

International Journal of Computational Geometry & Applications  
© World Scientific Publishing Company

## LAGUERRE-INTERSECTION METHOD FOR IMPLICIT SOLVATION

MICHELLE HATCH HUMMEL\*

*Department of Applied Mathematics and Statistics, Stony Brook University  
Stony Brook, NY 11794, USA  
mhumme@sandia.gov*

BIHUA YU

*Department of Applied Mathematics and Statistics, Stony Brook University  
Stony Brook, NY 11794, USA  
bihua.yu@stonybrook.edu*

CARLOS SIMMERLING

*Department of Chemistry and Laufer Center, Stony Brook University  
Stony Brook, NY 11794, USA  
carlos.simmerling@stonybrook.edu*

EVANGELOS A. COUTSIAS †

*Department of Applied Mathematics and Statistics and Laufer Center, Stony Brook University  
Stony Brook, NY 11794, USA  
evangelos.coutsias@stonybrook.edu*

Received May 02, 2016

Revised (July 15, 2017)

Communicated by Michelle Hatch Hummel

Explicit solvent molecular dynamics simulations of a macromolecule are slow as the number of solvent atoms considered typically increases by order of magnitude. Implicit methods introduce surface-dependent corrections to the force field, gaining speed at the expense of accuracy. Properties such as molecular interface surfaces, volumes and cavities are captured by Laguerre tessellations of macromolecules. However, Laguerre cells of exterior atoms tend to be overly large or unbounded. Our method, the inclusion-exclusion based Laguerre-Intersection method, caps cells in a physically accurate manner by considering the intersection of the space-filling diagram with the Laguerre tessellation. We optimize an adjustable parameter, the weight, to ensure the areas and volumes of capped cells exposed to solvent are as close as possible, on average, to those computed from equilibrated explicit solvent simulations. The contact planes are radical planes, meaning that as the solvent weight is varied, interior cells remain constant. We test the consistency of our model using a high-quality trajectory of HIV-protease, a dimer

\*Currently at Sandia National Laboratories, Albuquerque, NM 87123 USA

†Corresponding author.

2 *Hummel, Yu, Simmerling, Coutsiias*

with flexible loops and open-close transitions. We also compare our results with interval-arithmetic Gauss-Bonnet based method. Optimal solvent parameters quickly converge, which we use to illustrate the increased fidelity of the Laguerre-Intersection method over two recently proposed methods as compared to the explicit model.

*Keywords:* Laguerre tessellation; Voronoi tessellation; Implicit Solvation

## 1. Introduction

Laguerre tessellation (also known as the weighted Voronoi or power tessellation) is a well known mathematical tool used to partition the space containing data points into cells. Each data point is assigned the region which is closest to the given data point (generator of the cell) by the “power distance”. Laguerre methods have been widely used in biochemistry to find volumes of atoms, molecules, and amino acid residues (which we call simply “residues”) in proteins. They are useful for cavity location and in studying packing and deformation of proteins. They may also be used in protein structure prediction to check the quality of predicted structures. For example, residue volume may be considered an “intrinsic” property of amino acid type and is thus a predictable or checkable quantity<sup>12</sup>. Laguerre surfaces have been used as a quick and parameter free way to measure accessibility of atoms and residues for quantifying exposure of a molecule with solvent<sup>12</sup>. Residue contacts, which are important for protein structure and stability, have been studied using Laguerre surfaces. Interresidue contact surface areas are useful for protein structure analysis and prediction, and in studying structure-function relationships. Preferential contacts between amino acid species and atom-atom contact frequencies have also been found<sup>13–26</sup>. Differences in contact areas between model and reference structures have been used as a scoring function and for benchmarking protein structure prediction methods<sup>29</sup>. Contact surfaces have also been used to characterize interactions between multiple proteins, or proteins and ligands, with important applications in protein docking and formation of complexes<sup>27–30</sup>.

Surface area calculations play an increasingly important role in molecular dynamics simulations of macromolecules. For example, explicit solvent molecular dynamics simulations of a macromolecule are slow. The solute is imbedded in an adequately large periodic box filled with solvent and the system is simulated for a sufficiently long time so solvent molecules achieve thermal equilibrium. This typically increases the number of atoms in the system by an order of magnitude and furthermore the solvent viscosity significantly slows global dynamics such as protein folding<sup>36</sup>. Using implicit solvent in molecular dynamics is an attractive alternative due to the reduced number of atoms in the simulation, along with the reduced viscosity due to the instantaneous relaxation of solvent following solute conformational changes. Implicit solvent calculations also play an important role in the post-processing of MD trajectories that were carried out in explicit water as a means to estimate binding affinities (the MM-PBSA method<sup>21</sup>). Here implicit solvent is often advantageous since it directly provides an estimate of the free energy of (de)solvation, as compared to the enthalpy that is obtained using explicit solvent. This can be

valuable in some cases where solvation free energy is needed, since obtaining this value in explicit solvent requires proper averaging over all possible explicit solvent configurations for a given solute configuration, obtaining the entropic contribution as well as the appropriately averaged enthalpy. Popular implicit solvent models include the Poisson-Boltzmann and Generalized Born methods, however these model only the polar contribution to solvation energy. A more accurate treatment involves also including the nonpolar (hydrophobic and dispersion) contributions to solvation, which are typically considered as approximately proportional to the solvent accessible surface area of the solute. Several methods have been developed and are commonly used<sup>7 34 35</sup>. The speed and accuracy of these methods is vital to successful application, but due to the computational overhead many recent studies of protein folding omit nonpolar solvation despite the importance of the hydrophobic effect<sup>28</sup>.

The Laguerre cell is the region which is closest to the generator of the cell (atom or residue center, for example) by the power distance. The power distance is an appropriate quantity to use as larger areas are assigned to data points with larger weights (larger atoms or residues, for example). The Laguerre cell of a data point depends on its neighbors. This poses a problem for systems that are typically found in a medium such as solvent but for which the surrounding medium is not explicitly known. In this situation, Laguerre cells of points on the convex hull are unbounded and the cells of points otherwise in contact with the surrounding medium tend to be too large. To arrive at a geometrical description of a macromolecule that is both independent of the rapidly changing detailed structure of the surrounding solvent molecules (typically water) while at the same time encoding an "average" geometrical profile as seen by the medium is a challenge that must be overcome in order to enhance the accuracy of implicit solvation modeling.

Researchers have addressed this difficulty in a variety of ways. Some consider only cells in the bulk of the protein which are not affected by the surrounding (unknown) environment. Others surround the structure by a layer of water or an artificial environment of spheres with size equal to the average amino acid size<sup>14 31</sup>. However, this causes the size of the problem to increase, typically by an order of magnitude. Soyer et al., bounded Laguerre cells by only considering Laguerre facet vertices from tetrahedra of a small enough size<sup>33</sup>. Still cells near the boundary tend to be elongated or have fewer facets than those in the bulk of the protein. Cazals method 'Intervor' constructs interface surfaces by considering facets of dual Delaunay edges in the space filling diagram and discarding any edges that belong to large tetrahedra<sup>6 4</sup>. Discarding these edges causes pertinent information to be lost.

Another class of models prunes and truncates certain Laguerre facets. Mahdavi et al. construct intermolecular interface surfaces by considering truncated facets in the union of extended convex hulls<sup>24 25</sup>. The convex hulls are extended in a fashion to mimic a 1.4 Angstrom solvent radius. Ban and coworkers construct a series of interface surfaces based on a sequence of alpha complexes and truncate facets to

within alpha complex tetrahedra, but no attempt is made to mimic the effect of the solvent<sup>1 16</sup>.

McConkey et al. construct Laguerre-like cells by considering the surfaces of extended radical contact planes between neighboring atoms within a cutoff distance and the surface of an expanded sphere<sup>26</sup>. However, the algorithm miscalculates Laguerre volumes when the atom center lies outside its cell (what we term an “engulfing contact”), a situation that our algorithm treats correctly (See Section 3.2). Cazals et al. in their paper “Computing the volume of a union of balls: a certified algorithm”<sup>5</sup> provide a formalization of the McConkey’s algorithm which also correctly calculates engulfing contact volumes. While these algorithms compute the restriction of Laguerre cells to the space-filling model, there are three key differences with our model. McConkey’s and Cazals’ algorithms are based on the Gauss-Bonnet theorem whereas our algorithm is an inclusion-exclusion model which is an extension of Edelsbrunner’s algorithm for calculating atomic and molecular surface areas and volumes<sup>9</sup>. Inclusion-exclusion contributions extended to per-restriction quantities are given in equations (12)-(28) and (34)-(36). The inclusion-exclusion method is advantageous in that cycles of intersection points on the surface of the sphere do not need to be calculated which decreases algorithm complexity. The second key difference is that the former algorithms expand nonsolvent atoms by a solvent radius of 1.4 Å while our algorithm expands the squared radius. Expanding cells’ radii means that cells are bounded by extended radical planes rather than radical planes. This poses problems since the cells of atoms in the bulk, which should not be affected by the cutoff distance, are affected. Furthermore, cells of small atoms completely disappear for cutoff distances as small as 1.4 Å. The third difference is that in order to achieve double precision accuracy Cazals’ resorts to interval arithmetic and/or extended precision (“exact”) arithmetic which increases runtime by a factor of ten over the floating point models. As we are interested in calculating Laguerre-Intersection quantities at each step in a molecular dynamics (MD) simulation, speed is of utmost importance, and we can achieve desired accuracy using floating point operations. Remarkably, the accuracy of the results achieved by our method in the majority of cases is comparable to the extended precision results of the GB method, and highly more accurate than the GB method employing double precision arithmetic.

The proposed Laguerre-Intersection cell algorithm<sup>19</sup>, considers the intersection of Laguerre cells with expanded atoms. The contact planes are the radical planes, which means that as the solvent weight is varied, Laguerre cells stay constant. This method simulates the environment better than using the extended radical plane. The Laguerre cells are capped in a physical manner that enables the study of boundary facets, rather than truncation by convex hulls.

We calculate molecule specific optimal solvent parameters from explicit solvent HIV protease trajectory data. The HIVP dimer has two flexible loops covering the active site that exhibit open-closed dynamics<sup>17</sup>. We generated a 100 ns trajectory of HIV in explicit water to use as reference data (see Results). Despite the con-

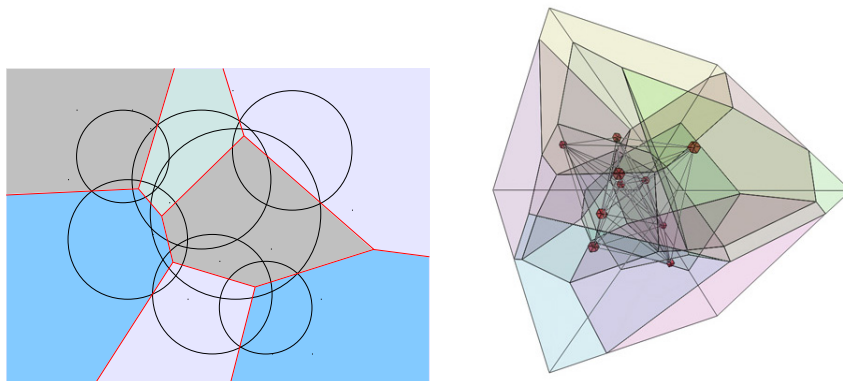


Fig. 1. 2d and 3d Laguerre diagram: 2d illustrates how the radical plane partitions space between atoms of different radii; 3d courtesy of Frederick Vanhoutte.

figurational variability, optimal solvent parameters converge quickly which we use to predict Laguerre quantities in the remainder of the trajectory. We show that these predicted quantities are closer to those found using explicit models than two current alternative methods. Further work is required to determine optimal solvent parameters for an arbitrary molecule.

## 2. Background

### 2.1. Laguerre and Laguerre-Intersection cells

Consider a molecule represented by a set of spheres or atoms  $\mathcal{A} \subset \mathbb{R}^3 \times \mathbb{R}$ . For  $p_i \in \mathcal{A}$  we write  $p_i = (p'_i, w_i)$  where  $p'_i$  is the center of the atom and  $w_i = r_i^2 = p''_i$  is the weight or squared radius of the atom. The Laguerre cell,  $L_i$ , of atom  $i$  is defined to be

$$L_i = \{x' \in \mathbb{R}^3 \quad : \quad |p'_i - x'|^2 - w_i \leq |p'_j - x'|^2 - w_j \quad \text{for all } p_j \in \mathcal{A}\} \quad (1)$$

and is a convex polytope (Fig. 1).

The quantity  $|p'_i - x'|^2 - w_i$  is called the power distance between  $p_i$  and  $x'$  and is written as

$$\pi(p_i, x') = |p'_i - x'|^2 - w_i. \quad (2)$$

A Laguerre cell is the set of points whose nearest neighbor by the power distance is  $p_i$ , and each Laguerre facet lies on the plane which is equi-powerdistant between two points and which is called the radical plane.

The weights of the atoms in  $\mathcal{A}$  may be increased or decreased by a certain amount  $w$  which we call the solvent weight. In this section, we call this modified set  $\mathcal{A}(w)$  where  $p_i(w) \in \mathcal{A}(w)$  has  $p'_i(w) = p'_i$  and  $w_i(w) = w_i + w$ . The Laguerre

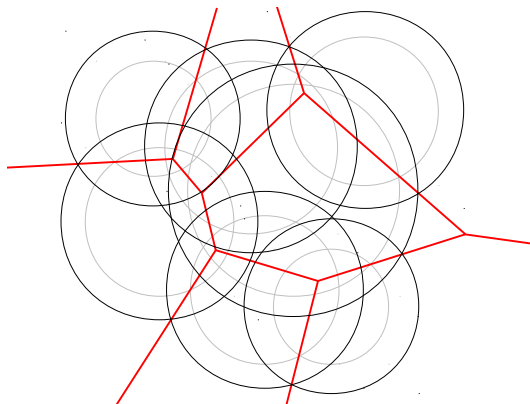


Fig. 2. Red bounds Laguerre cells of  $\mathcal{A}(w)$  (shown in black) and  $\mathcal{A}(0)$  (shown in gray). Note that the cells are the same for both weights. This is the reason that varying the weight rather than the radius in the space filling diagram gives better agreement between Laguerre-Intersection quantities and Laguerre quantities found using explicit solvent (See Section 5.5).

cell of  $p_i(w)$  is defined as

$$L_i(w) = \{x' \in \mathbb{R}^3 : |p_i(w) - x'|^2 - w_i(w) \leq |p_j(w) - x'|^2 - w_j(w) \quad \forall p_j(w) \in \mathcal{A}(w)\}. \quad (3)$$

Since  $L_i(w) = L_i(0)$  for all  $w$  (See Fig. 2) we simply write the Laguerre cell of atom  $i$  as  $L_i$ .

The Laguerre diagram,  $\mathcal{L}(\mathcal{A})$ , of  $\mathcal{A}$  is the collection of all Laguerre cells and their faces which we call Laguerre facets, segments, and nodes. A Laguerre facet is the intersection of two Laguerre cells and is a subset of the plane which is equidistant from the generator of the two cells. A Laguerre segment is the intersection of at least three Laguerre cells, and a Laguerre node is the intersection of at least four Laguerre cells. For  $p_i \in \mathcal{A}$  define

$$B_i(w) = \{x \in \mathbb{R}^3 : |p'_i - x|^2 - w_i(w) \leq 0\}. \quad (4)$$

The space filling model of  $\mathcal{A}$  with solvent weight  $w$  is defined as

$$\mathcal{B}(w) = \bigcup B_i(w) \quad (5)$$

We also define the following quantities:

- $LV_i$ : Volume of the Laguerre cell of atom  $i$ .
- $LS_i$ : Surface area of the Laguerre facets of atom  $i$ .

A residue is the collection of atoms in a single amino acid unit in the protein. The Laguerre volume of residue  $j$  (residual volume) is the sum of Laguerre volumes of atoms in residue  $j$ . The interresidue Laguerre surface area between residues  $j$  and  $k$  is the sum of areas of Laguerre facets which lie between atoms in residue  $j$  and atoms in residue  $k$ .

## 2.2. Laguerre diagram and regular tetrahedrization

The regular tetrahedrization is dual to the Laguerre diagram  $\mathcal{L}(\mathcal{A})$ . There is a one to one correspondence between the  $(3 - k)$ -faces in  $\mathcal{L}(\mathcal{A})$  and the  $k$ -simplices in  $\mathcal{T}(\mathcal{A})$ . Each node in  $\mathcal{T}(\mathcal{A})$  corresponds to a tetrahedron in  $\mathcal{T}(\mathcal{A})$  whose vertices are equipowerdistant to the node, which we call the characteristic point of the tetrahedron. Each segment in  $\mathcal{L}(\mathcal{A})$  corresponds to a triangle in  $\mathcal{T}(\mathcal{A})$  whose vertices are equipowerdistant to the power segment. For each facet in  $\mathcal{L}(\mathcal{A})$  there is an edge in  $\mathcal{T}(\mathcal{A})$  whose vertices are equipowerdistant to the power facet. Each cell in  $\mathcal{L}(\mathcal{A})$  corresponds to a point in  $\mathcal{T}(\mathcal{A})$ , namely the generator of the cell. We call a point whose center is on the convex hull of  $\mathcal{T}(\mathcal{A})$  exterior to the tetrahedrization. Otherwise a point is called interior to  $\mathcal{T}(\mathcal{A})$ . An edge is called exterior if both of its vertices are exterior to  $\mathcal{T}(\mathcal{A})$ . An edge is called interior if at least one of its vertices is interior. Note that the Laguerre cell of an exterior point is unbounded whereas the Laguerre cells of interior points are bounded. The Laguerre facets corresponding to exterior edges are unbounded whereas Laguerre facets of interior edges are bounded.

In this paper we take for granted robust algorithms for the calculation of the regular tetrahedrization and the  $\alpha$ -complex which is a subset of the the tetrahedrization.

## 3. Methods

### 3.1. Computation of Laguerre surface areas of interior facets

Interior Laguerre cells are bounded by Laguerre facets. Each of these facets correspond to an edge in the regular tetrahedrization. We call an edge  $e_{ij}$  if the vertices of that edge are  $p'_i$  and  $p'_j$  and call the Laguerre facet corresponding to that edge  $L_{ij}$ . The facets are convex polygons whose vertices are nodes in the Laguerre diagram, namely the characteristic points of the tetrahedra which surround that given edge. If a Laguerre cell has  $n$  facets, the volume may be divided into  $n$  pieces, one for each of its facets. The volume of the Laguerre cell of  $i$  corresponding to the edge  $e_{ij}$  is written  $LV_{ij}^{(i)}$  and the volume of the Laguerre cell of  $j$  corresponding to the edge is written  $LV_{ij}^{(j)}$ . We call the ordered list of tetrahedra which surround a given edge, i.e. the ordered list of tetrahedra in the edge's star, a *tetrahedra ring* or *tetraring*. A tetraring is called complete if the ring closes, otherwise it is called incomplete. Note that the tetraring of an interior edge is complete, whereas the tetraring of an exterior edge is incomplete (Fig. 3). In this section, we assume that the protein is surrounded by a layer of solvent. This means that all nonsolvent atoms are interior to the regular tetrahedrization as well as all edges which contain a nonsolvent atom as a vertex. We call these "nonsolvent edges". The computation of the Laguerre volumes and surfaces of each nonsolvent atom proceeds as follows:

- Regular tetrahedrization of all atoms is computed
- Characteristic points of tetrahedra are found
- For each nonsolvent edge

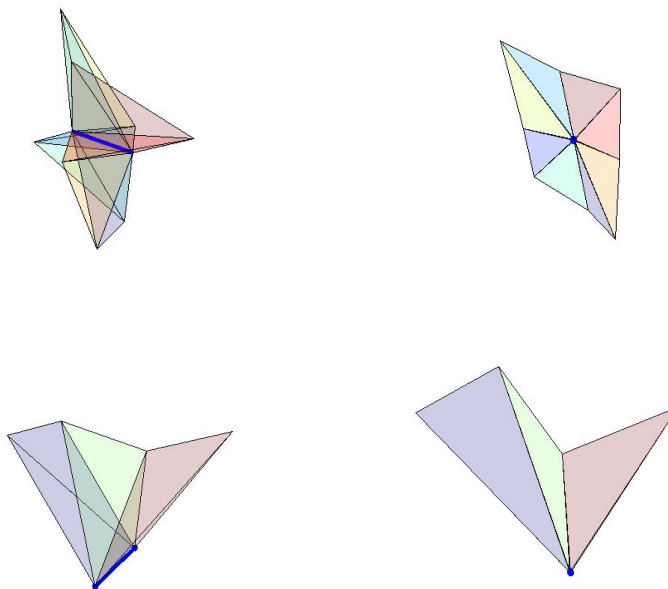


Fig. 3. Top row: Side and perpendicular views of the complete tetrahedral arrangement of an interior edge (shown in dark blue). Bottom row: Side and perpendicular views of an incomplete tetrahedral arrangement of an exterior edge (shown in dark blue).

- (1) Surface area of the corresponding Laguerre facet is computed and assigned to appropriate atoms
  - (2) Using the calculated area, corresponding Laguerre volumes are found and assigned to appropriate atoms
  - (3) Interresidual Laguerre surface areas are assigned
- Laguerre residue volumes are summed

### 3.1.1. Surface areas

When considering edge  $e_{ij}$ , we compute the surface area,  $LS_{ij}$ , of the facet,  $L_{ij}$ , between atoms  $i$  and  $j$ . The ordered vertices in the Laguerre facet are the characteristic points of the tetrahedra in the edge's tetrahedral arrangement. Note that the edge does not always intersect its Laguerre facet, but it is always perpendicular to it (Fig. 4). Next we define the characteristic point of an edge. The characteristic point,  $x_{ij} = (x'_{ij}, x''_{ij})$  of an edge,  $e_{ij}$  satisfies

$$\pi(p_i, x'_{ij}) = \pi(p_j, x'_{ij}) \quad (6)$$



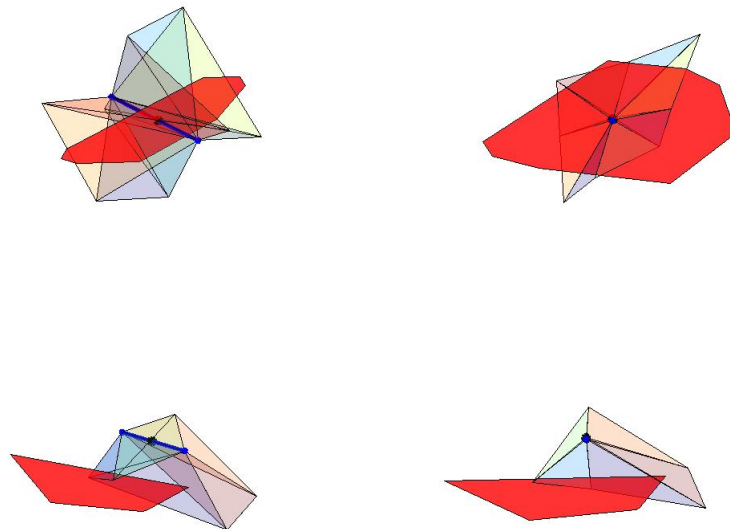


Fig. 4. Laguerre facet of edge represented in red. The vertices of the facet are the characteristic points of the tetrahedra in the edge's tetrahedron. Top row: Two views of a facet which is intersected by its corresponding edge. Bottom row: Two views of a facet which is not intersected by its corresponding edge.

with  $x''_{ij}$  minimal. The center of this point lies on the intersection of the line containing  $p'_i$  and  $p'_j$  and the plane of points which are equipowerdistant to  $p_i$  and  $p_j$ . The method computes the surface area as the sum of the areas of the triangles as shown in Fig. 5.

### 3.1.2. Volumes

The Laguerre volume contributions,  $L_{ij}^{(i)}$  and  $L_{ij}^{(j)}$ , are pyramids with base  $L_{ij}$ , and heights  $h_{ij}^i$  and  $h_{ij}^j$  respectively, with

$$h_{ij}^i = |x'_{ij} - p'_i| \quad (7)$$

$$h_{ij}^j = |x'_{ij} - p'_j|. \quad (8)$$

We can write

$$x'_{ij} = p'_j + t(p'_i - p'_j). \quad (9)$$

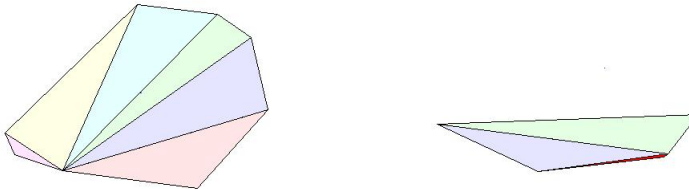


Fig. 5. Subdivision method for the two facets shown in Fig. 4.

Thus the position of the plane containing the edge's facet with respect to  $p'_i$  and  $p'_j$  is encoded in the value of  $t$  (Fig. 6). This gives

$$\begin{aligned}
 LV_{ij}^{(i)} &= \text{sign}(1-t) \frac{h_{ij}^i LS_{ij}}{3} \\
 LV_{ij}^{(j)} &= \text{sign}(t) \frac{h_{ij}^j LS_{ij}}{3}
 \end{aligned}
 \tag{10}$$

The term  $LV_{ij}^{(i)}$  is positive if the corresponding pyramid lies in  $L_i$  and negative otherwise (Figs. 7, 8). We call the latter case an “engulfing” contact; that is the generator of one cell lies in the Laguerre cell of another atom.

Formula 10 is in contrast to the volume calculation (Equation 7b) in McConkey's method<sup>26</sup> which does not consider signed volumes and thus returns incorrect quantities for engulfing contacts (For example, see Fig. 8).

### 3.2. Removing solvent

We would like to compute Laguerre-like volumes and surfaces of a molecule that is not surrounded by solvent, so that these quantities are as close as possible to the Laguerre volumes and surfaces that are found when the molecule is in its typical environment. In doing this, we come across two problems:

- The Laguerre cells are unbounded for certain atoms on the convex hull of  $\mathcal{A}'$ .
- The Laguerre cells of atoms that would otherwise be in contact with solvent are too large

This may be overcome by using Laguerre intersection cells,  $LI(w)$ . The Laguerre intersection cell of an atom is the intersection of the Laguerre cell of that atom with the expanded atom.

$$LI_i(w) = L_i \cap B_i(w).
 \tag{11}$$

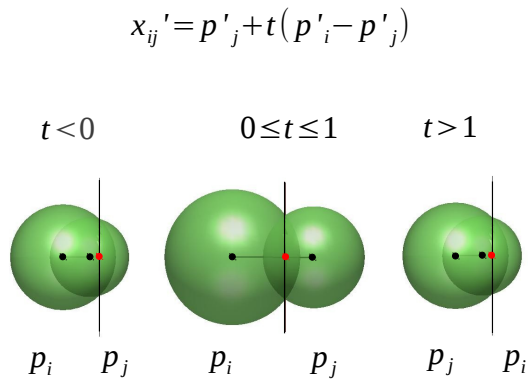


Fig. 6. The value of  $t$  is determined by the relationship of the equi-powerdistant plane with respect to the two edge points and indicates if the center of the generator of the cell is interior or exterior to its cell.

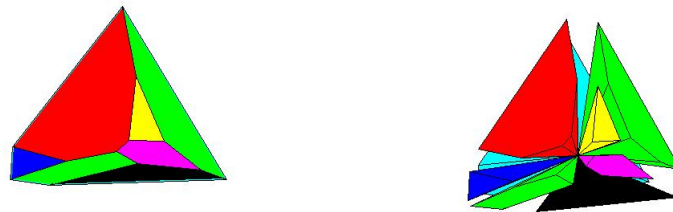


Fig. 7. Left: Laguerre cell that contains generator's center. Right: Pyramidal decomposition of Laguerre cell. All volumes are positive.

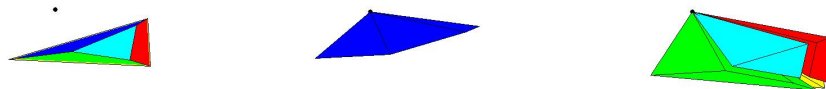


Fig. 8. Left: Laguerre cell that does not contain generator's center (black vertex). Center: Negative pyramidal volume. Right: Positive pyramidal volumes.

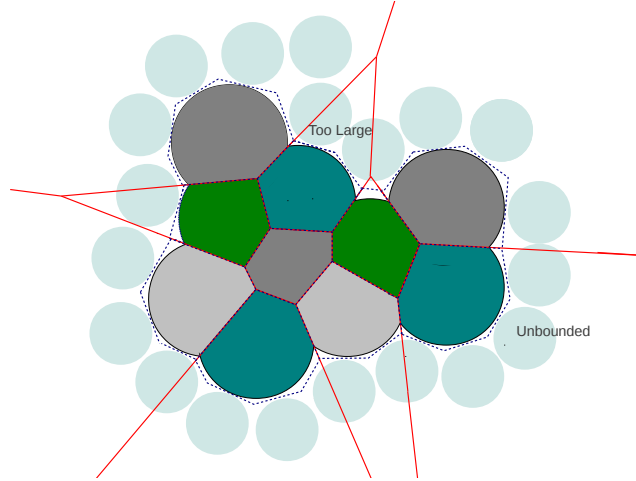


Fig. 9. Example of Laguerre-Intersection cells (shown in grays and greens). Solvent is represented by light blue spheres, boundary of Laguerre cells of molecule (minus solvent) are shown in red, and boundary of Laguerre cells of molecule in solvent is represented by the dotted line.

Exterior cells are bounded in a realistic way, and the cells of atoms which would otherwise be in contact with solvent are shrunk to a more appropriate size (Fig. 9). Laguerre cells are constant as a function of  $w$ , which means this parameter can be tuned to generate appropriate exterior Laguerre intersection cells while interior cells remain unchanged (Fig. 10).

Define  $LIS_i(w)$  and  $LIV_i(w)$ , be the surface area and volume of  $LI_i(w)$

$$\begin{aligned} LIS_i(w) &= \text{surf}(LI_i(w)) \\ LIV_i(w) &= \text{vol}(LI_i(w)). \end{aligned} \quad (12)$$

Just as the individual atomic surface area and molecular volume can be written as a sum of contributions corresponding to simplices in the alpha complex using the inclusion-exclusion formulas (See Appendix A),  $LIS$  and  $LIV$  can be split into contributions (Eqs. 13, 14) from simplices in the alpha complex which is a subset of the Delaunay tetrahedrization<sup>8</sup>. Define  $\mathcal{C}(w)$  to be the alpha complex of  $\mathcal{A}(w)$  with  $\partial\mathcal{C}(w)$  the set of simplices in the boundary of  $\mathcal{C}(w)$ . For  $T \subset \mathcal{A}(w)$ ,  $\sigma_T$  represents the simplex which is the convex hull of the centers of points in  $T$ . As we are working in three dimensions we only consider  $|T| \leq 4$ , where  $|T|$  is the number of elements in  $T$ . Recall that  $e_{ij} = \sigma_T$  for  $T = \{p_i, p_j\}$ . We also represent vertices and triangles as  $v_i$  and  $t_{ijk}$  in a similar manner.

The Laguerre-Intersection cell,  $LI$ , of an atom interior to  $\mathcal{C}$  is precisely the Laguerre cell  $L$ . The Laguerre-Intersection cell of an atom exterior  $\mathcal{C}(w)$  consists of polyhedral components (contributions from interior simplices) and spherical components (contributions from simplices on  $\partial\mathcal{C}(w)$ ).

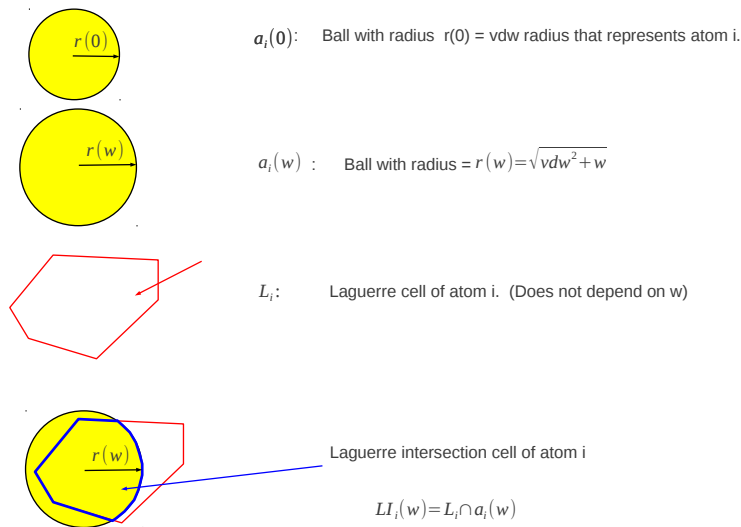


Fig. 10. Laguerre cell does not depend on  $w$  while the Laguerre-Intersection cell does depend on  $w$ .

The terms  $LIS_T^{(i)}(w)$  and  $LIV_T^{(i)}(w)$  are the surface area and volume contributions of the simplex  $\sigma_T$  to  $LIV_i(w)$  and  $LIS_i(w)$  (the analogs to  $S_T^{(i)}$  terms in A.3). We can also write  $LIV_T^{(i)}(w) = LIV_i^{(i)}(w)$  for  $T = \{p_i(w)\}$ ,  $LIV_T^{(i)}(w) = LIV_{ij}^{(i)}(w)$  for  $T = \{p_i(w), p_j(w)\}$ , etc., and likewise for surface areas (See Fig. 11). The terms  $P_T^{(i)}$  and  $F_T$  for  $|T| = 2$  are pyramidal volumes and facet surface areas that will be discussed shortly. We have

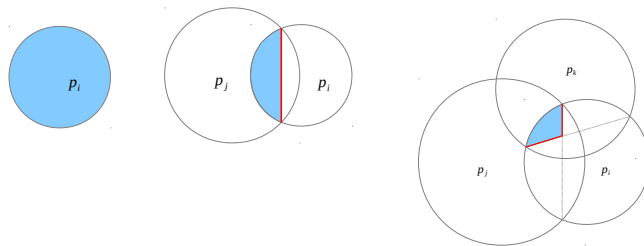


Fig. 11. Laguerre-Intersection cell volumes (blue),  $LIV_i^{(i)}$ ,  $LIV_{ij}^{(i)}$ ,  $LIV_{ijk}^{(i)}$ , and surfaces (red)  $LIS_{ij}$ , and  $LIS_{ijk}$ , to atom  $i$  from  $\partial C(w)$  simplices.

14 *Hummel, Yu, Simmerling, Coutsias*

$$\begin{aligned} LIV_i(w) &= \sum_{\sigma_T \in \partial\mathcal{C}(w)} (-1)^{k+1} c_T LIV_T^{(i)}(w) \quad |T| = k \\ &+ \sum_{\sigma_T \in \mathcal{C}(w)} P_T^{(i)}(w) \quad |T| = 2 \end{aligned} \quad (13)$$

and

$$\begin{aligned} LIS_i(w) &= S_i(w) + \sum_{\sigma_T \in \partial\mathcal{C}(w), k>1} (-1)^k c_T LIS_T^{(i)}(w) \quad |T| = k \\ &+ \sum_{\sigma_T \in \mathcal{C}(w)} F_T(w) \quad |T| = 2 \end{aligned} \quad (14)$$

where  $S_i(w)$  is the accessible surface area of atom  $i$  with solvent radius  $w$ , and  $c_T$ s are given for the following simplices

- $|T| = 1$ , i.e. a vertex  $v_i$ :  $c_T = \Omega_T$  is the fraction of the ball  $i$  outside the tetrahedra in the alpha complex. That is  $\Omega_T$  is the normalized outer solid angle subtended by the union of tetrahedra in  $\mathcal{C}$  which contain  $v_i$ .
- $|T| = 2$ , i.e. an edge  $e_{ij}$ :  $c_T = \Phi_T$  is normalized outer dihedral angle of the union of tetrahedra in  $\mathcal{C}$  which contain the edge  $e_{ij}$ .
- $|T| = 3$ , i.e. a triangle  $t_{ijk}$ :  $c_T$  is 1 if the triangle is singular and  $\frac{1}{2}$  if the triangle is regular. In other words,  $c_T$  is the fraction of  $V_T$  and  $S_T$  that is outside the union of tetrahedra in the alpha complex.

Here  $v_i = p'_i$ ,  $e_{ij} = \text{conv}(\{p'_i, p'_j\})$ , and  $t_{ijk} = \text{conv}(\{p'_i, p'_j, p'_k\})$ .

Since

$$S_i(w) = \sum_{\sigma_T \in \partial\mathcal{C}(w)} (-1)^{k+1} c_T S_T^{(i)}(w), \quad |T| = k \quad (15)$$

where  $S_T^{(i)}(w)$  is the contribution of  $\sigma_T$  to  $S_i(w)$ , we can write

$$\begin{aligned} LIS_i(w) &= c_i S_i^{(i)}(w) \\ &+ \sum_{\sigma_T \in \partial\mathcal{C}(w), k>1} (-1)^k c_T (LIS_T^{(i)}(w) - S_T^{(i)}(w)) \quad |T| = k \\ &+ \sum_{\sigma_T \in \mathcal{C}(w)} F_T \quad |T| = 2 \end{aligned} \quad (16)$$

### 3.3. Equations

#### 3.3.1. Vertices

The Laguerre volume and surface contributions from a vertex,  $v_i$ , are given by

$$c_i S_i^{(i)}(w) = \Omega_i r_i^2(w) \quad (17)$$

$$c_i LIV_i^{(i)}(w) = \frac{\Omega_i}{3} r_i^3(w). \quad (18)$$

### 3.3.2. Edges

The formulas for the volume and surface contributions from the intersection of two balls  $p_i$  and  $p_j$  are

$$c_T LIV_T^{(i)}(w) = \frac{\Phi_{ij}}{2} h_i^2 \left( r_i(w) - \frac{h_i}{3} \right) \quad (19)$$

$$c_T LIV_T^{(j)}(w) = \frac{\Phi_{ij}}{2} h_j^2 \left( r_j(w) - \frac{h_j}{3} \right) \quad (20)$$

$$c_T S_T^{(i)}(w) = \Phi_{ij} r_i(w) h_i \quad (21)$$

$$c_T S_T^{(j)}(w) = \Phi_{ij} r_j(w) h_j \quad (22)$$

$$c_T LIS_T^{(i)}(w) = \frac{\Phi_{ij}}{2} h_i (2r_i(w) - h_i) \quad (23)$$

$$c_T LIS_T^{(j)}(w) = c_T LIS_T^{(i)}(w). \quad (24)$$

### 3.3.3. Triangles

Let  $x_T = x_{ijk} = (x'_{ijk}, x''_{ijk})$  be the characteristic point of the triangle  $t_{ijk}$ . Let  $p$  be one of the two points of intersection on the spheres  $p_i(w)$ ,  $p_j(w)$ , and  $p_k(w)$ . We can write  $p = x'_{ijk} + \mathbf{n}x''_{ijk}$  where  $\mathbf{n}$  is the normal to the plane containing the triangle,

$$\mathbf{n} = \frac{(p'_j - p'_i) \times (p'_k - p'_i)}{|(p'_j - p'_i) \times (p'_k - p'_i)|}. \quad (25)$$

Let  $\sigma_{Tc}$  the tetrahedron defined by the centers of  $T$  and  $x_T$ . Then

$$\frac{1}{2} S_T^{(i)} = \phi_{ij} S_{ij}^{(i)} + \phi_{ik} S_{ik}^{(i)} - \omega_i S_i^{(i)} \quad (26)$$

$$\frac{1}{2} S_T^{(j)} = \phi_{jk} S_{jk}^{(j)} + \phi_{ji} S_{ji}^{(j)} - \omega_j S_j^{(j)} \quad (27)$$

$$\frac{1}{2} S_T^{(k)} = \phi_{ki} S_{ki}^{(k)} + \phi_{kj} S_{kj}^{(k)} - \omega_k S_k^{(k)} \quad (28)$$

where  $\phi_{ij}$  is the fractional inner dihedral angle of  $\sigma_{Tc}$  along edge  $\sigma_{ij}$ ,  $\omega_i$  is the fractional (inner) solid angle of  $\sigma_{Tc}$  subtended from  $p'_i$ , etc. Let  $x'_{ij}$ ,  $x'_{ik}$ ,  $x'_{jk}$  be the centers of the characteristic points of the triangle's edges. Then  $\frac{1}{2} LS_{ijk}^{(i)}$  is the sum of the areas of the triangles given by  $x'_{ijk}$ ,  $x'_{ij}$ ,  $p$ , and  $x'_{ijk}$ ,  $x'_{ik}$ ,  $p$ . We have

$$\frac{1}{2} LIV_T^{(i)} = \tilde{V}_i - \omega_i V_i^{(i)} + \phi_{ij} LIV_{ij}^{(i)} + \phi_{ik} LIV_{ik}^{(i)} \quad (29)$$

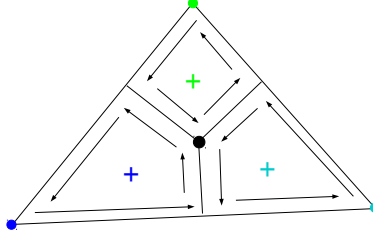


Fig. 12.  $p'_i$ ,  $p'_j$ ,  $p'_k$ , and  $x'_{ijk}$  represented by dark blue, light blue, green, and black dots, respectively. The point  $p$  is out of the page. The black arrows represent positive contributions to the volumes  $\tilde{V}_i$ ,  $\tilde{V}_j$ , and  $\tilde{V}_k$ .

where  $\tilde{V}_i$  is the signed volume of the solid with vertices at  $p'_i$ ,  $x_{ij}$ ,  $x_{ijk}$ ,  $x_{ik}$ , and  $p$ . More specifically, let

$$\mathbf{a} = x'_{ij} - p'_i \quad (30)$$

$$\mathbf{b} = x'_{ijk} - x'_{ij} \quad (31)$$

$$\mathbf{c} = x'_{ik} - x'_{ijk} \quad (32)$$

$$\mathbf{d} = p'_i - x'_{ik} \quad (33)$$

$$(34)$$

then

$$2\tilde{V}_i = \frac{h}{3} \left( (\mathbf{a} \times \mathbf{b}) + (\mathbf{c} \times \mathbf{d}) \right) \cdot \mathbf{n} \quad (35)$$

Equivalent equations hold for  $LIV_T^{(j)}$ ,  $LIV_T^{(k)}$ ,  $\tilde{V}_j$ , and  $\tilde{V}_k$ . See Figs. 12, 13, 14 for possible configurations.

### 3.3.4. Other contributions

We get additional contributions,  $F_T$  and  $P_T^{(i)}$ , from all edges in  $\mathcal{C}(w)$ . There are two types of these edges:

- edges interior to  $\mathcal{C}(w)$
- edges exterior to  $\mathcal{C}(w)$ .

Contributions from interior edges of  $\mathcal{C}(w)$  are computed as in section 3.1. That is

$$P_{ij}^{(i)} = LV_{ij}^{(i)} \quad (36)$$

and

$$F_{ij} = LS_{ij}. \quad (37)$$



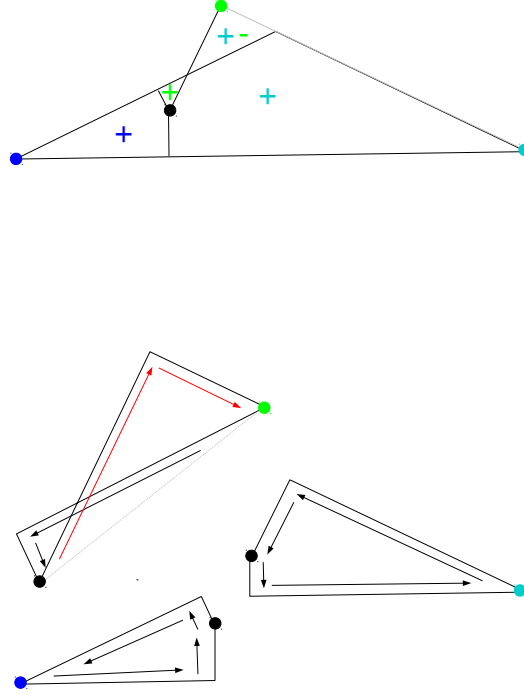


Fig. 13. Above: Sample configuration. Below: Black arrows represent positive contributions to  $\tilde{V}_i$ ,  $\tilde{V}_j$ , and  $\tilde{V}_k$ , while red arrows represent negative contributions.

The terms  $P_T^{(i)}$  and  $F_T$  are zero for all edges that are not part of at least one tetrahedron in  $\mathcal{C}(w)$ . To calculate the contribution from all other exterior  $\partial\mathcal{C}(w)$  edges, we must know the characteristic points of the triangles on the boundary of the  $\mathcal{C}(w)$  tetraring. The characteristic point of the boundary triangle may or may not be on the Laguerre facet, but it always indicates the direction of one of the facet's boundary segments (Fig. 15). For each exterior edge, the counterclockwise list of Laguerre nodes corresponding to tetrahedra in the alpha complex is already known. The surface area is computed by using signed surface areas as shown in pseudocode in the appendix. The volume contributions to  $LI_i$  and  $LI_j$  are found by multiplying the resulting surface by  $\frac{h_{ij}^i}{3}$  and  $\frac{h_{ij}^j}{3}$ , respectively. In this case, a sign is incorporated into the surface area which means the  $sign(1-t)$  and  $sign(t)$  terms in equation 10 are not needed.

Note that the  $\mathcal{C}(w)$  tetraring of an edge may not be connected. The pseudocode does not take this into account for simplicity reasons, while the actual implemen-

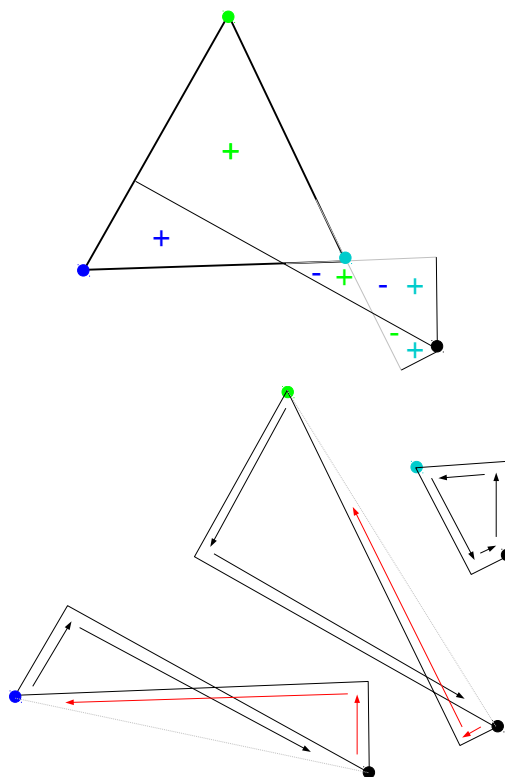
18 *Hummel, Yu, Simmerling, Coutsias*

Fig. 14. Above: Sample configuration. Below: Black arrows represent positive contributions to  $\tilde{V}_i$ ,  $\tilde{V}_j$ , and  $\tilde{V}_k$ , while red arrows represent negative contributions.

tation does.

## 4. Experiments: Accuracy of Algorithm

### 4.1. Monte-Carlo based estimates

We tested the accuracy of our Laguerre-intersection algorithm on a tryptophan dipeptide trajectory (1000 structures) with Monte-Carlo based estimates of Laguerre-Intersection volumes and surfaces. Tryptophan has a near-planar ring which was useful for testing the robustness of our algorithm.

Monte-Carlo Laguerre-Intersection volume estimates were obtained by generating  $N = 10^n$  with  $n = 1 : 7$  uniformly random points in the volume of each atom. The fraction  $f_{ij}$  of points in the volume of atom  $j$  in structure  $i$  which are closest to atom  $j$  by the power distance were tabulated. The Laguerre-Intersection volume

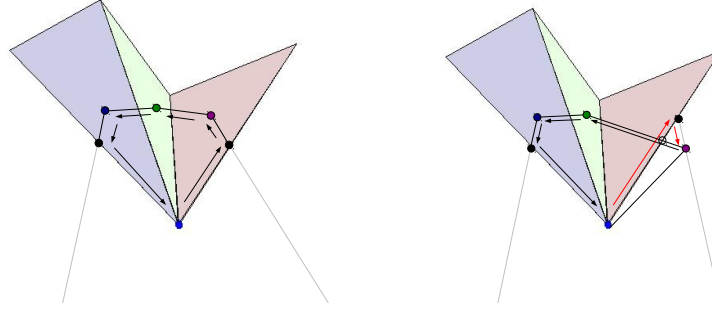


Fig. 15. Two examples of additional contributions from edges which are regular and exterior with respect to  $\mathcal{C}(w)$ . The royal blue points are the characteristic points of edges which are exterior to  $\mathcal{C}(w)$  and which points directly out of the paper. The triangles represent tetrahedra in the tetrahedra ring which are also in  $\mathcal{C}(w)$ . Dark blue, green, and purple dots represent the Laguerre nodes corresponding to the tetrahedra of the same color in the  $\mathcal{C}(w)$  tetrahedra ring. The black points are characteristic points of the boundary triangles. In the first case, both triangle characteristic points lie in the Laguerre facet. In the second case, one of the triangle characteristic points is not in the facet but is still needed to determine a line segment (shown in gray). In the first case all Laguerre nodes lie in the union of tetrahedra and all corresponding surface contributions are positive. In the second case, a Laguerre node lies outside the union of tetrahedra in  $\mathcal{C}(w)$ . The triangular surface that is bounded by the right black, purple, and clear dots was originally assigned during contributions from the boundary triangle, and must be subtracted to obtain the correct area.

estimate is

$$\widehat{LIV}_{ij} = f_{ij} \frac{4\pi r_j^3}{3} \quad (38)$$

where  $r_j$  is the radius of atom  $j$ .

Laguerre-Intersection atomic surface areas consist of spherical (solvent-accessible) portions and planar portions that lie on discs of intersection between two spheres. The solvent-accessible portions were estimated in a similar manner as Laguerre-Intersection volumes.  $N = 10^n$  with  $n = 1 : 7$  uniformly random points were generated on each sphere and the estimated surface areas are

$$\widehat{SAS}_{ij} = f_{ij} 4\pi r_j^2 \quad (39)$$

where  $f_{ij}$  is the fraction of points on the surface of atom  $j$  in structure  $i$  that lie closest to atom  $j$  by the power distance.

The planar Laguerre-Intersection surface area for atom  $j$  in structure  $i$  is

$$pLIS_{ij} = \sum_{k=1}^q LIS_{ijk} \quad (40)$$

where  $q$  is the number of alpha-complex neighbors and  $LIS_{ijk}$  is the portion of  $LIS_{ij}$  that lies on the disc of intersection between atom  $j$  and its neighbor  $k$  in

20 Hummel, Yu, Simmerling, Coutsias

structure  $i$ . Let  $A_{ijk}$  be the surface area of this disc and define

$$A_{ij} = \sum_{k=1}^q A_{ijk}. \quad (41)$$

Then  $M_{ij} = N \frac{A_{ijk}}{A_{ij}}$  random points are generated on the disc of intersection between atom  $j$  and its neighbor  $k$  in structure  $i$  and the estimated planar Laguerre-Intersection surface area is

$$\widehat{pLIS}_{ij} = \sum_{k=1}^q f_{ijk} A_{ijk} \quad (42)$$

where  $f_{ijk}$  is the fraction of points on the disc of intersection that are closest to atom  $j$  by the power distance.

Maximum and mean errors decrease as the number of test points increase (Fig. 16). Final mean errors for  $LIV$ ,  $SAS$ , and  $pLIS$  estimates are  $O(10^{-4})$ . Maximum relative errors for  $LIV$ ,  $SAS$ , and  $pLIS$  estimates are  $O(10^{-3})$ .

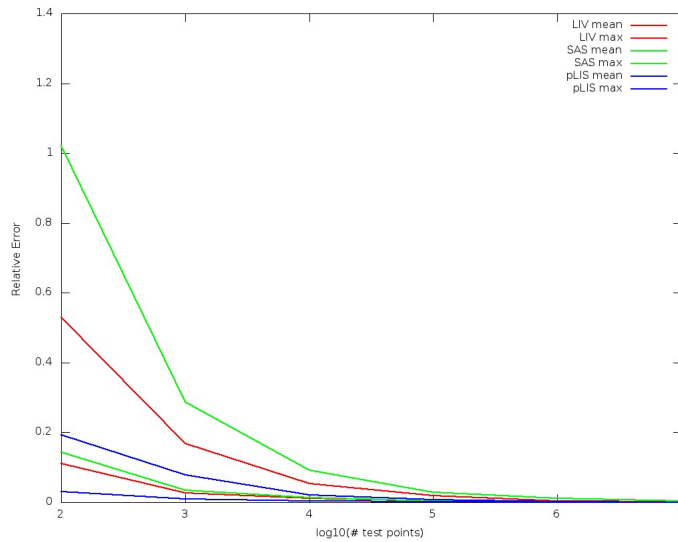


Fig. 16. Mean and maximum relative difference over all atoms between estimated and analytic Laguerre-Intersection atomic volumes and surfaces areas for 1000 structure tryptophan dipeptide trajectory.

An analysis of the accuracy of the estimates follows. Let  $LIV$  be the Laguerre-Intersection volume for an atom with radius  $r$ . Then the probability that a test

point  $x$  falls in  $LIV$  is

$$p = \frac{LIV}{4/3\pi r^3}. \quad (43)$$

Given  $n$  trials, the number of points,  $k$ , that fall within  $LIV$  follows a Binomial distribution with parameters  $p$  and  $n$ ,

$$k \sim B(p, n). \quad (44)$$

For large  $np$  and  $n(1-p)$ ,  $B(p, n)$  is approximated well by the normal distribution  $N(np, \sqrt{np(1-p)})$ .

Given a significance level,  $\alpha = .01$ , upper and lower limits,  $\bar{k}$  and  $\underline{k}$ , on  $k$  may be calculated with 99% confidence. Since the normal distribution is symmetric about the mean, we are 99% confident that the relative error of the estimate is bounded by

$$\left| \frac{\widehat{LIV} - LIV}{LIV} \right| \leq \left| \frac{\frac{\bar{k}}{n} - p}{p} \right|. \quad (45)$$

Note that the right hand side of (45) is a decreasing function of  $p$ .

We follow the same analysis for  $SAS$  and  $pLIS$ . For our trajectory,  $p \geq .174$ ,  $p \geq .056$ , and  $p \geq .180$  for  $LIV$ ,  $SAS$ , and  $pLIS$  respectively. This gives 99% probability that

$$\left| \frac{\widehat{LIV} - LIV}{LIV} \right| \leq 1.77 \times 10^{-3} \quad (46)$$

$$\left| \frac{\widehat{SAS} - SAS}{SAS} \right| \leq 3.34 \times 10^{-3} \quad (47)$$

$$\left| \frac{\widehat{pLIS} - pLIS}{pLIS} \right| \leq 1.73 \times 10^{-3} \quad (48)$$

which is consistent with the values in Fig. 16.

#### 4.2. Comparison with Gauss-Bonnet-type algorithm

We compared atomic Laguerre-Intersection volumes and surface areas calculated using our method with those found using *vorlume*<sup>3</sup>, the Gauss-Bonnet-type algorithm adopted in the Structural Bioinformatics Library (SBL) on a diverse set of 1013 proteins in the Protein Data Bank. With the latter algorithm, we used the ‘exact’ option which employs interval arithmetic with extended precision and sufficiently small intervals that the solution is considered ‘exact’. We compared the results from our algorithm and the single number reported by *vorlume* without the ‘exact’ option in the SBL, to test how our method compares to Gauss-Bonnet-type method when both employ double precision computations. We checked both algorithms’ results of surface and volume to see whether they fall in intervals provided

by the algorithm with ‘exact’ option and calculate the correct ratio, which is defined as the ratio of atoms for each protein whose value for surface or volume falls in the interval provided by *vorlume* with the ‘exact’ option.

Both algorithms do better on computing surfaces than on computing volumes. However, no matter whether surface or volume, our algorithm appears more robust and accurate. We identified the 20 proteins where each one of these two algorithms obtained its lowest correct ratios, respectively. The average correct ratio for *vorlume* to calculate the surface of its 20 most unfavorable proteins is 3.695%, while the correct ratio of our algorithm for the same 20 proteins is 96.44%. The average correct ratio for *vorlume* to calculate the volume of its 20 most unfavorable proteins is 2.316%, while the correct ratio of our algorithm for the same 20 proteins is 68.45%. The average correct ratio for our algorithm to calculate the surface of its 20 most unfavorable proteins is 94.99%, while the correct ratio of *vorlume* for the same 20 proteins is 37.52%. The average correct ratio for our algorithm to calculate the volume of its 20 most unfavorable proteins is 67.377%, while the correct ratio of *vorlume* for the same 20 proteins is 3.043%. We see that our algorithm’s results do not vary as much as the Gauss-Bonnet-type *vorlume* when it is calculating the surface/volume for the 2 different sets of 20 proteins.

If we consider the whole 1013 protein database which contain 3710687 atoms, the correct ratio for surface of our algorithm is  $3580002/3710687 = 96.48\%$  and the correct ratio for volume of our algorithm is  $3295669/3710687 = 88.81\%$ . Correspondingly, the correct ratio for surface of *vorlume* is  $1079062/3710687 = 29.08\%$  and the correct ratio for volume of *vorlume* is  $615823/3710687 = 16.60\%$ . In conclusion, our algorithm gives results more accurate and robust.

From the perspective of computing time, our algorithm has obvious advantages. We take the two modes of the Gauss-Bonnet-type algorithm *vorlume* in the SBL as a reference. The mode with option ‘exact’ will give results with extended precision at a considerable time cost while the mode without ‘exact’ uses double precision floats, which is less accurate than ‘exact’ model but much faster. Excluding the time for tetrahedralization from the total time budget, we find that over the 1013 protein database, which involved 3710687 atoms in total, the total time spent on computing surface and volume using *vorlume* with ‘exact’ option is 36982.902331s. The total time spent on computing surface and volume using *vorlume* without ‘exact’ option is 2948.505325s. The total time spent on computing surface and volume using our algorithm is 73.316s. Our algorithm is about 40 times faster than the much less accurate mode, while it is about as 504 times fast as the exact mode, with which we agree on the great majority of cases.

## 5. Experiments: Optimizing Solvent Parameter

We test the capability of the Laguerre-Intersection method for predicting explicit water Laguerre quantities. We train on molecular dynamics trajectory data to determine “optimal solvent parameters”; those parameters for which the Laguerre-

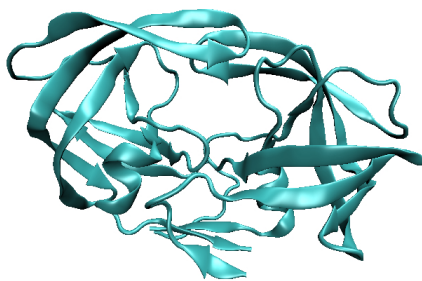


Fig. 17. Sample structure from the HIV protease molecular dynamics trajectory.

Intersection volumes and surface areas are, on average, as close as possible to the Laguerre volumes and surfaces of a molecule in its typical solvent environment.

### 5.1. Data

We work with solvated molecular dynamics trajectory data (2500 structures) of the HIV protease dimer (See Fig. 17) to determine optimal solvent parameters. The trajectory was initiated from the 1HVR crystal structure<sup>22</sup>, and generated using the Amber software<sup>2</sup> with the ff99SB protein force field<sup>18</sup> and the TIP3p water model<sup>20</sup>. Equilibration followed published protocol<sup>32</sup> with a 2fs time step, and the simulation coupled to a thermostat at 300K. A total of 100 ns of data were generated.

Optimal solvent parameters, which converge quickly, determine the capability of the implicit Laguerre-Intersection model to predict explicit water Laguerre quantities. We also compare our method to other implicit models.

First residual and atomic Laguerre volumes, atomic, interresidual, and residue-solvent surface areas (LV\_res, LV\_atom, LS\_atom, LS\_interres, LSAS\_res, respectively) were calculated for the solvated trajectory.

For various solvent parameters,  $w_k$  and  $r_k$ ,

$$\begin{aligned}w_k &= k \cdot dw \\ r_k &= k \cdot dr\end{aligned}$$

equivalent Laguerre-Intersection quantities were computed for each structure

24 Hummel, Yu, Simmerling, Coutsias

(LIV\_res, LIV\_atom, LIS\_atom, LIS\_interres, LISAS\_res, respectively). We set  $dw = .1 \text{ \AA}$  and chose  $dr$  such that

$$1.7^2 + k_{max} \cdot dw = (1.7 + k_{max} \cdot dr)^2. \quad (49)$$

This means that for an atom with atomic radius  $1.7 \text{ \AA}$  (the typical radius of carbon), the initial ( $k = 0$ ) and final ( $k = k_{max}$ ) total weights are the same using both methods.

Laguerre and Laguerre-Intersection quantities were compared and “optimal” solvent parameters were found. The optimal solvent weight (radius) is the value  $w_k$  ( $r_k$ ) that minimizes the error (Equations 50, 51).

We write the Laguerre volume of residue  $j$  in structure  $i$  as  $LIV\_res(i, j)$ , and the Laguerre-Intersection volume of residue  $j$  in structure  $i$  with solvent parameter  $w_k$  ( $r_k$ ) as  $LIV\_res(i, j, k)$ . We write the interresidual Laguerre surface area between residues  $j$  and  $jj$  in structure  $i$  as  $LS\_interres(i, j, jj)$ , and the interresidual Laguerre-Intersection surface area between residues  $j$  and  $jj$  in structure  $i$  with solvent value  $w_k$  ( $r_k$ ) as  $LIS\_interres(i, j, jj, k)$ . Equivalent formulas hold for the other Laguerre and Laguerre-Intersection quantities.

Using wildcard notation, an arbitrary Laguerre quantity and Laguerre-Intersection quantity are written as  $L^*$  and  $LI^*$ . Then with solvent parameter  $w_k$  ( $r_k$ ) 1-Norm and 2-Norm errors are

$$E_1(LI^*(k)) = \frac{1}{N} \sum_{i=1}^N \epsilon_{1,i}(LI^*(k)) \quad (50)$$

and

$$E_2(LI^*(k)) = \sqrt{\frac{1}{N} \sum_{i=1}^N \epsilon_{2,i}(LI^*(k))} \quad (51)$$

where  $N$  is the number of structures in the MD trajectory.

For  $LIV\_res$ ,  $LIV\_atom$ ,  $LIS\_atom$ , and  $LISAS\_res$

$$\epsilon_{1,i}(LI^*(k)) = \frac{1}{\bar{n}} \sum_{j=1}^n |LI^*(i, j, k) - L^*(i, j)| \quad (52)$$

and

$$\epsilon_{2,i}(LI^*(k)) = \frac{1}{\bar{n}} \sum_{j=1}^n (LI^*(i, j, k) - L^*(i, j))^2 \quad (53)$$

where  $n$  is the number of residues or atoms and

$$\bar{n} = \sum_{j=1}^n \mathbb{I}(j) \quad (54)$$

with



$$\mathbb{I}(j) = \begin{cases} 1 & \text{if } LI^*(i, j) \neq 0 \text{ or } L^*(i, j) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

For the first three quantities  $\bar{n} = n$ . The average Laguerre quantity is defined as

$$av(L^*) = \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{\bar{n}} \sum_{j=1}^n L^*(i, j) \right). \quad (55)$$

With *LIS\_interres* we have

$$\epsilon_{1,i}(LI^*(k)) = \frac{1}{\bar{n}} \sum_{j=1}^n \sum_{jj=j+1}^n |LI^*(i, j, jj, k) - L^*(i, j, jj)| \quad (56)$$

and

$$\epsilon_{2,i}(LI^*(k)) = \frac{1}{\bar{n}} \sum_{j=1}^n \sum_{jj=j+1}^n (LI^*(i, j, jj, k) - L^*(i, j, jj))^2 \quad (57)$$

where

$$\bar{n} = \sum_{j=1}^n \sum_{jj=j+1}^n \mathbb{I}(j, jj) \quad (58)$$

with

$$\mathbb{I}(j, jj) = \begin{cases} 1 & \text{if } LI^*(i, j, jj) \neq 0 \text{ or } L^*(i, j, jj) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The average interresidual surface area is defined as

$$av(LS\_interres) = \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{\bar{n}} \sum_{j=1}^n \sum_{jj=j+1}^n L^*(i, j, jj) \right). \quad (59)$$

In this paper we present the absolute errors divided by the average quantities. We do not compute the relative errors at each step because the exact quantities may be zero.

First we present optimal solvent parameters obtained from HIV protease trajectory data. We show these parameters converge quickly despite the configurational variability of the molecule. We compare the capability of the Laguerre-Intersection, McConkey<sup>26</sup>, and Cazals'<sup>4</sup> method to predict explicit solvent Laguerre quantities. After a brief overview of these methods, We will compare all three methods to explicit water models, followed by a comparison of the radius vs. weight method of expanding atoms.

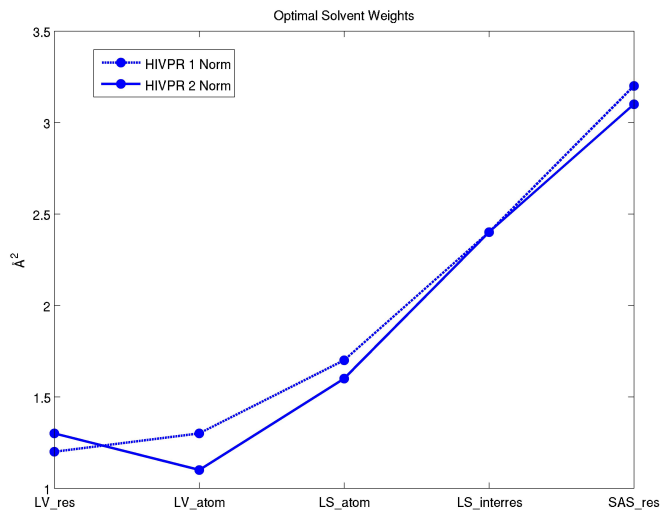


Fig. 18. Optimal solvent weights from the HIV protease trajectory.

## 5.2. Optimal solvent parameters

We plot the optimal solvent weights found from the HIV protease trajectory which has 2500 structures (See Fig. 18). The optimal solvent parameters for Laguerre volumes and Laguerre surfaces differ which is expected. One additional benefit of the weight method, is that the weighted Delaunay tetrahedrization is constant for all weights. This means the tetrahedrization (the time limiting step) only needs to be computed once per structure. When using the radius method, the tetrahedrization must be recalculated for each optimal solvent parameter.

Note that these are not “universal” parameters and are not necessarily optimal for an arbitrary molecule. Further work is needed to determine which values should be used in a given molecular dynamics simulation.

Optimal solvent parameters, as given by Equations 50, 51, 55, and 59, are recorded at each step in the trajectory. The first 100 out of 2500 iterations are shown in Fig. 19. We see that the optimal solvent parameters are immediately within one  $dw$  of the global optimal solvent parameters and soon become stationary with each step. Since the HIV protease dimer has flexible loops and open-close transitions we conjecture that most other proteins will exhibit similar or quicker convergence.

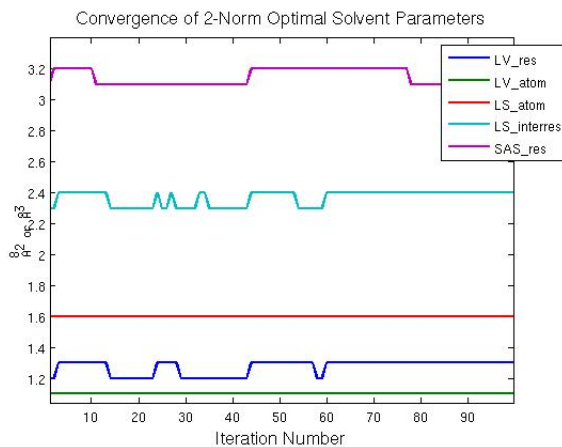


Fig. 19. Convergence of optimal solvent parameters.

### 5.3. Overview of current methods

#### 5.3.1. McConkey-Cazal method

McConkey-Cazals (MC) algorithm<sup>5 26</sup> constructs Laguerre-like cells by considering the surfaces of extended radical contact planes between neighboring atoms within a cutoff distance and the surface of an expanded sphere. Similar to our Laguerre-Intersection algorithm, this method considers the intersection of cells with the space filling diagram with atom radii expanded by a solvent radius of 1.4 Å. While the McConkey-Cazals method and the Laguerre-Intersection method are algorithmically different, (McConkey-Cazals method is based on the Gauss-Bonnet theorem while ours is an inclusion-exclusion method), another key difference is that the cells in McConkey-Cazals method are bounded by extended radical planes rather than radical planes. This poses problems since the cells of atoms in the bulk, which should not be affected by the cutoff distance, are affected. Furthermore, cells of small atoms completely disappear for cutoff distances as small as 1.4 Å.

#### 5.3.2. 'Intervor' for interface surfaces

Another Cazals' method, 'Intervor', is used to determine interface surfaces between molecules or a molecule and solvent<sup>6 4</sup> and locates atoms that are in direct contact with each other and those whose contact is mediated by water. This method is more qualitative in nature than the Laguerre-Intersection method, i.e. it does not claim to give contact areas that are similar to those of solvated systems but rather locate and describe topology of interactions.

A solvent radius of  $r = 1.4$  Å to all atoms and the alpha complex,  $\mathcal{C}(0)$ , is computed. Laguerre facets dual to edges not in  $\mathcal{C}(0)$  are thrown out. This is called

Quantity	Method	1-Norm / Average	2-Norm / Average
LS_interres	Laguerre-Intersection	.1416	.2027
	MC r=1.4	.3224	.5101
	Intervor	.6988	1.3150
LV_res	LI	.05695	.07097
	MC r=1.4	.4184	.5487
LS_atom	LI	.06794	.09994
	MC r=1.4	.4841	.5714
LV_atom	LI	.1039	.1636
	MC r=1.4	.7573	1.0937
SAS_res	LI	.1176	.1548
	MC r=14.	.2039	.2536

Table 1. Ratio of errors to average Laguerre quantities. See equations 50, 51, 55, 59.

the  $\alpha$  criterion. However, overly large facets still remain. Such facets are discarded according to 'condition  $\beta$ ':

$$\frac{\bar{\mu}}{w_e} > M^2 \quad (60)$$

where  $w_e$  is the weight of the smaller of the two balls in the edge and  $\bar{\mu}$  is the size of the largest weighted Delaunay tetrahedron that contains the edge. The value  $M$  is set to 5.

#### 5.4. Comparison with explicit model

We compare the Laguerre-Intersection, McConkey-Cazals<sup>26 5</sup> (MC), and Intervor<sup>4</sup> to explicit models for the HIV protease trajectory data. We see a clear improvement in accuracy using the Laguerre-Intersection method (See Table 1). SAS\_res is approximately thirty percent more accurate while LV\_res is about eight times more accurate when compared to the MC method. Intervor<sup>4</sup> gives norms about four times larger than the Laguerre-Intersection method and about two times larger than the MC method.

At this point we do not know what effect this improvement might have on the accuracy of statistical potentials developed with these models.

#### 5.5. Radius vs. weight method

In our simulations, we found that the weight rather than radius method gave smaller errors for atomic quantities and similar errors for residual quantities. This is expected as the Laguerre cell of the molecule does not vary as atom weights are uniformly increased (See Fig. 2). Residual errors decrease relative to atomic errors

for the radius method due to cancellation in atomic errors. Fig. 20 compares relative errors for Laguerre-Intersection quantities over the HIV protease trajectory.

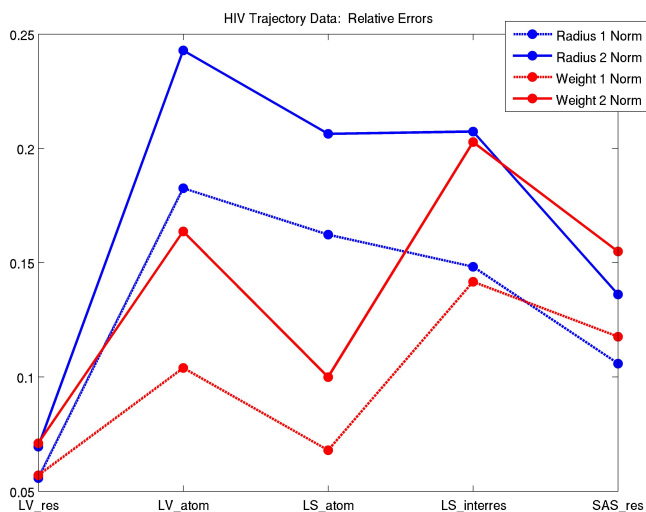


Fig. 20. Relative errors corresponding to optimal solvent parameters.

## 6. Conclusion

We developed an algorithm to compute the atomic and residual volumes and surface areas of the intersection of the Laguerre diagram with the space filling model of a molecule with applications to implicit solvation. While methods exist which calculate volumes and surface areas of the restriction of the Laguerre diagram to the space filling model, our algorithm is unique in that it is based on inclusion-exclusion rather than Gauss-Bonnet which decreases the complexity of the calculation. We also optimize an adjustable parameter, the weight, so the Laguerre-Intersection atomic and residual volumes and surface areas are as close as possible to Laguerre volumes and surface areas found using explicit solvent. We test our algorithm on an explicit water HIV protease molecular dynamics trajectory. We show that volumes and surface areas computed using our implicit method are 30% to 8 times closer to explicit quantities than those found using current models. We also show that varying the weight rather than the radius in the space-filling diagram gives much better agreement with explicit quantities.

## Acknowledgments

Michelle Hummel and Evangelos Coutsias were supported in part by NIH grant R01GM090205 and Stony Brook University research funds. Carlos Simmerling was supported in part by NIH grant R01GM107104. Evangelos Coutsias and Carlos Simmerling acknowledge support from the Laufer Center for Physical and Quantitative Biology at Stony Brook University. A portion of this work was carried out at the Department of Mathematics and Statistics, University of New Mexico.

## References

1. Yih-En Andrew Ban, Herbert Edelsbrunner, and Johannes Rudolph, *Interface surfaces for protein-protein complexes*, Journal of the ACM **53** (2006), 361–378.
2. David A. Case et al., *The Amber biomolecular simulation programs*, Journal of Computational Chemistry **26** (2005), 1668–1688.
3. F. Cazals, T. Dreyfus, and S. Loroit, *User manual: Space filling model surface volume*, Structural Bioinformatics Library (Accessed 2017), [http://sbl.inria.fr/doc/Space\\_filling\\_model\\_surface\\_volume-user-manual.html](http://sbl.inria.fr/doc/Space_filling_model_surface_volume-user-manual.html).
4. Frederic Cazals, *Revisiting the Voronoi description of protein-protein interfaces: Algorithms*, Lecture Notes in Computer Science **6282** (2010), 419–430.
5. Frederic Cazals, Harshad Kanhere, and Sebastien Loroit, *Computing the volume of a union of balls: a certified algorithm*, ACM Transactions on Mathematical Software **38** (2011).
6. Frederic Cazals, Flavien Prout, Ranjit P. Bahadur, and Joel Janin, *Revisiting the Voronoi description of protein-protein interfaces*, Protein Science **15** (2006), 2082–2092.
7. M. L. Connolly, *Analytical molecular surface calculation*, Journal of applied crystallography **16** (1983), 548–558.
8. Herbert Edelsbrunner, *Weighted alpha shapes*, Technical Report (1992).
9. ———, *The union of balls and its dual shape*, Proceeding of the ninth annual symposium on Computational Geometry, San Diego, CA, 1993.
10. Herbert Edelsbrunner and Ping Fu, *Measuring space filling diagrams and voids*, Proc. 28th Ann. HICSS (1995).
11. Herbert Edelsbrunner and Patrice Koehl, *The geometry of biomolecular solvation*, Discrete and Computational Geometry: MSRI publications **52** (2005), 241–273.
12. Jeremy Esque, Christophe Oguey, and Alexandre G. de Brevern, *A novel evaluation of residue and protein volumes by means of Laguerre tessellation*, J. Chem. Inf. Model **50** (2010), 947–960.
13. ———, *Comparative analysis of threshold and tessellation methods for determining protein contacts*, Journal of Chemical Information and Modeling **51** (2011), 493–507.
14. Mark Gerstein, Jerry Tsai, and Michael Levitt, *The volume of atoms on the protein surface: Calculated from simulation, using Voronoi polyhedra*, J. Mol. Biol. **249** (1995), 955–966.
15. K.D. Gibson and Harold A. Scheraga, *Volume of the intersection of three spheres of unequal size: a simplified formula*, Journal of Physical Chemistry **91**, 4121–4122.
16. Jeffrey J. Headd, Y.E. Andrew Ban, Paul Brown, Herbert Edelsbrunner, Madhuwanti Vaidya, and Johannes Rudolph, *Protein-protein interfaces: Properties, preferences, and projections*, Journal of Proteome Research **6** (2007), 2576–2586.
17. Viktor Hornak et al., *Hiv-1 protease flaps spontaneously open and reclose in molecular dynamics simulations*, PNAS **103** (2005), 915–920.

18. ———, *Comparison of multiple Amber force fields and development of improved protein backbone parameters*, Proteins: Structure, Function, and Bioinformatics **65** (2006), 712–725.
19. Michelle H. Hummel, *Delaunay-Laguerre Geometry for Macromolecular Modeling and Implicit Solvation*, Ph.D. thesis, University of New Mexico, New Mexico, 2014, Lobo-vault Repository, <http://hdl.handle.net/1928/25782>.
20. William L. Jorgensen et al., *Comparison of simple potential functions for simulating liquid water*, The journal of chemical physics **79** (1983), 926–935.
21. Peter A. Kollman et al., *Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models*, Acc. Chem. Res. **33** (2000), 889–897.
22. PY Lam et al., *Rational design of potent, bioavailable, nonpeptide cyclic ureas as hiv protease inhibitors*, Science **263** (1994), 380–384.
23. Jie Liang, Herbert Edelsbrunner, Ping Fu, Pamidighantam V. Sudhaker, and Shankar Subramaniam, *Analytical shape computation of macromolecules: I. molecular area and volume through alpha shapes*, PROTEINS: Structure, Function, and Genetics **33** (1998), 1–17.
24. Sedigheh Mahdavi, Ali Mohades, Ali Salehzadeh-Yazdi, Samad Jahandideh, and Ali Masoudi-Nejad, *Computational analysis of RNA-protein interaction interfaces via the Voronoi diagram*, Journal of Theoretical Biology **293** (2012), 55–64.
25. Sedigheh Mahdavi, Ali Salehzadeh-Yazdi, Ali Mohades, and Ali Masoudi-Nejad, *Computational structure analysis of biomacromolecule complexes by interface geometry*, Computational Biology and Chemistry **47** (2013), 16–23.
26. B.J. McConkey, V. Sobolev, and M. Edelman, *Quantification of protein surfaces, volumes and atom-atom contacts using a constrained Voronoi procedure*, Bioinformatics **18** (2002), 1365–1373.
27. Ray N, Cavier X, Paul JC, and Maignret B, *Intersurf: dynamic interface between proteins*, J Mol Graph Model **23** (2005), 347–354.
28. Hai Nguyen et al., *Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent*, Journal of the American Chemical Society **136**(40).
29. Kliment Olechnovic, Eleonora Kulberkyte, and Ceslovas Venclovas, *Cad-score: A new contact area difference-based function for evaluation of protein structural models*, Proteins **81** (2013), 149–162.
30. Nicolas Rayand, Xavier Cavin, Inria Lorraine, and Bernard Maignret, *Interactive poster: visualizing the interaction between two proteins*.
31. J.F. Sadoc, R. Jullien, and N. Rivier, *The Laguerre polyhedral decomposition: application to protein folds*, The European Physical Journal B **33** (2003), 355–363.
32. Yi Shang and Carlos Simmerling, *Molecular dynamics applied in drug discovery: The case of hiv-1 protease*, Methods in Molecular Biology **819** (2012), 527–549.
33. Alain Soyer, Jaques Chomilier, Jean-Paul Mornon, Remi Jullien, and Jean-Francois Sadoc, *Voronoi tessellation reveals the condensed matter character of folded proteins*, Physical Review Letters **85** (2000), 3532–3535.
34. Chunhu Tan, Yu-Hong Tan, and Ray Luo, *Implicit nonpolar solvent models*, The journal of physical chemistry **111** (2007), 12263–12274.
35. Jorg Weiser, Peter S. Shenkin, and W. Clark Still, *Approximate atomic surfaces from linear combinations of pairwise overlaps (lcpo)*, Journal of Computational Chemistry **20** (1999), 217–230.
36. Bojan Zagrovic and Vijay Pande, *Solvent viscosity dependence of the folding rate of a small protein: distributed computing study*, Journal of Computational Chemistry **24**

32 *Hummel, Yu, Simmerling, Coutsi*

(2003), 1432–1436.

### Appendix A. Inclusion-Exclusion Volume and Surface Area Formulas

The equations in sections 3.2 and 3.3 are extensions of the short inclusion-exclusion formulas for the atomic surface areas and volumes of a molecule discussed in <sup>9,10,23,11</sup>. We use the same notation as in 3.2.

The terms  $S_T$  and  $V_T$  are the surface area and volume respectively of the intersection of the balls in  $T$  (See Figure 21). The surface area and volume of the union of balls are <sup>9,10</sup>,

$$\mathcal{S} = \sum_{\sigma_T \in \partial \mathcal{C}} (-1)^{k+1} c_T S_T, \quad |T| = k \quad (\text{A.1})$$

$$\mathcal{V} = V + \sum_{\sigma_T \in \partial \mathcal{C}} (-1)^{k+1} c_T V_T, \quad |T| = k \quad (\text{A.2})$$

where  $V$  is the volume of the tetrahedra in the alpha complex and the coefficients,  $c_T$ , are given in Section 3.2.

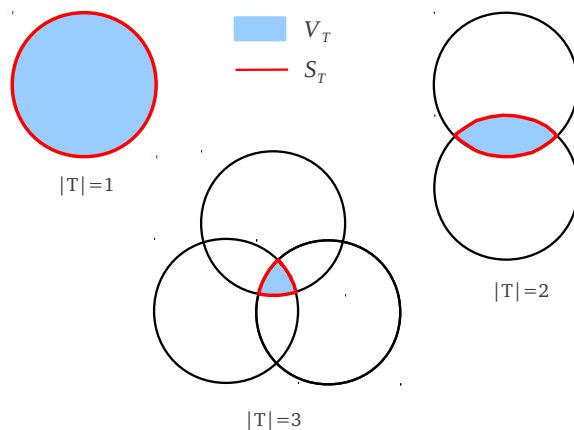


Fig. 21. Volume (blue) and surface area (red) contributions from simplices of dimensions one, two, and three.

The term  $\mathcal{S}$  gives the total surface area of the molecule, but we are also interested in the contribution of an individual atom,  $p_i$ , to the total surface area. Call this term  $\mathcal{S}^i$ . Then

$$\mathcal{S}^i = \sum_{\sigma_T \in \partial \mathcal{C}} (-1)^{k+1} c_T S_T^{(i)}, \quad |T| = k \quad (\text{A.3})$$



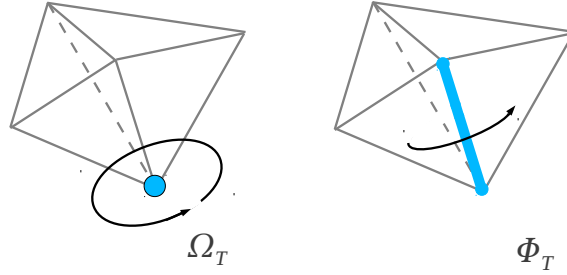


Fig. 22. Fractional outer solid angle and fractional outer dihedral angle.

where  $S_T^{(i)}$  is the contribution of  $S_T$  to  $S^i$  with

$$\sum_{i=1}^n S^i = S. \quad (\text{A.4})$$

These formulas will be given in subsequent sections. The sum A.3 may also be taken over all  $\sigma_T \in \partial C$  such that  $p_i \in T$  since  $S_T^{(i)} = 0$  if  $p_i \notin T$ .

### A.1. Equations

#### A.1.1. $|T| = 1$ : Volume and Surface Area

Consider  $T = \{p_i\}$ . The formulas for the volume and surface area of a ball are

$$V_T = \frac{4}{3}\pi p_i''^{3/2} \quad (\text{A.5})$$

$$S_T = 4\pi p_i'' \quad (\text{A.6})$$

$$S_T^{(i)} = 4\pi p_i''. \quad (\text{A.7})$$

Let  $\mathcal{I}$  be the set of tetrahedra in  $\mathcal{C}$  to which the vertex  $p_i$  is incident. For  $\sigma_I \in \mathcal{I}$  define  $\omega_I$  as the normalized inner solid angle subtended by the tetrahedron  $\sigma_I$  from the point  $p_i'$ . Then

$$\Omega_T = 1 - \sum_{\sigma_I \in \mathcal{I}} \omega_T^I \quad (\text{A.8})$$

$$(\text{A.9})$$

The normalized inner solid angle,  $\omega$ , of a tetrahedron subtended by the vectors

34 Hummel, Yu, Simmerling, Coutsias

$\mathbf{a} = p'_j - p'_i$ ,  $\mathbf{b} = p'_k - p'_i$ , and  $\mathbf{c} = p'_l - p'_i$  (See Figure 23) is given by the equation

$$\omega = \frac{1}{2\pi} \arctan \left( \frac{|\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})|}{abc + (\mathbf{a} \cdot \mathbf{b})c + (\mathbf{a} \cdot \mathbf{c})b + (\mathbf{b} \cdot \mathbf{c})a} \right) \quad (\text{A.10})$$

where  $a = |\mathbf{a}|$  and likewise for  $b$  and  $c$  (See Figure 23).

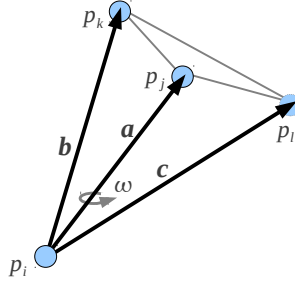


Fig. 23. Normalized inner solid angle,  $\omega$ , subtended by  $\mathbf{a} = p'_j - p'_i$ ,  $\mathbf{b} = p'_k - p'_i$ , and  $\mathbf{c} = p'_l - p'_i$ .

#### A.1.2. $|T| = 2$ : Volume and Surface Area

Consider  $T = \{p_i, p_j\}$ . The formulas for the volume and surface area of the intersection of the two balls,  $p_i$  and  $p_j$  are

$$V_T = \pi h_i^2 \left( \sqrt{p_i''} - \frac{h_i}{3} \right) + \pi h_j^2 \left( \sqrt{p_j''} - \frac{h_j}{3} \right) \quad (\text{A.11})$$

$$S_T^{(i)} = 2\pi \sqrt{p_i''} h_i \quad (\text{A.12})$$

$$S_T^{(j)} = 2\pi \sqrt{p_j''} h_j \quad (\text{A.13})$$

$$S_T = S_T^{(i)} + S_T^{(j)} \quad (\text{A.14})$$

where  $h_i$  and  $h_j$  are the heights of the spherical caps of  $p_i$  and  $p_j$  (See Figure 24).

The characteristic point,  $\mathbf{x} \in \mathbb{R}^3 \times \mathbb{R}$ , of the edge  $\sigma_T$  satisfies

$$\begin{aligned} \Pi(\mathbf{p}_i, \mathbf{x}) &= 0 \\ \Pi(\mathbf{p}_j, \mathbf{x}) &= 0 \\ \mathbf{x}' &= \mathbf{p}'_i + t(\mathbf{p}'_j - \mathbf{p}'_i) \end{aligned} \quad (\text{A.15})$$

for some scalar  $t$  (See Figure 6). Equations A.15 gives

$$p_i'^2 - p_j'^2 - 2\mathbf{x}' \cdot (\mathbf{p}'_j - \mathbf{p}'_i) - p_i'' + p_j'' = 0.$$

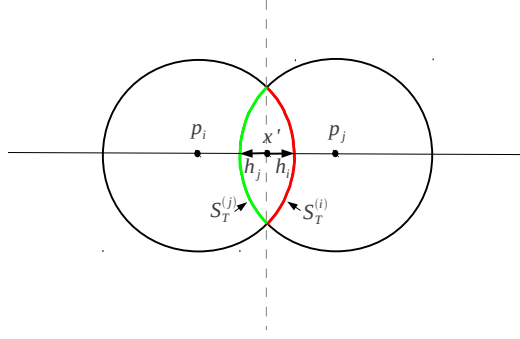


Fig. 24. Heights of the spherical caps and the partition of the surface area  $S_T$  between two atoms.

Define  $k = p_i'' - p_j''$  and  $\mathbf{a} = \mathbf{p}'_j - \mathbf{p}'_i$ . Then

$$\begin{aligned} t &= \frac{k - p_i'^2 + p_j'^2 - 2\mathbf{p}'_i \cdot (\mathbf{p}'_j - \mathbf{p}'_i)}{2a^2} \\ &= \frac{1}{2} \left( \frac{k}{a^2} + 1 \right) \end{aligned} \quad (\text{A.16})$$

and

$$\begin{aligned} h_i &= \sqrt{p_i''} - \text{sgn}(t)|\mathbf{x}' - \mathbf{p}'_i| \\ h_j &= \sqrt{p_j''} - \text{sgn}(1-t)|\mathbf{x}' - \mathbf{p}'_j|. \end{aligned} \quad (\text{A.17})$$

Let  $\mathcal{I}$  be the set of tetrahedra in  $\mathcal{C}$  to which the edge  $\sigma_T$  is incident. For  $\sigma_I \in \mathcal{I}$  define  $\phi_I$  as the normalized inner dihedral angle of  $\sigma_I$  along  $\sigma_T$ . Then

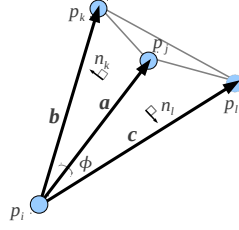
$$\Phi_T = 1 - \sum_{\sigma_I \in \mathcal{I}} \phi_T^I \quad (\text{A.18})$$

The normalized dihedral angle between planes with normals  $\mathbf{n}_k$  and  $\mathbf{n}_l$  is

$$\phi = \frac{\arccos(\mathbf{n}_k \cdot \mathbf{n}_l)}{2\pi}. \quad (\text{A.19})$$

Assume plane  $k$  is defined by the vectors  $\mathbf{a} = \mathbf{p}'_j - \mathbf{p}'_i$  and  $\mathbf{b} = \mathbf{p}'_k - \mathbf{p}'_i$  and the plane  $l$  is defined by the vectors  $\mathbf{a}$  and  $\mathbf{c} = \mathbf{p}'_l - \mathbf{p}'_i$ . Then

$$\begin{aligned} \mathbf{n}_k &= \frac{\mathbf{a} \times \mathbf{b}}{|\mathbf{a} \times \mathbf{b}|} \\ \mathbf{n}_l &= \frac{\mathbf{a} \times \mathbf{c}}{|\mathbf{a} \times \mathbf{c}|}. \end{aligned} \quad (\text{A.20})$$


 Fig. 25. Normalized dihedral angle  $\phi$  between planes with normals  $\mathbf{n}_k$  and  $\mathbf{n}_l$ .

### A.1.3. $|T| = 3$ : Volume and Surface Area

Consider  $T = \{p_i, p_j, p_k\}$ . The volume and surface area of the common intersection of three balls can be written as a weighted sum of the surface area of the single and the double intersections. If  $p_i, p_j$ , and  $p_k$  have a non-empty intersection then there are two points in common with the surfaces of all three balls. Call one of these points  $\mathbf{x}'$ , define  $p_x = (\mathbf{x}', 0) \in \mathbb{R}^3 \times \mathbb{R}$ , and let  $T_x = \{p_i, p_j, p_k, p_x\}$ . Let  $S_2$  be the set of edges defined by  $\sigma_{T_x}$  and  $S_1$  the set of vertices in  $\sigma_{T_x}$ .

The volume and surface areas (See Figure 26) of the intersection of  $p_i, p_j$ , and  $p_k$  are given by <sup>15</sup>

$$\frac{1}{2}V = V_{T_c} + \sum_{\sigma_t \in S_2} \Phi_t V_t - \sum_{\sigma_t \in S_1} \Omega_t V_t \quad k, l = (1, 2, 3) \quad (\text{A.21})$$

$$\frac{1}{2}\mathcal{S}_T^{(i)} = \Phi_{\{i,j\}} S_{\{i,j\}}^{(i)} + \Phi_{\{i,k\}} S_{\{i,k\}}^{(i)} - \Omega_{\{i\}} S_{\{i\}} \quad (\text{A.22})$$

$$\frac{1}{2}\mathcal{S}_T^{(j)} = \Phi_{\{j,k\}} S_{\{j,k\}}^{(j)} + \Phi_{\{j,i\}} S_{\{j,i\}}^{(j)} - \Omega_{\{j\}} S_{\{j\}} \quad (\text{A.23})$$

$$\frac{1}{2}\mathcal{S}_T^{(k)} = \Phi_{\{k,i\}} S_{\{k,i\}}^{(k)} + \Phi_{\{k,j\}} S_{\{k,j\}}^{(k)} - \Omega_{\{k\}} S_{\{k\}} \quad (\text{A.24})$$

$$\frac{1}{2}\mathcal{S}_T = \mathcal{S}_T^{(i)} + \mathcal{S}_T^{(j)} + \mathcal{S}_T^{(k)} \quad (\text{A.25})$$

where  $\Phi_{\{i,j\}}$  is the normalized dihedral angle of  $\sigma_{T_c}$  along the edge  $\sigma_{\{i,j\}}$ ,  $\Omega_i$  is the normalized solid angle of  $\sigma_{T_c}$  subtended from  $p'_i$ , and similarly for other combinations  $i, j$ , and  $k$ .

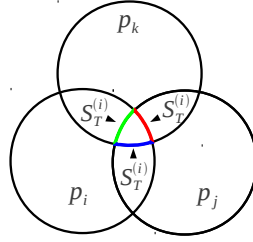
The point  $\mathbf{x}$  satisfies the following equations

$$|\mathbf{p}'_i - \mathbf{x}'|^2 - p''_i = 0 \quad (\text{A.26})$$

$$|\mathbf{p}'_j - \mathbf{x}'|^2 - p''_j = 0 \quad (\text{A.27})$$

$$|\mathbf{p}'_k - \mathbf{x}'|^2 - p''_k = 0 \quad (\text{A.28})$$

Let  $\mathbf{a} = \mathbf{p}'_j - \mathbf{p}'_i$  and  $\mathbf{b} = \mathbf{p}'_k - \mathbf{p}'_i$ , which gives the normal to the plane that contains  $\sigma_T$  as  $\mathbf{n} = \mathbf{a} \times \mathbf{b}$ . The characteristic point,  $\mathbf{x}_c$ , of the triangle  $\sigma_T$  was found in the


 Fig. 26. Partition of  $S_T$  between the three atoms.

computation of the alpha complex and satisfies  $\mathbf{x} = \mathbf{x}_c + h\mathbf{n}$  where  $h$  is the height of  $\mathbf{x}$  above the plane containing  $\sigma_T$ . Plugging this into equation A.26 gives

$$|\mathbf{p}'_i - \mathbf{x}_c|^2 - 2h(\mathbf{p}'_i - \mathbf{x}_c) \cdot \mathbf{n} + h^2 n^2 - p''_i = 0 \quad (\text{A.29})$$

The vector  $(\mathbf{p}'_i - \mathbf{x}_c)$  is orthogonal to  $\mathbf{n}$  which gives

$$h = \frac{\sqrt{p''_i - |\mathbf{p}'_i - \mathbf{x}_c|^2}}{n} = \frac{\sqrt{-\alpha_{\sigma_T}}}{n} \quad (\text{A.30})$$

where  $\alpha_{\sigma_T}$  is the size of the triangle  $\sigma_T$ .

Given a triangle,  $\sigma_T$ , its coefficient is given by

$$c_T = \begin{cases} 1, & \text{if } \sigma_T \text{ is singular} \\ \frac{1}{2} & \text{if } \sigma_T \text{ is regular} \end{cases} \quad (\text{A.31})$$

The triangle,  $\sigma_T$ , transitions from singular to regular when an incident tetrahedron becomes part of the alpha complex.

### A.2. $|T| = 4$ : Volume

Consider  $T = \{p_i, p_j, p_k, p_l\}$ . Let  $\mathbf{a} = p'_j - p'_i$ ,  $\mathbf{b} = p'_k - p'_i$ , and  $\mathbf{c} = p'_l - p'_i$ . The volume of the tetrahedron  $\sigma_T$  is given by

$$V_{tetra} = \frac{|\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})|}{6}. \quad (\text{A.32})$$

**Appendix B. Exterior Facet Area Computation Pseudocode**

```

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!COMPUTE AREA OF LAGUERRE INTERSECTION FACET OF EXTERIOR EDGE
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
!surf      =   area of the Laguerre intersection facet
!num_ring  =   number of nodes in Laguerre intersection facet
!nodes     =   num_ring x 3 matrix
!           =   contains coords of counterclockwise
!                 Laguerre nodes of tetrahedra in alpha complex
!                 with characteristic points of appropriate
!                 boundary triangles inserted
!x         =   center of characteristic point of edge
!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

surf=0
temp_vec=0
a=nodes(2,:)-x
do i=2,num_ring
  b=nodes(i,:)-x
  temp_vec=cross_product(b,a)
  if (dot_product(temp_vec,edge)>=0 ) then
    surf=surf+.5d0*sqrt(temp_vec)
  else
    surf=surf-.5d0*sqrt(temp_vec)
  end if
  a=b
end do

b=nodes(1,:)-x
temp_vec=cross_product(a,b)

if (dot_product(temp_vec,edge)>=0 ) then
  surf=surf+.5d0*sqrt(temp_vec)
else
  surf=surf-.5d0*sqrt(temp_vec)
end if
end if

```

**Appendix C. Glossary**

- $\pi$   
Power distance between weighted point and unweighted point;  $\pi(p, x') = |p' - x'|^2 - p''$ .
- $\sigma_T$   
A simplex which is the convex hull of point centers in set  $T$ .
- A**  
 $\mathcal{A}$   
Set of weighted data points (spheres) which represents atoms in a molecule.
- $\mathcal{A}'$   
Set of data point centers;  $\mathcal{A}' = \{p' \text{ such that } p \in \mathcal{A}\}$ .
- $\mathcal{A}(w)$   
Set of expanded points,  $\mathcal{A}(w) = \{p(w) \text{ such that } p \in \mathcal{A}\}$ .
- B**  
 $\mathcal{B}(w)$   
Space filling model of  $\mathcal{A}(w)$ . Equal to  $\bigcup_{p_i \in \mathcal{A}} B_i(w)$ .
- $B_i(w)$   
Closed ball with center at  $p'_i$  and weight  $p''_i$ .
- C**  
 $\mathcal{C}(w)$   
Alpha complex of data set with  $\alpha = w$ .
- $\partial\mathcal{C}$   
Boundary of the alpha complex  $\mathcal{C}$ .
- E**  
 $e_{ij}$   
Edge connecting the centers of  $p_i$  and  $p_j$ .
- F**  
 $F_T$   
Surface area of the intersection of the Laguerre facet corresponding to simplex  $\sigma_T$  with the interior of the alpha complex.
- L**  
 $\mathcal{L}(\mathcal{A})$   
Laguerre diagram of the data set  $\mathcal{A}$ .
- $L_i$   
Laguerre cell of atom  $i$ .
- $L_i(w)$   
Laguerre cell of  $p_i(w)$  in  $\mathcal{A}(w)$ .
- $L_{ij}$   
Laguerre facet corresponding to edge  $e_{ij}$ .
- $LI(w)$   
Set of Laguerre-Intersection cells.
- $LI_i(w)$

40 Hummel, Yu, Simmerling, Coutsias

Laguerre-Intersection cell of atom  $i$  with weight  $w$ .

$LIS_i(w)$

Surface area of  $LI_i(w)$ .

$LIS_T^{(i)}$

Contribution of the simplex  $\sigma_T$  to  $LIS_i$ .

$LIV_i(w)$

Volume of  $LI_i(w)$ .

$LIV_T^{(i)}$

Contribution of the simplex  $\sigma_T$  to  $LIV_i$ .

$LIV_i^{(i)}$

Equivalent to  $LIV_T^{(i)}$  for  $T = \{p_i\}$ .

$LIV_{ij}^{(i)}$

Equivalent to  $LIV_T^{(i)}$  for  $T = \{p_i, p_j\}$ .

$LIV_{ijk}^{(i)}$

Equivalent to  $LIV_T^{(i)}$  for  $T = \{p_i, p_j, p_k\}$ .

$LS_i$

Surface area of the Laguerre cell of atom  $i$ .

$LS_{ij}$

Surface area of facet  $L_{ij}$ .

$LV_i$

Volume of the Laguerre cell of atom  $i$ .

$LV_{ij}^i$

Volume contribution of the edge  $e_{ij}$  to the Laguerre cell of atom  $i$ .

**P**

$p = (p', p'')$

Weighted point (sphere) in  $\mathbb{R}^3 \times \mathbb{R}$  with center  $p'$  and weight (radius squared)

$p''$ .

$p'$

Unweighted point or location in  $\mathbb{R}^3$ .

$p''$

Weight or radius squared of point  $p$ . This can also be written as  $w$ .

$p(w)$

Expanded point  $p(w) = (p', p'' + w)$

$P_T^{(i)}$

Pyramidal volume contribution of  $\sigma_T$  to  $LIV_i$ .

**S**

$S_i(w)$

Accessible surface area of atom  $i$  with solvent weight  $w$ .

$S_T$

Surface area of the intersection of balls represented by points in  $T$ .

$S_i^{(i)}$

Equivalent to  $S_T^{(i)}$  where  $T = \{p_i\}$ .



- $S_T^{(i)}$  The contribution of  $S_T$  to  $S_i$ .
- T**
- $\mathcal{T}(\mathcal{A})$  Weighted Delaunay (Regular) tetrahedrization of the data set  $\mathcal{A}$ .
- $T$  Set of points corresponding to simplex in  $\mathcal{A}$ .
- $|T|$  Number of elements in the set  $T$ .
- $t_{ijk}$  A triangle with vertices at  $p'_i, p'_j, p'_k$ .
- V**
- $v_i$  Vertex.
- $V_T$  Volume of the intersection of balls represented by points in  $T$ .
- W**
- w** Weight or radius squared of a data point
- X**
- $x_{ij}$  Characteristic point of the edge  $e_{ij}$ .
- $x_{ijk}$  Characteristic point of the triangle  $t_{ijk}$ .
- $x_T$  Characteristic point of the simplex  $\sigma_T$ .  $x_T = (x'_T, x''_T)$  where  $x'_T$  is the center of the characteristic point and the weight  $x''_T$  is the size of the simplex.