A New Method for Estimating the Importance of Hydrophobic Groups in the Binding Site of a Protein

Matthew D. Kelly[†] and Ricardo L. Mancera*[,‡]

Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, U.K., and De Novo Pharmaceuticals, Compass House, Vision Park, Chivers Way, Histon, Cambridge CB4 9ZR, U.K.

Received June 18, 2004

Interactions between the hydrophobic regions of a binding site and those of a complementary ligand are often observed to provide the driving force for binding. We present a new method for the analysis of hydrophobic regions in the binding site of a protein that considers not only atom type but also the nonadditive effects arising from the shape and extent of a nonpolar region. The method has been parametrized using a purpose-built genetic algorithm to optimize its ability to identify those regions that are more likely to form a strong interaction with a nonpolar ligand group. We demonstrate the ability of this method to account for changes in the shape and extent of the exposed nonpolar surface, using both artificial and protein examples. The method is also able to rationalize differences in binding affinity for ligand−protein complexes with largely hydrophobic binding sites.

Introduction

Hydrophobic interactions play a crucial role in ligand−protein binding.[1] Most ligand binding sites contain at least one hydrophobic (nonpolar) region, with many demonstrating a clear preference for nonpolar ligands, as is the case with the nuclear hormone receptors[2,3] and fatty acid-binding proteins.[4,5] In view of their importance, a comprehensive analysis of hydrophobic regions in the binding site of a protein and an evaluation of their relative importance are essential steps in the rationalization of ligand−protein interactions and in structure-based drug design.

The hydrophobic interaction is traditionally seen as a solvent-induced force that drives two or more predominantly nonpolar molecules or surfaces to stick together when placed in aqueous solution. This is largely due to the entropic cost of restructuring the hydrogen-bonding network of water.[6−11] However, for a large hydrophobic object, it becomes impossible to maintain a hydrogen-bonding network in its vicinity resulting in the disruption of the structure of water and a stronger hydrophobic interaction.[12−18] The Lum−Chandler−Weeks theory of hydrophobicity can account for the transition that occurs from the hydrophobic hydration of small nonpolar solutes to the strong tendency for depletion of water near extended nonpolar surfaces of nanometer-length scale such as those in proteins.[19,20] Consequently, the computer simulation evidence[13−18] and recent theoretical developments[19,20] reveal the need to capture the stronger hydrophobic attraction that would arise between a ligand and a protein with a large or concave nonpolar surface.

The strength of the hydrophobic interaction is thus influenced not only by the polarity but also by the shape and extent of the exposed molecular surface. Currently, no simple theoretical model exists for calculating the hydrophobic force arising as a result of a protein surface beyond that afforded by empirical ligand−protein and protein−protein scoring functions, which do not consider the effects that arise with larger nonpolar surfaces of arbitrary shape.[1,21−26] However, a number of empirical and quantum mechanical approaches have been developed to estimate the lipophilic properties of small molecules.[27−35]

The molecular lipophilic potential (MLP)[24] was the first method designed to calculate the hydrophobic profile of a molecule in three dimensions. The development of the MLP was based on the finding that the partition coefficient (*P*) of a molecule, which represents its relative distribution over an octanol/water boundary, can be estimated from its chemical structure.[36] From the assumption that the log *P* is an additive property of the molecular fragments that make up a molecule, values for a wide variety of atom types and groups have been calculated.[37−39] Such hydrophobic parameters are used in the MLP, which calculates a "potential" for a point in space by summing the surrounding distance-weighted atom/fragment values. This concept of a MLP was later combined with the traditional comparative molecular field analysis (CoMFA) method[40] in the program HINT.[30] This modification was shown to produce a more comprehensive and informative description of the interaction properties of a ligand through the consideration of its hydrophobic behavior. However, while the MLP proved to be a useful tool for calculating the effects of substituting various chemical groups on a ligand, it is less effective for the analysis of the binding site of a protein. This is largely due to (1) the lack of parametrization (if at all possible) of the fragment-based approach to calculating log *P* for a protein surface and (2) the increasing effects of the shape and extent of a surface on the hydrophobicity of a protein surface region.

One possible approach to the identification of hydrophobic regions within the binding site of a protein is that afforded by the GRID method of Goodford.[41] In the GRID method, a probe group is positioned at the vertices of a grid superimposed onto the binding site of a protein.

* Author for correspondence: phone, +44 1223 238011; fax, +44 1223 238088; e-mail, Ricardo.Mancera@denovopharma.com. Present address: Western Australian Biomedical Research Institute, Curtin University of Technology, GPO Box U1987, Perth WA 6845, Australia. Phone: +61 8 9266 2133. Fax: +61 8 92667485. E-mail: R.Mancera@curtin.edu.au.

† University of Cambridge.

‡ De Novo Pharmaceuticals.

10.1021/jm049524q CCC: $30.25
Published on Web 01/15/2005

At each position, the interaction energy is calculated using an empirical force field. When a nonpolar probe such as a methyl group is used, those positions within the cavity where hydrophobic interactions dominate can be identified. In a parallel approach, SuperStar[42,43] is a knowledge-based method that uses data derived from a survey of a large number of crystal structures to calculate the likelihood of observing a particular interaction with a specific probe group. Likewise, when a nonpolar probe such as a methyl group is used, positions in the cavity where hydrophobic interactions would be favored may be identified. However, as with the MLP, the GRID and SuperStar methods do not explicitly take into account the effects of variations in shape and size of nonpolar molecular surfaces. Instead, these methods consider atom type to be the dominant factor in determining the hydrophobicity of a molecular surface.

One way to provide a simple account of surface properties is to compute the solvent-accessible surface area (SASA).[44] This approach provides a useful tool to gain insight into the overall extent of a hydrophobic region on a molecule or in the binding site of a protein but lacks any real account of the particular atom types that make up the binding site or their positions relative to one another. In addition, it provides no means of assessing the shape of the binding since it only calculates the relative accessibility of the contributing atoms.

A more comprehensive approach to calculate the hydrophobic properties of a molecule or the binding site of a protein is to simulate the behavior of solvent molecules in the vicinity of the molecule or binding site using molecular dynamics (MD).[11,12] Such an approach implicitly accounts for both surface properties and atom type in its calculation. However, these simulations are computationally expensive, and there is no standard method for defining the hydrophobicity of a molecular surface of arbitrary shape and size on the basis of the structural or dynamical properties of neighboring solvent molecules.

In the study of protein structure and stability, hydrophobicity scales for amino acids have also been derived from their relative distributions between being at the solvent accessible surface of a protein and being buried within its core.[45,46] Many such scales have been published describing a variety of rankings for the amino acids.[45−51] These inconsistencies generally arise as a result of variations in the definition of the surface and the data sets used in their parametrization. Experimental scales for amino acids have also been calculated from the free energy of transfer from water to octanol, with octanol approximating the interior of a protein.[52] Unfortunately, these scales of amino acid hydrophobicity are more suited to the analysis of the thermodynamics of protein folding than to the prediction of atom-specific interactions with a ligand.

The energetic benefit of shielding a hydrophobic surface from the solvent has been calculated specifically for ligand−protein interactions.[53] When the effect of hydrophobic substitutions on the binding affinity of ligand for antibiotics of the vancomycin group was measured, the favorable free energy of this shielding was calculated. The energies predicted by this approach were significantly higher than those generated by other methods such as solvent transfer, where the free energy of transferring a hydrophobic group from water to bulk hydrocarbon is measured.[54,55] The reason for this discrepancy is believed to be nonadditivity. For example, the burial of a hydrophobic group of a ligand in the binding site of a protein may increase the strength of neighboring polar interactions such as hydrogen bonds. This increase in strength is due to the motional restrictions placed on the ligand by the hydrophobic interaction. Such nonadditivity results in a greater increase in binding affinity than would be expected as a result of the hydrophobic substitution alone. However, such methods are designed for the prediction of protein−ligand binding affinities based on the size of hydrophobic surface buried as opposed to the analysis of binding sites and as such do not explicitly account for surface based properties such as shape and extent.

To overcome the limitations of current methods for analyzing the hydrophobicity of protein surfaces, we have developed a new method that scores the hydrophobicity of solvent-accessible atoms within the binding site of a protein based not only on atom type but also on the size and shape of the molecular surface surrounding that atom. In this paper, we also describe the purpose-built genetic algorithm that was used in the parametrization of the method in order to optimize its performance. We then proceed to validate the method and show its usefulness with a number of examples.

Methods

Representation of the Molecular Surface of the Binding Site of a Protein. The atoms comprising the binding site of a protein are identified by their proximity to either an existing ligand atom or a ghost atom positioned arbitrarily within the binding site. The default cutoff for the distance between a protein atom and any ligand atom is 5.0 Å. This distance may be extended when a ghost atom is used. Following the definition of the binding site, we can map a dot surface onto the binding site atoms (Figure 1a). This dot surface represents the path traced out by the center of a probe with a radius of 1.4 Å as it is rolled over the van der Waals surface of the binding site. A probe radius of 1.4 Å was chosen since it represents the size of a water molecule, this being consistent with calculations of the solvent-accessible surface area. The density of dots used to create the surface is approximately 1.7 dots Å^{-2}. All protein atoms lacking the accessibility to contribute to the dot surface are removed, and each remaining atom is assigned an initial weight based upon its classification into one of 11 groups (Table 1). To maximize the performance of the method, the initial weights assigned to the different classes of protein atoms were optimized using a purpose-built genetic algorithm (described below).

The value of the initial weight of each atom is projected onto those dots representing its probe-extended surface. The purpose of this projection is to enable the calculation of surface-derived effects relative to a position a ligand atom can occupy (i.e., the probe extended surface). Each dot now has an associated value corresponding to the underlying atom type. In the next stage of the algorithm, the value of each dot is updated to reflect the values and relative positions of the surrounding dots (Figure 1b−d). For each dot, the updated score (U) is calculated as follows:

$$U = I + \sum_{\text{neighbors}} \frac{I_j}{D_j + 1} A_j \qquad (1)$$

where I is equal to the initial value of the selected dot. I_j, D_j, and A_j are equal to the initial value, distance (Å), and angle penalty factor, respectively, of the jth member of the set of all dots within 6.0 Å of the selected dot. The distance and angle penalty factors are calculated relative to the dot for which the

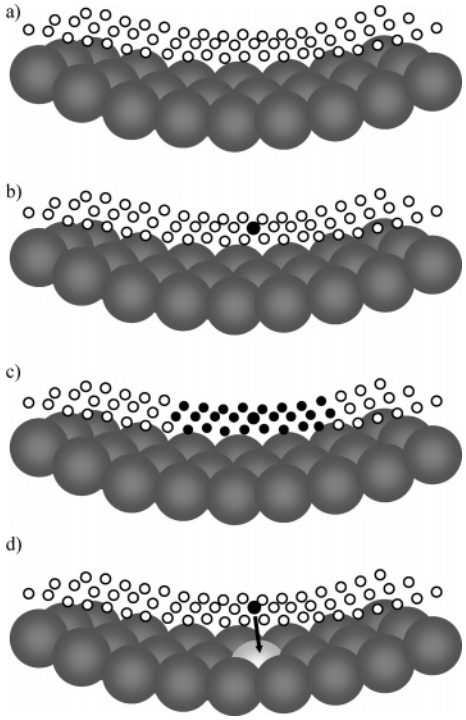

Figure 1. In our method, a dot surface is mapped onto the protein cavity (a) with an initial weight assigned to each dot based on the underlying atom type. For each dot forming the surface (b), the contribution from neighboring dots is calculated (c) and the final value projected back onto the underlying protein atom (d).

Table 1. Description of the 11 Classes of Protein Atom Defined in Our Method

atom type	description
	Carbons
nonpolar	carbon atom bonded only to carbon and hydrogen atoms
polar‘	all carbon atoms not classified as nonpolar
	Nitrogens
backbone	backbone nitrogen in all residue types
amide group	side chain nitrogens in glutamine and asparagine
charged	side chain nitrogens in lysine and arginine
ring	ring nitrogens in tryptophan and histidine
	Oxygens
backbone	backbone oxygen in all residues
amide group	side chain oxygens in glutamine and asparagine
carboxylate	side chain oxygens in aspartate and glutamate
hydroxyl	side chain oxygens in tyrosine, threonine and serine
	Sulfurs
sulfurs	all sulfur atoms (cysteine and methionine)

updated score is being calculated. The distance factor is simply the Euclidean distance between the centers of the two atoms, and the angle penalty factor represents the relative orientation of the neighboring dot. The angle penalty function can take a value between 0 and 1, with a higher value representing a dot on a surface oriented toward the selected dot and a lower one representing a dot on a surface oriented away from the selected dot.

Finally, for each binding site atom, the final score is calculated as the mean score of all dots representing its probe-extended surface. This final score denotes the calculated "hydrophobicity" of the atom. As the final score of an atom is not only dependent upon its own type but also dependent on that of neighboring atoms along with the extent and shape of the surrounding surface, it is very sensitive to its local environment. For example, a hydrophobic atom surrounded by other hydrophobic atoms would score higher than the same hydrophobic atom surrounded by polar atoms.

The resultant scoring of binding site atoms by our method is displayed on a 3D representation of the binding site, with all scored atoms depicted by spheres colored to represent their final scores. This visualization tool enables the quick and easy identification of regions of the binding site where it would be desirable to place a hydrophobic ligand group as well as those regions to avoid.

Optimization of Atomic Scores Using a Genetic Algorithm (GA). As indicated above, the initial weights assigned to each of the 11 atom types were optimized using a genetic algorithm in order to maximize the performance of the method. Genetic algorithms fall under the umbrella of evolutionary algorithms, which use an abstract representation of Darwinian evolution to maximize the fitness of a population. Making use of common genetic operators such as crossover and mutation, these methods explore the fitness landscape and "evolve" toward an optimal solution. Extensive reviews exist for this subject.[56−58] A crucial step in the implementation of such methods is the selection of a fitness function that accurately reflects the performance of a given solution. Therefore, a quantitative routine for assessing the performance of our method was required. For this purpose, we assessed the ability of our method to assign higher scores to those protein atoms experimentally observed to be close to nonpolar ligand atoms when compared to those atoms close to polar ligand atoms. The data set used was the "clean" subset of the CCDC/Astex data set.[59] This data set represents the largest populated structural classes of proteins, with each of the complexes having manually assigned protonation and tautomeric states. The clean subset comprises 220 ligand−protein complexes and is the result of removing those structures from the complete data set that contain either errors or inconsistencies in their PDB files, ligands with unlikely conformations, and complexes with severe clashes between protein and ligand atoms.[59] One of the potential pitfalls in optimizing a series of parameters for a particular data set is that the final parameters may be unsuitable for a different set of data. To test for this, the clean subset of 220 complexes was split into a training set of 165 complexes on which the initial weights were optimized and a test set of 55 complexes on which the optimal set of weights was tested for fitness. When the final solution of a set of complexes not involved in the optimization is tested, a measure of the method's generality can be gauged.

Every ligand atom within each complex of the data set was assigned to one of two categories, nonpolar or polar. Carbon atoms covalently bound only to carbon or hydrogen atoms were classified as nonpolar, and all other atoms were classified as polar. The "fitness" of our method, given a particular set of initial weights, was calculated as follows: The final scores of all protein atoms within a 4.0 Å radius of each nonpolar ligand atom were calculated and their values assigned to one population. A second population was created from the scores of all protein atoms within a 4.0 Å radius of each polar ligand atom. We note that due to the definitions of these two populations there is a degree of overlap between the two with some binding site atoms present in both populations. Following the generation of these two populations, we calculated the probability of there being no difference between them using the nonparametric Wilcoxon two-sample test. A nonparametric test was chosen due to the non-normal distribution of the data. Finally, the natural log of the probability was taken and the fitness of a particular set of parameters calculated as the magnitude of the negative of its value. The more negative the value, the greater the statistical significance that the two populations are different. To ensure that the nonpolar population contains, on average, protein atoms with scores greater than those members of the polar population, the means of the two populations were calculated, and in cases where the mean of the polar population was the greater, the natural log of the probability value was multiplied by −1. Following the maximization of the difference between the two populations of binding site atoms, the significance of this difference would provide a primary validation of the optimized parameters, since this difference provides a direct measure of the performance of the method.

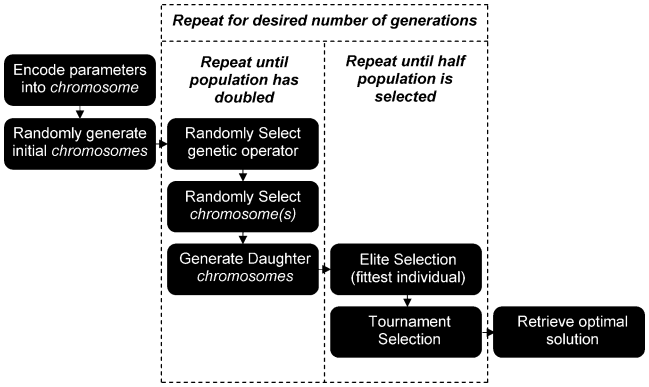

Figure 2. Flowchart describing the purpose-built genetic algorithm (GA) used to optimize our method.

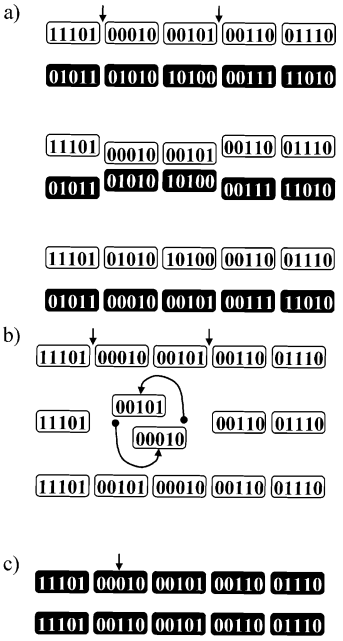

Figure 3. The three main operators used in the genetic algorithm were the following: (a) two-point crossover, in which two *gene* boundaries were selected at random on a pair of chromosomes and the *genetic* material between them was exchanged to generate two new daughter chromosomes; (b) flip operator, in which two *gene* boundaries are selected at random on a chromosome and the order of the genes between them was reversed to generate a single daughter chromosome; (c) mutation operator, in which a single *bit* is selected at random and its value reversed (i.e., 1 becomes 0 and 0 becomes 1).

The genetic algorithm used to optimize the parameters is detailed in Figure 2. The 11 initial parameters (weights) to be optimized were encoded into a "pseudochromosome" of 11 genes, each representing one of the parameters. Each gene consisted of a 5-bit binary string representing the range of integers from −16 to 15, inclusive. The three genetic operators used to generate new daughter chromosomes were the two-point crossover (Figure 3a), the flip (Figure 3b), and the point mutation (Figure 3c). Each of these operators has an associated probability that determines their chance of being used: 0.85 for crossover, 0.70 for flip, and 0.10 for mutation. Therefore, crossover and flip are the two main operators driving this genetic algorithm.

An initial population of 30 randomly generated chromosomes is used as the starting point for the genetic algorithm. Genetic operators act upon randomly selected chromosomes from this population until the population size has doubled to 60. On this expanded population, elite and tournament selections are used to select individuals to go through to the next generation. Elite selection is used to ensure the fittest chromosome always continues through to the next generation. In tournament selection, two chromosomes are selected at random and the fittest of the two goes through. This process allows some of the less-fit chromosomes to go through to the next generation.

In addition to the three genetic operators mentioned above, an additional novel feature was implemented in the genetic algorithm in order to maintain a diverse population. In test runs, the genetic algorithm was often observed to converge to a population of very low diversity with few unique solutions. In situations where the number of unique solutions is less than half the size of the population and fails to improve for five consecutive runs, the new operator is implemented. In this operation, a random population of 15 chromosomes is generated. This population is then "mated" through the crossover operator with the 15 fittest chromosomes of the current generation. As each interaction generates two daughter chromosomes, the resultant 30 new chromosomes are added to the current generation to create the complete population. Tournament and elite selections are then performed to generate the starting population of the next generation. This operator effectively injects fresh genetic material into a stagnant population to increase its diversity and hopefully leads to the creation of fitter individuals.

The genetic algorithm was run 5 times, each for 100 generations, and the overall fittest set of parameters taken. While genetic algorithms are powerful tools for finding a solution in the region of the global minimum, they often lack efficiency in identifying the absolute global minimum. As such, their combination with local search methods such as steepest descent[60] or the simplex method[61] can improve their performance. For this reason, each of the parameters from the fittest solution generated by the GA was sequentially incremented until the local optimal solution was found. This optimal set of weights was then encoded into our method for evaluating the importance of hydrophobic groups.

Results and Discussion

The first stage in the validation of our method was to assess its ability to identify those regions in a binding site where it would be more favorable to place nonpolar ligand atoms and those regions that are more favorable for polar ligand atoms. Since it was this capability that was optimized by the GA, its performance can be taken directly from the fitness function. With the optimal parameter set, the significance of the difference between the polar and nonpolar protein populations in the training set is $P = 7.614 \times 10^{-205}$. This is a highly significant result, which validates the method's ability to identify hydrophobic regions in the pocket. In addition, the final set of parameters also produced a highly significant result on the test set ($P = 1.250 \times 10^{-76}$) supporting the generality of the method.

The optimal weight values generated for the 11 atom types defined by our method are shown in Table 2. The more hydrophobic an atom type is, the larger the number, with negative numbers suggesting clear hydrophilic properties. As expected, the highest value is assigned to nonpolar carbons. This correlates with the relative abundance of this atom type around nonpolar ligand atoms when compared to polar ligand atoms. The only atom type showing a greater preference for nonpolar ligand atoms is sulfur. However, the relatively low frequency of sulfur atoms in the binding sites of the training set (0.8%) may affect the accuracy of this value. In the optimal parameter set, sulfur is given the second-highest value. The next-highest-ranking parameter is that of polar carbons. Polar carbons provide the major contribution (69.0%) to the SASA surrounding the nonpolar carbons in the binding sites studied and as

Table 2. The 11 Atom Types Defined in Our Method and the Optimized Weights Assigned to Them by the GA[a]

atom type	optimized parameter	% NP	% Pol	Δ[b]
	Carbons			
nonpolar	11	62.5	37.5	+25.0
polar	9	44.8	55.2	−10.4
	Nitrogens			
backbone	−19	41.3	58.7	−17.3
amide group	−3	40.6	59.4	−19.8
charged	−19	26.5	73.5	−46.9
ring	−25	29.4	70.7	−41.4
	Oxygens			
backbone	7	53.9	46.1	+7.8
amide group	−8	45.5	57.5	−15.0
carboxylate	−14	32.9	67.1	−34.1
hydroxyl	5	48.1	51.9	−3.9
	Sulfurs			
sulfurs	10	70.5	29.5	+40.9

[a] In addition, the relative occurrence of each type around nonpolar ligand atoms (% NP) compared to polar ligand atoms (% Pol) along with the difference between these values (Δ) is shown. [b] The Δ value represents the preference of that atom type for nonpolar ligand atoms.

such require a relatively high score due to the contribution that surrounding atoms make to the scoring of a binding site atom by our method. The next-highest-ranking parameter is that of backbone oxygens. Their position in the overall ranking is reflected by their unexpected preference for nonpolar ligand atoms (Table 2). Hydroxyl oxygen atom types are next in the rankings, because of a slight preference of these hydroxyl atoms for polar ligand atoms. Next in the ranking are the nitrogen and oxygen atoms of amide side chains. This similarity in their ranking is to be expected given their combined presence in the side chains of glutamines and asparagines. Carboxylate oxygen atom types are next in the rankings. While being the lowest-ranking class of oxygen atom, they are ranked higher than backbone nitrogen atoms due to the fact that they have a weaker preference for polar ligand atoms. The reason for this apparent discrepancy may be due to the positive reinforcement created by having a pair of carboxylate oxygen atoms in every acidic side chain, with the accompanying oxygen atom providing the major contribution (44.5%) to the surrounding SASA. On the other hand, the major contribution to the SASA surrounding the backbone nitrogen atoms comes from the top-ranking nonpolar carbons (45.6%). Consequently, the need to overcome this nonpolar influence may explain the requirement for a lower value for backbone nitrogen atoms compared to the carboxylate oxygen atoms. Charged nitrogens have the same weight value as backbone nitrogen atoms. The ranking of these two types of nitrogen atoms at the low end of the scale is supported by the fact that they have the greatest preference for polar ligand atoms when compared to nonpolar ligand atoms (Table 2). Although charged nitrogen atoms show the greatest preference for polar ligand atoms, ring nitrogen atoms lie at the lowest end of the scale. While ring nitrogen atoms show a clear preference for polar ligand atoms, charged nitrogen atoms have, in fact, a slightly larger preference. This apparent discrepancy may be again explained by consideration of the atom types contributing to the SASA surrounding the two classes of nitrogen. For charged nitrogen atoms, the greatest contribution comes from additional charged nitrogen atoms (62.0%), as is the case in the side chains of arginines; on the other hand, the dominant contribution for ring nitrogen atoms comes from nonpolar carbons (67.0%). Therefore, ring nitrogen atoms would be expected to require a more negative weight value to overcome this nonpolar contribution.

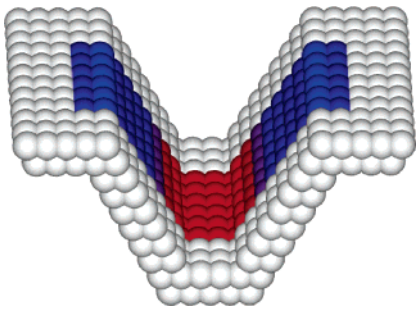

Figure 4. "V"-shaped artificial surface generated to test the ability of the method to account for different binding site shapes. The surface was created solely from nonpolar carbons so that any variations in score would derive from surface shape alone. Scoring of the atoms is represented by color, with the most hydrophobic atoms colored bright red and the least hydrophobic colored blue.

One of the motivations for developing this new method was to score the hydrophobicity of atoms within the binding site of a protein on the basis not only of their atom type but also of their local surface properties. To test the ability of the method to account for the shape and extent of the surrounding surface, an artificial atomic surface was created comprising a range of different surface shapes (Figure 4). Each atom forming this model is set as a nonpolar carbon, so any differences in the final scores can only derive from the surface properties. The atoms scored in this figure are colored to represent the full range of values for this model, with the highest scoring (most hydrophobic) atoms colored in red and the lowest scoring (least hydrophobic) colored in blue.

The "V"-shaped model depicted in Figure 4 demonstrates the effect of two surface-derived properties on the scores assigned to a protein atom: (1) the shape or curvature of the cavity surface in which the atom is found and (2) a crowding effect brought about by non-neighboring atoms coming into close proximity due to the three-dimensional architecture of the protein. As the coloring indicates, the most hydrophobic region is at the base of the "V". This is to be expected,[13] given the energetic costs of hydrating such an awkward hydrophobic surface: a severe enthalpic penalty arises from the solvent water molecules having to sacrifice a number of hydrogen bonds to fit into the narrow groove. This energetic cost is also accompanied by an entropic penalty due to the reduced reorientational freedom of the water molecules. In contrast, the least hydrophobic regions are predicted to be the ridges at the top of the "V". Again, this is to be expected considering that the water molecules hydrating this region would be able to straddle the ridge successfully thereby sacrificing minimal hydrogen bonds, resulting in a smaller enthalpic penalty (but there would still be a large entropic penalty). The surface between these two regions effectively represents a flat extended surface. Such a surface would be expected to have a hydration penalty somewhere between that of the ridge and the groove. This is accurately reflected by the scores assigned by our method. However, the scores for the atoms in this region are not

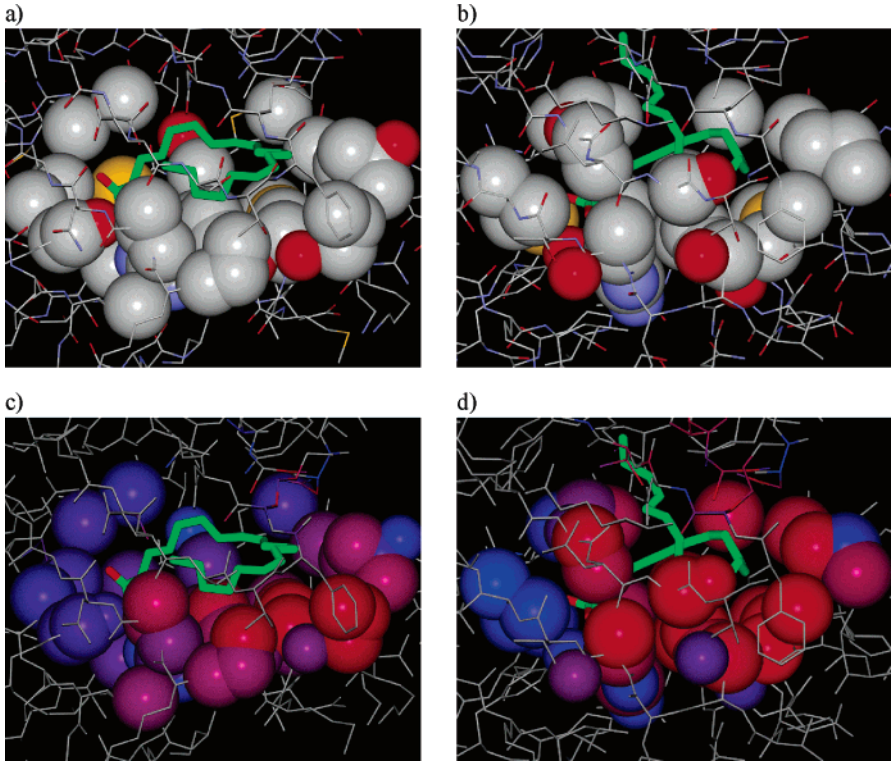

Figure 5. The largely hydrophobic binding sites of two lipid-binding proteins (1adl and 1fdq): (a) the ligand binds in a ringlike conformation to a flat surface in the binding site; (b) the ligand binds in a coiled conformation to a more curved surface; (c and d) the analysis of the different shaped binding sites by our method correctly reveals a more hydrophobic surface (brighter red) for the curved surface (d) as compared to the flatter surface (c) despite overall similarity in the atom types forming the two surfaces.

all uniform, which is due to another shape effect seen in this figure. From the base of the "V", the atoms remain hydrophobic as they move up the sides, then gradually become less so. This is the result of a crowding effect, where the atoms from the facing surface are contributing to the hydrophobicity due to their proximity. This effect can be explained by the increased enthalpic and entropic penalties associated with hydrating such a surface; that is, a water molecule would be required to sacrifice more hydrogen bonds and be subject to greater reorientational restrictions due to the proximity of the facing nonpolar surface. As the two faces of the "V" increase their separation, the crowding effect gradually disappears, and the only remaining effect is that of the extent of a flat surface, until the ridge is reached.

This ability of the method to account for local surface properties (such as shape) of hydrophobic binding sites was further tested by analyzing and comparing the ligand binding sites of two fatty acid-binding proteins: adipocyte lipid-binding protein and a fatty acid-binding protein found in the brain (PDB[62] codes 1adl and 1fdq, respectively).[4−5] In the first structure (1adl), the protein is complexed with arachidonic acid, which adopts a ringlike conformation and fits into a relatively flat region in the binding site. Conversely, the ligand in the second structure (1fdq), while differing only from the first ligand by a slightly extended aliphatic chain, adopts a coiled conformation in the binding site, reflecting the more curved nature of the binding site. Despite the difference in shape of the two surfaces, the compositions of atom types lining the binding sites are very similar (parts a and b of Figure 5). In such situations, scoring functions based on atom type alone would be unable to reveal the difference in hydrophobicity of the two surfaces. In addition, Lee and Richard's measure of the SASA[44] of atoms forming the two contrasting surface shapes reports a greater overall surface area for the flat pocket (1adl) when compared to the curved pocket (1fdq). The total SASA of all atoms forming the flat surface is 151.45 Å^2 with a mean SASA of 8.91 Å^2·atom^{-1} compared to a total SASA of 139.50 Å^2 and a mean SASA of 8.21 Å^2·atom^{-1} for the curved surface. Taking the nonpolar atoms alone, we calculate a total SASA for the flat surface of 136.06 Å^2 and a mean SASA of 9.07 Å^2·atom^{-1} compared to a total SASA of 121.91 Å^2 and a mean SASA of 8.12 Å^2·atom^{-1} for the curved surface. On the basis of these data, a method for estimating the hydrophobicity of a pocket based on the nonpolar SASA alone would incorrectly predict the flat surface as being more hydrophobic. However, through implicit consideration of local surface properties such as shape, our method is able to correctly identify the curved binding site as being the more hydrophobic (Figure 5c,d). As the coloring in parts e and f of Figure 5 demonstrates, the region around the aliphatic chain in the curved region of the binding site of 1fdq (Figure 5d) is scored as more hydrophobic (brighter red) than the corresponding region in the flat region of the binding site of 1adl (Figure 5c). Interestingly, the region in the curved area of the binding site of 1fdq around the acid group of the ligand is scored as being more polar than the corresponding region in the flat area of the binding site of 1adl. This suggests an increased complementarity between the protein and ligand in 1fdq, which is in fact reflected in the binding affinities of the two complexes. The ligand in the first complex (1adl) has a pK_i of 5.36,[4] while the ligand in the second complex (1fdq), which demonstrates a greater complementarity, has a pK_i of 7.27.[5]

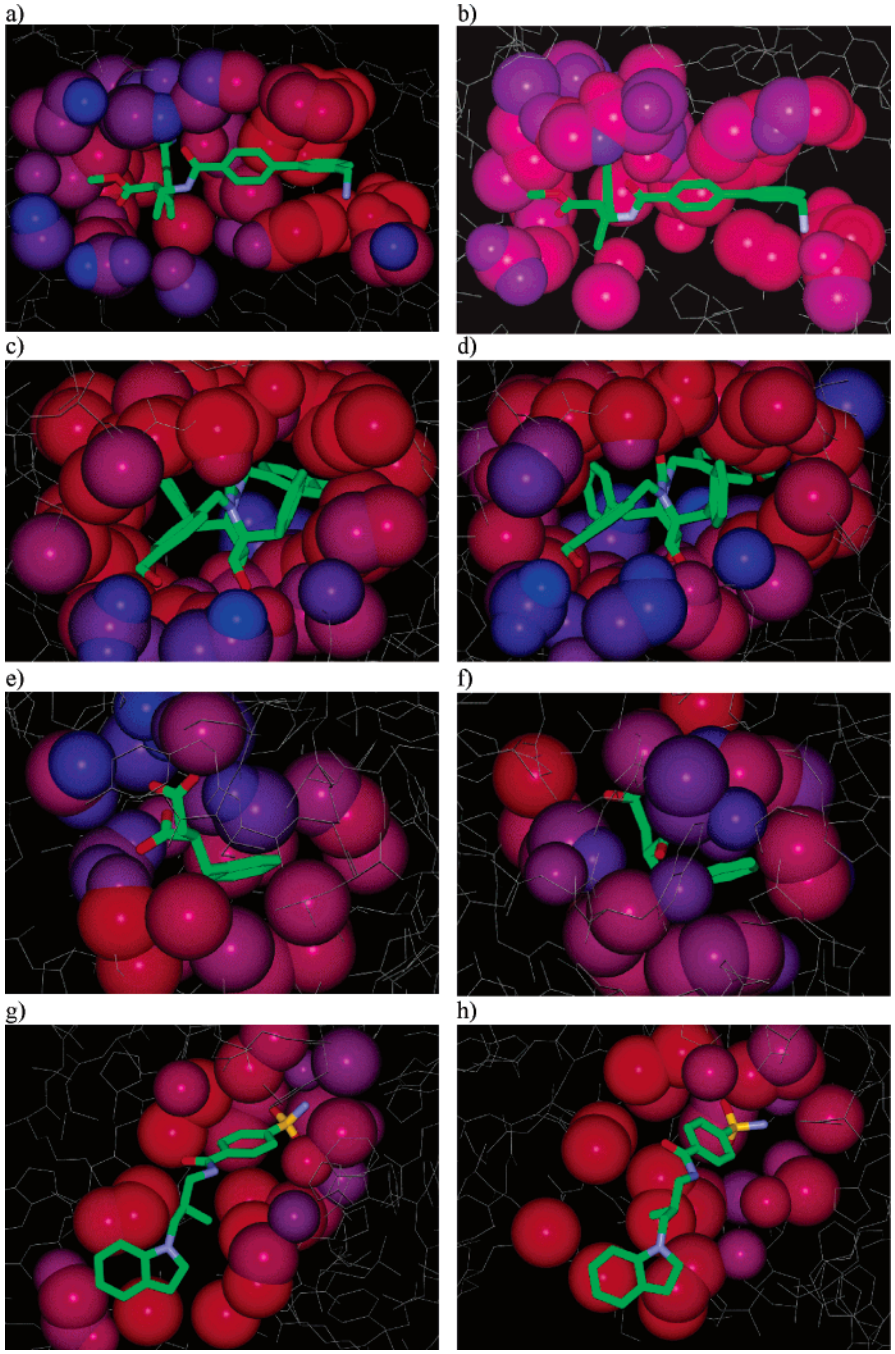

Figure 6. Analysis of four pairs of binding sites in which the same ligand binds with different affinity to the two proteins. In each case, the complex on the left demonstrates a greater complementarity between ligand and analyzed binding site, which correlates with a higher affinity interaction (Table 3): (a and b) factor Xa [1ezq] and trypsin [1f0u]; (c and d) HIV protease single mutant (V82F) [1met] and HIV protease double mutant (V82F, I84V) [1meu]; (e and f) carboxypeptidase A [1cbx] and serine carboxypeptidase II [1wht]; (g and h) carbonic anhydrase II [1if8] and carbonic anhydrase [1if7].

The relationship between the complementarity of a binding site and its ligand with the binding affinity was investigated further. We looked for examples where a partially hydrophobic ligand binds to the binding sites of two different proteins with differing affinity and assessed the ability of our method to account for the difference (Table 3). In each of the pairs of binding sites shown, the hydrogen-bonding interactions between the protein and ligand were either identical or very similar. The examples for this analysis were taken from the recently published PDBbind.[63] This database contains 800 manually inspected protein−ligand complexes, for which both experimental crystal structures and binding affinity data are available.

Table 3. The Four Pairs of Proteins Used To Demonstrate the Relationship between Binding Affinity (pK_i) and Complementarity[a]

protein description	PDB code	pK_i	complementarity
factor Xa	1ezq	9.05	1858.35
trypsin	1f0u	7.16	215.86
HIV protease (V82F)	1met	9.40	1239.27
HIV protease (V82F, I84V)	1meu	6.10	−127.49
carboxypeptidase A	1cbx	6.35	398.05
serine carboxypeptidase II	1wht	3.70	200.15
carbonic anhydrase II	1if8	10.52	−188.90
carbonic anhydrase II	1if7	9.64	−438.52

[a] The complementarity is calculated as the sum of protein atom scores within 4.0 Å of a nonpolar ligand atom minus the corresponding value for polar ligand atoms. The greater the value is, the greater the fit is between the ligand and analyzed pocket.

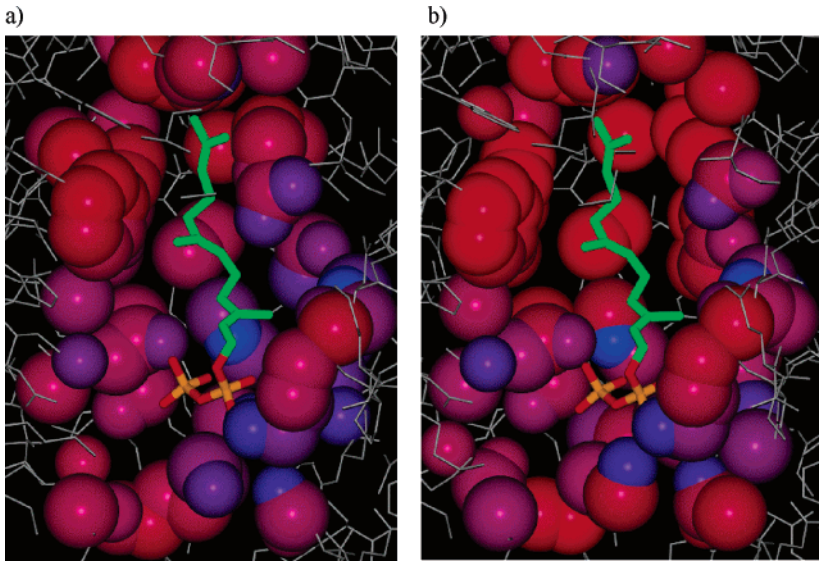

Figure 7. The unbound (a) and bound (b) conformations of farnesyltransferase are an example of induced fit, with the farnesyl pyrophosphate (FPP) ligand producing a rearrangement of the Arg202β. (In plate a, the ligand is superimposed onto the unbound conformation to highlight the binding site region.) This rearrangement results in a more hydrophobic binding site enabling the binding of the FPP substrate. The coloring of the binding site atoms represents their calculated hydrophobicity, with the more hydrophobic atoms colored bright red and the least hydrophobic colored blue.

Parts a and b of Figure 6 show the ligand binding sites of factor Xa and trypsin, respectively, both bound to the ligand RPR128515.[64] The analysis of the binding site of factor Xa indicates a greater complementarity with the ligand compared to the binding site of trypsin. For example, the protein atoms of the factor Xa binding site surrounding the nonpolar ring groups of the ligand are a brighter red (i.e., more hydrophobic) than the respective atoms in trypsin. In addition, those protein atoms of the binding site around the polar ligand groups are a brighter blue (i.e., more hydrophilic) in factor Xa than in trypsin. To quantify the complementarity between a ligand and an analyzed binding site, we have calculated the sum of protein atom scores within 4 Å of nonpolar ligand atoms and subtracted from this the corresponding value for polar ligand atoms. The greater the final value is, the greater the complementarity. The complementarity values for each of the examples shown are detailed in Table 3. The superior complementarity between both the hydrophobic and polar groups of the ligand and the binding site of Factor Xa is reflected by the difference in binding affinities. Factor Xa binds the ligand with a pK_i of 9.05[64] and Trypsin with a pK_i of 7.16.[64]

In parts c and d of Figure 6, the ligand binding sites of two mutant forms of HIV (a single mutant V82F and a double mutant V82F I84V, respectively) are shown, both bound to the ligand DMP323.[65] The additional mutation in the second structure results in a reduced nonpolar surface area in that region of the site. As in the previous example, the analysis of the binding site of the single mutant shows a greater complementarity with the ligand than the case of the double mutant. For example, the atoms forming the cavity surface in the left-hand side of the picture, which interact with two aromatic ring groups of the ligand, are scored as more hydrophobic in the single mutant (Figure 6c) than atoms in the corresponding region of the double mutant (Figure 6d). Again, this difference is reflected in the differing affinities of the two proteins for the ligand, with the single mutant binding the ligand with a pK_i of 9.40[65] and the double mutant binding the ligand with a pK_i of 6.10.[65]

Parts e and f of Figure 6 show the binding sites of carboxypeptidase A and serine carboxypeptidase II, respectively, each bound to the ligand L-benzylsuccinate.[66,67] In carboxypeptidase A, the binding site atoms surrounding the hydrophobic phenyl ring of the ligand are slightly more hydrophobic than the atoms in the corresponding region of serine carboxypeptidase II. A greater contrast is observed between the protein atoms surrounding the carboxylic acid groups of the ligand in the two binding sites, with a far more polar region predicted by our method in carboxypeptidase A. The greater complementarity evident in carboxypeptidase A is again reflected in the differing affinities of the two proteins for the ligand. Carboxypeptidase A binds the ligand with a pK_i of 6.35,[66] and serine carboxypeptidase II binds the ligand with a pK_i of 3.70.[67]

In parts g and h of Figure 6 a sulfamoyl-benzamide ligand is shown bound in two different conformations with carbonic anhydrase II.[68] In the first conformation, there are more high-scoring hydrophobic atoms around the nonpolar ligand groups than in the second conformation. In addition, the region of the binding site surrounding the polar sulfamoyl group is predicted as more polar in the first conformation. This greater complementarity is reflected in the differing affinities of the two proteins for the ligand. The ligand binds with a pK_i of 10.52[68] in the first conformation and with a pK_i of 9.64[68] in the second conformation.

Recently, a ligand-induced conformational change in the active site of farnesyltransferase (FTase) upon binding of the ligand farnesyl pyrophosphate (FPP) has been investigated.[69] The hydrophobicity of these binding sites was analyzed by generating a hydrophobicity map.[70−72] The maps were generated by evaluating the binding energy of a nonpolar probe sphere as the sum of van der Waals interactions plus the protein electrostatic desolvation in the continuum approximation.[72,73]

The inclusion of electrostatic desolvation enabled easy discrimination between neighboring hydrophobic and hydrophilic surface regions. The generated map revealed that the active site in the liganded complex[74] is much more hydrophobic than in the unliganded structure.[75] This is due to the movement of the side chain of Arg202β away from the FPP binding site, which enables the FPP substrate to bind. The region in the binding site around the terminal carbon of the FPP ligand was identified as having the best free energy of binding to the nonpolar probe, while in the unliganded FTase structure, the probe binds similarly over a wider area.

We compared the analysis of this ligand-induced conformational change with an analysis of the two conformations by our method. As the respective representations demonstrate (Figure 7), we correctly identify the liganded conformation as being more hydrophobic (brighter red) than the unliganded pocket. Our method also predicts the region surrounding the terminal carbon of the FPP ligand as being the most hydrophobic, with a more even distribution in the unliganded binding site, in agreement with the published results.

We can see that our method for assessing the relative importance of hydrophobic regions in the binding site of a protein is a simple but robust approach, superior to other simple approaches such as the use of SASA calculations. Our method is able to incorporate a number of properties that are all known to be important in defining the hydrophobicity of a nonpolar region: the polarity of the constituting atoms and the shape and extent of the local neighboring surface.

Conclusions

We have developed a new method to estimate the relative importance of hydrophobic regions within the binding site of a protein that considers both local chemical environment and surface properties such as extent, shape, and crowding. Following optimization with a purpose-built genetic algorithm, our method successfully distinguishes between those regions of the binding site more favorable for binding of nonpolar ligand atoms when compared to polar ligand atoms.

The method has also proved effective in predicting the variation in the hydrophobic effect due to changes in the shape of a surface using both artificial examples and protein binding sites. In addition, the complementarity between a ligand and our analysis of a binding site can be used to predict variations in the binding affinity of ligands to different binding sites, at least for cases where any present hydrogen-bonding interactions are similar. We have also shown that our method is capable of estimating correctly the changes in the hydrophobicity of a binding site as a consequence of a local conformational change that modifies the shape and extent of its nonpolar surface.

In summary, the method that we have developed and validated provides a fast and accurate tool for the estimation of the relative importance of hydrophobic groups in the binding site of a protein. Given the importance of hydrophobic interactions in ligand−protein binding, this method should prove its utility in the development of ligand design strategies in structure-based drug design.

References

(1) Ajay, A.; Murcko, M. A. Computational Methods to Predict Binding Free Energy in Ligand-Receptor Complexes. *J. Med. Chem.* **1995**, *38*, 4953−4967.

(2) Ekins, S.; Mirny, L.; Schuetz, E. G. A Ligand-Based Approach to Understanding Selectivity of Nuclear Hormone Receptors Pxr, Car, Fxr, Lxrα and Lxrβ. *Pharm. Res.* **2002**, *19*, 1788−1800.

(3) Meegan, M. J.; Lloyd, D. G. Advances in the Science of Estrogen Receptor Modulation. *Curr. Med. Chem.* **2003**, *10*, 1241−1253.

(4) LaLonde, J. M.; Levenson, M. A.; Roe, J. J.; Bernlohr, D. A.; Banaszak, L. J. Adipocyte Lipid-Binding Protein Complexed with Arachidonic Acid. Titration Calorimetry and X-Ray Crystallographic Studies. *J. Biol. Chem.* **1994**, *269*, 25339−25347.

(5) Balendiran, G. K.; Schnutgen, F.; Scapin, G.; Borchers, T.; Xhong, N.; Lim, K.; Godbout, R.; Spener, F.; Sacchettini, J. C. Crystal Structure and Thermodynamic Analysis of Human Brain Fatty Acid Binding Protein. *J. Biol. Chem.* **2000**, *275*, 27045−27054.

(6) Frank, H. S.; Evans, M. W. Free Volume and Entropy in Condensed Systems. *J. Chem. Phys.* **1945**, *13*, 507−532.

(7) Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* **1959**, *14*, 1−63.

(8) Nemethy, G. Structure of Water and of Aqueous Solutions. *Cryobiology* **1966**, *3*, 19−26.

(9) Blokzijl, W.; Engberts, J. B. F. N. Hydrophobic Effects. Opinons and Facts. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1545−1579.

(10) Pertsemlidis, A.; Saxena, A. M.; Soper, A. K.; Head-Gordon, T.; Glaeser, R. M. Direct Evidence for Modified Solvent Structure within the Hydration Shell of a Hydrophobic Amino Acid. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10769−10774.

(11) Rossky, P. J.; Karplus, M. Solvation. A Molecular Dynamics Study of a Dipeptide in Water. *J. Am. Chem. Soc.* **1979**, *101*, 1913−1937.

(12) Vaisman, I. I.; Brown, F. K.; Tropsha, A. Distance Dependence of Water Structure around Model Solutes. *J. Phys. Chem.* **1994**, *98*, 5559−5564.

(13) Wallqvist, A.; Berne, B. J. Molecular Dynamics Study of the Dependence of Water Solvation Free Energy on Solute Curvature and Surface Area. *J. Phys. Chem.* **1995**, *99*, 2885−2892.

(14) Wallqvist, A.; Berne, B. J. Computer Simulation of Hydrophobic Hydration Forces on Stacked Plates at Short Range. *J. Phys. Chem.* **1995**, *99*, 2893−2899.

(15) Lee, C. Y.; McCammon, J. A.; Rossky, P. J. The Structure of Liquid Water at an Extended Hydrophobic Surface. *J. Chem. Phys.* **1984**, *80*, 4448−4455.

(16) Lee, S. H.; Rossky, P. J. A Comparison of the Structure and Dynamics of Liquid Water at Hydrophobic and Hydrophilic Surfaces−A Molecular Dynamics Simulation Study. *J. Chem. Phys.* **1994**, *100*, 3334−3345.

(17) Grigera, J. R.; Kalko, S. G.; Fischbarg, J. Wall-Water Interface. A Molecular Dynamics Study. *Langmuir* **1996**, *12*, 154−158.

(18) Wallqvist, A. Polarizable Water at a Hydrophobic Wall. *Chem. Phys. Lett.* **1990**, *165*, 437−442.

(19) Lum, K.; Chandler, D.; Weeks, J. D. Hydrophobicity at Small and Large Length Scales. *J. Phys. Chem. B* **1999**, *103*, 4570−4577.

(20) Huang, D. M.; Chandler, D. Temperature and Length Scale Dependence of Hydrophobic Effects and Their Possible Implication for Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8324−8327.

(21) Glen, R. C.; Allen, S. C. Ligand-Protein Docking: Cancer Research at the Interface between Biology and Chemistry. *Curr. Med. Chem.* **2003**, *10*, 767−782.

(22) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A Review of Protein-Small Molecule Docking Methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151−166.

(23) Stahl, M.; Schulz-Gasch, T. Practical Database Screening with Docking Tools. *Ernst Schering Res. Found. Workshop* **2003**, *42*, 127−151.

(24) Audry, E.; Dubost, J. P.; Colleter, J. C.; Dallet, P. A. New Approach to Structure-Activity Relations: The "Molecular Lipophilicity Potential". *Eur. J. Med. Chem.* **1986**, *21*, 71−72.

(25) Vajda, S.; Sippl, M.; Novotny, J. Empirical Potentials and Functions for Protein Folding and Binding. *Curr. Opin. Struct. Biol.* **1997**, *7*, 222−228.

(26) Eisenberg, D.; McLachlan, A. D. Solvation Energy in Protein Folding and Binding. *Nature* **1986**, *319*, 199−203.

(27) Fauchere, J. L.; Quarendon, P.; Kaetterer, L. Estimating and Representing Hydrophobicity Potential. *J. Mol. Graphics* **1988**, *6*, 182−189.

(28) Croizet, F.; Langlois, M. H.; Dubost, J. P.; Braquet, P.; Audry, E.; Dallet, P.; Colleter, J. C. Lipophilicity Force Field Profile: An Expressive Visualization of the Lipophilicity Molecular Potential Gradient. *J. Mol. Graphics* **1990**, *8*, 153−155.

(29) Furet, P.; Sele, A.; Cohen, N. C. 3D Molecular Lipophilicity Potential Profiles: A New Tool in Molecular Modelling. *J. Mol. Graphics* **1988**, *6*, 182−189.

(30) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT: A New Method of Empirical Hydrophobic Field Calculation for CoMFA. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 545−552.
(31) Gaillard, P.; Carrupt, P.-A.; Testa, B.; Boudon, A. Molecular Lipophilicity Potential, a Tool in 3D QSAR: Method and Applications. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 83−96.
(32) Heiden, W.; Moeckel, G.; Brickmann, J. A New Approach to Analysis and Display of Local Lipophilicity/Hydrophobicity Mapped on Molecular Surfaces. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 503−514.
(33) Pixner, P.; Heiden, W.; Merx, H.; Moeckel, G.; Moller, A.; Brickmann, J. Empirical Method for the Quantification and Localisation of Molecular Hydrophobicity. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1309−1319.
(34) Rozas, I.; Martin, M. Molecular Lipophilic Potential on van der Waals Surfaces as a Tool in the Study of 4-Alkylpyrazoles. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 872−878.
(35) Suzuki, T.; Kudo, Y. Automatic Log P Estimation Based on Combined Additive Modelling Methods. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 155−198.
(36) Fujita, T.; Iwasa, J.; Hansch, C. A New Substituent Constant, Π, Derived from Partition Coefficients. *J. Am. Chem. Soc.* **1964**, *86*, 5174−5180.
(37) Rekker, R. F. In *The Hydrophobic Fragmental Constants*; Elsevier: New York, 1977.
(38) Hansch, C.; Leo, A. In *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley: New York, 1979.
(39) Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three-Dimensional Structural-Directed Quantitative Structure-Activity Relationships. Part I. Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565−577.
(40) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Recent Advances in Comparative Molecular Field Analysis (CoMFA). *Prog. Clin. Biol. Res.* **1989**, *291*, 161−165.
(41) Goodford, P. J. A Computational Procedure for Determining Energetically Favourable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(42) Verdonk, M. L.; Cole, J. C.; Taylor, R. SuperStar: A Knowledge-Based Approach of Identifying Interaction Sites in Proteins. *J. Mol. Biol.* **1999**, *289*, 1093−1108.
(43) Verdonk, M. L.; Cole, J. C.; Watson, P.; Gillet, V.; Willett, P. SuperStar: Improved Knowledge-Based Interaction Fields for Protein Binding Sites. *J. Mol. Biol.* **2001**, *307*, 841−859.
(44) Lee, B.; Richards, F. M. The Interpretation of Protein Structure: Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55*, 379−400.
(45) Pacios, L. F. Distinct Molecular Surfaces and Hydrophobicity of Amino Acid Residues in Proteins. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1427−1435.
(46) Rose, G. D.; Geselowitz, A. R.; Lesser, G. J.; Lee, R. H.; Zehfus, M. H. Hydrophobicity of Amino Acid Residues in Globular Proteins. *Science* **1985**, *229*, 834−838.
(47) Alkorta, I.; Villar, H. O. Quantum Mechanical Parameterisation for a Conformationally Dependent Hydrophobic Index. *Int. J. Quantum Chem.* **1992**, *44*, 203−218.
(48) Casari, G.; Sippl, M. J. Structure-Derived Hydrophobic Potential. Hydrophobic Potential Derived from X-ray Structures of Globular Proteins Is Able To Identify Native Folds. *J. Mol. Biol.* **1992**, *224*, 725−32.
(49) Cornette, J. L.; Cease, K. B.; Margalit, H.; Spouge, J. L.; Berzofsky, J. A.; Delist, C. Hydrophobicity Scales and Computational Techniques for Detecting Amphipathic Structures in Proteins. *J. Mol. Biol.* **1987**, *195*, 659−685.
(50) Frommel, C. The Apolar Surface of Amino Acids and Its Empirical Correlations with Hydrophobic Free Energy. *J. Theor. Biol.* **1984**, *111*, 247−260.
(51) Kantola, A.; Villar, H. O.; Loew, G. H. Atom Based Parameterization for a Conformationally Dependent Hydrophobic Index. *J. Comput. Chem.* **1991**, *12*, 681−689.
(52) Wimley, W. C.; White, S. H. Experimentally Determined Hydrophobicity Scale for Proteins at Membrane Interfaces. *Nat. Struct. Biol.* **1996**, *28*, 319−365.
(53) Williamson, M. P.; Williams, D. H. Hydrophobic Interactions Affect Hydrogen Bond Strengths in Complexes between Peptides and Vancomycin or Ristocetin. *Eur. J. Biochem.* **1984**, *138*, 345−348.
(54) Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. Extracting Hydrophobic Free Energies from Experimental Data: Relationship to Protein Folding and Theoretical Models. *Biochemistry* **1991**, *30*, 9686−9697.
(55) Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. Reconciling the Magnitude of the Microscopic and Macroscopic Hydrophobic Effects. *Science* **1991**, *252*, 106−109.
(56) Davis, L. D. *Handbook of Genetic Algorithm*s; Van Nostrand Rheinhold: New York, 1991.
(57) Goldberg, D. E. *Genetic Algorithms in Search Optimization and Machine Learning*; Addison-Wesley: Reading, MA, 1989.
(58) Holland, J. H. *Adaptation in Natural and Artificial Systems*; MIT Press: Cambridge, MA, 1992.
(59) Nissink, J. W.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A New Test Set for Validating Predictions of Protein-Ligand Interaction. *Proteins* **2002**, *49*, 457−471.
(60) Arfken, G. The Method of Steepest Descents. In *Mathematical Methods for Physicists*, 3rd ed.; Academic Press: Orlando, FL, 1985; pp 428−436.
(61) Nelder, J. A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308−313.
(62) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(63) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **2004**, *47*, 2977−2980.
(64) Maignan, S.; Guilloteau, J. P.; Pouzieux, S.; Choi-Sledeski, Y. M.; Becker, M. R.; Klein, S. I.; Ewing, W. R.; Pauls, H. W.; Spada, A. P.; Mikol, V. Crystal Structures of Human Factor Xa Complexed with Potent Inhibitors *J. Med. Chem.* **2000**, *43*, 3226−3232.
(65) Ala, P. J.; Huston, E. E.; Klabe, R. M.; McCabe, D. D.; Duke, J. L.; Rizzo, C. J.; Korant, B. D.; DeLoskey, R. J.; Lam, P. Y.; Hodge, C. N.; Chang, C. H. Molecular Basis of HIV-1 Protease Drug Resistance: Structural Analysis of Mutant Proteases Complexed with Cyclic Urea Inhibitors. *Biochemistry* **1997**, *36*, 1573−1580.
(66) Mangani, S.; Carloni, P.; Orioli, P. Crystal Structure of the Complex between Carboxypeptidase A and the Biproduct Analog Inhibitor L-benzylsuccinate at 2.0 ÅResolution. *J. Mol. Biol.* **1992**, *223*, 573−578.
(67) Bullock, T. L.; Branchaud, B.; Remington, S. J. Structure of the Complex of L-benzylsuccinate with Wheat Serine Carboxypeptidase II at 2.0-Å Resolution. *Biochemistry* **1994**, *33*, 11127−11134.
(68) Grzybowski, B. A.; Ishchenko, A. V.; Kim, C.-Y.; Topalov, G.; Chapman, R.; Christianson, D. W.; Whitesides, G. M.; Shakhnovich, E. I. Combinatorial Computational Method Gives New Picomolar Ligands for a Known Enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1270−1273.
(69) Ahmed, S.; Majeux, N.; Caflisch, A. Hydrophobicity and Functionality Maps of Farnesyltransferase. *J. Mol. Graphics Modell.* **2001**, *19*, 307−317.
(70) Nicholls, A.; Sharp, K. A.; Honig, B. Protein Folding and Association: Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons. *Proteins* **1991**, *11*, 281−296.
(71) Eisenhaber, F.; Argos, P. Hydrophobic Regions on Protein Surfaces: Definition Based on Hydration Shell Structure and a Quick Method for Their Computation. *Protein Eng.* **1996**, *9*, 1121−1133.
(72) Scarsi, M.; Majeux, N.; Caflisch, A. Hydrophobicity at Protein Surfaces. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 565−575.
(73) Scarsi, M.; Apostolakis, J.; Caflisch, A. Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions. *J. Phys. Chem. A* **1997**, *101*, 8098−8106.
(74) Park, H. W.; Boduluri, S. R.; Moomaw, J. F.; Casey, P. J.; Beese, L. S. Crystal Structure of Protein Farnesyltransferase at 2.25 Angstrom Resolution. *Science* **1997**, *275*, 1800−1804.
(75) Long, S. B.; Casey, P. J.; Beese, L. S. Cocrystal Structure of Protein Farnesyltransferase Complexed with a Farnesyl Diphosphate Substrate. *Biochemistry* **1998**, *37*, 9612−9618.

JM049524Q