysis of reported ligand−protein crystal structure complexes,[4] the cutoff distance used to calculate the contacts between the ligand and protein was 4.1 Å (only van der Waals contact). The ligand binding pocket is not rigid because these complexes are formed and reformed in a dynamic state. So one has to consider the contacts between the ligand and the protein atoms greater than the corresponding van der Waals contacts. If we calculate the number of contact distances that are greater than 4.1 Å and also $[WOS]_{L_i}{}^{M_i}$ greater than 0.5 Å^2, the value listed in the fourth column of Table 5 shows that almost 40% of the contacts whose distances are greater than 4.1 Å are also important in the binding of the ligands. For each complex, the longest distance in Å for the ligand−protease contacts with $[WOS]_{L_i}{}^{M_i}$ greater than 0.5 Å^2 is listed in the last column of Table 5. The important ligand−protease contacts are within a distance of 5.5 Å because the binding specificity and majority of the binding strength of a ligand is due to hydrogen bonds and hydrophobic interactions with the macromolecule site.

Visualization of Ligand−Protease Contacts. To display the atoms that are in contact for each ligand atom, we used the program MidasPlus.[16] To display the results of our calculations for any ligand atom (for example C1 in the 1hpx complex), in MidasPlus, we have to type the command “source the atom name (C1)”. Figure 6 shows a graphical output of the results for a ligand atom, C1, of the ligand in the complex (1hpx). The ligand KNI-272 is shown by thick green lines. The molecular surface for the atom C1 is shown by colored dots. The details of the interactions are shown on the window (right side) of Figure 6. In this window, the last column gives the occluded surface constant [OSC] for each contact and the last line gives the [NOSC] and [SEC] for this atom C1. The protease atoms that have favorable contacts (with the negative value of [OSC]) are color-coded by thick cyan lines. The other atoms that are in the same residue as the occluding atoms are represented by thick gray lines. The atom O of 149G makes unfavorable contact with the ligand atom C1. This atom is represented by thick magenta lines. If a dot of the surface of the ligand atom C1 makes favorable occluded contact, then it is colored cyan; if the dot makes unfavorable occluded contact, then it is colored magenta. The nonoccluded surface dots are represented by yellow dots. Thus, the molecular surface of the atom C1 is divided into favorable (cyan), unfavorable (magenta), and nonoccluded (yellow) surface patches.

Mapping Ligand−Protease Contacts onto the Ligand Molecular Surface. For each ligand−protease atom contact, the occluded dots are grouped into a patch. The dots in each patch are color-coded according to the value of $[OSC]_{L_i}{}^{M_i}$. Figure 7a shows a view of color-coded surface of the ligand from the flap region of the 1hiv complex. The color coding is from cyan to red for the occluded surface dots and yellow for the nonoccluded surface dots. The cyan surface patches make the most favorable contacts, followed by blue, purple, magenta, and red (the most unfavorable contacts). The two cyan patches are due to hydrogen bonds between the ligand

Table 3. Occluded Surface Constant for Each Protease Atom of the Five Ligand−Protease Complexes

occluding protein residue			[OSC]^{M_i}					occluding protein residue			[OSC]^{M_i}				
amino acid	sequence	atom name	1hiv	1hpv	1hpx	1hvi	1hv	amino acid	sequence	atom name	1hiv	1hpv	1hpx	1hvi	1hv
			Chain A								**Chain B**				
R	**8**	**CD**	**−0.03**	**0.00**	**0.00**	**−0.09**	**−0.09**	**R**	**108**	**CD**	**0.00**	**0.00**	**0.00**	**−0.06**	**−0.05**
R	**8**	**CZ**	**−0.22**	**0.00**	**−0.15**	**−0.32**	**−0.29**	**R**	**108**	**CZ**	**−0.14**	**0.00**	**0.00**	**−0.38**	**−0.33**
R	**8**	**NE**	**0.00**	**0.00**	**0.00**	**−0.15**	**−0.14**	**R**	**108**	**NE**	**0.00**	**0.00**	**0.00**	**−0.05**	**0.00**
R	**8**	**NH1**	0.07	**0.00**	0.11	0.08	**0.00**	**R**	**108**	**NH1**	0.13	**0.00**	**0.00**	0.13	0.12
R	**8**	**NH2**	0.18	**0.00**	0.14	0.36	0.34	**R**	**108**	**NH2**	0.11	0.16	**0.00**	0.36	0.41
L	**23**	**CD1**	**0.00**	**0.00**	**0.00**	**−0.03**	**−0.02**	**L**	**123**	**CD2**	**−0.08**	**−0.09**	**0.00**	**−0.10**	**−0.12**
L	**23**	**CD2**	**−0.08**	**−0.09**	**−0.09**	**−0.18**	**−0.18**	**D**	**125**	**CG**	0.31	0.33	0.34	0.32	0.24
D	**25**	**CG**	0.53	0.39	0.63	0.39	0.38	**D**	**125**	**OD1**	**0.00**	**−0.29**	**−0.28**	**−0.23**	**−0.23**
D	**25**	**OD1**	**−0.38**	**−0.28**	**−0.20**	**−0.27**	**−0.26**	**D**	**125**	**OD2**	**−0.03**	**−0.17**	0.24	**−0.37**	**0.10**
D	**25**	**OD2**	**−0.34**	0.23	**−0.21**	0.12	0.24	**G**	**127**	**C**	0.23	**−0.10**	0.78	0.84	0.85
G	**27**	**C**	0.77	0.71	0.07	**−0.07**	**−0.11**	**G**	**127**	**CA**	**−0.01**	**−0.04**	0.10	0.07	0.07
G	**27**	**CA**	0.06	0.08	**−0.05**	0.05	0.04	**G**	**127**	**O**	**−0.31**	0.32	**−0.25**	0.41	0.41
G	**27**	**O**	**−0.07**	**−0.09**	0.09	**−0.04**	**−0.45**	**A**	**128**	**CA**	0.27	**−0.09**	0.27	0.12	0.10
A	**28**	**CA**	0.12	0.14	0.21	0.26	0.23	**A**	**128**	**CB**	**−0.06**	**−0.13**	**−0.23**	0.32	0.17
A	**28**	**CB**	0.26	**−0.07**	0.17	**−0.24**	**−0.22**	**D**	**129**	**CG**	0.27	**0.00**	0.35	**−0.36**	**−0.41**
D	**29**	**CG**	0.35	**0.00**	**0.00**	**−0.01**	**−0.02**	**D**	**129**	**N**	**−0.56**	**0.08**	**−0.49**	**−0.44**	**−0.43**
D	**29**	**N**	**−0.65**	0.19	**0.00**	**−0.39**	**−0.40**	**D**	**129**	**OD1**	**0.00**	**0.00**	**0.00**	0.21	0.05
D	**29**	**OD1**	**0.00**	**0.00**	**0.00**	0.23	**0.00**	**D**	**129**	**OD2**	0.07	**0.00**	**0.00**	0.05	0.05
D	**29**	**OD2**	**0.00**	**0.00**	**0.00**	0.03	0.05	**D**	**130**	**C**	**0.00**	0.22	**−0.05**	**0.00**	**0.00**
D	**30**	**C**	**−0.06**	**0.00**	**0.00**	**0.00**	**0.00**	**D**	**130**	**CB**	**−0.03**	0.15	**−0.02**	**−0.03**	**−0.03**
D	**30**	**CB**	**0.00**	**0.00**	**0.00**	**−0.03**	**−0.03**	**D**	**130**	**CG**	**0.00**	0.14	**0.00**	**0.00**	**0.00**
D	**30**	**N**	0.16	0.12	**0.00**	**0.00**	**0.00**	**D**	**130**	**N**	**0.00**	0.54	0.07	**0.00**	**0.00**
D	**30**	**O**	0.08	**0.00**	**0.00**	**0.00**	**0.00**	**D**	**130**	**O**	0.05	0.11	0.07	**0.00**	**0.00**
V	**32**	**CG2**	**−0.09**	**−0.03**	**−0.15**	**−0.13**	**−0.11**	**D**	**130**	**OD2**	**0.00**	**−0.20**	**0.00**	**0.00**	**0.00**
I	**47**	**CB**	**−0.03**	**0.00**	**−0.02**	**−0.01**	**−0.01**	**V**	**132**	**CG1**	**0.00**	**−0.08**	**0.00**	**0.00**	**0.00**
I	**47**	**CD1**	**−0.12**	**0.00**	**−0.12**	**−0.05**	**−0.06**	**V**	**132**	**CG2**	**−0.11**	**0.00**	**−0.08**	**−0.12**	**−0.12**
I	**47**	**CG2**	**−0.06**	**−0.05**	**−0.03**	**−0.03**	**0.00**	**I**	**147**	**CB**	**−0.01**	**−0.01**	**0.00**	**−0.01**	**−0.01**
G	**48**	**C**	**−0.18**	**0.00**	**−0.12**	**−0.04**	**−0.03**	**I**	**147**	**CD1**	**−0.11**	**−0.09**	**−0.08**	**−0.06**	**−0.07**
G	**48**	**CA**	**−0.04**	**0.00**	**0.00**	**0.00**	**0.00**	**I**	**147**	**CG2**	**−0.03**	0.13	**0.00**	**−0.03**	**0.00**
G	**48**	**O**	0.65	0.18	0.13	**−0.47**	**−0.74**	**G**	**148**	**C**	**0.00**	**0.00**	**−0.17**	**−0.03**	**−0.04**
G	**49**	**C**	**−0.22**	**−0.14**	**−0.05**	**−0.13**	**−0.14**	**G**	**148**	**CA**	**0.00**	**0.00**	**−0.01**	**0.00**	**0.00**
G	**49**	**CA**	0.20	**−0.10**	**−0.06**	**−0.03**	0.05	**G**	**148**	**O**	**−0.36**	0.10	**−0.10**	**−0.33**	**−0.28**
G	**49**	**N**	0.07	**0.00**	**0.00**	**0.00**	**0.00**	**G**	**149**	**C**	**0.00**	**−0.06**	**−0.15**	**−0.14**	**−0.12**
G	**49**	**O**	0.12	**0.00**	**0.00**	**0.00**	**0.00**	**G**	**149**	**CA**	0.08	0.23	**−0.05**	0.16	0.12
I	**50**	**CA**	**−0.04**	**0.00**	**0.00**	**0.00**	**0.00**	**G**	**149**	**O**	**0.00**	**0.00**	0.04	**0.00**	**0.00**
I	**50**	**CB**	**0.00**	0.08	0.04	0.02	0.02	**I**	**150**	**CB**	0.03	0.03	0.02	0.01	0.02
I	**50**	**CD1**	**0.00**	0.10	**−0.05**	**−0.05**	**−0.09**	**I**	**150**	**CD1**	0.05	0.04	**−0.02**	**−0.11**	**−0.05**
I	**50**	**CG1**	**−0.04**	0.00	**−0.02**	**−0.02**	**0.00**	**I**	**150**	**CG1**	**−0.04**	**−0.05**	**−0.08**	**0.00**	**−0.03**
I	**50**	**CG2**	**−0.12**	0.19	0.07	**0.00**	**0.00**	**I**	**150**	**CG2**	**−0.03**	0.06	**−0.06**	0.10	**0.00**
I	**50**	**N**	0.07	**0.00**	0.06	0.15	0.15	**I**	**150**	**N**	0.15	**−0.27**	**0.13**	0.07	0.07
F	**53**	**CE1**	**−0.03**	**0.00**	**0.00**	**0.00**	**0.00**	**F**	**153**	**CE1**	**0.00**	**0.00**	**−0.02**	**0.00**	**0.00**
F	**53**	**CZ**	**−0.05**	**0.00**	**0.00**	**0.00**	**0.00**	**F**	**153**	**CZ**	**0.00**	**0.00**	**−0.03**	**0.00**	**0.00**
P	**81**	**C**	**−0.06**	**0.00**	**0.00**	**0.00**	**0.00**	**T**	**180**	**CG2**	**−0.04**	**−0.03**	**0.00**	**0.00**	**0.00**
P	**81**	**CB**	**−0.05**	**0.00**	**−0.08**	**−0.03**	**−0.03**	**P**	**181**	**CB**	**0.00**	**−0.03**	**0.00**	**−0.03**	**−0.05**
P	**81**	**CD**	**−0.02**	**0.00**	**−0.03**	**−0.02**	**−0.02**	**P**	**181**	**CD**	**−0.02**	**−0.03**	**0.00**	**−0.03**	**−0.03**
P	**81**	**CG**	**−0.13**	**−0.06**	**−0.10**	**−0.04**	**−0.04**	**P**	**181**	**CG**	**−0.14**	**−0.04**	**−0.04**	**−0.04**	**−0.03**
V	**82**	**CB**	**0.00**	**−0.02**	**0.00**	**−0.01**	**−0.01**	**V**	**182**	**CB**	**0.00**	**0.00**	**0.00**	**−0.01**	**−0.01**
V	**82**	**CG1**	**−0.05**	**0.00**	**−0.19**	**−0.24**	**−0.25**	**V**	**182**	**CG1**	**−0.08**	**0.00**	**−0.05**	**−0.19**	**−0.23**
V	**82**	**CG2**	**−0.20**	**−0.11**	**0.00**	**−0.04**	**−0.04**	**V**	**182**	**CG2**	**0.00**	**−0.09**	**0.00**	**−0.03**	**−0.03**
I	**84**	**CD1**	0.25	**−0.31**	**−0.25**	**−0.07**	**−0.04**	**I**	**184**	**CD1**	**−0.30**	−0.05	**−0.18**	**−0.05**	**−0.23**
I	**84**	**CG1**	**0.00**	**0.00**	**0.00**	**−0.09**	**−0.11**	**I**	**184**	**CG1**	**−0.01**	**0.00**	**0.00**	**−0.01**	**0.00**
I	**84**	**CG2**	**−0.08**	**−0.07**	**−0.06**	**−0.06**	**−0.05**	**I**	**184**	**CG2**	**−0.04**	**−0.06**	**−0.03**	0.06	**−0.10**

and the protease site. The hydrogen bonds are small but favorable, whereas the hydrophobic interactions are less favorable than the hydrogen bonds, but the surface areas of occlusion of hydrophobic atoms are more than that of the occluded surface areas for the hydrophilic atoms. As described earlier, the unprimed and primed groups are marked as P1, P1′, etc. This ligand is making a number of favorable contacts with the protease binding site particularly between P1′ and P3′ and between P1′ and P2. Figure 7b shows the color-coded molecular surface of the backside view of Figure 7a. This side of the ligand, which is closer to the active site of the protease molecule, makes more favorable contacts with the site than the other side (Figure 7a). The four favorable cyan patches are due to hydrogen bonds. The yellow patch near the top right-hand corner of the figure does not make any contact with the protease. So, if we want to improve the binding of this ligand, one can modify this region of the ligand. The molecular surface between P1 and P2′ makes good hydrophobic contacts with the protease site.

We superimposed the C-α atoms of the protease structures using the 1hiv protease structure as a template. In superimposing a structure with the template protease structure, only C-α atoms that have a difference in each of their backbone torsion angles (ϕ

Table 4

a. Favorably Occluding Residues in the Five Ligand−Protease Complexes

	amino acid sequence of favorably occluding residues (chain A)															
PDB code	R	L	D	G	A	D	D	V	I	G	G	I	T	P	V	I
1hiv	–	23	–	27	–	29	–	32	47	48	–	–	80	81	82	84
1hpv	–	23	25	–	28	–	–	32	–	–	–	50	80	81	82	84
1hpx	–	23	–	–	–	–	–	32	47	–	49	–	–	81	82	84
1hvi	8	23	–	27	–	29	30	32	47	48	49	–	–	81	82	84
1hvj	8	23	–	27	–	29	30	32	47	48	49	–	–	81	82	84

	amino acid sequence of favorably occluding residues (chain B)												
PDB code	L	D	D	D	V	I	G	G	I	F	P	V	I
1hiv	123	125	129	–	132	147	–	–	150	153	181	182	–
1hpv	123?	–	–	–	132	147	–	149	–	–	181	182	184
1hpx	–	–	129	–	132	147	148	149	150	153	181	182	184
1hvi	123	125	129	130	132	147	148	–	–	–	181	182	–
1hvj	123	–	129	130	132	147	148	–	–	–	181	182	184

b. Unfavorably Occluding Residues in the Five Ligand−Protease Complexes

	amino acid sequence of unfavorably occluding residues (chain A)									
PDB code	R	D	G	A	D	D	I	G	G	I
1hiv	8	25	–	28	–	30	–	–	49	50
1hpv	8	–	27	–	29	30	47	48	49	–
1hpx	8	25	27	28	–	–	–	48	–	50
1hvi	–	25	–	28	–	–	–	–	–	50
1hvj	–	25	–	28	–	–	–	–	–	50

	amino acid sequence of unfavorably occluding residues (chain B)									
PDB code	R	D	G	A	D	D	G	G	I	I
1hiv	108	–	127	128	–	130	148	149	150	184
1hpv	–	125	127	128	129	130	148	–	150	–
1hpx	–	125	127	128	–	130	–	–	–	–
1hvi	108	–	127	128	–	–	–	149	150	184
1hvj	108	125	127	128	–	–	–	149	150	–

Table 5. Analysis of the Number of Ligand−Protease Contacts

	number of contact distances			
PDB code	$[WOS]_{L_i}^{M_i}$ > 0.0 (Å^2)	$[WOS]_{L_i}^{M_i}$ > 0.5 (Å^2)	distance[a] > 4.1 (Å)	longest contact distance[a] (Å)
1hiv	457	160	55	4.9
1hpv	365	133	60	5.4
1hpx	333	145	64	5.3
1hvi	462	175	76	5.4
1hvj	457	169	74	5.5

[a] For $[WOS]_{L_i}^{M_i}$ > 0.5 (Å^2).

and ψ) of less than 20° were used. Figure 8 shows the results of the superimposition. Only the superimposed ligands are shown as color-coded stick representations (magenta for 1hiv, cyan for 1hpv, gray for 1hpx, greenish-yellow for 1hvi, and light brown for 1hvj). Also, we have displayed surface dots, which make favorable ligand−protease contacts. These surface dots are color-coded cyan, blue, purple, magenta, and red—the cyan dot being the most favorable contact and the red the least favorable (not unfavorable) contact. The superimposition of many ligand−protein complexes and the generation of common occluded surface patches will help investigators to generalize the nature of ligand−protein contacts for chemically divergent ligands. Also, this will help us to develop a few 3D pharmacophores using the cyan patches and the other favorable hydrophobic patches. These 3D pharmacophores could be used in searching for inhibitors in chemical database of small organic compounds.

Rank Ordering of Ligands. Is it possible to use the occluded surface-based constants to rank order the ligands binding to the same target? In Table 6 we have listed $[OSC]_L$, $[NOSC]_L$, the ligand total constant ($[LBC] = [OSC]_L + [NOSC]_L$), the solvent-exposed constant $[SEC]_L$, and the experimentally observed inhibitory constants for all the five ligands bound to HIV-1 protease. From the occluded surface constant, the ligand of 1hvj binds the strongest compared to other ligands. For the 1hpv ligand, since the $[OSC]_L$ is positive, it does not make favorable contacts with the protease. Also the 1hiv ligand slightly favors binding. For all the complexes, the nonoccluded surface constants ($[NOSC]_L$) of the ligands are positive. This means that some of the ligand atoms do not make good contacts with the protease. From the ligand total constant [LBC], the rank order of binding of the ligands agrees with the experimental data except for 1hpv. When the ligands are exposed to solvent, the $[SEC]_L$ is also positive because of the number of hydrophobic side chains present in these five ligands. On comparing the chemistry of these ligands, 1hpv ligand is less hydrophobic than the other ligands. This is clearly seen in the value of $[SEC]_L$ for 1hpv ligand (0.12) compared to 1hiv, 1hpx, 1hvi, and 1hvj ligands with $[SEC]_L$ values ranging from 18.2 to 26.1. Also between 1hvi and 1hvj ligands, 1hvi contains one extra hydroxyl group. The $[SEC]_L$ for 1hvi is less than that of 1hvj. The strength of binding of a ligand into the active site is the difference between [LBC] and $[SEC]_L$. The difference, which is listed in the sixth column in Table 6, predicts the best binder from the worst binder. The binding of either ligand of 1hpv or 1hiv to their respective site is not favorable. As described earlier and considering each residue, we also calculated the macromolecule (protease) stability con-

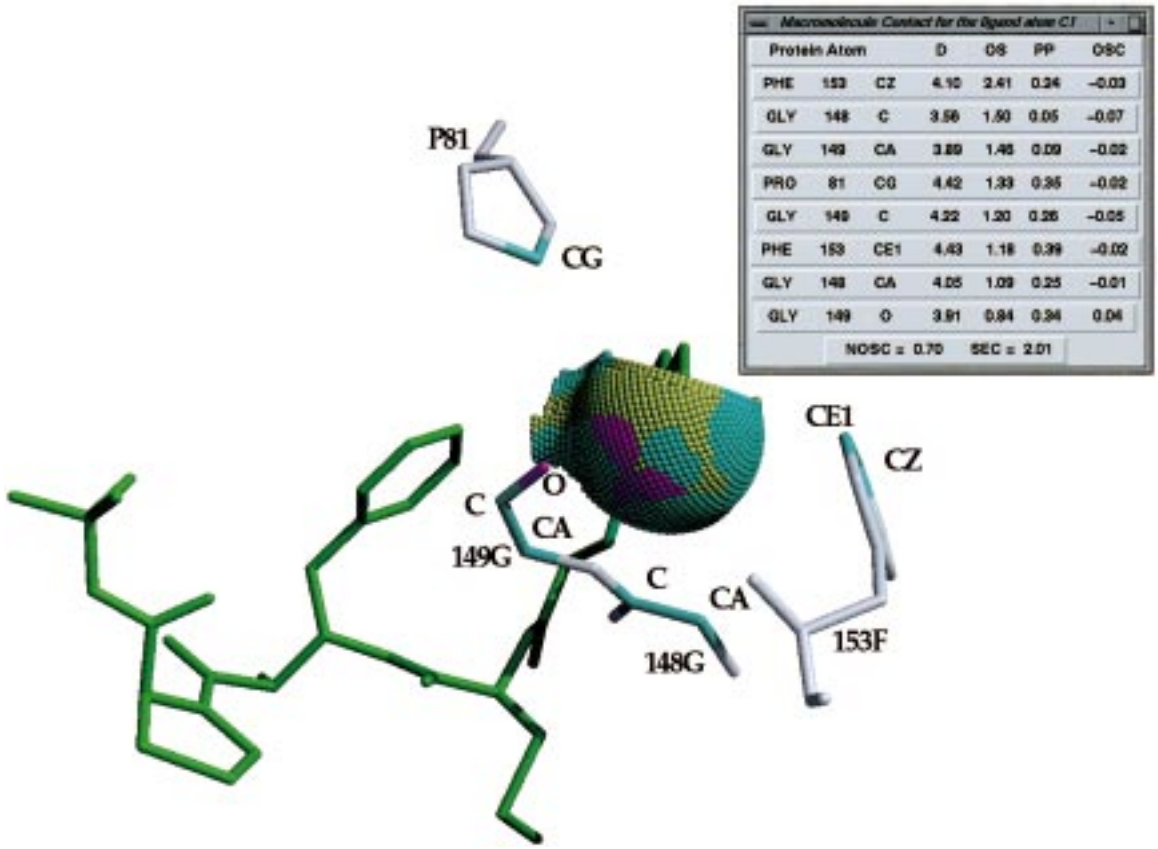

Figure 6. Visual representation of the occlusion of a ligand atom C1. Green thick lines represent the ligand. The cyan and magenta lines are the protease atoms involved in occluding the ligand atom. The other atoms in the occluding residues are thick gray lines. The molecular surface of the ligand atom C1 is represented by color-coded dots. Magenta color denotes unfavorable contact and cyan favorable contact. The window near the right-hand corner lists the actual value of the distance in Å, the occluded surface area [OS] in Å^2, the packing parameter [PP], and the occluded surface constant [OSC].

stant [MSC] without the ligand bound in order to determine the stability of the protease structures in the complexes. The [MSC] values are listed in the seventh column of Table 6. The protease structures for 1hvi and 1hvj complexes are much more stable than those of 1hiv, 1hpv, and 1hpx. In particular, the protease structure of 1hiv complex has to give-up favorable contacts between the residues in order to bind to the ligand. Taking into consideration the values listed in column 5 and [MSC], the rank ordering of these ligands except for ligand 1hpx qualitatively agrees with the experimental inhibitory constants. This may be due to entropic contribution to the ligand binding of the ligand in the 1hpx complex.

The binding of a ligand to a macromolecular target depends on (a) how often the ligand and the protein adopt the required conformation for binding? (b) how fast the ligand and the protein can desolvate in the conformation before binding? (c) how tight the ligand can bind to the site? and (c) how long the ligand can stay in the pocket? For these sets of compounds, the hydrogen bond interactions play some roll in the specificity of ligand binding. The solvation plays an important role for the binding of these ligands because of a number of hydrophobic side chains. The ligand in the 1hpv complex binds poorly to the pocket, and also the solvation is not favoring the ligand to stay in the pocket. On the other hand, the ligand in the 1hiv complex, forming more hydrogen bonds than the ligand in the 1hpv complex, binds slightly better than 1hpv. Also, the solvation of this ligand in the uncomplexed over the complexed state does not help the ligand to stay in the pocket. The protein (of 1hiv) has to change its conformation significantly to bind and hold on to the ligand.

Our method of calculating binding of ligands to a macromolecule site using occluded surfaces is different from other surface-based methods such as the program hydropathic interactions (HINT[21]) and the atomic solvation parameters (ASP) based on solvent-accessible surface area. In the case of HINT, for the calculation of the total interaction constant (TIC), the program uses the total average accessible van der Waals molecular surfaces of each atom in the ligand and also uses the empirically derived hydrophobic atom constants calculated from the solvent partition constants of small molecules. As pointed out earlier in this paper (refer to Table 1 and Figure 6) the molecular surface of each atom in a ligand could be divided into a few occluded surface patches, depending upon the contacts. The partial atomic charges have the necessary information about the chemical nature of the atom such as charged, polar, hydrophobic (small value of charge). In calculating the TIC, HINT does not include the strain energy involved with the macromolecule. In the ligand-binding studies, the macromolecule can change its conformation to bind to the ligand. As pointed out earlier by Ajaj and Murcko,[22] the ASP-based methods have the following major objections. First of all, the hydration of the polar groups on the surface is different from the buried polar groups. Second, the atoms were grouped into limited

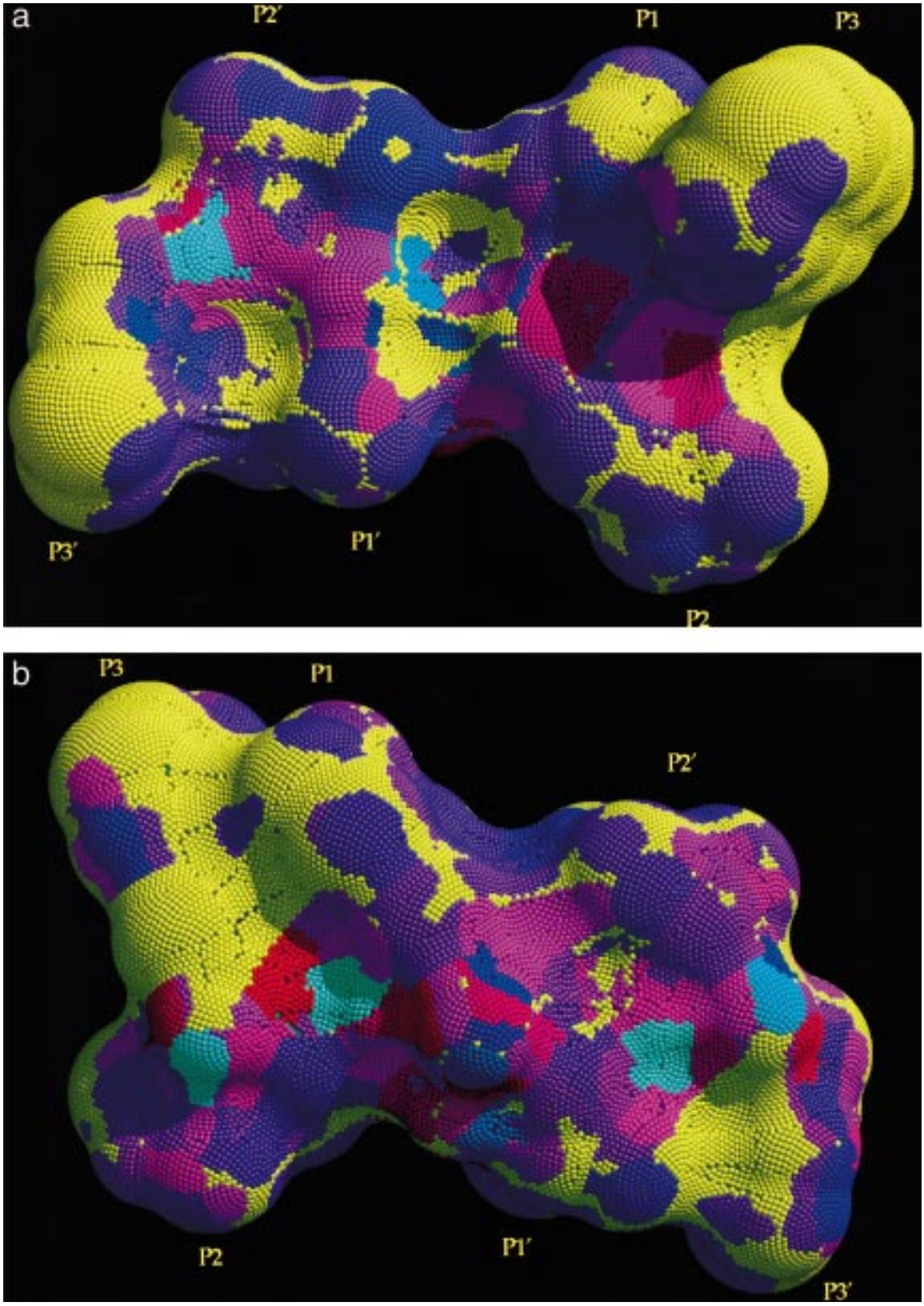

Figure 7. (a) Color-coded molecular surface of the ligand in the 1hiv complex. Cyan dots are the most favorable and red dots the least favorable contact regions. (b) Backside molecular surface view of panel a.

(five) atom types, and the parameters were derived using a training set of molecules of known (limited) number of crystal structures. In our method, we treat each atom (polar and nonpolar) differently depending upon its surrounding. The method described in this paper takes into account the various atom types by the partial atomic charges and also the chemical nature of the ligand atoms—acceptor, donor, both acceptor and donor, or greasy (hydrophobic). In our method, we do not derive the parameters based on a training set of molecules. Since we did not use any nonbonded term, our method,[23] currently, can be applied either to high-

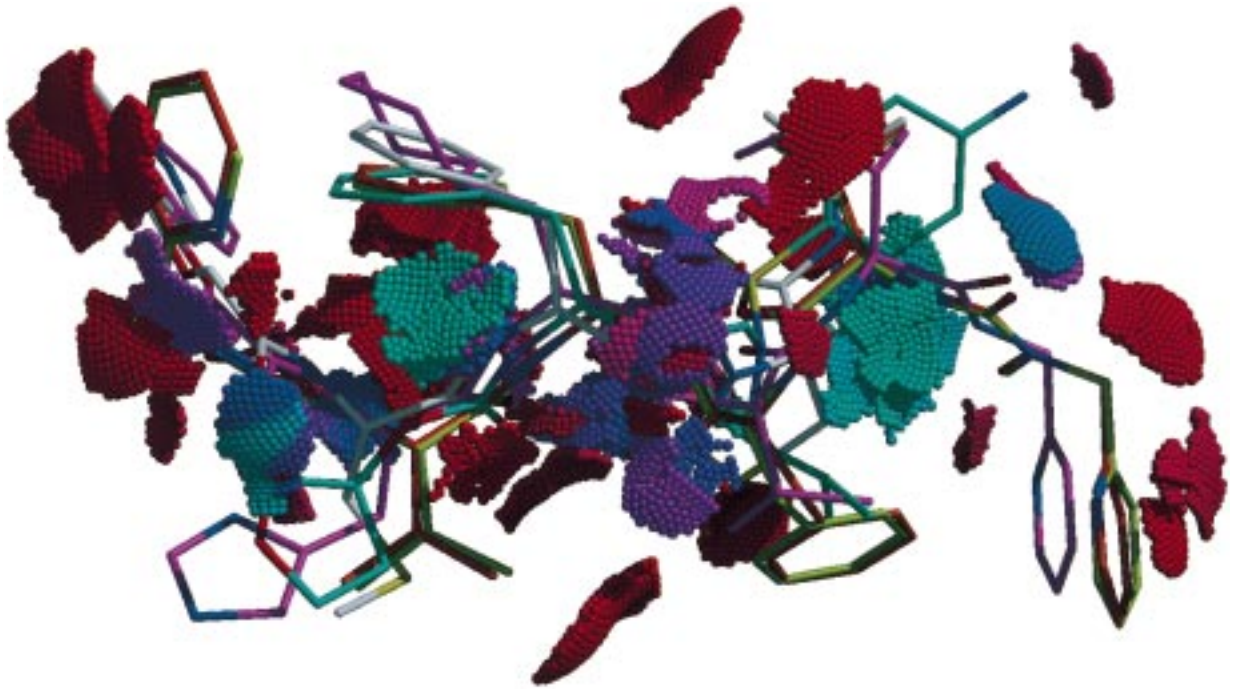

Figure 8. Superimposition of the five ligand−protease complexes. Only ligands are shown as color-coded bonds (magenta for 1hiv, cyan for 1hpv, gray for 1hpx, greenish-yellow for 1hvi, and light brown for 1hvj). For each ligand, only favorable contact dots are displayed. The cyan and red dots are respectively the most favorable and least favorable contacts of the favorable contacts.

Table 6. Occluded Surface-Based Ligand Constants for the Five Ligand−Protease Complexes

PDB code	$[OSC]_L$	$[NOSC]_L$	[LBC]	$[SEC]_L$	([LBC] − $[SEC]_L$)	[MSC]	K_i[a] (pM)
1hiv	−0.21	19.94	19.73	18.21	1.52	−307.53	2000.0
1hpv	1.10	2.99	4.08	0.12	3.96	−321.61	600.0
1hpx	−0.82	12.26	11.44	18.82	−7.38	−356.72	5.5
1hvi	−2.62	17.58	14.96	22.66	−7.70	−371.42	84.0
1hvj	−3.10	17.71	14.60	26.12	−11.52	−373.31	35.0

[a] Data taken from ref 17 for 1hiv, from ref 18 for 1hpv, from ref 19 for 1hpx, and from ref 20 for 1hvi and 1hvj.

resolution and well-refined X-ray crystal structures or to energy-minimized ligand−macromolecule complexes.

Conclusions

In this paper, the occlusion surface algorithm was used to quantify and visualize HIV-1 protease−ligand contacts in crystal structure complexes. This method is based on the occlusion of the van der Waals Connolly's molecular surface of ligand atoms by the van der Waals spheres of protein atoms surrounding the ligand. Unlike the method reported by Sobolev et al.,[11] our method, described in this paper, does not use an artificially expanded van der Waals sphere to calculate the surface in contact. We used the Connolly's van der Waals molecular surface and the normal to scoop out surrounding atoms, which are within a distance of 6.4 Å from any ligand atom. Using the partial charges on the ligand and protease atoms that are in contact and the occluded and nonoccluded surface areas, we have not only identified the favorable and unfavorable contacts but also the strength of the contacts.

Using our method of displaying ligand−macromolecule contacts, medicinal chemists would be able to focus on the "local" binding site of each atom in the ligand. In case medicinal chemists were not able to access a 3D graphics display system, they would be able to qualitatively estimate the nature of contacts each ligand atom makes with the macromolecule, using the ligand−protein contacts as shown in Table 1. We have clearly shown that the cutoff distance to identify critical ligand−protein contacts is 5.5 Å. From the residue-based ligand−macromolecule contacts (refer to Table 3), molecular biologists would be able to design rational protein design experiments to study the effect of amino acid change on ligand binding. In searching for new inhibitor leads for an enzyme, one can not only use the pharamcophore based on ligands but also look for conformations of ligands that would generate some of the critical surface patches as shown in Figure 8. This could help to reduce the number of compounds selected from a database of small molecule compounds. Unlike other surface-based methods, the occluded surface-based method clearly decomposes the binding of ligand into the occluded surface, nonoccluded surface, and solvent-exposed constants. Finally, we were able to qualitatively rank order the binding of the five protease ligands using the X-ray structure of the complexes. Further work is needed to incorporate the occluded surface-based method into molecular mechanics and dynamics calculations and structure-based ligand design programs.

Acknowledgment. The author acknowledges the National Cancer Institute for allocation of computing time and staff support at the Frederick Biomedical Supercomputing Center of the Frederick Cancer Research and Development Center. This project has been funded in whole or in part with federal funds from the

National Cancer Institute, National Institutes of Health, under Contract No. N01-CO-56000. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the United States Government.

References

(1) Kollman, P. A. Theory of Macromolecule-Ligand Interactions. *Curr. Opin. Struct. Biol.* **1994**, *4*, 240−245.
(2) Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeing, J.; Darby, G.; Jones, Y.; Stuart, D.; Stammers, D. High Resolution Structures of HIV-1 RT from Four RT-Inhibitor Complexes. *Nature Struct. Biol.* **1995**, *2*, 293−302.
(3) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPROT: a Program to Generate Schematic Diagrams of Protein−Ligand Interactions. *Protein Eng.* **1995**, *8*, 127−134.
(4) Wlodawer, A.; Erickson, J. W. Structure-based Inhibitors of HIV-1 Protease. *Annu. Rev. Biochem*. **1993**, *62*, 548−585.
(5) Bahar, I.; Jernigan, R. L. Interresidue Potentials in Globular Proteins and the Dominance of Highly Specific Hydrophilic Interactions at Close Separation. *J. Mol. Biol.* **1977**, *266*, 195−214.
(6) Lee, B.; Richards, F. M. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55*, 379−400.
(7) Esienberg, D.; McLachlan, A. D. Solvation Energy in Protein Folding and Binding. *Nature* **1986**, *319*, 188−203.
(8) Wallqvist, A.; Jernigan, R. L.; Covell, D. G. A Preference-Based Free- Energy Parametrization of Enzyme−Inhibitor Binding. Applications to HIV-1 Protease Inhibitor Design. *Protein Sci.* **1995**, *4*, 1881−1903.
(9) Horton, N.; Lewis, M. Calculation of the Free Energy of Association for Protein Complexes. *Protein Sci.* **1992**, *1*, 169−181.
(10) Gregoret, L. M.; Cohen, F. E. Novel Method for the Rapid Evaluation of Packing in Protein Structures. *J. Mol. Biol*. **1990**, *211*, 959−974.
(11) Sobolev, V.; Wade, R. C.; Vriend, G.; Edelman, M. Molecular Docking Using Surface Complementarity. *PROTEINS: Struct. Funct. Genet.* **1996**, *25*, 120−129.
(12) Pattabiraman, N.; Ward, K. B; Fleming, P. J. Occluded Molecular Surface: Analysis of Protein Packing. *J. Mol. Recognit.* **1995**, *8*, 334−344.
(13) DeDecker, B. S.; O'Brien, R.; Fleming, P. J.; Geiger, J. H.; Jackson, S. P.; Sigler, P. B. The Crystal Structure of a Hyperthermophilic Archaeal TATA-Box Binding Protein. *J. Mol. Biol.* **1996**, *264*, 1072−1084.
(14) Fleming, K. G.; Ackerman, A. L.; Engelman, D. M. The Effect of Point Mutations on the Free Energy of Transmembrane a-Helix Dimerization. *J. Mol. Biol.* **1997**, *272*, 266−275.
(15) Ratnaparkhi, G. S.; Ramachandran, S.; Udganokar, J. B.; Varadarajan, R. Discrepancies Between the NMR and X-ray Structures of Uncomplexed Barstar: An Analysis Suggests that Packing Density of Protein Structures Determined by NMR Are Unreliable. *Biochemistry* **1998**, *37*, 6958−6966.
(16) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. The MIDAS Display System. *J. Mol. Graph*. **1988**, *6*, 13−27.
(17) Thaisrivongs, S.; Watenpaugh, K. D.; Howe, W. J.; Tomich, P. K.; Lester, A. D.; Chong, K.-T.; Tomich, C.-S. C.; Tomasselli, A. G.; Turner, S. R.; Strohbach, J. W.; Mulichak, A. M.; Janakiraman, M. N.; Moon, J. B.; Lynn, J. C.; Horng, M.-M.; Hinshaw, R. R.; Curry, K. A.; Rothrock, D. J. Structure-Based Design of Novel HIV Protease Inhibitors: Carboxamide-Containing 4-Hydroxycoumarins and 4-Hydroxy-2-Pyrones as Potent Nonpeptide Inhibitors. *J. Med. Chem*. **1995**, *38*, 3624−3637.
(18) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. Crystal structure of HIV-1 Protease in Complex with VX-478, a Potent and Orally Bioavailable Inhibitor of the Enzyme. *J. Am. Chem. Soc.* **1995**, *117*, 1181−1182.
(19) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Liu, B.; Topol, I. A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. W. Structure of HIV-1 Protease with KNI-272, a Tight-Binding Transition-State Analogue Containing Allophenylnotstatine. *Structure* **1995**, *3*, 581−590.
(20) Hosur, M. V.; Bhat, T. N.; Kempf, D. J.; Baldwin, E. T.; Liu, B.; Gulnik, S.; Wideburg, N. E.; Norbeck, D. W.; Appelt, K.; Erickson, J. W. Influence of Stereochemistry on Activity and Binding Modes for C2 Symmetry-Based Diol Inhibitors of HIV-1 Protease. *J. Am. Chem. Soc.* **1994**, *116*, 847−855.
(21) Meng, E. C.; Kuntz, I. D.; Abraham, D. J.; Kellogg, G. E. Evaluating Docked Complexes with the HINT Exponential Function and Empirical Atomic Hydrophobicites. *J. Comput.-Aided Mol. Des*. **1994**, *8*, 299−306.
(22) Ajaj; Murcko, M. A. Computational Methods to predict binding free energy in Ligand−Receptor Complexes. *J. Med. Chem*. **1995**, *38*, 4953−4967.
(23) The program is a collection of several programs and runs on a SGI Indigo2 with an IRIX operating system (version 6.5). To get the charges for the ligand and the macromolecule, Biosym modeling software Insight II (version 97.0) is needed. Contact the author for additional information about the program.

JM980512C