IPSJ Transactions on Bioinformatics Vol. 3 2–9 (Feb. 2010)

Database/Software Paper

AltPS: A Structural Alignment Tool for Protein Surfaces Using Similarity of Local Atomic Environments

Ryoichi Minai[†1,*1] and Yo Matsuo[†1,*2]

We have developed an alignment tool for comparing protein local surfaces (AltPS). This program enables efficient exhaustive searches of the entire protein surfaces, using a feature vector for a surface atom with 6 to 18 elements to describe the geometrical and physicochemical properties in the local environment, without referring sequence or fold homology. AltPS runs on a personal computer with the input of a pair of PDB coordinates and outputs similarity scores between identified similar surfaces, alignments of the surface atoms, and corresponding superposed coordinates, based on cluster analysis of similar surface regions. In this report, we present some results on the application of AltPS to several protein pairs with similar functions to identify similar functional sites. AltPS can be downloaded from http://d-search.atnifty.com/research.html

1. Introduction

The three-dimensional (3D) structure and function of a protein are closely related to each other. It is therefore possible to predict function of proteins on the basis of structural similarities. In the annotation of a novel protein structure, a frequently used strategy to identify functional sites is to focus on the structural similarity between proteins having the same type of fold, even when no proteins with known functions can be found on the basis of sequence homology. Global alignment tools such as DALI [1], SSAP [2], VAST [3], and CE [4] are extensively employed for structural alignment during the annotation of such structural information. However, although these tools permit the alignment of the entire protein structure, they are not effective in determining the similar regions on the protein surface and the extent of similarity of these regions. On the other hand, it has been reported that proteins having different folds can bind with the same ligand and perform identical functions. A comparison of such proteins in terms of their surface regions having identical functions shows that these proteins share structural similarities [5)–7].

For structural comparisons focusing on protein surfaces without relying on the sequence or fold homology, several existing tools are currently available, such as 3D-surfer [8], eF-seek [9], Cavbase [10], Protein Functional Surfaces [11], and IsoCleft [12]. The targets of these programs, however, are restricted to potential function sites (e.g., ligand-binding sites or cavity regions), rather than the entire protein surface. Other methods that can compare the entire protein surface have also been reported, such as MolLoc [13] and SUMOMO [14], but these methods do not provide stand-alone software that can be conveniently used.

To address such issues, we have developed an alignment tool called AltPS that can compare the entire local surface of proteins and identify similar regions. This tool uses a feature vector to define the geometrical and physicochemical properties of individual atoms and their neighboring atoms on a protein surface, which makes it possible to perform comparative calculations on target regions and to evaluate structural similarity. AltPS is derived from our previously reported method [15] for calculating the similarity between ligand-binding sites, whose effectiveness has been validated. The alignment method reported in our previous study provides the best superposition of entire target regions, but does not identify similar partial regions on the target. For this reason, targets in that study needed to be confined to significant regions such as ligand-binding sites. The proposed method in this study, however, can be used to detect similar partial regions of entire target protein surfaces. This tool can be conveniently used for rapid calculations, even when information on the sequence or fold homology is not available. AltPS accepts input data in the protein data bank (PDB) format, runs calculations for determining similar regions of the proteins being compared, and produces superposed structures and atom alignments as output. In addition, this tool also generates a statistical score that describes the degree of similarity between the identified similar regions, which helps in the identification of important regions.

†1 Department of Supramolecular Biology, Yokohama City University
∗1 Presently with SINKEN Yamanashi Office
∗2 Presently with OncoTherapy Science, Inc.

2. Methods

2.1 Definition of Similarities in Local Atomic Environments

As described in our previous study [15], we have used a feature vector to concisely define the characteristics of a protein surface region. For any solvent-accessible atom (surface atom) i of a protein, its local physicochemical environment is represented by a vector (as shown below), wherein the quantity (a value computed from relative distances) of an adjacent surface atom j, which is a solvent-accessible atom and is located within $3d_c$ of atom i, is summed for each component of its physicochemical properties.

$$\mathbf{C}_i(k) = \sum_{j \in V^i} \left(h_{ij}^{AT1}(k), h_{ij}^{AT2}(k), ..., h_{ij}^{AT6}(k) \right) (k = 1, 2, 3) \quad (1)$$

In this formulation, V^i is the set of all adjacent surface atoms for atom i, $AT1$ to $AT6$ represents the physicochemical types (cation ($AT1$), anion ($AT2$), hydrogen-bond donor ($AT3$), hydrogen-bond acceptor ($AT4$), hydrophobic ($AT5$), and none of these types ($AT6$)) of atom j, as defined according to the PATTY (programmable atom typer) algorithm [16]. Here, $h_{ij}^{ATx}(k) = a\,(1 - |d_{ij} - (k-1)d_c|/d_c)$ if $|d_{ij} - (k-1)d_c|/d_c \leq 1$ and atom j is of type ATx; otherwise, $h_{ij}^{ATx}(k) = 0$. If atom j is both a donor and an acceptor, and $x = 3$ or 4, then $a = 0.5$; otherwise, $a = 1$. d_{ij} denotes the distance between atoms i and j, and d_c denotes the standard distance, whose default value is 3.2 Å. In addition, when the value of k decreases, $\mathbf{C}_i(k)$ is assumed to retain the characteristics of the immediate vicinity of atom i.

We use the Tanimoto coefficients (Tc) of vectors $\mathbf{C}_i(k)$ and $\mathbf{C}_j(k)$ to measure the degree of similarity between atoms i and j, which then gives the following two definitions: $s1_{ij} = Tc\,(\mathbf{C}_i(1), \mathbf{C}_j(1))$ and $s3_{ij} = \sum_{k=1}^{3} Tc\,(\mathbf{C}_i(k), \mathbf{C}_j(k))\,/3$. As compared to $s1$, $s3$ reflects the similarity in a larger surrounding environment. Here, Tc of two vectors $\mathbf{A}$ and $\mathbf{B}$ is represented as $Tc\,(\mathbf{A}, \mathbf{B}) = \mathbf{A} \cdot \mathbf{B} / \left\{ |\mathbf{A}|^2 + |\mathbf{B}|^2 - \mathbf{A} \cdot \mathbf{B} \right\}$.

2.2 Algorithm

For two proteins, q and t, to be investigated, their 3D data (PDB format) are input into AltPS. By using the procedure given below, the similar local surface regions of the proteins are extracted.

(1) The solvent-accessible surface area (ASA) of each protein atom is calculated and then solvent-accessible atoms (ASA > 0) are extracted as surface atoms. Moreover, the feature vectors, $\mathbf{C}_i(k)$, of these atoms are calculated by using Eq. (1).

(2) By considering individual surface atoms as central atoms, surface atoms with a radius of less than 5 Å are grouped into a "local area."

(3) Two local areas lq_i and lt_j are selected, in which the central atoms of lq_i and lt_j are surface atom i on protein q and surface atom j on protein t, respectively. If there is a strong similarity between the feature vectors of atoms i and j ($s3_{ij} \geq 0.8$), the 3D structure of lq_i will be superposed on lt_j, and the similarity is calculated as follows.

- **a.** The similarities, $s1_{ab}$, are calculated for each atom pair, where a is an atom of lq_i and b is an atom of lt_j.
- **b.** A pair of similar atom triads ($s1_{ab} \geq 0.85$) is taken from local areas lq_i and lt_j.
- **c.** The 3D coordinates of the atoms of lq_i are transformed by translation and rotation that best superpose its atom triad onto the atom triad of lt_j, by applying the Kabsch algorithm [17].
- **d.** The similarity between superposed local areas lq_i and lt_j, $S^{local} = \sum_{k \in lq_i} s1_{kl}/N$, are calculated. Here, k is an atom of lq_i, l is the atom of lt_j and is selected as a counterpart of k if it is located within 2.5 Å, and N is the number of constituent atoms in lq_i.
- **e.** If S^{local} is greater than or equal to 0.8, the pair lq_i and lt_j is defined as a "similar local area pair" and proceed to step (4).
- **f.** Another pair of similar atom triads is selected and repeat from step (c).

(4) Select another two local areas and repeat (3). Repeat this process until there are no more new pairs.

(5) Since the "similar local area pairs" in (3) and (4) are obtained as isolated fragments, they are subjected to the following clustering procedure.

When local areas lq_1 and lq_2 of protein q are selected from members of "similar local area pairs," they are clustered by imposing the following conditions.

- **a.** The distance between the central atoms of lq_1 and lq_2 should be less

than 10 Å.

b. When the (virtual) local areas lq_1 and lq_2 separately superposed onto each counterpart (designated as the "similar local area pair") on protein t by the method described in (3) are defined as lq'_1 and lq'_2, respectively, the distance between the central atoms of lq'_1 and lq'_2 should be less than 10 Å.

c. The difference between the interatomic distance of the central atoms in lq_1 and lq_2 and the interatomic distance of the central atoms in lq'_1 and lq'_2 should be less than 4 Å.

d. The difference between the angle formed by the normal vectors of lq_1 and lq_2 and the angle formed by the normal vectors of lq'_1 and lq'_2 should be less than 30°. Here, the direction of the normal vector is directed outward (toward the solvent region) from the central atom of the local area (which is defined as the origin).

e. When a merged region of lq_1 and lq_2 and a merged region of the counterparts of protein t are hypothetically superposed, the similarity S^{local} between these merged regions should be more than 0.7.

When lq_1 and lq_2 are clustered, the crossover of atoms between lq_1 and lq_2 is eliminated.

This process is repeated for all pairs in the local area of protein q that are the members of the "similar local area pairs," and they are clustered by single linkage clustering. In this manner, several surface regions, which are formed by the union of local areas of protein q, are obtained:$\{R^q_1, R^q_2, \cdots, R^q_N\}$.

Since each local area of protein q (if these areas are members of "similar local area pairs") has a counterpart local area in protein t, clusters of local areas of protein t are obtained by merging the counterpart local areas for each of $\{R^q_1, R^q_2, \cdots, R^q_N\}$. As a result, the surface regions of protein t are obtained:$\{R^t_1, R^t_2, \cdots, R^t_N\}$.

(6) Each surface area $R^q_k (k = 1, \cdots, N)$, which was obtained in (5), is superposed onto R^t_k by the same process as in (3), and the final similarity score $S_k = \sum_{(i,j)\in A^k} s1_{ij}/N^*_k$ for the pair of R^q_k and R^t_k is calculated. Here, i and j denote atoms of R^q_k and R^t_k, respectively; A^kdenotes aligned atom pairs between R^q_k and R^t_k (i and j are defined as an aligned atom pair when i and j are within a distance of 2.5 Å from each other and are of the same atom types according to PATTY); and N^*_k is the smaller of the numbers of constituent atoms between R^q_k and R^t_k. Here, the range of S values is between 0 and 1, with 1 indicating the highest similarity.

2.3 *Z*-score

When multiple pairs from similar regions are obtained, it may not be easy to determine which one of them has the highest similarity. The similarity score, S_k, as defined above, tends to become smaller as the size of a region increases. As the size of a region increases, the degree of freedom in geometry increases, reducing the proportion of matching atom pairs. In other words, even for the same value of S_k, we can predict that pairs with a larger number of constituent atoms (N^*) will tend to be more similar. In order to provide a standard by which we can systematically evaluate the degree of similarity, we adopt a statistical measure, namely, the Z-score.

First, we selected 12 types of proteins that have different folds, according to SCOP [18]. The surface regions of these proteins were randomly created, and their similarity (S) was calculated. Next, with these obtained results, we further calculated the mean $\overline{S}(N^*)$ and standard deviation $\sigma(N^*)$ of S, with respect to the different numbers of constituent atoms (N^*) of the regions. Finally, after calculating S for a region pair, we obtained the Z-score according to the following formula: $Z\text{-score} = \{S - \overline{S}(N^*)\}/\sigma(N^*)$.

3. Implementation

AltPS was programmed with C++ and tested on a Linux system (gcc version 4.2.3 installed). To run a calculation, the PDB file of the proteins to be searched for similar surface regions, is specified via the command line. The output files contain a list describing the similarities (S) and Z-scores of similar regions, the atomic alignment results of the similar regions, and the superposed PDB files (see **Fig. 1**). The threshold values for the number of constituent atoms (N^*) and the Z-score can be specified by the user in the command line options ($N^* = 30$ and Z-score = 4.0 by default).

(a) example output of the AltPS program

```
./altps -q 1IR3.pdb -t 1K3A.pdb
query  pdbfile = 1IR3.pdb
target pdbfile = 1K3A.pdb
output directory = ./
13 regions are detected. [Z-score >= 4, Size >= 30]
region list file: 1IR3_1K3A_1IR3A_0_1K3AA_0_region.list
```

(b) detected region list

1	1189	1208	922	0.74	65.2
2	346	348	298	0.84	34.1
3	588	598	470	0.77	43.3
4	52	49	36	0.71	7.25
5	33	44	22	0.62	4.3
6	35	37	24	0.67	5.17
7	37	36	29	0.76	6.35
8	37	38	24	0.6	4.48
9	35	35	22	0.58	4.04
10	34	39	23	0.61	4.27
11	36	30	21	0.63	4.11
12	43	44	27	0.57	4.67
13	38	37	26	0.65	5.08

Z-score
similarity score (S)
aligned atom number
atom number of region in target protein
atom number of region in query protein
region id

Fig. 1 Example of output from AltPS program: (**a**) console output of calculation for insulin receptor (PDB code 1ir3) and insulin-like growth factor 1 receptor (PDB code 1k3a) and (**b**) contents of detected region list file (1IR3_1K3A_1IR3A_0_1K3AA_0_region.list).

4. Results

Here, we demonstrated that AltPS can be robustly applied to diverse protein settings, which may come in the form of pairs with sequence homology, pairs with low sequence similarity but identical folds, or pairs with different folds. Calculations were performed on a computer equipped with a 2.2-GHz Athlon 64 single-core processor running on Linux.

4.1 ATP-binding Protein

Since ATP binds to many proteins with diverse sequences and folds, ATP-binding proteins are ideal candidates for investigating the structural similarity of protein surfaces. First, we calculated the surface similarity between the insulin receptor (PDB code 1ir3) and the insulin-like growth factor 1 (IGF1) receptor (PDB code 1k3a). These proteins share a high sequence identity (as high as 79.3%) and belong to the same protein family (protein kinases, catalytic subunit) according to the SCOP classification. The AltPS program calculated similar regions in response to the following command line: ./altps -q 1IR3.pdb -t 1K3A.pdb. As a result, 13 similar regions were detected in 8 min 41 s (cpu time). Figure 1 shows an output of this calculation. A result file of detected similar regions list (Fig. 1 (b)) was created as "1IR3_1K3A_1IR3A_0_1K3AA_0_region.list" in the output directory. The superposed PDB files of similar regions and input proteins (of query/target) for *region_id* 1 were "1IR3_1K3A_1IR3A_0_1K3AA_0_region_1_area_q/t.pdb" and "1IR3_1K3A_1IR3A_0_1K3AA_0_region_1_protein_q/t.pdb", respectively. In addition, a file containing the atomic alignment result for *region_id* 1 was created as "1IR3_1K3A_1IR3A_0_1K3AA_0_region_1_aligned_atoms.list".

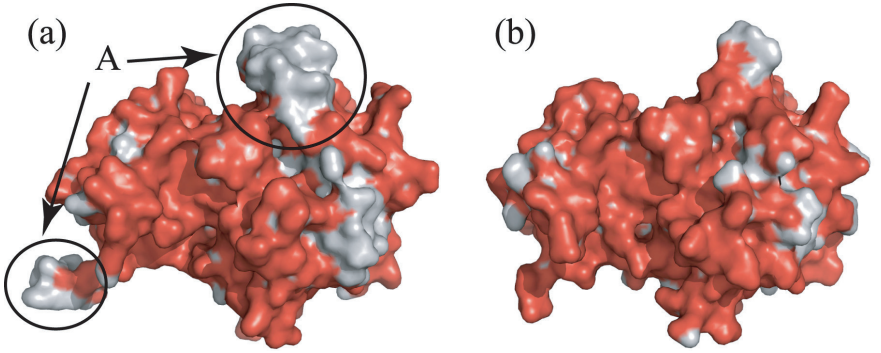

Fig. 2 Identified similar surface regions of (**a**) insulin receptor (PDB code 1ir3) and (**b**) insulin-like growth factor 1 receptor (PDB code 1k3a). Their sequence identity is 79.3%. The identified similar surfaces are highlighted in red (region ID 1 in Fig. 1 (b)). A denotes a surface region in which no similarity was detected in the calculation.

Upon calculation, a large region virtually covering the entire surface was identified (**Fig. 2**), which had a Z-score of 65.2. In the figure, the white surface of A represents the regions in which no similarity was detected in the calculation.

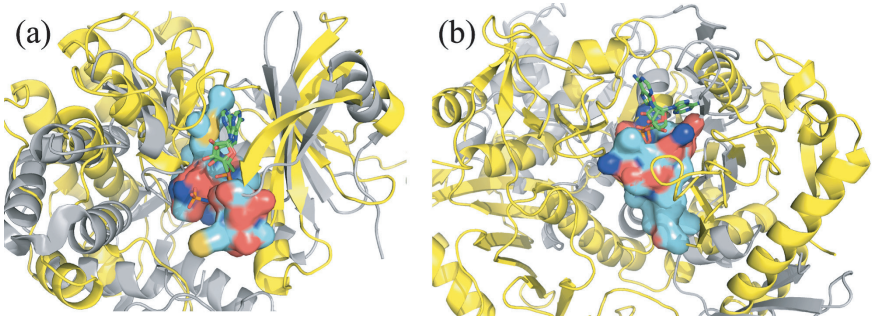

Fig. 3 Identification of ATP-binding site. (**a**) Similar surfaces identified for insulin receptor (PDB code 1ir3, yellow ribbon) and aminoglycoside 3'-phosphotransferase (PDB code 2b0q, gray ribbon). Ligands of the proteins (ANP and ADP) are represented in the stick form. (**b**) Similar surfaces identified for phosphoenolpyruvate carboxykinase (PDB code 1k3c, yellow ribbon) and ABC transporter (PDB code 1oxu c-chain, gray ribbon). Ligand of the proteins (ADP) is represented in the stick form. Similar surface regions identified by the proposed method are shown as Connolly surfaces.

As can be seen in the figure, it was visually confirmed that this region contains large structural differences in the actual surface structures. **Figure 3** shows the calculation results for cases in which there is no sequence homology. Figure 3 (a) shows the similar surface region that was identified for the insulin receptor (PDB code 1ir3) and aminoglycoside 3'-phosphotransferase (PDB code 2b0q). These proteins belong to different families according to the SCOP classification, but are both kinase-like proteins belonging to the same superfamily. The figure shows the identified surface region, which has a Z-score of 4.2. Figure 3 (b) shows the similar surface region that was identified for phosphoenolpyruvate carboxykinase (PDB code 1k3c) and the ABC transporter (PDB code 1oxu c-chain). Although these proteins contain different SCOP folds, they contain a similar region. AltPS detected 9 regions with a Z-score ≥ 3 and $N^* \geq 30$. The region shown in Fig. 3 (b) was the third-highest Z-score ($= 5.5$). The time required for the calculations of (a) and (b) was 7 min 38 s and 8 min 13 s, respectively. In both cases, the ATP-binding site (in particular, the phosphate-binding site) was identified as a similar region.

4.2 Proteins Having Different Folds

Both trypsin and subtilisin are serine proteases. Although these proteins have different folds, it is known that they have a common function (catalytic triad, Ser, His, Asp). We therefore used AltPS to compare the surface regions of trypsin (PDB code 1tpo) and subtilisin (PDB code 2prk). The required calculation time was 2 min 32 s. The surface region shown in **Fig. 4** was identified with a Z-score of 5.3. **Table 1** shows the results for the corresponding atomic alignments. An alignment between two amino acid residues in the catalytic triad (His57-His69 and Ser195-Ser224) was found. This result is consistent with the finding reported by Schmitt, et al. [10]. By using AltPS, we found a total of four regions with a Z-score ≥ 3 and $N^* \geq 30$, with the region shown in Fig. 4 receiving the highest Z-score ($= 5.3$).

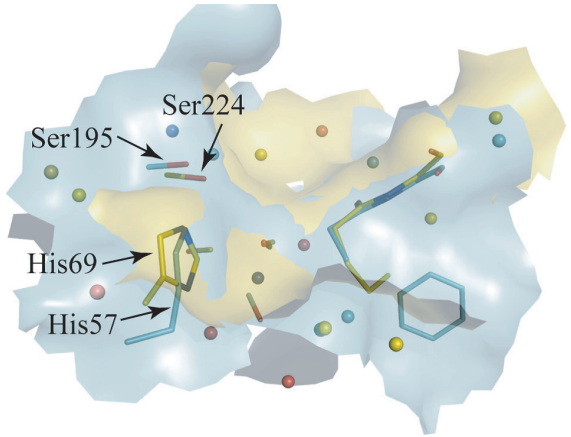

Fig. 4 Identification of catalytic site. A superposed image of similar regions identified for trypsin (PDB code 1tpo, cyan surface and stick) and subtilisin (PDB code 2prk, yellow surface and stick). Amino acid residues that constitute a catalytic triad are aligned (His57-His69 and Ser195-Ser224).

5. Discussion

AltPS can be used for performing comprehensive comparisons of entire protein surfaces and identifying novel similar regions. Since the search subject is not restricted to particular regions such as cavities or active sites, this tool can be conveniently employed for proteins even when information on their functional sites (e.g., ligand-binding sites) is not available. In order to prevent excessive cal-

Table 1 Output example of atomic alignment for similar surface regions of trypsin (1tpo) and subtilisin (2prk) in Fig. 4. Four similar regions were detected by using the proposed tool (AltPS). The alignment corresponds to one of these regions and has the highest Z-score (see Results).

1tpo					2prk				
Atom id	Atom name	Residue name	Residue number	Atom type	Atom id	Atom name	Residue name	Residue number	Atom type
184	SG	CYS	42	HYDROPHOBIC	1625	CE	MET	225	HYDROPHOBIC
286	CB	HIS	57	HYDROPHOBIC	519	CB	HIS	69	HYDROPHOBIC
287	CG	HIS	57	NONE	520	CG	HIS	69	NONE
289	CD2	HIS	57	HYDROPHOBIC	522	CD2	HIS	69	HYDROPHOBIC
290	CE1	HIS	57	HYDROPHOBIC	523	CE1	HIS	69	HYDROPHOBIC
291	NE2	HIS	57	POLAR(DONOR /ACCEPTOR)	524	NE2	HIS	69	POLAR(DONOR /ACCEPTOR)
606	CD1	LEU	99	HYDROPHOBIC	717	CD1	LEU	96	HYDROPHOBIC
1290	CB	SER	195	HYDROPHOBIC	1616	CB	SER	224	HYDROPHOBIC
1291	OG	SER	195	POLAR(DONOR /ACCEPTOR)	1617	OG	SER	224	POLAR(DONOR /ACCEPTOR)
1380	CG1	VAL	213	HYDROPHOBIC	1142	CB	ALA	158	HYDROPHOBIC
1384	C	SER	214	NONE	970	C	SER	132	NONE
1385	O	SER	214	ACCEPTOR	971	O	SER	132	ACCEPTOR
1386	CB	SER	214	HYDROPHOBIC	972	CB	SER	132	HYDROPHOBIC
1387	OG	SER	214	POLAR(DONOR /ACCEPTOR)	968	N	SER	132	DONOR
1389	CA	TRP	215	HYDROPHOBIC	975	CA	LEU	133	HYDROPHOBIC
1390	C	TRP	215	NONE	976	C	LEU	133	NONE
1391	O	TRP	215	ACCEPTOR	977	O	LEU	133	ACCEPTOR
1392	CB	TRP	215	HYDROPHOBIC	978	CB	LEU	133	HYDROPHOBIC
1395	CD2	TRP	215	HYDROPHOBIC	979	CG	LEU	133	HYDROPHOBIC
1398	CE3	TRP	215	HYDROPHOBIC	981	CD2	LEU	133	HYDROPHOBIC
1402	N	GLY	216	DONOR	982	N	GLY	134	DONOR
1403	CA	GLY	216	HYDROPHOBIC	983	CA	GLY	134	HYDROPHOBIC
1404	C	GLY	216	NONE	984	C	GLY	134	NONE
1405	O	GLY	216	ACCEPTOR	985	O	GLY	134	ACCEPTOR
1407	CA	SER	217	HYDROPHOBIC	987	CA	GLY	135	HYDROPHOBIC

culation load, it seems reasonable to devise strategies that focus first on the predicted functional regions, such as cavities. Unfortunately, however, this approach is likely to fail to identity important similar regions at the stage of functional site prediction. Even when the prediction of functional sites proceeds successfully, a subsequent analysis of similarity will depend heavily on how the regions are initially defined. For example, suppose that similarity exists in just a part of a predicted functional site. When the entire predicted region is subjected to a query for comparisons, the evaluation scores for similarity may consequently be grossly underestimated, resulting in the nonrecognition of partial similarity in the calculations.

The local area size (diameter: 10 Å) used in AltPS was set by considering the minimum size (around 10 Å) of the similar cavities detected in our previous study [15]. Moreover, other parameters, which are shown in the Algorithm (e.g., 5 Å, 2.5 Å, 10 Å, s3$\geq$0.8, etc.), were estimated on the basis of this local area size and empirically optimized through a number of test calculations. Since these parameters influence each other, the user cannot change them by means of a

parameter option. Besides, one major feature of AltPS is the computation of Z-scores. This statistical measure is introduced to compensate for the problem that can arise because S values tend to become smaller as the size of the identified similar regions increases. It can thus be used for the efficient comparison of similar regions. A high Z-score suggests a high probability that the identified regions are important similar regions. In practice, as illustrated by the examples in the Results section, the number of regions having high Z-scores ($\geq$4) is relatively small (approximately ten). Furthermore, functional sites (e.g., ligand-binding sites) were found in the regions with high Z-scores.

In several of the existing methods [13),14)], it is possible to compare local surface regions across the entire protein surface, similar to AltPS. However, these methods do not provide software that can be conveniently used on a personal computer. In contrast, AltPS runs as a stand-alone application on a local computer. As a result, it is can select or develop a computational environment depending on the user's needs. In AltPS, the geometrical and physicochemical properties of local surface regions are described by a vector, which can be an efficient means to narrowing down the number of regions during the initial comparison, thereby reducing the calculation load. In addition, AltPS uses only coordinate data of the surface atoms to describe protein surface structures, which also helps to economize calculations. In this manner, AltPS performs calculations by using a smaller set of representative points and feature vectors, as compared to triangulated mesh and spin images [19)], which are the commonly used methods for surface description. Although AltPS cannot accurately represent concavo-convex shapes in the local surface in this manner, it is most effective for obtaining the alignment of surface atoms.

This current version of AltPS was specifically designed for pair-wise calculations. Future development will be focused on improvements that allow more efficient searches of all protein structures in the PDB.

Acknowledgments The authors wish to thank Professor Akinori Kidera for his valuable comments and suggestions.

References

1) Holm, L. and Sander, C.: Dali: A network tool for protein structure comparison, *Trends Biochem. Sci.*, Vol.20, pp.478–480 (1995).
2) Orengo, C. and Taylor, W.: SSAP: sequential structure alignment program for protein structure comparison, *Methods Enzymol.*, Vol.266, pp.617–635 (1996).
3) Gibrat, J., Madej, T. and Bryant, S.: Surprising similarities in structure comparison, *Curr. Opin. Struct. Biol.*, Vol.6, pp.377–385 (1996).
4) Shindyalov, I. and Bourne, P.: Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, *Protein Eng.*, Vol.11, No.9, pp.739–747 (1998).
5) Weber, A., Casini, A., Heine, A., Kuhn, D., Supuran, C., Scozzafava, A. and Klebe, G.: Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition, *J. Med. Chem.*, Vol.47, pp.550–557 (2004).
6) Kinoshita, K., Sadanami, K., Kidera, A. and Go, N.: Structural motif of phosphate-binding site common to various protein superfamilies: all-against-all structural comparison of protein-mononucleotide complexes, *Protein Eng.*, Vol.12, No.1, pp.11–14 (1999).
7) Brakoulias, A. and Jackson, R.: Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: an automated all-against-all structural comparison using geometric matching, *Proteins*, Vol.56, No.2, pp.250–260 (2004).
8) Sael, L., La, D., Li, B., Rustamov, R. and Kihara, D.: Rapid comparison of properties on protein surface, *Proteins*, Vol.73, No.1, pp.1–10 (2008).
9) Kinoshita, K., Murakami, Y. and Nakamura, H.: eF-seek: prediction of the functional sites of proteins by searching for similar electrostatic potential and molecular surface shape, *Nucleic Acids Res.*, Vol.35, pp.W398–W402 (2007).
10) Schmitt, S., Kuhn, D. and Klebe, G.: A new method to detect related function among proteins independent of sequence and fold homology, *J. Mol. Biol.*, Vol.323, No.2, pp.387–406 (2002).
11) Binkowski, T. and Joachimiak, A.: Protein functional surfaces: global shape matching and local spatial alignments of ligand binding sites, *BMC Struct. Biol.*, Vol.8, p.45 (2008).
12) Najmanovich, R., Kurbatova, N. and Thornton, J.: Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites, *Bioinformatics*, Vol.24, pp.i105–i111 (2008).
13) Angaran, S., Bock, M., Garutti, C. and Guerra, C.: MolLoc: A web tool for the local structural alignment of molecular surfaces, *Nucleic Acids Res.*, Vol.37, pp.W565–W570 (2009).
14) Shrestha, N.L., Kawaguchi, Y. and Ohkawa, T.: SUMOMO: A Protein Surface

Motif Mining Module, *International Journal of Computational Intelligence and Applications*, Vol.4, No.4, pp.431–450 (2004).
15) Minai, R., Matsuo, Y., Onuki, H. and Hirota, H.: Method for comparing the structures of protein ligand-binding sites and application for predicting protein-drug interactions, *Proteins*, Vol.72, No.1, pp.367–381 (2008).
16) Bruce, L.B. and Sheridan, R.P.: PATTY: A programmable atom typer and language for automatic classification of atoms in molecular databases, *J. Chem. Inf. Comput. Sci.*, Vol.33, pp.756–762 (1993).
17) Kabsch, W.: A solution for the best rotation to relate two sets of vectors, *Acta Crystallographica Section A*, Vol.32, No.5, pp.922–923 (1976).
18) Murzin, A., Brenner, S., Hubbard, T. and Chothia, C.: SCOP: A structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol.*, Vol.247, No.4, pp.536–540 (1995).
19) Andrew, J.: Spin-Images: A Representation for 3-D Surface Matching, *Doctoral dissertation, Robotics Institute, Carnegie Mellon University* (1997).

(Received October 15, 2009)
(Accepted November 8, 2009)
(Released February 4, 2010)

(Communicated by *Tatsuya Akutsu*)

Ryoichi Minai received his M.S. degree from University of Tsukuba in 1998. From 2002 to 2008, he was a student in a doctoral program at Department of Supramolecular Biology, Yokohama City University. From 2008 to 2009, He worked as an assistant professor at University of Yamanashi. His current affiliation is SINKEN Yamanashi Office.

Yo Matsuo received his Ph.D. in Biophysics from Kyoto University in 1996. He worked at Fujitsu Laboratories, Ltd. (1990–1996), U.S. NIH/NCBI (1997–1999), PharmaDesign, Inc. (1999–2000), RIKEN (2000–2007), and Yokohama City University (2001–2007). His current affiliation is OncoTherapy Science, Inc.