Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2008) 1104–1112
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Pocket extraction on proteins via the Voronoi diagram of spheres

Donguk Kim [a], Cheol-Hyung Cho [b], Youngsong Cho [a], Joonghyun Ryu [a], Jonghwa Bhak [c], Deok-Soo Kim [d,*]

[a] *Voronoi Diagram Research Center, Hanyang University, Seoul, Republic of Korea*
[b] *Mechatronics and Manufacturing Technology Center, Samsung Electronics, Suwon, Republic of Korea*
[c] *National Genome Information Center (NGIC), KRIBB, Daejeon, Republic of Korea*
[d] *Department of Industrial Engineering, Hanyang University, Seoul, Republic of Korea*

Received 2 October 2007; accepted 2 October 2007
Available online 7 October 2007

Abstract

Proteins consist of atoms. Given a protein, the automatic recognition of depressed regions, called pockets, on the surface of proteins is important for protein-ligand docking and facilitates fast development of new drugs. Recently, computational approaches have emerged for recognizing pockets from the geometrical point of view. Presented in this paper is a geometric method for the pocket recognition which is based on the Voronoi diagram for atoms. Given a Voronoi diagram, the proposed algorithm transforms the atomic structure to meshes which contain the information of the proximity among atoms, and then recognizes depressions on the surface of a protein using the meshes.

Keywords: Voronoi diagram of atoms; Protein; Pocket; Binding site; Docking; Protein surface; Blending mesh

1. Introduction

Molecules such as protein, DNA, or RNA consist of atoms. Given the atomic structures of molecules, analyzing interactions between molecules is important for understanding their biological functions. An example is the interaction between a protein and a small molecule and this interaction is the basis of designing new drugs.

The study of molecular interactions, such as docking or folding, can be approached from a physicochemical and/or a geometrical point of view [30]. While the physicochemical approach is to evaluate and minimize the free energy between two molecules using, for example, the area of molecular surfaces, the geometric approach is to determine whether two molecules have geometrically meaningful features for interaction.

Interaction between a protein, called a *receptor*, and a small molecule, called a *ligand*, is usually done via some depressed regions, called *pockets*, on the surface of the receptor. Since the docking of chemicals into pockets to find an appropriate complementary ligand is essential in various aspects of drug design, identifying the potential pockets on a receptor is an important first step to efficient drug design. Considering the fact that chemical database usually contain a relatively large amount of data, automated methods are preferred for the generation of candidate pockets on receptors [24]. While the efforts on the physicochemical approach have been done since the early days of science, the efforts to understand the geometry of biological systems have started relatively recently [7,13,25,34].

Geometric approaches to recognize pockets on a protein usually involve the definition of surfaces on a protein. Initial studies on the problem primarily created a 3D spatial lattice of the space occupied by a protein and used simple techniques to reason the relative relations among the grids in the lattice to extract the exterior boundary of the protein, and then they recognized the depressed regions on the surface of the protein [6,14,15,36]. Recently, researchers have started to use more rigorously defined mathematical and computational tools related to the geometry among the atoms in a protein. α-shapes are one of the most successful efforts in this avenue. Since α-shapes can be used to represent the surface of a protein quite efficiently for a fixed size probe, it has been often used in the extraction of pockets [8]. Peters et al. and Liang et al. have even tried to explain the geometric characteristics between

* Corresponding author. Tel.: +82 2 2220 0472; fax: +82 2 2292 0472.
E-mail address: dskim@hanyang.ac.kr (D.-S. Kim).

1093-3263/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2007.10.002

extracted pockets and ligands to fit into the pockets using α-shapes [26,31]. However, the α-shape has certain limitations for various applications mainly due to the fact that it is defined from the centers of atoms, not from the surfaces of atoms and therefore it cannot take into account size difference among atoms.

In this paper, we will introduce another methodology to provide the definition of pockets on the surface of a protein from the geometric point of view and present an algorithm to recognize pockets automatically. Given a protein, the proposed algorithm first computes the *Voronoi diagram* of the atoms and constructs a *mesh structure* on the surface of the protein using the Voronoi diagram and a probe. We emphasize here that the Voronoi diagram of atoms is different from the Voronoi diagram of atom centers. The mesh structure defines the spatial proximity among the atoms on the surface of the protein with respect to the probe. Then, we define a *pocket primitive* which corresponds to each planar face of the convex hull of the protein. After pocket primitives are extracted, we evaluate the validity of boundaries between neighboring pocket primitives to test if two neighbors should be merged into a single pocket or not. Eventually, therefore, there will be a few pockets left on the surface of a receptor where each pocket corresponds to an appropriately depressed region.

2. Geometric models of a protein and related terminologies

A protein, consisting of atoms, can be viewed as shown in Fig. 1. The circles denote atoms constituting a protein and their radii are van der Waals radii. In the model, there are two kinds of surfaces associated with the protein: *solvent accessible surface* (SAS) and *molecular surface* (MS). SAS consists of a set of points on the space where the center of the probe is located when the probe is in contact with the protein without intersection with other atoms. The inner-most possible trajectories of points on the probe surface, then, define MS. SAS usually defines a *free-space* that a small molecule can move around without interfering with the protein and therefore plays a fundamental role for folding and/or docking [25]. On the other hand, MS, often called by another name *Connolly surface* after the name of the first researcher who developed an analytical algorithm for MS, conveniently defines the boundary between the interior and exterior volume of a protein so that the volume or the density of a protein can be calculated [5].

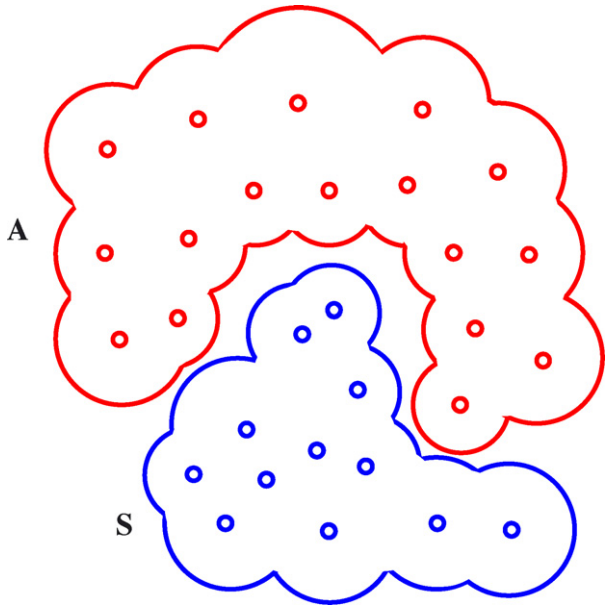

Fig. 2. A docking configuration between receptor A and ligand S.

Fig. 2 shows two molecules interacting with each other via a pocket defined on the molecular surface of molecule A. The molecule labeled A is a receptor and the small molecule labeled S is a ligand. Then, A and S interact with each other as the protruded region of S has been geometrically inserted into the depressed region, which is called a *pocket*, on the molecular surface of A.

Let $A = \{a_1, a_2, \ldots, a_n\}$ be a protein consisting of a number of atoms $a_i = (c_i, r_i)$ where $c_i = (x_i, y_i, z_i)$ and r_i define the center coordinates and the radius of an atom, respectively. In addition, suppose that $S = \{s_1, s_2, \ldots, s_m\}$ is a small molecule which also consists of a number of atoms s_j, defined similarly to a_i, and S will be docking with A. Note that $m \ll n$ in general since m is usually in the range of several dozens at most and n is between several hundreds and thousands. Usually a small molecule is approximated by the spherical probe $R = (c_R, r_R)$ and most investigations on the geometric properties for the

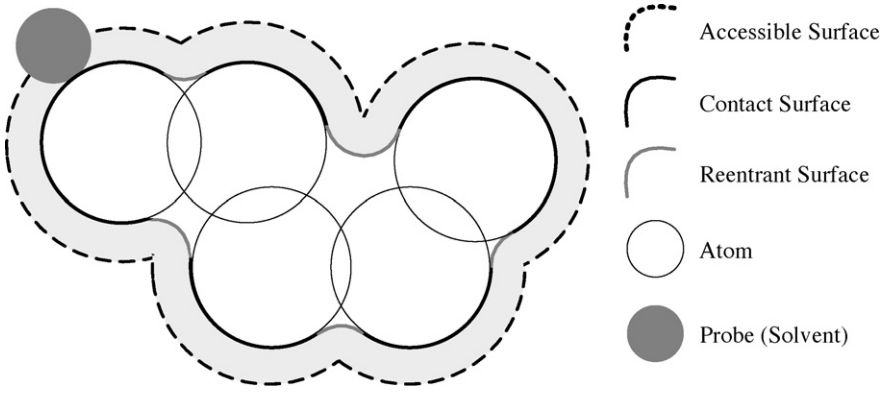

Fig. 1. The geometric model of a protein. The reentrant surface and the contact surface altogether define the molecular surface.

protein analysis are done using the probe. Note that R is defined as a minimum sized sphere enclosing all the atoms of S for simplicity. Taking the smallest enclosing sphere is a conservative approximation and therefore can possibly miss existing pockets.

Let π_j be a pocket where $\pi_j = \{a_{j1}, a_{j2}, \ldots, a_{jk}\}$ and these atoms altogether define a depressed region on the boundary B of A. The boundary $B = \{b_1, b_2, \ldots, b_l\}$ is the set of atoms of protein where some points on the van der Waals surface contribute to the molecular surface MS. In other words, B is a set of atoms $a_i \in A$ which is touched by the small molecule S without interfering with the protein A. Hence, $\pi_j \subseteq B \subseteq A$ and the set $\Pi = \{\pi_1, \pi_2, \ldots, \pi_p\}$ is the set of all possible pockets on B.

3. Topology for whole protein

To effectively and efficiently answer most geometric questions for a given protein, it is important to have a convenient tool to represent the spatial structure of a protein. In our research, we use the Voronoi diagram of atoms where the distance is defined from the boundaries, instead of the centers, of atoms.

The ordinary Voronoi diagram for a point set and its construction have been studied extensively and the properties are well-known in two and higher dimensions [29]. However, the construction of the Voronoi diagram for spheres, often referred to as an *additively weighted Voronoi diagram* [29], has not been explored sufficiently even though it has significant potential impact on diverse applications in the protein structure [1,12,28,33,35]. Due to the lack of appropriate algorithms and stable running codes for the Voronoi diagrams of spheres, most applications have instead adapted an ordinary Voronoi diagram of points, a power diagram, or an α-shape.

However, in many applications of proteins, an ordinary Voronoi diagram or a power diagram is an approximation of what is actually needed. For example, consider the computation of molecular surface on a protein. If the Voronoi diagram of spheres is available, computing the molecular surface becomes rather easier since all the information about the reentrant surface is immediately available from the Voronoi diagram. Hence, no additional search is necessary around the atomic structure. However, the ordinary Voronoi diagram of atom centers or the power diagram does not properly provide such information and therefore a potential global search is necessary to guarantee a correct result. Similar observations can be easily found in several recent studies [11,12,18]. Geode et al., for example, described an efficient algorithm to calculate the precise volume and density of a protein using the Voronoi diagram of atoms [12]. Kim et al. introduced various problems such as computing the molecular surface of a protein and locating the interaction interface in a polymer using the Voronoi diagram of atoms [16,18,19,21,23].

The Voronoi diagram of spheres is defined as follows. Associated with each atom $a_i \in A$, there is a corresponding *Voronoi region* VR_i for a_i, where $\mathrm{VR}_i = \{p | \mathrm{dist}(p, c_i) - r_i \leq \mathrm{dist}(p, c_j) - r_j, i \neq j\}$. Then, $\mathrm{VD}(A) = \{\mathrm{VR}_1, \mathrm{VR}_2, \ldots, \mathrm{VR}_n\}$ is called a *Voronoi diagram* for A. In this study, the topology of VD(A) is stored in a radial edge data structure which is the popular representation of topology for cell complexes [4]. The faces, edges, and vertices in the Voronoi diagram are denoted by $F^{\mathrm{V}} = \{f_1^{\mathrm{V}}, f_2^{\mathrm{V}}, \ldots\}$, $E^{\mathrm{V}} = \{e_1^{\mathrm{V}}, e_2^{\mathrm{V}}, \ldots\}$ and $V^{\mathrm{V}} = \{v_1^{\mathrm{V}}, v_2^{\mathrm{V}}, \ldots\}$, respectively. In this paper, the distance metric is the Euclidean distance from the boundary of an atom.

From the definition of a Voronoi diagram, a Voronoi vertex v^{V} is the center of an empty sphere tangent to four nearby atoms, while a Voronoi edge e^{V} is defined as a locus of points equidistant from the surfaces of three surrounding atoms. In addition, a Voronoi face f^{V} is a connected surface equidistant from two atoms with a greater distance from every other atom. Note that the face is always a hyperbolic surface unless it degenerates to a plane for two identical sized atoms.

Fig. 3 shows an example of VD(A) in the plane. In the figure, there are six atoms and therefore six Voronoi regions. In addition, there are 10 Voronoi edges and 5 Voronoi vertices in the model. Note that a Voronoi vertex corresponds to the center of an empty sphere tangent to three atoms. In addition, an edge is the locus of point equidistant from the circumferences of two 2D atoms. VR_6 is bounded by five Voronoi edges e_2, e_4, e_6, e_8, and e_{10}, while other Voronoi regions are unbounded. Note that all the points in VR_6 are closer to the circumference of a_6 than the circumference of every other atom in Euclidean distance.

Unlike other kinds of Voronoi diagrams, only a few reports are available on the construction of the Voronoi diagram of spheres. Aurenhammer discussed the transformation of the computation of the Voronoi diagram of d-dimensional spheres to that of a $(d+1)$-dimensional power diagram obtained from the convex hull in $(d+2)$-dimension [2]. Will discussed about the computation of Voronoi regions in the Voronoi diagram of three-dimensional spheres in his Ph.D. thesis [37]. Gavrilova, also in her Ph.D. thesis, reported several properties of the Voronoi diagram for spheres in arbitrary dimensions, including the shape of Voronoi regions, the nearest neighbors and the empty-sphere property [9,10]. Luchnikov et al. proposed a

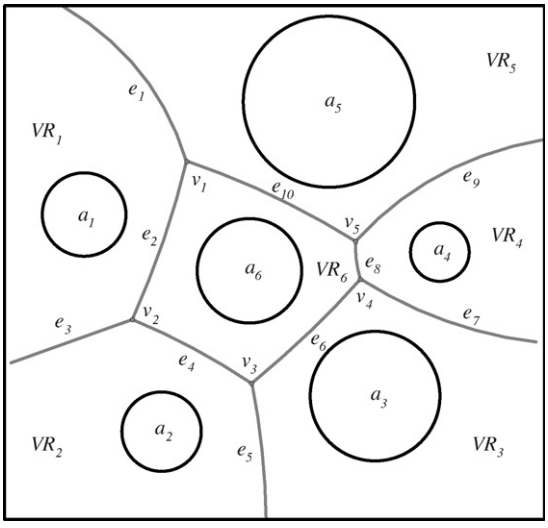

Fig. 3. A 2D example of a Voronoi diagram for circles. There are 6 atoms, 6 Voronoi regions, 10 Voronoi edges, and 5 Voronoi vertices.

practical idea of tracing edges which is simple yet powerful to obtain the desired diagram [27].

Recently, Kim et al. reported detailed characterizations of the edge-tracing algorithm and its full implementation for constructing the whole Voronoi diagram with discussions on various applications including the analysis of protein structures [17–21]. They showed that the whole Voronoi diagram can be constructed in $O(n^3)$ time in the worst-case. Kim and Kim reported another algorithm called region-expansion which uses the Voronoi diagram of sphere centers as its initial solution [23]. Kim et al. introduced the dual structure of the Voronoi diagram called quasi-triangulation and its compact data structure called interworld data structure for efficient storage and handling of the Voronoi diagram [22].

Boissonnat and Karavelas reported an algorithm to compute a Voronoi region using the convex hull of spheres transformed by inversion [3]. The convex hull topology, then, provides the topology of the region boundary.

4. Topology for surface atoms of a protein

Extracting pockets of a protein requires querying information on the surface behavior of the protein, and therefore a convenient representation of the connectivity among atoms contributing to the surface of the protein is essential. Even after the surface atoms are identified, recognizing pockets is still not an easy task from the computational point of view.

In the effort to extract pockets, we define a simpler geometric structure called a *mesh M* on the surface of a protein A as follows. Let $M = (V, E, F)$ be a mesh defined on the surface of a protein, where $V = \{v_1, v_2, \ldots\}$, $E = \{e_1, e_2, \ldots\}$ and $F = \{f_1, f_2, \ldots\}$ are sets of vertices, edges, and faces on M, respectively.

To define a mesh M on the surface of a protein A, we introduce a geometric operation among atoms called *blending*. A blending surface over a protein A consists of two kinds of blends as shown in Fig. 4: *rolling blends* and *link blends*. A rolling blend γ is defined by rolling the probe R between two atoms, and a link blend λ is defined among three neighboring atoms by placing R on the top of the atoms. Note that a rolling blend is a subset of torus and a link blend is a spherical triangle.

Suppose that two atoms a_i and a_j are in close proximity to each other so that they define a rolling blend γ_{ij}. Then, we define an edge e_{ij} between the centers c_i and c_j of the atoms a_i and a_j. If we apply this operation for all such pairs of atoms, we can obtain the definitions of E in M. Suppose that a triplet of atoms a_i, a_j and a_k defines a link blend λ_{ijk}. Then, we define a triangular face f_{ijk} using v_i, v_j, and v_k as the vertices. If we apply this operation for all such triplets of atoms, we can obtain the definitions of F in M. Since the edge e_{ij} corresponds to a Voronoi face, the edge set E can be constructed in a linear time of Voronoi faces. Similarly, the face f_{ijk} can be found in a linear time of Voronoi edges. Note that different probes produce different meshes even for an identical protein. Fig. 5 shows a 2D example illustrating a protein, the blending of the protein, and its mesh counterpart.

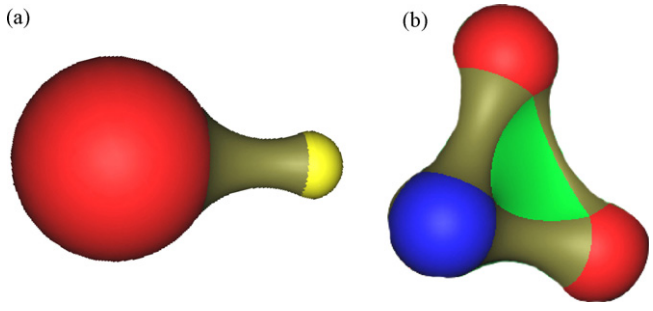

Fig. 4. Blending surfaces among atoms: (a) a rolling blend between two atoms with different radii and (b) a link blend among three atoms.

5. Extraction of pocket primitives

In this paper, let R_S be a probe for a small molecule S from which we want to define a pocket on the given protein A and let R_∞ be a hypothetical probe with an infinite radius. Let M_S and M_∞ be the mesh models defined by blending the protein A for probes with radii of R_S and R_∞, respectively. Then, M_∞ in fact corresponds to a mesh model bounded by faces defined by the centers of atoms with infinite Voronoi regions.

Let M_S and M_∞ be denoted by M_I and M_O meaning the inner and outer meshes, where $M_\mathrm{I} = (V_\mathrm{I}, E_\mathrm{I}, F_\mathrm{I})$ and $M_\mathrm{O} = (V_\mathrm{O}, E_\mathrm{O}, F_\mathrm{O})$. Fig. 6 shows examples of the inner and outer meshes for a protein. From the figure, which is a 2D analogy for our problem in 3D, we can make a simple observation as follows: for each edge of outer mesh, there is zero or one depressed region on the boundary of the protein. When an edge of the outer mesh coincides with one of the inner mesh, obviously no pocket is defined.

Although the problem in 3D is not as simple as its 2D counterpart, we can make a similar observation. For each face of the outer mesh of a protein, there can be a corresponding depressed region. A depressed region corresponding to the face of the outer mesh, however, may or may not have a clearly defined boundary since a pocket may be related to a few faces of the outer mesh. In such a case, the depressed region extracted from a face of the outer mesh cannot be defined to form a complete pocket but a pocket that has two or more faces of the outer mesh which altogether define a single pocket. Hence, we first introduce *pocket primitive* as a depressed region on M_I corresponding to each face of the outer mesh.

A face $f_\mathrm{O}^i \in F_\mathrm{O}$ has three associated vertices $v_\mathrm{O}^{i_1}$, $v_\mathrm{O}^{i_2}$, and $v_\mathrm{O}^{i_3}$. Since $M_\mathrm{O} \subseteq M_\mathrm{I}$ in the data structure perspective, there are always three vertices $v_\mathrm{I}^{i_1}$, $v_\mathrm{I}^{i_2}$, and $v_\mathrm{I}^{i_3}$ which coincide with $v_\mathrm{O}^{i_1}$, $v_\mathrm{O}^{i_2}$, and $v_\mathrm{O}^{i_3}$, respectively. Let $\tilde{\phi}_{(i_1,i_2)}$ be a path between $v_\mathrm{I}^{i_1}$ and $v_\mathrm{I}^{i_2}$ on the inner mesh M_I where $\phi_{(i_1,i_2)} = (v_\mathrm{I}^{i_1}, e_1, v_1, \ldots, e_{\phi_m}, v_\mathrm{I}^{i_2})$. Let $\phi_{(i_1,i_2)}$ be the shortest one among all possible paths between $v_\mathrm{I}^{i_1}$ and $v_\mathrm{I}^{i_2}$. The length of a path is defined as the sum of edge lengths on the path. Then the geometric meaning of the shortest path between two vertices is as follows: since two vertices are on M_O as well, the other vertices on the shortest path define depressions on M_I from the corresponding face of M_O. Hence, the shortest path defines the most upward wall separating two relatively deep depressions. $\phi_{(i_2,i_3)}$ and $\phi_{(i_3,i_1)}$ can be similarly defined.

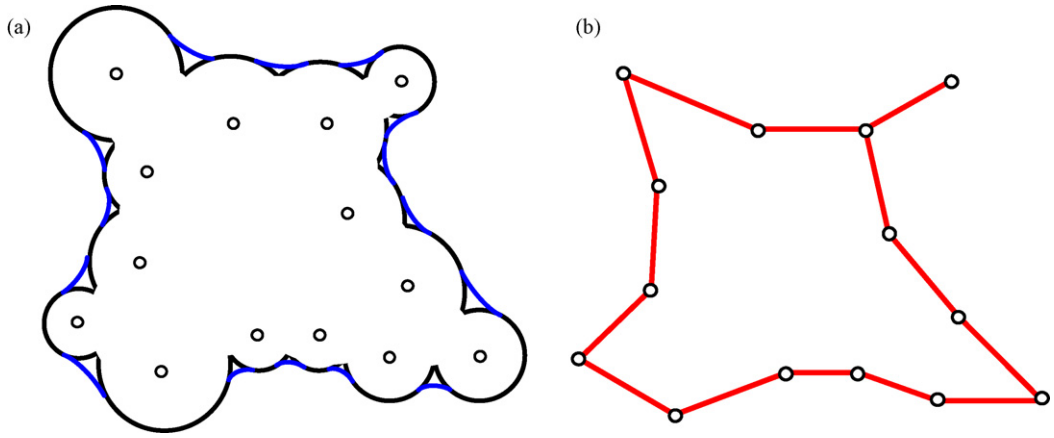

Fig. 5. An example of a blending mesh for a protein in 2D: (a) a blending surface and (b) a blending mesh.

Then, the face set $\tilde{F}_{\mathrm{I}}^{i}$ consisting of $f_{\mathrm{I}}^{h} \in F_{\mathrm{I}}$, where f_{I}^{h} denotes a face of M_{I} and is interior to the three shortest paths $\phi_{(i_1,i_2)}$, $\phi_{(i_2,i_3)}$, and $\phi_{(i_3,i_1)}$. Note that $\tilde{F}_{\mathrm{I}}^{i}$ forms a topologically triangular shaped depression on M_{I} from the boundary of M_{∞}. This depression is called a pocket primitive φ_i corresponding to a convex hull face $f_{\mathrm{O}}^{i} \in F_{\mathrm{O}}$ and is also represented as another graph $\varphi_i = (\tilde{V}_{\mathrm{I}}^{i}, \tilde{E}_{\mathrm{I}}^{i}, \tilde{F}_{\mathrm{I}}^{i})$. Note that $\tilde{F}_{\mathrm{I}}^{i}$ can be a null set in the worst-case meaning that no pocket primitive corresponds to the face.

6. Pocket recognition from pocket primitives

A pocket may consist of more than one pocket primitives. Hence, we check if two neighboring pocket primitives should be merged together to form a more meaningful depression based on an appropriate criterion. Let a *ridge* be the edge chain corresponding to the shortest path between two extreme vertices of a pocket primitive. Hence, a ridge plays a role of the boundary between two incident pocket primitives. Let a *mountains* be the edge chain separating two pockets. If a ridge is sufficiently high, it can be regarded as a mountains. Note that a pocket primitive always has three ridges and a pocket is surrounded by three or more mountains. Therefore, the boundary of a pocket primitive may or may not be the boundary of a pocket depending on conditions.

Suppose that a path ϕ^k, which is indeed a ridge, exists between φ^i and φ^j for an edge e_{O}^{k} of M_{O}. Then, we can define a certain measure to determine the discrepancy between two chains e_{O}^{k} and ϕ^k. Depending on the measure and its prescribed threshold value, two pocket primitives sharing the chains may or may not be merged. Even though there can be several ways to define such a measure, we use the average perpendicular distance of vertices on M_{I} from the edge e_{O}^{k}.

Let δ_k be the average distance between e_{O}^{k} and ϕ^k. If δ_k is larger than a prescribed value, we merge two neighboring pocket primitives sharing the chains. Otherwise, we regard ϕ^k as a mountain chain. As such a threshold value, in this paper, we use the average of δ_k.

7. Discussions

Shown in Fig. 7(a) is a dimer *Transcription regulation complex*, coded as 1BH8 in PDB [32], which consists of two chains as illustrated in different colors. Fig. 7(b) illustrates the blue chain only.

Fig. 8 illustrates an example of the pocket extraction on the surface of Fig. 7(b). After computing the Voronoi diagram of the protein, the convex hull of the protein is computed in Fig. 8(a). Shown in Fig. 8(b) is the surface blending of the protein using a predefined probe, and Fig. 8(c) shows the mesh structure of the

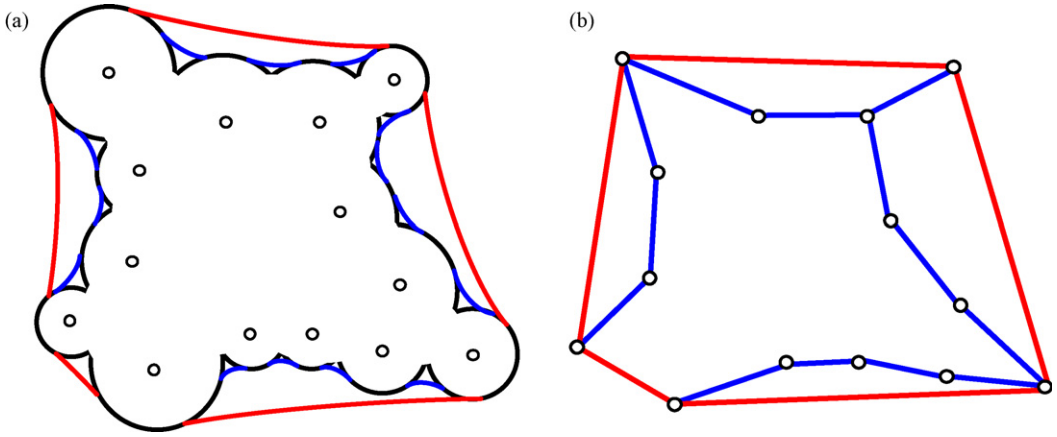

Fig. 6. Inner and outer meshes: (a) blending surfaces and (b) blending meshes.

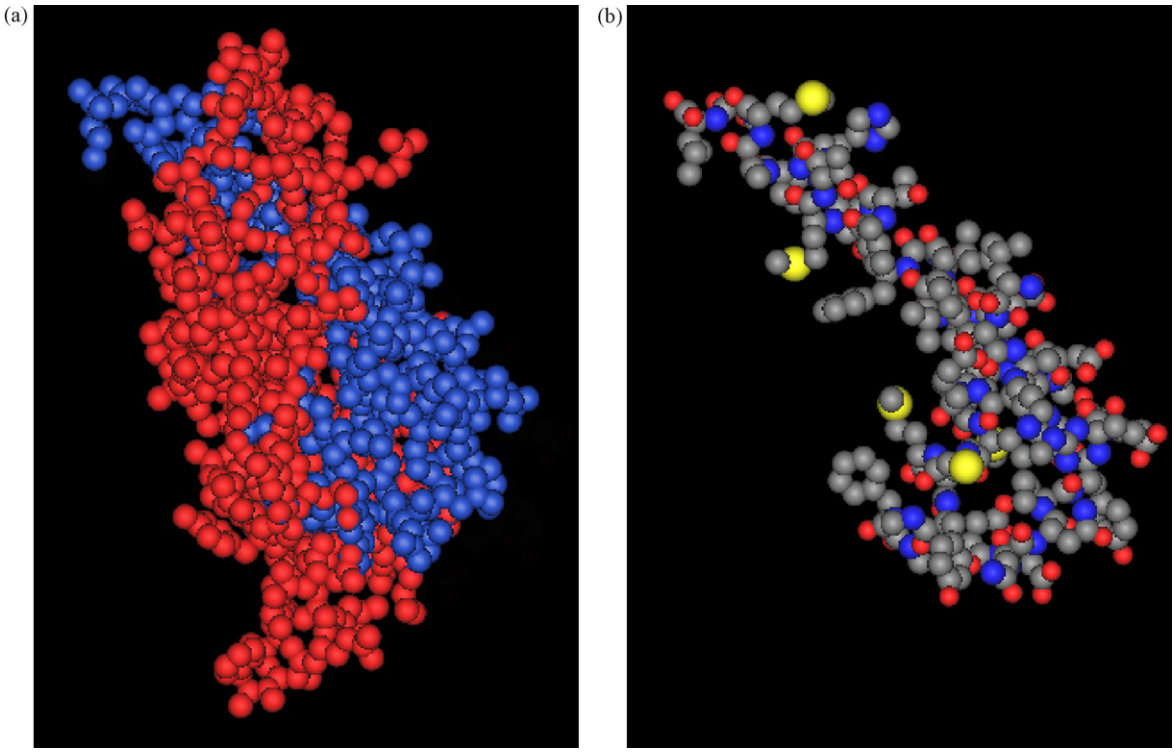

Fig. 7. The atomic structure of the protein 1BH8 [32] downloaded from PDB database: (a) two chains in different colors and (b) the blue chain only.

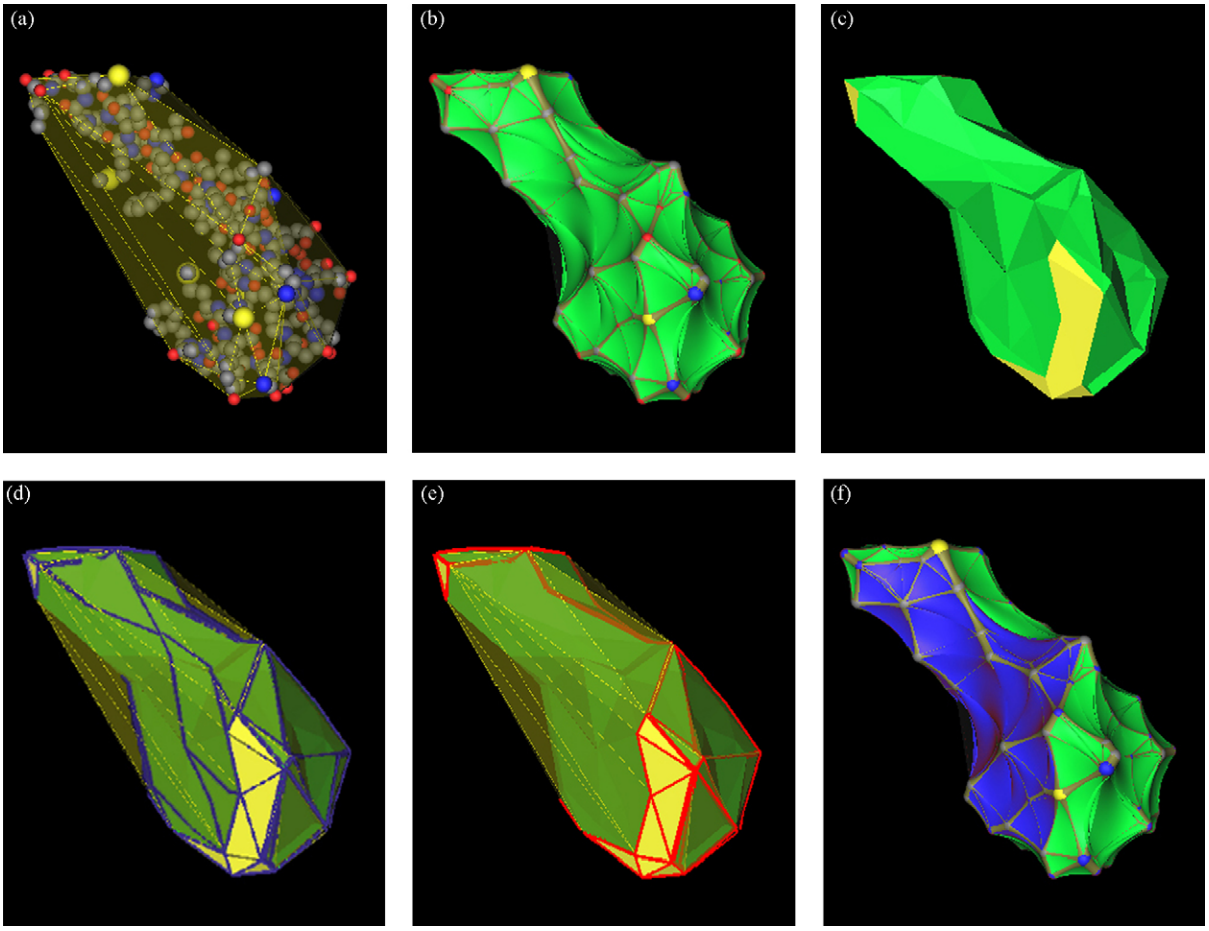

Fig. 8. The blue chain of 1BH8 and the procedure in the pocket extraction: (a) the convex hull, (b) the molecular surface, (c) the blending mesh, (d) the pocket primitives, (e) the boundary of pockets after merging, and (f) the largest pocket on the molecular surface.

surface atoms where the connectivity of the mesh illustrates the proximity among the atoms. Then, the pocket primitive corresponding to each face of the convex hull of the protein is shown in Fig. 8(d). In this example, we have chosen the mesh of the convex hull as an outer mesh M_O and the blending mesh for a probe with a radius 8 Å as an inner mesh M_I, since this probe gives the most convincing result. When a face on the outer mesh coincides with one of the inner mesh, we represented the face with yellow. After pocket primitives are extracted, we evaluate the qualities, the average depths, of ridges around all pocket primitives and merge the appropriate pocket primitive pairs. Fig. 8(e) shows the mesh structure on M_I of merged pockets while Fig. 8(f) illustrates the corresponding visualization using a molecular surface representation for the largest pocket in blue.

While the example in Fig. 8 is based on the convex hull faces, Fig. 9 shows the influence of probe size on the grains of pocket primitives by illustrating the pocket primitives based on outer meshes M_O's constructed by some probes with finite sizes. For example, the mesh shown in the left-most subfigure of Fig. 9(a) illustrates the topological proximity when the probe size is 60 Å, rather than ∞ used for the convex hull. The second subfigure in Fig. 9(a) shows an extracted pocket using the

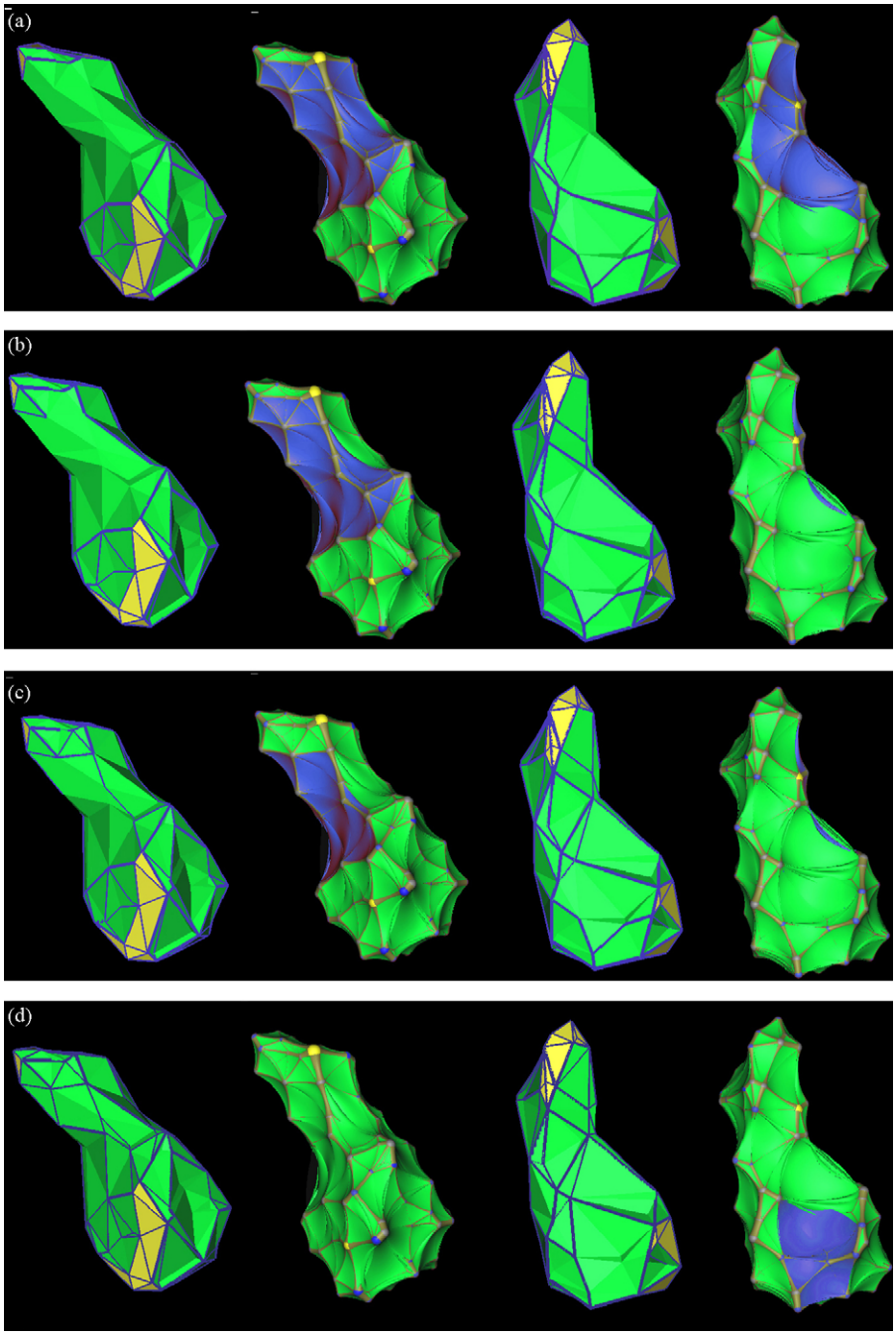

Fig. 9. The blending mesh and the largest pocket (blue surface) on the molecular surface of the blue chain for different probe sizes: (a) 60 Å, (b) 40 Å, (c) 30 Å, and (d) 20 Å. The radius of the probe is 8 Å for all inner meshes.

meshes in the first figure. It can be clearly shown that the extracted pocket in this figure is different from its counterpart in Fig. 8(f). The third and fourth models in Fig. 9(a) show the previous two models from a different direction. What are shown in Fig. 9(b–d) are the recognized pockets with the largest area on the protein, where the pockets are extracted from the mesh constructed from probes of radii 40, 30, and 20 Å's. When the radius of the probe is 50 Å, the pocket is identical to the case where the radius of the probe is 40 Å.

From the experiments, it can be observed that the number of pockets increases as the radius of the probe decreases and the resolution of pockets increases. In addition, there is a very strong tendency that an increasing probe radius causes the increase of the area of a pocket on the molecular surface. Fig. 9(c) and (d) even shows that the change of the probe size can even cause a drastic change of pockets. In the figures, the largest pockets for both cases have even completely different locations on the protein.

8. Conclusions

The search of pockets on the molecular surface of a protein is one of the most important starting points for a structure-based drug design. In this paper, we have provided the definition of pockets on a protein from the geometric point of view and presented an algorithm to automatically recognize pockets.

In the algorithm, we first computed the Voronoi diagram of atoms and constructed meshes, an inner and an outer mesh, using the topology information in the Voronoi diagram. Then, we defined a pocket primitive on the inner mesh for each face of the outer mesh. After the pocket primitives were extracted, we evaluated the validity of boundaries between neighboring pocket primitives to test if two neighbors should be merged into a single pocket or not. Eventually, there were a few pockets left on the surface of a receptor where each pocket corresponds to an appropriately depressed region. The algorithm in this paper has been implemented in C++ on Windows XP and test results for several probes are provided for an example protein.

There are three possible improvements of this research in the future. First, the structure of meshes can be more formally discussed. Second, the ∞-sized probe can be more realistically adjusted. Third, merging criteria between pocket primitives can be revised by applying more meaningful properties such as pocket depth and volume.

Acknowledgement

This research was supported by the Korea Science and Engineering Foundation (KOSEF) through the National Research Lab Program funded by the Ministry of Science and Technology, Korea (No. R0A-2007-000-20048-0).

References

[1] B. Angelov, J.-F. Sadoc, R. Jullien, A. Soyer, J.-P. Mornon, J. Chomilier, Nonatomic solvent-driven Voronoi tessellation of proteins: an open tool to analyze protein folds, Prot. Struct. Funct. Genet. 49 (4) (2002) 446–456.

[2] F. Aurenhammer, Power diagrams: properties, algorithms and applications, SIAM J. Comput. 16 (1987) 78–96.

[3] J.-D. Boissonnat, M.I. Karavelas, On the combinatorial complexity of Euclidean Voronoi cells and convex hulls of *d*-dimensional spheres, in: Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms, 2003, pp. 305–312.

[4] Y. Cho, D. Kim, D.S. Kim. Topology representation for the Voronoi diagram of 3D spheres, Int. J. CAD/CAM 5 (1) (2005) 59–68, (also available at http://www.ijcc.org).

[5] M.L. Connolly, Solvent-accessible surfaces of proteins and nucleic acids, Science 221 (1983) 709–713.

[6] J.S. Delaney, Finding and filling protein cavities using cellular logic operations, J. Mol. Graph. 10 (1992) 174–177.

[7] H. Edelsbrunner, M. Facello, J. Liang, On the definition and the construction of pockets in macromolecules, Discrete Appl. Math. 88 (1998) 83–102.

[8] H. Edelsbrunner, E.P. Mücke, Three-dimensional alpha shapes, ACM Trans. Graph. 13 (1) (1994) 43–72.

[9] M. Gavrilova, J. Rokne, Updating the topology of the dynamic Voronoi diagram for spheres in Euclidean *d*-dimensional space, Comput. Aided Geometr. Design 20 (4) (2003) 231–242.

[10] M. Gavrilova, Proximity and applications in general metrics, PhD thesis, Department of Computer Science, The University of Calgary, Calgary, Canada, 1998.

[11] M. Gerstein, J. Tsai, M. Levitt, The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra, J. Mol. Biol. 249 (5) (1995) 955–966, 6.

[12] A. Goede, R. Preissner, C. Frömmel, Voronoi cell: new method for allocation of space among atoms: elimination of avoidable errors in calculation of atomic volume and density, J. Comput. Chem. 18 (9) (1997) 1113–1123.

[13] A. Heifets, M. Eisenstein, Effect of local shape modifications of molecular surfaces on rigid-body protein–protein docking, Prot. Eng. 16 (3) (2003) 179–185.

[14] M. Hendlich, F. Rippmann, G. Barnickel, LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins, J. Mol. Graph. Model. 15 (6) (1997) 359–363.

[15] C.M. Ho, G.R. Marshall, Cavity search: an algorithm for the isolation and display of cavity-like binding regions, J. Comput. Aided Mol. Design 4 (1990) 337–354.

[16] C.-M. Kim, C.-I. Won, Y. Cho, D. Kim, S. Lee, J. Bhak, D.-S. Kim, Interaction interfaces in proteins via the Voronoi diagram of atoms, Comput. Aided Design 38 (11) (2006) 1192–1204.

[17] D.-S. Kim, Y. Cho, D. Kim, C.-H. Cho, Protein sructure analysis using Euclidean Voronoi diagram of atoms, in: Proceedings of the International Workshop on Biometric Technologies (BT2004), 2004, pp. 125–129.

[18] D.-S. Kim, Y. Cho, D. Kim, S. Kim, J. Bhak, S.-H. Lee, Euclidean Voronoi diagram of 3D spheres and applications to protein structure analysis, in: Proceedings of the First International Symposium on Voronoi Diagrams in Science and Engineering (VD2004), 2004, pp. 137–144.

[19] D.-S. Kim, Y. Cho, D. Kim, S. Kim, J. Bhak, S.-H. Lee, Euclidean Voronoi diagrams of 3D spheres and applications to protein structure analysis, Jpn. J. Indus. Appl. Math. 22 (2) (2005) 251–265.

[20] D.-S. Kim, Y. Cho, D. Kim, Edge-tracing algorithm for Euclidean Voronoi diagram of 3D spheres, in: Proceedings of the 16th Canadian Conference on Computational Geometry, 2004, pp. 176–179.

[21] D.-S. Kim, Y. Cho, D. Kim, Euclidean Voronoi diagram of 3D balls and its computation via tracing edges, Comput. Aided Design 37 (13) (2005) 1412–1424.

[22] D.-S. Kim, D. Kim, Y. Cho, K. Sugihara, Quasi-triangulation and inter-world data structure in three dimensions, Comput. Aided Design 38 (7) (2006) 808–819.

[23] D. Kim, D.-S. Kim, Region-expansion for the Voronoi diagram of 3D spheres, Comput. Aided Design 38 (5) (2006) 417–430.

[24] I.D. Kunts, Structure-based strategies for drug design and discovery, Science 257 (5073) (1992) 1078–1082.

[25] B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, J. Mol. Biol. 55 (1971) 379–400.

[26] J. Liang, H. Edelsbrunner, C. Woodward, Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design, Prot. Sci. 7 (1998) 1884–1897.
[27] V.A. Luchnikov, N.N. Medvedev, L. Oger, J.-P. Troadec, Voronoi–Delaunay analysis of voids in systems of nonspherical particles, Phys. Rev. E 59 (6) (1999) 7205–7212.
[28] J.C.G. Montoro, J.L.F. Abascal, The Voronoi polyhedra as tools for structure determination in simple disordered systems, J. Phys. Chem. 97 (16) (1993) 4211–4215.
[29] A. Okabe, B. Boots, K. Sugihara, S.N. Chiu, Spatial tessellations: concepts and applications of Voronoi diagrams, 2nd edition, John Wiley & Sons, Chichester, 1999.
[30] D. Parsons, J. Canny, Geometric problems in molecular biology and robotics, in: Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, Stanford University, California, USA, August 14–17, (1994), pp. 322–330.
[31] K.P. Peters, J. Fauck, C. Frömmel, The automatic search for ligand binding sites in protein of known three dimensional structure using only geometric criteria, J. Mol. Biol. 256 (1996) 201–213.
[32] RCSB Protein Data Bank Homepage. http://www.rcsb.org/pdb/.
[33] F.M. Richards, The interpretation of protein structures: total volume, group volume distributions and packing density, J. Mol. Biol. 82 (1974) 1–14.
[34] B.K. Shoichet, I.D. Kunts, Protein docking and complementarity, J. Mol. Biol. 221 (1991) 327–346.
[35] V.P. Voloshin, S. Beaufils, N.N. Medvedev, Void space analysis of the structure of liquids, J. Mol. Liq. 96–97 (2002) 101–112.
[36] R. Voorintholt, M.T. Kosters, G. Vegter, G. Vriend, W.G. Hol, A very fast program for visualizing protein surfaces, channels and cavities, J. Mol. Graph. 7 (4) (1989) 243–245.
[37] H.-M. Will, Computation of additively weighted voronoi cells for applications in molecular biology, PhD thesis, Swiss Federal Institute of Technology, Zurich, 1999.