Use of Amino Acid Composition to Predict Ligand-Binding Sites

Shinji Soga,† Hiroki Shirai,† Masato Kobori,† and Noriaki Hirayama*,‡

Molecular Medicine Research Laboratories, Drug Discovery Research, Astellas Pharma Inc., 21 Miyukigaoka, Tsukuba, Ibaraki 305-8585, Japan, and Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, Boseidai, Isehara, Kanagawa 259-1193, Japan

Received May 30, 2006

A novel method for predicting the binding sites for druglike compounds on the surface of proteins was developed on the basis of the specific amino acid composition observed at the ligand-binding sites of ligand−protein complexes determined by X-ray analysis. A profile representing the preference of each of the 20 standard amino acids at the binding sites of druglike molecules was obtained for a small set of high-quality complex structures. An index termed propensity for ligand binding (PLB) was created from these profiles. The PLB index was used to predict the propensity of binding for 804 ligands at all potential binding sites on the proteins whose structures were determined by X-ray analysis. If the sites with the first two highest PLB indices are taken into consideration, the successfully predicted sites reached a high percentage of 86. The PLB prediction is relatively simple, but the validation study showed that it is both fast and accurate to detect ligand-binding sites, especially the binding sites of druglike molecules. Therefore, the PLB index can be used to predict the ligand-binding sites of uncharacterized protein structures and also to identify novel drug-binding sites of known drug targets.

1. INTRODUCTION

The specific binding of a ligand to a therapeutic target molecule is the key to drug action. Each ligand binds preferentially to a specific site on the surface of the target molecule. The binding site is usually located in a concavity on the surface of the target protein whose configuration, in most cases, is highly specialized for the binding of a particular group of compounds.

Identification of the ligand-binding site for each specific protein molecule is crucially important when trying to find a suitable drug molecule for the target, but it is also important to understand the function of the protein.

Recently, the number of X-ray structures of ligand−protein complexes has increased markedly. A sufficient number of reliable ligand−protein complex structures are now available in the Protein Data Bank (PDB)[1] for use in the systematic study of the common characteristics of molecular interactions between ligands and proteins.

Since the binding site of a drug is considered to be highly specific to its molecular characteristics, the binding site must have a distinct character significantly different from those of similar concavities on the target molecule. Although each binding site is made up of specific amino acid residues from mostly noncontiguous regions of the protein, the site is expected to be richer in certain specific amino acid residues and poorer in others.

Several algorithms for predicting ligand-binding sites have been published over the past 10 years. They can be divided into three major categories: (i) geometric algorithms,[2] (ii) probe-mapping algorithms,[3] and (iii) physical potential algorithms.[4] Although these algorithms have each achieved a measure of success, none of them take the clustering of specific amino acids at the ligand-binding sites into account, which is an important effect that needs to be considered.

The compositions of the amino acids at the ligand-binding sites of the ligand−protein complexes whose structures were determined by X-ray crystallography were examined in this study. It is of particular interest to study the frequency of appearance at the binding sites for each of the 20 standard amino acids and check for specific amino acid compositions. Attempts to use this information for the prediction of ligand-binding sites were also made.

In this study, only ligand−protein complexes with drug or druglike compounds were considered. A detailed analysis of the amino acid compositions around the binding sites for these compounds revealed that there were clear propensities for the presence of specific amino acids at the binding site of each drug or druglike compound. A novel propensity for ligand binding (PLB) index was developed on the basis of the idiosyncratic amino acid profile of each ligand-binding site. The PLB index's ability to predict ligand-binding sites was tested on a systematically collected data set. Validation of this method indicated that it is both fast and accurate.

2. MATERIALS AND METHODS

In this study, the amino acid compositions of the binding sites of druglike compounds were analyzed using a training data set that consisted of the most accurate and diverse complex structures from the PDB. The PLB index was defined in terms of the specific profile of the amino acid composition observed at each binding site of the druglike compounds. The predictive power of the PLB index was then evaluated using a test data set consisting of a different set of accurate complex structures from the PDB (none of the

* Corresponding author tel.: +81 463 93 1121; e-mail: hirayama@is.icc.u-tokai.ac.jp.
† Astellas Pharma Inc.
‡ Tokai University School of Medicine.

10.1021/ci6002202 CCC: $37.00
Published on Web 01/23/2007

ones used in the training data set were included). The two data sets were compiled using the PDB data downloaded on July 20, 2005.

2.1. Identification of Concavities. Small organic molecules, especially drug molecules, typically bind at concavities on the surface of the protein. In this study, a program named Alpha Site Finder[5] implemented in the software system MOE[6] was used to detect concavities on the surface of proteins. Concavities identified by Alpha Site Finder are characterized by a cluster of small spheres called "α-spheres." The positions and characteristics of the spheres are calculated on the basis of the geometry and character of the protein surface. The cluster of α-spheres represents the shape and size of the concavity. Since Alpha Site Finder can map every depression on the protein surface, it is ideal for the purposes of this study. Alpha Site Finder usually identifies multiple concavities on each protein molecule; thus, the goal is to identify which of these is the most probable drug binding site.

2.2. Training Data Set. Since the training data set must represent typical drug−protein complexes, the candidates were carefully selected using the criteria described below.

2.2.1. High-Quality X-ray Structures. The amino acids at the binding sites are identified on the basis of the coordinates of ligand atoms other than hydrogen. Reliable positions of non-hydrogen atoms are essential for this study. For the training data set, in particular, the highest-quality X-ray structures were selected in the interests of accuracy.

In order to obtain the crystal structures in which all non-hydrogen atoms were unambiguously determined, structures were selected using the following criteria: a R_{free} value of less than 0.24, a resolution value of less than or equal to 2.5 Å, occupancy factors of 1.0 for all non-hydrogen atoms, and atomic displacement parameters of less than 30 Å^2 for all non-hydrogen atoms. If a protein was multimeric, only a monomer with the smallest atomic displacement parameters for their non-hydrogen atoms was considered.

2.2.2. Complexes with Druglike Ligands. The drug-likeness profile of a molecule, which comprises multiple molecular descriptors, is useful for determining how much like a drug the molecule is.[7] The ranges of values of the 14 descriptors given in Table 1 cover 85% of the drugs used clinically in Japan now. These values were used to eliminate nondruglike ligands. When 12 of the descriptor values for a particular ligand fell within these ranges, the ligand was considered to be a druglike molecule, and the corresponding complexes were taken into consideration.

There are many ligands with multiple phosphorus atoms in the PDB. However, from the drug-likeness viewpoint, they are not suitable. Therefore, all ligands containing more than one phosphorus atom per molecule were eliminated.

2.2.3. Nonredundant Structures. When a complex contained multiple identical ligands, only the one with the smallest average atomic displacement factor was considered. If a ligand formed complexes with multiple homologous proteins, only the complex structure with the smallest R_{free} was considered. The proteins in the training data set are considered to be sufficiently diverse since their maximum percent identity is 48%.

These selection criteria resulted in a training data set that includes 41 complex structures. The chemical structures of the ligands contained in the data set are listed in Figure 1, together with the PDB codes. As Figure 1 shows, the structures of ligands are chemically diverse.

Table 1. Value Distributions for 14 Molecular Descriptors That Apply to 85% of the Drugs Used Clinically in Japan[a]

descriptor	ranges	
weight	165	555
SlogP	−1.18	5.30
SMR	4.34	14.46
TPSA	13.0	165
density	0.73	0.99
vdw_area	165	497
vdw_vol	181	623
a_acc	1	7
a_don	0	6
a_hyd	6	26
KierA1	7.82	26.3
KierA2	3.13	11.8
KierA3	1.48	7.32
KierFlex	1.68	8.82

[a] weight: molecular weight; vdw_area: area of van der Waals surface calculated using a connection table approximation; vdw_vol: van der Waals volume calculated using a connection table approximation; density: molecular mass density (molecular weight divided by vdw_vol); a_acc: number of hydrogen-bond acceptor atoms (not counting acidic atoms but counting atoms that are both hydrogen-bond donors and acceptors such as −OH); a_don: number of hydrogen-bond donor atoms (not counting basic atoms but counting atoms that are both hydrogen-bond donors and acceptors such as −OH); a_hyd: number of hydrophobic atoms; SlogP:[9] calculated hydrophobicity by Crippen; SMR:[9] calculated molar refractivity by Crippen; TPSA:[10] topological polar surface area; KierA1, KierA2, KierA3, and KierFlex:[11] molecular connectivity indices.

2.3. Test Data Set. 2.3.1. Data Selection. A test data set of the complex structures was constructed in order to evaluate the predictive power of the PLB index. The proteins that are homologous to ones in the training data set were not included in the test data set. Sequence similarity was judged by a Basic Local Alignment Search Tool[8] search of which E-value threshold was below 1.0. The minimum value of percent identity between similar sequences was 48. By relaxing four of the training data set selection conditions, the proteins for the test data set were further narrowed down. First, the occupancy factors of the non-hydrogen atoms in a protein could be less than 1.0. Second, the atomic displacement factors of the non-hydrogen atoms in a ligand could be greater than 30 Å^2. Third, only the range of molecular weight was used to select molecules. Finally, ligands were allowed to contain multiple phosphorus atoms. If a protein was multimeric, only a monomer with the largest occupancy factors and the smallest atomic displacement parameters for their non-hydrogen atoms was considered. The test data set consisted of 756 complex structures with 804 ligands. Although the selection conditions were relaxed, the X-ray structures were still high-quality and many druglike molecules were included, making the test data set suitable for evaluating the PLB index.

2.3.2. Calculation of Concavities. All concavities in the protein structures were identified and catalogued. Alpha Site Finder found 15 892 concavities in 756 protein structures. Although the number of binding sites was 778, 804 of the ligands were bound. This means that multiple ligands were bound in some proteins. Since it was likely that some of the concavities were too small to accommodate ligands, they were not considered as possible binding sites. The number

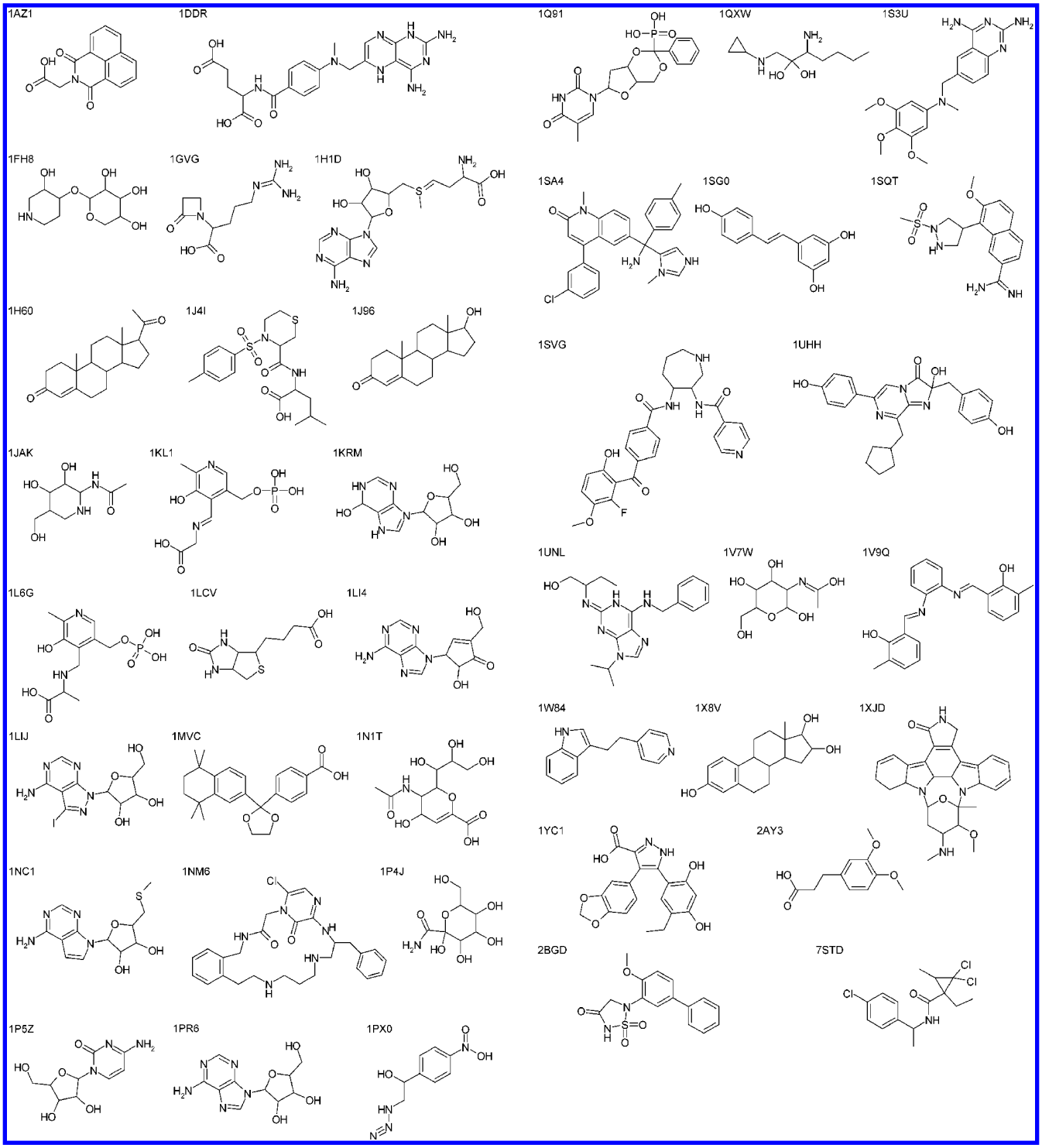

Figure 1. Chemical structures of the ligands in the training data set together with their PDB codes.

of α-spheres per concavity was used to judge size. Those concavities smaller than or equal to the minimum size of the smallest concavity in the training data set were considered trivial and eliminated. The number of concavities of realistic size for ligand binding was determined to be 15 232.

2.4. Specific Amino Acid Composition at Ligand-Binding Site. The amino acids surrounding each ligand in the training data set were examined. If any non-hydrogen atom in an amino acid was within 4.5 Å of any non-hydrogen atom of a ligand, it was expected that the amino acid would interact with the ligand. Therefore, 4.5 Å was used as the cutoff value to determine the amino acid composition of the binding site. MOE was used to calculate this distance. If the number of amino acids of type x at a binding site is represented as N_x, the composition of the amino acid of type x at the binding site, CA(x), is defined as

$$\mathrm{CA}(x) = N_x / \sum_{y=1}^{20} N_y \qquad (1)$$

Here, N_y denotes the number of amino acids of type y at the binding site. The denominator in the above equation is the total number of amino acid residues at the binding site.

Table 2. Amino Acid Compositions Normalized Using the 20 Standard Amino Acids[a]

x	CA(*x*)	SA(*x*)	RA(*x*)
A	0.050	0.072	0.701
C	0.021	0.013	1.650
D	0.065	0.064	1.015
E	0.061	0.064	0.956
F	0.080	0.041	1.952
G	0.058	0.073	0.788
H	0.056	0.025	2.286
I	0.050	0.050	1.006
K	0.028	0.060	0.468
L	0.087	0.084	1.045
M	0.037	0.020	1.894
N	0.041	0.051	0.811
P	0.010	0.049	0.212
Q	0.027	0.040	0.669
R	0.047	0.052	0.916
S	0.056	0.064	0.883
T	0.041	0.057	0.730
V	0.056	0.064	0.884
W	0.058	0.019	3.084
Y	0.068	0.041	1.672

[a] CA(*x*) denotes the composition of amino acid of type *x* at the ligand-biding sites of the proteins in the training data set. SA(*x*) denotes the composition of amino acid of type *x* on the surface of the proteins in the test data set. RA(*x*) denotes the ratio of CA(*x*) to SA(*x*).

The incidence of amino acid residues on the protein surface was also investigated. The amino acids on the surface of the protein were identified by calculating the solvent-accessible surface of each amino acid using a probe sphere with a radius of 1.4 Å. The solvent-accessible surface was calculated using MOE. If the probe came into contact with a non-hydrogen atom in a residue, the residue was regarded as being on the surface of the protein. If the number of amino acids of type *x* on the surface of a protein is N_{sx}, the rate of occurrence of amino acids of type *x* on the surface of the protein, SA(*x*), is defined as

$$\mathrm{SA}(x) = N_{sx} / \sum_{y=1}^{20} N_{sy} \tag{2}$$

Here, N_{sy} denotes the number of amino acids of type *y* on the surface of the proteins. The denominator is the total number of all amino acids on the surface of the protein. SA(*x*) was determined using all protein structures in the test data set. The normalized values of CA and SA using the 20 standard amino acids are given in Table 2.

The ratio of CA(*x*) to SA(*x*), designated as RA(*x*),

$$\mathrm{RA}(x) = \mathrm{CA}(x)/\mathrm{SA}(x) \tag{3}$$

is the rate of occurrence of an amino acid of type *x* at the ligand-binding site. RA(*x*) is also considered to be the preference factor for an amino acid of type *x* at the binding site. By using a linear combination of the RAs for all 20 standard amino acids, an index of PLB can be defined with respect to a particular concavity, *i*, as follows:

$$\mathrm{PLB}_i = \sum_{x=1}^{20} N_{i,x}\mathrm{RA}(x) \tag{4}$$

The weighting-factor, $N_{i,x}$, denotes the number of amino acids of type *x* found in concavity *i*.

In order to distinguish the binding site with the most potential from the other concavities on a protein, the PLB values should be normalized by all concavities in the protein. *Z*-scored PLBs are used for this purpose. If there are *M* concavities in a protein, the *Z*-scored PLB for concavity *i* is calculated as follows:

$$Z_{\mathrm{PLB}_i} = \frac{\mathrm{PLB}_i - \mu}{\sigma} \tag{5}$$

Here,

$$\mu = \frac{\sum_{i=1}^{M} \mathrm{PLB}_i}{M} \tag{6}$$

and

$$\sigma = \sqrt{\frac{\sum_{i=1}^{M} (\mathrm{PLB}_i - \mu)^2}{M}} \tag{7}$$

Hereafter, Z_{PLB_i} is designated simply as PLB. By use of the PLB index, it is possible to judge the probability of ligand binding for a given concavity. The concavity with the highest PLB index in a protein is the most probable site for ligand binding. Accordingly, a smaller PLB index implies that ligand binding is less probable.

3. RESULTS AND DISCUSSION

3.1. Frequency of Amino Acid Presence at the Ligand-Binding Sites and on the Surface of Proteins. The CAs and SAs for the 20 standard amino acids were calculated for all proteins in the training and test data sets, respectively. The results are shown in Figure 2.

The CA values for many of the amino acids are markedly different from their corresponding SA values. The RA values, shown in Figure 3, illustrate these differences clearly.

The amino acids are sorted in ascending order of RA in this figure. The effect is striking, and the implications are particularly intriguing. Since the frequency rates of the aromatic residues and Met at the ligand-binding sites were high, these residues can be considered binding-site-philic residues. By the same token, Pro, Lys, Gln, and Ala can be considered binding-site-phobic since they were not often found. These profiles clearly show that the amino acid composition at each ligand-binding site is highly specific. It follows, then, that these characteristic profiles could be used to predict ligand-binding site locations on the basis of the amino acid compositions around the concavities.

3.2. Prediction of Ligand-Binding Sites Using the PLB Index. The PLB indices were calculated for 15 232 concavities found in 756 proteins. Since most proteins have multiple concavities, the PLB index would be an ideal tool if it narrows down the options. The concavity with the highest PLB index can be considered to be the most probable ligand-binding site. There were 611 cases where the concavity with the highest PLB index corresponded to the true binding site, which is 79% of the true ligand-binding sites in the test data set. If the sites with the first two highest PLB indices are

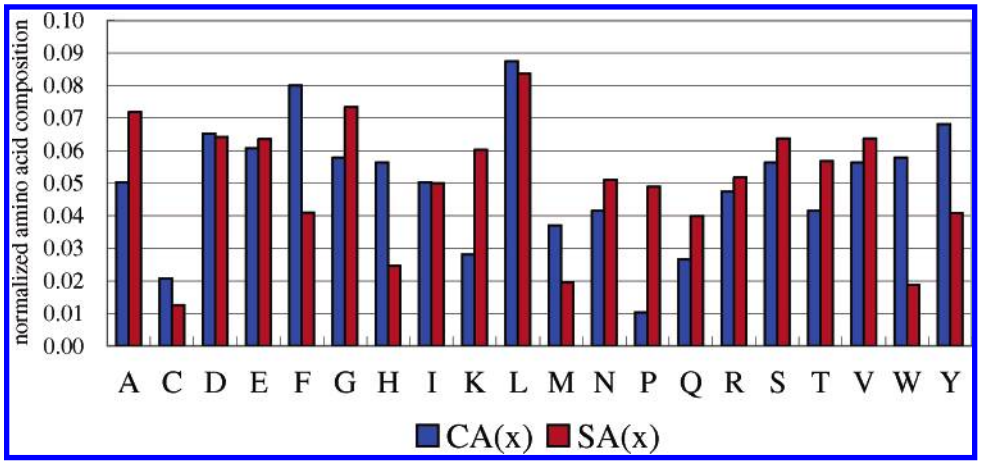

Figure 2. Normalized composition of the 20 standard amino acids. Blue: composition at the ligand-binding sites of the proteins in the training data set (CA). Red: composition on the surface of the proteins in the test data set (SA).

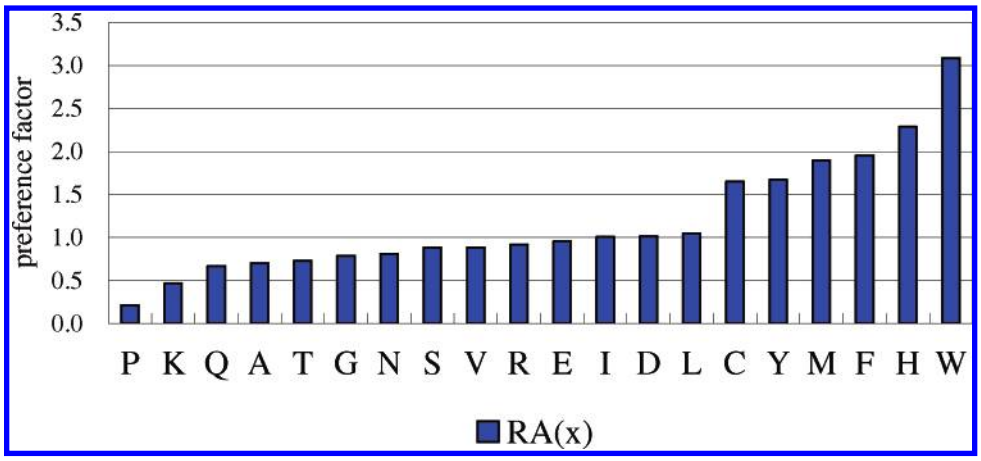

Figure 3. Preference factors for the 20 standard amino acids (RA = CA/SA).

taken into consideration, the successfully predicted sites reached to the higher percentage of 86. From a practical point of view, the latter selection is reasonable.

There were 1255 concavities with a PLB index greater than or equal to 1.2. Of these, 649 were true concavities. Hence, the PLB index prediction enrichment rate was calculated to be 649/1255/(778/15 232) = 10.12. This indicates that the PLB index is useful for the identification of ligand-binding sites.

Many nondruglike ligands were included in the test data set. Since the PLB index was derived from a training data set that contained many drug and druglike molecules, it is of great interest to see how well the PLB index predicted the binding sites of druglike compounds in the test data set. Each molecule's degree of "druglikeness" was judged using the druglikeness profile mentioned above. The number of true ligand-binding sites where druglike compounds bind is 126. The highest PLB index predicted 110 binding sites, which is 87% of the true binding sites. If the concavities with the top two PLB indices are taken into consideration, 120 binding sites can be identified. It covers as much as 95% of the true binding sites.

The results clearly show that the PLB index is useful for detecting not only small-molecule binding sites in general, but also binding sites specific to druglike molecules.

3.3. Two Typical Examples of Prediction. The following two examples illustrate how the PLB index is a more effective way to distinguish the true binding site from other concavities in a protein.

A total of 17 concavities were detected in the protein structure of a protein−ligand complex of carbonic anhydrase II (PDB code: 1OKL),[12] the eight largest of which are shown in Figure 4a. The concavities are represented by a cluster of α-spheres. Each α-sphere is classified as either hydrophilic or hydrophobic (red or white, respectively), depending on whether it is in a location conducive to hydrogen bonding or not.

The volume of a concavity can be expressed using several indices, such as the number of the α-spheres in the concavity, the number of protein atoms in contact with the α-spheres, and the number of amino acids in contact with the α-spheres. These indices were calculated for the 17 concavities and are listed in Table 3.

The concavity surrounded by the green circle is the largest with respect to the number of contact atoms. A closeup of the landscape around the concavity is shown in Figure 4b. The size and shape indicate that it could be a ligand-binding site. The PLB index of the concavity is, however, only 0.39. This value is significantly small, and it turns out that, indeed, the concavity was not used by the ligand. The highest PLB index, 2.01, was assigned to the concavity indicated by the red circle. The other indices for this concavity were not the highest, as is shown in Table 3. X-ray analysis revealed that the inhibitor was in actuality bound in this concavity, as

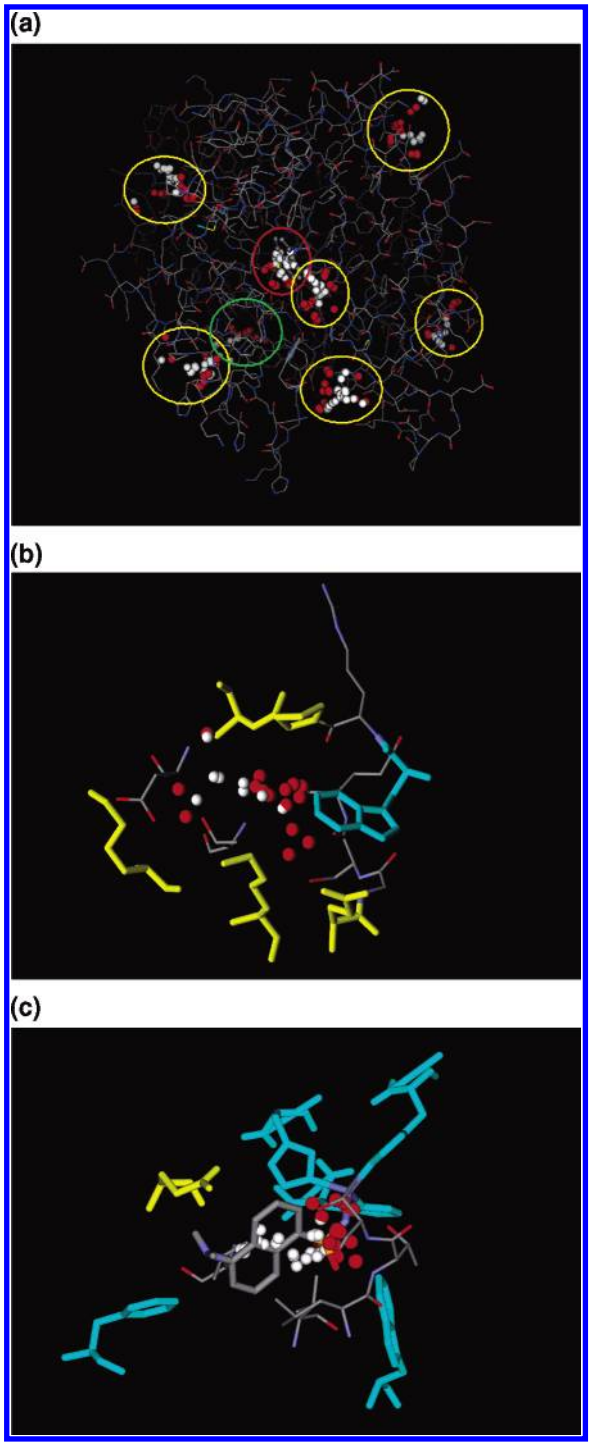

Figure 4. (a) The eight main concavities located in the protein structure of the complex between carbonic anhydrase II and a ligand of 5-(dimethylamino)-1-naphthalenesulfonamide (PDB code: 1OKL). The red and white spheres denote α-spheres (hydrophilic and hydrophobic, respectively). The concavities are circled. The carbon, nitrogen, oxygen, and sulfur atoms are gray, blue, red, and yellow, respectively. (b) A closeup of the largest concavity with respect to the number of atoms in contact with the α-spheres (circled in green in Figure 4a). In this figure, the red and white spheres are α-spheres, the binding-site-philic aromatic residues and Met are light blue, and the binding-site-phobic residues of Pro, Lys, Gln, and Ala are yellow. The PLB index of this concavity is 0.39. (c) A closeup of the true binding site circled in red in part a. The ligand is displayed by a thick stick model in the center of this figure. The red and white spheres denote α-spheres, the binding-site-philic aromatic residues and Met are light blue, and the binding-site-phobic residues of Pro, Lys, Gln, and Ala are yellow. The PLB index of this concavity is 2.01.

Table 3. Indices Characterizing the Concavity Sites Located in the Protein Structure of 1OKL

site	number of contact atoms around the concavity	number of amino acids around the concavity	number of α spheres in the concavity	PLB index
1	57	12	24	0.39
2	51	13	29	1.09
3	50	9	32	−0.59
4	50	12	42	1.88
5	47	11	45	1.35
6	47	10	27	0.25
7	45	10	31	−0.71
8	44	11	44	2.01 true binding site
9	39	8	20	−0.64
10	38	6	13	−0.63
11	30	7	15	−0.73
12	30	7	10	−0.80
13	28	7	11	−1.03
14	26	6	13	−1.34
15	25	8	12	0.33
16	24	6	11	−0.17
17	21	8	12	−0.64

depicted in Figure 4c. This result is indicative of the PLB index's usefulness for distinguishing the correct concavity among multiple possibilities.

Two distinct concavities were found in the protein structure of a ligand−protein complex of retinoic acid receptor γ-1 (PDB code: 1FCZ).[13] The size and shape of one concavity, shown in Figure 5a, made it look like a promising ligand-binding site.

The PLB index of 1.84 for this concavity, however, was appreciably smaller than the corresponding value (3.04) of the concavity shown in Figure 5b. Although there were some binding-site-phobic residues present in both concavities, the former contained five, including two Pro's, which are slightly more binding-site-phobic than the other three, and the latter contained two Lys's. In addition, there were eight binding-site-philic residues in the latter concavity compared with four in the former. The characteristics of the two concavities were distinctively different, and the PLB index indicated that the latter was much more likely to be a ligand-binding site. The inhibitor BMS181156 was actually bound at this site in the crystal structure of the complex.

Laskowski et al.[14] reported that ligand-binding sites tend to be associated with the largest clefts in the surface of a protein. However, the examples above clearly show that binding sites cannot be determined on the basis of size alone. The chemical properties resulting from amino acid composition significantly contribute to the determination of ligand-binding sites. The PLB index takes both the chemical properties and size of the concavity into consideration (eq 4), and its high rate of successful prediction reflects this.

4. CONCLUSIONS

Identification of the potential binding sites of small molecules on a particular target molecule is an important issue in drug discovery. On the basis of the assumption that drug-binding sites on target molecules have specific amino acid compositions, the compositions around the binding sites of drug or druglike compounds were determined by examining the X-ray structures of ligand-protein complexes. As expected, the amino acid compositions around the ligand-

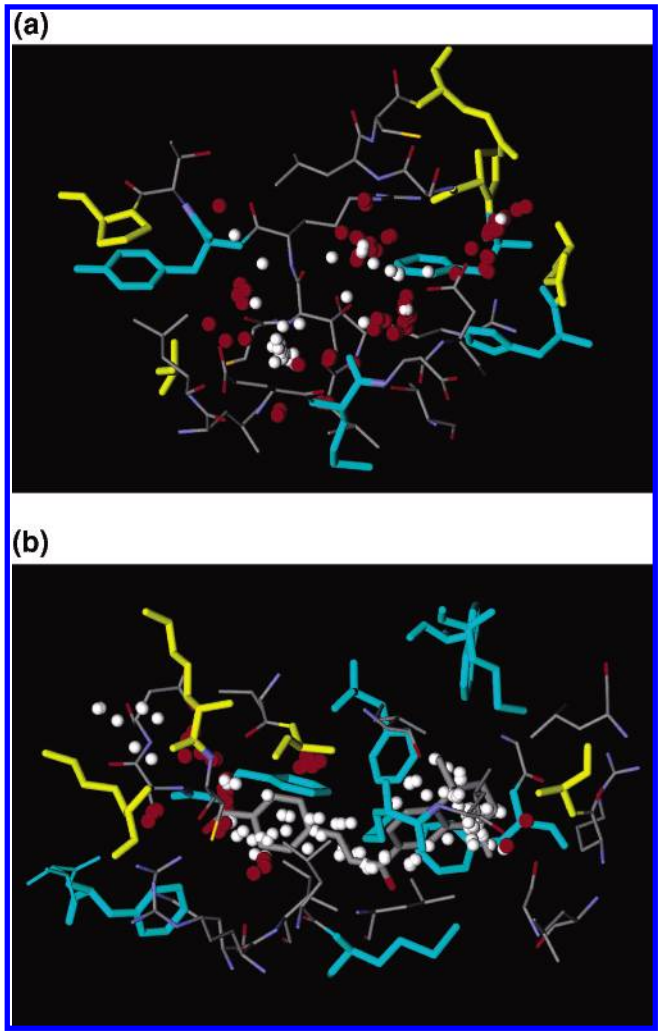

Figure 5. (a) One of the two distinct concavities located in the protein structure of the complex between retinoic acid receptor γ-1 and the ligand BMS181156 (PDB code: 1FCZ). The red and white spheres denote α-spheres; the binding-site-philic aromatic residues and Met are light blue, and the binding-site-phobic residues of Pro, Lys, Gln, and Ala are yellow. The PLB index of this concavity is 1.84. (b) One of the two distinct concavities with the largest PLB index, 3.04. This concavity corresponds to the true binding site of the ligand. The ligand is displayed by a thick stick model in the center of this figure. The red and white spheres denote α-spheres; the binding-site-philic aromatic residues and Met are light blue, and the binding-site-phobic residues of Pro, Lys, Gln, and Ala are yellow.

binding sites were markedly different from those on the surfaces of the proteins. The specific amino acid composition at each ligand-binding site was used to create a novel PLB index. The results of this study show that the PLB index is a good predictor of ligand-binding sites. It is also clear that the molecular interplay between the amino acids at a given site is crucial to the creation of the ligand-binding sites.

It is particularly interesting that the binding sites can be predicted accurately by the specific amino acid composition surrounding the concavities on the surface of proteins. From a practical point of view, this prediction method is useful because the accurate determination of the positions of the amino acids within the concavities would not necessarily be required. The method may be applicable to relatively low-resolution X-ray structures and those constructed using homology modeling. The PLB index would also be useful for identifying ligand-binding sites on novel target molecules with unknown ligands. It is clear that using the PLB index to predict ligand-binding sites could become a useful research tool for various aspects of drug discovery.

REFERENCES AND NOTES

(1) 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.
(2) Laskowski, R. A. SURFNET: A Program for Visualizing Molecular Surfaces, Cavities, and Intermolecular Interactions. *J. Mol. Graphics* **1995**, *13*, 323−330.
(3) Dennis, S.; Kortvelyesi, T.; Vajda, S. Computational Mapping Identifies the Binding Sites of Organic Solvents on Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4290−4295.
(4) An, J.; Totrov, M.; Abagyan, R. Comprehensive Identification of "Druggable" Protein Ligand Binding Sites. *Genome Inf.* **2004**, *15*, 31−41.
(5) Edelsbrunner, H.; Facello, M.; Fu, R.; Liang, J. Measuring Proteins and Voids in Proteins. *Proceedings of the 28th Annual Hawaii International Conference on Systems Science*; Publisher: Place of Publication, 1995; pp 256−264.
(6) *MOE (Molecular Operating Environment)*, version 2005.06; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2006.
(7) Horio, K.; Goto, J.; Hirayama, N. A Simple Method To Improve the Odds in Finding 'Lead-Like' Compounds from a Chemical Library. *Chem. Pharm. Bull.* Submitted.
(8) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403−410.
(9) Labute, P. A Widely Applicable Set of Descriptors. *J. Mol. Graphics Modell.* **2000**, *18*, 464−477.
(10) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714−3717.
(11) Hall, L. H.; Kier, L. B. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling. *Rev. Comput. Chem.* **1991**, *2*, 367−442.
(12) Nair, S. K.; Elbaum, D.; Christianson, D. W. Unexpected Binding Mode of the Sulfonamide Fluorophore 5-Dimethylamino-1-naphthalene Sulfonamide to Human Carbonic Anhydrase II. Implications for the Development of a Zinc Biosensor. *J. Biol. Chem.* **1996**, *271*, 1003−1007.
(13) Klaholz, B. P.; Mitschler, A.; Moras, D. Structural Basis for Isotype Selectivity of the Human Retinoic Acid Nuclear Receptor. *J. Mol. Biol.* **2000**, *302*, 155−170.
(14) Laskowski, R. A.; Luscombe, N. M.; Swindells, M. B.; Thornton, J. M. Protein Clefts in Molecular Recognition and Function. *Protein Sci.* **1996**, *5*, 2438−2452.

CI6002202