Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Parameter selection of pocket extraction algorithm using interaction interface*

KIM Chong-Min[1], WON Chung-In[1], RYU Joonghyun[2], CHO Cheol-Hyung[3], BHAK Jonghwa[4], KIM Deok-Soo†[1]

([1]*Department of Industrial Engineering, Hanyang University, Seoul 133-791, Korea*)

([2]*Voronoi Diagram Research Center, Hanyang University, Seoul 133-791, Korea*)

([3]*Samsung Electronics, Suwon, Gyeonggi-do 443-742, Korea*)

([4]*Korean Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-806, Korea*)

†E-mail: dskim@hanyang.ac.kr

Received May 20, 2006; revision accepted June 21, 2006

Abstract: Pockets in proteins have been known to be very important for the life process. There have been several studies in the past to automatically extract the pockets from the structure information of known proteins. However, it is difficult to find a study comparing the precision of the extracted pockets from known pockets on the protein. In this paper, we propose an algorithm for extracting pockets from structure data of proteins and analyze the quality of the algorithm by comparing the extracted pockets with some known pockets. These results in this paper can be used to set the parameter values of the pocket extraction algorithm for getting better results.

Key words: Pocket, Protein, Interaction interface, Protein interaction, Voronoi diagram
doi:10.1631/jzus.2006.A1492 **Document code:** A **CLC number:** TP393

INTRODUCTION

Being the bio era, a significant research effort has been devoted to the study of proteins since they are related to the critical body functions. There have been many studies on proteins, with the voluminous related data being currently available in various public repositories including the Protein Data Bank (PDB) (CATH, 2006; DALI, 2006; MSD, 2006; RCSB PDB, 2006; SCOP Database, 2006).

Given the atomic complexes of proteins, analyzing interactions between them is important for understanding their biological functions. The interaction between a protein and a small molecule is also one of the most important issues in designing new drugs.

The study of molecular interactions such as the docking of a protein with a ligand can be approached from a physicochemical and/or a geometrical point of view (Parsons and Canny, 1994). While the physicochemical approach is to find regions on the surface of a protein which minimize the potential energy between two molecules, the geometric approach is to determine whether two molecules have geometrically meaningful features for the interaction. Hence, both approaches may be used together for a better result.

A docking between a protein, called a receptor, and a small molecule, called a ligand, usually occur around depressed regions, called docking sites or pockets, on the surface of a receptor. Since designing a new drug requires finding a small chemical which can dock or bind at appropriate pockets on a protein, the recognition of pockets on proteins is one of the most fundamental processes in drug design.

Considering that chemical databases usually contain millions of chemical data entries, manually identifying pockets on the surface of a protein is

* Project supported by Creative Research Initiative from the Ministry of Science and Technology (MOST), Korea. BHAK Jonghwa is supported by Biogreen21 Fund and MOST Funds, Korea

time-consuming and error-prone. Therefore, the automatic recognition of pockets and the evaluation of the binding of a chemical to a pocket are rather important in the study of protein-ligand docking for the development of new drugs (Kuntz, 1992).

Contrary to the physicochemical approach, efforts to understand the geometry perspective of biological systems have started relatively recently (Agarwal *et al*., 2004; Edelsbrunner *et al*., 1998; Heifetz and Eisenstein, 2003; Lee and Richard, 1971; Liang *et al*., 1998; Peters *et al*., 1996; Shoichet and Kunts, 1991). Present consensus is that the geometry is as important and critical for biological systems in various important aspects as the physicochemical aspect of a molecule. Hence, research on the geometry in biological systems will provide new challenges as well as opportunities for the community of geometers.

In this paper, we will present an algorithm for extracting pockets from proteins structure data and compare the computed pockets with the real pockets from some known cases. The proposed algorithm is currently purely geometric and therefore some errors exist. Currently, the 3D coordinates of atoms for more than 30000 proteins are known and available through PDB (RCSB PDB, 2006).

Given a protein, the algorithm proposed in this paper first computes the Voronoi diagram of van der Waals atoms. Then, a β-shape is computed from the Voronoi diagram using a spherical probe (Kim *et al*., 2005a; 2006). The Voronoi diagram of atoms presented in this paper is similar to the ordinary Voronoi diagram for points in the sense that the Euclidean distance metric is used. However, it differs from the ordinary Voronoi diagram of points since the distance is measured from the surface of atoms, not from the centers of atoms.

PROTEIN STRUCTURE

A protein is a macromolecule consisting of a linear sequence of up to 20 different amino acids, with the distinct amino acids sequence determining the unique 3D structure of a protein. It is believed that the function of a protein is mostly determined by its unique 3D structure.

An amino acid consists of dozens of atoms. For example, a glycine is the smallest among the twenty amino acids consisting of 10 atoms (2 C's, 5 H's, 2 O's and a single N). On the other hand, *tryptophan*, the largest amino acid, consists of 27 atoms (11 C's, 12 H's, 2 O's and 2 N's). An amino acid has a carbon called an α-carbon C_α at its topological center. Around the C_α, hydrogen (H), an amino-group (NH_2), a carboxyl-group (COOH) and a side-chain (R) are attached. Note that a side-chain is also called an R-group where R is short for radical. Most of the 20 amino acids have identical global topology structure except for the side-chain. Therefore, a side chain determines a unique 3D structure and the functions of a protein. Consecutive amino acids are connected to each other via a peptide bond between their respective carboxyl- and amino-groups (Berg *et al*., 2002).

The structure of a protein is usually viewed from four different hierarchical levels. The linear sequence of amino acids is called the primary structure of a protein. Due to the interactions between the atoms in the primary structure, some amino acids form α-helices or β-sheets in 3D space with these descriptors being denoted as the secondary structure of a protein. When the secondary structure is appropriately folded in 3D, it is called the tertiary structure of a protein. The linear sequence of amino acids connected via peptide bonds forms a chain. Many proteins exist in nature as a set of more than one chain called the quaternary structure of a protein (Bourne and Weissig, 2003). A protein consisting of a single chain is called a monomer. When a protein consists of two, three, or four chains, it is called a dimer, a trimer, and a tetramer, respectively (Fig.1, see page 1496).

INTERACTION INTERFACE

It is known that the interactions among proteins affect various functions in a living cell. Therefore, biologists have been paying attention to the analyses of the interaction between proteins (Jones and Thornton, 1996; Tsai *et al*., 1997; Nooren and Thornton, 2003; Ofran and Rost, 2003; Sheinerman and Honig, 2002; Xu *et al*., 1997).

One of the most fundamental goals of protein analysis is to understand how proteins carry out various essential life processes and one of the main efforts of the analysis is to identify the protein structure of proteins and the interactions among proteins.

However, the properties and functions of many proteins have yet to be studied/determined, with it being generally agreed that inference of the function of a protein can be also helped by investigating its interactions with other proteins. The inference of protein function is based on the premise that the function of a protein can be discovered in part via its interactions with other proteins with known functions. Hence, the study of interactions can be vital to the understanding of proteins (Jones and Thornton, 1996).

An interaction defined between two or (more) proteins is called an inter-protein interaction, and an interaction between chains in a protein is called an intra-protein interaction. From a geometric point of view, the interaction interfaces of both types of interactions are identical.

Untrimmed interaction interface: IIF_∞

Suppose that we are given a proper tessellation of a whole space where an atom set A is defined. Examples of such tessellation may be a Voronoi diagram of atom centers, a power diagram of the atoms, or a Voronoi diagram of the atoms. In these tessellations, each face in the tessellation is defined by two nearby atoms in the set using a distance definition approximation for the tessellation.

Suppose that the set A consists of two or more groups where the unit of the group may be either a chain or a protein. Then, we collect a set F of faces of the tessellation where each face $f \in F$ is defined by two atoms from different groups in the set. Then, the set F is called an interaction interface among the groups in the set A. Note that this definition of interaction interface applies to any of the previously described schemes of the set A's space tessellation. For example, the subset of Voronoi faces of the ordinary Voronoi diagram of atom centers may be used as an interaction interface. In some previous studies of the interaction interface, the power diagram's faces were used to compute the interaction interface (Ban *et al*., 2004; Varshney *et al*., 1995).

In this paper, we define instead the interaction interface using the Voronoi diagram of atoms to make it more precise by fully reflecting the size differences among atoms in the set. Let $A=\{a_1, a_2, \ldots, a_i\}$, and $B=\{b_1, b_2, \ldots, b_j\}$ be two chains in a protein, where a_i and b_j are atoms with appropriate centers and radii. The interaction interface $IIF_\infty(A,B)$ between chains A and B is defined as

$$IIF_\infty(A,B)=\{p|dist(p, A)=dist(p, B)\}, \quad (1)$$

where $dist(p, A)$ denotes the minimum Euclidean distance from p to the surfaces of all van der Waals atoms in set A. Then, it can be easily shown that $IIF_\infty(A,B)$ is the subset of Voronoi faces in the Voronoi diagram $VD(A\cup B)$ of all atoms of $A\cup B$. Note that each face $f\in IIF_\infty(A,B)$ is defined by two atoms from two different chains. Hence, the following definition of $IIF_\infty(A,B)$ is equivalent to the definition in Eq.(1).

$$IIF_\infty(A,B)=\{f\in F^V|dist(p, A)=dist(p, B), \forall p\in f\}, \quad (2)$$

where F^V is the set of Voronoi faces in $VD(A\cup B)$. Therefore, $IIF_\infty(A,B)$ can be computed by collecting the appropriate faces from all the Voronoi faces in the Voronoi diagram $VD(A\cup B)$ by simply checking where each Voronoi face has two defining atoms from different groups. Hence, it takes $O(m)$ time in the worst case, where m is the number of Voronoi faces in the Voronoi diagram. Note that $IIF_\infty(A,B)$ extends to infinity. Hereafter, we will omit (A,B) whenever possible for notational convenience.

Fig.2 (see page 1496) shows a dimer 1R95 and the corresponding interaction interface IIF_∞. The interaction interface IIF_∞ computed from the Voronoi diagram of atoms extends to infinity as shown in Fig.2b and therefore is called an untrimmed interaction interface. Note that a protein 1R95 is found in an important bacterium called *Escherichia coli*, and is a dimer consisting of 2 groups, 194 amino acids, and 1468 atoms (except hydrogens) (Bilder *et al*., 2004).

Trimmed interaction interface: *IIF*

While an interaction interface is important for understanding protein function, the points on IIF_∞ which are far away from the protein, are biologically less significant (Creighton, 1999). Hence, we define a biologically more meaningful interaction interface which is relatively near the atoms of both chains.

Proteins usually exist in a solvent (which is mostly water) and a solvent molecule interacts with a protein in a rather complicated fashion if we consider its complete 3D structure. The water molecule, for example, is polarized due to its bonding among atoms

and the bend in its 3D structure and the orientation of a water molecule around the surface of a protein is rather stochastic. Hence, exact analysis of the interaction between the water molecule and a protein is infeasible. To simplify the model and the calculation, the usual practice is to approximate the solvent molecule with a small sphere enclosing the molecule. Note that the water molecule, H_2O, has an internal angle of 104.5° at ∠HOH and the common probe for the water molecule is a sphere with a radius 1.4×10^{-10} m (Berg *et al.*, 2002).

Suppose that a probe has a radius ρ. Then, the trimmed interaction interface $IIF(A,B)$ is defined as follows:

$$IIF(A,B)=\{p|dist(p, A)=dist(p, B)\leq\rho\}. \quad (3)$$

Hence, *IIF* can be computed by trimming off the less meaningful part of the surface from IIF_∞. This trimming is done using a spherical probe as a virtual trimmer, and the edges created by the trimming are called trimming edges. There are three types of trimming primitives in the computation of *IIF* as follows: circumscribing, voiding, and depilating.

REAL POCKET VS EXTRACTED POCKET

While the interaction interface is one of the most important constructs for a molecular structure, pockets are also important as well in the sense of understanding the functions of proteins and designing drugs (Kim *et al*., 2006). In this section, we analyze the quality of the pocket extraction algorithm to verify the validity of an extracted pocket of a protein via the presented algorithm by comparing the extracted pockets with some known pockets.

In PDB, there are several proteins which already contain ligands binding with the proteins. For example, the file of PDB code 1FKG shown in Fig.3 contains 1055 atoms. It turns out that 1FKG contains a protein named FK506 (blue) with a ligand 3-diphenyl-1-propyl-1-(3,3-dimethyl-1,2-dioxypentyl)-2-piperidine carboxylate (the HET ID of this ligand is SB3) (red). The protein contains 1022 atoms which constitute 107 residues, and the ligand contains 33 atoms. (Usually hydrogen atoms are not included in the data files in PDB since the constructed model via X-ray crystallography usually contains some errors and the hydrogen is relatively sensitive to this error. If it is necessary, the hydrogen atoms are usually generated via a computation. However, 1FKG contains some hydrogen atoms. If SB3 contains hydrogen atoms, their count is 67.)

Fig.3a shows the atoms of 1FKG and the interaction interface between the receptor and the ligand, and Fig.3b shows the interaction interface between the receptor and the ligand. Fig.3c shows the interaction interface with the atoms directly involved in the definition of the interaction interface. Fig.3d and Fig.3e respectively show these atoms for both receptor and ligand. Fig.3f then shows the interaction interface itself.

Therefore, the atoms in Fig.3d form a pocket which actually binds with the ligand.

We then call this pocket a real pocket *R*. Suppose that we remove the ligand from the original data and apply a pocket extraction algorithm on the protein data with the ligand removed. Then, we get an extracted pocket *X* from the protein without any a prior knowledge about the pocket on the protein.

For example, Fig.4a shows the receptor, FK506 binding protein, of the PDB data 1FKG after the ligand data is removed from 1FKG. Fig.4b shows all pockets computed using the Voronoi diagram of atoms of the data in Fig.4a and its dual structure β-shapes. The parameters of the outer and inner β-shapes are 50×10^{-10} m and 1.4×10^{-10} m, respectively. Note that the radius of a probe for a water-molecule is usually considered as 1.4×10^{-10} m. For the details of pocket extraction algorithm, please refer to (Kim *et al*., 2006). Then, the red region on the surface of the protein in Fig.4c is the largest pocket of the protein. It turns out that this extracted pocket *X* corresponds to the actual *R* pocket where the ligand SB3 binds.

EXPERIMENTS AND DISCUSSIONS

Since *R*-pocket is the real one in a protein and we compute *X*-pocket using an algorithm, there has to be some error associated with the extraction algorithm. Depending on the parameter of the algorithm, the success rate of the algorithm will be affected.

We performed some experiments by implement-

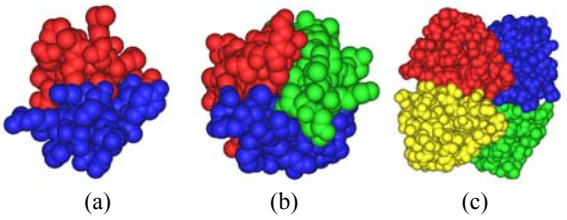

Fig.1 Examples of protein complexes. (a) A dimer (1A0M); (b) A trimer (1CE0); (c) A tetramer (1CSK). Shown in the parentheses are the PDB codes of the proteins

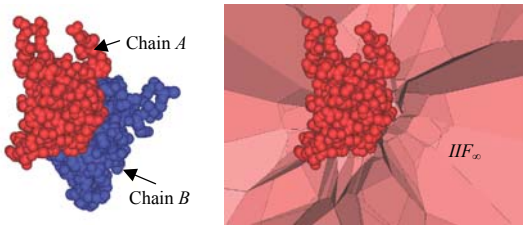

Fig.2 A dimer (1R95) and the corresponding IIF_∞. (a) Red spheres represent chain *A* and blue spheres represent chain *B*; (b) Pink surface represents IIF_∞

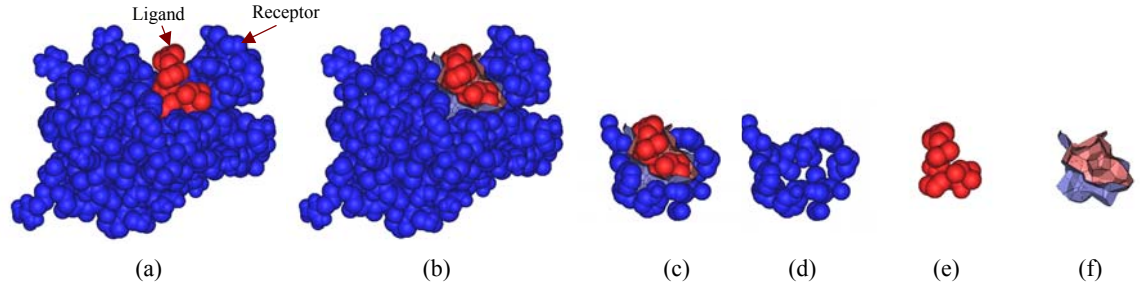

Fig.3 A model in 1FKG. (a) The configuration of a receptor (blue) and a ligand (red); (b) The receptor and the ligand with embedded interaction interface (*IIF*); (c) *IIF* with the interacting atom set; (d) The real pocket *R* which is the interacting atom set for the receptor; (e) Interacting atom set for the ligand; (f) *IIF* without atoms

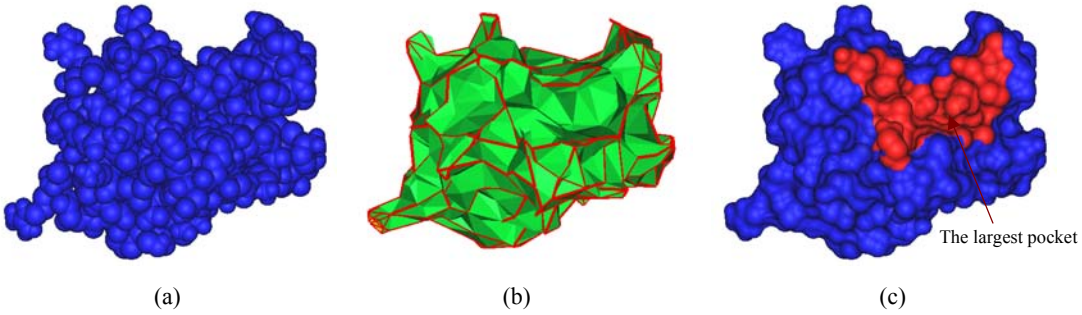

Fig.4 The receptor of a protein 1FKG and the corresponding pockets. (a) The receptor; (b) All pockets of the receptor; (c) The largest pocket on the molecular surface (red)

ing algorithms to construct the Voronoi diagram of atoms in the protein, to compute the interaction interface IIF_∞, the trimming of IIF_∞ to get a valid interaction interface *IIF*, and the β-shape. Then, we extracted *X*-pockets for various values of β, and then compared *X*-pockets with *R*-pocket.

Table 1 and Table 2 show the experimental results for 1FKG with various values of β for both outer β-shape and inner β-shape. In Table 1, the probe radius for inner β-shape I-Probe is fixed as 1.4×10^{-10} m which is the radius of a probe for a water-molecule. The size of the outer β-shape, however, varies. In fact, we reduced the outer probe size from the radius of 50000×10^{-10} m (which is considered to be sufficiently large) by halving it for each following experiment. The main observations were as follows:

(1) The number of extracted pockets increases as the outer probe size decreases.

(2) The *X*-pocket corresponding to the *R*-pocket does not change with respect to the size of the outer probe.

(3) Even though it is not clearly shown in Table 1, we found that the *X*-pocket always corresponds to the largest (in terms of the number of atoms in the pocket) pocket extracted from the given parameter.

Table 2 shows other interesting results in these

Table 1 Experimental results for a protein 1FKG (I-Probe radius is fixed at 1.4×10^{-10} m)

Size of O-Probe ($\times10^{-10}$ m)	Size of I-Probe ($\times10^{-10}$ m)	(1) Number of atoms in *R*-Poc.	Number of pockets	(2) Number of atoms in *X*-Poc.	(3) Number of matched atoms	Ratio (1)/(2)	Ratio (3)/(2)
50000	1.4	73	116	85	50	0.859	0.588
25000	1.4	73	116	85	50	0.859	0.588
12500	1.4	73	116	85	50	0.859	0.588
6250	1.4	73	116	85	50	0.859	0.588
3125	1.4	73	116	85	50	0.859	0.588
1563	1.4	73	118	85	50	0.859	0.588
781	1.4	73	118	85	50	0.859	0.588
391	1.4	73	122	85	50	0.859	0.588
195	1.4	73	132	85	50	0.859	0.588
98	1.4	73	140	85	50	0.859	0.588
49	1.4	73	170	85	50	0.859	0.588

Table 2 Experimental results for a protein 1FKG (O-Probe radius is fixed at 50×10^{-10} m)

Size of O-Probe ($\times10^{-10}$ m)	Size of I-Probe ($\times10^{-10}$ m)	(1) Number of atoms in *R*-Poc.	Number of pockets	(2) Number of atoms in *X*-Poc.	(3) Number of matched atoms	Ratio (1)/(2)	Ratio (3)/(2)
50	10.0	73	170	9	3	8.111	0.333
50	9.0	73	170	7	3	10.429	0.429
50	7.0	73	170	12	4	6.083	0.333
50	5.0	73	170	16	6	4.563	0.375
50	3.0	73	170	43	18	1.698	0.419
50	1.5	73	170	83	49	0.880	0.590
50	1.4	73	170	85	50	0.859	0.588
50	1.3	73	170	94	58	0.777	0.617
50	1.2	73	170	104	65	0.702	0.625
50	1.1	73	170	107	66	0.682	0.617
50	1.0	73	170	113	66	0.646	0.584

experiments. We fixed the radius of the outer probe as 50×10^{-10} m, then changed the inner probe size according to the following observations. To set the starting value of the size for the inner probe, we computed the minimum spheres enclosing the ligand and found its size was 7.61×10^{-10} m. So we changed the inner probe radius from 10.0×10^{-10} m down to 1.0×10^{-10} m. By this experiment, we found that the best size of the inner probe is between 1.0×10^{-10} m and 2.0×10^{-10} m. We then examined this interval more closely to find that the best size of the inner probe is around 1.2×10^{-10} m which is almost identical to the size of the hydrogen molecule. It is very interesting that the maximum match occurs when the inner probe radius is approximately the size of the hydrogen molecule. Note that the van der Waals radius of hydrogen is usually considered to be 1.2×10^{-10} m.

CONCLUSION

Voronoi diagram is known as an important mathematical and computational tool for many disciplines. While Voronoi diagram has been generalized in various directions, the computation of the Voronoi diagram for spheres has not been quite satisfactory in many aspects. Recently, the algorithms and their implementations have been known for successful computation of the Voronoi diagram of spheres (Kim, 2004; Kim *et al*., 2004a; 2004b; 2004c; 2005a; 2005b; 2005c). Since then, the Voronoi diagram of spheres has been applied to solve various problems in structural molecular biology (Kim, 2004; 2005; Kim *et al*., 2004b; 2004c; 2005a; 2005c; 2006). Since the investigation of molecular structure is interdisciplinary in nature, collaborations between

geometers and biologists are essential (Kim *et al*., 2004c; 2005c).

In this paper, we showed that the Voronoi diagram of spheres can be used for analysis of the interactions between chains in a protein, the interactions between proteins, and the interactions between any of the sets of atomic structures. Based on the interaction interface, we presented an approach to measure the quality of extracted pockets on proteins. We believe that this approach will provide important information for enhancing the quality of pocket extraction algorithms.

Like other applications in structural biology, this problem domain opens a new, interesting and important area of application for geometers in the CAD and CAGD communities. We expect more significant studies by them in this area to follow and hope that the Voronoi diagram of spheres will be one of the most fundamental computational tools used in this research.

References

Agarwal, P.K., Edelsbrunner, H., Harer, J., Wang, Y., 2004. Extreme Elevation on a 2-manifold. Proc. of the 20th Annual ACM Symposium on Computational Geometry, Brooklyn, New York, USA, p.357-365.

Ban, Y.E.A., Edelsbrunner, H., Rudolph, J., 2004. Interface Surfaces for Protein-protein Complexes. Proc. of the 8th Annual International Conference on Research in Computational Molecular Biology, San Diego, CA, p.205-212.

Berg, J.M., Tymoczko, J.L., Stryer, L., 2002. Biochemistry (5th Ed.). W.H. Freeman.

Bilder, P.W., Ding, H., Newcomer, M.E., 2004. Crystal structure of the ancient, Fe-S scafold IscA reveals a novel protein fold. *Biochemistry*, **43**(1):133-139.

Bourne, P.E., Weissig, H., 2003. Structural Bioinformatics. Wiley-Liss.

CATH Database, 2006. Http://www.biochem.ucl.ac.uk/bsm/cath/.

Creighton, T.E., 1999. Protein Folding. W.H. Freeman.

DALI Domain Dictionary, 2006. Http://www.ebi.ac.uk/dali/domain/.

Edelsbrunner, H., Facello, M., Liang, J., 1998. On the definition and the construction of pockets in macromolecules. *Discrete Applied Mathematics*, **88**(1-3):83-102. [doi:10.1016/S0166-218X(98)00067-5]

Heifetz, A., Eisenstein, M., 2003. Effect of local shape modifications of molecular surfaces on rigid-body protein-protein docking. *Protein Engineering*, **16**(3):179-185. [doi:10.1093/proeng/gzg021]

Jones, S., Thornton, J.M., 1996. Principles of Protein-protein Interactions. Proc. of the National Academy of Sciences of the United States of America, **93**(1):13-20. [doi:10.1073/pnas.93.1.13]

Kim, C.M., 2005. Interaction Interfaces in Proteins Using Voronoi Diagram of Atoms and Their Database. Master's Thesis, Hanyang University, Korea.

Kim, D.S., 2004. Euclidean Voronoi Diagram of Atoms and Protein. Proc. of the 1st BioGeometry Meeting—In Conjunction with the ACM Symposium on Computational Geometry. The Technical University of New York, New York, USA.

Kim, D.S., Cho, Y., Kim, D., 2004a. Edge-tracing Algorithm for Euclidean Voronoi Diagram of 3D Spheres. Proc. of the 16th Canadian Conference on Computational Geometry, p.176-179.

Kim, D.S., Cho, Y., Kim, D., Cho, C.H., 2004b. Protein Structure Analysis Using Euclidean Voronoi Diagram of Atoms. Proc. of the International Workshop on Biometric Technologies (BT2004), p.125-129.

Kim, D.S., Cho, Y., Kim, D., Kim, S., Bhak, J., Lee, S.H., 2004c. Euclidean Voronoi Diagrams of 3D Spheres and Applications to Protein Structure Analysis. Proc. of the 1st International Symposium on Voronoi Diagrams in Science and Engineering (VD2004), p.137-144.

Kim, D.S., Cho, C.H., Ryu, J., Kim, D., 2005a. Three dimensional beta shapes (Submitted to *Computer-Aided Design*).

Kim, D.S., Cho, Y., Kim, D., 2005b. Euclidean Voronoi diagram of 3D balls and its computation via tracing edges. *Computer-Aided Design*, **37**(13):1412-1424. [doi:10.1016/j.cad.2005.02.013]

Kim, D.S., Cho, Y., Kim, D., Kim, S., Bhak, J., Lee, S.H., 2005c. Euclidean Voronoi diagrams of 3D spheres and applications to protein structure analysis. *Japan Journal of Industrial and Applied Mathematics*, **22**(2):251-265.

Kim, D.S., Cho, C.H., Kim, D., Cho, Y., 2006. Recognition of docking sites on a protein using β-shape based on Voronoi diagram of atoms. *Computer-Aided Design*, **38**(5):431-443. [doi:10.1016/j.cad.2005.11.008]

Kuntz, I.D., 1992. Structure-based strategies for drug design and discovery. *Science*, **257**(21):1078-1082.

Lee, B., Richard, F.M., 1971. The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology*, **55**(3):379-400. [doi:10.1016/0022-2836(71)90324-X]

Liang, J., Edelsbrunner, H., Woodward, C., 1998. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Science*, **7**:1884-1897.

MSD (Macromolecular Structure Database), 2005. Http://www.ebi.ac.uk/msd/.

Nooren, I.M.A., Thornton, J.M., 2003. Diversity of protein-protein interactions. *The EMBO Journal*, **22**(14):3486-3492. [doi:10.1093/emboj/cdg359]

Ofran, Y., Rost, B., 2003. Analysing six types of protein-protein interfaces. *Journal of Molecular Biology*, **325**(2):377-387. [doi:10.1016/S0022-2836(02)01223-8]

Parsons, D., Canny, J., 1994. Geometric Problems in Molecular Biology. Proc. of the 2nd International Confer-

ence on Intelligent Systems for Molecular Biology. Stanford University, California, USA, p.322-330.

Peters, K.P., Fauck, J., Frommel, C., 1996. The automatic search for ligand binding sites in protein of known three dimensional structure using only geometric criteria. *Journal of Molecular Biology*, **256**(1):201-213. [doi:10.1006/jmbi.1996.0077]

RCSB PDB, 2006. Protein Data Bank Homepage. Http://www.rcsb.org/pdb/.

SCOP Database, 2006. Http://scop.berkeley.edu/.

Sheinerman, F.B., Honig, B., 2002. On the role of electrostatic interactions in the design of protein-protein interfaces. *Journal of Molecular Biology*, **318**(1):161-177. [doi:10.1016/S0022-2836(02)00030-X]

Shoichet, B., Kunts, I., 1991. Protein docking and complementarity. *Journal of Molecular Biology*, **221**(1):327-346. [doi:10.1016/0022-2836(91)80222-G]

Tsai, C.J., Lin, S.L., Wolfon, H.J., Nussinov, R., 1997. Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Science*, **6**:53-54.

Varshney, A., Brooks, F.P.Jr., Richardson, D.C., Wright, W.V., Manocha, D., 1995. Defining, Computing, and Visualizing Molecular Interfaces. Proc. of the IEEE Visualization ’95, Atlanta, GA, p.36-43.

Xu, D., Tsai, C.J., Nussinov, R., 1997. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Engineering*, **10**(9):999-1012. [doi:10.1093/protein/10.9.999]