

The weighted-volume derivative of a space-filling diagram

Herbert Edelsbrunner*[†] and Patrice Koehl*^{‡§}

*Department of Computer Science, Duke University, Durham, NC 27708-0129; [†]Raindrop Geomagic, Research Triangle Park, NC 27709; and [‡]Department of Structural Biology, Stanford University, Stanford, CA 94305-5126

Communicated by Michael Levitt, Stanford University School of Medicine, Stanford, CA, December 22, 2002 (received for review December 12, 2002)

Computing the volume occupied by individual atoms in macromolecular structures has been the subject of research for several decades. This interest has grown in the recent years, because weighted volumes are widely used in implicit solvent models. Applications of the latter in molecular mechanics simulations require that the derivatives of these weighted volumes be known. In this article, we give a formula for the volume derivative of a molecule modeled as a space-filling diagram made up of balls in motion. The formula is given in terms of the weights, radii, and distances between the centers as well as the sizes of the facets of the power diagram restricted to the space-filling diagram. Special attention is given to the detection and treatment of singularities as well as discontinuities of the derivative.

molecular dynamics | implicit solvent models | spheres | Delaunay triangulation

It is widely believed that characterizing the geometry of a protein molecule is essential for understanding its folding process as well as its interactions with other biomolecules and small ligands. Among all geometric measures, volume is probably the most fundamental property to study. The relevance of calculations of molecular volume is seen in the correlation of volume, or change in volume, with experimental factors such as energetics (1) as well as with physical chemical properties of the molecule such as charge, temperature, and pressure (2, 3). There is a large amount of current research that focuses on the calculations of standard atomic volumes (4–6). These atomic volumes have been considered as parameters that can be used to measure the quality of a three-dimensional protein structure (7) as well as probes to quantify structural variations in proteins that occur during evolution (8). Atomic volumes have also been used in packing (9–11) and solvation-energy calculations (12, 13). The latter energy function is central to computational biology, because solvation is an important factor determining the structure and thermodynamics properties of molecules in aqueous solution. All solvent effects on a molecule can be included in an effective solvation potential, $W = W_{\text{elec}} + W_{\text{np}}$, in which the first term accounts for the molecule–solvent electrostatics polarization, and the second term accounts for the molecule–solvent van der Waals interactions and for the formation of a cavity in the solvent. W_{elec} is usually described by a continuum model such as the generalized Born model (13, 14). Interestingly, an essential element of such a model is the knowledge of the volume associated with the individual atoms of the solute. Many solvation models describe the nonpolar part of the solvation energy W_{np} as a weighted sum of the solvent-exposed or -accessible surface area of each atom of the solute (15, 16). There is evidence, however, that for small solute the hydrophobic term W_{np} is not proportional to the surface area (17) but rather to the excluded volume of the molecule (18). A volume-dependent solvation term was originally introduced by Gibson and Scheraga (19) as the hydration shell model. Biomolecular simulations with implicit solvent that rely on iterative energy calculations such as energy minimization and molecular dynamics require efficient, exact algorithms for the evaluation of the solvation-energy function and its derivatives. In this article, we provide

a complete geometric description of the derivatives of the weighted volume of a molecule when its atoms are in motion.

Lee and Richards (20) define the accessible surface of a protein as the van der Waals surface of the molecule expanded by the radius of the solvent sphere about each atom center. The solvent-excluded volume is the volume enclosed by the accessible surface. Computational methods that evaluate the solvent-excluded volume can be divided into approximate and exact methods. Most of the approximate methods rely on numerical integration (21–23), whereas most analytical methods (9, 12, 24–27) point out that overlap volumes of four or more atoms do not occur in real molecules. An alternative approach for computing the volume and studying the packing of balls was originally developed by Voronoi (28) and later was applied to protein by Richards (9). It has been used successfully for the calculation of protein constituents, the description of protein motions, and the analysis of cavities in proteins (2, 8, 29, 30). A drawback of this approach is that it requires information on the first shell of hydration so that polyhedra may be defined for atoms at the surface of the molecule. It was also developed for equal-size balls and its applications to proteins with atoms that have different radii required approximations (30). Pavani and Ranghino (22) proposed a method for computing the volume of a molecule by using the inclusion–exclusion principle. In their implementation, only intersections of up to three balls were considered. This idea was generalized to any levels of intersections by Gibson and Scheraga (26), applying a theorem from mathematics that higher-order unions and intersections can always be reduced to lower-order unions and intersections (31). Doing the reduction correctly, however, remains computationally difficult and expensive. The alpha-shape theory solves this problem exactly by using Delaunay triangulations and their filtrations (32). Alpha shapes have been used to compute the surface area and volume of proteins as well as for detecting and measuring cavities in proteins (33). Fig. 1 provides a simple overview of the application of the alpha-shape theory to compute the volume of a protein.

The distinction between approximate and exact computation also applies to existing methods for computing the derivatives of the solvent-excluded volume with respect to atomic coordinates (34–36). All these methods have to take a large number of singularities into account, where approximations are usually required (36). This problem of singularities is even more acute for surface-area calculation (37, 38). The alpha-shape theory proposes a robust solution to this problem by implementing arbitrary precision arithmetic to avoid numerical problems and systematically resolving all singularities without explicitly perturbing the positions of the ball centers. The latter method is referred to as Simulation of Simplicity (39). In this article we describe an extension of the alpha-shape method that implements the weighted volume derivative theorem described below to provide an efficient, robust, exact computation of the derivatives of volumes. There is an inherent difficulty in using a potential based on volume for energy minimization or molecular dynamics. Although the volume is continuous in the position of the atoms, its derivatives may not be. We examine this issue within the framework of the alpha-shape method and relate discontinuities to

[§]To whom correspondence should be addressed. E-mail: koehl@csb.stanford.edu.

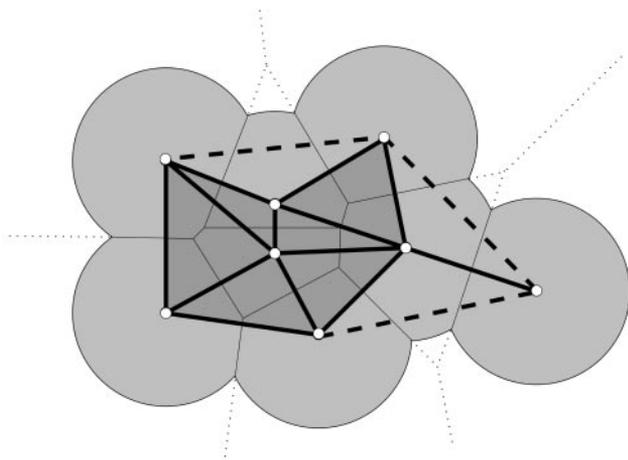


Fig. 1. Computing geometric properties of a molecule. As is common in biology, atoms are treated as intersecting balls, the union of which forms the space-filling diagram. Gerstein *et al.* (8) and Richards (9) proposed to study this union by using the Voronoi diagram (or more exactly the power diagram since the balls have different radii). That diagram divides the space into convex cells, one per atom. In the two-dimensional example shown here, the edges that separate these cells are shown as solid inside and dotted outside the union of disks. Note that the power cells of some surface atoms extend to infinity. The superimposed Delaunay triangulation (thick solid and dashed lines) is the dual of the power diagram obtained by drawing a line segment between two ball centers if their convex cells share a common edge. Despite their different appearance, the Delaunay triangulation and the power diagram contain exactly the same information. The key to connecting the power diagram to the molecule is to consider the intersection of the two; this defines a convex decomposition of the space-filling diagram (i.e. the light shaded area inside the disks, divided into convex regions by the solid part of the Voronoi edges). The dual of this decomposition is the dual complex (thick solid lines and shaded triangles). The dual complex is a subset of the Delaunay triangulation and contains all simplices (tetrahedra, triangles, and edges) that correspond to overlapping atoms. As demonstrated in ref. 32, the dual complex contains all information about a molecule required to compute its surface and volume. In this article we show that the same dual complex can be used to compute the derivative of the volume.

combinatorial changes in the subcomplex of the Delaunay triangulation that is dual to the space-filling diagram of the molecule.

Background and Results introduces the background and states the main result of this paper. *Derivation* uses geometric arguments to prove the formula for the weighted-volume derivative. *Discontinuities* and *Implementation* discuss the continuity of the derivatives and the implementation of the theorem, and *Discussion* concludes the article.

Background and Results

As common in biology, we model an atom as a ball and a molecule as the union of a finite collection of such balls. This union is referred to as the space-filling diagram of the molecule. In this section we explain how we approach the problem of computing the derivatives of the weighted volume of a three-dimensional space-filling diagram.

Geometric Structures. Let $B_i = (z_i, \rho_i)$ be the ball with center z_i and radius ρ_i for $0 \leq i < n$ and call $F = \bigcup_i B_i$ a space-filling diagram. The corresponding state is the point $\mathbf{z} \in \mathbb{R}^{3n}$ that lists the $3n$ center coordinates in sequence. The weighted volume of F is a map $W: \mathbb{R}^{3n} \rightarrow \mathbb{R}$, and its derivative at \mathbf{z} is a linear map $DW_{\mathbf{z}}: \mathbb{R}^{3n} \rightarrow \mathbb{R}$ defined by $DW_{\mathbf{z}}(\mathbf{t}) = \langle \mathbf{w}, \mathbf{t} \rangle$, where $\mathbf{w} \in \mathbb{R}^{3n}$ is the gradient of W at \mathbf{z} . Specifying the derivative is equivalent to giving the gradient.

Denote the sphere bounding B_i by S_i . The power distance of a point $x \in \mathbb{R}^3$ from S_i is $\pi_i(x) = \|x - z_i\|^2 - \rho_i^2$, and the power cell of S_i is the set of points P_i for which S_i minimizes the power distance. We assume or simulate the generic case, in which every

power cell is a three-dimensional convex polyhedron, and every facet, edge, and vertex belongs to exactly two, three, and four power cells. The power diagram consists of all power cells and their facets, edges, and vertices. As illustrated in Fig. 1, the power diagram decomposes the space-filling diagram into convex cells of the form $P_i \cap F$, and these cells share facets, edges, and vertices, which are the intersections of F with the facets, edges, and vertices of the power diagram. The dual complex K of that decomposition consists of all vertices, edges, triangles, and tetrahedra that are convex hulls of sphere centers with cells that intersect in nonempty cells, facets, edges, and vertices.

Sizes and Distances. We use fractions to express the sizes of geometric entities in the decomposition of the space-filling diagram. For example, $\beta_i = \text{vol}(P_i \cap F) / \text{vol}(B_i)$ is the fraction of the i th ball that belongs to its power cell. The volume of $F = \bigcup_i B_i$ is therefore $(4\pi/3) \sum_i \beta_i \rho_i^3$, and given real weights α_i , the weighted volume is $W = (4\pi/3) \sum_i \alpha_i \beta_i \rho_i^3$. The formula for the derivative of W needs fractions of disks, which we now discuss. Two spheres intersect in a circle $S_{ij} = S_i \cap S_j$, which bounds a disk B_{ij} . The fraction of the disk that belongs to the power diagram is

$$\beta_{ij} = \frac{\text{area}(P_i \cap P_j \cap F)}{\text{area}(B_{ij})}. \quad [1]$$

As explained in ref. 32, it is straightforward to compute these fractions from the dual complex K of F by using the principle of inclusion–exclusion. More details on how to perform this calculation are provided in *Appendix*. We will also need the area of the facet $P_i \cap P_j \cap F$, which we compute as $\beta_{ij} A_{ij}$. An expression for $A_{ij} = \text{area}(B_{ij})$ in terms of radii and distances between centers of spheres will be given in *Derivation* (see Eq. 7).

The derivative of W also requires the distance between the line L_{ij} passing through the centers z_i and z_j and the line L_{ijk} , the points of which have equal power distance from S_i , S_j , and S_k . This distance $\delta_{ij,k}$ is equal to the distance between the center z_{ij} of S_{ij} and the point y_{ijk} on L_{ijk} closest to z_{ij} . We note that y_{ijk} is the unique point in the plane of z_i , z_j , and z_k that has equal power distance to all three spheres.

Averages. The derivative of W also refers to the weighted-average vector from the center to the boundary of a facet $P_i \cap P_j \cap F$. The weight is the infinitesimal contribution to the area of the facet as we rotate the vector. That weight is constant over each edge and each arc in the boundary of the facet. We therefore compute the weighted-average vector as the weighted sum of the average vectors of the edges and arcs, with the weights being the areas of the subtended sectors. The average vector from the center to an edge $xy = P_i \cap P_j \cap P_k \cap F$ is the vector to the midpoint, $V_{xy} = (x + y)/2 - z_{ij}$. Consider next an arc ab on the circle S_{ij} , and let $\varphi = \varphi_{ab}$ be half the angle, as in Fig. 2. Assuming the center is the origin, the radius is 1, and the arc midpoint lies on the abscissa, we get the average by integration

$$v_{ab} = \frac{1}{2\varphi} \int_{-\varphi}^{\varphi} \cos t \, dt = \frac{\sin \varphi}{\varphi}. \quad [2]$$

To give a formula without the normalizing assumptions, we use the midpoint M_{ab} of the arc ab . The average vector is $V_{ab} = v_{ab}(M_{ab} - z_{ij})$.

In general, the facet $P_i \cap P_j \cap F$ has a boundary that consists of several edges and arcs. The weighted-average vector from the center to that boundary is the weighted sum of the individual averages,

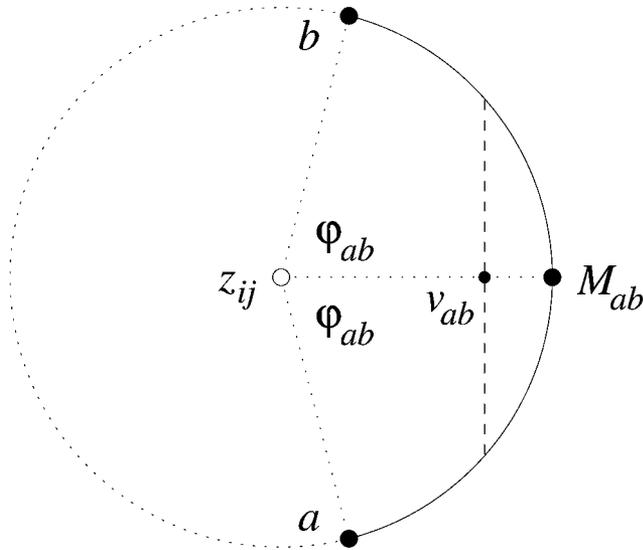


Fig. 2. The average of the solid arc ab lies on the line connecting the center z_{ij} and the midpoint M_{ab} of the arc.

$$V_{ij} = \frac{\sum_{xy} h_{xy} \|x - y\| V_{xy} + 2 \sum_{ab} \varrho_{ij}^2 \varphi_{ab} V_{ab}}{\sum_{xy} h_{xy} \|x - y\| + 2 \sum_{ab} \varrho_{ij}^2 \varphi_{ab}}, \quad [3]$$

where $h_{xy} = \delta_{ij,k}$ is the distance from z_{ij} to the line spanned by xy , and ϱ_{ij} is the radius of S_{ij} . The first sums range over all edges xy , and the second sums range over all arcs ab of the facet. Note that the denominator is twice the area of the facet. We provide a method to compute the weighted-average vector using inclusion–exclusion in *Appendix*.

Main Result. The main result of this article is a complete description of the weighted-volume derivative. We need some notation. The weight of the i th ball is $\alpha_i \in \mathbb{R}$, which may be positive or negative. We write $\zeta_{ij} = \|z_i - z_j\|$ for the distance between two centers and $U_{ij} = (z_i - z_j)/\zeta_{ij}$ for the unit vector between the same. Recall also the definitions of β_{ij} , A_{ij} , and V_{ij} given above.

Weighted-Volume Derivative Theorem. *The derivative of the weighted volume of the space-filling diagram with state \mathbf{z} is $DW_{\mathbf{z}}(\mathbf{t}) = \langle \mathbf{w}, \mathbf{t} \rangle$, where*

$$\begin{aligned} \begin{bmatrix} \mathbf{w}_{3i+1} \\ \mathbf{w}_{3i+2} \\ \mathbf{w}_{3i+3} \end{bmatrix} &= \sum_j \beta_{ij} A_{ij} (w_{ij} U_{ij} + x_{ij} V_{ij}), \\ w_{ij} &= \left[\frac{\alpha_i + \alpha_j}{2} + \frac{(\alpha_j - \alpha_i)(\varrho_i^2 - \varrho_j^2)}{2\zeta_{ij}^2} \right], \\ x_{ij} &= \frac{2(\alpha_i - \alpha_j)}{3\zeta_{ij}}, \end{aligned} \quad [4]$$

for $0 \leq i < n$. The sum is over all edges $z_i z_j$ in K .

We illustrate the formula with a small example consisting of three spheres that intersect in two points, the projection of which along the shared edge of the power diagram is shown in Fig. 3. In the unweighted case we have $\alpha_i = 1$, $w_{ij} = 1$, and $x_{ij} = 0$ for all i and j . The derivative thus simplifies to a sum of vectors between the centers: $\sum \beta_{ij} A_{ij} U_{ij}$. As shown by Csikós (40), this formula generalizes to dimensions different from three. We note that DW is not everywhere continuous. Specifically, it may not be continuous when two or more spheres coincide or three or more spheres intersect in a common circle (see *Discontinuities*).

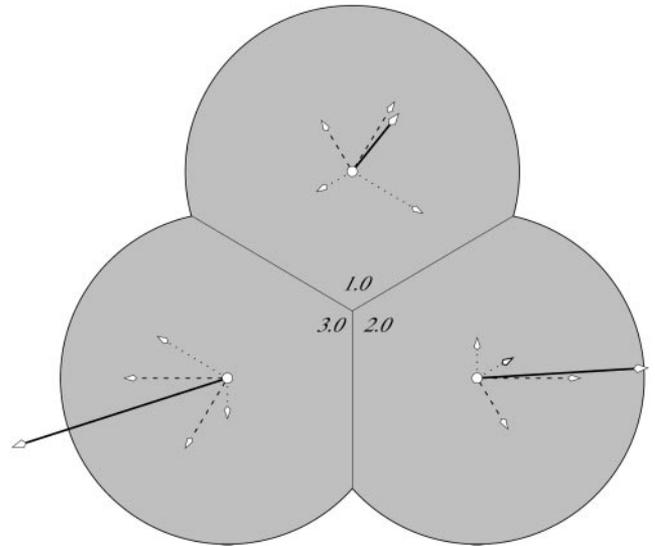


Fig. 3. Three equal-sized balls with different weights of 1.0, 2.0, and 3.0. For each pair we get components of the gradient normal to (dotted) and parallel to (dashed) the separating facet. The solid vectors are the sums of the components as well as the portions of the gradient that correspond to the different balls.

Derivation

In this section we derive the formula claimed in the *Weighted-Volume Derivative Theorem*. It is the sum of contributions from locally direction-preserving and locally distance-preserving components of the motion.

Direction-Preserving Motion. A direction is determined by two spheres and is preserved if one sphere stays put and the other moves along that line. To study the volume derivative for such a motion, we restrict our attention to two spheres S_i and S_j with nonempty intersection S_{ij} and define $F = B_i \cup B_j$. Let ζ_i and ζ_j be the signed distances of the centers from the plane of S_{ij} , as illustrated in Fig. 4. The sum of the two distances is $\zeta_{ij} = \zeta_i + \zeta_j = \|z_i - z_j\|$. We have $\varrho_i^2 - \varrho_j^2 = \zeta_i^2 - \zeta_j^2 = \zeta_{ij}(\zeta_i - \zeta_j)$, and therefore

$$\zeta_i = \frac{\zeta_{ij}}{2} + \frac{\varrho_i^2 - \varrho_j^2}{2\zeta_{ij}}, \quad [5]$$

and

$$\zeta_j = \frac{\zeta_{ij}}{2} + \frac{\varrho_j^2 - \varrho_i^2}{2\zeta_{ij}}. \quad [6]$$

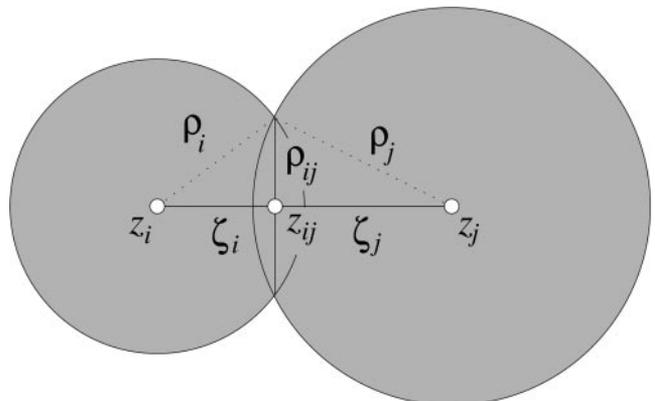


Fig. 4. The two spheres intersect in a circle with center z_{ij} and radius ϱ_{ij} .

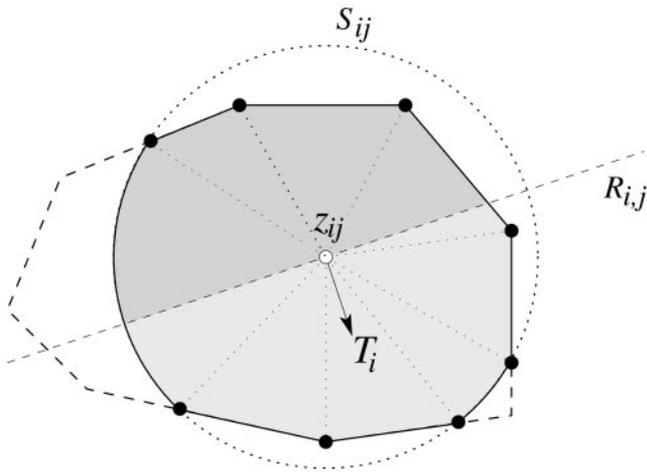


Fig. 5. Infinitesimal motion of a facet between two balls B_i and B_j . The two bounding spheres intersect in the dotted circle S_{ij} with center z_{ij} , perpendicular to the edge $z_i z_j$. The (shaded) facet $P_i \cap P_j \cap F$ between the corresponding two power cells is the intersection of the disk bounded by S_{ij} with the convex polygon common to the two cells. The solid boundary of the facet consists of edges and circular arcs. The only relevant component of the infinitesimal motion of the facet is a rotation around the line $R_{i,j}$ perpendicular to the edge $z_i z_j$ and the motion vector T_i . On one side of this line (dark shaded area), we trade volume of B_i for that of B_j , and on the other side (light shaded area) we do it the other way around.

It is geometrically obvious that the weighted-volume derivative is the area of the disk bounded by S_{ij} times the sum of the weighted derivatives of ζ_i and ζ_j . Because the square radius of the disk is $\varrho_{ij}^2 = \varrho_i^2 - \zeta_i^2$, the area of the disk is

$$A_{ij} = \pi(\varrho_i^2 - \zeta_i^2) = -\frac{\pi}{4\zeta_{ij}^2} \prod (\zeta_{ij} \pm \varrho_i \pm \varrho_j), \quad [7]$$

where the product ranges over the four different ways of assigning signs. The derivatives of ζ_i and ζ_j with respect to ζ_{ij} can be computed from Eqs. 5 and 6, and the weighted-volume derivative is

$$\begin{aligned} \frac{dW}{d\zeta_{ij}} &= A_{ij} \left[\alpha_i \frac{d\zeta_i}{d\zeta_{ij}} + \alpha_j \frac{d\zeta_j}{d\zeta_{ij}} \right] \\ &= A_{ij} \left[\frac{\alpha_i + \alpha_j}{2} + \frac{(\alpha_j - \alpha_i)(\varrho_i^2 - \varrho_j^2)}{2\zeta_{ij}^2} \right]. \end{aligned} \quad [8]$$

Alternatively, we could work analytically, compute the weighted-volume contributions of the two balls by integration, and get $dW/d\zeta_{ij}$ by differentiation. The result is the same.

Distance-Preserving Motion. The distance between the centers of two spheres S_i and S_j is preserved if one rotates about the other. We consider what happens to the facet $P_i \cap P_j \cap F$ that separates the two corresponding cells. As illustrated in Fig. 5, that facet is the intersection of a disk and a convex polygon. The boundary consists of edges and circular arcs. Let T_i normal to U_{ij} be the motion vector of S_i , and let $R_{i,j}$ be the line normal to both vectors that passes through the center z_{ij} of the disk. The infinitesimal motion of the facet consists of a rotation about $R_{i,j}$ composed with a translation along T_i , but only the rotation has a nonzero contribution to the volume derivative. On one side of the line we trade volume of B_i for that of B_j , and on the other side we do it the other way round. The net gain or loss is the accumulation of the contributions of the sectors over the edges and arcs in the boundary of the facet. Here we count a sector over an edge negative if the line defined by the edge separates z_{ij} and the facet.

We first consider an edge xy defined by S_i , S_j , and a third sphere S_k . We compute the derivative of the signed volume swept out by the triangle $z_{ij}xy$ with respect to the rotation angle θ ,

$$\frac{d\text{Vol}_{xy}}{d\theta} = \frac{h_{xy}}{3} \left\langle \frac{T_i}{\zeta_{ij}}, \int_x^y (t - z_{ij}) dt \right\rangle = \frac{h_{xy} \|x - y\|}{3\zeta_{ij}} \langle V_{xy}, T_i \rangle, \quad [9]$$

where $h_{xy} = \delta_{ij,k}$ is the distance from z_{ij} to the line passing through x and y , and $V_{xy} = (x + y)/2 - z_{ij}$ is the vector to the midpoint. Note that the result is two thirds the area of the triangle times the scalar product divided by the distance between the two ball centers.

Second, we consider a circular arc ab , which is part of the circle S_{ij} . We compute the derivative of the volume swept out by the disk sector $z_{ij}ab$ with respect to θ ,

$$\frac{d\text{Vol}_{ab}}{d\theta} = \frac{\varrho_{ij}^2}{3} \left\langle \frac{T_i}{\zeta_{ij}}, \int_a^b (t - z_{ij}) dt \right\rangle = \frac{2\varrho_{ij}^2 \varphi_{ab}}{3\zeta_{ij}} \langle V_{ab}, T_i \rangle, \quad [10]$$

where $2\varphi_{ab}$ is the angle of the arc, and V_{ab} is the average vector from the center to the arc. Again, the result is two thirds the area of the sector times the scalar product divided by the distance between the two ball centers. As explained in *Background and Results*, $V_{ab} = v_{ab}(M_{ab} - z_{ij})$, where M_{ab} is the midpoint of the arc and $v_{ab} = (\sin \varphi_{ab})/\varphi_{ab}$.

Assembling the Relations. Let \mathbf{t} be the velocity vector of the state \mathbf{z} and $T_i = [\mathbf{t}_{3i+1}, \mathbf{t}_{3i+2}, \mathbf{t}_{3i+3}]^T$ the velocity vector of z_i . Because the derivative is linear, we can decompose the motion into components and add the contributions. For every ordered pair i, j , we consider the direction-preserving component $\langle U_{ij}, T_i \rangle U_{ij}$ and the distance-preserving component $T_i - \langle U_{ij}, T_i \rangle U_{ij}$ of T_i .

The contribution of the direction-preserving component is the fraction of the disk B_{ij} that belongs to the power diagram times what is given in Eq. 8. This is $\beta_{ij} A_{ij} w_{ij} \langle U_{ij}, T_i \rangle$, where w_{ij} is given in *Weighted-Volume Derivative Theorem*. The contribution of the distance-preserving component is the weight difference times two thirds the area times the weighted average vector divided by the distance between the ball centers. The factor 2/3 is necessary to change the factor 1/2, which is used in the area computation of sectors, to 1/3, which arises in the volume computation of cones. The contribution is then $x_{ij} \beta_{ij} A_{ij} \langle V_{ij}, T_i \rangle$, where x_{ij} is given in *Weighted-Volume Derivative Theorem*. It is also the weight difference times the sum of contributions over all edges, given in Eq. 9, plus the sum of contributions over all arcs, given in Eq. 10. Adding the terms for all ordered pairs gives

$$DW_{\mathbf{z}}(\mathbf{t}) = \sum_{i=0}^{n-1} \sum_{j \neq i} \beta_{ij} A_{ij} \langle w_{ij} U_{ij} + x_{ij} V_{ij}, T_i \rangle \quad [11]$$

for the weighted-volume derivative. We can write this more succinctly as $DW_{\mathbf{z}}(\mathbf{t}) = \langle \mathbf{w}, \mathbf{t} \rangle$, where \mathbf{w} is as defined in *Weighted-Volume Derivative Theorem*. We have $\beta_{ij} = 0$ unless z_{ij} is an edge of K , which implies that the summation can be limited to the edges in K .

Discontinuities

The weighted-volume derivative, DW , maps a state \mathbf{z} to the gradient $\mathbf{w} = \mathbf{w}_{\mathbf{z}}$ of W at \mathbf{z} . From the formula in *Weighted-Volume Derivative Theorem*, we glean that DW is continuous except at states at which either two or more spheres coincide or three or more spheres intersect in a common circle. The set of such states is a $(3n - 3)$ -dimensional subset of \mathbb{R}^{3n} , which should be compared with the $(3n - 1)$ -dimensional subset at which the weighted-area derivative is discontinuous. To see that there are only these two types of discontinuities, we note that A_{ij} is everywhere continuous, x_{ij} , w_{ij} ,

and U_{ij} are continuous except when $z_i = z_j$, and β_{ij} and V_{ij} are continuous as long as the facet $P_i \cap P_j \cap F$ varies continuously. States at which that facet changes abruptly either have $z_i = z_j$ or contain at least one sphere, other than S_i and S_j , that contains the circle S_{ij} .

Every state \mathbf{z} at which DW is discontinuous has the property that every open neighborhood contains states with combinatorially different dual complexes. In other words, when \mathbf{z} passes through a discontinuity of DW , then the dual complex passes a moment at which it may undergo a combinatorial change. Note, however, that not every combinatorial change in the dual complex corresponds to a discontinuity in DW .

Implementation

To compute the volume and its derivatives with respect to the coordinates of the center of the balls, we first need the dual complex of the space-filling diagram. For that purpose, we have written a new version of the alpha-shape software, ALPHAVOL, specific to molecular simulation applications. ALPHAVOL first performs a regular triangulation of the set of centers of the balls by using the algorithm described by Edelsbrunner and Shah (41). Singularities are resolved by using the Simulation of Simplicity (39), whereas numerical imprecisions are avoided by using arbitrary precision arithmetic. The software is kept efficient through the application of a floating-point filter. The dual complex is computed by using the same care for singularities and numerical uncertainties. ALPHAVOL includes implementations of the volume and weighted-volume derivative formulas. Computation of the volume and its derivatives of a 250-residue protein by using ALPHAVOL requires ≈ 0.37 sec on a 1,000-Mhz Pentium processor. The 0.37 sec roughly breaks down to 0.11 sec for computing the regular triangulation, 0.03 sec for generating the dual complex, 0.17 sec to compute the weighted volume, and 0.06 sec to compute the weighted-volume derivatives.

The weighted-volume derivative formula, Eq. 4, was tested against the numerically estimated derivatives,

$$\Delta W_{3i+j} = \frac{W(\mathbf{z}_{3i+j} + \delta) - W(\mathbf{z}_{3i+j} - \delta)}{2\delta}, \quad [12]$$

where \mathbf{z}_{3i+j} is the j th coordinate of the i th atom center. The average relative difference between analytical and numerical derivatives was computed as a relative root mean square error:

$$\mu(W) = \frac{\sqrt{\sum_{i,j} (\mathbf{w}_{3i+j} - \Delta W_{3i+j})^2}}{\sqrt{\sum_{i,j} (\Delta W_{3i+j})^2}}. \quad [13]$$

Using $\delta = 0.0001$, we find $\mu(W)$ equal on the average to 5.1×10^{-8} and never larger than 9×10^{-8} for a collection of 100 proteins varying in size from 50 residues or ≈ 400 atoms to 500 residues or $\approx 4,000$ atoms. The good agreement between the analytical and numerical derivatives of the weighted volume shows that our approach and implementation are correct. ALPHAVOL is available online at <http://biogeometry.duke.edu/software/proshape>.

Discussion

The software ALPHAVOL, which computes the weighted volume of a union of balls and its derivatives with respect to the coordinates of the centers of the balls, provides a fast, accurate, and robust method for computing the interaction of water with nonpolar atoms of a molecule in a hydration shell model (19). ALPHAVOL implements analytical formulas for computing the weighted volume and its derivatives and explicitly deals with the problem of discontinuities. We have made preliminary steps toward inserting ALPHAVOL into the molecular dynamics software ENCAD (42) and GROMACS (43), but it is too early to say anything about the corresponding results.

Measuring the volume occupied by individual atoms in macromolecular structures has been the subject of scientific investigation for several decades. In recent years, this interest has increased, because volumes of specific atom types are used in a wide range of implicit solvent models. One of the most widespread uses of atomic volumes in molecular mechanics calculations is in the generalized Born model of electrostatic solvation (14), where the atom volume appears in the calculation of the desolvation of one atom by another atom. In the present implementation of the generalized Born model, ball overlaps are ignored, corrected by scaling factors, or dealt with by using a Gaussian representation of the ball. We expect that the formalism introduced in this article for computing volumes and their derivatives that explicitly and accurately deal with ball overlaps should prove useful for the generalized Born model.

Voronoi polyhedra have been widely used to compute the volume occupied by the atoms of proteins (4, 6–9) and recently DNA (5) in solution. One difficulty in applying the Voronoi-polyhedra approach to macromolecular structures is that the treatment of atoms at the surface requires special attention because their power cells are infinite. Possible solutions to this problem are the construction of a layer of solvent molecules around the macromolecule (44) or limiting the power diagram to the space-filling diagram (29, 33). This article describes the latter approach and extends it by providing exact expression for the derivatives of the atomic volume. These derivatives are expected to be useful for including potential based on packing in molecular mechanics or molecular dynamics calculation.

The mathematical tools used in this paper to express the volume derivative of a space-filling diagram have also been used to compute the weighted surface-area derivatives (R. Bryant, H.E., P.K., and M. Levitt, unpublished data).

Appendix

In this section we explain how to compute the fractions β_{ij} and the weighted-average vectors V_{ij} needed to express the weighted-volume derivative. The descriptions are relevant only in the context of implementing the derivative once the dual complex K of the space-filling diagram is known. We begin by introducing a notion of neighborhood within K : The *star* of a simplex $\tau \in K$ is the collection of simplices that contain τ , including τ itself, and the *link* of τ is the collection of faces of simplices in the star that are disjoint of τ . Writing the face relation between simplices with a smaller equal symbol, we can express these definitions more formally:

$$\text{St}\tau = \{\nu \in K \mid \tau \leq \nu\}, \quad [14]$$

$$\text{Lk}\tau = \{\omega \in K \mid \omega \leq \nu \in \text{St}\tau \text{ and } \omega \cap \tau = \emptyset\}. \quad [15]$$

For example, the star of an interior edge contains the edge together with a ring of triangles and tetrahedra around the edge. The link is the cycle of vertices and edges that surrounds the ring away from the interior edge; for each triangle in the star it contains the opposite vertex, and for each tetrahedron it contains the opposite edge.

Computing β_{ij} . Recall that B_{ij} is the disk spanned by the circle S_{ij} , which is the intersection of the spheres S_i and S_j . Formally, B_{ij} is the set of points $x \in \mathbb{R}^3$ for which $\pi_i(x) = \pi_j(x) \leq 0$. The number we are after is the fraction of that disk that belongs to the power diagram,

$$\beta_{ij} = \frac{\text{area}(P_i \cap P_j \cap F)}{\text{area}(B_{ij})}. \quad [16]$$

Given the dual complex K , we can use inclusion–exclusion to compute β_{ij} from simple pieces of the disk that are dual to simplices in the star of the edge z_{ij} . Let B_{ij}^k be the segment of the disk on z_k 's

side of the bisector. More formally, B_{ij}^k is the set of points $x \in B_{ij}$ for which $\pi_k(x) \leq \pi_i(x) = \pi_j(x)$. The general results in ref. 32 imply that β_{ij} can be computed by subtracting the segments B_{ij}^k from the disk B_{ij} and adding back their pairwise intersections,

$$\beta_{ij} = 1 - \frac{\sum_k \text{area}(B_{ij}^k) - \sum_{k,l} \text{area}(B_{ij}^k \cap B_{ij}^l)}{\pi \rho_{ij}^2}, \quad [17]$$

where ρ_{ij} is the radius of B_{ij} . The first sum ranges over all vertices z_k , and the second ranges over all edges $z_k z_l$ in the link of $z_i z_j$ in K .

Computing V_{ij} . Recall that V_{ij} is the weighted-average vector from the center z_{ij} of the disk B_{ij} to the boundary of the facet that is the portion of the disk belonging to the Voronoi diagram. The weight of each vector is its infinitesimal contribution to the area of the facet, which is constant along each edge and each arc. Hence, V_{ij} can be computed as the weighted sum of the average vectors of the edges xy and the arcs ab of the facet,

$$V_{ij} = \frac{\sum_{xy} h_{xy} \|x - y\| V_{xy} + 2 \sum_{ab} \rho_{ij}^2 \varphi_{ab} V_{ab}}{\sum_{xy} h_{xy} \|x - y\| + 2 \sum_{ab} \rho_{ij}^2 \varphi_{ab}}, \quad [18]$$

where h_{xy} is the distance from z_{ij} to the line spanned by the edge xy , and $2\varphi_{ab}$ is the angle of the arc ab . The first sums in the numerator and the denominator range over all edges xy , and the second sums range over all arcs ab of the facet. We note that the denominator is twice the area of the facet, which is $2\beta_{ij}A_{ij}$. We thus focus on the numerator, which we will express as an alternating sum using again inclusion–exclusion over the link of the edge $z_i z_j$ in K . We start with the average vector to the boundary of the entire disk, which is $N_{ij} = 0$. From this we subtract the contributions of disk segments and we add back contributions of pairs of disk segments.

Case 1 (disk segments). Let x and y be the points at which the sphere S_k intersects the circle S_{ij} . The disk segment B_{ij}^k is thus bounded by the edge xy and the arc ab connecting $a = x$ with $b = y$ along the circle as illustrated in Fig. 6 *Left*. We add the contribution of the edge and subtract the contribution of the arc. The former is twice the area of the triangle times the vector to the midpoint, which is $2\rho_{ij}^2 \cos \varphi_{ab} \sin \varphi_{ab} V_{xy} = \rho_{ij}^2 \sin(2\varphi_{ab}) V_{xy}$. Assuming $\varphi_{ab} \neq \pm 90^\circ$, the average vector of the arc can be computed as $V_{ab} = V_{xy}(\tan \varphi_{ab})/\varphi_{ab}$. This expression is undefined if $\varphi_{ab} = \pm 90^\circ$, in which case we may go back to Eq. 2 and get $V_{ab} = (2/\pi)(M_{ab} - z_{ij})$, where M_{ab} is the midpoint of the arc. The weight of this vector is twice the area of the disk sector defined

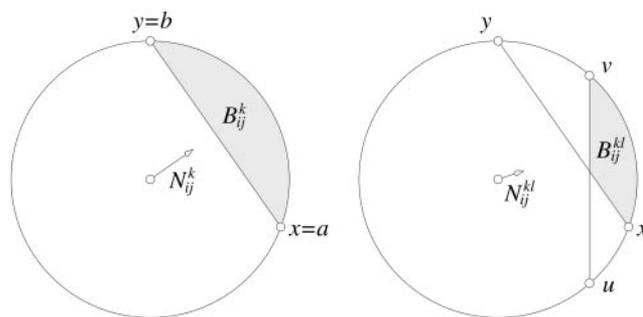


Fig. 6. The weighted average vector N_{ij}^k for a disk segment (*Left*) and the vector N_{ij}^{kl} for the intersection of a pair of disk segments (*Right*). Using inclusion–exclusion, we can compute the contribution of the union of the two disk segments as $N_{ij}^k + N_{ij}^l - N_{ij}^{kl}$ (see Eq. 20).

by ab and is given by $2\rho_{ij}^2 \varphi_{ab}$. The accumulated contribution of the disk segment therefore is

$$N_{ij}^k = \rho_{ij}^2 (2\varphi_{ab} V_{ab} - \sin(2\varphi_{ab}) V_{xy}), \quad [19]$$

where we take the contribution of the arc positive and that of the edge negative, in preparation for subtracting N_{ij}^k in the final sum. **Case 2 (pairs of disk segments).** Let x, y and u, v be the endpoints of the edges and arcs bounding the disk segments B_{ij}^k and B_{ij}^l , and define $B_{ij}^{kl} = B_{ij}^k \cap B_{ij}^l$. We only need to consider the case in which the endpoints alternate along the circle. As shown in Fig. 6 *Right*, the intersection point of the two edges lies inside the disk, and the boundary of B_{ij}^{kl} consists of one arc and two edges that both end at that intersection point. Let N_{ij}^{kl} be the weighted sum of average vectors taking the contribution of the arc positive and those of the edges negative, as before.

We get the weighted average vector over the boundary of the facet $P_i \cap P_j \cap F$ as

$$V_{ij} = \frac{N_{ij} - \sum_k N_{ij}^k + \sum_{k,l} N_{ij}^{kl}}{2\beta_{ij}A_{ij}}, \quad [20]$$

where $N_{ij} = 0$, the first sum ranges over all vertices z_k , and the second sum ranges over all edges $z_k z_l$ in the link of $z_i z_j$ in K .

This work was supported in part by National Science Foundation Grant CCR-00-86013. P.K. acknowledges support from the Department of Energy (DE-F603-95ER62135).

1. Privalov, P. L. & Gill, S. G. (1988) *Adv. Protein Chem.* **39**, 191–234.
2. Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.
3. Morild, E. (1981) *Adv. Protein Chem.* **34**, 93–166.
4. Tsai, J., Taylor, R., Chothia, C. & Gerstein, M. (1999) *J. Mol. Biol.* **290**, 253–266.
5. Nadassy, K., Tomás-Oliveira, I., Alberts, I., Janin, J. & Wodak, S. J. (2001) *Nucleic Acids Res.* **29**, 3362–3376.
6. Tsai, J. & Gerstein, M. (2002) *Bioinformatics* **18**, 985–995.
7. Pontius, J., Richelle, J. & Wodak, S. J. (1996) *J. Mol. Biol.* **264**, 121–136.
8. Gerstein, M., Sonnhammer, E. L. L. & Chothia, C. (1994) *J. Mol. Biol.* **236**, 1067–1078.
9. Richards, F. M. (1974) *J. Mol. Biol.* **82**, 1–14.
10. Rashin, A. A., Iofin, M. & Honig, B. (1986) *Biochemistry* **25**, 3619–3625.
11. Pettitt, B. M. & Karplus, M. (1985) *Chem. Phys. Lett.* **121**, 194–201.
12. Kang, Y. K., Nemethy, G. & Scheraga, H. A. (1987) *J. Phys. Chem.* **91**, 4105–4109.
13. Still, W. C., Tempczyk, A., Hawley, R. C. & Hendrikson, T. (1990) *J. Am. Chem. Soc.* **112**, 6127–6129.
14. Bashford, D. & Case, D. A. (2000) *Annu. Rev. Phys. Chem.* **51**, 129–152.
15. Eisenberg, D. & McLachlan, A. (1986) *Nature* **319**, 199–203.
16. Ooi, T., Oobatake, M., Nemethy, G. & Scheraga, H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3086–3090.
17. Simonson, T. & Brünger, A. T. (1994) *J. Phys. Chem.* **98**, 4683–4694.
18. Lum, K., Chandler, D. & Weeks, J. D. (1999) *J. Phys. Chem. B* **103**, 4570–4577.
19. Gibson, K. & Scheraga, H. (1967) *Proc. Natl. Acad. Sci. USA* **58**, 420–427.
20. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379–400.
21. Rowlinson, J. S. (1963) *Mol. Phys.* **6**, 517–524.
22. Pavani, R. & Ranghino, G. (1982) *Comput. Chem.* **6**, 133–135.

23. Gavezotti, A. (1983) *J. Am. Chem. Soc.* **105**, 5220–5225.
24. Richmond, T. J. (1984) *J. Mol. Biol.* **178**, 63–89.
25. Connolly, M. L. (1985) *J. Am. Chem. Soc.* **107**, 1118–1124.
26. Gibson, K. & Scheraga, H. (1987) *Mol. Phys.* **62**, 1247–1265.
27. Petitjean, M. (1994) *J. Comput. Chem.* **15**, 507–523.
28. Voronoi, G. F. (1908) *J. Reine Angew. Math.* **134**, 198–287.
29. Gellatly, B. J. & Finney, J. L. (1982) *J. Mol. Biol.* **161**, 305–322.
30. Goede, A., Preissner, R. & Frömmel, C. (1997) *J. Comput. Chem.* **18**, 1113–1123.
31. Kratky, K. W. (1978) *J. Phys. A Math. Gen.* **11**, 1017–1024.
32. Edelsbrunner, H. (1995) *Discrete Comput. Geom.* **13**, 415–440.
33. Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P. V. & Subramaniam, S. (1998) *Proteins Struct. Funct. Genet.* **33**, 1–29.
34. Kundrot, C. E., Ponder, J. W. & Richards, F. M. (1991) *J. Comput. Chem.* **12**, 402–409.
35. Gogonea, V. & Osawa, E. (1994) *J. Mol. Struct. THEOCHEM* **311**, 305–324.
36. Gogonea, V. & Osawa, E. (1995) *J. Comput. Chem.* **7**, 817–842.
37. Perrot, G., Cheng, B., Gibson, K. D., Vila, J., Palmer, K. A., Nayeem, A., Maigret, B. & Scheraga, H. A. (1992) *J. Comput. Chem.* **13**, 1–11.
38. Wawak, R. J., Gibson, K. D. & Scheraga, H. A. (1994) *J. Math. Chem.* **15**, 207–232.
39. Edelsbrunner, H. & Mücke, E. P. (1990) *Am. Chem. Soc. Trans. Graphics* **9**, 66–104.
40. Csikós, B. (1998) *Discrete Comput. Geom.* **20**, 449–461.
41. Edelsbrunner, H. & Shah, N. R. (1996) *Algorithmica* **15**, 223–241.
42. Levitt, M., Hirshberg, M., Sharon, R. & Daggett, V. (1995) *Comput. Phys. Commun.* **91**, 215–231.
43. Lindahl, E., Hess, B. & van der Spoel, D. (2001) *J. Mol. Mod.* **7**, 306–317.
44. Finney, J. L. (1975) *J. Mol. Biol.* **96**, 721–732.