BIOINFORMATICS

Vol. 18 no. 10 2002
Pages 1365–1373

Quantification of protein surfaces, volumes and atom–atom contacts using a constrained Voronoi procedure

B. J. McConkey, V. Sobolev and M. Edelman*

Department of Plant Sciences, Weizmann Institute of Science, Rehovot 76100, Israel

Received on December 20, 2001; revised on March 25, 2002; accepted on March 27, 2002

ABSTRACT

Motivation: Geometric representations of proteins and ligands, including atom volumes, atom–atom contacts and solvent accessible surfaces, can be used to characterize interactions between and within proteins, ligands and solvent. Voronoi algorithms permit quantification of these properties by dividing structures into cells with a one-to-one correspondence with constituent atoms. As there is no generally accepted measure of atom–atom contacts, a continuous analytical representation of inter-atomic contacts will be useful. Improved geometric algorithms will also be helpful in increasing the speed and accuracy of iterative modeling algorithms.

Results: We present computational methods based on the Voronoi procedure that provide rapid and exact solutions to solvent accessible surfaces, volumes, and atom contacts within macromolecules. Furthermore, we define a measure of atom–atom contact that is consistent with the calculation of solvent accessible surfaces, allowing the integration of solvent accessibility and inter-atomic contacts into a continuous measure. The speed and accuracy of the algorithm is compared to existing methods for calculating solvent accessible surfaces and volumes. The presented algorithm has a reduced execution time and greater accuracy compared to numerical and approximate analytical surface calculation algorithms, and a reduced execution time and similar accuracy to existing Voronoi procedures for calculating atomic surfaces and volumes.

Availability: Source code and executable programs are available from the authors.

Contact: brendan.mcconkey@weizmann.ac.il

INTRODUCTION

Molecular modeling algorithms seek to predict the structure of biomolecules and binding affinities of biomolecular interactions. These algorithms rely on various chemical and geometric approximations to model molecular structures and processes. An efficient method for characterizing geometrical relations between a set of points is the Voronoi tessellation (Voronoi, 1908), which subdivides space into regions associated with each point. The Voronoi algorithm determines contact planes between points, and the intersection of the contact planes surrounding a given point defines a polyhedron (a convex hull). The faces of the polyhedron specify the nearest-neighboring points and contacts between associated regions. A group of mutually consistent convex hulls partitioning space around a set of points defines a Voronoi diagram for the point set.

When applied to a macromolecule, the Voronoi procedure generates a set of polyhedra with a one-to-one correspondence with macromolecule atoms. These polyhedra may be used directly to quantify atom volumes, and indirectly to quantify solvent accessible surfaces and atom–atom contacts. In the original procedure, dividing planes separating polyhedra are placed midway between each contacting atom pair. This provides a mathematically exact solution and accounts for all space within the point set; however, when applied to protein structures it does not account for differences in atomic radii. Several modifications have been presented to account for different atom sizes. Richards' Method B (Lee and Richards, 1971) uses a ratio of the atomic radii for placing the dividing plane. This accommodates the different sizes of atoms, but also introduces a vertex error such that small volumes within the protein structure are not assigned. Another choice for the dividing plane is the radical plane (Gellatly and Finney, 1982), which accounts for varying atomic radii and is mathematically exact. Non-planar boundaries have also been suggested to account for differences in atom sizes (Gerstein *et al.*, 1995; Goede *et al.*, 1997).

Voronoi algorithms have been used previously to determine atomic volumes in liquids (Bernal and Finney, 1967), packing efficiencies within proteins and crystals (Richards, 1974; Dodd and Theodorou, 1991; Gerstein *et al.*, 1995; Tsai *et al.*, 1999) and atomic surfaces (Liang *et al.*, 1998). The surfaces of Voronoi polyhedra have been used to determine amino-acid contacts within a protein,

*To whom correspondence should be addressed.

and to develop scoring functions for proteins based upon residue–residue contacts (Zimmer *et al.*, 1998). However, the direct use of polyhedron surfaces for estimating inter-atomic contacts is limited, as there is no clear dependence of the contact area on separation distance, and the total surface area of a Voronoi polyhedron for a given atom type is not a constant.

A method frequently used to create Voronoi polyhedra constructs the hull from a set of points in dual space (Preparata and Shamos, 1985; O'Rourke, 1994), obtained by transforming the contact plane equations into a mathematically equivalent set of points. However, this cannot be used directly with all definitions of the contact plane, as the transformation of the contact planes is only valid if the atom center always falls within the convex hull (Preparata and Shamos, 1985). A Voronoi tessellation of a point set in three dimensions may also be obtained from a Delaunay tetrahedrization (Delaunay, 1934), the mathematical dual of the full Voronoi tessellation.

Methods for estimating the solvent accessible surface (SAS) (Lee and Richards, 1971) have used a finite set of points distributed on a sphere (Shrake and Rupley, 1973; Connolly, 1983; Richards, 1985; Eisenhaber *et al.*, 1995), arcs generated from the intersection of a sphere surface with a set of parallel planes (Richards, 1985; Gerstein, 1992), or exact analytical solutions (Richmond, 1984; Fraczkiewicz and Braun, 1998; Liang *et al.*, 1998). Given the appropriate choice of contact planes, the limiting vertices of the SAS lie on the edges of a Voronoi diagram and the SAS may be calculated from the intersection of Voronoi polyhedra with spheres surrounding each atom.

A computational problem related to calculation of the SAS is the estimation of atom–atom contact areas. Given two atoms A and B in close proximity, the spheres used in calculating the SAS may also be used to estimate an inter-atomic contact area. The region of overlap between the two spheres defines a contact area, where the surface of sphere A falling within sphere B defines the contact area of B with A (Sobolev and Edelman, 1995). If an area of contact may be assigned to more than one atom, the region can be partitioned. One method is to assign areas within the region of overlap to the closest contacting atom (Sobolev and Edelman, 1995), or alternately a weighting function may be used to divide the area according to inter-atomic distances between atoms (Delarue and Koehl, 1995). An advantage in using atom–atom contacts is that they may be readily determined from solved protein structures, permitting statistical analysis and comparison between known structures and models, whereas explicit energetic terms cannot be empirically determined for isolated atom pairs within proteins. A linear relation between free energy in protein structures and estimated atom–atom contact areas has been previously identified, and atom–atom contacts have been shown to reproduce the behavior of non-bonded interactions (de la Cruz and Calvo, 2001). Additionally, contact area based methods can be used to evaluate the accuracy of protein models (Delarue and Koehl, 1995; Abagyan and Totrov, 1997). Each of these methods use empirical solutions for the atom–atom contact area, and methods such as these may be aided by use of an analytical solution. We introduce a measure based on the Voronoi tessellation to provide a means to quantify inter-atomic contacts within macromolecules.

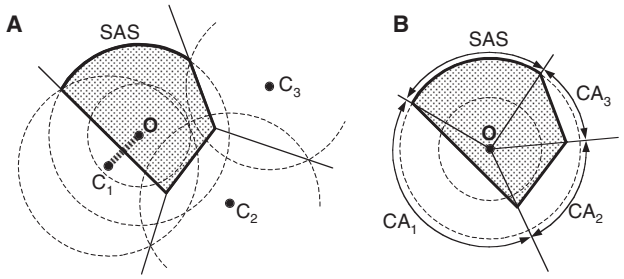

Fig. 1. Definition of Voronoi cells and atom–atom contacts. Diagrams are shown in two dimensions for clarity. **A.** The Voronoi cell (shaded) for atom O, using the expanded radical plane definition. Contact planes are determined from the intersection of expanded atom spheres, radius $r = r_{vdw} + r_{solvent}$. Regions of the polyhedron beyond the expanded sphere radius are truncated, and the intersection of the expanded sphere with the polyhedron defines the SAS. **B.** Atom–atom contact areas (CA_1, CA_2, and CA_3) of atoms C_1, C_2, and C_3 on center atom O. In three dimensions, the atom–atom contact area is the solid angle projection of a cell face onto the expanded sphere surface.

ALGORITHM

Definition of contact planes

Applied to a protein structure, the Voronoi procedure assigns a region in space to each protein atom, bounded by contact planes between neighboring atoms (Figure 1). Here, we have used definitions of contact planes that do not introduce a vertex error. In the presented algorithm, the contact plane definition may be selected as bisecting (Voronoi, 1908), radical plane (Gellatly and Finney, 1982), or extended radical plane (see below). Using the radical plane definition, the distance from an atom to the contact plane is given by

$$p_{ij} = \frac{d_{ij}^2 + r_i^2 - r_j^2}{2d_{ij}}, \tag{1}$$

where p_{ij} is the distance from atom i to the contact plane between atoms i and j, d_{ij} is the distance between atom centers, and r_i and r_j are the van der Waals radii.

The extended radical plane definition is similar to that of Gellatly and Finney (1982), but increases the radius of

each atom by the radius of a solvent molecule:

$$p_{ij} = \frac{d_{ij}^2 + (r_i + r_w)^2 - (r_j + r_w)^2}{2d_{ij}}, \qquad (2)$$

where r_w is 1.4 Å, the radius of a water molecule. This plane contains the circle of intersection of two spherical surfaces with radii $(r_i + r_w)$ and $(r_j + r_w)$, with centers separated by distance d_{ij}. The expanded radius $r_i + r_w$ is termed the contact radius. A similar formula was used previously to calculate solvent accessible surfaces (Fraczkiewicz and Braun, 1998). Contact planes for a given atom are calculated for all atoms within contact range, where the contact range for a pair of atoms is defined as the maximum distance such that a solvent molecule cannot pass freely between the atoms or, equivalently, as the sum of the contact radii.

Edge traversing algorithm

The presented algorithm constructs convex hulls for each atom in sequence. Given a set of possible contact planes surrounding an atom, a convex hull is constructed by traversing the edges of the hull from an initial vertex. Each vertex on the convex hull is calculated from the intersection of three planes of contact, and has three associated edges. Edges are treated as vectors emanating from a starting vertex, and the positive intersection of an edge vector with the next closest plane of contact identifies an adjacent vertex on the hull. The intersection of the edge vector with a plane terminates the given vector, and is the starting point for two new edge vectors. A list of unfinished edge vectors is maintained, and each newly created vector compared to the edge list. If a vector defined by the intersection of the same two planes is already present, the edge connects two known vertices, and both vectors are removed from the edge list. When the start and end vertices of all edge vectors have been identified, the convex hull is completed. Using this method, a single convex hull may be traversed in time $O(kN_C)$, where k is the number of edges and N_C is the number of contact planes. As contacts are limited to within a given radius, the time required for calculating all convex hulls within a protein is proportional to the number of protein atoms. The edge traversing algorithm eliminates the need to transform contact planes into dual space, in contrast to other methods (Preparata and Shamos, 1985; O'Rourke, 1994).

Information from previously calculated hulls is used in the construction of new hulls to increase the efficiency of the tessellation procedure. A list of previous atom–atom contacts is maintained, and this list used to reduce the number of possible contact atoms for new hulls. For a given atom, previous atoms are used in the hull construction only if they are known to have a non-zero contact area with the given atom, removing redundant atoms from the calculations. Additionally, selected vertices from previous hulls are saved as starting points for the generation of new convex hulls, removing the need to find initial vertices for the majority of atoms.

If no previous vertex has been generated for the current atom, an initial vertex is required for the convex hull. The contact plane closest to the current atom center will form a face on the convex hull, and this plane is used as a starting location. Lines of intersection between the closest plane and other contact planes are calculated, and the edge passing closest to the center atom is retained. Similarly, the points of intersection of this line with planes other than the first two are calculated, and the point closest to the origin is retained. This point lies on the surface of the hull.

If a convex hull contains a degeneracy, e.g. a vertex at the exact intersection of four or more edges, the edge traversing algorithm may give an unbounded solution. This is a rare occurrence (no degeneracies were found in any of the cases tested). A fail-safe has been incorporated into the algorithm, so that if the solution of a convex hull does not converge, a small random perturbation is applied to the coordinates of the central atom and the hull recalculated.

In contrast to the original Voronoi procedure, the region of space assigned to a given atom is limited to within the contact radius. Regions of the convex hull beyond the contact radius are truncated, and the defined atom cell is the intersection of the convex hull with a sphere. Additional vertices are calculated from the intersection of edge vectors with the sphere surface. Portions of the truncated polyhedron, or atom cell, on the surface of the sphere define the solvent accessible surface. If the entire polyhedron falls within the expanded radius of the atom, the atom is internal to the protein and has zero SAS. The calculated SAS is in agreement with the previous definition of the SAS (Lee and Richards, 1971) using the expanded radical plane definition (2).

The complete algorithm is outlined in Figure 2.

Calculation of atom–atom contacts and solvent accessible surfaces

For a cell associated with a given atom, the cell surface is comprised of the contact planes with neighboring atoms and the surface of the expanded atom sphere (the SAS). We define atom–atom contacts as the projection of the cell faces onto the surface of the expanded atom sphere (Figure 1). Together with the SAS, the contacts of a given atom sum to the total surface area of the expanded sphere. The contact areas and the SAS are optionally normalized to a percentage of the sphere surface, providing a distribution of atom contacts and solvent accessible surface for each atom. This permits direct comparison of contact distributions between atoms, including atoms of differing radii. This measure of contact

```
Procedure Voronoi_diagram {
    Index protein atoms to three-dimensional lattice;
    While all are polyhedra are not completed {
        Get next atom from index;
        Add atoms within contact range to contact list;
        Remove previous atoms not in contact;
        Calculate contact plane equations;
        If no initial vertex for current polyhedron {
            Identify contact plane closest to atom center;
            Identify closest edge on plane;
            Identify closest intersection point on edge;
            Store initial vertex;
        }
        Generate three edges from initial vertex;
        While unfinished edge vectors remain {
            Get next unfinished edge;
            For each contact plane on contact list {
                Get intersection of contact plane with edge;
                If (intersection  closest to start of edge)
                and (vector in positive direction) {
                    Keep intersection point as next vertex;
                }
            }
            Add new vertex to vertex list;
            Calculate edge vectors from new vertex;
            Add to edge list;
            Remove edge vector pairs defining same edge;
        }
        For each contact plane on contact list {
            If no initial vertex for contacting atom {
                store initial vertex for contact atom;
            }
        }
    }
    Calculate intersection of polyhedron with sphere;
    Calculate atom SAS, atom contacts, and/or volumes;
}
End Procedure.
```

Fig. 2. Edge traversing algorithm for constructing the Voronoi tessellation.

may be interpreted geometrically in terms of solid angles, with the normalized area of each face being $1/(4\pi)$ times the solid angle of the face subtended by the origin.

Areas for both the SAS and inter-atomic contacts are calculated as the sum of spherical triangles and spherical arc segments. The area of a spherical triangle can be calculated from the surface angles of the triangle; however, we found that in some cases this form of the area calculation has an excessive round-off error. Therefore, a mathematically equivalent form was used for the spherical triangle area based on the angles between the vertices subtended by the origin. Here, the area of a spherical triangle is given by

$$Area_T = 4r_i^2 * \mathrm{Arctan}\left\{\left(\tan\frac{S}{2}\tan\frac{(S-\angle AOB)}{2} \times \tan\frac{(S-\angle AOC)}{2}\tan\frac{(S-\angle BOC)}{2}\right)^{1/2}\right\} \quad (3)$$

where O is at the sphere center, A, B, and C are points on the surface of the sphere, r_i is the radius of the sphere, $\angle AOB$, $\angle AOC$, and $\angle BOC$ are the angles between the vertices subtended by the origin, and $S = (\angle AOB + \angle AOC + \angle BOC)/2$. The angles $\angle AOB$ may be calculated from the dot product of the vectors from the origin to each point defining the angle:

$$\angle AOB = \mathrm{Arccos}[(\overrightarrow{OA} \bullet \overrightarrow{OB})/(|\overrightarrow{OA}||\overrightarrow{OB}|)], \quad (4)$$

with similar formulas for $\angle AOC$ and $\angle BOC$. In the implementation of the area calculation procedure, the computationally time-consuming arccosine and tangent functions were eliminated between (3) and (4), and the tangent functions in equation (3) were expressed as a function of the cosines between vertices:

$$\begin{aligned} u &= ((1+\cos\angle AOB)(1+\cos\angle AOC) \times(1+\cos\angle BOC)/8)^{1/2} \\ v &= ((1+\cos\angle AOB)(1-\cos\angle AOC) \times(1-\cos\angle BOC)/8)^{1/2} \\ w &= ((1-\cos\angle AOB)(1+\cos\angle AOC) \times(1-\cos\angle BOC)/8)^{1/2} \\ x &= ((1-\cos\angle AOB)(1-\cos\angle AOC) \times(1+\cos\angle BOC)/8)^{1/2} \end{aligned} \quad (5a)$$

$$\begin{aligned} \tan\left(\frac{S}{2}\right) &= \left(\frac{1-u+v+w+x}{1+u-v-w-x}\right)^{1/2} \\ \tan\left(\frac{S-\angle AOB}{2}\right) &= \left(\frac{1-u+v-w-x}{1+u-v+w+x}\right)^{1/2} \\ \tan\left(\frac{S-\angle AOC}{2}\right) &= \left(\frac{1-u-v+w-x}{1+u+v-w+x}\right)^{1/2} \\ \tan\left(\frac{S-\angle BOC}{2}\right) &= \left(\frac{1-u-v-w+x}{1+u+v+w-x}\right)^{1/2} \end{aligned} \quad (5b)$$

The subdivision of faces into spherical triangles may be used for all contact areas that do not border the SAS. If an edge connects two vertices on the SAS boundary, the connecting line is no longer a great circle arc but lies on the circle of intersection of two spherical surfaces (the edge of a spherical cap). The additional area may be calculated as the difference between the pie-shaped spherical cap segment defined by the two arc points and the center of the spherical cap, and the spherical triangle defined by the same three points:

$$Area_{Arc} = r_i^2(\angle BAC)(1-\cos\angle AOB) - Area_T(ABC) \quad (6)$$

where A is the center of the spherical cap, B and C are two points on the cap edge, and $\angle AOB = \angle AOC$. From equations (3)–(5b) the area of the projection of

any cell surface may be calculated by subdivision into spherical triangles and spherical arcs. The normalized contact area of atom j on atom i is $A_{ij} = (\Sigma Area_T + \Sigma Area_{Arc})/4\pi r_i^2$, and the SAS may be calculated as $\mathrm{SAS}_i = 4\pi r_i^2(1 - \Sigma_j A_{ij})$. It would also be possible to calculate the SAS directly from the vertices around each solvent exposed region, using either the Gauss–Bonnet theorem (Connolly, 1983; Richmond, 1984) or subdivision into spherical triangles and arcs. It was found that the time of calculation for the contact areas was not the limiting factor in the speed of the algorithm, so a separate subroutine for calculating the SAS alone was not implemented.

Polyhedron vertices are ordered around each face, to facilitate the division of polyhedron faces into triangles and arcs. Starting with two points defining an edge of the polyhedron face, the other vertices may be ordered around the face according to the cosine of the angle between the given point and the starting edge. The face is divided into triangular regions, given a starting vertex and pairs of adjacent vertices in cyclic order. If the edge between two adjacent points borders the SAS, an additional term for the arc area is added.

Area correction for engulfing contacts

For covalently bonded atoms with dissimilar radii, it is possible that the radical plane or extended radical plane can result in an atom center falling outside the boundaries of its associated polyhedron. If the defined contact plane is on the opposite side of the center atom, the contact is termed an engulfing contact. In this case, the original projection of points to the sphere surface from the origin does not effectively partition the contact areas, as the area of the engulfing contact would increase in the presence of other contacts. We therefore introduce a correction by explicitly projecting the polyhedron vertices from a point inside the Voronoi polyhedron to the surface of the SAS sphere. The point within the polyhedron closest to the atom center is selected as the center of projection. If one engulfing contact is present, this point is in the center of the engulfing contact plane. Contact areas are calculated from the projected points as before, using the original atom center as the center of sphere center. Points defining the solvent accessible area are already on the SAS sphere, and are unchanged by this procedure.

The area of the engulfing contact is set to its contact area calculated in the absence of other contacts, i.e. the area of a sphere minus a spherical cap. A correction term is added for contacts bounding the engulfing contact. Given two adjacent vertices on the boundary of the engulfing contact and a second contact, the area encompassed by the edge of the spherical cap and a great circle arc between the two vertices is assigned to the non-engulfing contact.

Volume calculations

The volume of a Voronoi cell with zero solvent accessible surface was calculated from the surface area of each of the cell faces. Ordering the vertices around a cell face as $P_1, P_2, \ldots, P_n$, the area of the face j is calculated as

$$AF_j = \sum_{k=2\,\mathrm{to}\,n-1} |\overrightarrow{P_1P_k} \times \overrightarrow{P_1P_{k+1}}|/2. \qquad (7a)$$

The volume of the entire cell is calculated as the sum of pyramids from each face to the origin:

$$V_i = \sum_{j=1\,\mathrm{to}\,m} p_{ij} AF_j/3, \qquad (7b)$$

where the cell has m faces and p_{ij} is the distance from atom i to face j. The volume of surface atoms may be calculated from the faces of the polyhedron and the SAS. Two additional terms are present in the face area for points bounding the SAS:

$$\begin{aligned} AF_j = &\sum_{k=2\,\mathrm{to}\,n-1} |\overrightarrow{P_1P_k} \times \overrightarrow{P_1P_{k+1}}|/2 \\ &+ \sum_{k=1\,\mathrm{to}\,n-1} \delta_{k,k+1} r_j^2 \angle P_k O_f P_{k+1}/2 \\ &- \sum_{k=1\,\mathrm{to}\,n-1} \delta_{k,k+1} (|\overrightarrow{P_kP_{k+1}}|/2) \\ &\times (r_j^2 - |\overrightarrow{P_kP_{k+1}}|^2/4)^{1/2} \end{aligned} \qquad (8a)$$

where O_f is the center of the face, r_j is the distance from O_f to the intersection of the face with the sphere surface, $\delta_{k,k+1} = 1$ for vertices connected by the SAS boundary, and $\delta_{k,k+1} = 0$ otherwise. The volume assigned to the surface atom is

$$V_i = \sum_{j=1\,\mathrm{to}\,m} p_{ij} AF_j/3 + SAS(r_i + r_w)/3. \qquad (8b)$$

Implementation

The speed of program execution was increased using established methods. The initial search for atoms within contact range uses a lattice index, in which the protein is sub-divided into cubic regions and atoms are indexed to each region. The dimensions of each cube are 7.0 Å, slightly larger than twice the largest contact radius. The distance to the center atom is calculated for each atom within the same cube and for the 26 neighboring cubes.

Source code for the Voronoi procedure and for area and volume calculations were written in C, compiled with the gcc compiler, and executed on a 850 MHz Pentium III PC running the Linux operating system (Redhat 7.1). The publicly available area and volume calculation programs used for comparison were compiled and executed using the same system when source code was available. The

-O2 level of optimization was used for all program compilations.

The default atomic radii are those from Tsai *et al.* (1999), though user defined radii may also be specified. Other values for atomic radii (Chothia, 1975; Sobolev *et al.*, 1996) were also used to permit comparison with previous calculation methods.

RESULTS AND DISCUSSION

Solvent accessible surface calculations were compared to those obtained from algorithms using a numerical procedure, an approximate analytical procedure, and an exact analytical procedure. Three protein structures, 153l (O-glycosyl hydrolase, 1432 atoms), 1ads (aldose reductase, 2514 atoms), and 1mcp (immunoglobulin fab fragment, 3402 atoms), were selected from the Protein Data Bank (Bernstein, 1977; Berman *et al.*, 2000) and used for testing.

The numerical surface area estimate was based on the method of Shrake and Rupley (1973), which uses a finite number of points on the expanded sphere surface. The implementation of the numerical procedure (Sobolev *et al.*, 1996) was adapted from the WHAT IF molecular modeling package (Vriend, 1990). Results were compared for spherical shells containing 610, 1597, 4181, and 10946 surface points. The agreement between our method and the numerical algorithm is quite good, with a root mean square deviation (RMSD) of less than 0.31 Å^2 in all cases. Given an even distribution of points on the sphere surface, the numerical result should agree with the analytical when the number of points is large. This is the observed case, as the RMSD of the numerical estimate converges towards the analytical solution with an increasing number of points. The execution times of the new algorithm with the most rapid numerical algorithm were almost identical, deviating by a maximum of 0.01 s for the three test cases (Table 1).

The approximate analytical method for calculating surface areas (Gerstein, 1992) estimates the SAS by generating a set of parallel arcs on the expanded sphere surface (Richards, 1985). Differences in estimates between our method and the arc-based method were also small, with a maximum RMSD of less than 0.5 Å^2. The arc-based method had an execution time approximately three times longer than the presented method (Table 1).

A general purpose Voronoi tessellation giving an analytical measure of the solvent accessible surface was also used for comparison (Liang *et al.*, 1998). This program first calculates a Delaunay triangulation, which is then converted into the Voronoi diagram. Given the same atomic radii, the two programs provide almost identical results, with an RMSD of less than 0.001 Å^2 for each of the test structures. The procedure of Liang et al. has a considerably longer execution time; however, it can also provide additional information such as molecular surfaces and cavities.

Table 1. Comparison of execution times and RMSD between solvent accessible surface algorithms. The RMSD includes all atoms with a non-zero SAS. The same definition of atomic radii was used for each pairwise comparison. Protein structures were obtained from the Protein Data Bank

Method	PDB entry 153l Time (s)	153l RMSD (Å^2)	1ads Time (s)	1ads RMSD (Å^2)	1mcp Time (s)	1mcp RMSD (Å^2)
Vsurface[a]	0.25	—	0.45	—	0.61	—
Point-based[b]						
610[c]	0.25	0.29	0.44	0.30	0.60	0.31
1597[c]	0.47	0.14	0.81	0.14	1.09	0.15
4181[c]	1.02	0.07	1.79	0.07	2.43	0.07
10946[c]	2.54	0.03	4.49	0.03	6.41	0.04
Arc-based[d]	0.82	0.47	1.39	0.44	1.76	0.46
Analytical[e]	11.9	<0.001	21.4	<0.001	30.2	<0.001

[a]This work.
[b]Program of Sobolev and Edelman (1995), based on the method of Vriend (1990).
[c]Using a given number of surface points, as indicated.
[d]Program of Gerstein (1992), based on the method of Richards (1985).
[e]Program of Liang et al (1998). Time includes generation of Delaunay tetrahedrization, conversion to Voronoi diagram, and calculation of surface areas.

Volume calculations were compared with those obtained from an analytical method, and the effects of the different plane definitions evaluated (Table 2). The atom volume calculation method used for comparison (Gerstein *et al.*, 1995) implements a Voronoi procedure to generate atom volumes. Given the same choice of contact planes, this procedure and the presented algorithm should generate similar volume estimates for internal protein atoms. If bisecting contact planes are used in both algorithms, the procedures produce the same volume estimates (Table 2). There is a small deviation if the radical plane is used to define the contact planes. In this case, the contact plane definitions are no longer identical, as the method of Gerstein *et al.* (1995) uses a different definition of contact plane for contacts between covalently bonded atoms. The difference between the two algorithms using the radical plane definition is not large, with a maximum RMSD of 0.31 Å^3 between methods (Table 2). The speed of calculation of the new algorithm is approximately two and a half times faster than the previous Voronoi procedure. The intended use of the extended radical plane (2) is for calculation of the SAS and atom–atom contacts, but not volumes. However, the difference in volumes generated using it and the radical plane provides information on variation due to the contact definition, so a comparison has been presented in Table 2. Overall, the average RMSD between the two methods was 3.4 Å^3.

Table 2. Execution time and root mean square difference of volume calculations for proteins 153l, 1ads, and 1mcp. The RMSD between adjacent columns in the table is calculated from internal protein atoms (those with zero SAS). See text for contact plane definitions. Atomic radii as defined by Chothia (1975) were used to permit comparison between methods

Method	Contact plane[c]	PDB entry 153l Time (s)	Mean volume (Å^3)	RMSD (Å^3)	1ads Time (s)	Mean volume (Å^3)	RMSD (Å^3)	1mcp Time (s)	Mean volume (Å^3)	RMSD (Å^3)
Vvolume[a]	B	0.27	15.2 ± 7.21	<0.01	0.45	16.0 ± 7.7	<0.01	0.62	15.4 ± 6.9	<0.01
calc-volume[b]	B	0.69	15.2 ± 7.2		1.21	16.0 ± 7.7		1.53	15.4 ± 6.9	
Vvolume[a]	R	0.26	16.1 ± 7.2	0.12	0.45	16.7 ± 7.4	0.31	0.61	16.6 ± 6.5	0.18
calc-volume[b]	R	0.7	16.1 ± 7.2		1.18	16.7 ± 7.4		1.5	16.6 ± 6.5	
Vvolume[a]	R	0.26	16.1 ± 7.2	3.4	0.45	16.7 ± 7.4	3.4	0.61	16.6 ± 6.5	3.4
Vvolume[a]	X	0.24	17.3 ± 8.0		0.42	17.7 ± 8.0		0.55	17.6 ± 7.3	

[a]This work.
[b]Program of Gerstein *et al.* (1995).
[c]Contact plane type: B, bisecting plane; R, radical plane; X, extended radical plane. See text for definitions.

As the atom–atom contact measure is newly introduced, its accuracy can not be directly compared with other contact measures. The potential use of the measure may however be illustrated by the distribution of atom–atom contacts within the test proteins. The three test proteins, 153l, 1ads, and 1mcp, represent α-helical, mixed α/β, and β-sheet structures, respectively. Contact frequencies and area distributions were calculated for four sample atom–atom interactions within the protein structures. The contact distributions for backbone N/O, backbone N/N, sidechain C_β/C_β, and backbone N/C_β are shown in Figure 3. The atom–atom contact areas are expressed in SAS equivalent units of square Angstroms. The total observed contact frequencies, ranked in ascending order, are N/N, N/C_β, C_β/C_β, and N/O, consistent with the expected ranking of the energetic interactions for H-bond donor/donor, H-bond donor/hydrophobic, hydrophobic/hydrophobic, and H-bond donor/acceptor pairs. This ranking is further corroborated by the contact area distributions. Using the 90th percentile as a measure of the contact area range, the upper percentile limits of the four interaction types were 3, 5, 10, and 12 Å^2, respectively. The distributions of the atom pairs were comparable between the three test proteins for each interaction type, with the exception of the backbone N/O distribution. In this case, both the frequency and distribution were sensitive to the secondary structure of the protein. A greater frequency of N/O contacts was seen in the α-helical structure (153l) versus the β-sheet structure (1mcp), with the mixed α/β structure (1ads) intermediate between the two. A peak between 8 and 12 Å^2 was seen within each N/O distribution, representing hydrogen bonds between backbone N and O. The peak maximum is shifted by 2 Å^2 for 1mcp, representing a greater contact surface between N and O within β-sheets. A second peak between 4 and 7 Å^2 is seen within 153l and to a lesser extent 1ads, representing an additional contact between the first backbone nitrogen and third backbone oxygen in one turn of an α-helix. While this interaction would not be as strong as a direct hydrogen bond, it is still favorable due to partial charge interactions. Overall, the contact measure described here was correlated to the expected strength of interaction, and is thus likely to be a useful descriptor in statistical assessments of protein structure.

CONCLUSIONS

The described algorithm provides an effective method of quantifying atomic volumes and solvent accessible surfaces, and has a reduced execution time as compared to other methods. The method has a greater accuracy than numerical and approximate methods for calculation of solvent accessible surfaces, and comparable accuracy for calculating volumes.

Application of the original Voronoi algorithm to protein structures results in unbounded regions for some protein surface atoms, and also some polyhedra with regions extending beyond a distance of two atomic radii from the center of the associated atom. The introduced algorithm limits surface cells to within the radius of an expanded van der Waals sphere, reducing the estimates of volumes and surfaces for unbounded or extended polyhedra. The volume of atoms on the protein surface will still be greater than that calculated using explicit solvent molecules, and depending on the modeling task, more appropriate estimation of volumes of surface atoms can be attained if explicit solvent molecules are specified.

The atom–atom contact measure provides a means for quantifying inter-atomic contacts within a protein.

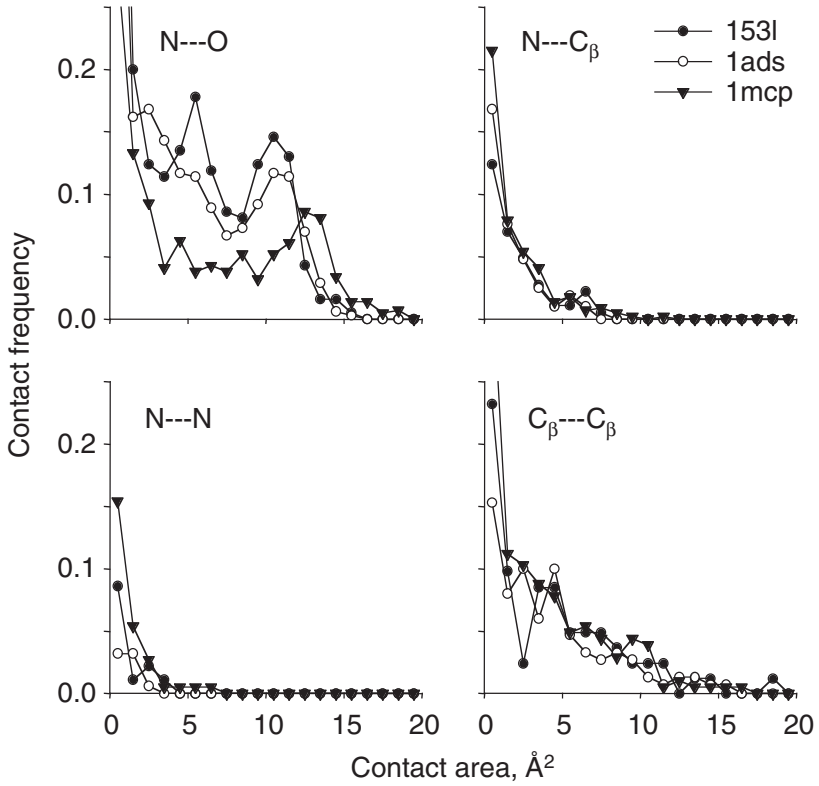

Fig. 3. Distributions of atom–atom contact areas within proteins 153l, 1ads, and 1mcp, for four sample atom pairs. N–O, contacts between backbone nitrogen and backbone oxygen; N–C_β, contacts between backbone nitrogen and sidechain beta carbon; N–N, contacts between backbone nitrogens; C_β–C_β, contacts between sidechain beta carbons. The contact area is expressed in solvent accessible surface equivalent units (Å^2) and the contact frequencies expressed as the probability of the given contact area per atom.

The method is an improvement over the direct use of Voronoi cell surfaces, as the total contact area of an atom (including the SAS) is constant and independent of the number and location of neighboring atoms. Additionally, the angular projection method introduces a more realistic distance dependency to the contact measure, as the projected area of a face element is inversely proportional to the square of the distance of the element from the atom center. The inverse dependence on distance is comparable to energetically based potentials, and thus the introduced contact measure may better approximate atom–atom interactions.

A contact surface between atoms bordering the solvent accessible surface will have a somewhat greater contact area than if explicit solvent atoms are specified. The use of explicit solvent atoms could remedy this, but would also introduce a random error into the calculations, as the contact areas become dependent upon the artificially created locations for solvent atoms. The inclusion of explicit solvent molecules greatly increases the computational complexity within protein folding and docking algorithms, and thus the contact measure without explicit solvent molecules may be preferred for these applications.

In summary, the presented algorithm for the Voronoi tessellation of a protein is faster than previous methods, and provides precise, continuous estimates of volumes, solvent accessible surfaces, and inter-atomic contacts. The introduced contact measure permits the quantification of atom–atom contacts, and may be an effective method of describing non-covalent interactions between protein atoms.

ACKNOWLEDGEMENTS

VS was supported by the State of Israel, Ministry of Absorption, Center for Absorption of Scientists.

REFERENCES

Abagyan,R.A. and Totrov,M.M. (1997) Contact area difference (CAD): a robust measure to evaluate accuracy of protein models. *J. Mol. Biol.*, **268**, 678–685.

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

Bernal,J.D. and Finney,J.L. (1967) Random close-packed hard-sphere model II. Geometry of random packing of hard spheres. *Discus. Faraday Soc.*, **43**, 62–69.

Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,Jr,E.E., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535–542.

Chothia,C. (1975) Structural invariants in protein folding. *Nature*, **254**, 304–308.

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

Connolly,M.L. (1993) The molecular-surface package. *J. Mol. Graph.*, **11**, 139–143.

de la Cruz,X. and Calvo,M. (2001) Use of surface area computations to describe atom–atom interactions. *J. Comput.-Aided Mol. Des.*, **15**, 521–532.

Delarue,M. and Koehl,P. (1995) Atomic environment energies in proteins defined from statistics of accessible and contact surface-areas. *J. Mol. Biol.*, **249**, 675–690.

Delaunay,B. (1934) Sur la sphere vide. *Izv. Akad. Nauk SSSR, Otdel. Mat. Est. Nauk*, **7**, 793–800. (see Preparata and Shamos, 1985).

Dodd,L.R. and Theodorou,D.N. (1991) Analytical treatment of the volume and surface-area of molecules formed by an arbitrary collection of unequal spheres intersected by planes. *Mol. Phys.*, **72**, 1313–1345.

Eisenhaber,F., Lijnzaad,P., Argos,P., Sander,C. and Scharf,M. (1995) The double cubic lattice method—efficient approaches to numerical-integration of surface-area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.*, **16**, 273–284.

Fraczkiewicz,R. and Braun,W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.*, **19**, 319–333.

Gellatly,B.J. and Finney,J.L. (1982) Calculation of protein volumes: an alternative to the Voronoi procedure. *J. Mol. Biol.*, **161**, 305–322.

Gerstein,M. (1992) A resolution-sensitive procedure for comparing protein surfaces and its application to the comparison of antigen-combining sites. *Acta Crystallogr. Sect. A*, **48**, 271–276.

Gerstein,M., Tsai,J. and Levitt,M. (1995) The volume of atoms on the protein surface—calculated from simulation, using Voronoi polyhedra. *J. Mol. Biol.*, **249**, 955–966.

Goede,A., Preissner,R. and Frommel,C. (1997) 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**, 1113–1123.

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

Liang,J., Edelsbrunner,H., Fu,P., Sudhakar,P.V. and Subramaniam,S. (1998) Analytical shape computation of macromolecules: I. Molecular area and volume through alpha shape. *Proteins*, **33**, 1–17.

O'Rourke,J. (1994) *Computational Geometry in C*. Cambridge University Press, Cambridge.

Preparata,F.P. and Shamos,M.I. (1985) *Computational Geometry. An Introduction*. Springer, New York.

Richards,F.M. (1974) The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.*, **82**, 1–14.

Richards,F.M. (1985) Calculations of molecular volumes and areas for structures of known geometry. *Method Enzymol.*, **115**, 440–465.

Richmond,T.J. (1984) Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.*, **178**, 63–89.

Shrake,A. and Rupley,J.A. (1973) Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.*, **79**, 351–371.

Sobolev,V. and Edelman,M. (1995) Modeling the quinone-B binding-site of the photosystem-II reaction-center using notions of complementarity and contact surface between atoms. *Proteins*, **21**, 214–225.

Sobolev,V., Wade,R.C., Vriend,G. and Edelman,M. (1996) Molecular docking using surface complementarity. *Proteins*, **25**, 120–129.

Tsai,J., Taylor,R., Chothia,C. and Gerstein,M. (1999) The packing density in proteins: Standard radii and volumes. *J. Mol. Biol.*, **290**, 253–266.

Voronoi,G. (1908) Nouveles applications des parametres continus a la theorie des formes quadratique. *J. Reine Angew. Math.*, **134**, 198–287. (see Preparata and Shamos, 1985).

Vriend,G. (1990) What If—a molecular modeling and drug design program. *J. Mol. Graph.*, **8**, 52–56.

Zimmer,R., Wohler,M. and Thiele,R. (1998) New scoring schemes for protein fold recognition based on Voronoi contacts. *Bioinformatics*, **14**, 295–308.