Identification of the Druggable Concavity in Homology Models Using the PLB Index

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, 143 Shimokasuya, Isehara, Kanagawa 259-1143, Japan

Received July 3, 2007

Identification of the druggable concavity, in which drug-like molecules are highly inclined to bind, is an important step in structure-based drug design. We previously proposed an index named PLB (propensity for ligand binding), which is based on the amino acid composition characteristically observed at the small molecule binding sites in the X-ray structures of the complexes between proteins and drug-like small molecules. The PLB index was proven to be useful in identifying the druggable concavities in the quality X-ray structures of proteins. Here, we apply the PLB to predicting the druggable concavity in target proteins using the structures of homologous proteins constructed by homology modeling. In this study, we assembled a set of reference proteins that were accurately determined by X-ray analysis in forms of complexes with drug-like small molecules. Homology models for the reference protein were constructed using multiple homologous proteins as templates. The PLB index was then used to predict the druggable concavity. If the template protein in a complex with a drug-like small molecule was used, the druggable concavity was predicted well, with a prediction rate of 78%. When only the apo protein was available as the template, the practical prediction rate was 71%. Interestingly, even when the percent sequence identity between the reference and template proteins was lower than 30, the PLB index could successfully identify the druggable concavity in some cases. This study demonstrates the practical value of applying the PLB index to identifying the drugabble concavity in the homology model.

1. INTRODUCTION

The specific interaction between a drug molecule and its corresponding target protein is a crucial event in the action of the drug. The surface of target proteins usually contains multiple concavities to which small molecules may bind. Generally, however, a particular drug molecule will bind at a very specific concavity to initiate the subsequent drug action. Identification and characterization of such specific concavities, defined as druggable concavities in this paper, are indispensable to structure-based drug discovery.

We have investigated the binding sites, i.e., concavities, of drug-like molecules in various target proteins whose structures were accurately determined by X-ray analysis. Results showed that the amino acids clustering at these concavities are highly specific, leading us to develop a simple discrimination index named PLB (propensity for ligand binding).[1] Validation studies using relatively high-quality X-ray structures demonstrated that the PLB index is a suitable means of identifying the specific concavity to which a drug-like molecule should bind.

Many drug discovery projects still suffer a lack of suitable high-quality X-ray structures of the target molecule. Against this background, however, the rapid increase in high-quality X-ray structures in the Protein Data Bank (PDB)[2] has markedly facilitated the identification of proteins homologous to the target protein from among the thousands-strong pool of proteins in the PDB. Further support is provided by the homology modeling method, which is also now a reliable means[3] of constructing a reasonable three-dimensional structure of an object protein by the best use of sequence similarity and X-ray structure of a template protein. Given the opportunities offered by these advances, the ability to predict or identify druggable concavities in structures constructed by the homology modeling method would be useful.

We were particularly interested in examining whether the PLB index can be applied to this challenging problem. Here, we investigated the possibility of predicting druggable concavities in homology models using the PLB index.

2. MATERIALS AND METHODS

Figure 1 shows a schematic diagram of the flow of druggable concavity prediction based on homology modeling.

2.1. Use of Quality Protein Structures. A set of 13 416 quality protein structures in the PDB was downloaded on Feb. 21, 2007 and termed the quality dataset. Although not compared in detail in this study, the highest quality structures were used to avoid various expected ambiguities caused by low-quality structures. Quality dataset structures were required to meet two criteria, an X-ray diffraction resolution of less than or equal to 2.5 Å, to ensure that the data were derived from quality structures, and an R_{free} value of less than or equal to 0.24, to ensure adequate concordance between the diffraction data and the structure.

2.2. Selection of Reference Structures. A set of complexes between nonredundant proteins and small molecules

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

10.1021/ci7002363 CCC: $37.00
Published on Web 10/13/2007

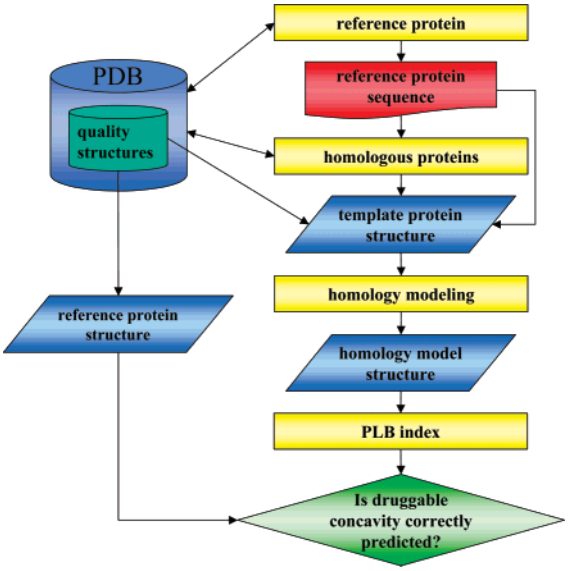

Figure 1. Flow chart of druggable concavity prediction in the homology model.

from the quality dataset was selected and used as reference structures throughout the study. The following four criteria were used to extract the reference structures. First, the small molecule bound in the reference protein had to be a drug-like molecule, as determined using criteria reported previously.[4] Second, occupancy factors and atomic displacement factors of all non-hydrogen atoms of drug-like molecules had to be 1.0 and less than 30 Å^2, respectively. This ensured that the atomic positions of the drug-like molecules were unambiguously determined and, hence, that, on the basis of the atomic positions, the amino acids around the concavity where the drug-like molecules were bound could be clearly identified. Third, the proteins had to be nonredundant, as determined using the "Non-redundant PDB chain set (NR-PDB)" resource (http://www.ncbi.nlm.nih.gov/Structure/VAST/nrpdb.html): this resource clusters all PDB chains into groups of similar chains by sequence similarity and assigns a specific "Group ID" to each group, which was then used to judge nonredundancy. A *p*-value of 10×10^{-7} was used to judge sequence similarity by BLAST.[5] Where there were multiple proteins with the same group ID, we selected structures which included more drug-like ligand. Fourth, each reference protein had to have multiple homologous proteins in the quality dataset with greater than 80% sequence length alignment with the reference protein. In addition, at least one such protein must be a complex with a small molecule. The strict requirements of the second criterion were dropped for the atomic parameters of the small molecule in the homologous proteins.

On examination, only 15 reference structures were found to fulfill the above requirements (Table 1). The number of homologous proteins for each reference protein is given in Table 2, in which the PDB codes are arranged according to the secondary structure content in ascending order. Table 2 also gives the number of homologous proteins complexed with small molecules. The number of homologous proteins without small molecules is given in Table 4. Since these homologous proteins were used as templates to construct the homology models corresponding to the reference structures, they are named as template proteins hereafter. As shown in Tables 2 and 4, various homologues were selected with percent identities from 20s to 90s.

2.3. Homology Modeling. The homology model corresponding to the reference protein was constructed by the homology modeling method using the X-ray coordinates of a template protein and the amino acid sequence of the reference protein. We used the homology modeling algorithms implemented in the software system MOE.[6] The effect of small molecules bound in the template protein was not taken into account when the model structure was built. Since there are multiple template proteins with different sequence similarities, the degree of concordance between a reference structure and its homology model should vary.

2.4. PLB and R Indices. Specific amino acids tend to cluster at a druggable concavity in a protein. In the previous study,[1] we found that aromatic residues and Met appeared highly frequently at the druggable concavities. On the contrary, the occurrence rates of Pro, Lys, Gln, and Ala at the druggable concavities were significantly low. On the basis of this observation, we introduce the PLB index, which identifies the druggable concavity among other concavities in the protein. The PLB index is defined as follows:

$$\mathrm{PLB}_i = \sum_{x=1}^{20} N_i(x)\,\mathrm{RA}(x) \qquad (1)$$

RA(x) is the ratio of the occurrence rate of amino acid x at the druggable concavities to the occurrence rate of amino acid x on the surface of the proteins. $N(x)$ denotes the number of amino acid x found in a concavity i. Values are summed for all 20 standard amino acids. To distinguish the most druggable concavity from other concavities on a protein surface, the PLB values should be normalized by all concavities in the protein. Z-scored PLBs are used for this purpose. Hereafter Z-scored PLB is designated simply as PLB. Use of the PLB index allows the druggability of each concavity in the protein to be estimated. In this study, the PLB index was applied to identify the druggable concavity in the homology model in order to check how the PLB index can be applied to the homology models, starting from multiple template structures with different sequence similarity.

If the degree of sequence homology between a template and its reference protein is small, the structural discrepancy between the homology model and the reference structure is expected to be large. This means that the concavity in the homology model is to some extent deformed compared to the druggable concavity in the reference structure. This in turn requires evaluation of how accurately the druggable concavity in the reference structure was reconstructed in the homology model. A simple method is to compare the amino acids occurring around these concavities: if there are n amino acids at the druggable concavity in a reference protein and m corresponding amino acids at the relevant concavity in a homology model, the ratio of m/n indicates the goodness of modeling. This ratio, designated as the R index, was used to evaluate the goodness of concavities formed in the homology model, using amino acids located within 4.5 Å of non-hydrogen atoms of the drug-like molecule.

Table 1. Fifteen Reference Proteins Extracted from the Protein Data Bank[2]

PDB code	chain ID	group ID	protein name	enzyme code	ligand name[a]	drug-likeness[b]
1ZUA	X	2	aldo−keto reductase family 1 member B10	1.1.1.−	TOL	14
1E0X	A	33	endo-1,4-β -xylanase A precursor	3.2.1.8	X2F-XYS	13
1BK9		57	phospholipase A2, acidic	3.1.1.4	PBP	12
1TU6	A	59	cathepsin K precursor	3.4.22.38	FSP	14
1W4P	A	73	ribonuclease pancreatic precursor	3.1.27.5	UM3	13
1JZF	A	78	azurin precursor		RTB	13
1YMS	A	126	β-lactamase CTX-M-9a		NBF	14
2WEA		128	penicillopepsin	3.4.23.20	PP6	12
1HEE	A	174	carboxypeptidase A1 precursor	3.4.17.1	ZN-LHY	13
1WBI	A	198	avidin-related protein 2 precursor		BTN	14
1CXV	A	218	collagenase 3 precursor	3.4.24.−	CBP	14
1H4G	A	237	glycoside hydrolase		FXP	13
1TT1	A	473	glutamate receptor, ionotropic kainate 2 precursor		KAI	12
2CYB	A	478	tyrosyl-tRNA synthetase	6.1.1.1	TYR	13
1H60	A	678	pentaerythritol tetranitrate reductase		FMN	14

[a] Abbreviations for ligands used in the PDB. [b] The drug-likeness of small molecules complexed with the proteins was judged using the 14 descriptors in a previous paper.[4] The ranges of these descriptors were calculated to cover 85% of all drugs now used clinically in Japan. The number of the descriptors whose values were within the relevant ranges was used as an index of drug-likeness. For example, a drug-likeness index of 12 means that 12 of 14 descriptors had values within the above ranges.

Table 2. Number of Homologous Proteins Complexed with Small Molecules[a]

% identity[b]	1CXV	1TU6	1JZF	1ZUA	1H60	1W4P	1HEE	1WBI	2WEA	1E0X	1BK9	1TT1	1YMS	1H4G	2CYB	total
20%s	0	2	0	0	3	0	0	6	8	0	0	0	0	0	1	20
30%s	0	0	0	5	4	1	0	4	1	3	0	0	2	0	0	20
40%s	0	7	0	11	5	1	0	0	0	1	3	0	4	3	1	36
50%s	6	4	0	1	1	0	0	1	3	1	2	4	0	1	1	25
60%s	3	0	0	0	0	0	0	1	1	0	0	0	1	0	0	6
70%s	0	0	0	8	0	1	0	0	0	0	0	0	0	0	0	9
80%s	0	0	0	0	0	0	0	2	0	0	0	1	1	0	0	4
90%s	1	2	1	0	6	0	2	0	1	5	0	0	3	0	0	21
total	10	15	1	25	19	3	2	14	14	10	5	5	11	4	3	141

[a] The PDB code of the reference proteins is written at the top of the table. PDB codes were arranged according to the secondary structure content of the structure in ascending order. [b] Value of the % identity is the range of percent identity between a reference and a homologous protein. 20%s means the range between 20% and 30%.

Table 3. Prediction Results of Druggable Concavities in Homology Models Constructed from Template Structures with Small Molecules[a]

% identity[b]	1CXV	1TU6	1JZF	1ZUA	1H60	1W4P	1HEE	1WBI	2WEA	1E0X	1BK9	1TT1	1YMS	1H4G	2CYB	av
20%s		1.00			0.67			0.83	1.00						1.00	0.90
30%s				0.80	1.00	1.00		0.50	1.00	0.67			0.50			0.75
40%s		0.71		1.00	1.00	1.00				1.00	1.00		0.25	0.67	1.00	0.83
50%s	0.83	1.00		1.00	1.00			1.00	1.00	1.00	1.00	0.75		1.00	1.00	0.92
60%s	1.00							1.00	1.00				1.00			1.00
70%s				1.00		1.00										1.00
80%s								1.00				1.00	1.00			1.00
90%s	1.00	1.00	1.00		1.00		1.00		1.00	1.00			1.00			1.00
av	0.90	0.87	1.00	0.96	0.95	1.00	1.00	0.79	1.00	0.90	1.00	0.80	0.64	0.75	1.00	

[a] Values indicate the success rate of prediction. For 1CXV, there were six homology models with a 50%s identity, of which the druggable concavities of five were predicted successfully, giving a success rate of 0.83. The average means the successful prediction rate averaged for each row or column. [b] See footnote *b* of Table 2.

3. RESULTS AND DISCUSSION

3.1. Identification of the Druggable Concavity in Homology Models Constructed from Template Proteins with Small Molecules. Using a template protein that was homologous to a reference protein, a model structure was constructed using the homology modeling method. The concavity formed in the homology model was then compared to that in the reference structure using the *R* and PLB indices. If we set $R \geq 0.5$ and PLB ≥ 1.2, the druggable concavities of reference protein structures are adequately predicted. These thresholds were therefore used in this study. Evaluation was conducted on the following basis. If the template structure contained a small molecule in the relevant concavity, the protein was expected to be modeled more or less to optimally recognize the small molecule. In the homology model constructed from such a template structure, the druggable concavity should be reconstructed better than that modeled from the template structure without a bound small molecule. It was therefore expected that the latter case would produce a significantly worse prediction rate.

Table 4. Number of Homologous Proteins in the Apo State

% identity[a]	1CXV	1TU6	1JZF	1ZUA	1H60	1W4P	1HEE	1WBI	2WEA	1E0X	1BK9	1TT1	1YMS	1H4G	2CYB	total
20%s	1	1	0	1	0	0	0	1	5	0	0	0	1	0	1	11
30%s	0	2	1	1	0	4	1	2	3	9	0	3	7	1	2	36
40%s	0	3	1	2	0	0	3	0	0	6	9	0	7	5	1	37
50%s	0	3	0	0	0	0	1	0	0	0	8	1	2	2	0	17
60%s	2	0	2	0	0	1	2	2	0	1	1	0	0	1	0	12
70%s	0	0	1	1	0	2	0	0	0	1	0	0	0	0	0	5
80%s	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	3
90%s	0	1	10	0	0	6	4	0	0	1	2	1	2	1	0	28
total	3	10	15	5	0	13	11	5	8	18	20	5	22	10	4	149

[a] The PDB code of the reference protein is written at the top of the table. PDB codes are arranged according to the secondary structure content of the structure in ascending order. The column indicates the range of percent identity between the reference protein and homologous protein. 20%s means the range between 20% and 30%.

Table 5. Prediction Results of Druggable Concavities in Homology Models Constructed from the Apo Template Structures[a]

% identity[b]	1CXV	1TU6	1JZF	1ZUA	1H60	1W4P	1HEE	1WBI	2WEA	1E0X	1BK9	1TT1	1YMS	1H4G	2CYB	av
20%s	0.00	1.00		0.00				0.00	0.60				1.00		1.00	0.55
30%s		1.00	0.00	1.00		0.50	1.00	1.00	1.00	0.44		0.67	0.86	1.00	1.00	0.72
40%s		0.67	0.00	0.50			1.00			0.83	1.00		0.86	1.00	1.00	0.86
50%s		1.00					1.00				1.00	1.00	1.00	1.00		1.00
60%s	1.00		0.50			1.00	1.00	1.00		1.00	1.00			1.00		0.92
70%s			1.00	1.00		0.50				0.00						0.60
80%s													1.00			1.00
90%s		1.00	0.50			0.67	1.00			1.00	1.00	1.00	1.00	1.00		0.75
av	0.67	0.90	0.47	0.60	-	0.62	1.00	0.80	0.75	0.61	1.00	0.80	0.91	1.00	1.00	

[a] Cell values indicate the same as in Table 3, i.e., the success rate of prediction. [b] See footnote *b* of Table 2.

Table 6. Homologous Structures of 2CYB

reference			homologous proteins				
PDB code	chain ID	% identity	PDB code	chain ID	l igand name	protein name	species
2CYB	A	54.6	1J1U	A	TYR	tyrosyl-tRNA synthetase	Methanococcus jannaschii
2CYB	A	40.0	2CYC	B	TYR	tyrosyl-tRNA synthetase	Pyrococcus horikoshii
2CYB	A	22.6	1R6T	A	TYM	tryptophanyl-tRNA synthetase, cytoplasmic	Homo sapiens
2CYB	A	43.1	2CYA	A		tyrosyl-tRNA synthetase	Aeropyrum pernix
2CYB	A	38.6	1N3L	A		tyrosyl-tRNA synthetase, cytoplasmic	Homo sapiens
2CYB	A	37.6	1Q11	A		tyrosyl-tRNA synthetase, cytoplasmic	Homo sapiens
2CYB	A	22.5	1R6T	B		tryptophanyl-tRNA synthetase, cytoplasmic	Homo sapiens

Using the template proteins complexed with small molecules, 141 homology models were constructed as shown in Table 2. The percent identity ranges varied widely, from 20%s to 90%s in general, but were unfortunately not equally wide in all cases. The rates of successfully predicted druggable concavities are given in Table 3. Where multiple homologous proteins occurred in the same bin of percentage, their success rates were averaged. For instance, in the case of 1CXV, there were six homologous proteins in the bin of 50%s. *R* and PLB indices for five of the homology models were equal to or greater than 0.5 and 1.2, respectively, giving a success rate in this bin of $5/6 = 0.83$. In the 60%s bin, however, the indices for all homology models exceed the threshold values, so the success rate was 1.0. Apart from three exceptions, the druggable concavities were predicted reasonably well, with 38 of 49 sampling points showing a ratio of 1.0. The prediction rate was a markedly high 78%. Further, even in those cases in which the percent identity was less than 30, the prediction rate was greater than 0.83 in 4 of 5 sampling points in the table. A reasonably good prediction rate was obtained in these difficult situations. These results unequivocally demonstrate that the prediction of druggable concavities using the PLB is highly accurate for template structures having small molecules in the relevant concavities.

3.2. Identification of the Druggable Concavity in Homology Models cConstructed from Template Proteins without Small Molecules. Predicting the druggable concavity in homology models derived from apo protein structures is highly challenging. Using template proteins in the apo state, 149 homology models were constructed by homology modeling as given in Table 4. Since 1H60 has no homologous apo protein in the quality dataset, the total number of proteins in this table is 14. The rates of successfully predicted druggable concavities are given in Table 5. Of 56 points evaluated, a total of 37 had a prediction rate of 1.0. The complete success rate was 66%. From a practical standpoint, points with prediction rates greater than 0.83 can be regarded as successfully predicted, increasing the success rate marginally to 71%. These results demonstrated that if a quality

```
2CYB_A  27   ETKEKPRAYYGYEPSGE-IHLGHWMTVQKLMDLQEA-GFEIIYLIADLHAYLNEKGTFEE  84
             E K+    Y G  PS E  H+GH++       LQ+     ++   D   YL +  T ++
1R6T_B  55   ENKKPFYLYTGRGPSSEAXHVGHLIPFIFTKWLQDVFNVPLVIQXTDDEKYLWKDLTLDQ  114

2CYB_A  85   IAEVADYNKKVFIALGLDESRAKFVLGSEYQ-LSRDYVLDVLKMARITTLNRARRSMDEV  143
                A  N K  IA G D ++        +Y   S  +  +V+K+ +  T N+ +      +
1R6T_B  115  AYGDAVENAKDIIACGFDINKTFIFSDLDYXGXSSGFYKNVVKIQKHVTFNQVK----GI  170

2CYB_A  144  SRRKEDPMYSQMIYPLMQAL--------DIAHLGVD---LAVGGIDQRKIHMLARENLPR  192
                +   + ++ +P +QA           I     D   L    IDQ       R+  PR
1R6T_B  171  FGFTDSDCIGKISFPAIQAAPSFSNSFPQIFRDRTDIQCLIPCAIDQDPYFRXTRDVAPR  230

2CYB_A  193  LGYSSPVCLHTPILVGLDGQKMSSSKGNYISVRDPPEEVERKIRKAYCPAGVVEENPILD  252
             +GY  P  LH+     L G   S       I + D  ++++ K+ K +  +G
1R6T_B  231  IGYPKPALLHSTFFPALQGPNSS------IFLTDTAKQIKTKVNK-HAFSG---------  274

2CYB_A  253  IAKYHILPRFGKIYVERDAKFGGDVE------YASF--------EELAEDFKSGQLHPLD  298
                       G+  +E   +FGG+ +       Y +F        E++ +D+ SG     +
1R6T_B  275  ----------GRDTIEEHRQFGGNCDVDVSFXYLTFFLEDDDKLEQIRKDYTSGAXLTGE  324

2CYB_A  299  LKIAVAKYLNMLLEDARKR  317
             LK A+ + L  L+ + + R
1R6T_B  325  LKKALIEYLQPLIAEHQAR  343
```

Figure 2. Sequence alignment between 2CYB_A and 1R6T_B in BLAST format. The red rectangle indicates the ligand-binding site in the reference protein.

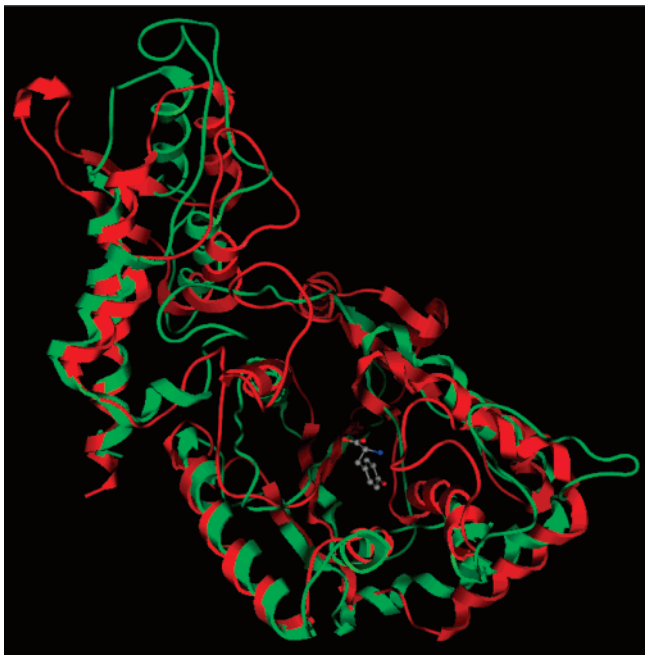

Figure 3. Reference structure (PDB code: 2CYB_A) (red) and homology model (green) constructed from 1R6T_B. The small molecule in the reference structure is shown by a ball-and-stick model.

template protein structure in the apo state was available, the PLB index could adequately predict the druggable concavity in the drug−target protein whose structure was constructed by homology modeling.

3.3. Prediction Rate vs Secondary Structure Content. Performance of the PLB index in predicting the druggable concavity appears reasonably good from a practical point of view. We anticipated that the percent identity would be closely related to the prediction rate. To our surprise, however, druggable concavities were successfully predicted in some cases in which the percent identity was markedly low, e.g., 1YMS, 1H4G, and 2CYB; in these cases, the relevant concavities were perfectly predicted even though the template proteins were in the apo state. We were particularly intrigued by this result. Subsequent evaluation of potential explanations indicated the role of the secondary structures. Using the definition of secondary structures by Kabsch and Sander,[7] contents of the secondary structure in the reference proteins were 0.46, 0.47, 0.5, 0.5, 0.51, 0.52, 0.54, 0.56, 0.57, 0.57, 0.58, 0.61, 0.62, 0.63, and 0.65 for 1CXV, 1TU6, 1JZF, 1ZUA, 1H60, 1W4P, 1HEE, 1WBI, 2WEA, 1E0X, 1BK9, 1TT1, 1YMS, 1H4G, and 2CYB, respectively. Although no direct relationship between the secondary structure content and prediction rate was seen, the secondary structure content of the above three proteins exceeded 60%, indicating that the druggable concavities could be identified by the PLB index provided that the secondary structure content was high, even if the percent identity was low. This trend appeared to hold generally in the set of reference proteins. In the case of 1TU6, however, the prediction rates were markedly high over a wide range of percent identities in spite of the significantly low secondary structure content. Detailed analysis of the role of secondary structure content in the prediction of druggable concavity therefore awaits the accumulation of a quality dataset in the PDB. Until then, however, this general trend can be considered a rule of thumb.

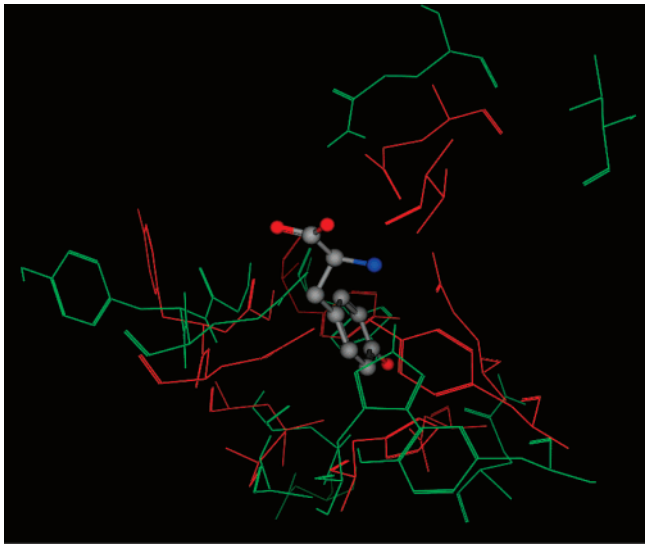

Figure 4. Close-up of the druggable concavity in Figure 3. The reference structure and homology model are shown by red and green lines, respectively. The small molecule is shown by a ball-and-stick model.

3.4. Typical Prediction of the Druggable Concavity in a Homology Model. In this example, the reference protein is tyrosyl-tRNA synthetase from *Archaeoglobus fulgidus* (2CYB). The homologous proteins used as template structures are given in Table 6. Since the percent identity is low, i.e., 22.5%, tryptophanyl-tRNA synthetase from human cytoplasma (1R6T) seems to be a practical example. The sequence alignment used for homology modeling is shown in Figure 2, and the homology model and X-ray structure of tyrosyl-tRNA synthetase are superimposed in Figure 3. Although the overall structures correspond reasonably well, the dispositions and conformations of amino acids around the concavity differ significantly, as shown in the expanded view (Figure 4). Nevertheless, prediction of the druggable concavity was successful. The reason for this may be that the PLB index does not depend on the detailed structures around the concavity and is good at capturing the spirit of the druggable concavity. This is clearly the most advantageous aspect of prediction using the PLB index.

4. CONCLUSIONS

In the postgenomic era, structural information regarding disease-related proteins is now increasing at an explosive

rate. This in turn has resulted in ever higher demands for identification of the druggable concavity in drug-target proteins from sequence data alone. Thanks to recent technical advances, the number of protein structures determined by X-ray analysis is increasing, facilitating the identification of X-ray structures reasonably homologous to a target amino acid sequence. The modeling of a protein structure from the amino acid sequence has been one of the most attractive topics in protein science, but, despite much effort, ab initio modeling from the sequence only remains challenging. Against this background, the homology modeling method, which provides the best use of the three-dimensional structure of homologous proteins, has now advanced to practical use. The availability of a suitable methodology that could exploit these data to predict the drug-binding concavity in the target protein would greatly assist drug discovery.

In this paper, we investigated the use of the PLB index to meet this need. The PLB index, based on the specific amino acid compositions at drug-binding concavities in drug-binding proteins, serves as an index to assess the druggability of the relevant concavity in the drug−target protein. Results showed that the PLB index successfully identified the druggable concavity for a set of quality reference proteins selected for this study. For template structures with a small molecule bound in the relevant concavity, the success rate of prediction was a markedly high 78%. Moreover, the success rate was 71% even in more practical cases in which the template structure was not a complex with a small molecule. The present study demonstrates that the PLB index can be used to identify the druggable concavity in homology models. Although the essential condition for the proper application of the PLB index is now a high-quality template structure, the expanding pool of quality structures in the PDB will clearly increase the applicability of this method.

REFERENCES AND NOTES

(1)) Soga, S.; Shirai, H.; Kobori, M.; Hirayama, N. Use of amino acid composition to predict ligand-binding sites. *J. Chem. Inf. Model.* **2007**, *47*, 400−406.
(2)) Berman, H. M.; Westbrook, J.; Fenz, 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.
(3)) Marti-Renom, M. A.; Madhusudhan, M. S.; Fiser, A.; Rost, B.; Sali, A. Reliability of assessment of protein structure prediction methods. *Structure* **2002**, *10*, 435−440.
(4)) Horio, K.; Goto, J.; Muta, H.; Hirayama, N. A Simple method to improve the odds in finding 'lead-like' compounds from a chemical library. *Chem. Pharm. Bull.* **2007**, *55*, 980−984.
(5)) 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.
(6)) *MOE (Molecular Operating Environment*), version 2006.0801; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2006.
(7)) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577−2637.

CI7002363