

Alpha shape and Delaunay triangulation in studies of protein-related interactions

Weiqiang Zhou and Hong Yan

Submitted: 24th August 2012; Received (in revised form): 24th October 2012

Abstract

In recent years, more 3D protein structures have become available, which has made the analysis of large molecular structures much easier. There is a strong demand for geometric models for the study of protein-related interactions. Alpha shape and Delaunay triangulation are powerful tools to represent protein structures and have advantages in characterizing the surface curvature and atom contacts. This review presents state-of-the-art applications of alpha shape and Delaunay triangulation in the studies on protein–DNA, protein–protein, protein–ligand interactions and protein structure analysis.

Keywords: *Delaunay triangulation; alpha shape; protein–DNA interactions; protein–protein interactions; protein–ligand interactions; protein structure analysis*

INTRODUCTION

The study of proteins, one of the essential elements in biological organisms, has attracted scientists' attention for hundreds of years since it was first recognized by Antoine Fourcroy in the 18th century. More importantly, protein-related interactions such as protein–DNA, protein–protein and protein–ligand interactions take place in almost every living organism. As the key process of gene inheritance, protein–DNA interaction involves DNA replication, DNA transcription and nucleosome remodeling [1–4]. Protein–protein interaction plays a crucial role in cellular function [5, 6] and studies of protein–protein interaction provide a better understanding of the functional organization of the proteome [7]. Among the protein-related interactions, protein–ligand interaction has been paid the most attention because it is related to the understanding of protein functions and drug development [8, 9].

As tremendous progress is being made in structural biology, more high-resolution 3D molecular

structures are becoming available. There is a growing demand for a 3D geometric model to represent and study protein-related interactions. Here, we review the alpha shape [10] and its underlying model Delaunay triangulation [11]. The 3D Delaunay triangulation, applied for the first time in the analysis of protein structure by Singh *et al.* [12], is a unique partition of 3D space with non-overlapping tetrahedrons. In a protein structure, the Delaunay triangulation is a powerful tool to represent the relationship between the neighboring atoms. Singh *et al.* derives five classes of every four neighboring C_{α} atoms and develops a four-body potential to evaluate sequence–structure compatibility for solving the inverse protein folding problem. The alpha shape is proposed by Edelsbrunner and Mücke for computing the 3D structure of a finite point set in space based on the Delaunay triangulation. It is first applied to the field of structural biology by Liang *et al.* [13, 14] to compute the molecular area and volume and detect inaccessible cavities in proteins. In the

Corresponding author. Weiqiang Zhou, Department of Electronic Engineering, City University of Hong Kong, Tat Chee Avenue 83, Hong Kong. Tel: +852-9062-9845; Fax: +852-3442-0437; E-mail: kenandzhou@hotmail.com

Weiqiang Zhou is a PhD student at the Department of Electronic Engineering, City University of Hong Kong. His research in bioinformatics focuses on prediction of biomolecular interactions using mathematical models and machine learning techniques.

Hong Yan is a professor at the Department of Electronic Engineering, City University of Hong Kong. His research in bioinformatics focuses on analysis of genomic data and prediction of biomolecular interactions using advanced signal processing and pattern recognition techniques. Prof. Yan is a fellow of IEEE and IAPR.

past, Poupon presented a review focusing on the application of the Voronoi tessellation in studying protein structure and interaction [15]. However, great development has taken place in the field of structural biology where a review of the state-of-the-art methods is required. We therefore present a review of the recent applications of alpha shape and Delaunay triangulation in the studies of protein–DNA, protein–protein, protein–ligand interactions and some other protein structure studies.

ALPHA SHAPE AND DELAUNAY TRIANGULATION

Delaunay triangulation or Delaunay tessellation, the dual shape of the Voronoi diagram enables a unique division of the space to be made based on nearest neighbors. We can define the Delaunay triangulation based on its duality with a Voronoi diagram. Given a set of points $S = \{P_i | i = 1, 2 \dots, n\}$, the Voronoi diagram (Figure 1A) is the set of cells, V_i defined by:

$$V_i = \{P | d(P, P_i) \leq d(P, P_j), \forall j \neq i\}. \quad (1)$$

where $d(P, P_i)$ is the distance between P and P_i . In other words, V_i is the locus of the points closer to P_i than any other points in S . Then, the Delaunay triangulation can be computed as the dual shape of the Voronoi diagram (Figure 1B). Using a 3D Delaunay triangulation, one can easily get the atom contacts which are represented by the edges in the Delaunay triangulation (Figure 2A).

Alpha shape can be derived from Delaunay triangulation, which offers a concrete definition of a shape to represent the structure of a set of points. Two versions of alpha shape have been developed: the basic alpha shape (Figure 1C) and the weighted alpha shape (Figure 1D). For the basic alpha shape, we consider a set of none weighted points M and a k -simplex σ_T is defined as subset $T \subseteq M$ of size $|T| = k + 1$, where k is 0, 1, 2 or 3. A ball b with $b \cap M = 0$ is defined as an empty ball. The Delaunay triangulation can be obtained as the collection of all the k -simplex that have empty open balls b with $T = \partial b \cap M$ where ∂b is the boundary of ball b . We can see that the Delaunay triangulation contains the vertices, edges, triangles and tetrahedrons which are represented by the 0-simplex, 1-simplex, 2-simplex and 3-simplex, respectively. Therefore, the alpha shape can be obtained as a subset of the Delaunay triangulation which is controlled by the value of α . Define an α -ball as an open ball with radius α , $0 \leq \alpha \leq \infty$,

where a 0-ball is a point and a ∞ -ball is an open half space. Similarly, an α -ball B is defined as empty if $B \cap M = 0$. For $0 \leq k \leq 2$, a k -simplex σ_T is said to be α -exposed if there is an empty α -ball B with $T = \partial B \cap M$, where ∂B is the boundary of α -ball B . Then, the alpha shape of points set M can be represented by a polytope which consists of all the α -exposed σ_T for $0 \leq k \leq 2$. We can see that the Delaunay triangulation is actually a collection of alpha shapes with $0 \leq \alpha \leq \infty$. The definition of the weighted alpha shape is similar, but now we consider a set of weighted points W . First, we define the idea of orthogonal and suborthogonal: two points P_1 and P_2 with radii r_1 and r_2 are said to be orthogonal if $|P_1 - P_2|^2 = r_1^2 + r_2^2$ while they are defined as suborthogonal if $|P_1 - P_2|^2 > r_1^2 + r_2^2$. For a given value of α , the weighted alpha shape contains all the k -simplex σ_T such that there is an α -ball B orthogonal to the points in σ_T and suborthogonal to the other points in W . Using alpha shape to represent the surface of the protein structure (Figure 2B), one can easily extract the geometric properties of the protein surface.

Several libraries have been established to calculate Delaunay triangulation and alpha shape such as the Qhull [16] and the Computational Geometry Algorithm Library (CGAL) [17]. Qhull is developed by Barber *et al.* based on the Quickhull algorithm which offers a programmable library to compute the convex hull, Delaunay triangulation and Voronoi diagram. CGAL is an open source project which aims to provide efficient and reliable geometric algorithms. Compared with Qhull, CGAL offers a wider range of geometric computation including polygons, convex hull, Delaunay triangulation, Voronoi diagram, alpha shape and mesh generation.

PROTEIN–DNA INTERACTION

Early studies of protein–DNA interactions aim to detect genetic codes in the DNA sequence [18, 19]. However, as more 3D structure of protein–DNA complexes have become available, researchers find it difficult to encode protein–DNA interaction using simple codes due to the various spatial relationships between protein and DNA [20]. Therefore, more attention is being paid to the geometric properties in protein–DNA interactions. Traditional methods in studying protein–DNA interactions usually focus on pairwise atom–atom distance potential. Robertson and Varani [21] apply a distance-

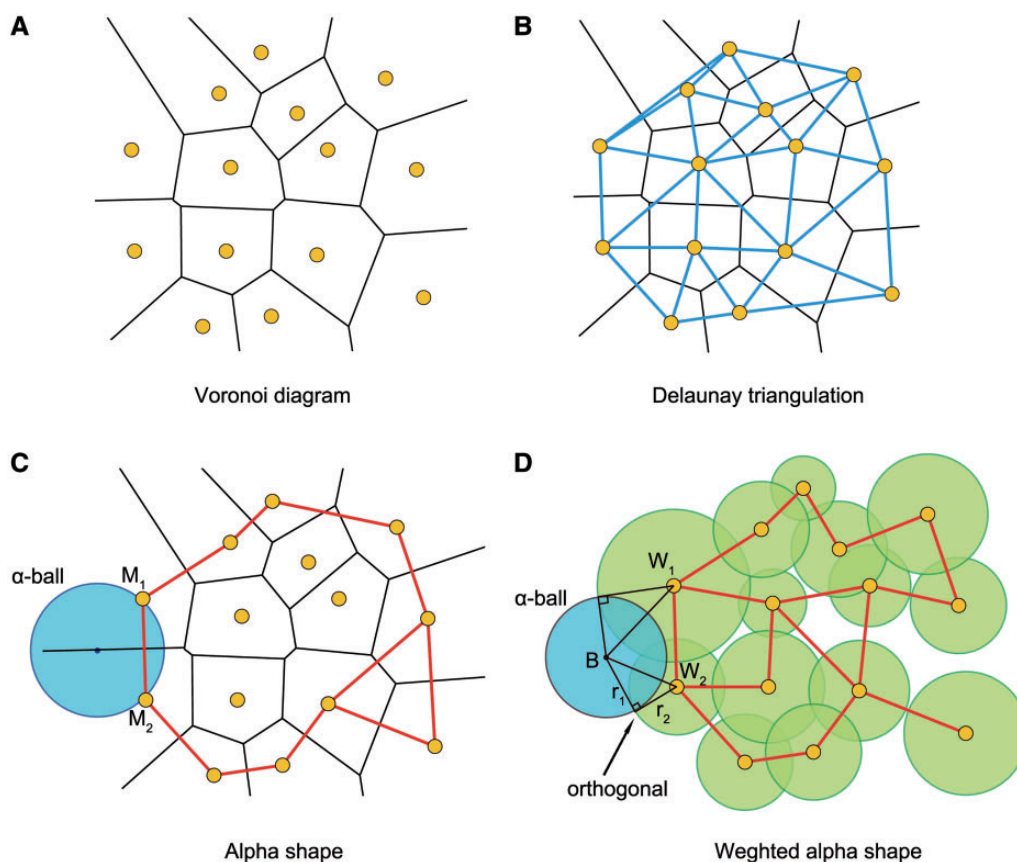


Figure 1: Construction of the Delaunay triangulation and alpha shape. **(A)** Voronoi diagram for a set of points. The Voronoi cell for a point is the locus closer to the given point than the other points. **(B)** Delaunay triangulation is the dual shape of the Voronoi diagram which can be obtained by connecting all the points that share common Voronoi faces. The edges of Delaunay triangulation (bold segments) represent the connection network of the points. **(C)** The basic alpha shape of a set of non-weighted points. The dark coloured sphere is an empty α -ball with its boundary connects M_1 and M_2 . The segment between M_1 and M_2 defines an edge of the alpha shape. **(D)** The light coloured spheres represent a set of weighted points. The dark coloured sphere represents an α -ball B which is orthogonal to W_1 and W_2 . Define r_1 and r_2 as the radii (weight) of point B and W_2 . B is orthogonal to W_2 because the distance between B and W_2 meets the condition $|B - W_2|^2 = r_1^2 + r_2^2$.

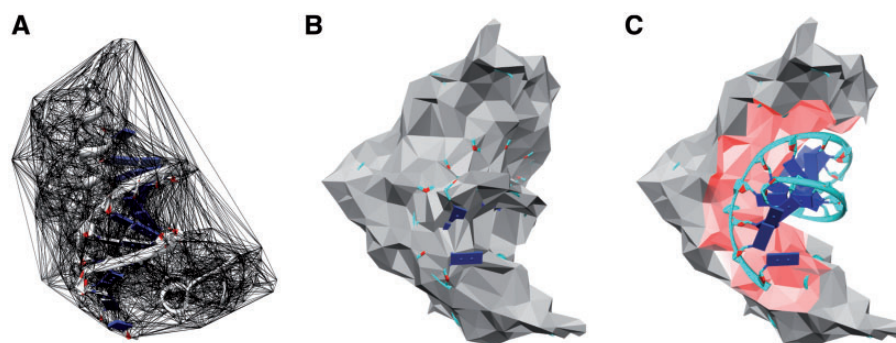


Figure 2: 3D Delaunay triangulation and alpha shape of the molecular structures. **(A)** Delaunay triangulation of a protein–DNA complex. The edges in the Delaunay triangulation represent the atom contacts in the protein–DNA complex. **(B)** Alpha shape of a protein–DNA complex. The alpha shape characterizes the geometric properties of the complex surface. **(C)** Alpha shape of the protein in the complex. The interface of the protein–DNA complex can be obtained by computing the difference between the alpha shapes of the complex and the protein. This figure is adopted from diagrams in [22].

dependent statistical potential method by considering the interface atom–atom distance in a protein–DNA complex, which shows a good performance in discriminating the native protein–DNA complexes from the docking decoy complexes.

Recently, Zhou and Yan [22] apply alpha shape to represent the interface of protein–DNA complexes and develop a new discriminatory function based on surface curvature. They use alpha shape to represent the surface of the protein–DNA complex (Figure 2B) and the corresponding protein (Figure 2C), respectively. The interface of the complex can be obtained by computing the difference between these two alpha shapes. For the interfaces, they apply solid angle to represent the surface curvature of the interface atoms. The discriminatory function is established based on conditional probability:

$$S = - \sum_i \ln \frac{P_c((S_{a_i}, r_i, a_i) | C)}{P((S_{a_i}, r_i, a_i))}. \quad (2)$$

where S_{a_i} stands for the solid angle of the interface atom i , r_i stands for the residue type and a_i stands for the atom type. The solid angle is defined as follow: Let OABC be the vertices of tetrahedron with origin at O subtended by the triangular face ABC and ϕ_{ab} , ϕ_{bc} , ϕ_{ac} be the dihedral angle between OAC and OBC, OAB and OAC, and OAB and OBC, respectively. The solid angle of O can then be calculated as $\Omega = \phi_{ab} + \phi_{bc} + \phi_{ac} - \pi$. The solid angle of interface atom i can be obtained by summing up the solid angle values of all tetrahedrons each of which has a vertex at atom i in the alpha shape. $P_c((S_{a_i}, r_i, a_i) | C)$ represents the low count corrected [23] probability of the correct protein–DNA interface having a set of features (S_{a_i}, r_i, a_i) while $P((S_{a_i}, r_i, a_i))$ stands for the probability of any structure having a set of features (S_{a_i}, r_i, a_i) .

The results [22] show that the curvature–dependent potential (average z -score -7.38) outperforms the distance–dependent potential (average z -score -6.8) in discriminating the native protein–DNA complexes from the low RMSD decoy complexes. They also demonstrate that the curvature–dependent method shows good performance (average z -score -8.17) in discriminating the native complexes from the high surface–complementarity scored [24] decoy complexes. A further study of the interface features of the protein–DNA complex by Zhou and Yan [25] using alpha shape and support vector machine also shows good results in predicting unbound DNA-binding proteins. These studies indicate that the

surface curvature plays an important role in protein–DNA interaction.

Using alpha shape and Delaunay triangulation, one can make a concise reconstruction of the protein–DNA structure. They are not only useful in representing the geometric properties of the structure but also powerful tools in representing the relationships among atoms or residues in the protein–DNA structure. Mathe *et al.* [26] use Delaunay triangulation to represent the four neighboring residues in the protein structure and develop a four-body potential to predict the transactivation activity of missense mutations in the DNA-binding domain of tumor suppressor TP53. In their work, they define the centers of mass of the residue side chains as the vertexes in Delaunay triangulation and represent the four-nearest-neighbor residues using the quadruplets in the Delaunay triangulation. The four-body potential can be represented as follows:

$$q_{ijkl} = \log \frac{f_{ijkl}}{p_{ijkl}}. \quad (3)$$

where i , j , k and l are the four amino acids that compose the quadruplet; f_{ijkl} is the frequency of the quadruplet in the training set and p_{ijkl} is the frequency of random occurrence of the quadruplet.

PROTEIN–PROTEIN INTERACTION

A variety of methods have been developed for the prediction of protein–protein interaction, mostly based on feature extraction and machine learning methods to perform prediction [27]. Bradford and Westhead [28] extract a set of features including surface shape, conservation, electrostatic potential, hydrophobicity, residue interface propensity and solvent accessible surface area (ASA) to identify protein–protein binding sites based on SVM. Porollo and Meller [29] propose the use of the difference in observed and predicted relative solvent accessibility as a feature and apply different machine learning methods to predict protein–protein interactions. Therefore, the key issue in the study of protein–protein interaction is actually the problem of protein–protein surface or interface representation and feature extraction.

Ban *et al.* [30] use Delaunay triangulation to extract the interface of protein–protein complexes and define different level-of-focus hierarchy of the interface. Using the level-of-focus hierarchy of the interface, they develop a function to distinguish hot-spot residues from neutral residues. The hot-spot residues

are a small number of residues that contribute a large amount of binding energy in the interaction which are usually located in the interface [31]. The function is defined as follows:

$$h(R) = \sum_{i=0}^k w_i \cdot \text{area}(p_i). \quad (4)$$

where R is a residue, p_i is a polygon in the interface generated by the side-chain atoms in R , w_i is the fraction of the interface surface that belongs to R before p_i is removed. The function shows a better performance than the buried surface area method and show equal performance with the physical model proposed by Kortemme and Baker [32] in predicting the hot-spot and neutral residues (Table 1).

Similar to the Delaunay triangulation, alpha shape also provides a good representation of the protein surface. Albou *et al.* [33] apply alpha shape to reconstruct the surface of proteins and use the connectivity provided by the alpha shape to extract protein binding sites. The method shows a good signal-to-noise ratio (SNR) with a value of 4.9 and a maximum overlap of the binding site residue with a value of 65.9%, which outperforms a previous approach proposed by Jones and Thornton [34] with SNR 2.2 and maximum overlap 56%. They also develop a local surface curvature $C(a)$ to characterize the surface atoms of the protein:

$$C(a) = \sum_{i \in \text{surfacepatch}} \frac{\Omega(i)}{d(a,i)}. \quad (5)$$

where $d(a,i)$ is the distance over the surface in the alpha shape between atom a and i , (i) is the solid angle in the alpha shape of atom i . The local surface curvature of a residue is calculated as the means of $C(a)$ for all the atoms in the residue. This feature shows good correlation with the ASA [35] by Pearson correlation coefficient of 0.86 for atoms and 0.89 for residues, which is useful to analyze the protein-protein interface properties.

Table 1: Comparison of accuracy for different protein binding site prediction methods^a

Method	Hot-spot residues identified (%)	Neutral residues identified (%)
Ban <i>et al.</i>	72.4	72.6
BAS	65.0	64.3
Kortemme and Baker	79.0	68.0

^aThe results are adopted from [30].

Protein-protein interaction can also be characterized by the contact patterns in the interface. Khashan *et al.* [36] applies almost Delaunay triangulation to analyze the contacts of the residues in the interface of protein-protein complexes. Almost Delaunay triangulation is developed by Bandyopadhyay *et al.* [37] to solve the problem of imprecision of atom coordinates in molecular structures. A robust structure can be obtained by almost Delaunay triangulation which shows tolerance to a small change of the coordinates in the molecular structure. According to this property, protein-protein contact patterns in different complexes can be revealed by the subgraphs in the almost Delaunay triangulation. Khashan *et al.* use the side chain centroids to represent the protein-protein complex in residue level and calculate the interaction residue network using almost Delaunay triangulation. These centroids and the edges connecting them form the contact patterns in protein-protein interaction.

PROTEIN-LIGAND INTERACTION

A major issue in the study of protein-ligand interaction is ligand binding sites prediction. The most common approach is to detect the pockets or cavities in the protein surface, which is also known as the geometric-based method. Depending on the types of algorithms used, the geometric-based methods can be divided into three categories: grid scanning, probe sphere and alpha shape. PocketPicker [38] and LIGSITE^{csc} [39] are widely used grid scanning methods which map the protein onto a 3D grid and scan for pockets from different directions (Figure 3A). PASS [40] is a representative probe sphere method which uses probe spheres to fill pockets layer by layer and select active site points to represent ligand binding sites (Figure 3B). A typical alpha shape method CAST was developed by Liang *et al.* [41] using alpha shape to represent the protein surface and they apply discrete triangles flow to detect the pockets (Figure 3C). Compared to the other methods, alpha shape-based methods rely on the original geometric information from the protein structure and do not rely on other information such as the grid size and searching directions.

Recently, Guilloux *et al.* [42] used the alpha sphere for the detection of pockets (Figure 3D) in the protein surface and developed Fpocket. Alpha sphere is the same concept as α -ball mentioned before. In 3D space, an alpha sphere contains four

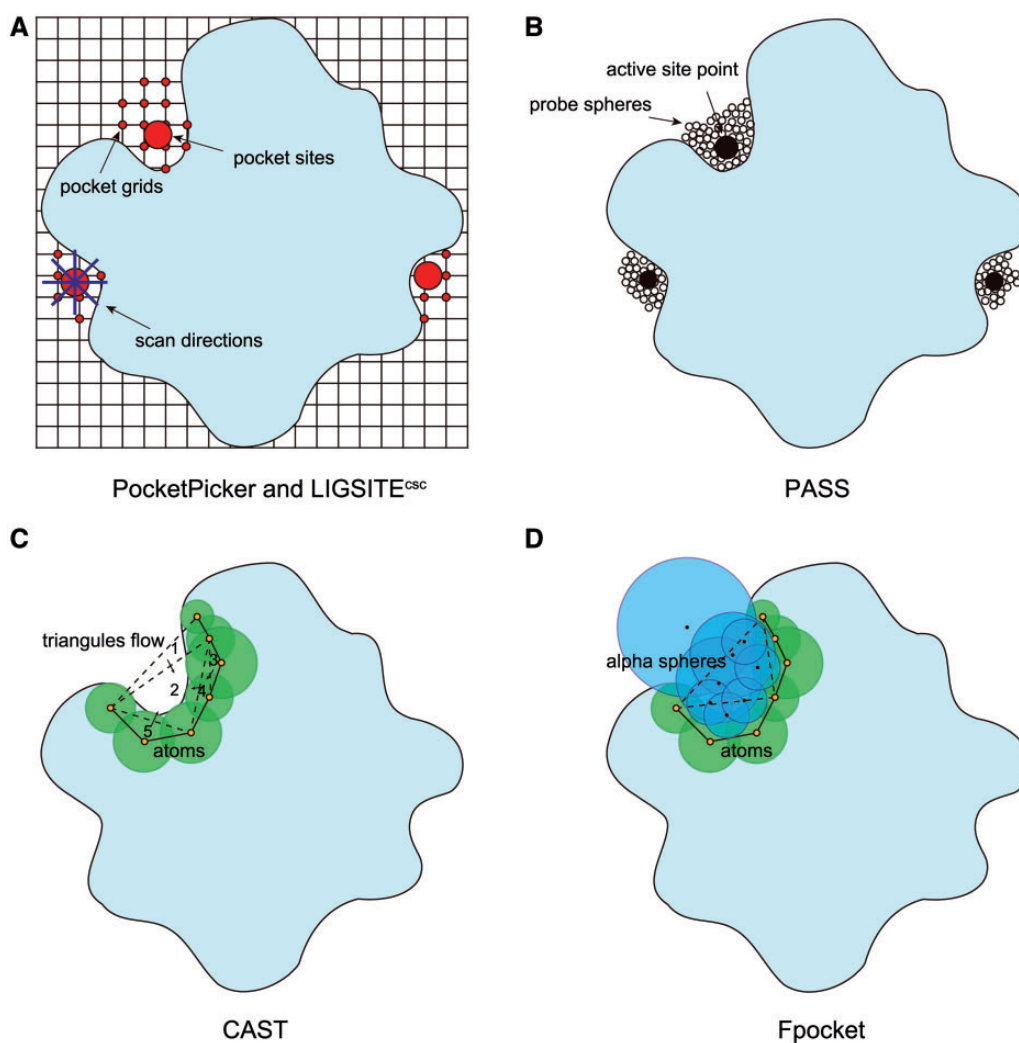


Figure 3: Ligand binding sites detection methods. **(A)** PocketPicker and LIGSITE^{csc} are based on grid scanning. The protein is mapped to a grid system and the pockets can be obtained by scanning the grid points near the protein surface in different directions. **(B)** PASS is based on probe sphere method. This method uses probe spheres to fill the pockets layer by layer and active site points to represent pockets. **(C)** CAST applies alpha shape to represent the protein structure and uses triangles flow to identify empty triangles (triangle 2 in dotted segments) to represent pockets. **(D)** Fpocket uses alpha spheres with different radii to represent different roughness of the protein surface. The pockets can be identified by clustering the alpha spheres with proper radii.

atoms on the protein surface which have equal distance to the alpha sphere center. Therefore, the radius of an alpha sphere reflects the local curvature defined by the four atoms that it contains. The basic idea of Fpocket is to find alpha spheres which have proper radii to represent the clefts and cavities on the protein surface. The process for detecting the ligand binding sites in Fpocket is shown below:

- (1) Identify proper alpha spheres using maximum and minimum radius thresholds.
- (2) Clustering alpha spheres from step 1 to obtain potential ligand binding sites.

- (3) Rank the potential ligand binding sites according to their properties such as number of alpha spheres, hydrophobic density and polarity.

The results show that Fpocket outperforms the previous methods in predicting ligand binding sites in both bound and unbound protein structures (Table 2). Fpocket shows a very high success rate in detecting the true ligand binding sites within its top three predicted binding sites, which indicates good prediction power.

Prediction of potential binding sites is only the first step in protein–ligand interaction study. We are

more interested in the molecular function and properties of these binding sites. An extended version of CAST, established by Dundas *et al.* [43] is called CASTp. CASTp further examines the output residues of CAST and annotates the function information of these residues from different databases [44–46]. Tseng and Li introduce a concept of split pocket to identify the functional surface of the protein. They use the same strategy in CAST to compute the pockets in the protein surface but with flexible probe (solvent) radii. In the weighted alpha shape model, if we consider solvent molecules, the corresponding radius for an atom i is $r_i + r_s$ where r_i is the radius of the atom and r_s is the radius of the solvent molecule. Tseng and Li assigned different probe radii according to the physicochemical properties [47, 48] of the atoms. Using split pocket, they found that function surfaces tend to be better conserved in evolution than the other regions of the protein. This result demonstrates that shape analysis is useful to infer protein functions and classify different protein families.

In protein–ligand interaction, hydrogen bonding plays a significant role and has been proven to be the major force maintaining the stability of protein–ligand complexes [49]. Recently, Zhou and Yan [50] used alpha shape to reveal geometric patterns of hydrogen bonding in protein–ligand interaction. They analyze the curvature of the donor, hydrogen and acceptor atoms in 1072 hydrogen bonds which are represented by the solid angles extracted from the alpha shape of the protein–ligand complexes. The hydrogen bonds can be classified into four types [50]: $D_P H_S A_{LS}$, $D_{PS} H_S A_{LS}$, $D_L H_S A_{PS}$ and $D_{LS} H_S A_{PS}$, where D, H and A stand for donor, hydrogen and acceptor atoms, respectively, subscripts P and L represent protein and ligand, respectively, and subscript S indicates that the atom is on the

surface of the alpha shape. Results show that hydrogen and acceptor atoms are spatially matched in $D_P H_S A_{LS}$ and $D_L H_S A_{PS}$ hydrogen bonds while donor and acceptor atoms are spatially matched in $D_{PS} H_S A_{LS}$ and $D_{LS} H_S A_{PS}$ hydrogen bonds. They show that the spatially matched hydrogen bonds have average association energy of -10.65 kcal/mol while the unmatched ones -9.97 kcal/mol. That means that spatially matched hydrogen bonds require larger energy in the association process than the unmatched ones. This result demonstrates that the spatially matched hydrogen bonds show larger bond strength, which indicates these hydrogen bonds play a more important role than the others in protein–ligand interaction.

OTHER APPLICATIONS

Alpha shape and Delaunay triangulation are not only useful in the study of protein-related interactions but also applicable to other studies of protein structures. Here, we review some promising applications of alpha shape and Delaunay triangulation to the studies of protein packing, structural alignment and channel detection in protein structure.

Protein packing

Distance-dependent potential is a useful method in studying protein packing. Samudrala and Moult [51] propose the use of pairwise atom–atom distance potential and develop an all-atom distance-dependent discriminatory function for the prediction of protein structure. However, there are disadvantages for considering all atom pairs when analyzing a large protein structure or a large scale of protein structure data. It is computationally expensive and it may contain redundant information. In the past, one way to reduce the atom pair data is to use a residue pair which considered only C_α or C_β atom contacts instead of all atom contacts in a protein structure. However, these approaches were mostly based on distance cut off criteria which were proven to contain atom pairs with no significant physical interaction [52, 53]. Li *et al.* [54] introduce the idea of alpha contact instead of distance cut off to analyze residue contact in protein structure. They use the edges in alpha shape to identify the contact of atoms within a protein structure, which show good performance in discriminating native structures from decoys. Later, Zomorodian *et al.* [55] demonstrate that using alpha shape to filter the set of all residue

Table 2: Percentile success rates in ligand binding sites prediction for 48 unbound/bound protein structures^a

Method	Top 1		Top 3	
	Unbound	Bound	Unbound	Bound
Fpocket	69	83	94	92
PocketPicker	69	72	85	85
LIGSITE ^{csc}	71	79	–	–
PASS	60	63	71	81
CAST	58	67	75	83

^aThe results are adopted from [42].

pairs leads to a significant reduction of the number of pairs without loss of information and discrimination power.

Another challenge in protein packing is the prediction of the tertiary structure. While previous studies focus on the prediction of protein secondary structure. Singh *et al.* [12] first establish a four-body potential based on Delaunay triangulation. They demonstrate that the tetrahedral motifs found in their work show a strong correlation with the protein secondary structure. Recently, Day *et al.* [56] represented the contact in the protein structures with Delaunay triangulation and found tertiary motifs by clustering the relative packing group.

Structural alignment

Structural alignment is an important tool to study the homology between different proteins and reveal biological functions. Ilyin *et al.* [57] apply Delaunay triangulation in structural alignment and have developed TOPOFIT. TOPOFIT is based on the superimposition of the Delaunay triangulation patterns defined by the contact structures [58]. Using a maximum topological match threshold, TOPOFIT is able to obtain a good balance between low RMSD and large alignment length. However, similar to protein packing, structural alignment faces a computational complexity problem. In order to reduce the computational complexity, Roach *et al.* [59] propose the use of the edges in Delaunay triangulation to represent the relations among the C_α atoms. They divide the edges into three classes according to their length, and denote short, intermediate and long interactions. Using this transformation, the structure alignment between two 3D proteins is changed into a 1D string sequence alignment problem. They show that the sequential representation greatly reduces the computational complexity with little loss of information.

Channel detection

The first algorithm HOLE for the study of the channel inside a protein was developed by Smart *et al.* [60]. HOLE starts from a user defined point and finds the route by squeezing a ball through the channel. Petrek *et al.* [61, 62] develop CAVER and MOLE, which use convex hull and Voronoi diagram to represent the protein structure, respectively. Both CAVER and MOLE apply Dijkstra's graph search algorithm to detect channels inside a protein. Compared to CAVER, MOLE is able to give

smoother channel profiles and more precisely localized channel bottlenecks while significantly reducing the errors. Recently, Yaffe *et al.* [63] propose the application of alpha shape and medial axis to study the protein structure and develop MolAxis to identify the channels in proteins. They use alpha shape to approximate a subset of the medial axis for the complement of the protein. Using this subset of medial axis, the channels are detected by the corridors which are probable routes taken by small molecules passing through channels. They demonstrate that MolAxis outperforms the previous algorithms in the following aspects: the detected channels from MolAxis show a better balance between length and clearance due to the global optimization approach of the pathways; using fewer vertices for approximation, MolAxis requires a much shorter running time than the previous algorithms; MolAxis provides an adjustable resolution which guarantees smaller errors.

DISCUSSION

The state-of-the-art applications in the studies of protein-related interactions and protein structure analysis demonstrate the advantages of alpha shape and Delaunay triangulation in solving these structural biology problems. Alpha shape and Delaunay triangulation are robust techniques for reconstructing the molecular surface and characterizing the curvature of surface atoms, representing the atom contacts inside or between molecules and extracting hierarchically structural information of the molecules. Although alpha shape and Delaunay triangulation are useful in tackling many structural biology problems, they also have disadvantages in some studies such as molecular volume and surface area computation. These disadvantages are caused by the structural composition of the alpha shape or the Delaunay triangulation. In contrast, The Voronoi diagram (dual shape of the Delaunay triangulation) is a good model for molecular volume computation [15]. The Voronoi diagram provides an accurate measure of molecular volume using Voronoi cells. Compared with alpha shape and Delaunay triangulation, a better model in computation of molecular surface area is the Connolly surface [64] which is also known as solvent accessible surface because the Connolly surface model offers a smoother representation of the molecular surface that it can estimate the molecular surface area precisely. However, as

discussed earlier, the alpha shape and Delaunay triangulation can be used for characterizing the curvature of molecular surface atoms while it would be difficult to do so with the Voronoi diagram or the Connolly surface. The aforementioned discussion shows that the advantages and disadvantages of different geometric models are actually complementary to each other. In other words, taking advantages of one model may overcome the disadvantages of another model. For example, if we need the information of both surface atom curvature and molecular volume, we can employ both alpha shape and Voronoi diagram rather than using either of them alone.

Some biomolecular interaction-related problems cannot be solved using the geometric methods discussed earlier. They include the discovery of evolutionary conservation and free energy calculation, which require techniques of sequence alignment and molecular dynamics. BLAST [65] is a well-known sequence alignment method capable of searching for conserved domain in both DNA and protein sequences. A widely employed software system for molecular dynamics is Amber [66] which makes use of force fields for protein simulation and free energy calculation. For some protein-related studies, a combination of different methods is needed. For example, Capra *et al.* [67] integrate sequence- and geometric-based methods in the detection of protein ligand binding sites. Schmidtke *et al.* [68] improve Fpocket with molecular dynamics and develop MDpocket to identify cavities in protein structures. Therefore, it is useful to combine geometric approaches with sequence- and energy-based methods for solving difficult problems.

CONCLUSION

With the increased computer power, it is now possible to analyze large-scale 3D molecular structures. Alpha shape and Delaunay triangulation provide robust geometric models for such analysis. Alpha shape and Delaunay triangulation can be applied to the characterization of surface curvatures and atom contacts in protein structures and many studies of protein-related interactions. In this review, we have summarized the most recent development of alpha shape and Delaunay triangulation in the studies of protein–DNA, protein–protein, protein–ligand interactions and protein structure analysis. However, many structural biology problems are still unsolved

and we expect that alpha shape and Delaunay triangulation can be useful in tackling these problems. For example, much more research is needed to study protein flexibilities and conformational space. Alpha shape and Delaunay triangulation may provide useful mathematical models for these tasks.

Key Points

- As more 3D protein structures become available, there is a growing demand for 3D geometric models to represent and study protein-related interactions.
- Alpha shape and Delaunay triangulation have been applied successfully in the studies of protein–DNA, protein–protein, protein–ligand interactions and protein structural analysis.
- Alpha shape and Delaunay triangulation can provide useful mathematical models to tackle unsolved structural biology problems, such as the investigation of protein flexibilities and conformational space.

FUNDING

This work is supported by the Hong Kong Research Grants Council [Project CityU 123809].

References

1. Cartharius K, Frech K, Grote K, *et al.* MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 2005;**21**:2933–42.
2. Johnson PF, Mcknight SL. Eukaryotic transcriptional regulatory proteins. *Annu Rev Biochem* 1989;**58**:799–839.
3. Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, *et al.* Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem* 2004;**73**:39–85.
4. Stormo GD. DNA binding sites: representation and discovery. *Bioinformatics* 2000;**16**:16–23.
5. Burgoyne NJ, Jackson RM. Predicting protein interaction sites: binding hot-spots in protein–protein and protein–ligand interfaces. *Bioinformatics* 2006;**22**:1335–42.
6. Russell RB, Aloy P. Targeting and tinkering with interaction networks. *Nat Chem Biol* 2008;**4**:666–73.
7. Stelzl U, Worm U, Lalowski M, *et al.* A human protein–protein interaction network: a resource for annotating the proteome. *Cell* 2005;**122**:957–68.
8. Amzel LM. Structure-based drug design. *Curr Opin Biotechnol* 1998;**9**:366–9.
9. DeWeese-Scott C, Moutl J. Molecular modeling of protein function regions. *Proteins* 2004;**55**:942–61.
10. Edelsbrunner H, Mucke EP. Three-dimensional alpha-shapes. *ACM Trans Graph* 1994;**13**:43–72.
11. Delaunay B. Sur la sphère vide, *Izvestia Akademii Nauk SSSR. Otdelenie Matematicheskii i Estestvennyka Nauk* 1934;**7**: 793–800.
12. Singh RK, Tropsha A, Vaisman II. Delaunay tessellation of proteins: Four body nearest-neighbor propensities of amino acid residues. *J Comput Biol* 1996;**3**:213–21.

13. Liang J, Edelsbrunner H, Fu P, *et al.* Analytical shape computation of macromolecules: I. Molecular area and volume through alpha shape. *Proteins* 1998;**33**:1–17.
14. Liang J, Edelsbrunner H, Fu P, *et al.* Analytical shape computation of macromolecules: II. Inaccessible cavities in proteins. *Proteins* 1998;**33**:18–29.
15. Poupon A. Voronoi and Voronoi-related tessellations in studies of protein structure and interaction. *Curr Opin Struct Biol* 2004;**14**:233–41.
16. Barber CB, Dobkin DP, Huhdanpaa H. The Quickhull algorithm for convex hulls. *ACM Trans Math Software* 1996;**22**:469–83.
17. CGAL. *Computational Geometry Algorithms Library*. <http://www.cgal.org>.
18. Fickett JW. Recognition of protein coding regions in DNA sequences. *Nucleic Acids Res* 1982;**10**:5303.
19. Schneider TD, Stormo GD, Gold L, *et al.* Information content of binding sites on nucleotide sequences. *J Mol Biol* 1986;**188**:415.
20. Pabo CO, Nekludova L. Geometric analysis and comparison of protein–DNA interfaces: why is there no simple code for recognition? *J Mol Biol* 2000;**301**:597–624.
21. Robertson TA, Varani G. An all-atom, distance-dependent scoring function for the prediction of protein–DNA interactions from structure. *Proteins* 2007;**66**:359–74.
22. Zhou W, Yan H. A Discriminatory function for prediction of protein–DNA interactions based on alpha shape modeling. *Bioinformatics* 2010;**26**:2541–48.
23. Sippl MJ. Calculation of conformational ensembles from potentials of mean force—an approach to the knowledge-based prediction of local structures in globular-proteins. *J Mol Biol* 1990;**213**:859–83.
24. Gabb HA, Jackson RM, Sternberg MJE. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* 1997;**272**:106–20.
25. Zhou W, Yan H. Prediction of DNA-binding protein based on statistical and geometric features and support vector machines. *Proteome Sci* 2011;**9**:S1.
26. Mathe E, Olivier M, Kato S, *et al.* Predicting the transactivation activity of p53 missense mutants using a four-body potential score derived from Delaunay tessellations. *Hum Mutat* 2006;**27**:163–72.
27. Ezkurdia L, Bartoli L, Fariselli P, *et al.* Progress and challenges in predicting protein–protein interaction sites. *Brief Bioinformatics* 2009;**10**:233–46.
28. Bradford JR, Westhead DR. Improved prediction of protein–protein binding sites using a support vector machines approach. *Bioinformatics* 2005;**21**:1487–94.
29. Porollo A, Meller J. Prediction-based fingerprints of protein–protein interactions. *Proteins* 2007;**66**:630–45.
30. Ban YEA, Edelsbrunner H, Rudolph J. Interface surfaces for protein–protein complexes. *J ACM* 2006;**53**:361–78.
31. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *J Mol Biol* 1998;**280**:1–9.
32. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein–protein complexes. *Proc Natl Acad Sci USA* 2002;**99**:14116–21.
33. Albou LP, Schwarz B, Poch O, *et al.* Defining and characterizing protein surface using alpha shapes. *Proteins* 2009;**76**:1–12.
34. Jones S, Thornton JM. Prediction of protein–protein interaction sites using patch analysis. *J Mol Biol* 1997;**272**:133–43.
35. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 1971;**55**:379–400.
36. Khashan R, Zheng W, Tropsha A. Scoring protein interaction decoys using exposed residues (SPIDER): a novel multibody interaction scoring function based on frequent geometric patterns of interfacial residues. *Proteins* 2012;**80**:2207–17.
37. Bandyopadhyay D, Snoeyink J. Almost-Delaunay simplices: Robust neighbor relations for imprecise 3D points using CGAL. *Comput GeometryTheoryAppl* 2007;**38**:4–15.
38. Weisel M, Proschak E, Schneider G. PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chem Cent J* 2007;**1**:7.
39. Huang BD, Schroeder M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct Biol* 2006;**6**:19.
40. Brady GP, Stouten PFW. Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des* 2000;**14**:383–401.
41. Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci* 1998;**7**:1884–97.
42. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* 2009;**10**:168.
43. Dundas J, Ouyang Z, Tseng J, *et al.* CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 2006;**34**:W116–8.
44. Berman HM, Westbrook J, Feng Z, *et al.* The Protein Data Bank. *Nucleic Acids Res* 2000;**28**:235–42.
45. Gasteiger E, Gattiker A, Hoogland C, *et al.* ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 2003;**31**:3784–8.
46. McKusick VA. Mendelian inheritance in man and its online version, OMIM. *AmJ Hum Genet* 2007;**80**:588–604.
47. Gerstein M, Tsai J, Levitt M. The volume of atoms on the protein surface—calculated from simulation, using Voronoi polyhedra. *J Mol Biol* 1995;**249**:955–66.
48. Li AJ, Nussinov R. A set of van der Waals and Coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins* 1998;**32**:111–27.
49. Babine RE, Bender SL. Molecular recognition of protein–ligand complexes: applications to drug design. *Chem Rev* 1997;**97**:1359–472.
50. Zhou W, Yan H, Hao Q. Analysis of surface structures of hydrogen bonding in protein–ligand interaction using alpha shape model. *Chem Phys Lett* 2012;**545**:125–31.
51. Samudrala R, Moulton J. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. *J Mol Biol* 1998;**275**:895–916.
52. Taylor WR. Multiple sequence threading: An analysis of alignment quality and stability. *J Mol Biol* 1997;**269**:902–43.
53. Bienkowska JR, Rogers RG, Smith TF. Filtered neighbors threading. *Proteins* 1999;**37**:346–59.

54. Li X, Hu CY, Liang J. Simplicial edge representation of protein structures and alpha contact potential with confidence measure. *Proteins* 2003;**53**:792–805.
55. Zomorodian A, Guibas L, Koehl P. Geometric filtering of pairwise atomic interactions applied to the design of efficient statistical potentials. *Comput Aided Geometric Des* 2006;**23**:531–44.
56. Day R, Lennox KP, Dahl DB, *et al.* Characterizing the regularity of tetrahedral packing motifs in protein tertiary structure. *Bioinformatics* 2010;**26**:3059–66.
57. Ilyin VA, Abyzov A, Leslin CM. Structural alignment of proteins by a novel TOPOFIT method, as a superimposition of common volumes at a topomax point. *Protein Sci* 2004;**13**:1865–74.
58. Chan HS, Dill KA. Origins of structure in globular-proteins. *Proc Natl Acad Sci USA* 1990;**87**:6388–92.
59. Roach J, Sharma S, Kapustina M, *et al.* Structure alignment via delaunay tetrahedralization. *Proteins* 2005;**60**:66–81.
60. Smart OS, Goodfellow JM, Wallace BA. The pore dimensions of gramicidin-A. *Biophys J* 1993;**65**:2455–60.
61. Petrek M, Otyepka M, Banas P, *et al.* CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics* 2006;**7**:316.
62. Petrek M, Kosinova P, Koca J, *et al.* MOLE: A Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure* 2007;**15**:1357–63.
63. Yaffè E, Fishelovitch D, Wolfson HJ, *et al.* MolAxis: Efficient and accurate identification of channels in macromolecules. *Proteins* 2008;**73**:72–86.
64. Connolly ML. Analytical molecular-surface calculation. *J Appl Crystallogr* 1983;**16**:548–58.
65. Johnson M, Zaretskaya I, Raytselis Y, *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res* 2008;**36**:W5–9.
66. Case DA, Cheatham TE, Darden T, *et al.* The Amber biomolecular simulation programs. *J Comput Chem* 2005;**26**:1668–88.
67. Capra JA, Laskowski RA, Thornton JM, *et al.* Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. *PLoS Comput Biol* 2009;**5**:e1000585.
68. Schmidtke P, Bidon-Chanal A, Luque FJ, *et al.* MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* 2011;**27**:3276–85.