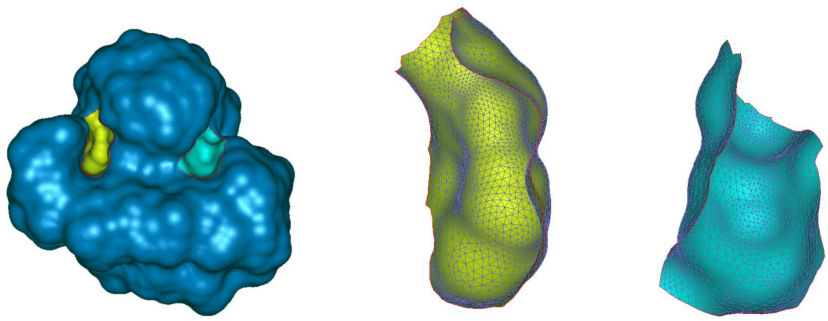
# Cavities on the Surfaces of Macromolecules

Ho-Lun Cheng and Xinwei Shi

School of Computing
National University of Singapore
{hcheng, shixinwe}@comp.nus.edu.sg

**Abstract.** We present an approach to define and extract the cavities on the surfaces of macromolecules. Each cavity is represented by a triangular mesh enclosing a depression on a molecule such as a protein and a DNA. These surface patches would facilitate the study of ligand docking problem and similarity matching of proteins.



(a) (b) (c)

**Fig. 1.** Two cavities on the surface of a Thrombin protein molecule. Figure (a) illustrates the cavities on the surface with different colors. The red curves are the boundaries of the cavities. Figure (b) and (c) illustrate the zoomed views of the two cavities.

## 1 Introduction

The geometric shapes of proteins play a deterministic role for their functions. Identification and extraction of shape features such as convexity and concavity on protein surfaces are important for computer aided drug design and protein recognition [1, 9]. In this paper, we define and construct the *cavities* on macromolecular surfaces as surface patches that enclose some depressions on the molecules.

The approach to define and extract protein surface cavities in this paper is based on the technical notion of the *skin surface* and *alpha shape*[5, 7]. The skin surface, defined by Edelsbrunner, works as a surface model for molecules with a number of distinct properties such as smoothness, decomposability and complementarity. These properties are desirable in molecular modeling applications and suggest the molecular skin model outruns the existing molecular models, for example, the molecular surface model [5]. A skin surface is a smooth 2-manifold specified by a set of spheres and has intrinsic relationships with the alpha shape defined by the same set of spheres. The alpha shape, proposed by Edelsbrunner as well, is a powerful geometric tool in biological applications. The *pockets* in the alpha shape is used to search active sites in proteins [7]. Each pocket characterizes the atoms that may implicate an active site. The *cavities* in this paper extend the definition of the pocket. Each cavity specifies a region on a molecular surface that would be the interface of protein-protein interactions.

**Outline.** Section 2 contains the necessary geometric background of the skin surface and alpha shape. In Section 3, we define the cavity and extract the surface patch of each cavity from a molecular surface mesh. Finally, some experimental results and future works are presented in Section 4.

## 2 Background

This section introduces the skin model and the alpha shape of molecules, with the aim to develop the definition of the cavity on the molecular surfaces in the next section.

### 2.1 Molecular Skin

A skin surface is defined by a set of weighted points

$$B = \{b_i = (z_i, w_i) \in \mathbb{R}^3 \times \mathbb{R} \mid i = 1..n\},$$

where $b_i$ is a weighted point with $z_i$ as its position and $w_i$ as its weight. The skin surface $F_B$ defined by $B$ is the boundary of an infinite family of spheres derived from $B$ by convex combination and shrinking operation. We omit the details of the skin definition. Readers can refer to Edelsbrunner for the details [4, 5].

The skin surface is a closed $C^1$ continuous surface with continuous maximum principle curvature. It can be decomposed by the *mixed complex,* $M_B$*,* into a finite collection of quadratic patches, namely, sphere patches and hyperboloid patches. The mixed complex is a collection of polyhedra that partitions the $\mathbb{R}^3$ space. The portion of the skin surface in each polyhedron is a quadratic patch.

To model a molecule with the skin surface, we consider each atom as a weighted point $b_i \in B$. That is, the position $z_i$ is the center of an atom, and its weight $w_i$ is $\sqrt{2}$ times the summation of the atom's van der Waals radius with the radius of the probe sphere, which is usually chosen as 1.4 Angstrom

to represent the water as the solvent. Then, the skin surface, $F_B$, represents a surface model of the molecule. The Figure 1 (a) shows the molecular skin model of a protein molecule.

### 2.2 Alpha Shapes

The space filling model of molecules are commonly used in computational biology literature also [8]. In this model, an atom is represented by a spherical ball and a molecule is the union of balls. The space filling model are always studied on the base of its dual shape, namely, alpha shape.

An alpha shape specified by a set of weighted points is the underlying space of a subcomplex of the Delaunay triangulation parameterized by a real value $\alpha$. Recall that $B$ is a set of weighted points. We denote $B(\alpha)$ as the set of the points with the weights grown by $\alpha$, that is, the weight $w_i$ for each $b_i$ increases with $\alpha$. The alpha complex specified by $B$ and $\alpha$, $K_{B(\alpha)}$, is a subcomplex of its Delaunay triangulation $D_B$, namely,

$$K_{B(\alpha)} = \{\delta_X \in D_B | \bigcup B(\alpha) \cap \nu_X \neq \emptyset\},$$

in which $\bigcup B(\alpha)$ is the union of the balls in $B(\alpha)$ and $\nu_X$ is the Voronoi cell of the Delaunay simplex $\delta_X \in D_B$. We denote the alpha shape as $|K_{B(\alpha)}|$. Each simplex in $D_B$ enters the alpha complex as the value of $\alpha$ increases from $-\infty$ to $\infty$. At the same time, the alpha shape $|K_{B(\alpha)}|$ changes from an empty set to the convex hull of the centers of weighted points in $B$. The dark triangle and the line segment in Figure 2 illustrate an example of the alpha shape specified by a set of disks in $\mathbb{R}^2$.

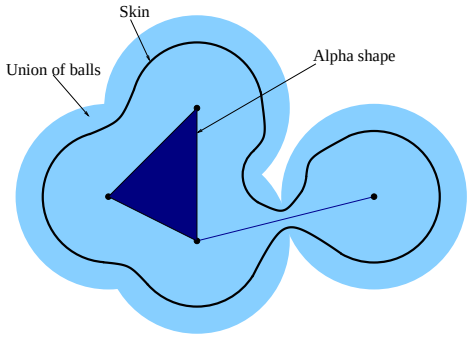



**Fig. 2.** The relationship among the union of ball, the skin and the alpha shape. The light region in the figure is the union of four disks, the dark curve are the corresponding skin and the dark triangle and the line segment are the alpha shape.

We denote $K_B$ as the alpha complex of $B$ when the value of $\alpha$ is zero. It is obvious that the alpha shape $|K_B|$ is contained in the union balls in

$B$. Moreover, the alpha shape $|K_B|$ is enclosed by the skin surface $F_B$ also and the skin surface is enclosed in the union of balls in $B$ as well. Formally, $|K_B| \subset \text{Bdy}(F_B) \subset \bigcup B$, in which $\text{Bdy}(F_B)$ is called the skin body, which is the space enclosed by the skin surface $F_B$. In particular, these three spaces have same topological properties: $|K_B| \simeq \text{Bdy}(F_B) \simeq \bigcup B$, in which we denote $X \simeq Y$ if two topological space $X$ and $Y$ are *homotopy equivalent* [3, 5]. The topological space $X$ is homotopy equivalent to $Y$ means $X$ and $Y$ are connected in the same way and we can construct a map that deforms $X$ to $Y$. This relationship is illustrated in the Figure 2. Based on this relationship, we define the *cavities* on the skin surface by investigating the pockets in its alpha shape.

## 3 Cavities on the Molecular Surfaces

The cavities on a molecular surface are defined on the base of the pockets in space filling model. We first introduce the concept of the pockets. Then, we define the cavities on the skin surface.

### 3.1 Pockets

A pocket in a union of balls is a portion of the complementary space of $\bigcup B$ that have some narrow entrances from the outside [6]. Figure 3 illustrates an example of the pocket of an alpha shape in $\mathbb{R}^2$.

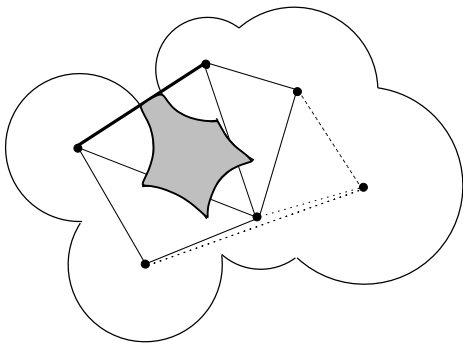


**Fig. 3.** A pocket in an alpha shape specified by 6 disks. The dark region is the pocket of the union of disks, three triangles are the pockets in the alpha shape and the bold edge is its mouth.

A pocket, $P_i$, can be represented by a collection of tetrahedra in the Delaunay triangulation $D_B$ that is not a simplex in $K_B$, that is, $P_i \subset D_B - K_B$. In this paper, we use $P_i$ to refer to a pocket in the alpha shape $|K_B|$. In Figure 3, the three triangles with solid edges forms the pocket of the alpha shape specified by

the 6 disks. The pockets are constructed efficiently by investigating the entering order of the tetrahedra in $D_B$ into the alpha complex $K_{B(\alpha)}$ as $\alpha$ increases [6].

We consider the boundary of a pocket $P_i$. It is a collection of triangles and is denoted as $\partial P_i$. The triangles in $\partial P_i - K_B$ are the interfaces between the pockets and space outside the union of balls. Each connected component in $\partial P_i - K_B$ is called a *mouth* of the pocket $P_i$ and is denoted as $M_{ij}$, in which $j = 1 \dots t$ and $t$ is the number of the mouths of $P_i$. In $\mathbb{R}^2$, each mouth is a chain of Delaunay edges. See Figure 3 for an example. The thick edge is the mouth of the pockets in the alpha shape.

### 3.2 Cavities

With the pocket definition, we define the cavities on the skin surface $F_B$ and specify its representation in this section.

Since the skin surface $F_B$ is homotopy equivalent to the union of balls $\bigcup B$, a depression on the surface $F_B$ would correspond to a concave feature on the boundary of the union of balls. A pocket $P_i$ in the alpha shape $|K_B|$ always encloses such a depression. In particular, these depressions open up to the outside with narrow mouths. We define the cavities on the skin surface as the portion enclosing such depressions, that is, $F_B \cap \bigcup P_i$.

As a result, a cavity is a surface patch on $F_B$ clipped by the pocket $P_i$. It is interesting that only the boundary of $P_i$ may intersect the skin surface $F_B$. Specifically, only the mouths of $P_i$ intersect the skin surface. The reason is that the skin surface is a subset of the complementary space of the alpha shape because of $|K_B| \subset \text{Bdy}(F_B)$. For the cavity specified by $P_i$, it is enclosed by $\bigcup P_i$ and opens up to the outside at the mouths of $P_i$ as well. Thus, a cavity is a connected surface patch. Its boundaries are the intersection of $F_B$ and the mouths of $P_i$, that is, $\bigcup_{j=1\dots t} F_B \cap M_{ij}$. We propose the following lemma to specify the boundaries of a cavity.

**Lemma 1** *The mouths of a pocket $P_i$ intersect the skin surface $F_B$ with topological circles.*

Proof. We consider the intersection of a mouth $M_{ij}$ with the skin surface $F_B$. According to the mouth definition, $M_{ij}$ is a topological disk consisted of a triangle or a chain of triangle connected by shared edges.

We firstly prove that the mouth $M_{ij}$ is not enclosed by the skin surface $F_B$. Since the mouths are the entrance from the outside to the inner part of the pockets, we have $M_{ij} \cap \{\mathbb{R}^3 - \bigcup B\} \neq \varnothing$. Moreover, because the skin body is a subset of the union of balls in $B$, that is, $\text{Bdy}(F_B) \subset \bigcup B$, then, we have $\{\mathbb{R}^3 - \bigcup B\} \subset \{\mathbb{R}^3 - \text{Bdy}(F_B)\}$. Thus, we have $M_{ij} \nsubseteq \text{Bdy}(F_B)$.

Then, we show the boundaries of the mouth $M_{ij}$ is enclosed by the skin surface $F_B$. The boundary of $M_{ij}$, denoted as $\partial M_{ij}$, consists of a loop of edges in $K_B$ because each edge is a face a triangle in $K_B$. Thus, $\partial M_{ij} \subset |K_B|$. Since $|K_B| \subset \text{Bdy}(F_B)$, then we have $\partial M_j \subseteq \text{Bdy}(F_B)$.

As a result, $M_{ij} \cap F_B$ is a set of topological circles. ◻

### 3.3 Extraction of the Cavities

Based on the skin decomposition, the intersection of each mouth $M_{ij}$ with the skin surface $F_B$ are loops of arcs intersecting triangles with a quadratic patch. The cavity on $F_B$ specified by the pocket $P_i$, $F_B \cap \bigcup P_i$, is a surface patch with the boundaries $\{F_B \cap M_{ij} | j = 1, 2, \ldots, t\}$. We can extract the cavities on a skin surface by computing the intersection of mouths and the skin surface.

Instead of computing the intersection curve $\gamma = F_B \cap \partial P_i$ analytically, we compute an approximation of the curve $\gamma$ using the skin mesh. The reason is the cavities extracted from a skin mesh are represented by a triangular mesh, which can facilitate fast combinatorial algorithm for applications such as locating active sites on protein surface in similarity matching.

The skin mesh is generated efficiently by the algorithm proposed in [2]. We implement the computation of the pockets using the algorithm proposed in [6]. The boundaries of each cavity are computed by intersecting the skin mesh with the mouths of the pockets. We obtain a chain of triangles in the skin mesh representing the boundaries of the cavity. Then, we explore the triangles along this boundary toward the direction to the inner part of the pocket to extract the triangular mesh of the cavities.

## 4 Results and Discussion

In this paper, we define the cavities on the surface of macromolecules and represent them as a triangular mesh. Figure 1 illustrates an example of the cavities on a protein surface.

The cavities can be taken as an extension of the pockets in alpha shapes. We consider the applications from two aspects. Firstly, we are aiming at a new method for the ligand docking problem [10]. The complementarity property of the skin surface tells us we could find a set of weighted points in the complementary space of the union of $B$ to specify $F_B$. Thus, we may define the protrusions on the molecular skin surface as well. With the notions cavities and protrusion, we can predict the conformation of docked proteins by matching the cavity surface patches and protrusion patches. Secondly, we may use the cavities to compare the similarity of two different proteins[9]. Since similar shape of the active sites on two different proteins' surfaces may suggest similarities of protein functions, we can compare two proteins by computing the geometric similarity of the cavity patches on the surface of proteins.

## References


1. Adrian H.Elcock, D. S., and McCammon, J. Computer simulation of protein-protein interactions. *Journal of Physical Chemistry B 105* (2001), 1504–1518.
2. Cheng, H.-L., and Shi, X. Guaranteed quality triangulation of molecular skin surfaces. In *Proceedings of IEEE Visualization 2004* (to appear).

3. Edelsbrunner, H. The union of balls and its dual shape. In *Proceedings of the ninth annual symposium on Computational geometry* (1993), ACM Press, pp. 218–231.
4. Edelsbrunner, H. Smooth surfaces for multiscale shape representation. *Proc. Sympos. Found. Software Techn. Theoret. Comput. Sci.* (1995), 391–412.
5. Edelsbrunner, H. Deformable smooth surface design. *Discrete Computational Geometry 21* (1999), 87–115.
6. Edelsbrunner, H., Facello, M., and Liang, J. On the definition and the construction of pockets in macromolecules. Tech. rep., 1995.
7. Edelsbrunner, H., and Mucke, E. P. Three-dimensional alpha shape. *ACM Transaction on Graphics 13*, 1 (Jan. 1994), 43–72.
8. Leach., A. R. *Molecular Modelling.* Longman, Harlow, England, 1996.
9. Maria Rosen, Shuo Liang Lin, H. W., and Nussino, R. Molecular shape comparisons in searches for active sites and functional similarity. *Protein Engineering 11*, 4 (1998), 263–277.
10. R.D., T., P.J., J., and J.W., E. A review of protein-small molecule docking methods. *Journal of Computer-Aided Molecular Design 16* (2002), 151–166.