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

Regine S. Bohacek* and Colin McMartin

Pharmaceuticals Division, CIBA-GEIGY Corporation, Summit, New Jersey 07901. *Received May 28, 1991*

The accessible surface, described by Lee and Richards (the L&R surface: *J. Mol. Biol.* 1971, *55*, 379), has remarkably useful properties for displaying ligand-protein interactions. The surface is placed one van der Waals radius plus one probe radius away from the protein atoms. The ligands are displayed in skeletal form. With a suitable probe radius, those parts of the ligand in good van der Waals contact with the protein binding site are found superimposed on the L&R surface. Display of the surface using parallel contours therefore provides a very powerful guide for interactive drug design because only ligand atoms lying on or close to the surface are in low-energy contact. The ability of the surface to accurately display steric complementarity between ligands and proteins was optimized using data from small molecule crystal structures. The possibility of displaying the chemical specificity of the binding site was also investigated. The surface can be colored to give precise information about chemical specificity. Electrostatic potential, electrostatic gradient, and distance to hydrogen-bonding groups were tested as methods of displaying chemical specificity. The ability of these methods to describe the complementarity actually observed in the interior of proteins was compared. High-resolution crystal data for ribonuclease and trypsin was used. The environment surrounding extended peptide chains in the protein was treated as a virtual binding site. The peptide chain served as a virtual ligand. This large sample of experimental data was used to measure the correlation between type of ligand atom and the calculated property of the nearest binding site surface. The best correlation was obtained using hydrogen-bonding properties of the binding site. Using this parameter the surface could be divided into three separate zones representing the hydrophobic, hydrogen-bond-acceptor, and hydrogen-bond-donor properties of the binding site. The percentage of hydrophobic ligand atoms found to lie closest to the hydrophobic protein surface was 91%. The equivalent scores for ligand hydrogen-acceptor atoms and hydrogen-donor atoms found at the corresponding complementarity zone were 94% and 91%. The surface zones can be readily displayed using three colors. To test the method on real ligand/binding site interactions, nine thermolysin-inhibitor complexes of known structure were evaluated using the parameters and criteria derived from the protein-packing study and a correlation between complementary contacts and logarithm of potency was obtained which had an r^2 of 0.99. Stereo images showing a colored contour representation of the L&R surfaces of binding sites of enzymes and skeletal structures of docked ligands provide a graphic summary of key features of a receptor site. This provides a powerful tool for the design of novel ligands.

Introduction

We are now in an era of drug design where experimental techniques, such as X-ray diffraction and solution NMR, are providing structures of target binding sites at atomic resolution. However, even when the binding site has been accurately determined, the design of drugs is still a very complicated process. The interaction between a ligand and a protein binding site is usually highly complex, typically involving several hundred nonbonded contacts between ligand atoms and binding site atoms. Careful examination of the interactions between ligand and binding site can yield a wealth of data (see, for example, Janin and Chothia[1]). However for drug design it is useful to have a graphical tool which summarizes key features of the site which are not readily discernible on visual inspection of the molecules in skeletal form. Visual display of the surface of the protein or binding site can therefore be an extremely helpful aid to understanding the docking of inhibitors and ligands.

Two very different types of molecular surface are currently used in molecular modeling applications (see Figure 1), namely the van der Waals and the Lee and Richards accessible surface (L&R surface). In the first category are surfaces placed one van der Waals radius away from the molecule[3-6] (see Figure 1a). The second type of surface is the solvent-accessible surface of Lee and Richards[2] where the surface is a van der Waals radius plus a probe radius away from the molecule (Figure 1b).

These two types of surface have very different properties for visualizing intermolecular interactions. With a van der Waals surface as shown in Figure 1a the ligand atoms lie a van der Waals radius away from the surface of the enzyme. With this type of surface it is only possible to see if ligand atoms are in van der Waals contact with the enzyme if the van der Waals surface of the ligand is also displayed. Although it is feasible to do this, examination of two three-dimensional surfaces of complex shape for steric complementarity is not simple. It is also difficult to examine the two surfaces for chemical complementarity since both surfaces will be multicolored.

As shown in Figure 1b, the L&R surface indicates the positions of ligand atoms in van der Waals contact. If a ligand is represented in skeletal form all the atoms which are in good van der Waals contact will lie very close to the surface. The surface therefore acts as a powerful guide showing where atoms should be placed to lie in good contact. In addition, if the surface can be colored to show complementarity, it should be easy to see which part of the surface is close to an atom and to examine the surface color for complementarity.

Areas of the surface and pockets in the surface which are not occupied can also be readily seen. These indicate opportunities for extending the ligand. The position of

(1) Janin, J.; Chothia, C. The Structure of Protein-Protein Recognition Sites. *J. Biol. Chem.* **1990**, *27*, 16027-16030.
(2) Lee, B.; Richards, F. M. Interpretation of Protein Structure—Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55*, 379-400.
(3) Richards, F. M. Areas, Volume, Packing and Protein Structures. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151-176.
(4) Connolly, M. J. Solvent-Accessible Surfaces of Proteins and Nucleic Acids. *Science* **1983**, *221*, 709-713.
(5) Weiner, P. K.; Langridge, R.; Blaney, J. M.; Schaeffer, R.; Kollman, P. A. Electrostatic Potential Molecular Surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 3754-3758.
(6) Ho, C. M. W.; Marshall, G. R. Cavity Search: An Algorithm for the Isolation and Display of Cavity-like Binding Region. *J. Comput.-aided Mol. Des.* **1990**, *4*, 337-354.

0022-2623/92/1835-1671$03.00/0 © 1992 American Chemical Society

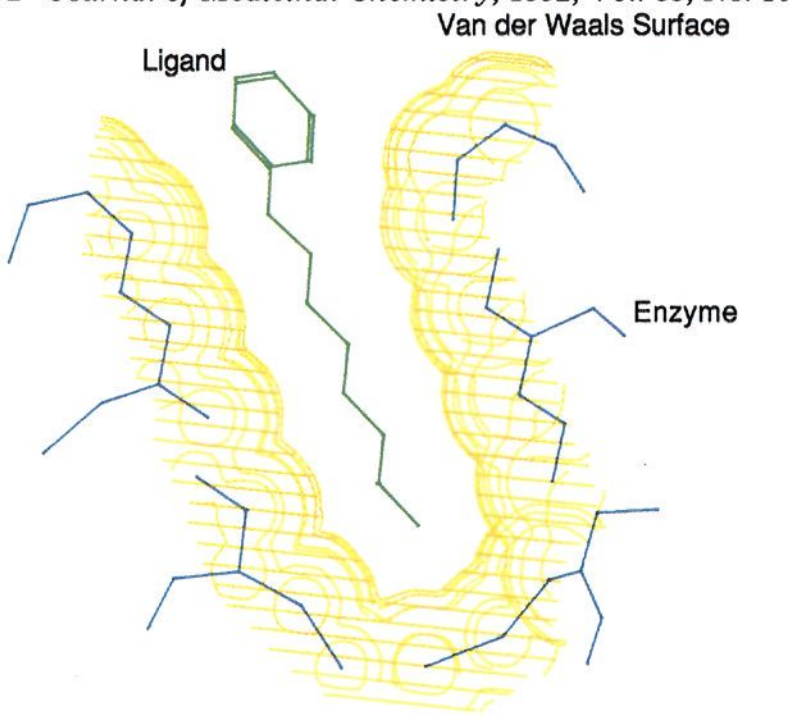

Figure 1. Illustration of the use of binding site surfaces to display steric complementarity between ligand and binding site: (a, top) van der Waals surface (ligand, green; enzyme, blue; surface, yellow); (b, bottom) Lee and Richards accessible surface. The Lee and Richards accessible surface acts as a guide for the positions of the centers of ligand atoms in van der Waals contact with the enzyme.

the surface shows places where the extending groups must lie (and the regions where atoms must not be placed).

These surfaces are potentially valuable tools, but if they are to be useful for decision making in drug design, it is important to know whether they are reliable and how to optimally parameterize them both sterically and for display of chemical complementarity. No information on either of these questions is currently available for the Lee and Richards surface, and we felt it would be worthwhile to investigate in detail the specific application of this surface to ligand–protein binding site interactions. Two questions have been addressed. (1) What probe radius to use and how reliable is the surface as a guide to positioning of ligand atoms? (2) Can the surface be used to represent binding site specificity (i.e. to reliably indicate which types of ligand atom should be placed on a given part of the surface to optimize binding)?

To answer the question on probe radius and steric reliability we used experimentally derived crystal-packing data for small molecules. Crystal data of molecules that

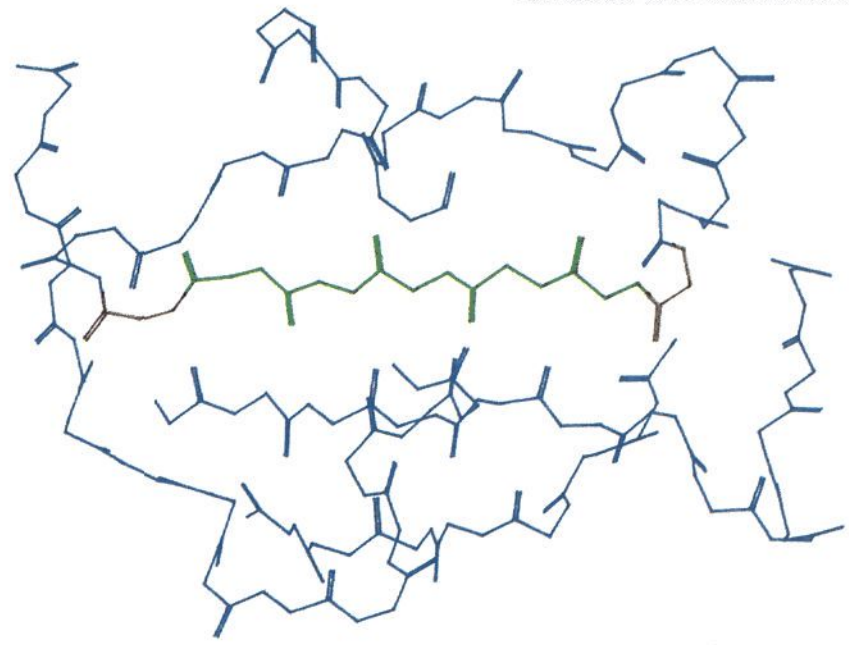

Figure 2. The formation of a virtual ligand/virtual binding site pair from a protein. The crystal structure of the backbone of ribonuclease A is shown.[20] The amino acids shown in red are removed and the sequence which lay between them (green) is treated as a ligand while the remaining structure serves as a binding site (blue). A Lee and Richards surface is placed on this binding site and properties of surface points (e.g. electrostatic potential) are computed using data from the binding site atoms only. These properties are then quantitatively evaluated for their correlation with the atom type of the ligand atom found in contact with the surface. The ligand/binding site contacts are divided into two classes, buried, and solvated, depending on distance from the solvent surface of the protein.

display no internal folding was used. This ensured that each atom was in nonbonded contact with surrounding molecules. The choice of probe radius is critical since too large or small a value will displace the surface away from the position where ligand atoms in good contact with the binding site must lie.

A successful answer to the second question, which concerns the display of chemical specificity, would greatly extend drug-design applications of the surface by directing the choice of functional groups for novel ligand molecules. This question was also investigated using experimental data. The simplest way to tell whether a property of the binding site at the L&R surface (e.g. electrostatic potential) is a useful way to characterize specificity is to compare the type of ligand atoms found close to the surface according to reliable experimental data with the property calculated for that part of the surface. This procedure answers questions such as how often negatively charge atoms lie in a region of positive potential.

To apply this method for assessing the usefulness of surface properties we needed a large amount of data where the complementarity of the interactions could be expected to be high. For the reasons described later we generated this learning set of virtual ligand/binding sites from proteins of well-defined structure (see Figure 2).

Once the optimal method for displaying steric and chemical specificity had been determined, it was applied to ligand/enzyme complexes. The ligand/enzyme data *were not included* in the learning sets and served to validate our conclusions.

The usefulness of the surface for studying ligand enzyme interactions was evaluated in two ways: qualitatively by visual inspection and quantitatively using thermolysin inhibitors where both structure and potency are known. A parallel contour representation[7] previously shown to be

(7) Bohacek, R. S.; Guida, W. C. A Rapid Method for the Computation, Comparison, and Display of Molecular Volumes. *J. Mol. Graphics* **1990**, *7*, 113–117.

Table I. van der Waals Radii Used for the Generation of the L&R Accessible Surface

atom	type	radius	atom	type	radius
C	sp_3	1.65 (1.80)[a]	O	carbonyl	1.30 (1.60)
CH	sp_3	1.85 (1.85)	O	other	1.35 (1.65)
CH_2	sp_3	2.00 (1.92)	N	sp_2	1.75 (1.75)
CH_3	sp_3	2.00 (2.00)	N	sp_3	1.75 (1.85)
C	sp_2	1.50 (1.85)	H	hetero	1.00 (1.00)
CH	sp_2	1.70 (1.85)	S		2.55 (2.00)

[a] Radii in angstroms based on contact distances from crystal-packing data quoted by Hopfinger.[9] AMBER radii are given in parentheses for comparison.

effective for van der Waals surfaces was used.

A preliminary communication of part of this work has been presented.[8] Reference to the use of a surface similar to the L&R surface for studying steric aspects of ligand binding site interactions has recently been made (unpublished results of Barry quoted in ref 6), but no details were published.

Steric Properties of the Surface

(i) Choice of Probe Radius. The position of the L&R surface of an enzyme depends on the van der Waals radius of the enzyme atom and the radius of the probe. The distance between a ligand atom and an enzyme atom in van der Waals contact depends on the van der Waals radii of the enzyme and ligand atoms

$$\text{Distance of enzyme atom to surface} = V_e + V_p$$

$$\text{Distance of enzyme atom to ligand atom} = V_e + V_l$$

V_e, V_p, and V_l are the van der Waals radii of the enzyme, probe, and ligand atoms, respectively.

These relationships show that a ligand atom in optimal contact can only lie on the surface when the probe radius equals the van der Waals radius of the ligand atom. The choice of radius is critical because changing the radius will displace the surface away from the ligand atoms, thus making it unreliable as a docking guide. Because the ligand atoms have different van der Waals radii, the choice of a single probe radius is necessarily a compromise. The choice is further complicated since atoms forming a hydrogen bond are brought closer together than the sum of the van der Waals radii.

We therefore examined the consequences of choosing different probe radii on the steric match between the surface and a ligand. Small molecule crystal packing data was found to be very well suited to this purpose. The molecules were tightly packed and were sufficiently simple to ensure that every atom formed good contacts with neighboring molecules. It was found to be convenient and adequate to use structures which only display heteroatom hydrogens; i.e. united atom carbons were used.

A molecule of each crystal was chosen to represent a ligand. The surrounding molecules of the packed structure represent the binding site. Using van der Waals radii based on contact distances (see Table I), a van der Waals surface is placed on the binding site. Distances of different ligand atoms from this van der Waals surface of the crystal cavity are shown in Figure 3. This distance is equal to the probe radius that would place the L&R surface of the binding site on the ligand atom. The distance distribution

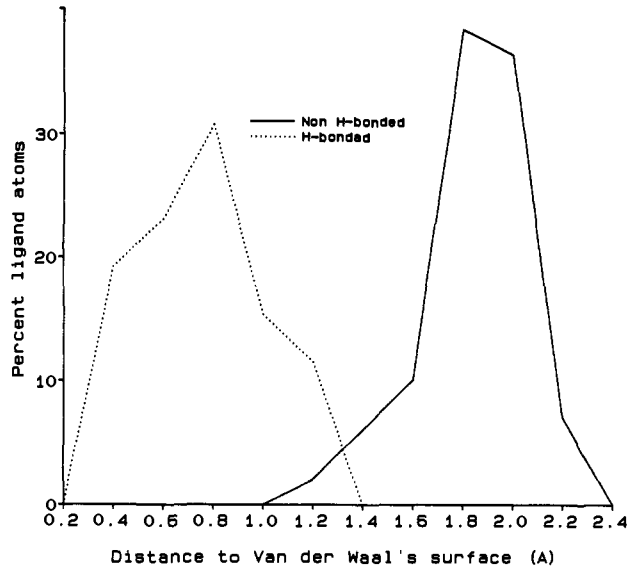

Figure 3. Distribution frequency of distance of ligand atoms beyond the van der Waals surface of the binding site. Small molecule crystal data was used to prepare 15 virtual ligand/binding site pairs. The distance represents the probe radius that would place the ligand atom on an L&R surface. Ligand atoms directly involved in hydrogen bonding (n = 26) are compared with atoms which do not hydrogen bond (n = 100). Ligand atoms attached to hydrogen atoms which hydrogen bond are excluded.

in Figure 3 is bimodal. The longer distances represent van der Waals contact and as would be expected the shorter distances occur where hydrogen bonding is taking place.

The trough occurs at about 1.4 Å and by choosing a compromise probe radius of 1.4 Å it is possible to create a surface which places nearly all ligand atoms which are in van der Waals contact on the solvent side of the surface. With this probe radius, atoms which form hydrogen bonds penetrate the surface and lie on the side closest to enzyme atoms. Using a radius of 1.4 Å it is possible to assign all of the hydrogen-bonding atoms and 98% of the non-hydrogen-bonding atoms correctly to the appropriate side of the surface.

Since the probe radius is necessarily a compromise, in general the atoms of the ligand will not lie exactly on the surface. The data in Figure 3 can be used to estimate the distance from the accessible surface where ligand atoms are likely to be found. With a probe radius of 1.4 Å, hydrogen-bonding-ligand atoms penetrate the surface by on average 0.6 Å. Atoms which do not hydrogen bond lie at a similar distance on the ligand side of the surface. These features of the surface make it easy to orient a molecule correctly with respect to the binding site.

(ii) Topological Properties of the Surface. The L&R surface has different topological features from a van der Waals surface. The van der Waals surface consists of spherical areas which are convex toward the solvent with radii equal to the van der Waals radii (Figure 1a). The L&R surface is composed of spherical segments which have as radius the sum of the van der Waals radius plus the probe radius. This more shallow curvature leads to parts of the surface being relatively flat and featureless. However the surface gives much sharper definition of grooves and crevices in the binding site than a van der Waals surface (see Figure 1b and Figures 7 and 8). The width of the grooves and crevices can be very narrow (zero width can be obtained if two surfaces touch). If surfaces touch, a ligand atom must lie close to the interface if it is to avoid high-energy van der Waals contact with atoms of the binding site. In fact the entire space on the solvent side

(8) Bohacek, R. S.; McMartin, C. Display and Analysis of Protein Binding-Site Topologies Using Accessible Surfaces. *J. Mol. Graphics* 1989, *3*, 173.

(9) Hopfinger, A. J. *Conformational Properties of Macromolecules*; Academic Press: New York, 1973.

of the surface represents a region of free volume where ligand atom centers can lie without being in repulsive van der Waals contact with the binding site.

These properties of the L&R surface make it useful for suggesting the shape of the skeletal structure of ligands which have steric complementarity to the binding site.

Chemical Specificity of Binding Sites

Having optimized the positioning of the L&R surface, we investigated ways of displaying binding site properties. To be useful, the displayed property must provide reliable guidance about the ligand groups expected to be found in different regions of the site.

A number of different ways of calculating binding site properties were compared. Each property was computed and correlated with the type of ligand group or atom found near the surface in complexes with experimentally determined structures.

(i) Choice of Ligand Binding Site Learning Set. The most appropriate choice for a learning set would clearly be structures of ligand–protein complexes. We found that the number of high-resolution complexes was rather small. In addition not all of these complexes have high affinity and there is no reason to suppose that they provide examples of optimum complementarity. We also wanted to retain at least some of the data on actual ligand–protein complexes to validate the results of our investigation of optimum surface display.

We therefore decided to seek a more extensive alternative data set and to reserve the data on ligand–protein complexes to test the surface once it had been parameterized. The packing of the interior of proteins is likely to be close to optimum in terms of complementarity and is likely to represent a good model for protein–ligand interactions. High-resolution protein structures were therefore used and ligand/binding site pairs were created by taking extended chain sequences of peptide as virtual ligand and the surrounding protein as virtual binding site (Figure 2).

The choice of extended chain sequences was deliberate and was made to avoid having virtual ligands which had internal interactions. Display of binding site properties is likely to be most effective for ligand atoms which are not strongly internally bonded, e.g. through salt bridges or hydrogen bonds. Where a ligand is internally bonded, its environment is part ligand and part binding site.

(ii) Method for Evaluating Surface Properties. Two ways of evaluating the match between surface properties could be considered: (a) Points on the surface could be selected and correlated with the types of ligand atoms found near each point. (b) Ligand atoms types could be compared to the property of the nearest surface point.

Since the ligand atoms form a small unambiguously defined set, we employed procedure using the following algorithm: (1) Calculate the L&R surface of the binding site and the L&R *solvent* boundary of the combined ligand–binding site complex. (2) Select all ligand atoms in contact with the binding site (L_i). (3) Establish the point on the accessible surface of the enzyme which lies closest to each ligand site (S_i). (4) Decide if the ligand atom and the surface point is in contact with the solvent boundary computed in step 1. (5) Compare the type of each virtual ligand atom (L_i) with the properties of the nearest enzyme surface point (S_i) (e.g. electrostatic potential due to binding site atoms or distance from hydrogen-bonding atoms of the binding site). (6) Select the surface property which best reflects ligand atom type.

To select ligand atoms in contact with the binding site a contact distance of 4.0 Å or less was chosen as being slightly greater than the sum of van der Waals radii of carbon atoms. This distance ensured that the atoms selected would be contact atoms and gave a large enough set of ligand atoms for subsequent analysis.

Table II. Percent of Ligand Atoms Located at Specific Complementary Regions of the L&R Surface of Binding Sites[j]

ligand atom type[b]	number of ligand atoms[c]	surface zone of binding site[a] (% ligand atoms found in zone) H acceptor[d]	H donor[e]	Hydrophobic[f]
carbonyl O	67	0	*94*[g]	6
all polar[h] H	87	*92*	1	7
HN	83	*92*	1	7
HO	4	*100*	0	0
NH	76	*89*	3	8
OH	4	50	25	25
all nonpolar[i] C	150	7	2	*91*
CH_3	54	4	0	*96*
aromatic C, CH	29	3	3	*93*
aliphatic CH	23	4	4	*92*
aliphatic CH_2	44	16	2	*82*
carbonyl C	76	13	53	34
all polar C	179	37	23	40

[a] The receptor surface is divided into three types of zone based on H-bonding environment (see Figure 6 for basis of zone separation). For each ligand atom the zone of the binding site surface to which it is closest is determined and statistics are compiled. [b] The table shows ligand atoms sorted according to chemical type (down) and to observed surface zone (across). [c] Total number of ligand atoms of a given type in contact with the surface; only contacts not exposed to solvent are used for this table. [d] H-acceptor surface: seeks H-donor ligand atom, less than 3.0 Å from a binding site oxygen. [e] H-donor surface: seeks H-acceptor ligand atom, less than 2.6 Å from a binding site polar hydrogen and more than 3.0 Å from a binding site oxygen. [f] Hydrophobic surface: greater than the above distances from binding site H and O. [g] Results in italic indicate results for the expected complementary surface zone for the ligand atom type. [h] Polar H are H atoms covalently bonded to N or O. [i] Nonpolar C are carbon atoms not covalently bonded to N, O, or carbonyl carbon. [j] Data from virtual ligand/binding site pairs derived from proteins; complementary regions determined using hydrogen-bonding properties of the binding site.

Preliminary studies showed that complementarity was strongest in regions which were buried (i.e. where water was excluded from direct contact). Ligand and binding site contact atoms that were more than 3.2 Å away from the *solvent*-accessible surface of the ligand–enzyme complex were classified as buried and tested first. This distance was chosen as being a few percent greater than the typical distance between two hydrogen-bonded oxygens, which is 2.85 Å. Using this cutoff, about 50% of the total number of contacts were found to be buried. As will be shown below, this definition proved to be effective because the complementarity of nonsolvated contacts was found to be accurately described using a simple algorithm for computing surface properties.

(iii) Ligand Atom Type Classification. In order to evaluate complementarity between a protein surface and a ligand, it is necessary to choose some property of ligand atoms or ligand functional groups to compare with the surface property being tested. Properties which suggest themselves are charge, dipole moment, hydrophobicity, hydrogen-bonding potential, and functional type of atom. In this study correlation with the ligand charge obtained from AMBER was tested initially, but the results were unsatisfactory. The reasons for this became apparent when correlation of each major functional type of ligand atom and local surface property was examined (see Table II for example).

When investigating surface electrostatic potential and electrostatic gradient, a number of patterns emerged.

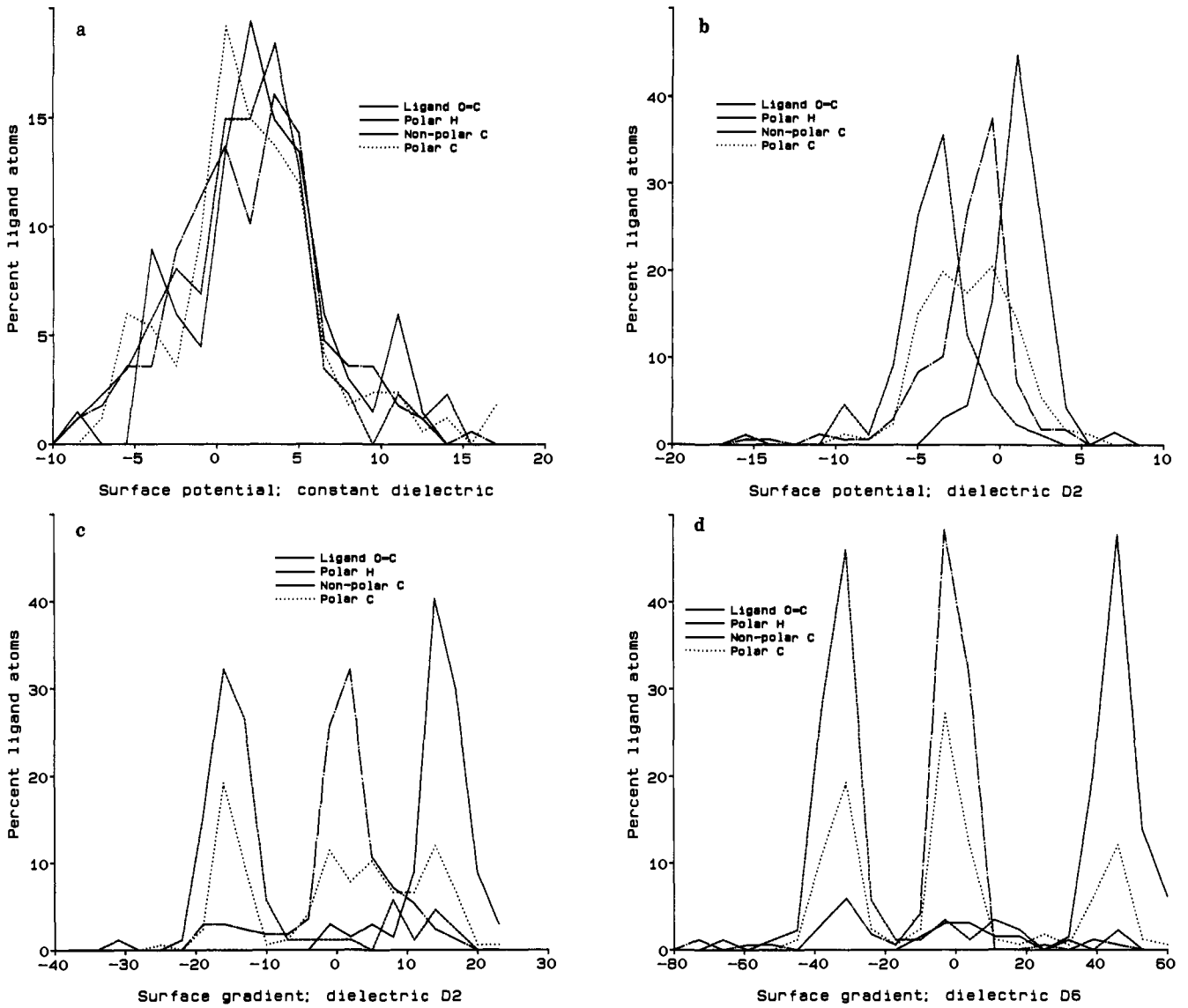

Figure 4. Distribution frequency of different virtual ligand atom types with respect to electrostatic properties found at the nearest point of the L&R surface. Protein data was used as explained in Figure 2. Only buried ligand/binding site contacts are shown. Electrostatic potential (a) without and (b) with distance-dependent dielectric to the power 2 are compared with electrostatic gradient normal to the surface with (c) distance-dependent dielectric to the power 2. The effect of using highly distance-dependent dielectric to the power 6 is also shown (d). Very good resolution of hydrogen-bond-donor and -acceptor and hydrophobic ligand atoms is obtained using the gradient normal to the surface and a highly distance-dependent dielectric.

Hydrophobic atoms were found most often at regions of low potential or gradient. Carbonyl oxygens were found in regions of positive potential and hydrogen atoms in regions of negative potential. So far this is consistent with a classification based on ligand atom charge. However for other polar ligand atoms classification by charge was not helpful. For example protonated amide nitrogens, although negatively charged, are most frequently found in areas of the cavity with a negative potential. This happens because the proton, which is positively charged, is attracted to these regions by hydrogen bonding. Similar mismatching of charge and surface potential occurred for protonated hydroxyls and for carbonyl carbons.

Surface displays based on the distance to local hydrogen-bonding groups of the binding site were also investigated (see below for details). The ligand atoms were found to be grouped in a very similar manner to that observed for the electrostatic surfaces and the surface property did not correlate with ligand atom charge.

It is therefore clear that complementarity cannot be described in terms of ligand atom charge and is more effectively described by grouping the atoms into equivalent types of functionality. For all the surfaces tested the best results were obtained by classifying the ligand atom types into the following groups: (Ia) polar hydrogens, hydrogens attached to heteroatoms; (Ib) polar atoms carrying a hydrogen, heteroatoms with attached hydrogens; (II) electronegative atoms, oxygen and nitrogen with lone pair; (III) nonpolar atoms, all carbons except for those attached to a heteroatom or to a carbonyl carbon; (IV) polar carbons, carbons excluded from category III.

This scheme separated atoms into three classes (I, II, and III) which we found could be well differentiated by local surface property (see below and Table II) and a residual class (IV) which was not well correlated with any surface property we tested.

(iv) Evaluation of Surface Properties Based on Charge. The relationship between various electrostatic surface properties and ligand atom type is examined in the frequency plots of Figure 4a–d. Binding site potential (Figure 4a) gives very little discrimination between ligand atom types. There is some improvement with a distance-dependent dielectric (Figure 4b). Because the electrostatic force is long range, there is a tendency for

anions or cations to dominate the potential over a large area, and this is probably the reason for the lack of correlation between surface potential and ligand atom type and for the improvement obtained using a distance-dependent dielectric.

It is arguable that the electrostatic gradient should be more relevant for determining complementarity. Most of the polar interactions in our data set involve ligand dipoles. The net potential energy depends on the potential gradient of the electrostatic field and the distance of the ligand atoms forming a dipole. The ligand dipole therefore samples the potential of the binding site at two points 1–2 Å apart. The ligand atom dipoles often lie in a direction roughly perpendicular to the surface. For this reason a better approach may be to use the electrostatic gradient as a surface property. The component of the electric gradient vector normal to the surface was computed and found to give a better prediction of ligand atom type than electrostatic potential (Figure 4c). The ligand atom types are however still not completely separated.

The progressive improvement seen in Figure 4, parts a–c, clearly suggests that the shorter range potential calculations more accurately reflect the processes determining complementarity. To examine this further a very high distance dependence was tested using a dielectric depending on high powers of the distance. With a power of 6, three of the ligand atoms types separate almost completely into three tightly clustered zones (Figure 4d). This confirms that the interactions determining complementarity are very short range since with a power of six only binding site atoms very close to the surface will contribute to the potential. Again this procedure does not separate the polar carbon atoms. This lack of separation may be due to the fact that these atoms have intermediate properties; i.e. they carry a charge but cannot hydrogen bond. It might also result from the polar carbons being biased toward the polar regions of the binding sites through direct attachment to hydrogen-bonded heteroatoms.

Surface potential of binding sites has been used previously with a van der Waals surface[5] and shown to reflect large-scale features of ligand binding site interaction (e.g. orientation of charged groups). In the present study this method does not appear to be effective for a description of complementarity at the atomic level.

The dramatic effects of shortening the range of the electrostatic field were not expected. However, it has been shown previously that using short range electrostatic effects can be beneficial in protein minimizations.[10]

(v) Hydrogen Bonding and Hydrophobic Surface Properties. An alternative description of complementarity might be achieved using hydrogen-bonding properties of the surface. Those parts of the enzyme surface where hydrogen bonds can be formed can be expected to be well hydrated, i.e., hydrophilic. The parts of the enzyme surface which lack hydrogen-bonding groups are likely to be hydrophobic. The parts of the enzyme surface which can form hydrogen bonds should display specificity for the complementary type of ligand hydrogen-bonding group.

Distance to the nearest hydrogen-bond-acceptor or -donor atom was found to be highly effective as a method of characterizing the surface (Figure 5a,b). It can be seen that there is a marked tendency for the point on the surface lying closest to a ligand atom to be of complementary type.

This method of classifying the surface also provides the following way of defining hydrophobic areas:

Definition 1, a region of the surface which is not close to hydrogen-bonding groups of the binding site.

Inspection of Figure 5a,b shows that the nonpolar ligand carbons are well assigned by this definition since they lie close to points on the binding site surface which are some distance from hydrogen-bond-acceptor and hydrogen-bond-donor binding site atoms.

This definition was compared to two alternative, more conventional definitions:

Definition 2, a region of the surface where there is a nearby hydrophobic enzyme atom.

Definition 3, a region of the surface where there is a cluster of nearby hydrophobic atoms.

Definitions 2 and 3 were tested using the protein data (Figure 5c,d). Although there is a tendency for the hydrophobic ligand atoms to separate, as expected it is clear that definitions 2 and 3 are much less effective than definition 1 for describing regions of the binding site where nonpolar ligand atoms are found.

The effectiveness of hydrogen-bonding distance to resolve major hydrogen-bonding and hydrophobic ligand atom types is shown in the bivariate plots of Figure 6a,b. The lines indicated on the figure can be used to classify the surface into areas of different atom type with hardly any overlap. Table II shows the effectiveness of this classification of surface points on a variety of ligand atom subtypes.

The results of Table II are obtained using the van der Waals radii of Table II and are dependent on the choice of van der Waals radii. With the radii of AMBER[5] the predictions for hydrophobic groups were considerably reduced. AMBER radii give prediction rates of 88% for CH_3 groups and 64% for CH_2 groups compared to 96% and 82% for the same groups using the radii of Table I.

(vi) Influence of Solvent. The results obtained so far for complementarity are for contacts which are not strongly influenced by solvent. The types of ligand atom in contact with binding site atoms and simultaneously within hydrogen-bonding range of solvent have also been examined. Solvation of a ligand atom is considered to occur if distance to the solvent-accessible surface of the complex is less than 3.2 Å, as in section ii.

A large number of virtual binding site–ligand interactions are in the vicinity of solvent. Correlation between distance of the surface from hydrogen-bonding atoms of the binding site and the position of nonpolar ligand carbons remains very effective, but the relationship to hydrogen-bonding ligand atoms is weaker. This shows that if the surface is polar then a polar ligand group is expected, but at a nonpolar surface, either polar or nonpolar atoms can be found. Presumably the solvent stabilizes the polar ligand group even when it is at an apparently nonpolar binding site surface.

Application to Ligand Binding Site Complexes

The best correlation between ligand atom type and surface property was obtained using distance from the surface to enzyme hydrogen-bonding atoms and this was therefore adopted as the method of coloring the surface. It has a second advantage over an electrostatic calculation, namely that knowledge of atomic charges is not required. The problem of assigning charges is not a trivial one and different force fields actually use different charges. Another feature of this coloring protocol is that it supports a simple interpretation in terms of local hydrogen bonds and it is therefore easy to discuss and analyze results by locating the responsible receptor functional group.

(10) Whitlow, M.; Teeter, M. M. An Empirical Examination of Potential Energy Minimization Using the Well-Determined Structure of the Protein Crambin. *J. Am. Chem. Soc.* 1986, *108*, 7163–7172.

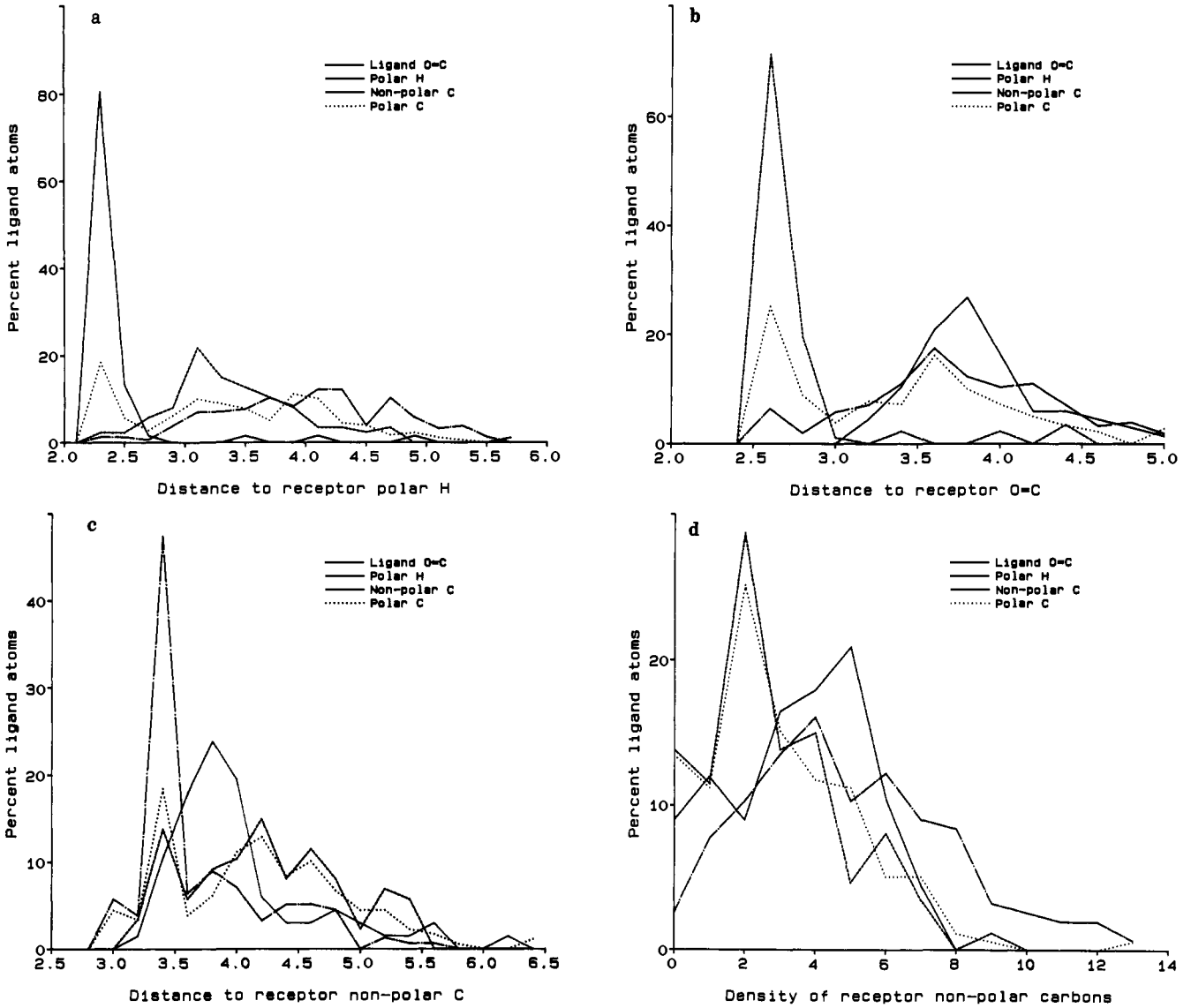

Figure 5. Distribution frequency of different virtual ligand atom types with respect to distance of the L&R surface from binding site hydrogen-bonding groups (a and b) and hydrophobic groups (c). Protein data was used as explained in Figure 2. Only buried contacts are shown. Part (d) shows the relationship between ligand atom type and density of binding site hydrophobic atoms measured by counting the number of such atoms within 5.0 Å of the surface point nearest to the ligand atom.

Figures 7 and 8 show the surface represented using two orthogonal sets of parallel contours. The contours are placed 0.5 Å apart. This method of showing the surface is graphically very effective and in conjunction with viewing as a stereo pair gives a powerful sense of depth and overall shape. At a graphics terminal, with real-time rotation and line intensity depth queuing, the contour representation works very well even if not viewed as a stereo pair. This is helpful for those unable to view in stereo. Nonstereo display is of course greatly enhanced with real-time rotation, allowing the relationship of the ligand to the surface to be readily observed.

The parameters indicated by the results of Figures 5 and 6 and Table II are used to color the surface (see Experimental Section for details).

(i) Trypsin–Trypsin Inhibitor. Figure 7 shows the L&R surface of the extended binding site of trypsin.[11] The fragment of bovine pancreatic trypsin inhibitor which contacts trypsin is shown with carbon atoms in white and heteroatoms colored red and blue. The major features of the site are clearly displayed as described in the legend. The specificity cation binding pocket is revealed as a tunnel lying to the back left of the image.

(ii) Thermolysin–Inhibitor Complexes. Surfaces were prepared for the thermolysin binding site and contacts to the inhibitor ZF^{P}LA were examined.[12] Major peptide binding specificity pockets of the site are displayed in Figure 8a. Figure 8b shows hydrophobic and hydrogen-bonding interactions which are critical for the design of inhibitors.[13] Figure 8b also shows very clearly the way in which atoms involved in hydrogen bonding penetrate the surface while those which do not hydrogen bond lie

(11) Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen and Its Complexes with Inhibitors. *Acta Crystallogr.* **1983**, *39*, 480–490.

(12) Holden, H. M.; Tronrud, D. E.; Monzingo, A. F.; Weaver, L. H.; Matthews, B. W. Slow- and Fast-Binding Inhibitors of Thermolysin Display Different Modes of Binding: Crystallographic Analysis of Extended Phosphonamidate Transition-State Analogues. *Biochemistry* **1987**, *26*, 8542–8552.

(13) Matthews, B. W. Structural Basis of the Action of Thermolysin and Related Zinc Peptidases. *Acc. Chem. Res.* **1988**, *21*, 333–340.

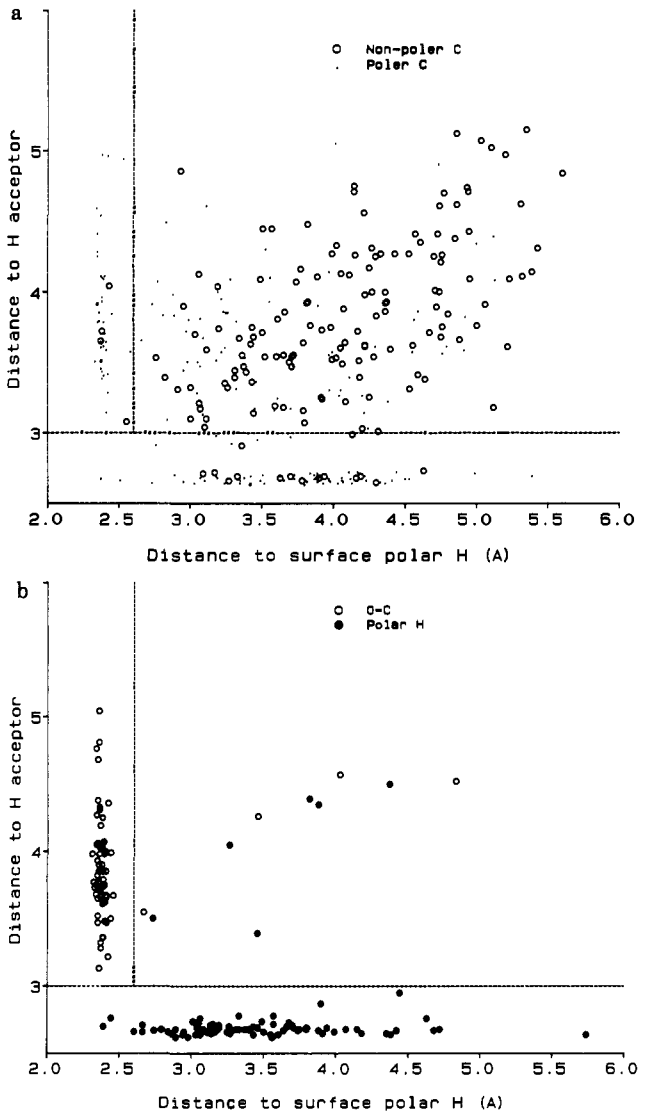

Figure 6. Bivariate plots showing the distribution of virtual ligand atom types over a surface characterized by distance to binding site hydrogen-bond-acceptor atoms and binding site hydrogen-donor atoms. Protein data was used as explained in Figure 2. Only buried contacts are shown: (a) distribution of carbon atoms (b) distribution of oxygen and hydrogen atoms. Hydrogen-donor and -acceptor and hydrophobic ligand atoms cluster into three separate areas shown by the dashed lines. This classification provides a powerful way of coloring the surface to predict ligand atom type.

entirely on the ligand side of the surface.

(iii) Quantitative Assessment of Ligand Binding. The surface coloring algorithm optimized using proteins was validated using thermolysin–inhibitor complexes. It was not certain that the parameters obtained for the packing of protein chains would apply to a ligand–protein binding site interaction. It was possible that the method, depending as it does on hydrogen bonding and a rather special way of defining hydrophobicity, might be less useful in the latter situation.

The practical value of the method will depend on its relationship to binding affinity; i.e. does an increased number of complementary contacts enhance binding or not? To answer this question, complementarity of ligand–binding site interactions was quantitatively scored. Results for 10 thermolysin–inhibitor complexes are given in Table III.

It is not expected that each of these inhibitors will be totally complementary to the receptor since some of them have low potency. However a correlation between number of complementary ligand surface contacts and log potency might be expected if the method of depicting surface properties is valid.

Results of statistical analysis of the data are shown in Table III and in Figure 9a,b. Although the statistical correlations for all molecules are reasonable, $ZG^P(O)LL$ is clearly an outlier and its exclusion was found to markedly improve the statistical values.

$ZG^P(O)LL$ is the only example in our set where two hydrogen-bonding groups in contact with each other are mismatched. This inhibitor has an ester oxygen in place of an amide NH in a critical H-bonding region of the receptor. The mismatch occurring here is known to be very costly in terms of free energy.[14] Formation of the complex will result in dehydration of two polar groups (ligand amide oxygen and binding site carbonyl) which cannot hydrogen bond to each other and are in a state of electrostatic repulsion in the docked structure. The method of assessing complementarity being used to produce the data of Table III only scores correct matches. It would be interesting to include a penalty for mismatches of hydrogen bonds in the complementarity scoring method, but with only one example in the data set, the penalty cannot at present be parameterized.

The results of Table III show that correlation between log potency and total predicted contacts is highest and that the correlation for buried nonpolar carbon contacts is also highly statistically significant.

The results of the analysis of interactions between protein chains, based on a large amount of data (Figures 5 and 6, Table II), showed that nonpolar carbon contacts of all types and buried hydrogen-bonding contacts are best predicted. A multiple linear regression of the thermolysin–inhibitor data was therefore carried out using two explanatory variables: total complementary nonpolar carbon contacts (NPHO) and buried complementary hydrogen bonding contacts, where the ligand provides a hydrogen donor or a hydrogen acceptor (NHBOND).

all compounds		
log (K_i)	3.1 − (0.51 × NPHO)	3.1 − (0.21 × NHBOND)
significance (p)	0.038	0.39

$n = 10$; sd = 1.21; $p = 0.045$; $r^2 = 0.59$

excluding $ZG^P(O)LL$		
log (K_i)	3.82 − (0.65 × NPHO)	3.82 − (0.29 × NHBOND)
significance (p)	0.0001	0.0003

$n = 9$; sd = 0.21; $p = 0.000002$; $r^2 = 0.99$

The results show that potency is highly correlated with complementarity for the series of compounds excluding the outlier. The compounds span a wide range of potencies and occupy the specificity pockets in different ways. However, the number of compounds where structural data were available for this test is somewhat limited. In addition a number of factors other than complementarity can be expected to influence potency.

For example the strain energy imposed on the ligand when docked to the receptor may play a part in determining potency. In addition thermolysin is a zinc metalloprotease and zinc–ligand interaction energies can be expected to play an important part in determining potency.

The method used in the present study indicates that complementarity is a major factor in determining potency.

(14) Bartlett, P. A.; Marlowe, C. K. Evaluation of Intrinsic Binding Energy from a Hydrogen Bonding Group in an Enzyme Inhibition. *Science* **1987**, *235*, 569–571.

Table III. Complementary between Ligand and Enzyme in Thermolysin–Inhibitor Complexes

ligand[c]	potency:[a] K_i, μM	fraction of correctly predicted ligand atoms[b]							
		nonpolar (all)[d]	polar, buried[e]			polar, solvated[f]			
		C	H	O	C	H	O	C	total
ZF^PLA	0.000068	10/12[g]	2/2	3/3	1/3	1/2	2/4	3/4	22/30
ZG^PLL	0.0091	7/10	2/2	2/2	2/3	0/2	2/4	4/4	19/31
phosphoramidon	0.028	6/9	3/3	2/2	2/4	0/3	2/7	7/5	20/35
CLT	0.05	7/11	2/2	1/1	1/4	1/2	3/4	2/4	17/28
HONH-BAGN	0.43	6/7	0	0	0/1	2/2	3/4	3/6	14/20
BAG	0.75	6/7	0	1/1	0/2	1/4	1/2	3/5	12/21
THIO	1.8	5/6	0	1/1	0/2	1/2	1/2	2/3	10/16
RETRO	2.3	5/6	0	1/1	0/1	1/2	1/2	3/4	11/16
$ZG^P(O)LL$	9.0	8/10	1/1	3/4	2/3	0/2	1/3	3/4	18/27
P-Leu-NH_2	21.3	2/4	1/1	3/3	0/2	1/3	0/1	0/0	7/14
all compounds[g]									
r		−0.73	−0.62	−0.15	−0.40	0.10	−0.59	−0.43	−0.79
p		0.016	0.056	0.7	0.25	0.78	0.07	0.22	0.007
without $ZG^P(O)LL$[g]									
r		−0.94	−0.65	−0.35	−0.67	0.30	−0.56	−0.46	−0.94
p		0.0002	0.058	0.35	0.048	0.43	0.11	0.22	0.0001

[a] From data compiled by Matthews.[13] [b] Each ligand atom is assessed for complementarity to the free volume surface of the binding site using parameters based on the results shown in Figures 5 and 6 and Table II and described in Experimental Section. The numerator shows the number of ligand atoms which are in contact with the complementary enzyme surface, i.e. hydrophobic or hydrogen bonding, and the denominator gives the total number of contacts. [c] X-ray structures of the ligand–thermolysin complexes were used. Cbz = carbobenzoxy; ZF^PLA = Cbz-Phe^P-L-Leu-L-Ala; ZG^PLL = Cbz-Gly^P-L-Leu-L-Leu; phosphoramidon = *N*-[(α-L-rhamnopyranosyloxy)hydroxyphosphinyl]-L-Leu-L-Trp; CLT = *N*-(1-carboxy-3-phenylpropyl)-L-Leu-L-Trp; HONH-BAGN = HONH-(benzylmalonyl)-L-Ala-Gly-*p*-nitroanilide; BAG = (2-benzyl-3-mercaptopropanoyl)-L-alanylglycinamide; THIO = thiorphan, *N*-[(*S*)-2-(mercaptomethyl)-1-oxo-3-phenylpropyl]glycine; RETRO = retrothiorphan (((*R*)-1-(mercaptomethyl)-2-phenylethyl)amino)-3-oxopropanoic acid; $ZG^P(O)LL$ = Cbz-Gly^P-(O)-L-Leu-L-Leu; P-Leu-NH_2 = *N*-phosphoryl-L-leucinamide. [d] Nonpolar carbon atoms. [e] Polar ligand atoms which are not in contact with solvent. [f] Polar solvated ligand atoms. [g] r = correlation between the number of correctly predicted atoms (numerator) and the logarithm of potency. p = significance of the linear regression.

With further work on a larger set of compounds it may be possible to develop a method for quantitative estimation of potency based on the complementarity scoring.

Discussion

The L&R surface has a number of useful properties for studying docking. When represented using parallel contours, it provides a powerful visual guide to the orientation of bound ligand molecules. The surface reveals the overall topology of the binding sites and provides a detailed picture of the shapes of pockets and the way different parts of a ligand fit into them.

The L&R surface can be used to display specificity with a high degree of reliability. The surface can display regions for hydrophobic, hydrogen-bond-acceptor, and hydrogen-bond-donor properties in a manner which very accurately reflects the complementarity observed in proteins (Table II).

The hydrogen-bonding-based method for displaying chemical complementarity between the binding site and functional groups of ligands was the most effective method tested in the present study (Table IV). The test data were chosen on the basis of availability of an adequate set of experimental crystal structure coordinates and not because they represent specific types of functionality for which the method might be most appropriate.

The analysis of the learning data set we used has provided a considerable amount of information on how to assign the areas of different specificity and led to the following conclusions: (a) Hydrogen bonding is much more powerful than electrostatic gradient or potential for describing complementarity on an atom-by-atom basis (Table IV). (2) Exact distance cutoffs can be assigned for the depiction of hydrogen-bonding areas of the surface (Figure 6). (3) van der Waals radii based on contact distances are essential for the adequate description of complementarity. (4) Hydrogen bonding provides a very effective means for assigning surface hydrophobicity provided the optimal definition is used. Absence of nearby hydrogen-bonding enzyme groups is a much more effective means of assigning hydrophobic surface areas than presence of nearby hydrophobic enzyme groups (Figure 5).

Table IV. Comparison of the Effectiveness of Different Binding Site Surface Properties for Display of Regional Binding Site Specificity for Three Major Ligand Atom Types

surface property of binding site	efficiency for indicating ligand atom type	ligand atom types resolved
(1) Electrostatic Potential		
(a) normal dielectric	low[a]	1, 2, 3[b]
(b) distance-dependent dielectric	low	1, 2, 3
(2) Electrostatic Gradient		
(a) distance-dependent dielectric	moderate[c]	1, 2, 3
(b) D^6 dielectric	high[d]	1, 2, 3
(3) Distance to Hydrogen-Bonding Atoms[e]		
(a) hydrogen donor	high	1 from 2, 3
(b) hydrogen acceptor	high	2 from 1, 3
(c) bivariate using 3a and 3b	high	1, 2, 3
(d) bivariate using 3a and 3b with AMBER van der Waals radii	moderate	1, 2, 3
(4) Distance to Hydrophobic Atoms	moderate	1, 2 from 3
(5) Clustering of Hydrophobic Atoms	moderate	1, 2 from 3

[a] Low: less than 50% of one of ligand atom types is assigned correctly. [b] 1, hydrogen donor; 2, hydrogen acceptor; 3, nonpolar carbon; 1, 2, 3 means each type is resolved from the other two; 1 from 2, 3 means that 2 and 3 are not resolved from each other. [c] Moderate: less than 80% of one of the ligand atom types is assigned correctly. [d] High: more than 90% of each of the ligand atom types is assigned correctly. [e] Contact van der Waals radii (Table I) are used unless otherwise stated.

The display of hydrophobic properties provides information which is virtually impossible to obtain by visual inspection of skeletal structures without a surface. We have often observed hydrophobic parts of a ligand to be in regions of the binding site which were not obviously

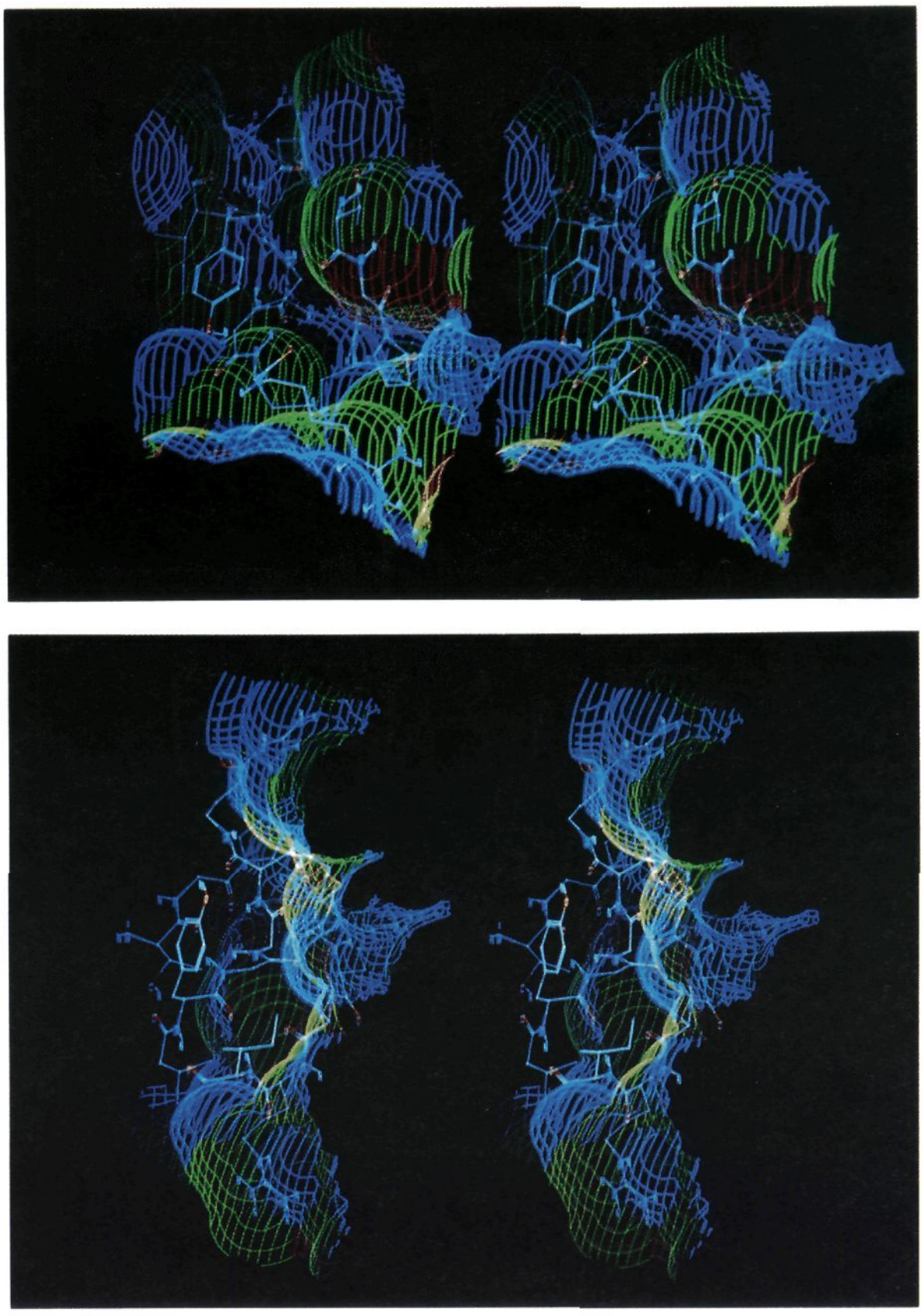

Figure 7. Contour plot of the L&R surface of trypsin. The crystal structure for the trypsin–inhibitor complex[11] was used. A probe radius of 1.4 Å was used and the surface colored using the H-bond distance criteria described in Table II. The hydrophobic surface is yellow and the H-acceptor and H-donor surfaces are blue and red. The fragment of trypsin inhibitor involved in binding is shown with carbon atoms white, nitrogen atoms blue, and oxygen atoms red: (a, top) The binding site is revealed as a twisting groove, running vertically from the top of the surface with a number of small binding pockets and crevices and a major deep pocket which is the cation binding site. The polar backbone atoms of the inhibitor lie in the red and blue regions near the foot of the groove. There are extended hydrophobic areas on the walls of the groove where amino acid side chain atoms can from hydrophobic contact. (b, bottom) Side view of the binding site showing the deep P1 specificity pocket on the right.

hydrophobic; i.e. hydrophobic groups of the enzyme were not significantly present. The display of surface hydrophobicity based on absence of nearby hydrogen-bonding groups is frequently observed to indicate that the surface near these ligand atoms is hydrophobic.

Although our results show that the surface coloring we propose is useful for indicating the type of ligand atom expected at various points of the binding site, a number of features known to be involved in specificity of binding are not shown by the surface. Use of the surface should be supplemented by knowledge of properties of the site such as ionic charge, bound water molecules, and potential flexibility.

Flexibility of binding sites where it exists[15] always poses a major problem for modeling studies. Where the binding site is flexible and the geometries of the main states of the binding site are known, the accessible surface can be used to compare the different states and to design drugs for specific states of the site.

The use of the L&R accessible surface can be combined with other techniques for probing the binding site such as GRID[16] to obtain detailed energy maps for the interaction

(15) Knowles, J. R. Enzyme Catalysis: Not Different, Just Better. *Nature* **1991**, *350*, 121–124.
(16) Goodford, P. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.

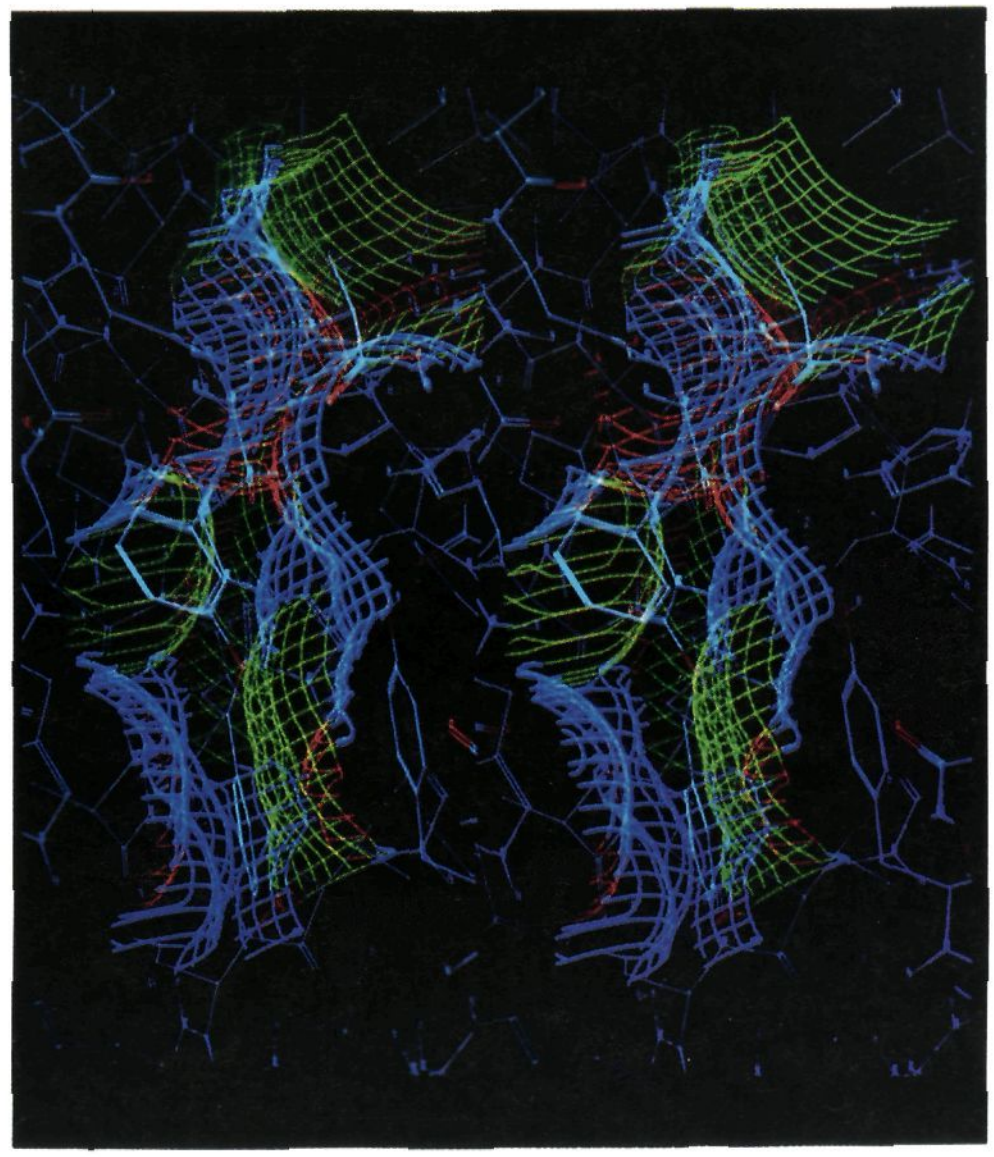

Figure 8. The thermolysin–ZF^{p}LA complex showing the L&R surface of the binding site. The colors and parameters used for the surface are the same as in Figure 7, except that ligand nitrogens are white and hydrogens blue. Part of the enzyme structure is shown with carbon atoms in dark blue, oxygen atoms red, and nitrogen blue. Part a (top) shows an extended view of ZF^{p}LA[12] fitting into the site. The yellow areas show four side-chain binding pockets: on the bottom right is the S2 site with a phenyl ring (perpendicular to the plane of the image); in the center left is the S1 site also containing a phenyl ring; the main specificity site, S1′, lies in the back of the top left hand side of the image; the S2′ site is a groove on the upper right and contains an alanine side chain. Part b (bottom) shows a closer view of the S1′ specificity pocket. The hydrophobic leucine side-chain atoms can be seen fitting closely into the yellow area. This figure also shows the formation of hydrogen bonds between the amide atoms of the inhibitor and the enzyme. The carbonyl of the inhibitor penetrates the complementary red zone of the surface and forms two hydrogen bonds with the guanidine group of Arg 203. Polar inhibitor hydrogens penetrate the blue zone of the surface. The hydrogen of the amide group (front left) forms a hydrogen bond with the carbonyl oxygen of Asn 112 and the hydrogen of the phosphoramidate group forms a hydrogen bond with Ala 113 (back left). These bonds are known to be very important for inhibitor potency.[14]

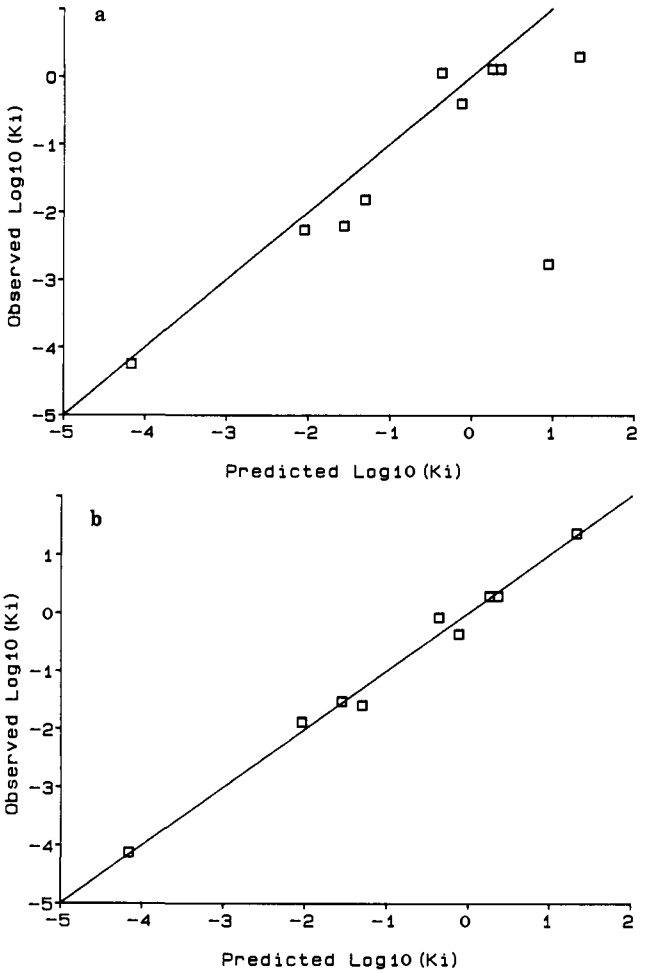

Figure 9. Results of linear regression analysis of the relationship between logarithm of observed binding constants of thermolysin inhibitors and numbers of hydrophobic contacts and nonsolvated hydrogen-bonding contacts between ligand and enzyme (see the text): (a) all inhibitors of Table III, (b) all inhibitors except ZG^{P}(O)LL.

of specific functional groups of the ligand with the site and DELPHI[17] to evaluate electrostatic potential with a continuum solvent model.

In conclusion, the surface of Lee and Richards is a valuable graphical aid for studying the steric properties of binding sites. The surface can also be used to accurately display complementary hydrogen-bonding and hydrophobic properties of the site. Thus this surface is a useful tool at the early stages of the conception and design of novel drugs directed to protein targets of known structure.

Experimental Section

Data Sources. Our preliminary investigations of the L&R accessible surface used AMBER values for van der Waals radii.[5] Subsequently it was found that the prediction of chemical specificity could be substantially improved by reducing the radius of oxygen (see results). The radii finally selected and used to generate the data reported here are based on those obtained from crystal geometry.[9] The values used are given in Table I. For electrostatic calculations, charges were assigned according to the OPLS functions recommended for proteins.[18]

Molecules were represented with united atom carbons and heteroatom hydrogens. The proteins ribonuclease[19-21] and trypsin[22-24] were obtained from the Brookhaven Protein Data Bank[25a,b] (5RNS and 1NTP). These structures are the result of extensive investigations using different techniques, and the data files contain explicit coordinates for all non-carbon hydrogens. Crystal structures for the following small compounds were taken from the Cambridge Crystal Data Base (Cambridge, U.K.): acetophenone,[26] benzene,[27] benzoic acid,[28] benzophenone,[29] benzoquinone,[30] cresol,[31] glycine,[32] *N*-acetylglycine,[33] *N*-methylacetamide,[34] oxamide,[35] pentane,[36] phthalic acid,[37] suberic acid,[38] and tetramethylammonium maleate.[39] Non-carbon hydrogens

(17) Gilson, M. K.; Sharp, K. A.; Honig, B. H. Calculating the Electrostatic Potential of Molecules in Solution: Method and Error Assessment. *J. Comput. Chem.* **1988**, *9*, 327–335.
(18) Jorgenson, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Protein Energy Minimization for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
(19) Wlodawer, A.; Borkakoti, N.; Moss, D. S.; Howlin, B. Comparison of Two Independently Refined Models of Ribonuclease-A. *Acta Crystallogr.* **1986**, *B42*, 379–387.
(20) Wlodawer, A.; Sjolin, L. Structure of Ribonuclease A: Results of Joint Neutron and X-ray Refinements at 2.0-Å Resolution. *Biochemistry* **1983**, *22*, 2720–2728.
(21) Borkakoti, N.; Moss, D. S.; Stanford, M.; Palmer, R. A. Ribonuclease-A: Least-Squares Refinement of the Structure at 1.45 Å Resolution. *Acta Crystallogr.* **1982**, *B38*, 2210–2217.
(22) Kossiakoff, A. A. Use of the Neutron Diffraction-H/D Exchange Technique to Determine the Conformational Dynamics of Trypsin. *Basic Life Sci.* **1984**, *27*, 281–304.
(23) Kossiakoff, A. A. Protein Dynamics Investigation by the Neutron Diffraction-Hydrogen Exchange Technique. *Nature* **1982**, *296*, 713–721.
(24) Kossiakoff, A. A.; Spences, S. A. Determination of the Protonation States of Aspartic Acid-102 and His-57 in the Tetrahedral Intermediates of Serine Proteases: Neutron Structure of Trypsin. *Biochemistry* **1981**, *20*, 6462–6474.
(25) (a) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, T.; Shinamouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535–542. (b) Abol, E. E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. In *Crystallographic Databases—Information Content, Software Systems, Scientific Applications*; Allen, F. H., Bergerhoff, G., Sievers, R., Eds.; Data Commission of the International Union of Crystallography: Bonn/Cambridge/Chester, 1987; pp 107–132.
(26) Tanimoto, Y.; Kobayashi, H.; Nagakura, S.; Saito, Y. The Crystal Structure of Acetophenone at 154 Degrees K. *Acta Crystallogr.* **1973**, *B29*, 1822–1826.
(27) Bacon, G. E.; Curry, N. A.; Wilson, S. A. A Crystallographic Study of Solid Benzene by Neutron Diffraction. *Proc. R. Soc. London Ser. A* **1964**, *279*, 98–110.
(28) Feld, R.; Lehmann, M. S.; Muir, K. W.; Speakman, J. C. The Crystal Structure of Benzoic Acid: Redermination with X-rays at Room Temperature; a Summary of Neutron-Diffraction Work at Temperatures Down to 5K. *Z. Kristallogr.* **1981**, *157*, 215–231.
(29) Fleischer, E. B.; Sung, N.; Hawkinson, S. Crystal Structure of Benzophenone. *J. Phys. Chem.* **1968**, *72*, 4311–4312.
(30) Bolhuis, F. Van; Kiers, C. T. Refinement of the Crystal Structure of p-Benzoquinone at −160 Degrees C. *Acta Crystallogr.* **1978**, *34*, 1015–1016.
(31) Bois, C. Low Temperature Structure of *p*-Cresol. *Acta Crystallogr.* **1970**, *26*, 2086–2992.
(32) Marsh, R. E. A Refinement of the Crystal Structure of Glycine. *Acta Crystallogr.* **1958**, *11*, 654–663.
(33) Mackay, M. F. N-Acetylglycine (Neutron). *Cryst. Struct. Commun.* **1975**, *4*, 225–228.
(34) Katz, J. L.; Post, B. The Crystal Structure and Polymorphism of N-methyl-acetamide. *Acta Crystallogr.* **1960**, *13*, 624–628.
(35) Ayerst, E. M.; Duke, J. R. C. Crystal Structure of Oxamide. *Acta Crystallogr.* **1954**, *7*, 588–590.
(36) Norman, N.; Mathisen, H. The Crystal Structures of Lower Paraffins. *Cryst. Acta Chem. Scand.* **1964**, *18*, 353–360.
(37) Bates, R. B.; Cutler, R. S. Phthalic Anhydride. *Acta Crystallogr.* **1977**, *33B*, 893–895.
(38) Housty, J.; Hospital, M. Crystal Structure of Suberic Acid. *Acta Crystallogr.* **1964**, *17*, 1386–1390.
(39) Drobez, S.; Golic, L.; Leban, I. Structure of Tetramethylammonium Hydrogen Maleate. *Acta Crystallogr.* **1985**, *41*, 1503–1505.

were deleted. The molecules were packed into a unit cell taking into account space group, and this cell was repeated by translation to ensure that at least one molecule was surrounded by other molecules (using SHELXTL PLUS software, Siemens Analytical X-ray Instruments Incorporated, Madison, WI). A fully surrounded molecule serves as ligand and the remaining molecules as a binding site. Tetramethylammonium maleate gives rise to an anion and a cation binding binding site. After transformation to MACROMODEL format,[40] double bonds and net charges were assigned interactively at a graphics terminal.

The following enzyme–inhibitor complexes were obtained from the Brookhaven Data Bank[25] (The numbers and letters after the chemical names are the data bank code names.): thermolysin complexed with ZF^{P}LA (Cbz-Phe^{P}-L-Leu-L-Ala; 5TMN),[12] ZG^{P}LL (Cbz-Gly^{P}-L-Leu-L-Leu; 5TMN),[41] ZG^{P}(O)LL (Cbz-Gly(O)-L-Leu-L-Leu; 6 TMN),[41] phosphoramidon (1TLP),[42] CLT [*N*-(1-carboxy-3-phenylpropyl)-L-Leu-L-Trp; 1TMN],[43] HONH-BAGN [HONH-(benzylmalonyl)-L-Ala-Gly-*p*-nitroanalide; 5TLN],[44] and P-Leu-NH_2 (*N*-phosphoryl-L-leucinamide; 2TMN).[42] Additional inhibitors complexed to thermolysin, but for which only the inhibitor coordinates have been published, were thiorphan,[45] retro-thiorphan,[45] and BAG [(2-benzyl-3-mercaptopropanoyl)-L-alanylglycinamide].[46] For the analysis, polar hydrogens were added using MACROMODEL.[40] Glu 143 was protonated and His 231 was made positive in accordance with published findings.[13] The structures phosphoramidon and HONH-BAGN, which have resolutions greater than 2.3 Å, were energy minimized. The inhibitors for which the complete thermolysin–inhibitor complex coordinates are not available were minimized using the active site of 5TMN and MACROMODEL[40] with the AMBER force field.[47] During the energy minimization, Glu 143, Leu 202, Leu 133, and Val 139 and the inhibitors were unconstrained, whereas the other atoms of the active site were constrained with a 10 kJ/Å^2 force constant. The high-resolution structures, ZF^{P}LA, ZG^{P}LL, ZG^{P}(O)LL, CLT, and P-Leu-NH_2, were used without further energy minimization.

Contour Plots. The procedure used for generating the L&R accessible surface was similar to that described for van der Waals surfaces by Bohacek and Guida.[7] A 3-D matrix of bit points is generated by turning on all the points less than the van der Waals radius plus the probe radius from the center of the atom. The points that are off represent the volume on the solvent side of the accessible surface. The size of the box is determined from a typical ligand or template which must be chosen or formed to cover the parts of the binding site of interest.

A maze routine[7,48] is then used to convert this information into two orthogonal sets of parallel contours. These sets of contour points are used to display the surface on a graphics terminal. They are also used for the computational analysis of complementarity as described below.

The sets of contour points are trimmed using the model ligand or template atom so that parts of the enzyme site remote from the template are not shown. This allows attention to be focused on areas of the surface selected to be of specific interest.

Steric Complementarity. To assess steric complementarity a fully surrounded molecule in a simple small molecule crystal array was used as a virtual ligand. For each ligand atom (i) the distance to the closest point on a van der Waals surface of the binding site, i.e. the remaining molecules, was calculated by finding the receptor atom (j) giving the smallest value of

$$P_i = D_{ij} - V_j$$

D_{ij} is the distance between atoms i and j, and V_j is the van der Waals radius of receptor atom j. The values of P_i can then be used to determine a suitable compromise probe radius for the L&R accessible surface as described in the methods.

Chemical Complementarity. The value of displaying various properties of the binding site on the L&R surface for prediction of ligand atom–binding site complementarity was assessed by examining the interactions between chains in proteins. Parts of the protein were treated as a virtual ligand. An L&R accessible surface was formed using the remaining atoms as a binding site. The ability of this surface to predict the character of local ligand atoms was then investigated. For each ligand atom the nearest point on this surface was found. Various ways of computing properties at this surface point based entirely on the geometry and type of binding site atoms were then investigated to see which properties were most able to predict the actual physicochemical nature of the ligand atom. The actual procedure involved a number of further steps to ensure a suitable data set as described below.

Virtual Ligands and Binding Sites from Proteins. The coordinates of ribonuclease and trypsin were obtained from the Brookhaven Protein Data Base. The solvent designations for ribonuclease were ignored because the assignment of protons is in many cases not based on experimental data. The following amino acid pairs were removed to form a virtual ligand/binding site pair. Each pair of amino acids leads to one ligand binding site pair, the ligand being the chain lying between the amino acids (Figure 2, green), the binding site the protein structure lying outside the amino acids (Figure 2, blue): ribonuclease A—Ser 18, Ser 23; Pro 42, Ala 52; Ser 59, Cys 65; Cys 72, Tyr 76; Tyr 76, Cys 84; Cys 95, Ala 102; Ala 102, Cys 110; Pro 114, Ser 123; trypsin—Asn 79, Ala 85; Pro 92, Asp 102; Lys 107, Ser 116; Ser 49, His 57; Pro 28, Ser 37; Ser 195, Cys 201; Ser 116, Ile 121; Ile 121, Cys 128; Gly 203, Val 213; Val 213, Gly 219; Asn 223, Cys 232; Ser 61, Gly 69; Asp 153, Leu 158; Leu 158, Asp 165.

Sequences which had extended chain conformation with no obvious internal hydrogen bonding were chosen to act as ligands. They therefore exemplify situations where external interactions are likely to play a maximum role in determining the chemical complementarity of the ligand. The atoms of the ligand comprised all the atoms of the amino acids including the α amino amide nitrogen and hydrogen at the N-terminal of the strip and the carbonyl carbon and oxygen at the C-terminal end. Each strip gave rise to an individual binding site which consisted of the remaining atoms minus those belonging to the two amino acids adjacent to the strip, the peptide bond being split in the same manner as that used for the ligand. The extra amino acids were removed because otherwise the surface would be too close to terminal ligand atoms, which would only be a covalent bond distance away from binding site atoms.

Computation of Nearest Surface Point (NSP). To evaluate the effectiveness of various surface properties it was necessary to have a method to identify the point on the binding site ac-

(40) Mohamadi, F.; Richards, N. G.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrikson, T.; Still, C. MACROMODEL—An Integrated Software System for Modelling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comput. Chem.* **1990**, *11*, 440–467.

(41) Tronrud, D. E.; Holden, H. M.; Matthews, B. W. Structures of Two Thermolysin-Inhibitors Complexes That Differ by a Single Hydrogen-Bond. *Science (Washington, D.C.)* **1987**, *235*, 571–574.

(42) Tronrud, D. E.; Monzingo, A. F.; Matthews, B. W. Crystallographic Structural Analysis of Phosphoramidates as Inhibitors of Transition-state Analogs of Thermolysin. *Eur. J. Biochem.* **1986**, *157*, 261–268.

(43) Monzingo, A. F.; Matthews, B. W. Binding of N-Carboxymethyl Dipeptide Inhibitors to Thermolysin Determined by X-ray Crystallography: A Novel Class of Transition-State Analogues for Zinc Peptidases. *Biochemistry* **1984**, *23*, 5724–5729.

(44) Holmes, M. A.; Matthews, B. W. Binding of Hydroxamic Acid Inhibitors to Crystalline Thermolysin Suggests a Pentacoordinate Zinc Intermediate in Catalysis. *Biochemistry* **1981**, *20*, 6912–6920.

(45) Roderick, S. L.; Fournie-Zaluski, M. C.; Roques, B. P.; Matthews, B. W. Thiorphan and retro-Thiorphan Display Equivalent Interactions When Bound to Crystalline Thermolysin. *Biochemistry* **1989**, *28*, 1493–1497.

(46) Monzingo, A. F.; Matthews, B. W. Structure of a Mercaptan-Thermolysin Complex Illustrates Mode of Inhibition of Zinc Proteases by Substrate-Analogue Mercaptans. *Biochemistry* **1982**, *21*, 3390–3394.

(47) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Algona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.

(48) Gonzalez, R. C.; Wintz, P. *Digital Image Processing*; Addison-Wesley Publishing Company: Reading, MA, 1977.

cessible surface which lies closest to each ligand atom. This enabled a comparison to be made between the type of the ligand atom and the local property of the surface. Only ligand atoms in contact with the binding site were used. Each ligand atom was tested to see if it lay within 4.0 Å of a binding site atom. Those further away were designated as out of contact and not processed further. For those in contact, the nearest surface point was found by examining the coordinates in the appropriate contour file. A file was written listing each contact ligand atom and the coordinates of the NSP. Using this data it is simple to compile statistics comparing the atom type of each contact ligand atom to the computed binding site property at the NSP.

Elimination of Local Solvent Influence. The ligand atom/NSP list was pruned to include only ligand atom/binding site surface contacts where solvation would not be expected to have a local influence. This was done by preparing a Lee and Richard's accessible surface for the whole protein. This surface shows all the points where there is space for a solvent molecule. If there is a point on this solvent surface close to either a ligand atom or the NSP of the receptor surface for that atom, then clearly solvent is liable to influence the interaction and may over-rule the effects of binding site atoms in determining the nature of the ligand atom. All cases where the binding site surface NSP or the ligand atom were less than 3.2 Å from a solvent surface point were omitted from the ligand atom/NSP list. A second ligand atom/NSP list was prepared including solvated points.

Calculation of Surface Properties. Properties for NSP's were calculated using equations derived from the electrostatic potential and using distance to binding site atoms of various types.

Surface potential ($\mathbf{SP}_{ni}$) at point i is

$$\mathrm{SP}_{ni} = 1000\sum_i Q_j/R_{ij}^{n+1}$$

$$R_{ij} = D_{ij}/3$$

where D_{ij} is the distance between surface point i and binding site atom j, Q_j is the charge of binding site atom j and n is the degree of distance dependence of the dielectric; i.e. $n = 0$ for a constant dielectric of 1.

Gradient (G_i) at point i normal to the surface was calculated by first computing the unit vector ($\mathbf{V}_i$) pointing from the surface point to the binding site atom responsible for the surface. The gradient was then

$$G_i = \mathbf{V}_i\sum_i \mathbf{R}_{ij}\cdot Q_j/R_{ij}^{n+3}$$

$$\mathbf{R}_{ij} = \mathbf{D}_{ij}/3$$

$$R_{ij} = D_{ij}/3$$

where $\mathbf{D}_{ij}$ is the vector connecting point i to binding site atom j, and D_{ij} is the distance between the point and the atom.

Hydrogen-bonding/hydrophobic environment was assessed by finding the distance of the closest binding site atom of type x to the surface point. For hydrogen acceptors x was any oxygen, for hydrogen donors x was a polar hydrogen (i.e. one attached to a non-carbon atom). x was also tested as a carbon or as a nonpolar carbon (i.e. a carbon not attached covalently to an oxygen or nitrogen or to a carbonyl carbon). The nearest surface point was also characterized according to the density of nonpolar binding site groups by calculating the number of nonpolar receptor carbon atoms lying within a distance of 5.0 Å.

Quantitative Assessment of Ligand/Binding Site Complementarity. Files of data giving the coordinates and atom types of a ligand and a binding site were used and the number of successful hydrophobic and hydrogen-bonding contacts was counted. A contour description of the free volume surface of the binding site is computed using the method already described with a probe radius of 1.4 Å. Each point on the surface is then classified by measuring the distances to the nearest binding site polar hydrogen (HDIST) and to an oxygen (ODIST). The surface point is then designated H acceptor if ODIST is less than 3.0 Å, H donor if HDIST is less than 2.6 Å and ODIST greater than or equal to 3.0 Å, or hydrophobic if HDIST is greater than or equal to 3.0 Å and ODIST greater than or equal to 2.6 Å.

Each ligand atom is then examined to find the nearest surface point using the criteria already described in Experimental Section to determine contact and solvation.

The ligand atom contacts are scored as being complementary if (a) a ligand polar hydrogen is in contact with a hydrogen-acceptor surface point or (b) a ligand oxygen is in contact with a hydrogen-donor surface point or (c) a ligand carbon is in contact with a hydrophobic surface point.

Acknowledgment. We thank Dr. Frank H. Clarke and Dr. Wayne C. Guida for their encouragement, helpful discussions, and advice.